The induction contribution to the lattice energy of organic crystals

Gareth W. A. Welch

2010

This thesis is submitted to UCL in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Acknowledgements

I would like to thank my supervisor, Professor Sally Price, for her support during my research, and especially during the writing up phase. It has been her guidance and leadership which has steered my Ph.D. from its beginnings to its end.

I also wish to thank Dr. Maurice Leslie, whose timely updating of DMACRYS from FORTRAN77 to FORTRAN90 and early retirement during this project have been both a help and a hindrance. Without his unique knowledge of the code and visits to London, the adaptations I have made would never have been achievable. The discussion and support I have received from Dr. Panos Karamertzanis has been absolutely invaluable, especially with regards to programming but also his scientific rigor. Likewise, Professor Anthony Stone and Dr. Alston Misquitta have not only developed the theories which I have used, but have provided innumerable hours of discussion and advice. It has been a pleasure to be part of the CPOSS group, and I am proud to have co-authored papers with many of them.

Declaration

I, Gareth William Alan Welch, confirm the work presented in this thesis is my own. Where information has been derived from other sources, I confirm this has been indicated in the thesis.

Abstract

A recently developed method for generating distributed, localized atomic polarizabilities from the *ab initio* molecular charge density is used to assess the importance of the induction energy in crystal structures of small organic molecules. Two models are first contrasted based on large cluster representing the crystalline environment: one using the polarizability model in which induced multipoles are evaluated in response to the electrostatic field due to atomic multipoles; the other is a complementary procedure in which the same cluster is represented by atomic point-charges and the molecular charge density is calculated *ab initio* in this environment. The comparable results of these two methods show that the contribution to the lattice energy from the induction term can differ significantly between polymorphic forms, for a selection of organic crystal structures including carbamazepine and oxalyl dihydrazide, and 3-azabicyclo[3,3,1]nonane-2,4-dione. The observed charge density polarization of naphthalene in the crystalline state is also reproduced.

This demonstrates that explicit inclusion of the induction energy, rather than its absorption into an empirically fitted repulsion-dispersion potential, will improve the relative ordering of the lattice energies for computed structures, and that it needs to be included in crystal structure prediction. Hence, the distributed atomic polarizability model was coded into the lattice-energy minimization program DMACRYS (which was developed as a Fortran90 recoding of DMAREL) to allow the induction energy to be calculated.

Publications

Main work

Is the induction energy important for modeling organic crystals? Welch, G. W. A.; Karamertzanis, P. G.; Misquitta, A. J.; Stone, A. J.; Price, S. L. *J. Chem. Theory Comput.* **2008,** *4* (3), 522-532

Modelling organic crystal structures using distributed multipole and polarizability models intermolecular potentials Price, S. L.; Leslie, M., Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M, *Phys. Chem. Chem. Phys.* **2010**, **12**, 8478-8490

Other contributions to the CSP literature

Modeling the interplay of inter- and intra-molecular hydrogen bonding in conformational polymorphs. Karamertzanis, P. G.; Day, G. M.; Welch, G. W. A.; Kendrick, J.; Leusen, F. J. J.; Neumann, M. A.; Price, S. L. *J. Chem. Phys.* **2008,** *128*

Modelling the induction energy for seven polymorphs of two compounds which exhibit changes in the inter- and intra-molecular hydrogen bonding configuration between polymorphs. Contributions to methodology, analysis and electrostatic potential plots.

Significant progress in predicting the crystal structures of small organic molecules - a report on the fourth blind test. Day, G. M.; Cooper, T. G.; Cruz Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A.; Schweizer, B. *Acta Crystallogr. , Sect. B* **2009,** *65* (2), 107-125.

Significantly involved in the Price group search methodology for all four blind test molecules. As well as performing calculations and handing the large search with custom potential for XIII (see chapter X), I oversaw and automated searches and performed reminimizations for large sets of crystal structures (20000+) with different potential models, as well providing general computational and scientific expertise and support.

A first principles solution of the crystal structure of C6Br2ClFH2. Misquitta, A. J.; Welch, G. W. A.; Stone, A. J.; Price, S. L. *Chem. Phys. Lett.* **2008,** *456* (1-3), 105-109

Specifically relating to the custom potential derived for XIII from the CCDC Blind Test, I provided scripting solutions and performed reminimizations on a large set of structures using custom potentials while they were still being developed against the CCDC deadline.

Can the Formation of Pharmaceutical Cocrystals Be Computationally Predicted? I. Comparison of Lattice Energies. Issa, N.; Karamertzanis, P. G.; Welch, G. W. A.; Price, S. L. *Cryst. Growth Des.* **2009,** *9* (1), 442-453

Can the Formation of Pharmaceutical Co-Crystals Be Computationally Predicted? II. Crystal Structure Prediction. Karamertzanis, P. G.; Kazantsev, A. V.; Issa, N.; Welch, G. W. A.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L. *J. Chem. Theory Comput.* **2009,** *5* (5), 1432-1448

I assisted Nizar Issa in preparing and conducting co-crystal searches, especially in relation to caffeine (which is disordered). As well as proving directly to the scientific content, this publication is the result of extensive training and support that I have provided to the author, in terms of both science and computer/computational issues.

Search for a predicted hydrogen bonding motif - A multidisciplinary investigation into the polymorphism of 3-azabicyclo[3.3.1]nonane-2,4-dione. Hulme, A. T.; Johnston, A.; Florence, A. J.; Fernandes, P.; Shankland, K.; Bedford, C. T.; Welch, G. W. A.; Sadiq, G.; Haynes, D. A.; Motherwell, W. D. S.; Tocher, D. A.; Price, S. L. *J. Am. Chem. Soc.* **2007,** *129* (12), 3649-3657

Following a universally unsuccessful prediction of 'BUQQOT' in a previous Blind Test a thorough study was undertaken to identify polymorphs that exhibited the 'dimer' configuration that

was predicted, instead of the 'catemer' configuration that was observed. Using my expertise with ORIENT *I demonstrated that the dimers were easily broken by the presence of another molecule to form a chain structure, and that the electrostatic potential around the hydrogen-bond donor sites was actually much weaker than usually expected. By employing higher level methods than were standard to the group at the time, I provided a line of argument that justified why the predictions had focused on the wrong structural motif.*

Unrelated publications

Wavepacket dynamics study of Cr(CO)(5) after formation by photodissociation: relaxation through an (E circle plus A) circle times e Jahn-Teller conical intersection. Worth, G. A.; Welch, G.; Paterson, M. J. *Mol. Phys.* **2006,** *104* (5-7), 1095-1105

Following a study by Robb et al. that identified the relaxation through a conical intersection of chromium pentacarbonyl by means of a second-order Jahn-Teller mechanism, I implemented a fitting algorithm as well as a new exp-6 potential form to compliment the harmonic and quadratic potential. Using these forms I fitted three adiabatic potential energy surfaces to a series of data points over 5 coupled vibrational modes, which were subsequently used to a perform wavepacket dynamics study that proved the relaxation mechanism was in fact through a pseudo-Jahn-Teller mode

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Motivation: Computer Modeling of the Organic Solid state

The study and modelling of condensed-phase materials is a broad area of research with far reaching applications; from understanding basic physical and chemical processes, to simulating complex systems such as protein docking, to designing materials at the molecular level for highly specific purposes. A challenging subset of this problem is the field of predicting the crystal structures of organic molecules.[1] Crystal structure prediction (CSP) aims to predict both the observed and as yet unobserved polymorphs of molecules, in order to identify systems where a thermodynamically more stable structure may appear to compete with metastable crystal in production or storage. This is very important in the pharmaceutical industry because the dissolution rate, and hence bioavailability of a drug, will vary with polymorph[2]. CSP may also be used to computationally screen molecules for favourable solid-state properties, such as density of energetic materials[3], prior to their synthesis.

There are a variety of applications for the ability to correctly predict crystal structures of small organic molecules. Sometimes there is not enough information available to solve a crystal structure, either because the single crystal data available is poor, or even completely unavailable if a suitable crystal cannot be grown. CSP can provide possible structures to guide the solution of the structure from powder X-ray data. In other cases, identifying that polymorphs exist with a predictive search can guide an experimental search for new polymorphs, by suggesting which crystallization method might nucleate the first sample[4]. Competing crystal forms can be a serious problem for pharmaceuticals, where drugs are licensed to contain a specific polymorph as this determines the bioavailability and other

physical properties. The anti-HIV drug Ritonavir is a perfect example of a very expensive, high profile polymorphism problem.[5] A more stable polymorph suddenly appeared when the drug was in production, and since they were no longer able to reliably manufacture the licensed pharmaceutical form, the drug had to be reformulated. A computational search might, in principle, have suggested the more stable structure during the drug development phase. This could have helped the search for this polymorph; the appearance of form II was linked to the presence of a degradation product as a solid impurity[5]. Hence, CSP search is a very useful complement to the solid state characterization that is needed for developing any organic material[1].

1.2. Progress in Crystal Structure Prediction

By starting with a simple molecular diagram, CSP can produce a number of possible crystal structures by minimizing the lattice energy of thousands of trial configurations. If we assume that the lattice energy predicts the relative thermodynamic stability, the point with the lowest lattice energy (global minimum) will correspond to the most stable observed crystal form. Any predicted structure within 5 kJ mol^{-1} of the global minimum is sufficiently close in energy that it may represent an observable polymorph. In practice this lattice energy landscape contains clusters of structures close to the global minimum that do not represent experimental structures, and sometimes the global minimum does not correspond to the known most stable form.

CSP can be successful at predicting crystal structures, as shown by the successes in the Cambridge Crystallographic Data Centre (CCDC) blind tests.[6-9] The most recent[9] blind test was particularly successful with each structure being correctly predicted by at least two groups, while one group was able to correctly predict all four crystal structures[10]. It is hard

to assess the difficulty of predicting a given crystal structure from the molecular diagram, as crystal energy landscapes can differ in the number and diversity of packing motifs within the energy range of possible polymorphism, even between isomers[11].

Two cases where the prediction of possible polymorphs with very different hydrogen-bonding motifs lead to a coordinated experimental search are carbamazepine[12] and 3-azabicyclo[3.3.1]nonane-2,4-dione.[13] The known stable forms of these molecules are comprised of dimer units held by two hydrogen-bonds in the first case, and a catemeric (chain-like) structure in the latter. The first crystal structure predictions were that the thermodynamic form of each would be comprised of catemers and dimers respectively, but these were not found experimentally. Thus it is important to establish why these predicted structures were not found, and this thesis contributes to showing that the relative stabilities were insufficiently accurate in the initial predictions.

1.3. Approaches Needed to Improve Crystal Structure Prediction

There are many factors that should be considered when predicting which crystal structures will be observed and which will not. Kinetics, and particularly nucleation, is a significant issue as it is believed that the structure of the initial crystalline nuclei in solution will determine the bulk structure. Competing rates of nucleation for different forms should be a significant factor, which brings with it the issue of kinetics and solvent effects for micro and nanoscopic particles. Since most CSP methodologies start with trial periodic lattices which are subsequently relaxed, the question of whether a given lattice energy minimum could actually be formed by some physical process is ignored.

Even if the thermodynamically stable structure can be crystallized, current CSP methods based on lattice energy are only predicting the most stable static perfect crystal at 0 K. Factors such as entropy and zero-point motion at crystallization temperature are completely neglected[14]. Accurately modelling the thermodynamic free energy would surely increase the reliability of CSP, but it is challenging and computationally expensive to implement.[15,16] Any model for these terms relies on the accuracy of the model for intermolecular interactions, which is used to calculate the lattice energy.

1.4. Development of More Accurate Modelling of the Lattice Energies of Organic Molecules

Since the method of identifying observable polymorphs is relying on correctly predicting the relative lattice energies of the hypothetical structures, accurate models for the intermolecular interactions are essential. A successful prediction may rely on differences of less than a kJ mol^{-1} between very different crystal structures. Some advances have been made in the use of periodic electronic structure calculations to model organic crystal structures quantum mechanically[10,17] , although these periodic DFT calculations still require a corrective dispersion term, and these calculations are very expensive. However, most crystal structure modelling relies on a model intermolecular potential i.e. an analytical model for the intermolecular forces.

Model intermolecular potentials usually explicitly model the electrostatic interactions and the short-range repulsive and long-range attractive dispersion forces. The electrostatic interactions are either modelled by atomic point charges or atomic multipole moments, which have been derived from *ab initio* calculations of the charge density of the molecule in isolation. Usually the model for all the remaining intermolecular interactions is derived by

empirical fitting to reproduce a set of organic crystal structures and other properties, such as heats of sublimation. [18-24] Although these empirical potentials are usually atom-atom repulsion-dispersion potentials of the exp-6 functional form, assumed to be transferable between different molecules, their accuracy is limited. Most critically the empirical fitting will attempt to absorb all the approximations in the calculations, such as the neglect of other contributions to the intermolecular forces, thermal expansion, etc etc.[14] The time spent in parameterising these models is recovered by their transferability and the speed at which they can be used in a periodic crystal lattice, or other large systems such as biological processes where there may be hundreds of thousands of interactions to compute. If the model intermolecular potential can be accurately determined for a specific molecule rather than being generic to many systems, it would be much more accurate than the current empirical models.

In addition to the repulsions, dispersion and electrostatic intermolecular forces, there is a contribution due to the induction energy. The induction energy comes from the polarization of the molecular charge density as it distorts from the ground-state configuration of the isolated molecule in response to the surroundings, i.e. the presence of other molecular charge densities nearby. This is difficult to include in model potentials which have the assumption of pairwise additivity; i.e. that the interaction of a many-body system is simply the sum of the interaction between all pairs of particles. The electrostatic energy is strictly pairwise additive, but the induction energy is not because the fields due to the neighbouring molecules can cancel to a very large extent. At present the induction energy is absorbed into the empirical repulsion-dispersion potential, as are all non-electrostatic terms (and indeed all electrostatic terms that are not adequately modelled by the point-charges or multipoles with which the parameters were determined). This thesis

aims to improve the accuracy of the calculation of the crystal lattice energies by studying the induction energy in organic crystals.

1.5. Current Models for the Induction Energy

Polarization effects have been modelled for ionic systems of spherical ions for a long time, where the strong electric fields mean that polarization effects are significant. This has been achieved usually by the use of the shell model,[25] which treats the ion as two points connected by a spring. In this model the stiffness of the spring is in effect an isotropic polarizability tensor that determines how far the outer 'shell' site moves from the fixed nucleus site, in order to describe an induced dipole. This simple model has been widely used in the modelling of ionic systems[25,26], and in recent times has been implemented into force-fields for use in modelling water and biological systems[27,28].

Although the shell model has been used very successfully for a range of systems, it has two main weaknesses. Firstly, it is unrealistic to describe the polarizability of the charge density in the region of a covalently bonded atom in an isotropic manner. Charge density will tend to move along bond axes and to regions where it is stabilised, such as delocalised π-systems or electronegative atoms, and is less able to move perpendicular to the bonding plane. This anisotropy will be influenced by the specific chemical environment, which brings forward the second weakness of this model. If the assumption that atomic polarizability could be modelled isotropically were sufficient, by what means could the spring constants be determined? The fitting of empirical potentials to organic structures has runs into problems of having ill-determined parameters. They may be fitted to experimental observations either for specific systems or transferred from a more general dataset, in

conjunction with some repulsion-dispersion model potential, and are rarely derived computationally to explicitly describe the interaction.

A popular model of polarizability is the Applequist[29] model, which originally used isotropic atomic polarizabilities that are semi-empirical and derived from a combination of experimental and fitted values. This model can also be used with anisotropic polarizabilities. However, it does suffer from several problems, particularly where it treats atoms within a molecule as distinct particles which interact with one another through multipole interactions. Bonded atoms are separated by such a short distances that the multipole expansion is invalid, and must be corrected for by some empirical means.

Although polarizability models such as these have been used with some degree of success for systems of ions and the smallest of molecules, it has largely been ignored in larger simulations such as modelling crystal lattices. This is partly due to the difficultly in accurately modelling the polarizability across a molecule. The error of using simple models in complex systems like organic crystal structures could be greater than the differences in lattice energy, making it more accurate to neglect it. The use of empirically fitted potentials means that induction energy is already absorbed into the parameters that define the model, in an average sense. As long as these potentials are used, any model of the induction energy will include an unknown amount of double counting. There is also a high computational expense in dealing with the polarization effect, since each site is polarized by the field due to surrounding sites, which changes the fields within the structure. This in turn requires the interactions to be recalculated, and so on until they converge.

An improved model would allow for anisotropically polarizable sites, be derived directly from the molecular properties rather than fitted to available experimental data, and allow

structures to be relaxed with polarization. A method has been published[30-32] which allows for anisotropic polarizability terms to be derived directly from a molecular properties calculation, and may be used with a distributed multipole model. In addition, a scheme has been developed[33,34] that allows a repulsion-dispersion model to be derived that is specific to a system and excludes induction. This would mean that the entire intermolecular interaction would be calculated with terms derived from the *ab initio* wavefunction. The method is described briefly in section 2.4 and full details can be found in the referenced material. Thus we now have a method of calculating distributed polarizabilities that could be used in crystal structure modelling for organic molecules.

An evaluation of the importance of the induction energy has been made by the 'Pixel' method[35]. The semi-classical density sums SCDS-Pixel method uses numerical integration over a crystal structure, in which the molecules are represented by the *ab initio* charge density of the isolated molecule. The volume of the crystal structure is divided into units, 'pixels', to which are assigned properties relating to nearby atoms. Charge density is allocated to these units, which interact to give the exact electrostatic energy,[36] and polarizability may be modelled by distributing the atomic polarizability over pixels with charge density associated with that atom.[37] Difficulty still comes from determining the atomic polarizability, and how it ought to be distributed. The SCDS-Pixel method allows the intermolecular interaction energies to be separated into individual components, and has shown that induction is a significant contribution to the lattice energy in many organic crystals. Although this method does offer accurate lattice energies, it does not readily allow structural relaxation to minimize the lattice energy. Hence, it is timely to use distributed polarizability models to confirm that induction energy could be a significant discriminator between different organic crystal structures. Distributed polarizability models could then be implemented into lattice energy minimization codes.

1.6. The project

This PhD has been part of a collaborative project with Dr. Alston Misquitta at Cambridge. Under the guidance of Professor Anthony Stone, Dr. Misquitta has developed a method using SAPT(DFT) and Coupled Kohn-Sham theory[38-40] to calculate frequency-dependent distributed polarizabilities[30,31,41,42] for small organic molecules. As stated in section 2.2.4 these polarizabilities may also be used to evaluate the dispersion interaction[42]. Part of the project has been to facilitate the production of completely *ab initio* potentials where all terms are derived from the molecular charge density[33,43]. My role has been to assess the use of distributed polarizability models for organic crystals structure prediction. This required modifying the DMACRYS code to include the distributed polarizabilities, and so calculate the induction energy of a crystal lattice.

The main thrust of this thesis is to investigate the effects of including the induction energy in crystal structure prediction, and to implement the model into the lattice energy minimizer DMACRYS. Chapter 2 outlines the background of calculating interaction energies for small organic molecules that are the target of CSP, and by describing the polarizability model of Misquitta and Stone (WSM model). Chapter 2 also includes a summary of the codes used to research and implement the WSM model. Chapter 3 demonstrates the effect of using a high quality wavefunction, such as that used to determine the polarizability model, on the electrostatic energy and therefore relative lattice energy of putative crystal structures. An improved method[44] of determining distributed multipoles from wavefunctions that include diffuse Gaussian functions is tested, and leads to a change to the default parameters of the distribution algorithm. Chapter 4 contains details of testing the distributed polarizability model by using clusters of molecules to represent the crystal

environment, and compares the energies from these clusters using the distributed anisotropic polarizability model with an *ab initio* calculation of a molecular charge density surrounded by point charges. The agreement between these two different models gives us confidence that the polarizability model can represent the charge density polarization of an anisotropic field in a cluster of molecules and is worthwhile implementing into DMACRYS. I use a cluster model to study oxalyl dihydrazide, naphthalene and carbamazepine, and demonstrate how including the induction energy can improve the relative rankings of observed polymorphs and reproduce observed charge density polarization. In chapter 5, I describe how the rank 1 WSM model is coded in to DMACRYS and validate it against the cluster models. Chapter 6 presents work where minima from a search for a polyhalogenated compound used in the most recent blind test, $C_6Br_2ClFH_2$, are treated using the implemented polarization model with a custom *ab initio* derived[43] repulsion-dispersion potential from which induction effects have been excluded. This leads in to chapter 7, which discusses the further research necessary to improve the devised induction model, and to use it while relaxing crystal structures.

Reference List

1. Price, S. L. Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism. *Accounts Chem. Res.* **2009,** *42* (1), 117-126.

2. Price, S. L. The computational prediction of pharmaceutical crystal structures and polymorphism. *Adv. Drug Deliver. Rev.* **2004,** *56* (3), 301-319.

3. Bernstein, J. *Polymorphism in Molecular Crystals;* Clarendon Press: Oxford, 2002.

4. Llinas, A.; Goodman, J. M. Polymorph Control: past, present and future. *Drug Discover. Today* **2008,** *8* (5/6), 198-210.

5. Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. Ritonavir: An extraordinary example of conformational polymorphism. *Pharmaceut. Res.* **2001,** *18* (6), 859-866.

6. Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooij, W. T. M.; Price, S. L.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. A test of crystal structure prediction of small organic molecules. *Acta Crystallogr. , Sect. B* **2000,** *56*, 697-714.

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

8. Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P. A third blind test of crystal structure prediction. *Acta Crystallogr. , Sect. B* **2005,** *61* (5), 511-527.

9. Day, G. M.; Cooper, T. G.; Cruz Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A.; Schweizer, B. Significant progress in predicting the crystal structures of small organic molecules - a report on the fourth blind test. *Acta Crystallogr. , Sect. B* **2009,** *65* (2), 107-125.

10. Neumann, M. A.; Leusen, F. J. J.; Kendrick, J. A Major Advance in Crystal Structure Prediction. *Angew. Chem. ,Int. Ed.* **2008,** *47* (13), 2427-2430.

11. Barnett, S. A.; Johnson, A.; Florence, A. J.; Price, S. L.; Tocher, D. A. A systematic experimental and theoretical study of the crystalline state of six chloronitrobenzenes. *Cryst. Growth Des.* **2008,** *8* (1), 24-36.

12. Florence, A. J.; Johnston, A.; Price, S. L.; Nowell, H.; Kennedy, A. R.; Shankland, N. An automated parallel crystallisation search for predicted crystal structures and packing motifs of carbamazepine. *J. Pharm. Sci.* **2006,** *95* (9), 1918-1930.

13. Hulme, A. T.; Johnston, A.; Florence, A. J.; Fernandes, P.; Shankland, K.; Bedford, C. T.; Welch, G. W. A.; Sadiq, G.; Haynes, D. A.; Motherwell, W. D. S.; Tocher, D. A.; Price, S. L. Search for a predicted hydrogen bonding motif - A multidisciplinary investigation into the polymorphism of 3-azabicyclo[3.3.1]nonane-2,4-dione. *J. Am. Chem. Soc.* **2007,** *129* (12), 3649-3657.

14. Price, S. L. Quantifying intermolecular interactions and their use in computational crystal structure prediction. *CrystEngComm* **2004,** *6*, 344-353.

15. Day, G. M.; Price, S. L.; Leslie, M. Atomistic calculations of phonon frequencies and thermodynamic quantities for crystals of rigid organic molecules. *J. Phys. Chem. B* **2003,** *107* (39), 10919-10933.

16. Karamertzanis, P. G.; Raiteri, P.; Parrinello, M.; Leslie, M.; Price, S. L. The Thermal Stability of Lattice Energy Minima of 5-Fluorouracil: Metadynamics as an Aid to Polymorph Prediction. *J. Phys. Chem. B* **2008,** *112* (14), 4298-4308.

17. Karamertzanis, P. G.; Day, G. M.; Welch, G. W. A.; Kendrick, J.; Leusen, F. J. J.; Neumann, M. A.; Price, S. L. Modeling the interplay of inter- and intramolecular hydrogen bonding in conformational polymorphs. *J. Chem. Phys.* **2008,** *128* (24), art-244708.

18. Williams, D. E. Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals. *J. Comput. Chem.* **2001,** *22* (11), 1154-1166.

19. Williams, D. E. Improved intermolecular force field for crystalline oxohydrocarbons including O-H...O hydrogen bonding. *J. Comput. Chem.* **2001,** *22* (1), 1-20.

20. Williams, D. E.; Houpt, D. J. Fluorine Nonbonded Potential Parameters Derived From Crystalline Perfluorocarbons. *Acta Crystallogr. , Sect. B* **1986,** *42* (JUN), 286-295.

21. Williams, D. E. Nonbonded Interatomic Potential-Energy Functions and Prediction of Crystal-Structures. *Acta Crystallogr. , Sect. A.* **1984,** *40*, C95.

22. Williams, D. E.; Cox, S. R. Nonbonded Potentials For Azahydrocarbons: the Importance of the Coulombic Interaction. *Acta Crystallogr. , Sect. B* **1984,** *40* (8), 404-417.

23. Cox, S. R.; Hsu, L. Y.; Williams, D. E. Nonbonded Potential Function Models for Crystalline Oxohydrocarbons. *Acta Crystallogr. , Sect A.* **1981,** *37* (MAY), 293-301.

24. Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *J. Phys. Chem.* **1996,** *100* (18), 7352-7360.

25. Catlow, C. R. A.; Norgett, M. J. Shell-Model Calculations of Energies of Formation of Point-Defects in Alkaline-Earth Fluorides. *J. Phys. C Solid State* **1973,** *6* (8), 1325-1339.

26. Lindan, P. J. D.; Gillan, M. J. Shell-Model Molecular-Dynamics Simulation of Superionic Conduction in Caf2. *J. Phys,-Condens. Mat.* **1993,** *5* (8), 1019-1030.

27. Crowe, R. W.; Santry, D. P. Polarization and Dipole-Moment of Lattice Water in Hydrogen-Bonded Crystals. *Chem. Phys. Lett.* **1973,** *22* (1), 52-55.

28. Alfredsson, M.; Brodholt, J. P.; Hermanson, K.; Vallauri, R. The use of a point polarizable dipole in intermolecular potentials for water. *Mol. Phys.* **1998,** *94* (5), 873-876.

29. Applequist, J.; Carl, J. R.; Fung, K. K. Atom Dipole Interaction Model for Molecular Polarizability - Application to Polyatomic-Molecules and Determination of

Atom Polarizabilities. *Journal of the American Chemical Society* **1972,** *94* (9), 2952-&.

30. Misquitta, A. J.; Stone, A. J. Distributed polarizabilities obtained using a constrained density-fitting algorithm. *J. Chem. Phys.* **2006,** *124* (2), 024111.

31. Misquitta, A. J.; Stone, A. J. Accurate induction energies for small organic molecules: 1. Theory. *J. Chem. Theory Comput.* **2008,** *4* (1), 7-18.

32. Misquitta, A. J.; Stone, A. J.; Price, S. L. Accurate induction energies for small organic molecules. 2. Development and testing of distributed polarizability models against SAPT(DFT) energies. *J. Chem. Theory Comput.* **2008,** *4* (1), 19-32.

33. Stone, A. J.; Misquitta, A. J. Atom-atom potentials from ab initio calculations. *Int. Rev. Phys. Chem.* **2007,** *26* (1), 193-222.

34. *CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form,* University of Cambridge (2007), http://www-stone.ch.cam.ac.uk/programs.html#CamCASP: 2007

35. Gavezzotti, A. Calculation of lattice energies of organic crystals: the PIXEL integration method in comparison with more traditional methods. *Z. Kristallogr.* **2005,** *220* (5-6), 499-510.

36. Gavezzotti, A. Calculation of intermolecular interaction energies by direct numerical integration over electron densities. I. Electrostatic and polarization energies in molecular crystals. *J. Phys. Chem. B* **2002,** *106* (16), 4145-4154.

37. Gavezzotti, A. Calculation of intermolecular interaction energies by direct numerical integration over electron densities. 2. An improved polarization model and

the evaluation of dispersion and repulsion energies. *J. Phys. Chem. B* **2003,** *107* (10), 2344-2353.

38. Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *J. Chem. Phys.* **2005,** *123* (21), 214103.

39. Misquitta, A. J. Interaction energies from a symmetry-adapted perturbation theory based on density functional theory. *Abstracts of Papers of the American Chemical Society* **2004,** *228*, U201.

40. Misquitta, A. J.; Szalewicz, K. Intermolecular forces from asymptotically corrected density functional description of monomers. *Chem. Phys. Lett.* **2002,** *357* (3-4), 301-306.

41. Misquitta, A. J.; Szalewicz, K. Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers. *J. Chem. Phys.* **2005,** *122* (21), art-214103.

42. Misquitta, A. J.; Stone, A. J. Dispersion energies for small organic molecules: first row atoms. *Mol. Phys.* **2008,** *106* (12-13), 1631-1643.

43. Misquitta, A. J.; Welch, G. W. A.; Stone, A. J.; Price, S. L. A first principles solution of the crystal structure of C6Br2ClFH2. *Chem. Phys. Lett.* **2008,** *456* (1-3), 105-109.

44. Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005,** *1* (6), 1128-1132.

Chapter 2. Theory of Intermolecular Forces

The computational study of intermolecular interactions brings with it the opportunity to gain new insights into the physical world and the theories that attempt to describe it. It also allows us to perform experiments using computers that will save huge efforts in practical experiments to achieve. Intermolecular interaction models are used to simulate systems such as liquids, nanotubes and protein docking. For crystal structure prediction, intermolecular interactions are responsible for the thermodynamic stability of the lattice, growth rates, mechanical stability and other properties of interest. If we can accurately and reliably predict the properties of crystal structures, we can identify which potential pharmaceuticals will be difficult to crystallize or which may suffer from converting to polymorph with different bioavailability. The quality of the simulation is determined by how accurately the interaction energies are modelled, but these interactions have anisotropic and quantum mechanical components to them, some of which are not pairwise additive, so the number of computations required may increase exponentially with system size. Any simulation must be sufficiently large for meaningful results: molecular dynamics in a biological system may require molecules with hundreds of atoms interacting in a periodic system with tens of thousands of water molecules. This makes very high accuracy in all areas very expensive in terms of CPU time, and so approximations are made which reduce the simulation time at the expense of accuracy. For crystal structure prediction, we make the approximation of an infinite lattice, but an interaction radius of 15 Å is usually sufficient for the electrostatic and repulsion-dispersion interactions. This is combined with a Ewald sum for a periodic lattice, which efficiently accounts for the long-range electrostatic interactions. Often, models are created for a specific type of simulation, e.g.

enzyme docking or modelling ice, such that they are accurate for certain properties but poorly reproduce others, and so a careful selection must be made when designing a simulation.

This chapter describes the intermolecular interaction models used throughout this thesis. The quantum mechanical origins of the energy are outlined in section 2.2, and a formulation for the polarizability tensors in sections 2.2.3 and 2.5. This is followed by a description of the model potentials used and, where relevant, how they are derived from their quantum mechanical origins. This is done to establish the spherical tensor formulation for the models for electrostatic fields and polarizabilities, described by Stone[1] and used in this thesis to model the induction energy. Finally, the programs in which the models are implemented and used for work in this thesis are described in section 2.6.

2.1. Modelling Intermolecular Forces

For an n-body system the interaction between the bodies comprises of many two-body terms, plus three- and four- up to n-body terms. The many-body expansion of the intermolecular potential may be written[1]

$$U(\mathbf{R}_1,\mathbf{R}_2,\cdots,\mathbf{R}_n)=\sum_{i<j}U^{(2)}(\mathbf{R}_i,\mathbf{R}_j)+\sum_{i<j<k}U^{(3)}(\mathbf{R}_i,\mathbf{R}_j,\mathbf{R}_k)+\cdots + \sum_{i<j<\cdots<n}U^{(n)}(\mathbf{R}_i,\mathbf{R}_j,\cdots\mathbf{R}_n). \quad (1)$$

In this expression, $\mathbf{R}_i$ is the vector describing the position and orientation of molecule *i* and $U^{(n)}$ is the n-body interaction potential of the system. This expansion is exact, but it is not feasible to calculate all the terms and is only practical if it can be truncated at low order. Indeed, most modelling truncates at second-order, making the pairwise additive approximation, e.g. that the energy of three molecules A, B and C is expressed as a sum of two-body interactions of the pairs AB + BC + AC. Some interactions, such as the

electrostatic interaction, are strictly pair-wise additive, i.e. only contribute to the $U^{(2)}$ term. The repulsion and dispersion are approximately pairwise additive, whereas the induction energy is not. Induction is a response of the molecular charge density to the electric fields due to the surrounding molecules, which modifies the electric fields within the crystal to cause a further response, and so on. This is implicitly a many-body interaction and is not pairwise-additive.

Thus, the largest contributor to the lattice energy is usually the leading term, $\sum_{i<j} U^{(2)}(\mathbf{R}_i, \mathbf{R}_j)$, however in solids and liquids each molecule is surrounded by a coordination sphere of molecules, whose fields contribute constructively and destructively to produce the field experienced by the central molecule. In systems where the electrostatic fields are strong, such as from ions, the *n*-body terms will not be negligible[2] and the higher terms are required. In most organic solids we assume pairwise additivity, and the induction is included in the empirically fitted model used to describe the short-range repulsive and long-range attractive forces. By fitting the parameters of such a model to reproduce experimentally observed properties, the overall effect of the many-body forces can be reproduced although the 'repulsion' and 'dispersion' energies will not reflect their proper interpretations from perturbation theory as they include all interactions not modelled explicitly.

2.2. Physical Origin and Definition of Contributions

The interaction energy of a pair of molecules can be decomposed into physically distinct contributions which then lend themselves to calculation with different theoretical models. This decomposition is most easily seen through intermolecular perturbation theory, sometimes known as the polarization expansion.

2.2.1. Intermolecular Perturbation Theory

For a system of two molecules, A and B, the Schrödinger equation of the isolated molecule A is[1]

$$\left(H^A - E_n^A\right)\phi_n^A = 0, \tag{2}$$

and expressed similarly for B. For the dimer, the Hamiltonian may be written

$$H = H^A + H^B + \lambda H'. \tag{3}$$

Here the total Hamiltonian is the sum of the isolated Hamiltonians plus a perturbation, H'. This contains the interactions expressed in atomic units:

$$H' = \sum_{\alpha}\sum_{j}\frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} - \sum_{\alpha}\sum_{j\neq\alpha}\frac{Z_\alpha}{r_{\alpha j}} - \sum_{\beta}\sum_{i\neq\beta}\frac{Z_\beta}{r_{\beta j}} + \sum_{i}\sum_{j}\frac{1}{r_{ij}}, \tag{4}$$

where $R_{\alpha\beta}$ is the distance between nuclei α and β (within molecules A and B respectively); $r_{\alpha j}$ is the distance between nucleus α and electron *j*; likewise for $r_{\beta i}$ and r_{ij} are the separation of the electrons *i* and *j* (of molecules A and B respectively). In the absence of the perturbation H' (i.e. $\lambda = 0$), the ground-state wavefunction for the dimer is $\phi_0 = \phi_0^A\phi_0^B$ with energy $E_0 = E_0^A + E_0^B$. This is only valid at long-range because the wavefunction is not antisymmetric with respect to electron exchange between A and B. In this case, polarization theory is only asymptotically valid and the exchange and repulsive terms that result in the repulsive wall will be missing from the description.

Using standard Raleigh-Schrödinger perturbation theory the wavefunction and the energy can be expanded[3] as the following series

$$\phi = \phi_0 + \lambda\phi_{pol}^{(1)} + \lambda^2\phi_{pol}^{(2)} + \cdots, \tag{5}$$

$$E = E_0 + \lambda E_{pol}^{(1)} + \lambda^2 E_{pol}^{(2)} + \cdots, \tag{6}$$

where the interaction energy in polarization theory is defined as

$$E_{\text{int}} = E_{pol}^{(1)} + E_{pol}^{(2)} + E_{pol}^{(3)} + \cdots, \tag{7}$$

and

$$(H^{(0)} + \lambda H')\phi = E\phi\,. \tag{8}$$

The expansion can be grouped in orders of λ; $H'\phi^{(1)} = E^{(1)}\phi^{(1)}$ and $H'\phi^{(2)} = E^{(2)}\phi^{(2)}$ etc., each of which can be examined separately from the others. In Rayleigh-Schrödinger perturbation theory, the first order polarization energy is written in as

$$E_{pol}^{(1)} = \left\langle \phi_0 \left| H' \right| \phi_0 \right\rangle. \tag{9}$$

The first-order energy is the expectation of the ground-state wavefunction, which is defined to be the electrostatic energy. The second order perturbation energy can also be expressed in this notation:

$$\begin{aligned} E_{pol}^{(2)} &= \sum_{k\neq 0} \frac{\left|\left\langle \phi_0^A \phi_0^B \left| H' \right| \phi_k^A \phi_0^B \right\rangle\right|^2}{E_0^A - E_k^A} \\ &\quad + \sum_{l\neq 0} \frac{\left|\left\langle \phi_0^A \phi_0^B \left| H' \right| \phi_0^A \phi_l^B \right\rangle\right|^2}{E_0^B - E_l^B} \\ &\quad + \sum_{\substack{k\neq 0 \\ l\neq 0}} \frac{\left|\left\langle \phi_0^A \phi_0^B \left| H' \right| \phi_k^A \phi_l^B \right\rangle\right|^2}{E_0^A + E_0^B - E_k^A - E_l^B} \\ &= E_{ind,pol}^{(2)} + E_{disp,pol}^{(2)} \end{aligned}, \tag{10}$$

where *k* and *l* denote the quantum states of the monomer A and B respectively. The second order energy contains the leading terms for the induction and dispersion energies. Induction energy is the change in energy between the ground state charge density of molecule A with a field from molecule B, expressed in terms of the excited states of A, and *vice versa*. Dispersion is described in the limits only of excited states on both molecules, arising from instantaneous correlation between fluctuating charge densities, thus the perturbation expansion involves excited states in both molecules.

2.2.2. Electrostatic Energy

The first-order energy has a physical interpretation that can be found by expressing it as the expectation of the interaction operator

$$E_{estat}^{(1)} = \left\langle \phi_0 \left| H' \right| \phi_0 \right\rangle = E_{pol}^{(1)}. \tag{11}$$

This is the interaction between two charge densities, so can also be rewritten as an integral over each of their volumes[3]

$$E_{estat}^{(1)} = \int\int \rho_A(r_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \rho_B(r_2) d\mathbf{r}_1 d\mathbf{r}_2, \tag{12}$$

where the charge distribution $\rho_A(r_1)$ for molecule A is given by the expression

$$\rho_A(r) = \sum_\alpha Z_\alpha \delta(r - R_\alpha) - \rho_A^{el}(r). \tag{13}$$

The Dirac delta function, $\delta(r - R_\alpha)$, represents the contribution of the nucleus α at position R_α, with charge Z_α, and $\rho_A^{el}(r)$ is the electronic charge distribution of molecule A. Thus (12) shows that $E_{pol}^{(1)}$ can be defined by the interaction of charge densities of the isolated molecules. This term is very orientationally dependent for non-spherical molecules and so plays a major role in determining the structure adopted by the solid state.

2.2.3. Induction

As seen in equation (10), the term for the induction energy is part of $E_{pol}^{(2)}$ and is given by the expression

$$E_{ind}^{(2)} = E_{ind,polA \leftarrow B}^{(2)} + E_{ind,polB \leftarrow A}^{(2)}, \tag{14}$$

where $E_{ind,A \leftarrow B}^{(2)}$ is the induced energy of A in the field of B, and *vice versa* $(A \leftrightarrow B)$. Using Rayleigh-Schrödinger perturbation theory this is expanded to

$$E_{ind}^{(2)} = -\sum_{k\neq 0} \frac{\left|\left\langle \phi_0^A \phi_0^B \left| H' \right| \phi_k^A \phi_0^B \right\rangle\right|^2}{E_k^A - E_0^A} - \sum_{l\neq 0} \frac{\left|\left\langle \phi_0^A \phi_0^B \left| H' \right| \phi_0^A \phi_l^B \right\rangle\right|^2}{E_l^B - E_0^B}, \quad (15)$$

which comprises the first two terms of equation (10). The first term involves the ground state charge distribution of B that produces a field, causing a change in the charge distribution of A (described by the excited states ϕ_k^A), which interacts with the field to provide a lowering of the energy. The second term is the same, but that it is the charge density of B that responds to the field produced by the ground state of A.

In reality the effect does not stop at this level: it is intuitive that the change in the charge density of A will result in a change in the field experienced by B, and hence the charge density of B will change in response, changing the field experienced by A, and so on until the effect of induction is converged. This effect would be instantaneous between all *n* molecules and occur as the crystal was constructed. Each of these iterative interactions is accounted for by increasingly higher orders in the energy expansion, but here the leading term for the induction energy is only the initial interaction between one molecule in its ground state and the perturbation of the other, as modelled by a superposition of excited states.

Each region of the molecular charge distribution has a characteristic susceptibility to an applied field which can be expressed in terms of a frequency-dependent density susceptibility (FDDS):[4]

$$\alpha_A(r,r';\omega) = 2\sum_{k\neq 0} \frac{E_k^A - E_0^A}{(E_k^A - E_0^A)^2 - \omega^2} \left\langle \phi_0^A \left| \hat{\rho}_A(r) \right| \phi_k^A \right\rangle \left\langle \phi_k^A \left| \hat{\rho}_A(r') \right| \phi_0^A \right\rangle. \quad (16)$$

This describes the linear response to a frequency dependent perturbation, where $\hat{\rho}_A(r) = \sum_{a \in A} \delta(\mathbf{r} - \mathbf{r}_a)$ is the electronic charge density. For static polarizabilities, $\omega = 0$ and the expression cancels to

$$\alpha_A(r, r'; 0) = 2 \sum_{k \neq 0} \frac{\left\langle \phi_0^A \middle| \hat{\rho}_A(r) \middle| \phi_k^A \right\rangle \left\langle \phi_k^A \middle| \hat{\rho}_A(r') \middle| \phi_0^A \right\rangle}{E_k^A - E_0^A}, \tag{17}$$

and so the induced energy of A due to the field of B can be written[3] as a double integral of the molecular charge densities and the FDDS:

$$E_{ind,A \leftarrow B}^{(2)} = -\sum_{k \neq 0} \frac{\left| \left\langle \phi_0^A \phi_0^B \middle| H' \middle| \phi_k^A \phi_0^B \right\rangle \right|^2}{E_k^A - E_0^A} = -\frac{1}{2} \iint \frac{\alpha_A(\mathbf{r}, \mathbf{r}'; 0) \rho_B(\mathbf{r}) \rho_B(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \tag{18}$$

This relationship is developed into a model intermolecular potential in terms of distributed polarizabilities in section 2.5. The major aim of this thesis is to test and implement this model derived and tested by Misquitta & Stone for van der Waals dimers of small molecules.[4-8] The implementation will allow the induction energy of crystal structures of small organic molecules to be calculated and iterated to self consistency.

2.2.4. Dispersion[1]

Dispersion is a universal attractive force with a purely quantum mechanical origin, and cannot be described by classical physics. It arises from the instantaneous correlation of fluctuations in the charge densities. The zero-point motion of electrons in one molecule creates a temporary dipole which induces a correlated dipole in the neighbouring molecules. In terms of perturbation theory, this concept can be expressed by the equation

$$U_{disp} = -\sum_{\substack{k \neq 0 \\ l \neq 0}} \frac{\left| \left\langle \phi_0^A \phi_0^B \middle| H' \middle| \phi_k^A \phi_l^B \right\rangle \right|^2}{E_k^A - E_0^A + E_l^B - E_0^B}. \tag{19}$$

Notice that in this expression the charge densities of both molecules are described by a mixing of excited states, whereas only one of the molecules distorts for induction. Equation (19) resembles the form of the induction energy (15), and hence could be expressed as a product of the FDDS (16), except that the denominator contains terms for both molecules. This problem can be solved by using the mathematical identity

$$\frac{1}{A+B} \equiv \frac{2}{\pi}\int_0^\infty \frac{AB}{(A^2+v^2)(B^2+v^2)}dv, \tag{20}$$

used by Casimir and Polder[9]. Following the work of Misquitta[4], equation (19) can now be written

$$\begin{aligned} U_{disp} &= -\frac{2}{\pi}\int_0^\infty \sum_{k_A\neq 0^A} \frac{\left\langle \phi_0^A \middle| H' \middle| \phi_k^A \right\rangle \left\langle \phi_k^A \middle| H' \middle| \phi_0^A \right\rangle E_{k0}^A}{(E_{k0}^{A\,2} - \omega^2)} \sum_{m_A\neq 0^A} \frac{E_{l0}^B \left\langle \phi_0^B \middle| H' \middle| \phi_l^B \right\rangle \left\langle \phi_l^B \middle| H' \middle| \phi_0^b \right\rangle}{(E_{l0}^{B\,2} - \omega^2)} d\omega \\ &= -\frac{1}{2\pi}\int_0^\infty \alpha^A(r_1, r_1'; i\omega)\alpha^B(r_2, r_2'; i\omega)d\omega \end{aligned} \tag{21}$$

where $E_{k0}^A = E_k^A - E_0^A$. Dispersion is therefore the product of the molecular polarizabilities integrated over imaginary frequencies, which is a difficult concept that arises from the mathematical derivation. The well-behaved nature of polarizabilities at imaginary frequencies, which decrease monotonically from the static polarizabilities at $\omega = 0$, to zero as $\omega \to \infty$ means that the functions are easily evaluated with numerical quadrature techniques[1].

2.2.5. Exchange-Repulsion[1]

The exchange and repulsion energies are of opposite sign, but they both act at short range, and are usually modelled together as a single repulsive function. The repulsion part comes from the Pauli exclusion principle, which forbids electrons of the same spin to occupy the same space. The exchange part comes from the indistinguishable nature of electrons that

allows them to exchange between molecules. Long-range perturbation theory fails at short-range when the charge distributions overlap due to the lack of correct antisymmetrization in the wavefunction ϕ, which neglects the indistinguishability. This can be corrected using an antisymmetrizer, $\mathcal{A}$, in the perturbation theory, now called Symmetrized Rayleigh-Schrödinger (SRS) perturbation theory[3]. The *n*th order correction to the energy in SRS is defined as:

$$E_{SRS}^{(n)} = \frac{\left\langle \phi_0 \left| H' \hat{A} \right| \phi_{pol}^{(n)} \right\rangle}{\left\langle \phi_0 \left| \phi_{pol}^{(n)} \right\rangle \right.}, \tag{22}$$

and the *n*th order exchange energy is defined as the remainder of this and the polarization energy, i.e.:

$$E_{exch}^{(n)} = E_{SRS}^{(n)} - E_{pol}^{(n)}. \tag{23}$$

Further details may be found in Ref. [3]. The major part of the exchange energy appears in $E_{exch}^{(1)}$. At second-order, the exchange analogues to the induction and dispersion energies can be defined as: $E_{exch-ind}^{(2)}$ and $E_{exch-disp}^{(2)}$. Higher-order terms are usually neglected.

2.2.6. Other Short-Range Terms

Additional interaction at shorter intermolecular distances are penetration, charge-transfer and also a damping of induction and dispersion interactions. Penetration arises at short range as the difference between the true electrostatic energy of the overlapping charge distributions, as in (12), and the long-range model, such as distributed multipoles, that neglects the extent of the charge distribution. It is additive, usually attractive except at very short distances, and decays exponentially so is usually absorbed into the exchange-repulsion terms for modelling. Charge-transfer describes the transfer of electron density from high energy occupied orbitals on one molecule to the lowest unoccupied orbital of another. This term is non additive and is always stabilising to the structure, otherwise the

electron density would not transfer. It must be small or it would describe a form of covalent bonding and so intermolecular perturbation theory would no longer be valid. Damping is a corrective factor to account for the exchange-induction and exchange-dispersion, which are the differences between calculating the induction and dispersion interactions with exchange taken into account, and calculating them using the non-expanded long range approximation.

2.2.7. Summary

At long range the intermolecular interaction energy comprises of electrostatic, dispersion and induction terms, which can be modeled using the properties of the isolated molecule's charge density. At short range, the interactions are due to the overlapping molecular charge densities and cannot be modeled analytically from perturbation theory.

When deriving model potentials to use in simulations, we can use these theories to get the electrostatic term relatively easily as described in 2.3.2. This thesis is concerned primarily with the evaluation of the induction energies using atomic polarizabilities[4,8] of the type described in section 2.4. The polarizabilities that are used for this are also used to calculate dispersion[6,10], and allow for model interaction potentials to be fitted to computational data that are specific to a particular molecule, rather than using a large number of crystal structures and fitting to experimental observations.

2.3. Model Intermolecular Pair Potentials

2.3.1. Currently Used Empirically-Fitted Model Potentials

At the start of this project, the interaction model generally used by the Price group for crystal structure prediction consisted of an electrostatic model from a routine distributed multipole analysis (DMA, described in 2.3.3) performed on an MP2/6-31G(d,p) wavefunction calculation of the isolated molecule, plus an empirical potential which nominally represents the 'repulsion-dispersion' interaction but includes all other terms which are absorbed into the pairwise additive approximation of using an atom-atom based exp-6 function:

$$U(R,\Omega)=\sum_{\iota\in M,\kappa\in N}(A_\iota A_\kappa)^{1/2}\exp\left(-\frac{(B_\iota+B_\kappa)}{2}R_{\iota\kappa}\right)-\frac{(C_\iota C_\kappa)^{1/2}}{R_{\iota\kappa}^6}, \qquad (24)$$

where i and k are the atoms of molecules M and N, and ι and κ refer to the atom types of i and k. This 'FIT' potential had been derived by fitting the parameters of the function such that known crystal structures were reproduced in conjunction with the electrostatic model[11,12]. As such, all non-electrostatic interactions and the penetration energy are absorbed into this functional form. This potential is limited to C,H,N,O atoms[12], although some terms for Cl[13] and Br[14] have been determined for use in some specific applications. Hence the induction energy was included only in an average way into the model potential, by the empirical fitting.

2.3.2. Electrostatic Models

Intermolecular interactions of organic molecules are usually dominated by the electrostatic force, so there is a need to have a computationally efficient model of the molecular charge

density that can represent the non-uniform distribution around the molecule. This is usually done with multipole moments, which express a non-spherical charge distribution by an expansion of charge separation.

2.3.2.1. Cartesian Definition of Multipole Moments[1]

Total charge is the zero-order multipole moment, defined as $q = \sum_a e_a$, where e_a is the charge on particle *a*, including all nuclei as well as electrons. The first-order moment is the dipole; the separation of two equal and opposite charges along a vector, such as in hydrogen fluoride. This can be expressed similarly as[1]:

$$\hat{\mu}_\alpha = \sum_a \mathrm{e}_\mathrm{a} \mathbf{r}_{a\alpha} \tag{25}$$

where $\mathbf{r}_a$ is the vector position of particle *a*. Being an operator, the expectation value can be found in the normal way:

$$\mu_\alpha = \langle \phi | \hat{\mu}_\alpha | \phi \rangle = \int \rho_n(\mathbf{r}) r_\alpha \rho_n(\mathbf{r}) d^3\mathbf{r} \tag{26}$$

where α is a Cartesian axis x, y or z. The next in this series is a quadrupole, such in the molecule N_2, which is defined as[1]:

$$\hat{\Theta}_{\alpha\beta} = \sum_a e_a \left(\frac{3}{2} r_\alpha r_\beta - \frac{1}{2} {r_a}^2 \delta_{\alpha\beta} \right) \tag{27}$$

The quadrupole definition has two properties of note. Firstly it is traceless, so that $\Theta_{xx} + \Theta_{yy} + \Theta_{zz} = 0$. Secondly, it is symmetric with permutations of the indices ($\Theta_{yx} = \Theta_{yx}$) hence there are only 5 independent components. The next multipole, octopole, is defined as eight charges arranged in a three-dimensional array. Higher order multipole moments than this are outside of this implied three-dimensional analogy, but moments of rank n can be generalised as

$$\xi^{(n)A}_{\alpha\beta\ldots\upsilon} \equiv (-)^n \frac{1}{n!}\sum_a e_a r_a^{2n+1} \frac{\partial}{\partial r_{a\alpha}} \frac{\partial}{\partial r_{a\beta}} \ldots \frac{\partial}{\partial r_{a\upsilon}} \left(\frac{1}{r_a}\right) \tag{28}$$

This definition is not changed by any permutation of the suffixes, and is traceless for any pair of suffixes. As such, a multipole moment of rank *n* has 2*n*+1 independent components, the same as the number of spherical harmonics of rank *n*.

2.3.2.2. Spherical Tensor Definition of the Multipole Moments[1]

For many applications the spherical tensor formulation is more convenient to use than the Cartesian.[15] The expansion of $1/r_{ab}$ for interacting sites can be written in the form

$$\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \sum_{lm} \frac{r_<^l}{r_>^{l+1}} (-1)^m C_{l,-m}(\theta_1, \varphi_1) C_{lm}(\theta_2, \varphi_2) \tag{29}$$

In this expansion $r_<$ and $r_>$ are the smaller and larger of the two distances respectively, and *C* are spherical harmonics. If the terms are defined as $\mathbf{r}_1 = \mathbf{A} + \mathbf{a}$ and $\mathbf{r}_2 = \mathbf{B} + \mathbf{b}$, where **A**, **B** are the vectors describing the centre of mass of molecules A and B, and **a**, **b** are vectors describing their particles relative to the centre of mass (see Figure 1), the expansion can be rewritten as

$$\frac{1}{|\mathbf{R} + \mathbf{b} - \mathbf{a}|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^m R_{l,-m}(\mathbf{a} - \mathbf{b}) I_{lm}(\mathbf{R}) \tag{30}$$

where $R_{l,-m}(\mathbf{r})$ and $I_{km}(\mathbf{r})$ are regular and irregular spherical harmonics, resepectively. This expression is only valid for $R > |\mathbf{a} - \mathbf{b}|$. Using the standard addition theorem for spherical harmonics, and by introducing the multipole moment operator[1] $\hat{Q}^{A(G)}_{l_1 m_1} = \sum_{a \in A} e_a R_{lm}(\mathbf{a})$, the perturbation may be expressed in the following form using Wigner 3*j* symbols:

$$
\begin{aligned}
H' &= \frac{1}{4\pi\varepsilon_0}\sum_{a\in A}\sum_{b\in B}\frac{e_a e_b}{|\mathbf{R}+\mathbf{b}-\mathbf{a}|} \\
&= \frac{1}{4\pi\varepsilon_0}\sum_{l_1 l_2}\sum_{m_1 m_2 m}(-1)^{l_1}\left(\frac{(2l_1+2l_2+1)!}{(2l_1)!(2l_2)!}\right)^{1/2} \\
&\times \sum_{a\in A} e_a R_{l_1 m_1}(\mathbf{a})\sum e_b R_{l_2 m_2}(\mathbf{b}) I_{l_1+l_2,m}(\mathbf{R})\begin{pmatrix} l_1 & l_2 & l_1+l_2 \\ m_1 & m_2 & m \end{pmatrix} \\
&= \frac{1}{4\pi\varepsilon_0}\sum_{l_1 l_2}\sum_{m_1 m_2 m}(-1)^{l_1}\left(\frac{(2l_1+2l_2+1)!}{(2l_1)!(2l_2)!}\right)^{1/2} \\
&\times \hat{Q}^{A(G)}_{l_1 m_1}\hat{Q}^{B(G)}_{l_2 m_2} I_{l1+l2,m}(\mathbf{R})\begin{pmatrix} l_1 & l_2 & l_1+l_2 \\ m_1 & m_2 & m \end{pmatrix}
\end{aligned}
\tag{31}
$$

The superscript *G* in this expression refers to the fact that this is in the global coordinate system. The molecular properties are usually defined prior to knowing its orientation in the global axies, and these may be expressed in a local axis system fixed in the molecule by means of the transformation

$$
\hat{Q}^{(L)}_{lk} = \sum \hat{Q}^{(G)}_{lm} D^{l}_{mk}(\mathbf{\Omega}), \tag{32}
$$

for which $\mathbf{\Omega}$ is the rotation between the global and local systems, and $D^{l}_{mk}(\mathbf{\Omega})$ is the Wigner rotation matrix for this rotation. The perturbation may be written in this local frame, for which the superscript *L* is dropped:

$$
H' = \frac{1}{4\pi\varepsilon_0}\sum_{l_1 l_2}\sum_{k_1 k_2}\begin{pmatrix} l_1+l_2 \\ l_1 \end{pmatrix}\hat{Q}^{A}_{l_1 k_1}\hat{Q}^{B}_{l_2 k_2}\overline{S}^{k_1 k_2}_{l_1 l_2 l_1+l_2} R^{-l_1-l_2-1}. \tag{33}
$$

The distance dependence is given by *R*. The relative orientation of the two molecules A and B is given by the $\overline{S}^{k_1 k_2}_{l_1 l_2 l_1+l_2}$ term. These functions become the interaction tensors, T, and are tabulated by Stone.[16] The multipole moment operator still contains complex spherical harmonics, but may be transformed to real functions using the identity $\hat{Q}_{l,k} = \sum_{\kappa} X_{k,\kappa}\hat{Q}_{l,\kappa}$ where $X_{k,\kappa}$ are the transformation coefficients. The perturbation may then be expressed more simply as

$$H' = \sum_{tu} \hat{Q}_t^A T_{tu}^{AB} \hat{Q}_u^B \,, \tag{34}$$

The interaction tensor T_{tu}^{AB} is the spherical tensor form, and includes the relative orientation of the local molecular axis systems. The subscripts have been condensed into a single index, *t* or *u*, which take the values 00, 10, 11s, 11c, 20, 21s, 21c, 22c, ... Where *t* and *u* are 00, i.e. interacting charges, then T_{tu} is simply $1/R_{AB}$. For higher ranking multipoles the tensor is more complex; for interacting dipoles it is

$$T_{1\alpha,1\beta} = \frac{3r_\alpha^a r_\beta^b + c_{\alpha\beta}}{R^3} \,, \tag{35}$$

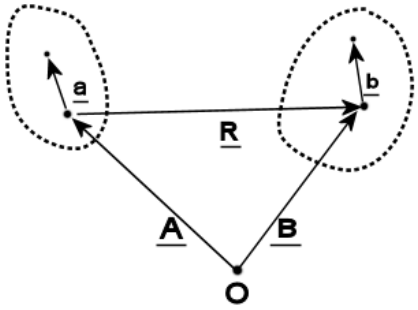
where r_α^a is the directional cosine, the component α of the unit vector in direction from *a* to *b*, expressed in the local axis system of *a*, and similarly for r_β^b. $c_{\alpha\beta}$ is the dot product of the unit vectors that define local axis systems for sites *a* and *b*.

Substituting (34) into (11) leads us to an expression for the electrostatic energy in terms of multipole moments that is:

$$\begin{aligned} E_{estat} &= \left\langle \phi_0 \left| \hat{Q}_t^A T_{tu}^{AB} \hat{Q}_u^B \right| \phi_0 \right\rangle \\ &= \sum_{tu} Q_t^A T_{tu}^{AB} Q_u^B \end{aligned} \tag{36}$$

The electrostatic energy is therefore calculated by summing over the interaction between the multipole moments.

Figure 1: Defining the position vectors of interacting molecules A and B (molecular centre of mass), and charges a and b within the molecules, relative to a global origin.



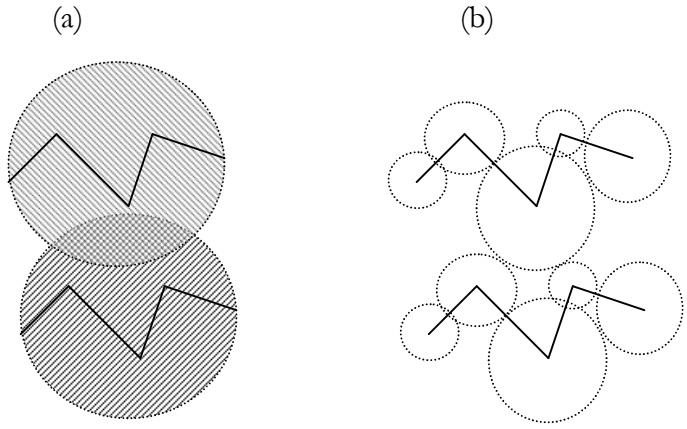
2.3.3. Distributed Multipole Moments[1]

It is possible to represent the electrostatic potential around a molecule by an expansion of multipole moments about its centre. At long range this is exact, since the charge distribution appears approximately spherical as the distance approaches infinity. At shorter range the anisotropic nature of charge density is more apparent, which is not easily represented by a central multipole expansion. A better representation is a multiple expansion about several sites in the molecule, so as to represent, for example, the lone-pairs on oxygen at one end of the molecule and a methyl at the end of a hydrocarbon chain at the other, by different multipole moments on the oxygen and carbon sites, respectively. This multi-site multipole description can be achieved with the distributed multipole expansion, where the molecule is divided into discrete regions, usually each atom, each with its own expansion. Such a scheme allows the electrostatic interaction of the multipoles to be calculated for short intermolecular distances. In the central multipole expansion, a sphere of divergence centred on the multipole site encompasses all of the electrons in the molecule. Where these spheres overlap, as in most crystal structures due to the close contacts, the multipole expansion is not valid, and is slowly convergent where the spheres are in contact. With the distributed description, a sphere is centred on each multipole site and has a radius determined by the charge distribution being represented at that site. This allows molecules to be much closer and in a greater range of orientations without any of

the spheres overlapping (Figure 2). This need for distribution to interaction sites fits in very well with the atom-atom approach to the intermolecular potential. The interaction sites, which are normally the nuclear sites, will be denoted *a* and *b* from here after, to be consistent with traditional notation in the theory of intermolecular forces applied to pairs of molecules.

Distributed moments may be determined in a number of ways, and may be derived from a computational analysis of the charge density or by fitting to experimental data. However, measuring multipole moments experimentally[17-19] is challenging and the data is insufficient to recommend any particular distribution of multipoles across specific sites[1].

Figure 2: Spheres of divergence for the (a) central and (b) distributed multipole models. The molecular orientations are the same in each case, but in the central model the spheres overlap and the multipole description does not converge, whereas the spheres in the distributed model do not overlap.



A systematic way to determine the distributed moments is by the distributed multipole analysis (DMA) of Stone[20]. This approach uses the way in which we express the molecular

wavefunction in terms of Gaussian functions, which are usually centred on the nuclei. The charge density involves the sum of products of these functions, which are also Gaussian functions centred on a point between the sites from which they are a product:

$$\exp[-\alpha(\mathbf{r}-\mathbf{a})^2]\exp[-\beta(\mathbf{r}-\mathbf{b})^2] = \exp\left[-\frac{\alpha\beta}{\alpha+\beta}(\mathbf{a}-\mathbf{b})^2\right]\exp[-(\alpha+\beta)(\mathbf{r}-\mathbf{p})^2], \qquad (37)$$

where $\mathbf{p}$ is a point $(\alpha\mathbf{a}+\beta\mathbf{b})/(\alpha+\beta)$, about which the product is centred between the points $\mathbf{a}$ and $\mathbf{b}$ which are the centres of the original functions, usually the atomic nuclei. A multipole expansion can represent this function at $\mathbf{p}$, and moved to the arbitrary site $\mathbf{x}$ with the formula

$$Q_{lk}(\mathbf{x}) = \sum_{k=0}^{l}\sum_{q=-k}^{k}\left[\binom{l+m}{k+q}\binom{l-m}{k-q}\right]^{1/2} \times Q_{kq}(\mathbf{p})R_{l-k,m-q}(\mathbf{x}-\mathbf{p}), \qquad (38)$$

as used by Stone[21].

The position to which the expansions are recentred at must be chosen by some means, and one method is to simply move the moments to the nearest[20] multipole site (usually a nucleus). This is very computationally efficient, but has been shown to be very dependent on the basis set used. It will produce unphysical moments with increasing basis set size[22] due to the presence of diffuse functions that may be used to describe special regions far from the site about which the function is centred. A recent improvement to this method has been to implement a real-space partitioning scheme that integrates the electron density at points on a grid around the atoms[22]. This density due to the tails of diffuse functions is then allocated proportionally amongst atoms to which it is nearby. This modified method of determining multipoles is tested as my first investigation in chapter 3. Since the development of the polarizability model requires larger basis sets with diffuse functions, it is important for the multipole moments to be consistent with changes in basis set, and physically relevant to the atomic type and local environment.

2.4. Polarizability

The charge density around an atom or molecule is not static, but responds to external fields that attract or repel the electrons. As well as multipole moments calculated for a molecule with no external fields (static moments), moments can be calculated for the induced density distribution due to a field. The induction energy is defined by (14) and (15) which can be interpreted as induced moments on molecule A interacting with the static moments on molecule B, and vice versa.

The advantages for using a distributed multipole-moment model for organic molecules apply also to the polarizability. The anisotropic nature of the charge distribution is more accurately described, as electrons have more freedom of movement in some regions of the molecule than others. The Raleigh-Schrödinger perturbation expressions for the second-order energy from (10) can written using the spherical harmonic multipole expansion[1] and site-site polarizabilities by

$$\begin{aligned} E_{ind}^{(2)} &= -\sum_{n\neq 0} \frac{\left\langle 0\left|\sum_{at} \hat{Q}_t^a V_t^a\right|n\right\rangle\left\langle n\left|\sum_{bu} \hat{Q}_u^b V_u^b\right|0\right\rangle}{E_n - E_0} \\ &= -\tfrac{1}{2}\sum_{abtu} V_t^a \alpha_{tu}^{ab} V_u^b \end{aligned} \tag{39}$$

where the polarizability tensor is given by

$$\alpha_{tu}^{ab} = \sum_{n\neq 0} \frac{\left\langle 0\left|\hat{Q}_t^a\right|n\right\rangle\left\langle n\left|\hat{Q}_u^b\right|0\right\rangle + \left\langle 0\left|\hat{Q}_u^b\right|n\right\rangle\left\langle n\left|\hat{Q}_t^a\right|0\right\rangle}{E_n - E_0}. \tag{40}$$

2.4.1. Calculation of Distributed Polarizabilities

The frequency dependent polarizabilities for real eigenstates are defined in (16) in terms of the charge distribution of the whole molecules. If $\hat{\rho}_A(r)$ and $\hat{\rho}_B(r)$ are replaced by the multipole moment operators $\hat{Q}_t^A$ and $\hat{Q}_u^B$ respectively, the frequency dependent polarizabilities can be found by integrating over space using the frequency-dependent density susceptibility (FDDS) from equation (16):

$$\alpha_{tu}^{AB}(\omega) = \iint \alpha(\mathbf{r},\mathbf{r}'|\,\omega)\hat{Q}_t^A(\mathbf{r})\hat{Q}_u^B(\mathbf{r}')d^3\mathbf{r}d^3\mathbf{r}', \tag{41}$$

The FDDS comes from the equations of coupled Kohn-Sham (CKS) theory[23,24], in the form[1]

$$\alpha(\mathbf{r},\mathbf{r}'|\,\omega) = \sum_{iv,i'v'} C_{iv,i'v'}(\omega)\phi_i(\mathbf{r})\phi_v(\mathbf{r})\phi_{i'}(\mathbf{r}')\phi_{v'}(\mathbf{r}')\,. \tag{42}$$

In this expression ϕ_l is a molecular orbital, with the subscripts *i* and *i'* (*v* and *v'*) labelling the occupied (and virtual) orbitals. $C_{iv,i'v'}(\omega)$ is a coefficient defined by the electric and magnetic Hessians of CKS theory[23,25]. Now equation (41) can be recast as

$$\alpha_{lm,l'm'}(\omega) = \sum_{iv,i'v'} C_{iv,i'v'}(\omega)\int \hat{Q}_{lm}(\mathbf{r})\rho_{iv}(\mathbf{r})d^3\mathbf{r}\int \hat{Q}_{l'm'}(\mathbf{r}')\rho_{i'v'}(\mathbf{r}')d^3\mathbf{r}' \tag{43}$$

where $\rho_{iv} = \phi_i\phi_v$, and the multipole moment operators have been replaced by their equivalent of real spherical harmonics.

Some manner of partitioning scheme is required to obtain polarizabilities for individual sites, and this is achieved by using density-fitting to expand the orbital products, ρ_{iv}, in terms of an auxiliary basis set $\{\chi\} = \{\chi^a, \chi^b, \ldots\}$, where $\{\chi^a\}$ is a set of basis functions centred on site *a*, etc. Hence the transition density is approximated by

$$\begin{aligned}\tilde{\rho}_{iv}(\mathbf{r}) &= \sum_a \sum_{p\in a} D_{iv,p}\chi_p(\mathbf{r}) \\ &= \sum_a \tilde{\rho}_{iv}^a(\mathbf{r})\end{aligned} \tag{44}$$

where $\tilde{\rho}_{iv}^{a}(\mathbf{r})$ is the contribution of the auxiliary functions of site *a* to the sum, and $D_{iv,p}$ is a coefficient determined by fitting.

When the density is partitioned in this way, the FDDS can be approximated similarly:

$$\alpha(\mathbf{r},\mathbf{r}'|\,\omega) \approx \sum_{a,b} \alpha^{a,b}(\mathbf{r},\mathbf{r}'|\,\omega)\,, \tag{45}$$

where

$$\alpha^{a,b}(\mathbf{r},\mathbf{r}'|\,\omega) = \sum_{p\in a, q\in b} \tilde{C}_{pq}(\omega)\chi_p(\mathbf{r})\chi_q(\mathbf{r}'), \tag{46}$$

and the coefficients are transformed as

$$\tilde{C}_{pq}(\omega) = \sum_{iv,i'v'} D_{iv,p} C_{iv,i'v'} D_{i',v',q}\,. \tag{47}$$

By substituting (45) into (43), the distributed polarizability for sites (a,b) is

$$\alpha_{lm,l'm'}^{a,b}(\omega) = \sum_{p\in a, q\in b} \tilde{C}_{pq}(\omega) \int \hat{Q}_{lm}(\mathbf{r}-\mathbf{a})\chi_p(\mathbf{r}) d^3\mathbf{r} \int \hat{Q}_{l'm'}(\mathbf{r}'-\mathbf{b})\chi_q(\mathbf{r}') d^3\mathbf{r}'. \tag{48}$$

To obtain a basis-independent and meaningful distribution scheme, a constrained density fitting procedure[4] is used that forces auxiliary basis functions centred on a site to describe the density local to that site only: an over-complete description of a system, using functions on one site to describe regions at distant sites, is a known problem in quantum chemistry[26].

2.4.1.1. Localized Polarizabilities

Distributed polarizabilities consist of many non-local terms that describe the response of a wavefunction to a perturbation in one region, and the subsequent change elsewhere of the charge density because of this. Calculation of the induction energy for a large system involving these cross-site terms will be computationally expensive and, because each term refers to specific sites in the molecule and their relation to each of the other sites, they cannot be used as an approximation for other systems. By localizing these terms, each site

has a polarizability that describes how it responds in the specific environment it is in, but not explicitly linked to other atoms in the molecule.

It is possible to transform the cross-site terms so that they refer only to one site, incorporating the response of the density to the field at that site, and that due to the response of the site's neighbours. One example of such a scheme has been developed by Le Sueur and Stone[27]. This approach retains the overall molecular polarizabilities of the molecule, although the convergence of the model is degraded when polarizabilities are shifted.

The procedure works by representing the response at site *b* due to a perturbation at site *b'* by instead the using sites *a* and *a'*. The field at *b* is expressed by a Taylor expansion about *a,* and the induced moments at *b'* by a multipole expansion about *a'*:

$$\alpha^{a}_{l_a m_a, l_{a'} m_{a'}} = (-1)^{l_a + l_{a'}} \left[\frac{(2l_a + 1)!}{(2l_b)!(2l_a - l_b)!} \right]^{1/2} \left[\frac{(2l_{a'} + 1)!}{(2l_{b'})!(2l_{a'} - l_{b'})!} \right]^{1/2} \times W_{l_b m_b, l_a m_a}(\mathbf{r}_a - \mathbf{r}_b) W_{l_{b'} m_{b'}, l_{a'} m_{a'}}(\mathbf{r}_{a'} - \mathbf{r}_{b'}) \alpha^{b}_{l_b m_b, l_{b'} m_{b'}} \tag{49}$$

$$W_{l'm', lm}(\mathbf{r}) = (-1)^{m} \begin{pmatrix} l - l' & l' & l \\ m - m' & m' & m \end{pmatrix} R_{(l-l')(m-m')}(\mathbf{r}), \tag{50}$$

where $R_{lm}(\mathbf{r})$ is a regular spherical harmonic. By choosing *a* = *a'* the polarizabilities $\alpha^{b}_{l_b m_b, l_{b'} m_{b'}}$, referring to two distant sites, can be removed and replaced instead with a set of polarizabilities $\alpha^{a}_{l_a m_a, l_a m_a}$, i.e. now referring to only one site. For this procedure to be effective the polarizabilities should not be moved too far ($|\mathbf{r}_a - \mathbf{r}_b|$ must be small) to retain the convergent properties of the multipole expansion, and restrictions must be placed to ensure the total charge is conserved:

$$\sum_{a\in A} \alpha^{aa'}_{l_a m_a,00} = 0. \tag{51}$$

These charge conservation rules are detailed by Stone[27], and fully implemented into the localization routines of ORIENT, which applies the procedure of Le Sueur and Stone[27] that preserves the charge-flow polarizabilities, and additionally tests that they have indeed been kept to. During the course of this thesis other localization methods have been developed and implemented in the WSM scheme[8].

Localizing the polarizabilities inevitably has the effect of reducing the accuracy of the model. The final stage in the WSM method is to fine-tune the localized polarizabilities to reproduce a set of point-to-point polarizability data generated by the SAPT(DFT) calculation[8].

2.4.2. Induced Moments and Induction Energy

By using the localised polarizabilities in a distributed multipole formulation, the induced moment multipoles on a polarizable site can be calculated and these can then be used to calculate the induction energy. The induced moment is the product of the electrostatic field and atomic polarizability

$$\Delta Q_t^a = \alpha_t^a V_t^a , \tag{52}$$

and subsequently, the induction energy is the product of the electrostatic field and the induced moments,

$$E_{ind} = \sum_{a\in A}\sum_{t} \Delta Q_t^a V_t^a . \tag{53}$$

Solving these equations once will give the first-order induction energy.

To recover higher order contributions then equation (52) must be solved again using contributions to the field from both the static and induced multipole moments. The calculation is iterated so that each molecule responds to each other, until the induced moments converge and there is no further change in the induction energy.

2.4.3. Damping Close-Contact Interactions

The charge densities of individual molecules will overlap slightly in organic crystal structures, especially where hydrogen-bonding gives rise to intermolecular distances that are less than the sum of the relevant atomic van der Waals radii. In these instances there will be some contribution to the interaction from exchange-induction, resulting in a damping of the induction energy as calculated using the atomic multipoles and polarizabilities. When applying the iterative procedure, the induced moments will be slow to converge where the separation between sites is small, and in some cases will diverge rapidly. A damping function is used in an attempt to compensate for the divergence of the multipole expansion although little is known about damping functions for induction[6]. We use a Tang-Toennies damping function, which has been used to damp multipole expansions of the dispersion energy. Studies on molecules including water, formamide and benzene show that it does not correct fully for the limitations of the multipole model,[8] but no better form has been proposed. The Tang-Toennies damping function has the form[28]

$$f_n(\beta R_{ab}) = 1 - \left(\sum_{k=0}^{2n} \frac{(\beta R_{ab})^k}{k!} \right) \exp\left(-\beta R_{ab}\right), \tag{54}$$

where n is the sum of the ranks of multipoles u and v. It has been effective in reducing the singular behaviour of the induction energy when inter-site distances are particularly short.[8] The damping expression is used in the calculation of the induced moments, using

$\beta = 2\sqrt{2I_X}$, where I_X is the first vertical ionization potential of the molecule in atomic units[7].

The damping has to be applied in an atom-atom form using the parameter derived for the whole molecule, which is a simplistic extrapolation from studies on small polyatomic systems.[8] A damping parameter based on the specific atom-atom interactions would be preferable, and could be implemented into the scheme described in this thesis if they could be determined. This thesis is evaluating the induced moments and induction energy in crystals of larger molecules than those used to develop the WSM distributed polarizability model, including strong hydrogen bonding and halogen groups.

2.5. Non-Empirical Potentials

The rationale behind non-empirical potentials is to improve the quality of chemical simulations by using functional forms and parameters that are derived from chemical theory, rather than from experiment. As previously mentioned, empirical models such as the Williams potential, will have absorbed the damped induction energy along with other terms into the repulsion-dispersion functional form. Such potentials can be very effective: they are quick to evaluate, the parameters come from a database of values, and for crystal structure prediction there have been many successes using empirical potentials.

By moving to model the induction energy explicitly for organic crystals, it must be removed from the generic repulsion-dispersion model. This cannot be done without re-determining all of the parameters. It would also require an experimental way to validate the contribution to the lattice energy due to induction. Alternatively, we can look to the theory and use accurate *ab initio* calculations to develop functional forms and parameter sets for

every contribution to the intermolecular potential. Ideally these would be determined for each system studied. Using the specifically designed potential should give much more accurate results than an empirical potential for a given system. Such potentials have been used in the past for crystal structure prediction of chlorobenzenes[29], oxalic acid[30] and of some amides[31], where the effects of induction have been assumed to be negligible when comparing polymorphs.

2.6. Programs Used in this Thesis

The following subsections describe the main codes used for testing and development in this thesis. Due to the collaborative nature of the research group, a number of other codes are referred to in the text when describing the methods used to generate the data I then work with. These codes are listed in Appendix A. Any subroutines or functions that I have written or significantly modified, along with some utility codes I have created to reformat the output of one program for input into another, are listed in Appendix B,

2.6.1. Cluster Calculations and Visualisation: ORIENT

ORIENT[32] is a program written to carry out calculations on systems of interacting molecules, clusters and surfaces. It implements a site-site potential to calculate the electrostatic energy using a sum over interacting multipole moments as in equation (36), up to R^{-6}. ORIENT also implements an atom-atom exp-6 repulsion-dispersion potential, and induction interactions using distributed multipoles and atomic polarizabilities described in section 2.5, including damping. The program is designed to model small clusters of molecules, and includes non-central forces and torques resulting from interactions. Molecular properties are defined in a molecule-fixed axis system, and whole molecules and their properties are translated and rotated about their centres of mass using Cartesian coordinates and Euler angles to create systems and clusters. A full description of

capabilities of ORIENT can be found on Stone's website.[32] Equations for energy, interaction tensors, distributed properties and molecular rotation are given in Stone's book.[1] A key feature is that ORIENT already supports anisotropic, distributed atomic polarizabilities to calculate induction energies, and to iterate the calculation. However, the code was only intended for work with small clusters of molecules. Some modifications were made ORIENT to allow the scaling up to the large clusters I have used in this thesis.

Throughout this thesis, ORIENT is used to calculate the interaction energy, including induction if polarizabilities are provided, of large clusters of molecules generated from the crystallographic cell. These clusters are used to validate the coding of the induction energy into DMACRYS, by comparing the electrostatic interaction energies of the crystal and the cluster approximation, and induced moments which arise in each case.

ORIENT also provides a facility to produce interactive 3-D plots of a van der Waals surface, showing the electrostatic potential due to the distributed multipole model. Van der Waals radii are defined as those in ref [33] and polar hydrogen sites (H-N and H-O) have a radius of zero[34]. Plots of the electrostatic and induction energies, using a +1 point charge as a probe, are produced in this way throughout this thesis.

2.6.2. Lattice Energy Minimization with Anisotropic Atom-Atom Potentials: DMACRYS

DMACRYS, formerly DMAREL, is a lattice energy minimization code that has progressed from working with small rigid polyatomics to complex molecular crystals. It is this code that my work has been primarily concerned with adapting, so that induction energies for molecular crystals may be evaluated. It works with rigid-body molecules in an orthonormal

Cartesian axis system, using distributed multipole moments and an exp-6 repulsion-dispersion model potential to calculate the lattice energy, forces and torques.

It takes as input SHELX or FDAT files that contain cell parameters and fractional coordinates for atoms in the crystallographic asymmetric unit cell, as well as a set of symmetry cards that relate to the crystal space group. This is interpreted by the utility program NEIGHCRYS, which generates the input file for DMACRYS, including all molecules generated by symmetry elements present in the cell, using a defined axis system. Rigid molecules are defined from the atomic coordinates using a maximum length of a covalent bond in the molecule, and the user defines the molecule-fixed axis system using three atomic sites: two to give the x axis, and a third to define the xy plane, assuming that the z axis will complete an orthogonal, right-handed system. Internally, the program defines that the global z axis lies along the crystallographic *c* axis, the global x axis is parallel to the reciprocal *a* axis, and the global y axis is defined to give a right-handed orthogonal set. The relation between the molecule-fixed axes, the crystallographic cell axes, and the global internal axes is shown by Figure 4 page 70. Once the molecule-fixed axes are chosen any molecular properties which depend on the axes (i.e. multipole moments and polarizabilities) must be calculated in the same axis system centred about the molecular centre of mass in order to be compatible. Where a molecule in the cell is generated by an inversion operation, it is necessary that the right-handed axis system is maintained in order to calculate the intermolecular interactions, forces and torques. The result of maintaining a right-handed convention is that some molecular properties (i.e. multipole moments and polarizabilities that are odd powers of z in the spherical tensor operator) must have a change of sign for molecules generated by an inversion operator.

NEIGHCRYS also collates the required terms for the exp-6 repulsion-dispersion potential from a database. It is able to identify the atomic types by their bonding environment; for the simple FIT[35,36] potential this is just the basic atomic type, except for hydrogen where it distinguishes between H-O, H-N and H-C hydrogen atoms. It also includes the more complex Williams (WILL01)[37] potential, which has different types defined for more elements, as well as shortening the covalent bonds to hydrogen by 0.1 Å to correct for the discrepancy between the measured position of the electrons by X-ray diffraction and the actual position of the hydrogen nucleus. If neither of these schemes is sufficient to distinguish different atom types, it is also possible to provide a customized list of atomic types, which would be used for custom made or *ab initio* derived potentials.

Intermolecular interactions are calculated in direct space up to a cut-off distance (usually 15 Å by default) using distributed multiple moments up to $l = 4$ (hexadecapole) and repulsion dispersion potential. Electrostatic interactions are summed up to the limit R^{-5} and long-range electrostatic interactions, i.e. charge-charge, charge-dipole and dipole-dipole, are summed over direct space within the cut-off sphere and also additionally subject to an Ewald summation over reciprocal space, which includes the non-negligible contribution to the energy from the long-range interactions.

DMACRYS calculates the non-central forces, torques, and second derivatives, which are due to the multipole interactions. The forces and torques are transferred to the molecular centres of mass and used to determine the strains on the rigid molecules, with the result of determining translation or rotation of the molecules they may be subject to. The change in the crystal structure is expressed as a vector, $\boldsymbol{\delta}$, which comprises of the six rotational and translational components of each molecule in the unit cell, and six strain matrix elements.

Using this information it is possible to express the intermolecular lattice energy as a function of a small change in the lattice parameters using a power series:

$$U(r') = U(r) + \boldsymbol{\delta}^{\mathbf{T}}.\mathbf{g} + \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\delta}^{\mathbf{T}} \bullet \mathbf{W} \bullet \boldsymbol{\delta} \tag{55}$$

where $\mathbf{g}$ is the vector of first derivatives, and $\mathbf{W}$ is a matrix of second derivatives, which can be calculated analytically. Using this equation the displacement from the equilibrium is estimated using $\boldsymbol{\delta} = -\mathbf{W}^{-1} \cdot \mathbf{g}$, which gives the search vector to a modified Newton-Raphson algorithm that minimizes the lattice energy. For the minimization procedure the second derivative matrix is only calculated periodically, and is updated using an algorithm to save computation time. Once a stable minimum is found the matrix is recalculated explicitly in cases where the derivatives are required for determining properties of the structure, such as the *k=0* rigid body phonon frequencies.

2.6.2.1. Ewald Summation

Lattice summations take the form of $1/r^n$. Where n > 3 these sums are absolutely convergent, otherwise they are only conditionally convergent, i.e. the result depends upon the order in which the terms are summed and surface properties of the crystal. The conditionally convergent sums consist of the charge-charge (n=1), charge-dipole (n=2) and dipole-dipole (n=3) interactions, which are implemented in DMACRYS using Cartesian coordinates, and the charge-quadrupole (n=3) interaction is also conditionally convergent but is not included in the DMACRYS Ewald summation code.

The Ewald method uses a reciprocal space sum that is computationally more efficient. The electrostatic potential at a point in space, **r**, due to an infinite lattice of point charges is given by the expression:

$$V_E(\mathbf{r}) = \frac{1}{V_0 \varepsilon_0} \sum_{\mathbf{k}\neq 0}^{\infty} A(k) \sum_{j}^{N} exp\left(-i\mathbf{k}\cdot\left(\mathbf{r}_j - \mathbf{r}\right)\right) + \frac{1}{4\pi\varepsilon_0} \sum_{j}^{\infty} B_0\left(\left|\mathbf{r}_j - \mathbf{r}\right|\right) \quad , \tag{56}$$

where

$$A(k) = \left\{\frac{1}{k^2}\right\} exp\left(-k^2/4\xi^2\right), \tag{57}$$

$$B_0(u) = \left\{\frac{1}{u}\right\} erfc(\xi u), \tag{58}$$

$$erfc(\xi u) = \frac{2}{\sqrt{\pi}} \int_{\xi u}^{\infty} exp\left(-s^2\right) ds . \tag{59}$$

In these formulae, N is the number of point charges in the unit cell, ξ is the Ewald convergence parameter that determines the weighting of the reciprocal space part included in the sum through the function $A(k)$, V_0 is the volume of the unit cell and **k** is the reciprocal lattice vector. This system may be adapted for point charges and dipoles by employing the operator

$$\hat{M}_j = C_j - \mathbf{D}_j \cdot \hat{\nabla} . \tag{60}$$

The application of this operator leads to the expression

$$V_E(\mathbf{r}) = \frac{1}{V_0 \varepsilon_0} \sum_{\mathbf{k}\neq 0}^{\infty} A(k) \sum_{j}^{N} \left(C_j - i\mathbf{D}_j \cdot \mathbf{k}\right) exp\left(-i\mathbf{k}\cdot\left(\mathbf{r}_j - \mathbf{r}\right)\right) + \frac{1}{4\pi\varepsilon_0} \sum_{j}^{\infty} \left(C_j B_0\left(\left|\mathbf{r}_j - \mathbf{r}\right|\right) - \mathbf{D}_j \cdot \left(\mathbf{r}_j - \mathbf{r}\right) B_1\left(\left|\mathbf{r}_j - \mathbf{r}\right|\right)\right) \quad , \tag{61}$$

where $B_1\left(\left|\mathbf{r}_j - \mathbf{r}\right|\right)$ is derived from the recurrent formula

$$B_l(u) = \frac{1}{u^2}\left\{(2l-1) B_{l-1}(u) + \frac{\left(2\xi^2\right)^l}{\xi\sqrt{\pi}} \exp\left(-\xi^2 u^2\right)\right\}. \tag{62}$$

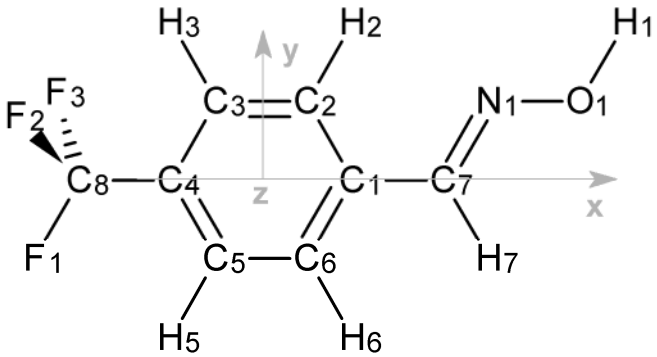
The energy of the system, per cell, is therefore

$$U = \frac{1}{2V_0\varepsilon_0}\sum_{\mathbf{k}\neq 0}^{\infty} A(k)\left|\sum_{j}^{N}\left(C_j - i\mathbf{D}_j\cdot\mathbf{k}\right)exp\left(-i\mathbf{k}\cdot\mathbf{r}_j\right)\right|^2 + \frac{1}{4\pi\varepsilon_0}\sum_{l}^{2}\sum_{i}^{N}\sum_{j>i}^{\infty} G_{ji}^{l}\left(\mathbf{r}_{ji}\right)B_l\left(r_{ji}\right) + \tfrac{1}{2}\sum_{i}^{N}\phi_i^s \quad (63)$$

G_{ji}^{l} are scalar functions defined in [38], and the final sum over ϕ_i^s is a correction for the self-interaction of sites. The first sum, which is over the square of the modulus of the reciprocal space terms, is difficult to factor out into individual contributions to the energy at a site for the purposes of damping. Testing has shown (section 5.1) that the contribution of the Ewald sum to the induction energy is small, and so the field contributions due to charges and dipoles are summed over direct space only, as for the higher multipole interactions.

2.7. An Explicit Example of Molecular Properties for Calculating Electrostatic and Induction Contributions to the Lattice Energy

Figure 3: (E)-4-(Trifluoromethyl)benzaldehyde oxime, with the molecule-fixed axes shown.



I use (E)-4-(Trifluoromethyl)benzaldehyde oxime (ADEJUP[39] with the minor component of the CF_3 disorder removed) (Figure 3) as an example of a small molecule with a large degree of planarity, with a CF_3 group that places atoms out of the plane. Molecular

properties were calculated by CamCASP using the PBE0-AC/Sadlej charge density for the experimental molecular structure, hence the atoms are not truly planar, and the method is described fully in Chapter 4 and the references therein. The refined rank 1 polarizabilities are shown in Table 1, tabulated to show only the 6 independent components.

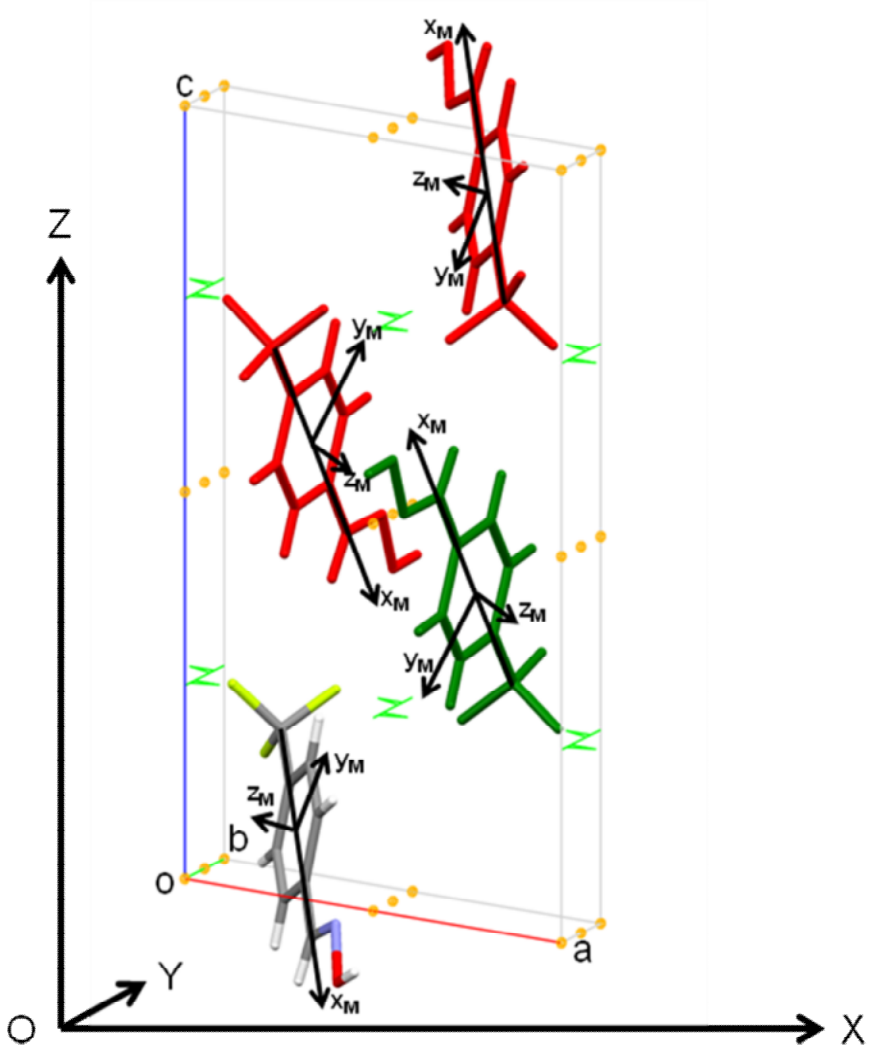
The choice of axes places the ring in the xy-plane, defining the x-axis along C8 – C1 with C3 to define the xy plane. As such the polarizabilities in x and y (11c and 11s) are the dominant terms in the polarizability tensors. The crystal is in the $P2_1/c$ spacegroup, and inversion operation requires some of the off-diagonal terms to undergo a change of sign, which are italicised in Table 1.

The fluorine atoms lie along the x-axis. F1 is in the xy plane and has the greatest polarizability in the y direction, then x. Of the off-diagonal terms, xy, i.e. along the bond direction, is significantly larger than the xz or yz terms. For F2 and F3, the greatest component is in the z-axis, reflecting their position above and below the plane of the molecule, and so the zx and zy off-diagonal components are relatively large. The hydrogen atoms also have relatively large xy terms since they are polarizable along their bond axis. Most of the larger polarizability values can be rationalised by charge density moving along the direction of covalent bonds.

Table 1: Independent polarizability tensor components for (E)-4-(Trifluoromethyl)benzaldehyde oxime. Diagonal components are listed in the first three columns, followed by the off-diagonal components. Those components that undergo a sign change with inversion are italicised.

	Polarizability tensor components / α_0^3					
	α(10,10)	α(11*c*,11*c*)	α(11*s*,11*s*)	α(*10,11c)*	α(*10,11s)*	α(11s,11c)
C1	3.12	17.27	12.99	*0.52*	*-0.34*	-6.55
C2	4.75	16.68	7.76	*0.08*	*0.24*	9.50
C3	4.44	19.36	7.89	*-0.35*	*-0.11*	-7.55
C4	2.77	18.83	9.39	*-0.02*	*-0.65*	0.35
C5	4.68	10.13	9.28	*-0.08*	*0.11*	5.40
C6	5.03	17.55	8.04	*-0.09*	*-0.02*	-3.25
C7	2.34	22.52	7.00	*0.38*	*-0.17*	-4.13
C8	3.94	2.94	4.56	*-0.01*	*-0.04*	0.41
O1	4.60	7.86	6.06	*0.16*	*-0.06*	-1.17
N1	4.93	18.55	7.04	*0.72*	*-0.47*	-3.57
F1	3.06	3.77	4.15	*-0.05*	*0.08*	-0.59
F2	4.64	3.91	3.53	*0.72*	*0.57*	0.14
F3	4.11	3.91	3.92	*-0.37*	*-0.49*	0.17
H1	1.06	1.12	1.02	*-0.01*	*-0.02*	0.04
H2	2.57	1.67	2.66	*-0.02*	*-0.13*	-1.27
H3	1.84	-1.00	1.81	*0.06*	*0.28*	1.00
H5	1.85	0.93	2.23	*0.05*	*0.15*	-1.68
H6	2.17	-0.29	2.14	*0.01*	*0.01*	0.72
H7	2.13	-0.65	3.34	*-0.32*	*-0.06*	0.40

Figure 4: Illustration of the various axes systems for the $P2_1/c$ structure of (E)-4-(Trifluoromethyl)benzaldehyde oxime . The Cartesian global axis system is related to the crystallographic axis system so the Z corresponds to c, X is parallel to bxc (and is not along a as β=99.3°) and Y completes a right handed orthonormal axis system (for this monoclinic cell, Y is parallel to b). The molecule in the input asymmetric unit cell is coloured by element, the molecule related by a 2-fold screw axis with identical multipole moments in dark green and the molecules related by an inversion centre or glide plane which have symmetry-related multipole moments in red. The light green lines denote the screw axes, the orange balls show the inversion centres, and the glide plane, in the ac plane, has been omitted for clarity.



Reference List

1. Stone, A. J. *The theory of intermolecular forces;* Oxford University Press: 2002.

2. Wilson, M.; Madden, P. A. Anion Polarization and the Stability of Layered Structures in Mx_2 Systems. *Journal of Physics-Condensed Matter* **1994,** *6* (1), 159-170.

3. Jeriorski, B.; Szalewicz, K. Symmetry-adapted perturbation theory. In *Handbook of molecular physics and quantum chemistry Volume 3: Molecules is the physico-chemical environment: Spectroscopy, dynamics and bulk properties*, 2002.

4. Misquitta, A. J.; Stone, A. J. Distributed polarizabilities obtained using a constrained density-fitting algorithm. *J. Chem. Phys.* **2006,** *124* (2), 024111.

5. Misquitta, A. J. Interaction energies from a symmetry-adapted perturbation theory based on density functional theory. *Abstracts of Papers of the American Chemical Society* **2004,** *228*, U201.

6. Stone, A. J.; Misquitta, A. J. Atom-atom potentials from ab initio calculations. *Int. Rev. Phys. Chem.* **2007,** *26* (1), 193-222.

7. Misquitta, A. J.; Stone, A. J. Accurate induction energies for small organic molecules: 1. Theory. *J. Chem. Theory Comput.* **2008,** *4* (1), 7-18.

8. Misquitta, A. J.; Stone, A. J.; Price, S. L. Accurate induction energies for small organic molecules. 2. Development and testing of distributed polarizability models against SAPT(DFT) energies. *J. Chem. Theory Comput.* **2008,** *4* (1), 19-32.

9. Casimir, H. B. G.; Polder, D. The Influence of Retardation on the London-Vanderwaals Forces. *Physical Review* **1948,** *73* (4), 360-372.

10. Misquitta, A. J.; Stone, A. J. Dispersion energies for small organic molecules: first row atoms. *Mol. Phys.* **2008,** *106* (12-13), 1631-1643.

11. Cox, S. R.; Hsu, L. Y.; Williams, D. E. Nonbonded Potential Function Models for Crystalline Oxohydrocarbons. *Acta Crystallogr. , Sect A.* **1981,** *37* (MAY), 293-301.

12. Williams, D. E.; Cox, S. R. Nonbonded Potentials For Azahydrocarbons: the Importance of the Coulombic Interaction. *Acta Crystallogr. , Sect. B* **1984,** *40* (8), 404-417.

13. Hsu, L. Y.; Williams, D. E. Intermolecular Potential-Function Models for Crystalline Perchlorohydrocarbons. *Acta Crystallogr. , Sect A.* **1980,** *36* (MAR), 277-281.

14. Day, G. M.; Cooper, T. G.; Cruz Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A.; Schweizer, B. Significant progress in predicting the crystal structures of small organic molecules - a report on the fourth blind test. *Acta Crystallogr. , Sect. B* **2009,** *65* (2), 107-125.

15. Stone, A. J. Appendix E: Cartesian-Spherical Conversion Tables. In *The Theory of Intermolecular Forces*, 2nd ed.; Clarendon Press: Oxford, 2002.

16. Stone, A. J. Appendix F: Interaction Functions. In *The Theory of Intermolecular Forces*, 2nd ed.; Clarendon Press: Oxford, 2002.

17. Coppens, P. Electron density distribution in cyanuric acid ii. Neutron diffarction study at liquid nitrogen temperature and comparison of X-ray and neutron diffraction results. *Acta Crystallogr. , Sect. B.* **1971,** *27* (1), 146-158.

18. Martin, A.; Pinkerton, A. A. Charge Density Studies Using CCD Detectors: Oxalic Acid at 100 K Revisited. *Acta Crystallogr. , Sect. B* **1998,** *54* (4), 471-477.

19. Munshi, P.; Row, T. N. G. Topological analysis off charge density distribution in concomitant polymorphs of 3-acetylcoumarin, a case off packing polymorphism. *Cryst. Growth Des.* **2006,** *6* (3), 708-718.

20. Stone, A. J. Distributed Multipole Analysis, Or How to Describe A Molecular Charge-Distribution. *Chemical Physics Letters* **1981,** *83* (2), 233-239.

21. Stone, A. J.; Alderton, M. Distributed multipole analysis - Methods and applications (Reprinted from Molecular Physics, vol 56, pg 1047-1064, 1985). *Molecular Physics* **2002,** *100* (1), 221-233.

22. Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation* **2005,** *1* (6), 1128-1132.

23. Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965,** *140* (4A), 1133-&.

24. Lee, A. M.; Handy, N. C.; Colwell, S. M. The Density-Functional Calculation of Nuclear Shielding Constants Using London Atomic Orbitals. *Journal of Chemical Physics* **1995,** *103* (23), 10095-10109.

25. Misquitta, A. J.; Stone, A. J. Distributed polarizabilities obtained using a constrained density-fitting algorithm. *Journal of Chemical Physics* **2006,** *124* (2), 024111.

26. Mayer, I.; Surjan, P. R. Improved Intermolecular SCF Theory and the BSSE Problem. *International Journal of Quantum Chemistry* **1989,** *36* (3), 225-240.

27. Lesueur, C. R.; Stone, A. J. Localization Methods for Distributed Polarizabilities. *Molecular Physics* **1994,** *83* (2), 293-307.

28. Tang, K. T.; Toennies, J. P. An Improved Simple-Model for the Vanderwaals Potential Based on Universal Damping Functions for the Dispersion Coefficients. *Journal of Chemical Physics* **1984,** *80* (8), 3726-3741.

29. Day, G. M.; Price, S. L. A nonempirical anisotropic atom-atom model potential for chlorobenzene crystals. *J. Am. Chem. Soc.* **2003,** *125* (52), 16434-16443.

30. Nobeli, I.; Price, S. L. A Non-Empirical Intermolecular Potential for Oxalic Acid Crystal Structures. *J. Phys. Chem. A* **1999,** *103* (32), 6448-6457.

31. Mitchell, J. B. O.; Price, S. L. A systematic nonempirical method of deriving model intermolecular potentials for organic molecules: Application to amides. *J. Phys. Chem. A* **2000,** *104* (46), 10958-10971.

32. *ORIENT: a program for studying interactions between molecules,* version 4.6; University of Cambridge, 2006

33. Bondi, A. Van der Waals Volumes and Radii. *J. Phys. Chem.* **1964,** *68* (3), 441-451.

34. Buckingham, A. D.; Fowler, P. W. A Model for the Geometries of Van der Waals Complexes. *Can. J. Chem.* **1985,** *63* (7), 2018-2025.

35. Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *J. Phys. Chem.* **1996,** *100* (18), 7352-7360.

36. Beyer, T.; Price, S. L. Dimer or catemer? Low-energy crystal packings for small carboxylic acids. *J. Phys. Chem. B* **2000,** *104* (12), 2647-2655.

37. Williams, D. E. Improved intermolecular force field for crystalline oxohydrocarbons including O-H...O hydrogen bonding. *J. Comput. Chem.* **2001,** *22* (1), 1-20.

38. Smith, W. Point multipoles in the Ewald summation (Revisited). *CCP5 Newsletter* **1998,** *46*, 18-30.

39. Jia, J.; Wang, X. Z.; Zhang, Y.; Wang, J. W. (E)-4-(Trifluoromethyl)benzaldehyde oxime. *Acta Crystallographica Section E* **2006,** *62* (7), o2683-o2684.

Chapter 3. Improving the Accuracy of Modelling the Electrostatic Interactions

Modelling interactions between all charged particles in a many-body system is hugely expensive computationally, and so atomic point charges or multipole moments are often used. The accuracy of the model depends upon the distance between the interacting particles, the number of expansions used and the quality of the original data from which the model is derived. Where the particles are far apart the electrostatic potential arising from each atom is approximately as if it were spherical and point charges may be as accurate as more complex models. However at short distances the electrostatic potential around the molecule changes rapidly and higher-order multipole expansions are required to reproduce the potential accurately. When highly directional interactions are present in a system, such as hydrogen bonds, then a high quality wavefunction is needed to accurately describe the electron density at the interaction sites, but using large diffuse basis sets has revealed a weakness in the traditional method of generating distributed multipoles[1]. This chapter is an investigation into a new implementation of the Distributed Multipole Analysis (DMA) that is designed to be more stable with large basis sets.

3.1. Distributed Multipole Analysis

The need to distribute the molecular multipole moments over many sites, usually nuclei, rather than at a single site is discussed in chapter 2.3, where the DMA method is also described. There are many ways in which multipoles can be determined. By starting with a charge density, measured experimentally by X-ray techniques or calculated *ab initio*, various partitioning schemes can be used to assign density to multipole sites. Mulliken Analysis

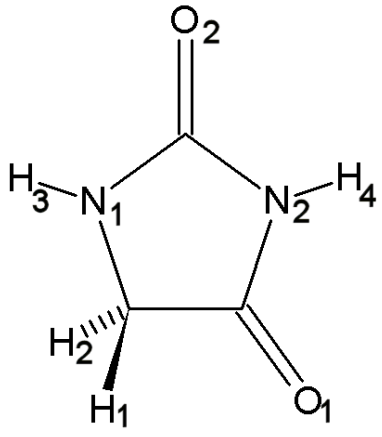
uses the basis functions centred on each atom to determine the charge density associated with that atom[2]. Being reliant upon the basis functions themselves this is very sensitive to the choice of basis set, and although the multipole expansion may still be valid, the charges are inadequate to describe the electrostatic potential around the atom. Another common partitioning scheme used is Bader's Atoms In Molecules[3], where the charge density of the molecule is divided into regions at zero flux in the gradient vector field of the electron density. This scheme usually has the effect of defining complex shapes, but by being based on charge density the method is relatively insensitive to the basis set.

Stone's Distributed Multipole Analysis[4] is a systematic way to determine multipole moments from an *ab initio* calculations. As described in chapter 2.3.3 this approach uses the way in which we express the molecular wavefunction in terms of Gaussian functions centred on arbitrary points. The functions are multiplied together, and expanded as a series of multipoles which are then moved to the nearest defined multipole site. While this is computationally very efficient, it becomes very unstable with large numbers of diffuse functions[5] as the distance increases between the charge density and the multipole expansion that represents it. For the triple-ζ quality basis sets used in this thesis, many atoms are assigned physically meaningless charges greater than 1 in magnitude, and even the wrong sign from what one would intuitively expect. An adaptation of the DMA method resolves these problems by defining a method to determine distributed multipoles that converge with basis set.

3.2. A New Method of Distributed Multipole Analysis (DMA)

The GDMA program is an implementation of the DMA method[4,6] used on the electron density matrix computed by *Gaussian*[7]. A new version, GDMA2.1, has recently been released[5] that uses a grid-based numerical integration over real space in order to both give a more physically intuitive description of each multipole site. The original method, GDMA1.2, was defined in 1985[6], and this chapter investigates the differences between the original implementation and the new method implemented 20 years later. The basis set dependence of the new version was investigated[5] using carbon monoxide and formamide, and this chapter supplements this by showing how the adjustable parameters affect the multipoles and interaction energies of a larger system: hydantoin[8]. This molecule, in the C2/c spacegroup, was used by the Cambridge Crystallographic Data Centre (CCDC) in the 2005 blind test[9]. I then present the case of how the changes in the DMA analysis affect the relative lattice energies of a crystal structure search of carbamazepine, when higher quality basis sets are used than usual for the Price group.

Figure 5: The hydantoin molecule.



3.2.1. GDMA 2.1

For non-diffuse functions the procedure is the same as outlined in section 2.3.3, since only the diffuse functions are treated in a new way. A function is considered diffuse when the exponents α,β in equation (37) are less than a given value (4 is the value), and numerical integration is applied over the space covered by the function. A grid of points is constructed around the multipole sites over which the integration takes place. Each grid point is connected to nearby multipole sites by weighting values assigned to it which describes the connection of that point to nearby multipole sites, determined by distance and the 'radius' parameters of the sites. When the electron density has been calculated for a grid point, it is transferred to the nearby sites according to these weightings. In this way the multipole moments evaluated at each site are a better reflection of the electron density in the vicinity of that site, and as such the distributed moments are more in keeping with chemical intuition.

3.2.2. GDMA Parameters

The implementation of the quadrature introduces several parameters. SWITCH determines which functions are integrated, set by default to 4. Setting this to zero results in the same multipole moments as calculated by GDMA 1.2, while using a large value of 100000, therefore to reasonably include all core functions, produced multipoles that varied in the 5^{th} and 6^{th} significant figures from the default setting. The calculation took significantly longer by two orders of magnitude as more functions were integrated, so that a value of 4 is probably optimal in most cases.

The integration grid cannot be finely adjusted, since the program automatically uses the next grid size available if the entered values do not correspond to one of the programmed

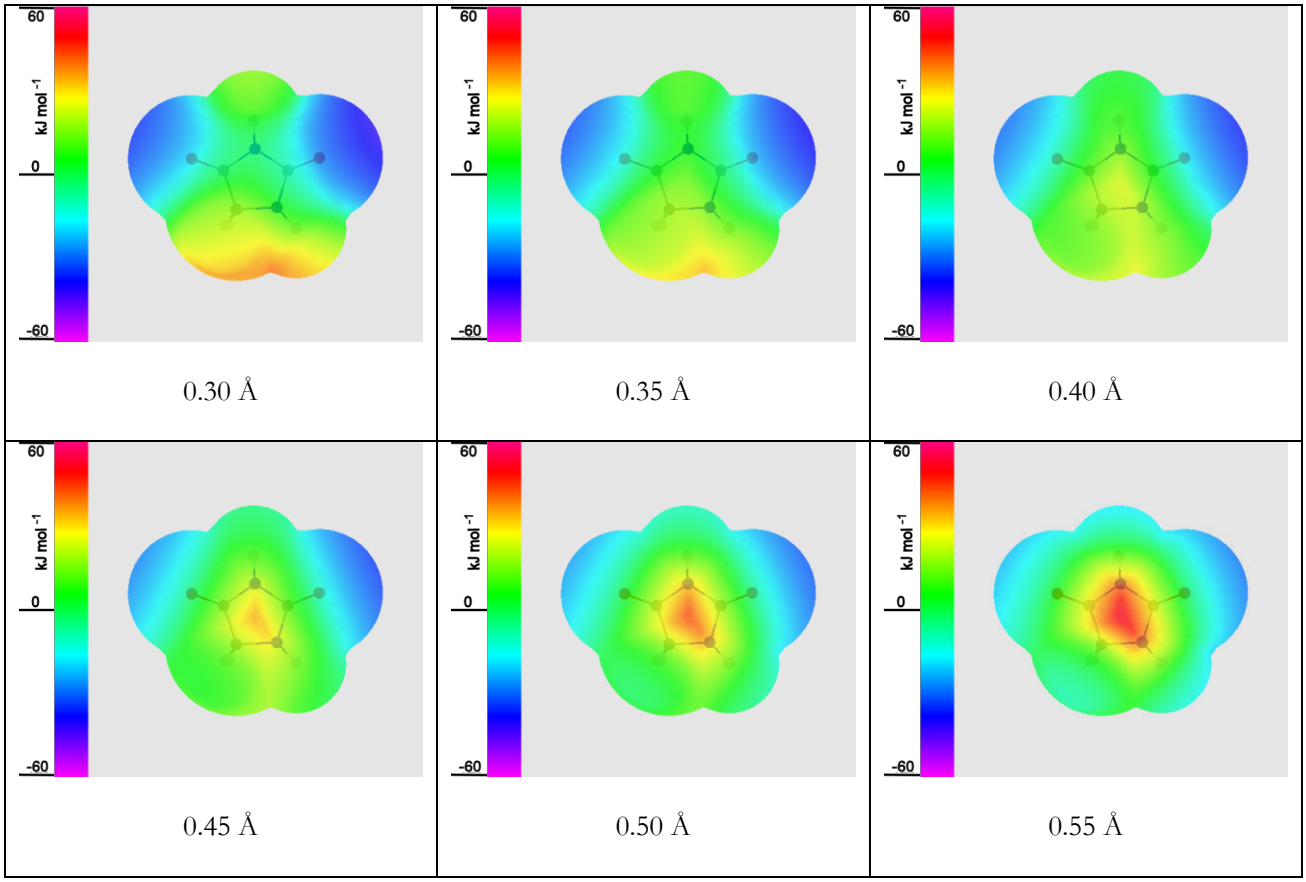
sizes. Quadrupling or halving the default value produced variations of the multipoles in up to the 3rd significant figure but usually much less. The use of the integration scheme is enough to implement the new method of the DMA, which is fairly insensitive to quality of the grid or the choice of when to use the grid as long as the most diffuse functions are treated with it.

3.2.3. Weighting to Atomic Sites

The grid points are weighted by their distance from nearby atoms and their RADIUS values. By default each atom has a radius, r_X^{GDMA} of 0.65 Å except for hydrogen sites which, following an investigation outlined below, has r_H^{GDMA} = 0.325 Å so as to attract less electron density to these sites. In practice this radius parameter is not a length but is used to define a ratio that weights the contribution of the electron density in regions of space to nearby atoms.

Figure 6 shows the trend this causes in the electrostatic potential due to the moments truncated at rank 0. The effect of the redistribution of charge density by the algorithm can be seen; using r_H^{GDMA} = 0.30 Å the oxygen sites are surrounded by negative potential and the hydrogen sites at the bottom of the plots are positive, while using r_H^{GDMA} = 0.55 Å the hydrogen sites are more neutral and the oxygen sites are much less negative due to electron being represented by large point charges on hydrogen sites. Charge decays with $1/r$ so the distribution of charge across the sites will affect the convergence of the multipole expansion if molecules are surrounded by hydrogen atoms carrying large charges.

Figure 6: The electrostatic energy of a unit charge probe on the 1.5x van der Waals surface of hydantoin. A rank 4 multipole analysis using GDMA2.1 with varying values for r_H^{GDMA} is truncated at rank 0, and the 3D surface is probed with a unit point charge. The scale is ±60 kJ mol^{-1}.



Because intermolecular interactions in organic crystals can be short, less than the sum of the van der Waals (vdW) radii, it is important that the multipole expansion does converge rapidly. Table 2 lists the minimum and maximum potential on 1.5x vdW surface of a truncated multipole expansion, and the difference from the rank 4 GDMA1.2 reference surface. Examining the first column, it appears that the truncated charges can generate a reasonable representation of the original rank 4 expansion if r_H^{GDMA} is reduced from 0.65 Å to 0.30 Å, with refinements given by including the higher ranking multipoles.

Table 2: The range of electrostatic potential values on the 1.5x van der Waals surface of hydantoin, in kJ mol^{-1}, when probed with a unit point charge. The rank 4 multipole expansion is truncated to the stated rank, and the difference with the complete GDMA1.2 rank 4 expansion is given as a percentage.

H radius	Rank 0		Rank 2		Rank 4	
	min	max	min	max	min	max
0.30 Å	-37.965 (-2.1%)	47.732 (2.4%)	-37.597 (-3.0%)	50.725 (8.7%)	-39.304 (1.4%)	46.130 (-1.1%)
0.35 Å	-35.444 (-8.6%)	40.485 (-13.2%)	-37.676 (-2.8%)	51.145 (9.7%)	-39.251 (1.2%)	46.209 (-0.9%)
0.40 Å	-32.871 (-15.2%)	33.108 (-29.0%)	-37.781 (-2.6%)	51.250 (9.9%)	-39.225 (1.2%)	46.261 (-0.8%)
0.45 Å	-30.220 (-22.1%)	43.531 (-6.6%)	-37.860 (-2.4%)	51.093 (6%)	-39.172 (1.0%)	46.261 (-0.8%)
0.50 Å	-27.489 (-29.1%)	55.477 (19.0%)	-37.965 (-2.1%)	50.751 (8.8%)	-39.146 (0.9%)	46.209 (-0.9%)
0.55 Å	-24.653 (-36.4%)	67.738 (45.3%)	-38.070 (-1.8%)	50.200 (7.7%)	-39.120 (0.9%)	46.130 (-1.1%)
0.60 Å	-21.687 (-44.1%)	79.763 (71.1%)	-38.149 (-1.6%)	49.491 (6.1%)	-39.094 (0.8%)	46.130 (-1.1%)
0.65 Å	-28.067 (-27.6%)	91.551 (96.3%)	-38.227 (-1.4%)	48.624 (4.3%)	-39.041 (0.7%)	46.209 (-0.9%)

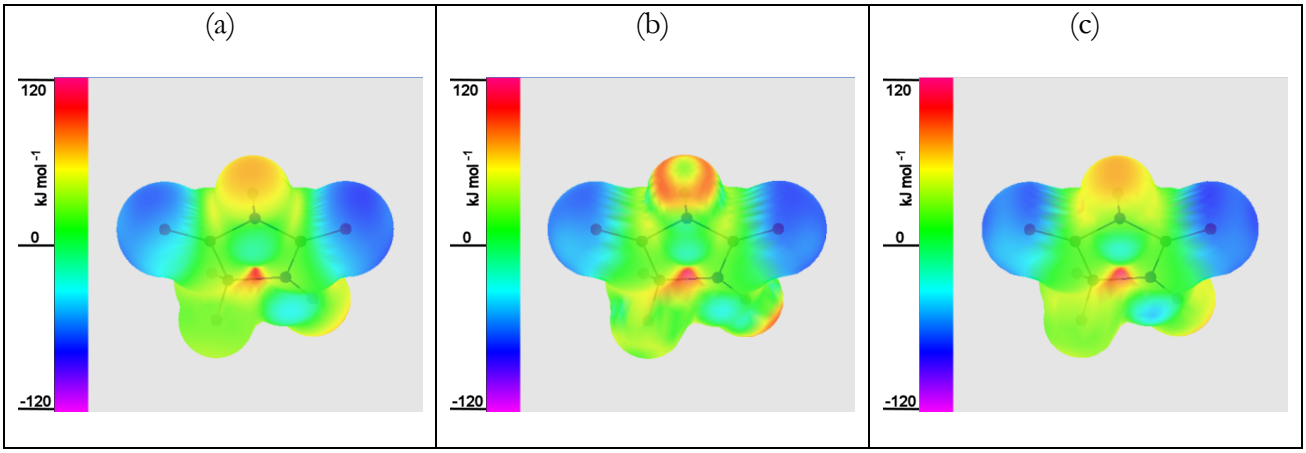
Table 3: The charge on atoms of hydantoin taken from an expansion up to rank 4, derived from an MP2/6-31G(d,p) wavefunction calculation.

Version	O1	O1	N1	N2	C1	C2	C3	H1	H2	H3	H4
1.2	-0.76	-0.81	-0.54	-0.59	1.06	0.97	-0.05	0.03	0.03	0.32	0.33
2.1 (r_H^{GDMA}=0.65)	-0.32	-0.32	0.39	0.39	0.34	0.32	0.97	-0.45	-0.44	-0.43	-0.42
2.1 (r_H^{GDMA}=0.35)	-0.35	-0.38	-0.21	-0.13	0.25	0.20	0.09	0.07	0.07	0.18	0.20

The atomic charges taken from a rank 4 DMA are presented in Table 3 to illustrate how the integration method with equal weightings produces undesirable terms. GDMA 1.2 gives fairly large negative charges on the electronegative oxygen sites and slightly less for nitrogen, while the two carbon atoms that are bonded to the oxygen are correspondingly positive as are the hydrogen sites. H3 and H4 are bonded to nitrogen and are an order of magnitude more positive than the C-H hydrogen sites. This is in line with our chemical intuition. Compare this to the v2.1 charges with r_H^{GDMA} = 0.65 and not only are the nitrogen sites positively charged, but all of the hydrogen sites carry equal and significant negative charges. This is clearly against chemical intuition and is undesirable even though the whole expansion produced equivalent potential energy surfaces. By reducing the weighting of hydrogen atoms when distributing the integrated electron density, the sign and magnitude of the charges become chemically reasonable; indeed they may even be preferable since assigning nearly unit charges to the oxygen and carbon sites in particular, as in the v1.2 method, may be considered to be unrealistic.

Final evidence of the need to reduce the relative weight of hydrogen sites is displayed in Figure 7 by means of the electrostatic potential on the 0.8x van der Waals surface. This is very close to the atoms, but it highlights an important feature that can significantly affect the interaction energy in certain geometries. The potential around the H-N proton, which is at the top of the image, tilted towards the reader, is positive and strongest in the direction along the N-H bond when using multipoles from GDMA1.2. Distributing the electron density with equal weighting between sites results in a very different potential in this region, which is now almost neutral along the N-H axis. Polar hydrogens such as this are frequently involved in hydrogen bonding, which is a very short, highly directional intermolecular interaction. Hydrogen bonding networks are often the main source of electrostatic stability in crystal lattices where hydrogen-bonding donors and acceptors are available, so it is important to accurately reproduce the potential in these regions. The effect on interaction energies are discussed in the following section, while the importance of accurately modelling these regions is further highlighted in section 3.4

Figure 7: Electrostatic energy of hydantoin rank 4 multipole expansion probed with a unit point charge, plotted on the 0.8x van der Waals surface. The NH group is tilted 45 degrees towards the viewer; (a) GDMA 1.2, (b) GDMA 2.1 r_H^{GDMA} = 0.65 Å for all atoms, (c) GDMA 2.1 r_H^{GDMA} = 0.35 Å. The scale is +/- 120 kj mol^{-1}



3.3. Packing Interactions

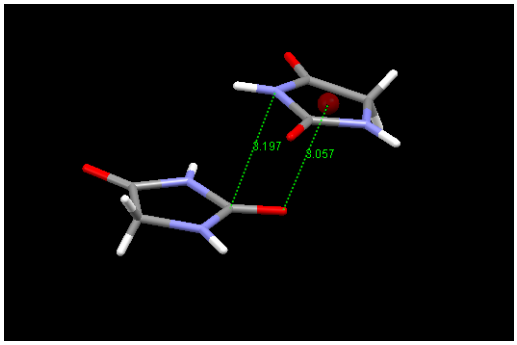
3.3.1. Molecular Dimers in Crystal Geometries

The intermolecular interactions in the hydantoin crystal[10] can be divided into three groups: inter-layer interactions due to the stacked nature of the layers; hydrogen bonding between two adjacent molecules in the same layer; non hydrogen bonding interactions between two adjacent molecules in the same layer. These are illustrated in Figure 8. For each of these molecular arrangements, the electrostatic energy has been calculated for the pair and listed in Table 4.

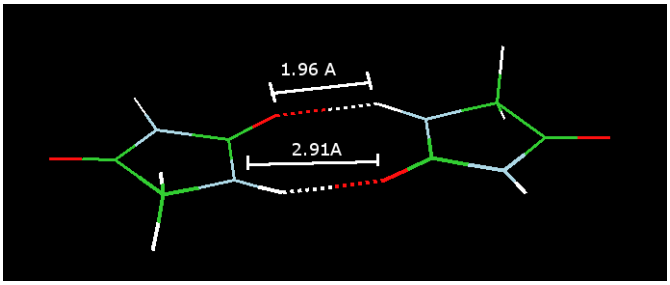
For the inter-layer interactions the shortest atom-atom distance between layers is 3.2 Å between nitrogen and oxygen, which greater than the sum of their van der Waals radii so it is expected that the multipole expansion is reasonably converged. As such, the interaction energies differ by ~1 % between versions 1.2 and 2.1.

Where the two molecules are hydrogen bonded the electrostatic energy varies more significantly. An oxygen-nitrogen distance of 2.91 Å is certainly less than the sum of van der Waals radii; a region which has been seen to differ between the three electrostatic models, particularly around the N-H protons (Figure 7). The original GDMA was designed to optimise the convergence of the expansion, so in this region, by definition, it is the best of the three models for calculating the energy. Hence improvement in GDMA2.1 model should lead to better agreement with the GDMA1.2 calculated values. Changing the hydrogen radius from the default slightly increases the magnitude of the interaction energy, seen in Table 4.

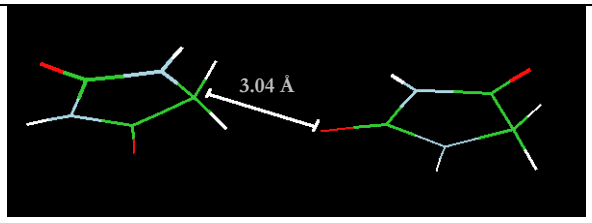
Figure 8: Close contacts between hydantoin molecules in the crystal. (a) the separation between layers, where the molecules are inverted and rotated (red sphere is the centroid); (b) the hydrogen bond motif showing O···H 1.96 Å and O···N 2.91 Å; (c) non hydrogen-bonding close contact between O···CH_2 of 3.04 Å



(a)



(b)



(c)

The non H-bonding energies reveal a huge relative error by using the optimised radius, and the default value seems to be much better. This highlights the importance of judgement when comparing the models, since although the value in the third column is more than

three times that in the first, the difference is one third of a kilojoule, which can be tolerated relative to the energy for hydrogen bonding.

Table 4: Electrostatic energies (kJ/mol) for hydantoin dimers.

	GDMA 1.2	**GDMA 2.1 (H 0.65 Å)**	**GDMA 2.1 (H 0.35 Å)**
Inter-layer (Figure 8 a)	-12.268	-12.389	-12.392
H-bonding (Figure 8 b)	-52.727	-49.079	-54.311
Non H-bonding (Figure 8 c)	0.157	0.170	0.533

3.3.2. Effect on Lattice Energies

The difference in energies seen in Table 4 is almost negligible for pairs of molecules, but lattice calculations involved hundreds or thousands of such interactions. The effect of varying both the DMA method and basis on the lattice energy of the crystal is shown in Table 5. Using the original method, increasing the size of the basis set has the effect of stabilising the lattice by 12 kJ mol^{-1}. This is quite a significant increase, and the implications to relative lattice energies of such an improvement in the model are described for carbamazepine in the following section. As well as this increase, it is apparent that the individual contributions to the total energy differ between the two basis sets by both order of magnitude and sign, which is due to the increase in diffuse functions as described above. Comparison between the GDMA1.2 and the default GDMA2.1 implementation, the latter has

much greater stability in the component energies between basis sets, which is what version 2.1 was designed to achieve. However, for these multipoles there is a significant reduction in the stability of the lattice which is due largely to an underestimation of hydrogen-bonding interactions. When r_H^{GDMA} is reduced to 0.35 Å as indicated by examining the potential energy plots in section 3.3.2, the component energies remain stable with basis set and the total electrostatic energy matches that for GDMA1.2.

3.3.3. Conclusion

Having tested the new implementation of the GDMA method using hydantoin, it is clear that there is one parameter for which it is very sensitive. The atomic radii, which are used as a ratio to weight points on the grid to atomic sites, have a significant influence on where the charge density is expanded about. By leaving all atoms to have an equal weighting, too much charge density becomes associated with hydrogen atoms and this leads to unphysical properties, such as negative charge on hydrogen atoms and positive charge on nitrogen atoms. This is significantly improved by reducing the radius associated with hydrogen atoms to be approximately half that of the other atoms. The code has been modified accordingly by the author so that $r_H^{GDMA} = 0.325$, being half the default value of other sites, and is used throughout the rest of this thesis.

Table 5: DMAREL lattice energies of hydantoin in the experimental crystal structure, using different electrostatic models, in kJ/mol.

	GDMA version / **Basis Set (MP2)**	**1.2**	**2.1** $r_H^{GDMA} = 0.65$	**2.1** $r_H^{GDMA} = 0.35$Å
Intermolecular charge-charge energy	6-31G(d,p)	-83.045	6.555	-39.834
	6-311++G(2d,2p)	-29.250	6.738	-41.946
Total charge-dipole energy	6-31G(d,p)	5.121	13.195	-18.822
	6-311++G(2d,2p)	-45.454	14.149	-21.926
Total dipole-dipole energy	6-31G(d,p)	5.029	-64.906	4.538
	6-311++G(2d,2p)	2.863	-67.726	5.135
Higher multipole interaction energy	6-31G(d,p)	1.160	-22.738	-17.536
	6-311++G(2d,2p)	-11.738	-30.113	-23.509
Total electrostatic energy	**6-31G(d,p)**	**-71.735**	**-67.894**	**-71.872**
	6-311++G(2d,2p)	**-83.579**	**-76.952**	**-82.246**
Total isotropic repulsion-dispersion energy		-25.022	-25.022	-25.022

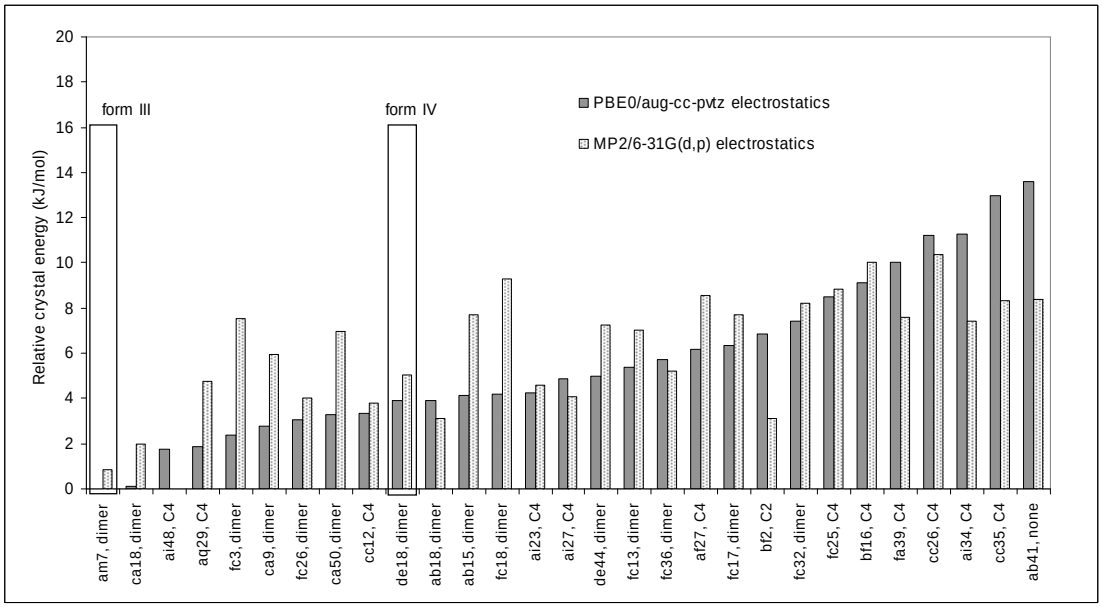
3.4. Effect of Basis Set on the Relative Lattice Energies of Carbamazepine

A previous theoretical search for carbamazepine polymorphs[2] used an MP2/6-31G** basis set to derive the electrostatic model. The observed crystal structures are based on hydrogen-bonded dimers, and these were found to be reasonably low in energy by the

original search. However the global minimum structure (ai8) was based on a hydrogen-bonded chain-based structures. The original rigid-body search was followed by a DMAFLEX[11] minimization, in which the hydrogen-bonding protons were allowed to relax in the crystal structure from their gas-phase positions. This procedure makes small changes to a set of defined bond lengths, angles and torsions, and performs an *ab initio* calculation and GDMA analysis for the new structure. The modified molecule is pasted into the crystal structure, which is allowed to relax as a rigid-body system until the code determines that another conformational change is indicated. Following this procedure, the known forms of carbamazepine were still ranked as less stable than un-observed chain-based crystal structures. I performed a study on the 29 lowest energy structures, using a triple-ζ quality basis set for the distributed properties and the updated GDMA2.1 analysis, and present the change in the relative energy in Figure 9.

The sensitivity to the quality of wavefunction used for the electrostatic model is seen by the change of relative ordering of the minima. Dimer-based structures become preferred instead of the catemeric structures. The unobserved catemer-based structures bf2, ai34 and ab41 are relatively destabilized, whereas the dimer-based structures am7, fc3 and fc18 have moved to become much more favourably ranked. In particular the structure am7, which corresponds to the known thermodynamic form III, becomes the global minimum with the new electrostatic model. The structure corresponding to form IV also becomes more relatively stable than before, reducing the lattice energy difference between these polymorphs. Thus, even prior to considering the induction energy in this work, the relative ordering given by having a more accurate charge density as a basis for the electrostatic model, shows that the known dimer structure is more stable than any hypothetical catemer structure.

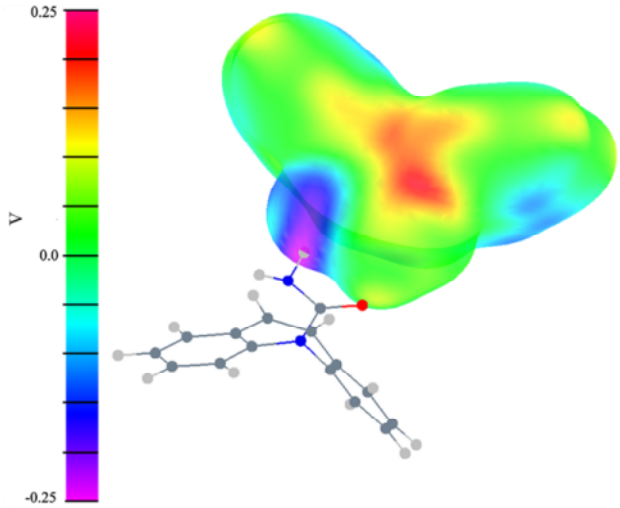
Figure 9: Carbamazepine relative crystal energies of the 29 lowest energy structures from a rigid body search and then refined using DMAFLEX to generate the structures used for this plots. Predicted known forms III and IV are indicated.



This result can be rationalised by examining the electrostatic potential energy surface of carbamazepine at the region of the main intermolecular interaction. Figure 10 displays the difference in the electrostatic potential energy between the two models, with a second molecule drawn in the location of a companion molecule in the dimer configuration. A surface is plotted that corresponds to the van der Waals surface scaled by 1.3 and with no radius on the polar hydrogen sites. The more expensively produced charge density generates a more negative potential at the point where the corresponding polar hydrogen atom is found, and more positive in the region of the corresponding oxygen. The difference is highly localised to this region to favour the dimer interaction, while many of the other regions that appear different, work against the catemer structures by weakening the attractive and strengthen repulsive electrostatic interactions. For instance, catemer structures tend to involve the edges of the C_6 rings, which are now more positive, docking

over the region above the molecule, which is also more positive and seen as a red blob in Figure 10.

Figure 10: Difference in electrostatic potential with theory and basis set for carbamazepine, shown in relation to a second molecule in the dimer-based interaction. The potential was calculated from distributed multipole moments derived from the GAUSSIAN03 charge density, and plotted using a unit charge probe as E(PBE0 aug-cc-pVTZ) – E(MP2/6-31G). The surface is defined by the van der Waals radii scaled by 1.3. The potential calculated from the PBE0 wavefunction is more negative where it interacts with a polar hydrogen atom and more positive where it interacts with an oxygen atom, hence strengthening the intermolecular interation.**



3.5. Conclusions

As computing resources become more and more powerful the method used to calculate a wavefunction for a small molecule may be improved, using more diffuse basis sets. By doing so, as the electronic calculation is systematically improved the original method of generating DMAs exhibits poor stability in terms of the multipoles produced. By treating the more diffuse Gaussian functions differently and allocating charge to sites using a

weighted grid, the derived multipoles exhibit much greater stability and convergence with basis set than before.

There is a cost of this stability, which is the introduction of adjustable parameters in the method. I have shown that the distributed multipoles are quite insensitive to the parameters except for the RADIUS , which is used to determine the ratio of charge density that is allocated to multipole sites. It makes intuitive sense that hydrogen atoms should not attract density to them, since their electrons are engaged in a single covalent bond, and representing large amounts of charge density by expansion on hydrogen atoms affects the interaction energies. As a result of this investigation the default radius for hydrogen sites has been changed by Stone from 0.65 Å to 0.325 Å in GDMA2.2[12].

By investigating the effect of using a better quality model of the molecular charge distribution, it has been shown that the predicted relative lattice energies of carbamazepine may be dramatically altered. This leads to an experimentally observed structure becoming the most favourable, and structures based on the observed dimer motif are relatively more favourable because important interactions are strengthened.

3.5.1. Future Potential Models

It is assumed that as the model intermolecular potential is improved, then crystal structure prediction will become more successful at identifying observable polymorphs by comparing relative lattice energies. Electrostatic interactions require good quality calculations of the wavefunction and an efficient method to derive a multipole description. GDMA2.2 allows us to generate physically plausible multipole moments that show

systematic improvement with the basis sets that are required to calculate higher order properties such as polarizability.

Other forces are presently described by an empirically fitted exp-6 repulsion-dispersion model, and it is known that this poorly reproduces lattices held together by dispersive forces between stacked aromatic rings. No other interactions, such as induction, are explicitly modelled but instead are absorbed into the fitting of the exp-6 model. Expanding the model to include induction, and therefore also to explicitly separating it from the repulsion-dispersion model, should further improve the relative stability of observable polymorphs. However, this also necessitates that a new repulsion-dispersion model is derived so that induction is not double counted. A method of generating such a potential has been derived by Misquitta and co-workers[13-17]. In the following chapter the induction energy between different carbamazepine polymorphs is estimated using the WSM polarizabilities, as part of an investigation into exactly which model should be coded into DMACRYS.

Reference List

1. Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005,** *1* (6), 1128-1132.

2. Mulliken, R. S. Electronic Population Analysis on Lcao-Mo Molecular Wave Functions .1. *J. Chem. Phys.* **1955,** *23* (10), 1833-1840.

3. Bader, R. F. W. *Atoms in Molecules. A Quantum Theory;* Clarendon Press: Oxford, 1990; Vol. 22.

4. Stone, A. J. Distributed Multipole Analysis, Or How to Describe A Molecular Charge-Distribution. *Chemical Physics Letters* **1981,** *83* (2), 233-239.

5. Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation* **2005,** *1* (6), 1128-1132.

6. Stone, A. J.; Alderton, M. Distributed Multipole Analysis - Methods and Applications. *Mol. Phys.* **1985,** *56* (5), 1047-1064.

7. *Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo,* R. *Gomperts,* R. *E. Stratmann, O. Yazyev, A. J. Austin,* R. *Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi,* R. L. *Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004,* 2004

8. Yu, F. L.; Schwalbe, C. H.; Watkin, D. J. Hydantoin and hydrogen-bonding patterns in hydantoin derivatives. *Acta Crystallogr. , Sect. C* **2004,** *60*, O714-O717.

9. Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P. A third blind test of crystal structure prediction. *Acta Crystallogr. , Sect. B* **2005,** *61* (5), 511-527.

10. Collins, A.; Cooper, R. I.; Watkin, D. J. Structure matching: measures of similarity and pseudosymmetry. *J. Appl. Crystallogr.* **2006,** *39*, 842-849.

11. Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction. II. Flexible molecules. *Mol. Phys.* **2007,** *105* (2-3), 273-291.

12. *GDMA2.2,Stone, A. J.,* http://www-stone.ch.cam.ac.uk/programs.html#GDMA 2006,

13. Misquitta, A. J.; Stone, A. J. Accurate induction energies for small organic molecules: 1. Theory. *J. Chem. Theory Comput.* **2008,** *4* (1), 7-18.

14. Misquitta, A. J.; Stone, A. J.; Price, S. L. Accurate induction energies for small organic molecules. 2. Development and testing of distributed polarizability models against SAPT(DFT) energies. *J. Chem. Theory Comput.* **2008,** *4* (1), 19-32.

15. *CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form,* University of Cambridge (2007), http://www-stone.ch.cam.ac.uk/programs.html#CamCASP: 2007

16. Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies

from time-dependent density-functional calculations. *J. Chem. Phys.* **2005,** *123* (21), 214103.

17. Misquitta, A. J.; Szalewicz, K. Intermolecular forces from asymptotically corrected density functional description of monomers. *Chem. Phys. Lett.* **2002,** *357* (3-4), 301-306.

Chapter 4. Testing the Importance of Induction Energy for Organic Crystal Structures

4.1. Testing the Importance of the Induction Energy

The Williams-Stone-Misquitta (WSM) method[1,2] allows us to obtain distributed polarizabilities from the *ab initio* properties of isolated molecules that are optimal at a given rank. From comparisons with SAPT(DFT) induction energies of a variety of dimers, ranging from HF to benzene,[2] we know that the damped WSM models are able to describe not just the long-range induction energy, but also an induction energy at short-range, even in the most testing area of hydrogen bonding contacts. These models result in errors of 2-7% of the dimer interaction energy at typical contact distances. However, for condensed phases the errors in total induction energy may be smaller than for van der Waals dimers because of the large number of longer range interactions, for which the WSM models are extremely accurate. Therefore, these polarizability models should give us a very powerful tool for computing the induction energy of an organic crystal.

Another way of approximating the induced moments of the crystalline phase is by *ab initio* evaluation of the molecular charge density, with the field of the surrounding molecules represented by point charges. When done self-consistently, we obtain an electronic response to point charge field model (SCERP), which does include some of the effects of electron penetration, because the point charges are fitted to the electrostatic potential close to the van der Waals[3] surface. This model is also limited by the accuracy that can be attained by the point charge model.

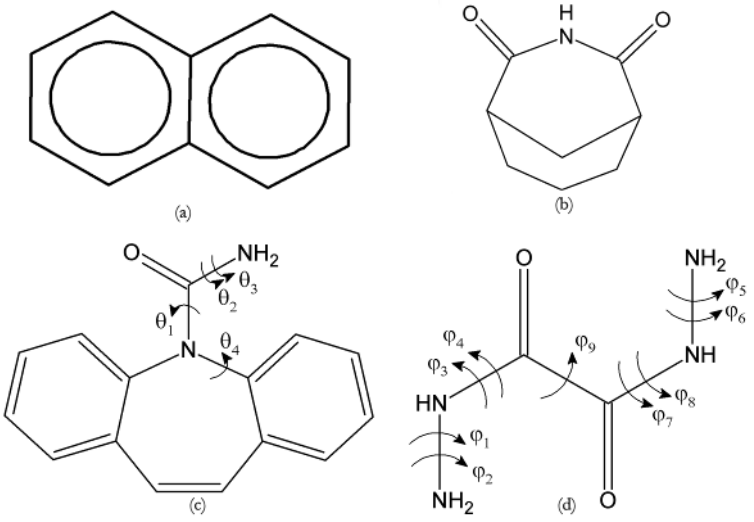
The WSM polarizability model has been validated for dimers against SAPT(DFT) energies.[2] SCERP provides an independent test of the polarizability models in the condensed phase. This paper uses these models to estimate the induction effects in a range of molecular crystals. Our purpose is to establish whether the contributions are sufficiently important that we should implement such models in the crystal structure modelling program DMAREL.[4,5]

The aim of this chapter is to determine the importance of the induction energy in organic crystal structures, particularly its relevance to the field of organic crystal structure and polymorph prediction,[6] for which promise for aiding the design of new materials and the selection of solid form for pharmaceutical development[7] is severely compromised by uncertainties in the estimation of relative lattice energies.

Four contrasting examples are considered for the effect of induction energy in their crystal lattices. The methods used to calculate the term are described in sections 4.2.5 and 4.2.6. The results are given in section 4.3 for the molecules naphthalene, oxalyl dihydrazide and carbamazepine, for which the molecular structures are shown in Figure 11. The naphthalene crystal is investigated as a non-polar system. Charge density studies[8] have shown a change in the electron distribution in the region of the C-H bond involved in a C-H $\cdots\pi$ interaction in the crystal structure. The other examples are all tests of the differences in induction energy corresponding to different types of hydrogen bonding, as the electrostatic fields involved in hydrogen bonding are amongst the strongest in crystal structures of neutral organic molecules. The relative induction energies of the 5 different polymorphs of oxalyl dihydrazide[9] are examined because of the plurality of hydrogen bonding geometries sampled, including one with a significant intramolecular component. The relative induction energies for sets of experimentally observed and hypothetical crystal structures of carbamazepine and 3-azabicyclo[3,3,1]nonane-2,4-dione are computed, to

investigate whether modelling induction could improve the prediction of relative lattice energies of crystal structures based on doubly hydrogen-bonded dimers or chain motifs. In both cases, the predictions that the two types of crystal structure were energetically competitive inspired extensive polymorph screening studies to search for the alternative motif.[10,11] For carbamazepine, all known polymorphs are based on a doubly hydrogen-bonded amide dimer (although it does adopt a catemer in a solid solution[12]), whereas the catemer is marginally more stable according to current modelling.[13-15] On the other hand, 3-azabicyclo[3,3,1]nonane-2,4-dione adopts an imide catemer in all its solid forms,[11] although many of the participants in the 2001 international blind test of crystal structure prediction[16] predicted a dimer structure as more stable.

Figure 11: Molecules used in this investigation. (a) Naphthalene, (b) 3-azabicyclo[3,3,1]nonane-2,4-dione, (c) carbamazepine and (d) oxalyl dihydrazide. Arrows indicate angles which have been refined by DMAFLEX, double arrows indicate that two atoms independently have a torsion angle defined along the same bond.



4.2. Method

4.2.1. Approximating the Crystalline Environment Using Clusters

In order to show how the induction energy may be significant in distinguishing between polymorphs, it was necessary to implement some model of the electrostatic fields experienced by a molecule in a lattice structure, so that the WSM polarizabilities could be used to determine the induced moments and therefore the induction energy. I achieved this by modelling the lattice as a finite cluster of whole unit cells, where a central molecule is polarizable.

The clusters are generated using the Crystal2Cluster tool, which generates the molecular translation and rotation parameters to describe a cluster of the required size, for the crystal structure of interest. Numerical experimentation has shown that a cluster in which a central molecule is surrounded to at least 15 Å in all directions is large enough to converge the electrostatic energy of a molecule in the centre, to that of an infinite lattice calculation using DMAREL. This typically means using a cluster of 5×5×5 unit cells. These clusters were then used with the WSM polarizabilities using the ORIENT program, and also used in the comparative SCERP scheme using Gaussian. These methods are outlined in the following sections.

4.2.2. Choice of Crystal Structures and Cluster and Molecular Models

For our calculations we use centrosymmetric crystal structures, from which the clusters are built. The crystal structure used for naphthalene was the 100 K X-ray structure.[8] The molecular structure was optimized *in vacuo* at the MP2/6-31G** level, and then pasted into the experimental structure by minimizing the RMS overlap of the carbon atoms. Finally the crystal structure was relaxed to a lattice energy minimum using DMAREL with a

distributed multipole model derived from a PBE0/Sadlej wavefunction calculation, and the FIT repulsion-dispersion potential. The clusters used in the induction energy calculations were of 5×5×5 unit cells.

For oxalyl dihydrazide, the five experimental crystal structures[9] were refined to account for the error in the X-ray determination of the proton positions that are important in the multiple hydrogen-bonds in these crystals.[16] This DMAFLEX refinement[13] optimized the lattice energy, including the MP2/6-31G** intramolecular conformational penalty, with respect to the nine torsions shown in Figure 11, and the crystallographic cell parameters and molecular positions using a distributed multipole model derived from an MP2/6-31G** wavefunction calculation, and the FIT potential. The conformational differences between the polymorphs are shown schematically in Figure 12 and in detail in Table 6. The cluster sizes were 9×7×5 unit cells for α and ε; 7×5×7 unit cells for γ and δ; and 9×5×9 unit cells for β polymorphs, to give suitable supercells containing between 490 and 980 molecules that conformed to our requirement of 15 Å of molecules surrounding the polarizable molecule.

Figure 12: The two major intramolecular conformations of oxalyl dihydrazide. The β, γ, δ and ε polymorphs contain stretched intramolecular hydrogen bonds, indicated by a dashed line. The torsion angles for all five polymorphs are given in Table 6.

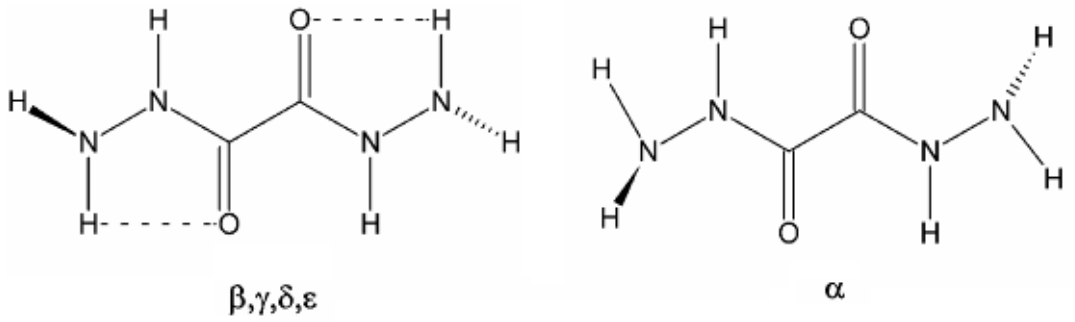


Table 6: The conformations of oxalyl dihydrazine within the models for the polymorphs used in this study, defined by the torsion angles indicated in Figure 11

polymorph	RMS[a] (Å)	conformational details								
		φ_1 (°)	φ_2 (°)	φ_3 (°)	φ_4 (°)	φ_5 (°)	φ_6 (°)	φ_7 (°)	φ_8 (°)	φ_9 (°)
α	0.207	160.04	-85.67	-179.29	-7.45	85.77	-160.30	180.24	6.31	-179.55
β	0.588	-75.90	38.13	185.31	-3.95	-42.95	71.37	-182.86	0.76	181.40
γ	0.188	-76.34	37.07	-182.09	-0.73	38.41	-72.27	174.40	0.28	-178.16
δ	0.500	93.58	-19.12	177.27	3.76	-93.09	19.88	-176.54	-1.790	179.638
ε	0.210	-97.60	14.41	-174.08	-2.66	97.85	-13.96	173.85	3.29	-180.20

[a] Root mean square discrepancy of the 15-molecule coordination sphere between the experimental and minimized crystal structure

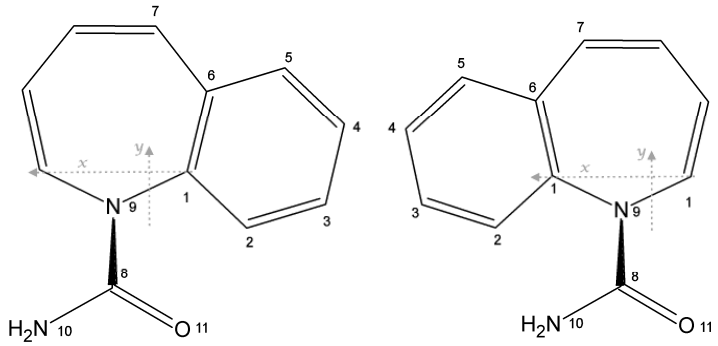
The bicyclic structure of 3-azabicyclo[3,3,1]nonane-2,4-dione[17] makes it essentially rigid, therefore we used the *in vacuo* MP2/6-31G** optimized molecular structure. A set of 6 low-energy crystal structures[11] generated using this molecular conformation were considered to represent a range of packing arrangements within 3 kJ mol^{-1} of the global minimum lattice energy. We also examined the minimum obtained by starting with the 297 K experimental crystal structure[17] and minimized with the same MP2/6-31G**/FIT model. The set of structures include both the observed catemer and doubly hydrogen bonded dimer motifs in a range of space groups. The 5×5×5 unit cell clusters contained 250, 500 or 1000 molecules.

For carbamazepine, we used 14 of the DMAFLEX[13] relaxed structures described in section 3.4 which had been low energy crystal structures in the rigid body search.[10] These structures covered a wide range of packing including those corresponding to known forms III and IV. The carbamazepine clusters used in the polarizability calculations consisted of 5×5×5 unit cells, and contained 250 to 1000 molecules.

4.2.3. Calculation of the Polarizabilities

The distributed atomic polarizabilities were calculated using the WSM method with the recommended asymptotically corrected PBE0 density functional and Sadlej basis set. For naphthalene, oxalyl dihydrazide and 3-azabicyclo[3,3,1]nonane-2,4-dione this was done for the whole molecule, held rigid at the conformation from the crystal structure. However, the carbamazepine molecule was too large for the WSM polarizability analysis due to computational limitations, which included hardware, i.e. the amount of RAM available to the machines at the time, software i.e. the size of the matrices on which the code code could operate, and a different scheme was adopted for this molecule. The distributed multipoles were calculated in GAUSSIAN03 using the non-asymptotically corrected PBE0 functional with the Sadlej basis set. The difference between the corrected and uncorrected functionals is insignificant for the calculation of electrostatic energies using distributed multipole moments, but the correction is essential for accurate polarizabilities. The polarizabilities were constructed from two molecular fragments, indicated in Figure 13. The geometries of the fragments were held rigid at the MP2/6-31** optimized geometry *in vacuo* of carbamazepine, except the positions of the hydrogen atoms added in place of the 6-membered ring, which were optimized at the same level of theory. Although polarizability is a molecular property, influenced by all sites, it has been necessary to make the approximation of transferability for the polarizabilities calculated for these smaller molecules to the larger structure. This approximation seems reasonable when comparing the polarizabilities for sites between the two fragments, in Table 9 page 119.

Figure 13: Fragments of carbamazepine used to calculate its atomic polarizabilities. The atom numbering is used to identify sites in Table 9 page 119, and indicates the polarizabilities used for the whole molecule.



The WSM polarizability models used in this chapter are identified as: L1, comprised of dipole-dipole polarizability terms only; L2, which is the L1 model plus dipole-quadrupole and quadrupole-quadrupole terms; and L2/1, which is a mixed model where all atoms have L2 terms except for hydrogen atoms, which are only represented by the L1 terms.

4.2.4. Calculating Induced Moments Using ORIENT Clusters

Once the supercell clusters have been generated and the polarizabilities calculated, an induction energy calculation can be performed. This is done by using the distributed multipole model for all molecules in the cluster, and by placing distributed polarizabilities on the atomic sites if a molecule at the centre of the cluster. ORIENT is able to calculate the electrostatic fields produced at each polarizable site by the surrounding molecules and evaluate the resulting induced moments. This is only the first step, and to continue the induced moments must be applied to all molecules in the cluster. Once this is done, the electrostatic fields due to the modified multipole model, now polarized, are used to re-

calculate the induced moments at the polarizable sites. This process is repeated until the induced moments converge, which are then used to calculate the induction energy of the crystal lattice. ORIENT implements the Tang-Toennies damping function, and unless stated as 'undamped' the following parameters were used: oxalyl dihydrazide forms α 1.625; β 1.667; δ 1.650; ε 1.649; γ 1.657; naphthalene 1.547; carbamazepine 1.510; 3-azabicyclo[3,3,1]nonane-2,4-dione 1.674.

4.2.5. Calculating Induced Moments Using Self-Consistent Electronic Response to Point Charges (SCERP)

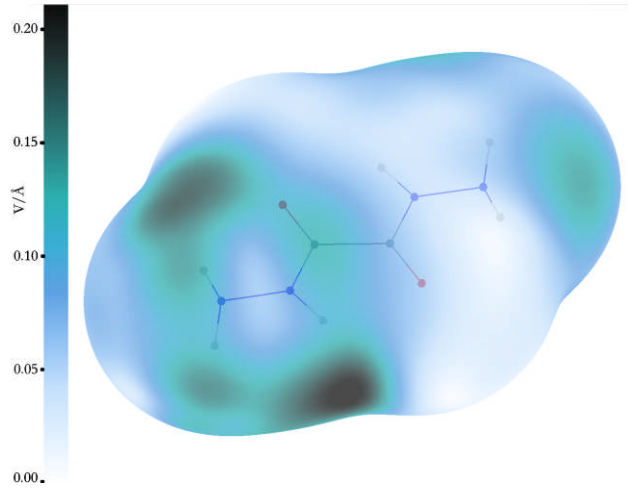
SCERP is an alternative method of evaluating the effect of induction on the charge distribution directly using the Gaussian03 *ab initio* package.[18] The CHELPG[3] potential derived charges, which are fitted to a grid of points between the van der Waals atomic radii and 2.8 Å from the nuclei, were obtained for the isolated molecule from an aug-cc-pVTZ charge density with the PBE0 functional. These charges were placed on all the atomic sites in the cluster modelling the crystal, except the central molecule. This molecule was described using aug-cc-pVTZ basis functions, and a PBE0 functional; the charge density of the molecule within the field of the cluster was calculated. This polarized charge density was analyzed by GMDA2.2[19] to obtain a set of multipoles that correspond to $Q_u^b + \Delta Q_u^b$, and hence the induced multipole moments (up to hexadecapole) obtained by subtraction of the multipoles obtained from the in vacuo wavefunction. The potential derived charges of the polarized charge distribution were then used in a further cluster calculation, and the process repeated until the calculated induction energy had converged.

This method is more computationally expensive than using the multipole expansion, and cannot be used for lattice energy minimization because there is no mechanism in place to relax the arrangement of the cluster to simulate the infinite crystal lattice, but can be used

for testing aspects of the polarizability model. The resources required are almost independent of the number of charges used, and so very large clusters could be used to check convergence with cluster size. Although the use of point charges to model electrostatic field is relatively crude, they are used here to induce a second-order response of the molecular charge density, and within the self-consistent nature of the process. A comparison of the molecular electrostatic field for charges and multipoles is made below. Penetration effects from the overlap of the charge distributions in the cluster are absent, except in so far as they are included in the fitting of the potential derived charges to points close to the molecule.

If the polar hydrogen sites are considered to have a van der Waals radius of zero,[20] the region of interaction with surrounding nuclei in hydrogen bonding arrangement may be approximated by the van der Waals surface scaled by 1.8. In Figure 14 I present a comparison of the electrostatic field at this surface when calculated using multipole moments or point charges, in terms of the norm of the difference vectors at 19814 points on the surface for α oxalyl dihydrazide. The mean difference is 0.08 V/Å (standard deviation 0.04 V/Å) which is less than 9 %, of the largest field, 0.92 V/Å, with the multipole moments. The highly localized nature of the error in the electrostatic field can be plainly seen in Figure 14, as dark blemishes around the hydrogen sites. For both the α and the ε polymorphs these regions coincide with the shortest hydrogen bonds seen in any of the crystal structures in this work, hence we anticipate the largest errors in our calculations to be for these crystals.

Figure 14: Error in the electrostatic field around α oxalyl dihydrazide. The plot shows the norm of the difference in the electrostatic field vectors, calculated from distributed multipole moments and point charges. The surface is the van der Waals surface scaled by 1.8, which is accessible by the hydrogen-bonding protons. The maximum field difference displayed is 0.226 V/Å.



4.2.6. Calculating the Induction Energy of a Crystal Lattice from Induced Moments

This classical polarization model for the induction energy is[1,21]

$$E_{ind,d-class}(A) = \tfrac{1}{2}\sum_{a\in A}\sum_{B\neq A}\sum_{b\in B}\sum_{tu}\Delta Q_t^a\, f_{(tu)}(\beta R_{ab})T_{tu}^{ab}Q_u^b\ , \tag{64}$$

where the omission of the superscript on the LHS implies that ΔQ_t^a are the converged induced moments. If the damping function is set to unity, this equation is almost identical to the expression for the electrostatic energy,

$$E_{electrostatic}(A) = \tfrac{1}{2}\sum_{a\in A}\sum_{B\neq A}\sum_{b\in B}\sum_{tu}Q_t^aT_{tu}^{ab}Q_u^b\ , \tag{65}$$

and this can be exploited to estimate the induction energy of the crystal using the routines already implemented in DMAREL[5] that evaluate this function and perform the lattice summations.

Equation (64) has only one molecule bearing just the induced moments interacting with the electrostatic field of the rest of the crystal. This cannot be directly calculated by DMAREL, which assumes that all symmetry related sites bear equal (or inverted) multipole moments. However, substituting $\left(Q_t + \Delta Q_t/2\right)$ into equation (65) gives

$$\begin{aligned}
&\tfrac{1}{2}\sum_{a\in A}\sum_{B\neq A}\sum_{b\in B}\sum_{tu}\left(Q_t^a + \frac{\Delta Q_t^a}{2}\right)T_{tu}^{ab}\left(Q_u^b + \frac{\Delta Q_u^b}{2}\right) \\
&= \tfrac{1}{2}\sum_{a\in A}\sum_{B\neq A}\sum_{b\in B}\sum_{tu}\left(Q_t^a T_{tu}^{ab} Q_u^b + \frac{\Delta Q_t^a T_{tu}^{ab} Q_u^b}{2} + \frac{\Delta Q_u^b T_{tu}^{ab} Q_t^a}{2} + \frac{\Delta Q_t^a T_{tu}^{ab} \Delta Q_u^b}{4}\right) \qquad (66) \\
&= E_{electrostatic}(A) + E_{ind,d-class}(A) + \Delta E_{error}(A).
\end{aligned}$$

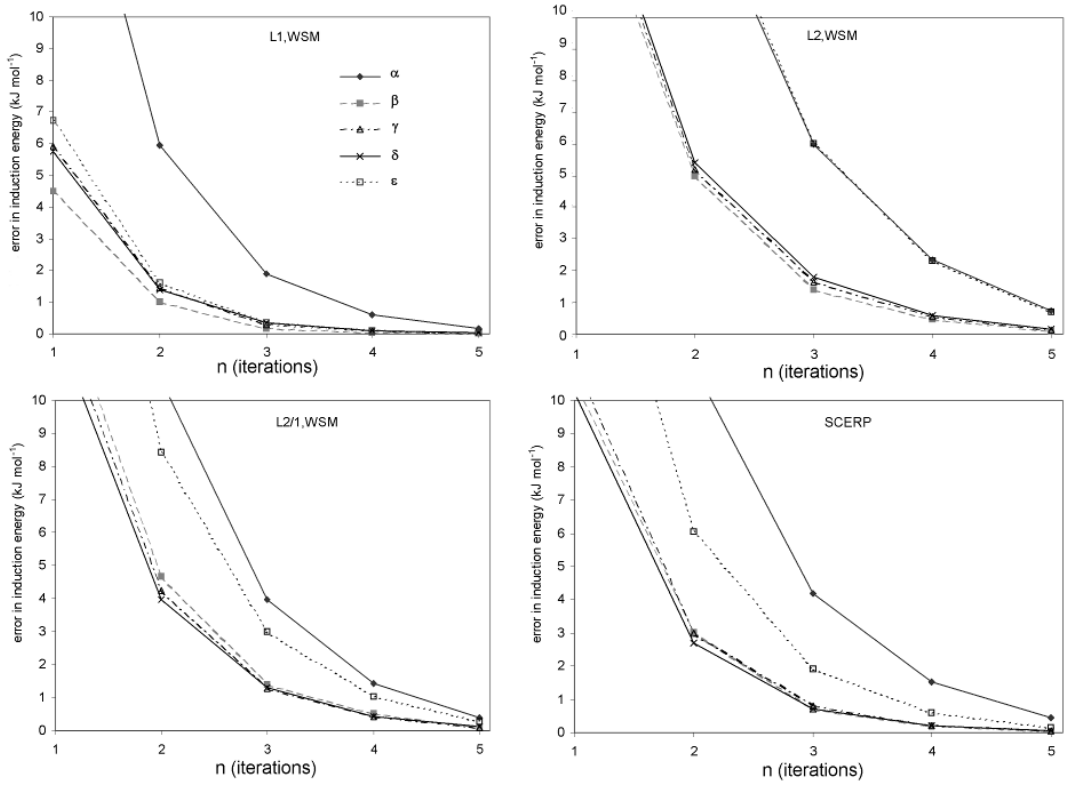
Thus, the induction energy can be calculated from three evaluations of the "electrostatic" contribution to the lattice energy, one where all molecules have the distributed multipole moments $\left(Q_t + \Delta Q_t/2\right)$ to get $E_{electrostatic} + E_{ind,d-class} + \Delta E_{error}$, a second with distributed multipole moments $\Delta Q_t/2$ to give ΔE_{error}, and a third using only Q_t to give $E_{electrostatic}$. All three evaluations use Ewald summation for the charge-charge, charge-dipole and dipole-dipole terms, and sum all the other contributions in direct space for all molecules whose centre of mass is within 15 Å. There is no facility to include damping of the electrostatic interactions in DMAREL, however the definition of the induction energy (64) requires that the damping function is included for each iteration of the interaction of induced and static multipole moments. The necessary damping is present for the iterative procedure which evaluates the induced moments in the cluster, but is not applied in the final energy calculation. This necessary approximation is made because of the current limitations of DMAREL. This method of evaluating the induction energy can only be performed at each given crystal structure, and cannot be used for optimizing the crystal structure.

4.3. Results

4.3.1. Effect of Iterations

The iteration procedure makes a significant difference to the calculated induction energy, stabilizing the crystal. Figure 15 shows how the induction energy converges over 5 iterative cycles, for three WSM models and SCERP. After only one iteration, the induction energy is at least 10 kJ mol^{-1} less than the converged value in most of the examples. It is notable that both the SCERP and polarizability induction models have very similar convergence properties, agreeing with the observation from modelling small molecules, that around half a dozen iterations are required for self-consistent polarization. The rank 1 model converges rapidly, but higher-ranking polarizabilities do require damping. In practice, Figure 15 shows that the infinite summation can be truncated, depending on the model, to 5-8 iterations, which is sufficient to achieve convergence within less than 0.5 kJ mol^{-1} for the systems studied.

Figure 15: Convergence of $E_{ind,d-class}$ for polymorphs of oxalyl dihydrazide for several induction energy models. The plot shows the error, $E_{ind,d-class}^{(2-n+1)} - E_{ind,d-class}^{(2-\infty)}$, in the induction for different truncations of the infinite sum.



4.3.2. Naphthalene

Induction is important not only for hydrogen-bonded systems. The crystal structure of naphthalene has been previously analyzed for experimental evidence of induced changes in the charge density.[8] The SCERP point charge model predicts an induction energy of −1.9 kJ mol^{-1} for the 100 K experimental crystal structure, using the molecular geometry optimized *in vacuo*. Although small in absolute terms, this is 31% of the electrostatic energy. A damped WSM2/1 polarizability model estimates the induction energy to be 25% of the electrostatic energy. In comparison, the SCERP induction energy for oxalyl dihydrazide polymorphs is 18–38% of the electrostatic energy. Thus, in relative terms,

even the charge density of naphthalene is significantly affected by the surrounding molecules in the lattice. By analyzing the change in electrostatic energy due to the induced moments interacting with a unit charge probe, we may indirectly observe the change in charge distribution caused by the crystalline environment.

Figure 16: Induced electrostatic energy surface for naphthalene. The energy is calculated from the SCERP model, for the van der Waals + 1.1 Å surface that is accessible by short-contact nuclei. The atom numbering system reflects the symmetry of contacts within the crystal structure, not of the isolated molecule. The energy is calculated using a unit charge probe, and ranges from -5.23 kJ mol-1 to +6.82 kJ mol-1.

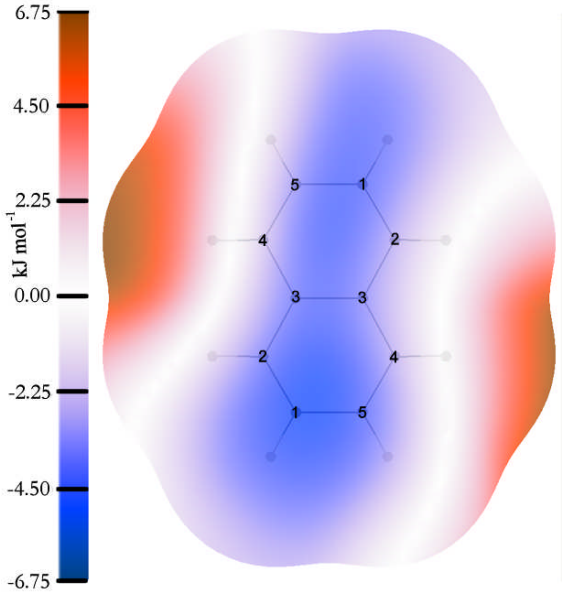


Figure 16 show a plot of the change in the electrostatic energy using the SCERP induced moments, on the van der Waals plus 1.1 Ångstrom surface that is sampled by the atomic sites of the surrounding molecules. The anisotropic nature of the induction is clear. The increased electrostatic potential around the C(4)-H bond, in contrast to the C(2)-H bond, shows that the close contact with the π electron cloud of the surrounding molecules in the crystal has significantly polarized this bond, as observed in the experimental charge density.[8]

4.3.3. Oxalyl dihydrazide

The RMSD[22] between the refined structures and the experimental crystal structures was about 0.2 Å for the α, γ and ε polymorphs and less than 0.6 Å. for β and δ, for all non-hydrogen atoms in a 15-molecule cluster (Table 6). The model for the ε polymorph is more dense that the experimental structure, resulting in one short N···N distance of 2.73 Å. A comparison of the intermolecular electrostatic energy for all of clusters of oxalyl dihydrazide is within 0.5 kJ mol^{-1} of the infinite lattice value, shown in Table 7.

Experimentally, it has been difficult to fully characterize the relative stability of these polymorphs of oxalyl dihydrazide, due a self-reaction that takes place prior to melting.[9] However, lattice-energy methods that only model the intermolecular repulsion, dispersion and electrostatic forces, including the conformational energy differences from *ab initio* gas phase calculations, predict that the lattice energy of the α form is approximately −110 kJ mol^{-1}, whereas the other four forms range from −130 to −138 kJ mol^{-1} (Table 7). Such a large energy difference is considered to be outside the range of possible polymorphism.[23]

Table 7: Intermolecular electrostatic and total lattice energy (kJ mol^{-1}) of oxalyl dihydrazide polymorphs

	$E_{electrostatic}$ [a]		ΔE_{intra} [b]	$E_{crystal}$ [c]
	Cluster	DMAREL		
α	-157.27	-156.63	56.34	-110.219
β	-128.34	-128.25	1.11	-138.326
γ	-127.65	-127.60	4.37	-135.729
δ	-122.77	-123.35	6.87	-130.216
ε	-148.53	-148.49	7.52	-137.778

[a] The electrostatic model used is distributed multipole moments derived from the PBE0 Sadlej charge density; [b] Intramolecular energy calculated at the MP2/6-31G** level, referenced to the energy of the *in vacuo* structure, optimized at the same level of theory; [c] Sum of the DMAREL electrostatic energy, exp-6 respulsion-dispersion energy, and intramolecular energy.

By including a correction for the induction energy of the lattice, the predicted lattice energy of the α form becomes comparable with that of the β, γ and δ polymorphs. It seems apparent that modelling charge density polarization for polymorphs that exhibit different intra- and intermolecular hydrogen bonding is important. This issue has been explored further using electronic structure calculations on oxalyl dihydrazide and other polymorphic systems.[16]

Table 8: Lattice parameters of oxalyl dihydrazide structures used in this work, and their relation to experimental values.

	Lattice parameter	DMAFLEX	Abs diff[a]	% diff
α	a / Å	3.5158	-0.1063	-2.93 %
	b/ Å	6.6530	-0.1792	-2.62 %
	c/ Å	9.1963	0.0669	0.73 %
	β /°	102.2718	2.9738	2.99 %
β	a/ Å	4.0455	0.2835	7.54 %
	b/ Å	10.8263	-0.8257	-7.09 %
	c/ Å	4.9620	-0.6570	-11.69 %
	β /°	90.9521	-1.8409	-1.98 %
γ	a/ Å	5.0710	-0.0085	-0.17 %
	b/ Å	14.0052	-0.6627	-4.52 %
	c/ Å	6.9463	-0.0882	-1.25 %
	β /°	114.7818	0.6218	0.54 %
δ	a/ Å	3.9710	0.3102	8.47 %
	b/ Å	13.2728	-1.2772	-8.78 %
	c/ Å	5.0628	-0.0018	-0.04 %
	β /°	123.4397	4.4337	3.73 %
ε	a/ Å	5.1815	-0.1827	-3.41 %
	b/ Å	3.7805	-0.0607	-1.58 %
	c/ Å	11.7785	-0.5406	-4.39 %
	β /°	107.9730	-1.0260	-0.94 %

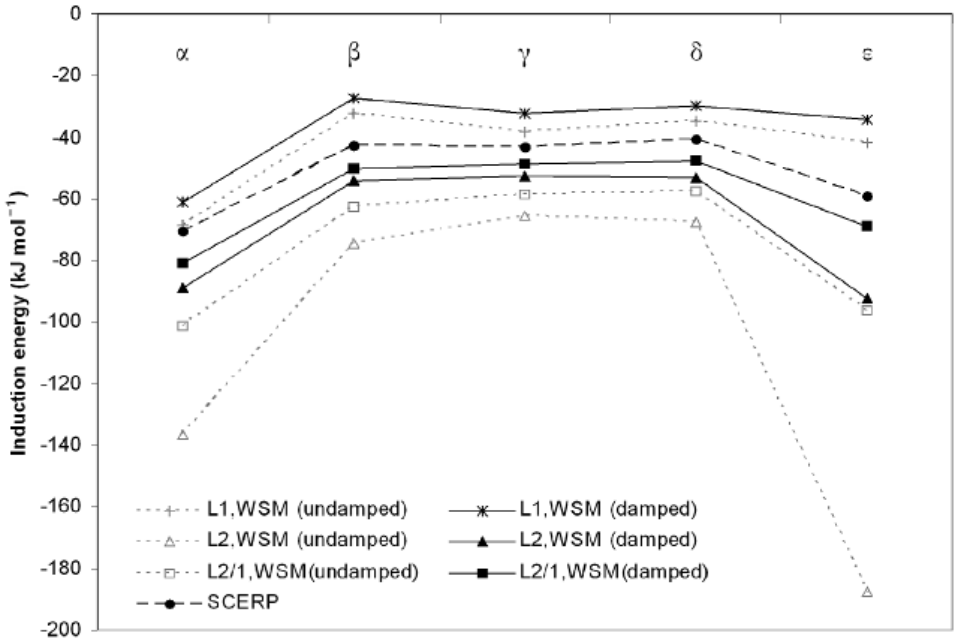
[a] Absolute difference is DMAFLEX − experimental

The energies calculated using SCERP with WSM models are shown in Figure 17. Using the SCERP model the α structure is stabilized the most, followed by ε. The β, γ and δ

structures are stabilized less than the ε form, but by similar magnitudes to one another. Each of the WSM models follow the SCERP results except the L2 models, which find the relative polarization of the structures to be ε>α>>β, γ, δ. This deviation can be explained by examining the crystal structures.

In these five polymorphs there are very short intermolecular contacts, the shortest being an N···H-N contact in ε of 1.77 Å, with a N···N distance of 2.73 Å, which are significantly less than the sum of their van der Waals radii and somewhat less than their experimental values, considerably increasing the uncertainty in the induction energies for this particular polymorph. In the α form, the situation is better with N···H-N distances 1.83 and 2.80 Å respectively. For these two systems in particular, the effect of having higher polarizability terms on the hydrogen sites is to give implausible induction energies when damping is not included, but this is resolved by removing the higher terms on hydrogen, leaving only rank 1 polarizabilities on these sites. The other three forms all have X···H distances greater than 1.85 Å and do not show divergence of the induction energy. The results relating to oxalyl dihydrazide strongly suggest that an iterated, damped polarizability model, based on L1 model or mixed L2/L1 model, agrees reasonably well with the self-consistent electronic response to point charges method.

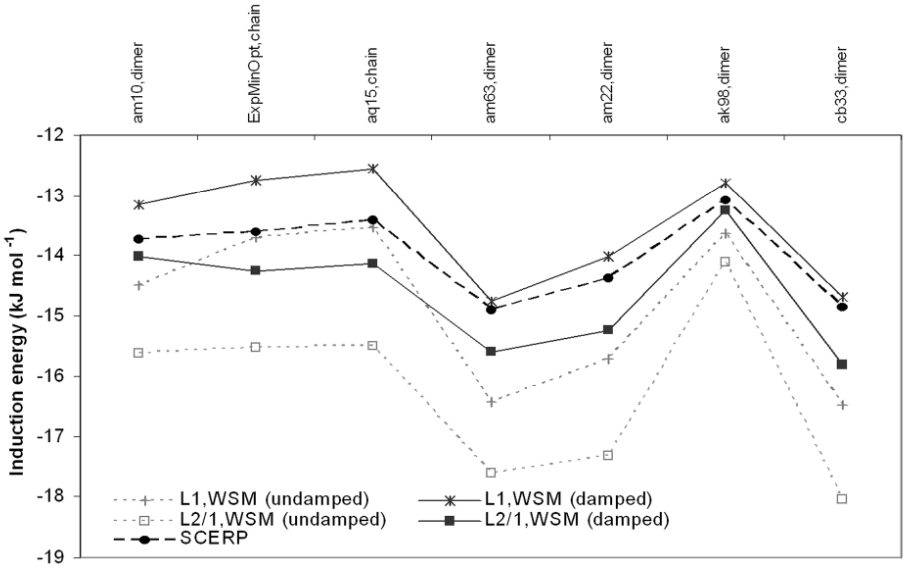
Figure 17: The induction energy of oxalyl dihydrazide for various WSM polarizability schemes.



4.3.4. 3-Azabicyclo[3,3,1]nonane-2,4-dione

3-Azabicyclo[3,3,1]nonane-2,4-dione presents several challenges in terms of our polarizability calculations: the size of the molecule, in terms of basis functions required and associated computational limits, as well as the volume of space to be sampled for the point-to-point polarizabilities and the C2 symmetry in the molecule. Despite this, and the fact that symmetry of the molecules is not explicitly enforced by SITUS at any stage, after refinement and localization the resulting polarizabilities are reassuringly symmetric.

Figure 18: Induction energies for 3-azabicyclo[3,3,1]nonane-2,4-dione. The crystal structures are ordered left-to-right by decreasing lattice stability, as calculated from the distributed static multipole + empirical repulsion-dispersion potential.



We find the induction energy for 3-azabicyclo[3,3,1]nonane-2,4-dione to be 33-36% of the electrostatic energy, with good agreement between SCERP and the damped WSM models (Figure 18). For the crystal structures considered, the induction energy varies by less than 3 kJ mol^{-1}, but this is significant relative to the difference in lattice energies of these structures calculated using a repulsion-dispersion model potential,[24] which range from −95.08 to −97.64 kJ mol^{-1}.

Hence, more realistic modelling of the intermolecular interactions to include the induction energy would certainly re-rank the structures. However, the observed chain-hydrogen bonding motif is not favoured relative to many of the competitive dimer structures, and there is no clear-cut correlation with the hydrogen-bonding motif. Hence, neglect of the induction energy does not appear to be the only problem in modelling the relative stability of crystal structures of 3-azabicyclo[3,3,1]nonane-2,4-dione. It may be that the empirical

repulsion-dispersion potential does not model the interactions between these molecules very accurately, or that there is some kinetic aspect to the crystal growth which is not accounted for by the models used.

4.3.5. Carbamazepine

The rank 1 polarizabilities which were calculated for the fragments of carbamazepine in Figure 13 are listed in Table 9. The environment of the carbon atoms 1-8 are largely the same in each case, except for the spacial proximity to the nitrogen or oxygen of the imide group. The polarizabilities reflect this by being very similar for all labelled atoms. There is a change of sign in the off-diagonal terms which follow the rules for inversion shown in Table 10.

4.3.5.1. Re-Ranking Due to Induction

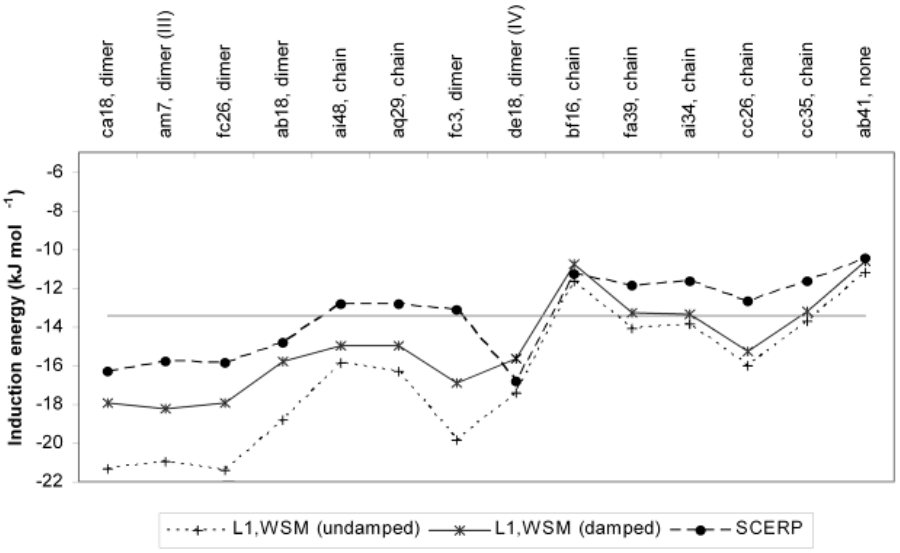
Despite the additional assumptions, there is still reasonably good agreement in the relative induction energies between SCERP and the damped L1 polarizability model, accounting for an increase in stability of 10.5 – 18.2 kJ mol^{-1} in the lattice energy. Both models find that the dimer-based structures, and particularly the experimental forms III and IV, are stabilized more by induction than the chain-based structures, and all hydrogen bonded structures more than the structure (ab41) with no hydrogen-bonding (Figure 9). This is significant, as the published crystal structure predictions[10] for carbamazepine found that a structure with a hydrogen-bonded chain motif was more stable than the experimentally known dimer based structures. Improving the modelling of the electrostatic energies by using distributed multipoles from the better charge distribution used in the current work also alters the relative stabilities, favouring the most stable observed polymorph form III. Hence, more accurate modelling of the electrostatics and adding the induction clearly gives

a significant energy lowering to the most stable dimer based structures, which is in accord with experiment.

Table 9: Localized dipole-dipole polarizabilities derived from the two molecular fragments from carbamazepine (Figure 13). Italicized values were not used in the work.

		CO fragment			NH2 fragment		
	Q_{ls}	10	11c	11s	10	11c	11s
C1	10	8.40964	-6.38059	-0.63136	8.32961	5.37695	-0.28298
	11c	-6.38059	18.2973	-0.18501	5.37695	18.7080	-0.88010
	11s	-0.63136	-0.18501	14.3840	-0.28298	-0.88010	12.5378
C2	10	9.40761	-6.11886	-0.96389	8.36570	5.27708	-0.96397
	11c	-6.11886	15.5801	0.23032	5.27708	15.0561	-0.64310
	11s	-0.96389	0.23032	5.62917	-0.96397	-0.64310	5.57362
C3	10	6.57784	-3.54568	-0.05483	7.23791	3.32652	-0.12180
	11c	-3.54568	12.2044	-1.58251	3.32652	12.0866	1.86013
	11s	-0.05483	-1.58251	10.5551	-0.12180	1.86013	10.6218
C4	10	6.33095	-2.05589	-0.58270	5.91204	2.81588	-1.52775
	11c	-2.05589	8.83260	1.23153	2.81588	9.58192	-0.30910
	11s	-0.58270	1.23153	13.1491	-1.52775	-0.30910	12.7272
C5	10	8.09373	-4.58729	-0.46761	7.14971	4.31559	-0.94157
	11c	-4.58729	13.6670	0.56426	4.31559	13.2160	-0.38222
	11s	-0.46761	0.56426	8.10629	-0.94157	-0.38222	9.23846
C6	10	5.29561	-6.37881	-0.66653	6.18645	5.56108	-0.38540
	11c	-6.37881	18.0670	-1.03093	5.56108	16.8127	-0.13428
	11s	-0.66653	-1.03093	14.8632	-0.38540	-0.13428	14.3687
C7	10	5.22380	-1.15772	-1.14327	*5.69914*	*0.65990*	*-1.12674*
	11c	-1.15772	14.6161	1.05936	*0.65990*	*14.9953*	*-1.89426*
	11s	-1.14327	1.05936	10.8107	*-1.12674*	*-1.89426*	*10.2504*
C8	10	8.63517	-2.51614	4.79658	*8.25499*	*0.96267*	*4.73088*
	11c	-2.51614	9.14583	-0.91720	*0.96267*	*9.45985*	*1.24418*
	11s	4.79658	-0.91720	6.00270	*4.73088*	*1.24418*	*4.97555*
N9	10	5.32670	-0.79347	2.83959	*4.96203*	*1.54720*	*3.07380*
	11c	-0.79347	11.5588	0.88526	*1.54720*	*11.5141*	*-0.89165*
	11s	2.83959	0.88526	6.61628	*3.07380*	*-0.89165*	*5.87135*
N10	10	5.69640	-0.74738	1.34147	5.61351	-0.93765	1.34076
	11c	-0.74738	8.04773	-2.45118	-0.93765	6.18010	0.44634
	11s	1.34147	-2.45118	5.52778	1.34076	0.44634	5.82469
O11	10	8.36953	1.78883	1.82821	7.49731	1.29836	0.99698
	11c	1.78883	6.09074	1.02358	1.29836	8.08473	1.59716
	11s	1.82821	1.02358	7.80493	0.99698	1.59716	7.55187

Figure 19: Induction energies for crystal structures of carbamazepine. The structures are ordered left-to-right by decreasing lattice stability, as calculated from the distributed multipoles described, plus an empirical[24] repulsion-dispersion potential. The lattice-energy range for the structures shown is 16 kJ mol^{-1}. The horizontal line indicates the average induction energy with the SCERP model to illustrate the discrimination of structural motifs by the polarizability model.



4.4. Conclusion

4.4.1. How Important is the Induction Energy for Organic Crystals?

Two very different models for estimating the induced moments in organic crystals have been compared: an *ab initio* response to an applied field due to point charges representing the crystal environment, and the use of distributed polarizabilities in the field arising from a distributed multipole representation of the surrounding molecules. The induction energy contribution to the lattice energy, evaluated from these induced moments, is significant. Over this diverse range of crystal structures, the models agree that the induction energy is often between 20 – 40% of the electrostatic contribution to the lattice energy. This order of magnitude is consistent with estimates of the induction energy relative to the electrostatic

energy for small polyatomic molecules,[25-28] using equally rigorous or better models for the induction energy, although often the polarization is not iterated to self-consistency. It is also comparable with the modelling of the induction contribution to lattice energies of neutral organic molecules derived by the PIXEL method,[29-31] where experimental atomic polarizabilities are evenly distributed over the atomic charge density.

More importantly, for different known and predicted crystal structures that otherwise have very similar lattice energies, the two models agree on the relative magnitude of the induction energy. In the case of oxalyl dihydrazide, including the intermolecular induction is essential for the calculated relative lattice energies to be consistent with the experimental observation of the polymorphs. This is an extreme case, as the intermolecular induction for the α polymorph compensates for the intramolecular hydrogen bonding in the other conformational polymorphs. In the case of carbamazepine, the induction energy favours the observed doubly hydrogen bonded dimer based structures over the hypothetical catemer based structures. The differences in the induction energies of the low energy computed structures of 3-azabicyclo[3,3,1]nonane-2,4-dione cannot be so simply ascribed to the hydrogen bonding motif, but this reflects the relative weakness of the hydrogen bonds for this imide, which forms a plastic phase.[35] In each of these comparisons of known and hypothetical crystal structures, the differences in induction energies are small, only a few kilojoules per mole, but this is sufficient to provide a significant reordering of the relative stability of structures that are virtually equi-energetic according to models which do not explicitly model the induction.

To correctly model crystal lattice energy, intermolecular potentials require a re-parameterization of the entire repulsion-dispersion potential: adding the induction energy to lattice energies calculated using an empirically fitted potential involves a high degree of double counting. This is sufficient to lead to structures which are too dense if we attempt

to minimize crystal structures with an induction term in addition to potentials which have been empirically-fitted without the explicit inclusion of induction. It is also important that the model for the induction can be readily implemented in a program that minimizes lattice energies of organic crystal structure.

4.4.2. Practical Consideration for Using Polarizability Models in the Organic Solid State

A local polarizability model can be implemented in lattice energy minimization packages that use distributed multipole moments. It appears to be feasible to calculate WSM polarizabilities from a reasonable quality *ab initio* charge density for quite large molecules, with 3-azabicyclo[3,3,1]nonane-2,4-dione probably being the limit with current resources. This is an acceptable limitation, given that the transferable polarizability model calculated from fragments of carbamazepine gave reasonable results compared with the SCERP calculations that used the complete molecule. Thus, it seems that transferable polarizability models could be derived for use in modelling larger molecules.

The induction energy does depend on the order of the polarizabilities included. We have noted some anomalous behaviour where rank 2 polarizabilities are used on hydrogen, particularly when involved in short contacts within the crystal structure (most notably on oxalyl dihydrazide ε). Given the small amount of charge density associated with polar hydrogen atoms, it seems reasonable that polarizabilities for these sites should be limited to rank 1 for applications to dense systems. The differences between L2 and L1 WSM models for the other atoms are comparable to those between them and the SCERP model.

The error in modelling charge overlap effects in particularly short hydrogen-bonding geometries probably explains the larger variance with polarizability model observed in our

oxalyl dihydrazide induction energies, relative to those for the 3-azabicyclo[3,3,1]nonane-2,4-dione crystals which do not have such short contacts. The WSM polarizability model does not account for any density overlap effects, and we have shown that damping is required in order to avoid unreasonable energies for the shorter intermolecular contacts found in hydrogen-bonded crystal structures. We have already shown[2] the agreement between SAPT(DFT) induction energies and WSM models to be very good, and so the WSM polarizability method of modelling the induction energy has a firm foundation. This investigation has shown that damped polarizability models are also suitable for modelling the induction energy in large clusters representing crystals, with many-body effects, qualitatively different field anisotropy and short contacts.

We consistently find that the *ab initio* SCERP model falls midway between the L1 and L2/L1,WSM models, and that the relative ordering of the energies is consistent. The SCERP model has also approximated the electrostatic field around the molecules (Figure 5), which does lead to significant errors in the hydrogen bonding region. Thus, we conclude that we cannot at present model the induction energy more accurately than the range indicated by the differences between the SCERP and the L1 and L2/L1 WSM models. However, it is clear that the induced moments will need iterating to self-consistency (Figure 15).

Reference List

1. Misquitta, A. J.; Stone, A. J. Accurate induction energies for small organic molecules: 1. Theory. *J. Chem. Theory Comput.* **2008,** *4* (1), 7-18.
2. Misquitta, A. J.; Stone, A. J.; Price, S. L. Accurate induction energies for small organic molecules. 2. Development and testing of distributed polarizability models against SAPT(DFT) energies. *J. Chem. Theory Comput.* **2008,** *4* (1), 19-32.
3. Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles From Molecular Electrostatic Potentials - the Need For High Sampling Density in Formamide Conformational-Analysis. *J. Comput. Chem.* **1990,** *11* (3), 361-373.

4. Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. The Relaxation of Molecular Crystal Structures Using a Distributed Multipole Electrostatic Model. *J. Comput. Chem.* **1995,** *16* (5), 628-647.

5. *DMAREL 3.02,* 2001

6. Gavezzotti, A. Ten years of experience in polymorph prediction: what next? *CrystEngComm* **2002,** *4* (61), 343-347.

7. Price, S. L. The computational prediction of pharmaceutical crystal structures and polymorphism. *Adv. Drug Deliver. Rev.* **2004,** *56* (3), 301-319.

8. Oddershede, J.; Larsen, S. Charge density study of naphthalene based on X-ray diffraction data at four different temperatures and theoretical calculations. *J. Phys. Chem. A* **2004,** *108* (6), 1057-1063.

9. Ahn, S. Y.; Guo, F.; Kariuki, B. M.; Harris, K. D. M. Abundant polymorphism in a system with multiple hydrogen-bonding opportunities: Oxalyl dihydrazide. *J. Am. Chem. Soc.* **2006,** *128* (26), 8441-8452.

10. Florence, A. J.; Johnston, A.; Price, S. L.; Nowell, H.; Kennedy, A. R.; Shankland, N. An automated parallel crystallisation search for predicted crystal structures and packing motifs of carbamazepine. *J. Pharm. Sci.* **2006,** *95* (9), 1918-1930.

11. Hulme, A. T.; Johnston, A.; Florence, A. J.; Fernandes, P.; Shankland, K.; Bedford, C. T.; Welch, G. W. A.; Sadiq, G.; Haynes, D. A.; Motherwell, W. D. S.; Tocher, D. A.; Price, S. L. Search for a predicted hydrogen bonding motif - A multidisciplinary investigation into the polymorphism of 3-azabicyclo[3.3.1]nonane-2,4-dione. *J. Am. Chem. Soc.* **2007,** *129* (12), 3649-3657.

12. Florence, A. J.; Leech, C. K.; Shankland, N.; Shankland, K.; Johnston, A. Control and prediction of packing motifs: a rare occurrence of carbamazepine in a catemeric configuration. *CrystEngComm* **2006,** *8* (10), 746-747.

13. Karamertzanis, P. G.; Price, S. L. Energy Minimization of Crystal Structures Containing Flexible Molecules. *J. Chem. Theory Comput.* **2006,** *2* (4), 1184-1199.

14. Cabeza, A. J. C.; Day, G. M.; Motherwell, W. D. S.; Jones, W. Amide pyramidalization in carbamazepine: A flexibility problem in crystal structure prediction? *Cryst. Growth Des.* **2006,** *6* (8), 1858-1866.

15. Cabeza, A. J. C.; Day, G. M.; Motherwell, W. D. S.; Jones, W. Importance of molecular shape for the overall stability of hydrogen bond motifs in the crystal structures of various carbamazepine-type drug molecules. *Cryst. Growth Des.* **2007,** *7* (1), 100-107.

16. Karamertzanis, P. G.; Day, G. M.; Welch, G. W. A.; Kendrick, J.; Leusen, F. J. J.; Neumann, M. A.; Price, S. L. Modeling the interplay of inter- and intramolecular hydrogen bonding in conformational polymorphs. *J. Chem. Phys.* **2008,** *128* (24), art-244708.

17. Howie, R. A.; Skakle, J. M. S. 3-Azabicyclo[3.3.1]nonane-2,4-dione: CSP2001 structure prediction test case No. 1. *Acta Crystallogr. , Sect. E* **2001,** *57*, o822-o824.

18. *Gaussian 03,* Gaussian Inc.: Wallingford CT, 2003

19. Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005,** *1* (6), 1128-1132.

20. Buckingham, A. D.; Fowler, P. W. A Model for the Geometries of Vanderwaals Complexes. *Can. J. Chem.* **1985,** *63* (7), 2018-2025.

21. Stone, A. J. *The Theory of Intermolecular Forces;* 1st ed.; Clarendon Press: Oxford, 1996.

22. Chisholm, J. A.; Motherwell, S. COMPACK: a program for identifying crystal structure similarity using distances. *J. Appl. Crystallogr.* **2005,** *38*, 228-231.

23. Bernstein, J. *Polymorphism in Molecular Crystals;* Clarendon Press: Oxford, 2002.

24. Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *J. Phys. Chem.* **1996,** *100* (18), 7352-7360.

25. Chipot, C.; Luque, F. J. Fast evaluation of induction energies: a second-order perturbation theory approach. *Chem. Phys. Lett.* **2000,** *332* (1-2), 190-198.

26. Chipot, C.; Angyan, J. G. Continuing challenges in the parametrization of intermolecular force fields. Towards an accurate description of electrostatic and induction terms. *New J. Chem.* **2005,** *29* (3), 411-420.

27. Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. Transferable ab initio intermolecular potentials. 1. Derivation from methanol dimer and trimer calculations. *J. Phys. Chem. A* **1999,** *103* (48), 9872-9882.

28. Jansen, G.; Hattig, C.; Hess, B. A.; Angyan, J. G. Intermolecular Interaction Energies By Topologically Partitioned Electric Properties. 1. Electrostatic and Induction Energies in One- Center and Multicenter Multipole Expansions. *Mol. Phys.* **1996,** *88* (1), 69-92.

29. Gavezzotti, A. Calculation of lattice energies of organic crystals: the PIXEL integration method in comparison with more traditional methods. *Z. Kristallogr.* **2005,** *220* (5-6), 499-510.

30. Gavezzotti, A. Calculation of intermolecular interaction energies by direct numerical integration over electron densities. I. Electrostatic and polarization energies in molecular crystals. *J. Phys. Chem. B* **2002,** *106* (16), 4145-4154.

31. Gavezzotti, A. Calculation of intermolecular interaction energies by direct numerical integration over electron densities. 2. An improved polarization model and the evaluation of dispersion and repulsion energies. *J. Phys. Chem. B* **2003,** *107* (10), 2344-2353.

Chapter 5. Implementing the Polarizability Model in a Crystal Structure Modelling Code

5.1. Introduction

Following the work done in chapter 4 to test the WSM model for calculating the effect of polarizability in clusters of organic molecules, it is desirable to be able to use it when calculating the crystal lattice energy with DMACRYS. This chapter describes how I have implemented the L1 WSM model into the existing code, taking advantage of the program features that already exist and adding the damping function. Results with systems such as oxalyl dihydrazide and naphthalene, show that the dipolar polarizability model is worth implementing for a wide range of organic systems. Extending the model to include quadrupolar terms would dramatically increase the computational expense of the energy calculation and require the additional separation and storage of the relevant field contributions from the electrostatic energy summation. This is a fairly large task, but the framework for it is in place alongside the dipole implementation.

It has been shown in chapter 4 that it is necessary to iterate the calculation of the induced moments until the induction energy converges. Previously this has been performed using large clusters of molecules generated from the crystallographic unit cell. The electric field due to this cluster was evaluated at the atomic sites of one molecule in the central unit cell, and polarizabilities on these sites give the corresponding induced moments. By means of a script, these were captured from the output and added to the static moments on the surrounding molecules so as to recalculate the electric field of the same physical structure with updated multipole moments. The converged induced moments were then used to

evaluate the induction contribution to the lattice energy. This procedure cannot be used to minimize the lattice, however, and is only good for the induction at a single structure.

The iterative procedure has been coded into DMACRYS, which is described below in section 5.4, and calculates the induced moments to update the static multipoles internally. The Tang-Toennies damping function, equation (54), used in the ORIENT cluster calculations in section 4.3, has also been coded into DMACRYS. This damping function is not only used for the induction, but is known to be effective in damping the dispersion interaction[1]. There is the option to use this for completely *ab initio* potentials including damped dispersion in the future.

The first step in using the WSM model is to be able to calculate the induction energy for a static crystal lattice. The procedure was designed so it would be possible to optimize crystal structures using numerical gradients for induction when using a non-empirical potential. Aside from the lack of non-empirical potentials to avoid double counting the induction energy, there are some technical issues that have hampered my attempts to relax a crystal structure with induction, which are discussed in chapters 6 and 7.

5.2. Data Input and Inversion Symmetry

The input for the DMACRYS code is minimally changed when an induction calculation is required, NEIGHIND requests an additional input file which contains the atomic polarizabilities. The best practice is to calculate the distributed multipoles and polarizabilities from the same wavefunction using CamCASP[2], and they must be expressed in the same molecular axis system to be input into DMACRYS. The format of the polarizability file is directly related to the input file containing the multipole moments, and

is purposefully generated by a similar set of procedures as the existing multipole file so as to be consistent for the user. The polarizability input file for the first three atoms of the blind test molecule XIII (Figure 23) are shown in Figure 20 as an example. NEIGHIND matches up all of the properties for each atom and the output is a DMACRYS input file.

Figure 20: Example of the polarizability input file format. Each entry contains four lines. The first line consists of the atom number, atom label, x y z Cartesian coordinates relative to the molecular centre of mass, and the rank of polarizabilities input. The remaining three lines are the lower triangle of the symmetric polarizability matrix.

```
1    BrBR_1____       -0.957846 -5.432957 -0.000009       RANK  1
 26.27810
  0.00166   17.97020
  2.04770   -0.00047   17.39480

2    BrBR_2____       -0.958960  5.432649  0.000066       RANK  1
 26.27810
  0.00166   17.97020
  2.04770   -0.00047   17.39480

3    ClCL_1____       -3.895727  0.000047 -0.000040       RANK  1
 17.30710
 -0.00058   11.69060
 -0.71240    0.00050   12.30990
```

When a molecule is generated by an inversion centre, some of the properties must undergo a sign change so that the right-handed molecule-fixed axis system is preserved. NEIGHIND already handles this for the multipole moments, changing the sign of those with an odd order in the z-axis (e.g. Q_{10} = z, Q_{21s} = xz, Q_{21s} = yz, etc.). For the polarizabilities this is also necessary, and I have ensured that those with an odd order in the z-axis are inverted. The inversion matrix for dipolar and quadruplar polarizabilities is given in Table 10, where the presence of a '—' in the dipolar terms correspond to the italicised columns in Table 1.

NEIGHIND was programmed so it could process the quadrupolar polarizability terms correctly, however DMACRYS has only been modified to use dipolar polarizabilites.

Table 10: Those polarizabilities, α, that require a sign changes when the molecule is inverted are marked by a dash.

α_{tu}		Dipolar			Quadrupolar				
	u \ t	10	11c	11s	20	21c	21s	22c	22s
Dipolar	10		–	–	–			–	–
	11c	–				–	–		
	11s	–				–	–		
	20	–				–	–		
Quadrupolar	21c		–	–	–			–	–
	21s		–	–	–			–	–
	22c	–				–	–		
	22s	–				–	–		

5.3. Calculating Electrostatic Fields

DMACRYS already contains code to maintain a list of molecules whose centres of mass fall within a defined cut-off radius. This list is processed by the routine `frcnst` such that all pair-pair interactions within that range are summed over, to give the energy of the system. The electrostatic interaction energy (36) for a pair of molecules is summed over the lattice of molecules. By recasting the expression to use the electrostatic field at an interaction site due to the surrounding molecules, we have the expression

$$\begin{aligned} E_{estat} &= \tfrac{1}{2}\sum_{\substack{a\in A\\ b\in B}}\sum_{tu} Q_t^a T_{tu}^{ab} Q_u^b \\ &= \sum_{a\in A}\sum_{t} Q_t^a V_t^a \end{aligned} \quad . \tag{67}$$

where V_t^a is the electrostatic field at site *a* of a molecule in the central unit cell. However, dmarel was not designed to calculate the total field at each site in this way as part of the

electrostatic energy summation. Separating the required data from the electrostatic energy, so as to resolve the fields at sites due to all interactions, was a significant undertaking. The fields are calculated at sites *a* due to all interacting multipoles, and are stored in the array `gfield as` 10, 11s and 11c components. The product of a field and a polarizability tensor is an induced multipole, and the subsequent product of the induced multipoles with the electrostatic field gives the induction energy as described in section 2.4. Future implementation of rank 2 polarizabilities would additionally require the l=2 components of the field at each site to be calculated and stored.

The equation for the induction energy (section 2.4.2) also uses the electrostatic fields:

$$E_{ind} = \sum_{a \in A} \sum_{t} \Delta Q_t^a V_t^a \,. \tag{53}$$

By amending the routines that calculate the multipole interaction energies, the fields are tabulated for each polarizable site during the normal lattice energy calculation.

The electrostatic terms are separated within the code, into charge-charge, charge-dipole, dipole-dipole, and higher multipole interactions, which are summed up to the direct-space cut-off. The terms up to dipole-dipole are also summed over an infinite lattice by using an Ewald summation, except for the charge-quadrupole interaction which is always summed in direct space instead of using the Ewald code in DMACRYS. As described in section 2.6.2.1, this summation in reciprocal space is for long-range interactions. The individual contributions to the field due to reciprocal space are difficult to factor out from the Ewald sum so that they may be damped appropriately, and it would be preferable to be able to neglect them.

Table 11: First order induced moments for α-oxalyl dihydrazide (see Figure 12) resulting from the non-damped electrostatic field due to 15 Å direct space cut-off; and the same system using Ewald summation.

Atom	Sum	Q10/ e a_0	Q11s/ e a_0	Q11c/ e a_0
C1	Direct	0.011733	0.027841	-0.064738
	Ewald	*0.011770*	*0.027534*	*-0.064475*
C2	Direct	-0.014003	-0.030117	0.065996
	Ewald	*-0.014140*	*-0.030442*	*0.066650*
O1	Direct	0.066538	0.105565	-0.086672
	Ewald	0.066629	0.104710	-0.086053
O2	Direct	-0.062365	-0.106257	0.086981
	Ewald	*-0.062615*	*-0.105935*	*0.087123*
N1	Direct	0.110803	0.066820	-0.154249
	Ewald	*0.110960*	*0.062823*	*-0.157660*
N2	Direct	0.017572	-0.055687	-0.138902
	Ewald	*0.017930*	*-0.057649*	*-0.138539*
N3	Direct	-0.020377	0.055590	0.132344
	Ewald	*-0.020762*	*0.057164*	*0.132390*
N4	Direct	-0.109162	-0.079153	0.156327
	Ewald	*-0.109361*	*-0.075847*	*0.160176*
H1	Direct	0.000954	0.012444	-0.056538
	Ewald	*0.001106*	*0.012425*	*-0.056680*
H2	Direct	0.051501	-0.037121	-0.028993
	Ewald	*0.051946*	*-0.037542*	*-0.029196*
H3	Direct	-0.033412	0.016085	-0.040690
	Ewald	*-0.033327*	*0.015819*	*-0.040629*
H4	Direct	0.033850	-0.016220	0.049477
	Ewald	*0.033781*	*-0.016109*	*0.049504*
H5	Direct	-0.055111	0.042221	0.034164
	Ewald	*-0.055567*	*0.042598*	*0.034572*
H6	Direct	0.001429	-0.010666	0.055076
	Ewald	*0.001284*	*-0.010707*	*0.055373*

The induced moments for α-oxalyl dihydrazide, calculated using the field from direct space only and also including the reciprocal space contribution, and shown in Table 11. For the tabulated induced moments, the difference in induction energy of the lattice is 0.2 kJ mol^{-1}, and the converged induction energies using the two different summations come to -56.64

and -57.12 kJ mol^{-1}, respectively. In this strongly polarized system the difference is less than 1% of the induction energy, and for less strongly interacting molecules the difference will be correspondingly less. Therefore the electrostatic fields used to calculate the induced moments and induction energy are only summed in direct space, as Table 11 shows that the reciprocal space contribution to the induction energy is negligible.

5.3.1. Damping

Some form of damping is required to account for the effect of orbital overlap on the multipole expansion, and thereby prevent the model potential producing catastrophically too short intermolecular contacts. The Tang-Toennies damping function (54) takes three parameters: the sum of the angular momentum of the interacting multipoles (n), the distance between the sites (R), and a damping factor (β). The expansion allows for higher-order multipoles to be damped by using the previous evaluation as a starting point. First and second derivatives are easily evaluated at the same time, so this function can be used for calculating numerical derivatives of the dispersion energy as part of the normal minimization with analytical derivatives in the future.

$$
\begin{aligned}
f_n(\beta R) &= 1 - \left(\sum_{k=0}^{n} \frac{(\beta R)^k}{k!} \right) e^{-\beta R} \\
f_n'(\beta R) &= -\left(\sum_{k=0}^{n} \frac{k \beta^k R^{(k-1)}}{k!} \right) e^{-\beta R} + \left(\sum_{k=0}^{n} \frac{(\beta R)^k}{k!} \right) \beta e^{-\beta R} \\
f_n''(\beta R) &= -\left(\sum_{k=0}^{n} \frac{k(k-1) \beta^k R^{(k-2)}}{k!} \right) e^{-\beta R} + 2\left(\sum_{k=0}^{n} \frac{k \beta^k R^{(k-1)}}{k!} \right) \beta e^{-\beta R} \\
&\quad + \left(\sum_{k=0}^{n} \frac{(\beta R)^k}{k!} \right) \beta^2 e^{-\beta R}
\end{aligned}
\quad , \tag{68}
$$

The function is implemented in `damping.f90`; the functional part of the code is listed in Figure 21

Figure 21: Tang-Toennies function in damping.f90, calculating damping function and its derivatives efficiently, as an example of coding. The 'case' control structure means it is straightforward to add further damping functions as required.

```
scal=1
g=1.d0

select case(typ)
case(0)
  g=1.d0       ! Damping
  g1=0.d0      ! First derivative
  g2=0.d0      ! Second derivative

case(1)
!  Tang-Toennies
    if (new) then
      !  Starting from scratch
      br=scal*r
      ebr=exp(-br)
      s=(1.d0+br)*ebr
      z=ebr*br
      fn=1.d0
      do k=2,n
        fn=fn+1.d0
        z=z*br/fn
        s=s+z
      end do
    else
      !  continuing from previous calculation
      do k=n0+1,n
        fn=fn+1.d0
        z=z*br/fn
        s=s+z
      end do
    endif
    !  z is now exp(-bR)*(bR)^n/n!
    n0=n
    g=(1.d0-s)
    g1=scal*z
    g2=g1*scal*(fn/br-1d0)
end select
```

Once implemented, damping is easily applied to the fields since the distance, R, and ranks *t,u* are readily available for each interaction, and the sum becomes

$$\mathbf{V}_t^a = \sum_{a \in A} \sum_{B \neq A} \sum_{b \in B} \sum_{tu} f_n(\beta R_{ab}) T_{tu}^{ab} Q_u^b \,, \tag{69}$$

where $\mathbf{V}_t^a$ is an array of field components *t* for which there is a polarizability tensor of that order, for each atom *a*. The normal summation of the energy already includes $T_{tu}^{ab} Q_u^b$ as a term, with the orientation parts of the T_{tu}^{ab} tensor stored in a lookup table. Where the field is multiplied by a charge to calculate the electrostatic interaction energy, it is also scaled by the damping function and summed into the total field at site *a*. The entire procedure is described in the following section.

At present, molecules are given a single damping parameter, which is used for all types of interacting atoms. It is likely that certain atoms or functional groups will need to be damped differently, and this is discussed further in chapter 6. If a damping parameter can be determined for individual atoms then this could be incorporated into the modified atom-atom based code.

5.4. The Iterative Procedure to Achieve Induced Moments

The induced moments and induction energy are calculated for a fixed cell. For each set of lattice parameters, the new `iterate` subroutine, which controls use of the induced moments, calls the function that calculates the direct-space electrostatic sum. This allows the electrostatic field to be captured as described above, and summed into a `field` array described by equation (69). Once the sum over interacting sites is complete, the `iterate` routine proceeds to calculate the induced moments using equation (52) and the induction energy by equation (53).

The iteration cycle is described more fully by the following series of steps:

1. **Sum the electrostatic field at each polarizable site, and store.**

The new `iterate` subroutine calls `frcnst`. This is the main subroutine that controls the procedure of cycling over each atom-atom pair and calculating the interaction energy of the multipole moments, and the relevant repulsion-dispersion potential. The electrostatic field at each site in the asymmetric unit cell is calculated for each interaction with multipole moments on those sites. While the energy is summed, the field is stored in the `gfield` array in the global coordinate system. If requested by the DAMP keyword, then the `damping` function is called with the atom-atom distance and the rank of interaction, and returns a value by which the field is scaled before it is summed into the total.

2. **Rotate the field from the lab frame to the local molecular frame.**

Once this sum over interaction sites is complete, the stored fields are rotated to the local molecular axis frame. This is a trivial step since the code has calculated the relevant transformations which are used to rotate the multipoles for the electrostatic energy, and are already available in the program. The electrostatic field is retained separately in `sfieldloc` for later use, as it is multiplied with the induced moments to calculate the energy.

3. **Take the product of the field at each site with the corresponding polarizability**

The electric field components are multiplied with their corresponding polarizabilities so that induced multipole moments are produced, following the equation $\Delta Q_t^a = \alpha_t^a A_t^a$. These induced moments replace any previous induced moments, and are multiplied with the original electrostatic fields stored in `sfieldloc` to give the induction energy using equation (53).

4. **Monitor the change in the induction energy**

If the calculated induction energy has changed by greater than a preset tolerance, then the new induced moments are summed to the original static moments. The tolerance by which

the induction energy is tested is 10^{-10} kJ mol^{-1}. This is higher than the accuracy of the polarizability model, however it is necessary to have this level of consistency in order to reduce the noise of the potential energy surface for the numerical gradients calculation (see section 7.2). If the energy change is less than the tolerance, then the procedure goes to step 6.

5. Recalculate the field with induced moments

The newly calculated induced moments are summed to the original static moments, $\Delta Q_t^a + Q_t^a$, which are then used to recalculate the electric fields at each site. The procedure then returns to step 2.

6. Return the converged induction energy to the program

When the induction energy has converged within the tolerance level, the static multipole moments are restored, and the value of the induction energy is returned to the calling function.

The convergence of the induction energy is demonstrated in Table 12, which shows the energy at each iteration for a series of closely related structures that are generated when the code makes small, systematic changes to the lattice parameters to build the second derivative matrix **W** (actually $\mathbf{W}^{-1}$, see equation 55 and surrounding text). Further detail about the numerical derivatives and the problems of minimizing the lattice are described in section 7.2. Each column terminates when the energy change is less than the tolerance setting. As expected, iteration of the energy is vital. The second iteration at search step 1 recovers 15 kJ mol^{-1} more than the first step; around one third more energy. For the structure at step 1 the energy has converged to within 0.05 kJ mol^{-1} by the 7th iteration, and has converged to 10^{-10} kJ mol^{-1} after 17 iterations.

Where the starting induced moments are zero, it may sometimes require more than 200 iterations to converge the energy to this level of consistency. By using the induced moments of a closely related structure as a starting point, subsequent calculations are able to converge with fewer iterations. The effect of using the previous moments is seen by looking along the top row of Table 12. The induction energies for steps 2-8 start at -56.63 kJ mol^{-1} as a good approximation of the induction energy, and converge within a few iterations. An example of where many more iterations are needed is shown by Table 13. The ε-oxalyl dihydrazide structure is a challenging one because of the short intermolecular contacts that come about in the optimized lattice cell (section 4.3.3), and requires 110 iterations to converge the induction energy at the starting point. By using the converged induced moments as a guess, the number of iterations needed is more than halved at each subsequent step.

Table 12: Induction energy (kJ mol^{-1}) for α-ODH, as an illustration of the convergence behaviour for 8 closely related structures that are automatically tried by DMACRYS when calculating the numerical second derivatives.

	Structure							
Iteration	1	2	3	4	5	6	7	8
1	-39.585215	-56.631929	-56.632239	-56.631577	-56.631379	-56.632032	-56.631924	-56.631739
2	-51.106391	-56.631983	-56.632307	-56.631375	-56.631429	-56.632230	-56.631807	-56.631739
3	-54.852538	-56.632014	-56.632338	-56.631285	-56.631451	-56.632298	-56.631762	
4	-56.025312	-56.632014	-56.632343	-56.631240	-56.631456	-56.632334	-56.631739	
5	-56.428732		-56.632347	-56.631235	-56.631460	-56.632343	-56.631739	
6	-56.559939		-56.632343	-56.631226	-56.631465	-56.632347		
7	-56.607075		-56.632347	-56.631226	-56.631465	-56.632347		
8	-56.622780		-56.632356					
9	-56.628606		-56.632343					
10	-56.630573		-56.632347					
11	-56.631330		-56.632343					
12	-56.631577		-56.632352					
13	-56.631685		-56.632343					
14	-56.631712		-56.632352					
15	-56.631721		-56.632347					
16	-56.631730		-56.632352					
17	-56.631730		-56.632352					

Table 13: Induction energy for ε-ODH, showing the converged energy and number of iterations to achieve that

Step	1	2	3	4	5	6	7	8
# Iterations	110	35	25	46	31	46	40	4
Energy (kJ mol^{-1})	-30.662799	-30.662673	-30.662609	-30.663232	-30.662986	-30.662609	-30.662799	-30.662799

5.5. Validating Induced Moments and Energies

The implementation of the induction method may be validated by comparing the results with those for alternative methods used in chapter 4. The ORIENT cluster model calculates induced moments in response to the electrostatic field, which may be compared at each iteration to the response calculated within DMACRYS. The induced moments can also be used to calculate induction energies. There are limitations on the agreement between the DMACRYS model and the previous cluster models, since there is a difference between the infinite lattice system and the clusters comprised of unit cells. The clusters that I use for comparison are generated using the SHELX file to determine the rotation angle, axis and the translation of the centre of mass of the molecule, to reproduce their positions in unit cells. As well as this source of numerical error in the atomic positions, there may be molecules that lie on the boundary of the cut-off sphere that are included in one system and not in the other, which account for the small differences between the methods. However they are closely enough related for the comparison to be made.

Table 14 contains the induced moments as the first order response to the electrostatic field of α-oxalyl dihydrazide, for which the atom numbering system is shown in . Most of the variation is in the 4th decimal place, which is approaching the level of numerical accuracy.

Figure 22: Atom labelling system of oxalyl dihydrazide.

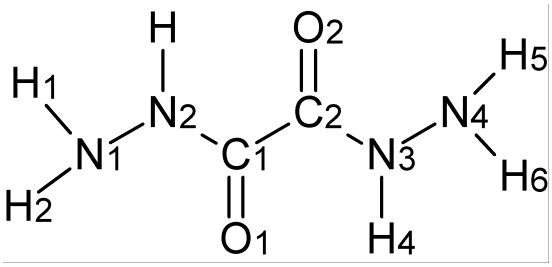


Table 14: Induced dipoles (e a_0) of α-ODH after 1 iteration within the lattice (DMACRYS) and in a cluster (ORIENT), without damping.

	DMACRYS			ORIENT		
	Q_{10}/ e a_0	Q_{11s}/ e a_0	Q_{11c}/ e a_0	Q_{10}/ e a_0	Q_{11s}/ e a_0	Q_{11c}/ e a_0
C1	0.011733	0.027841	-0.064738	0.011742	0.02842	-0.065728
C2	-0.014003	-0.030117	0.065996	-0.014138	-0.029496	0.066829
O1	0.066538	0.105565	-0.086672	0.061098	0.101591	-0.081531
O2	-0.062365	-0.106257	0.086981	-0.057212	-0.101727	0.081499
N1	0.110803	0.066820	-0.154249	0.099745	0.054187	-0.150971
N2	0.017572	-0.055687	-0.138902	0.017378	-0.056653	-0.137295
N3	-0.020377	0.055590	0.132344	-0.020471	0.057560	0.131122
N4	-0.109162	-0.079153	0.156327	-0.097899	-0.064956	0.152779
H1	0.000954	0.012444	-0.056538	0.001445	0.010895	-0.052042
H2	0.051501	-0.037121	-0.028993	0.046172	-0.031898	-0.031174
H3	-0.033412	0.016085	-0.040690	-0.024848	0.005421	-0.030516
H4	0.033850	-0.016220	0.049477	0.024728	-0.005281	0.037208
H5	-0.055111	0.042221	0.034164	-0.050429	0.037843	0.035788
H6	0.001429	-0.010666	0.055076	0.000807	-0.009004	0.050602

The RMSD of these two sets of moments is 0.0062266 e a_0, i.e. the value of the multipoles are very similar. The RMSD of the electrostatic potential on the 1.8x van der Waals surface is only 1.14902 kJ mol^{-1} over 6542 points probed with a unit point charge. In terms of the induction energy of the crystal, using either the induced dipoles from that ORIENT cluster, or from DMACRYS are, for the first-order induced moments: -46.738 and -46.524 kJ/mol, respectively.

This level of consistency is reassuring, but it is a requirement that it extends to the converged case. Iterated moments are presented in Table 15, for which the lattice induction energies are -68.572 and -68.951 kJ/mol. The RMSD for this set is 0.002078 e a_0 which remains very small after the two models have been through 18 iterations. By including the same damping function with the ORIENT and DMACRYS models, the converged moments (Table 16) have an RMSD of 0.001182 e a_0.

Table 15: Induced dipoles (e a_0) of α-ODH, converged within the lattice (DMACRYS) and within the cluster (ORIENT), without damping.

	DMACRYS			ORIENT		
	Q_{10}/ e a_0	Q_{11s}/ e a_0	Q_{11c}/ e a_0	Q_{10}/ e a_0	Q_{11s}/ e a_0	Q_{11c}/ e a_0
C1	0.018362	0.064135	-0.090914	0.018548	0.067232	-0.094800
C2	-0.023106	-0.070033	0.097363	-0.023010	-0.068310	0.095857
O1	0.103124	0.187303	-0.111544	0.103864	0.191988	-0.117050
O2	-0.095326	-0.190748	0.114843	-0.095380	-0.191680	0.115779
N1	0.152781	0.112745	-0.218841	0.154319	0.119672	-0.217440
N2	0.022721	-0.063155	-0.199107	0.022053	-0.060430	-0.201880
N3	-0.026518	0.061698	0.192909	-0.026340	0.062289	0.193440
N4	-0.154151	-0.136553	0.228251	-0.154700	-0.138400	0.224000
H1	0.001260	0.021134	-0.082394	0.001510	0.021708	-0.082600
H2	0.069163	-0.04537	-0.041353	0.068983	-0.044800	-0.042190
H3	-0.047535	0.024607	-0.057489	-0.048340	0.025361	-0.058400
H4	0.049461	-0.026494	0.071455	0.049731	-0.025800	0.071723
H5	-0.074888	0.052702	0.048528	-0.075550	0.053434	0.048593
H6	0.002169	-0.019630	0.081400	0.001977	-0.019510	0.080254

Table 16: Induced dipoles (e a_0) of α-ODH, converged within the lattice (DMACRYS) and within the cluster (ORIENT), with damping.

	DMACRYS			ORIENT		
	Q_{10}/ e a_0	Q_{11s}/ e a_0	Q_{11c}/ e a_0	Q_{10}/ e a_0	Q_{11s}/ e a_0	Q_{11c}/ e a_0
C1	0.017530	0.061483	-0.091009	0.017607	0.062563	-0.093027
C2	-0.021599	-0.063835	0.092360	-0.021879	-0.063419	0.093928
O1	0.091855	0.174015	-0.105187	0.090396	0.174913	-0.106627
O2	-0.084091	-0.174770	0.104736	-0.082753	-0.174987	0.105613
N1	0.135881	0.093419	-0.206008	0.133155	0.091212	-0.206403
N2	0.022221	-0.062658	-0.192525	0.021986	-0.063602	-0.194171
N3	-0.026030	0.062960	0.183746	-0.026283	0.065672	0.186120
N4	-0.133749	-0.110020	0.210508	-0.132072	-0.107315	0.211238
H1	0.001606	0.018114	-0.073594	0.001974	0.018479	-0.073315
H2	0.059657	-0.037586	-0.041705	0.059813	-0.037666	-0.042235
H3	-0.035037	0.011577	-0.043176	-0.035068	0.010906	-0.043310
H4	0.035539	-0.012000	0.053124	0.035499	-0.010982	0.053326
H5	-0.065670	0.045553	0.047248	-0.066112	0.046162	0.047738
H6	0.001628	-0.016149	0.071496	0.001216	-0.016307	0.071175

Table 17: The damped first order induced moments for (E)-4-(Trifluoromethyl)benzaldehyde oxime. The left hand side of the table are for the molecule, and the right hand side for the inverted molecule.

	Original molecule			Inverted molecule		
	Q_{10}/ e a_0	Q_{11s}/ e a_0	Q_{11c}/ e a_0	Q_{10}/ e a_0	Q_{11s}/ e a_0	Q_{11c}/ e a_0
C2	0.001900	0.010471	0.004350	-0.001900	0.010471	0.004350
C3	0.004145	0.037103	0.000295	-0.004145	0.037103	0.000295
C4	0.004613	0.028232	0.025668	-0.004610	0.028230	0.025668
C5	-0.005102	0.001543	0.020470	0.005130	0.001564	0.020482
C6	-0.005617	-0.043144	0.050001	0.005615	-0.043160	0.050004
C7	-0.007941	-0.052190	-0.028664	0.007940	-0.052188	-0.028663
C8	-0.001536	0.014712	-0.015638	0.001536	0.014712	-0.015638
C9	-0.001244	-0.032281	0.040536	0.001243	-0.032288	0.040537
F1	-0.001912	0.005281	-0.001297	0.001912	0.005281	-0.001297
F2	-0.000283	0.008764	0.001113	0.000283	0.008764	0.001113
F3	0.000450	0.022173	0.007918	-0.000450	0.022173	0.007918
O1	0.008507	0.028989	-0.036682	-0.008497	0.028930	-0.036587
H1	0.006383	0.001986	-0.021279	-0.006383	0.001986	-0.021279
H2	0.004007	0.000322	-0.001893	-0.004007	0.000322	-0.001893
H3	-0.003508	0.002817	0.010534	0.003508	0.002817	0.010534
H4	-0.008039	-0.014863	0.012644	0.008039	-0.014863	0.012644
H5	-0.004391	-0.006741	-0.011053	0.004391	-0.006741	-0.011053
H6	0.000304	0.002133	0.017328	-0.000304	0.002133	0.017328
N1	-0.010543	0.055466	0.055495	0.010606	0.055194	0.055628

As a further test, I present data of (E)-4-(Trifluoromethyl)benzaldehyde oxime (Figure 3) which was introduced in 2.7. The implementation must be able to correctly handle crystal space groups that have an inversion centre, and this is proven by Table 17. The induced moments for the enantiomers are calculated independently, so the tight agreement of the moments and change of sign in the Q_{10} moment confirms that my implementation is correct in this respect. The most strongly polarized atom is the nitrogen, in the Q_{11s} and Q_{11c} directions, i.e. in the plane, because of its involvement in hydrogen bonding. DMACRYS estimates the damped, first order induction energy to be −4.15 kJ mol^{-1} while an ORIENT cluster calculation estimates −4.62 kJ mol^{-1}. The converged induction energies are estimated to be −4.86 kJ mol^{-1} and −5.02 kJ mol^{-1}, respectively.

5.6. Conclusions

Following the success of the WSM polarizability model that was tested in chapter 4, I have implemented the induced dipole model into DMACRYS using an iterative method. With this complete, it is now possible to routinely calculate the induction energy of a crystal structure from the WSM polarizabilities available. The induction energy is converged to a high level of consistency so that it can be used to calculate numerical derivatives of the lattice energy with respect to the lattice parameters, which in principle will allow structures to be relaxed with induction. For some technical reasons this is not yet feasible, and this is discussed further in chapter 7.

I have compared the induced moments and induction energies from ORIENT clusters with those calculated using DMACRYS, and shown that they are consistent within small error. DMACRYS internally generates symmetry-related atomic positions in a periodic system. In order to accurately predict the crystal lattice energy, adding the induction energy as I have

described is not sufficient. Intermolecular interaction potentials are needed that do not include any of the induction energy, otherwise an unknown amount of double-counting will skew the results. Such a potential was created[3] for the 2007 blind test, and in chapter 6 I present my part in testing and using that potential and a subsequent evaluation of the induction energy of some crystal structures.

Reference List

1. Tang, K. T.; Toennies, J. P. An Improved Simple-Model for the Vanderwaals Potential Based on Universal Damping Functions for the Dispersion Coefficients. *J. Chem. Phys.* **1984,** *80* (8), 3726-3741.

2. *CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form,* University of Cambridge (2007), http://www-stone.ch.cam.ac.uk/programs.html#CamCASP: 2007

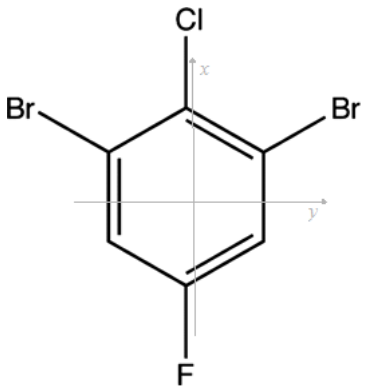
3. Misquitta, A. J.; Welch, G. W. A.; Stone, A. J.; Price, S. L. A first principles solution of the crystal structure of C6Br2ClFH2. *Chem. Phys. Lett.* **2008,** *456* (1-3), 105-109.

Chapter 6. Crystal Structure Prediction with *ab initio* Potential

6.1. Introduction

The Cambridge Crystallographic Data Centre (CCDC) has held blind tests[1-4] in the prediction of organic crystal structures to allow the community to assess their ability to predict the observed (but unknown to the participants) structure starting with only the molecular diagram. Each time the target molecules are presented in categories, such as small and rigid or flexible, and are chosen to offer a selection of computational challenges. In the most recent blind test (2007)[4] the target molecules included a poly-halogenated benzene structure, Figure 23, identified as XIII.

Figure 23: Blind test molecule XIII, $C_6Br_2ClFH_2$



This system is challenging for several reasons. Firstly it is a small, planar, rigid molecule, and such systems are often able to pack and stack in a large number of similar ways, and can be disordered.[5] Such molecules often produce large numbers of related crystal structures close to the global minimum[6,7] and so it can be very challenging to predict the most stable structure. Secondly, it is well known that halogen atoms are anisotropic in their

electron distribution[8,9], and so the model intermolecular potential should model the anisotropy in the repulsive part of the potential. There is a lack of well tested anisotropic atom-atom potentials, although potential parameters have been derived for the chlorobenzenes.[10]

The recent development of CamCASP[11] allowed Dr Alston Misquitta to generate a custom, anisotropic atom-atom intermolecular potential for XIII from the molecular charge density.[12] My contribution to the crystal structure prediction was to prepare a suitable set of crystal structures with an approximate potential, and then refine their relative lattice energies using the custom anisotropic atom-atom potential. The potential was developed with CamCASP for use with ORIENT, and I ensured it was correctly translated for use with DMACRYS. At the time, we were unable to model the induction energy in DMACRYS, and so the induction energy had to be omitted the model potential. Following the programming of the induction into DMACRYS (Chapter 5), it became possible to use this system to test whether the addition of the induction contribution would change the relative lattice energies of some of the low energy structures produced in the search.[12] Since the repulsion-dispersion potential was generated from *ab initio* data, there would be no double-counting of the induction energy in this case.

6.2. Conducting the search

As described in the related Letter[12] the search strategy was to generate a large number of approximate structures, and to reduce the number being considered by reminimizing them with increasingly better model potentials. Two initial searches were conducted to give an overview of the typical packing behaviour. One used MOLPAK,[13] with DMA electrostatic model from an MP2/6-31G(d,p) wavefunction computation and the FIT potential, and

another using CrystalPredictor[14] with a point-charge model and the Williams01 potential. CrystalPredictor uses a number generation algorithm that should sample the packing space more completely than MOLPAK, which generates initial structures by building a co-ordination sphere that is dependent upon the crystalline space group being generated. CrystalPredictor had not been used by the group for searches before and so it was important to verify that it found all of the minima that MOLPAK does. The CrystalPredictor search produced some 20,000 crystal structures with one molecule in the asymmetric unit cell, before clustering of duplicated structures. Later analysis showed that all of the circa 3000 structures produced by MOLPAK were reproduced by this method.

When the CrystalPredictor data set was reminimized using the same multipole moments and potential used for the MOLPAK search, and duplicates and transition state structures removed, there were approximately 4500 stable minima remaining. The 266 structures within 5 kJ mol^{-1} of the global minimum were used to quickly assess the trial *ab initio* potentials as they were developed and refined. The final potential was used to reminimize the lowest 1200 unique minima to find the global minimum.

6.3. Overview of the ab initio Potential

The custom potential[12] was created using CamCASP[11] using the approach of deriving an *ab initio* atom-atom potential described in a recent review.[15] The molecular geometry was found by an *in vacuo* optimization of the MP2/6-31G(d,p) wavefunction using GAUSSIAN[16]. This geometry was held rigid throughout the investigation, and the planar C_{2v} symmetry was exploited in deriving the repulsion-dispersion potential. Molecular properties were obtained using the asymptotically corrected[17] PBE0[18] exchange-correlation function with the Sadlej pVTZ[19,20] basis set (as used for all the polarizability calculations in this thesis).

The first ionisation energy, required by the asymptotic correction, was calculated by a delta-DFT calculation using the PBE0 functional in GAUSSIAN. This also provides the estimated damping parameter β=1.635, using the relation in section 2.4.3. The DMA is used for the electrostatic energy, and polarizabilities at imaginary frequencies, $\alpha(i\omega)$, are used for the C_6, C_7, C_8 dispersion coefficients[21].

In order to fit the short range potential terms, a set of 1400 dimer geometries were generated. At each geometry, SAPT(DFT) was used to calculate the total intermolecular potential, as defined by[22-25]

$$U \approx E_{elst}^{(1)}(KS) + E_{exch}^{(1)}(KS) + E_{disp,d}^{(2)} \tag{70}$$

These energies were used to parameterize an anisotropic atom-atom potential of the form

$$U = G\sum_{a\in A}\sum_{b\in B}\exp(-\alpha_{ab}(R_{ab} - \rho_{ab}(\Omega_{ab}))) + \mathrm{DMA} - \sum_{ab} C_6^{ab}/R^6 \,. \tag{71}$$

The anisotropy in the atom-atom repulsion, $\rho_{ab}(\Omega_{ab})$, is a two term Legendre expansion in the angles between covalent bond vectors and the atom-atom vectors. The form of this expansion was determined by fitting the atom-atom overlaps between the molecules in the set of dimer geometries. Then, using the assumption that the short range repulsion is proportional to the total overlap between the molecules, the proportionality constants were determined by fitting to the SAPT(DFT) interaction energies. In principle, the repulsive term is modelling the exchange-repulsion and penetration energies. However, since the dispersion term is an undamped C_6 approximating the damped C_6, C_7, C_8 dispersion, some errors in this approximation are being absorbed into the repulsion term. Since this was the first test of this approach to a completely non-empirical intermolecular potential, various trial model potentials were generated with different methods of fitting the model potential to the SAPT(DFT) energies during the course of the blind test. The complete method has been published[12].

6.4. The Crystal Energy Landscape

As the errors in the potential were reduced by fine-tuning the fitting procedure, the 266-molecule test set was repeatedly reminimized until it was clear that further potential refinements were unlikely to change the relative energies of the most structures. We had intended to reminimize the full set of approximately 4500 structures with the custom potential in case there were any large shifts in energy. Due to time pressure I took the strategy of reminimizing structures in order of increasing energy from the global minimum. After 1200 minimizations the energy gap between new structures and the global minimum had reached 12 kJ mol^{-1} and no lower energy putative structure were being sufficiently stabilised to compete for submission as the experimental structure. The 10 lowest-energy stable structures are listed in Table 18.

Table 18: Lattice details for the lowest 10 unique structures from the XIII search. The three submitted structures are highlighted in bold. The lowest in energy, af395, corresponds to the experimental structure.

Label	Space group	a (Å)	b (Å)	c (Å)	a (°)	b (°)	g (°)	Cell density (g cm^{-3})	Lattice energy (kJ mol^{-1})
af395	**$P2_1/c$**	**3.8052**	**13.7907**	**20.339**	**90**	**134.528**	**90**	**2.5171**	**-91.41**
ab43	**P-1**	**7.9509**	**11.4832**	**9.1399**	**80.861**	**101.55**	**35.362**	**2.4592**	**-90.71**
af62	$P2_1/c$	8.8629	6.7961	15.7847	90	54.773	90	2.4659	-90.2
af50	$P2_1/c$	8.0542	6.7534	15.7693	90	113.105	90	2.4276	-89.69
db456	C2/c	13.2885	8.9616	13.6942	90	72.475	90	2.4631	-89.39
xx646	**I-4**	**20.2489**	**20.2489**	**3.7629**	**90**	**90**	**90**	**2.4826**	**-89.26**
db59	C2/c	15.9934	9.1018	13.0163	90	124.606	90	2.4561	-89.25
db155	C2/c	16.1883	6.7752	15.6994	90	68.004	90	2.3992	-88.65
af8	$P2_1/c$	3.7882	13.5148	16.3292	90	113.459	90	2.4973	-88.64
db26	C2/c	17.8463	6.7696	15.9088	90	124.452	90	2.4169	-88.63
aq797	$P2_12_12_1$	3.8222	13.5104	15.0057	90	90	90	2.4716	-88.34

Figure 24: Two views of the overlay of the experimental (black) and predicted af395 (grey) crystal structures of $C_6Br_2ClFH_2$, taken from[12]. The two structures are hard to separate as the overlay is nearly perfect.

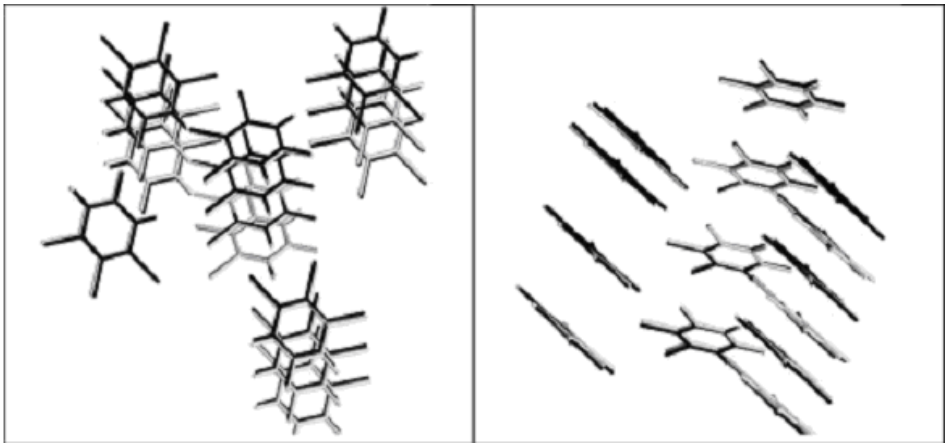
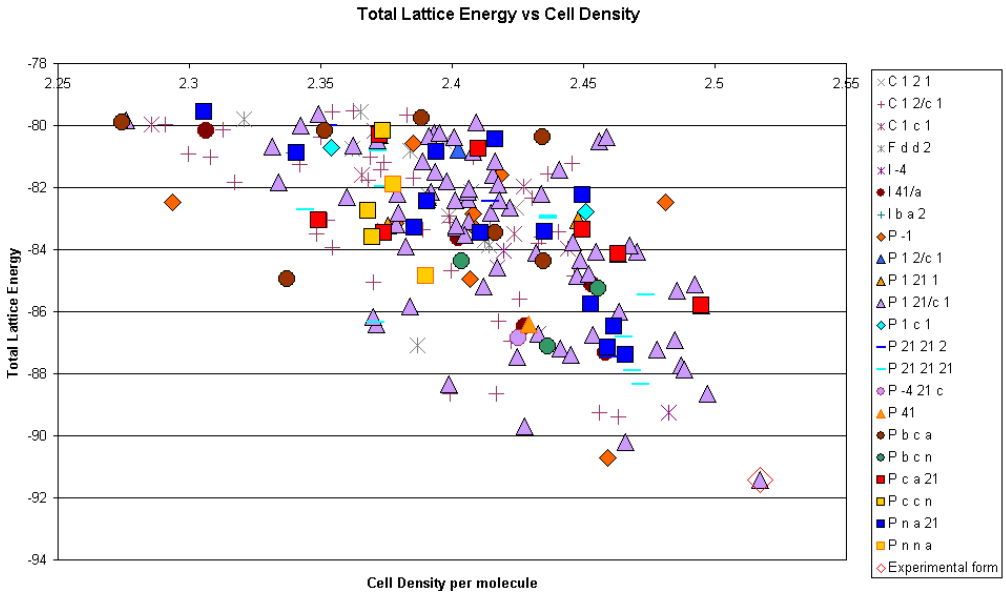


Figure 25: Crystal energy landscape of XIII after minimization using the custom potential (excluding the induction). The experimental form corresponds to the lowest energy predicted structure.



Although the crystal structure prediction methodology at the time did not include any induction contribution, the experimental structures were successfully predicted with an overlay of the 15 molecules shown in Figure 24 of 0.15 Å. There were 35 structures within 5 kJ mol^{-1} of the global minimum and 154 structures within 10 kJ mol^{-1} (see Figure 25). The experimental structure is clearly predicted to be the most stable in lattice energy, although the next minimum is only 0.6 kJ mol^{-1} less stable than it.

This success in predicting the known structure as the global minimum in the lattice energy implies that the crystallization was under thermodynamic control, and that the order of the free energies at the crystallization temperature is the same as at 0 K.

6.4.1. The Polarizability and Induction Energies of XIII

The distributed polarizabilities for XIII had been calculated as part of the derivation of the dispersion coefficients. A comparison of the isotropic atomic polarizabilities of oxalyl dihydrazide and XIII is given in Table 19, along with the total isotropic molecular polarizabilities. The C-H carbon atoms in XIII are comparable in polarizability to the CHNO atoms in α-oxalyl dihydrazide, but the other carbon atoms are significantly more polarizable and overall XIII is twice as polarizable.

The damped induction energy when a unit point charge is positioned on the 1.8 times van der Waals surface of XIII is shown in Figure 26. The induction energy is lowest near the fluorine, -12.6 kJ mol^{-1}, and particularly large above the plane of the ring, -70.0 kJ mol^{-1}, near the carbons bonded to chlorine and fluorine. By comparison, the induction energy of α-oxalyl dihydrazide when probed over the same surface ranges from -25.1 to -53.9 kJ mol^{-1}, i.e. it is not as polarizable as XIII. A casual inspection of Figure 26 may seem to imply

that the bromine sites are not particularly polarizable, but bromine has a significantly larger van der Waals radius other than the atoms in the molecule. Hence, other atoms in van der Waals contact with bromine will have their nucleus within the surface plotted in Figure 26, whereas other van der Waals contacts may be on or outside the surface.

The electrostatic contribution to the global minimum lattice energy is only 12% for the experimental crystal structure of XIII. In addition, there are no strongly directional electrostatic fields produced around XIII, as there would be around hydrogen-bonding groups. Smaller electrostatic fields will limit the magnitude of induction in the crystal.

Table 19: Isotropic atomic and total polarizabilities of α-ODH, and XIII.

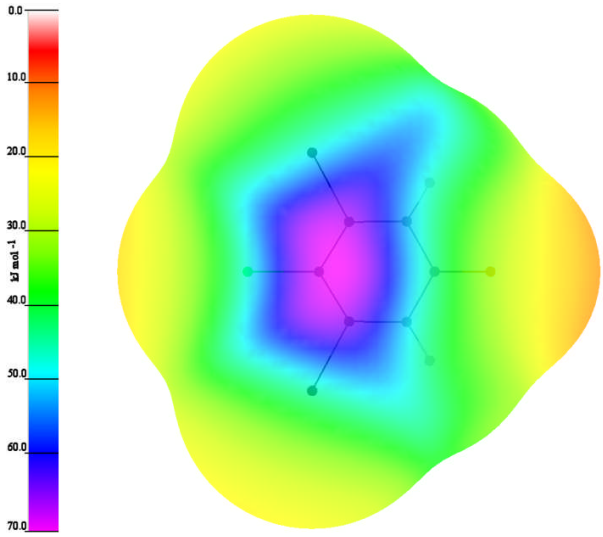
α-oxalyl dihydrazide (Chapter 4) (units $4\pi\varepsilon_0 a_0^3$)														
C1	C2	O1	O2	N1	N2	N3	N4	H1	H2	H3	H4	H5	H6	Total
8.58	8.47	7.08	7.21	7.21	6.94	6.87	7.26	1.48	2.03	1.46	1.56	2.12	1.44	69.70
XIII (units $4\pi\varepsilon_0 a_0^3$)														
Br	Cl	F	C(Br)	C(Cl)	C(F)	C(H)	H							Total
20.55	13.77	2.77	12.65	14.19	10.06	7.79	2.41							127.93

The SAPT(DFT) method was used to find the induction contribution to the dimer interactions energies. For the 1400 geometries, the induction contributed typically around 3% and not more than 10% of the total interaction energy, which was negligible enough to make the approximation of ignoring it for the fitting for the blind test. However, it is substantial enough for the induction in crystals of XIII to be a good test of the later implementation of the induction in DMACRYS.

It is now possible to evaluate the induction energy within crystal structures. The different atom-atom contacts within the low energy structures could result in difference in the induction energy that would change the order of stability. The success in the blind test

would have been due to a cancellation of errors if another low energy structure was reranked as the global minimum when induction was included.

Figure 26: The induction energy probed using a damped unit point charge on the 1.8x vdW surface of XIII. The energy range is from -12.6 to -70.0 kJ mol^{-1} The molecule is orientated so that the Cl atom is closest to the scale bar.



6.4.2. The Induction Contribution to the Lattice Energy of XIII Crystal Structures

I have used DMACRYS to estimate the induction in some of the crystal structures whose lattice energies are close to the global minimum, using a rank 1 WSM polarizability model with a damping parameter of $\beta = 1.635$. The induction energy for the global minimum and four other low energy structures are given in Table 20.

Table 20: Lattice energy of the low-energy structures close to the global minimum, and their additional stabilisation by induction energy.

Search Code	Static lattice energy (kJ mol^{-1})	E_{ind} (kJ mol^{-1})	Number of DMACRYS iterations	ORIENT induction energy (kJ mol^{-1})
af395 (minimum)	-91.412	-8.10	284	-8.22
ab43	-90.707	-1.83	23	-1.89
db495	-89.385	-1.29	26	-1.35
db59	-89.247	-1.22	21	-1.29
af8	-88.643	-2.26	177	-2.39

Although the induction energy seems to further stabilise the observed crystal structure, these estimates may not be valid. The induction calculations only converged for these five of the ten lowest energy structures considered. In order to investigate the convergence problem, induction energy calculations were performed for the crystal lattice using DMACRYS, and for a molecule within a cluster of unit cells using ORIENT. An example of a successful convergence of the induction energy is given in Figure 27, which shows the induction energy of the DMACRYS and ORIENT models of ab43. After only two or three steps, it appears that the two systems are different, but this error is due to the slightly different configurations: DMACRYS summed within a cut-off in the lattice, and ORIENT using a cluster of unit cells. However, after a further few iterations the energies converge to within 0.04 kJ mol^{-1}. The first-order induced moments are also in good agreement between the two methods, as shown in Table 21.

In contrast, Figure 28 shows the divergence of the induction energy for structure af62. This structure is a valid lattice energy minimum using a good model for the intermolecular forces, and we would expect the induction energy to converge to a similar value to those shown in Table 20, not for it to approach infinity. In Figure 28 the ORIENT energies are

the interaction of the induced moments with the previous iteration's static plus induced moments, so the disagreement with the DMACRYS energies is not a cause for concern. (The additional calculations needed to calculate the induction energy in ORIENT were not worthwhile performing for a clearly diverging system). Thus, the agreement between DMACRYS and the ORIENT calculations implies that the non-convergence of the induction energy is not a problem with the implementation into DMACRYS, but is a problem with the model for the induction energy.

Figure 27: The induction energy for the XIII structure ab43. DMACRYS and ORIENT are equally able to model the induction energy, which converges rapidly in this case.

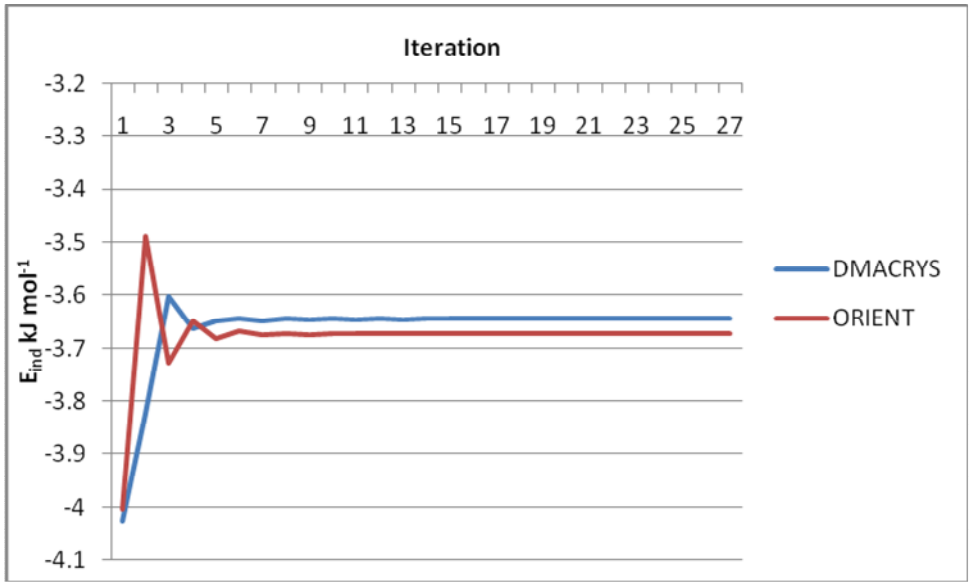
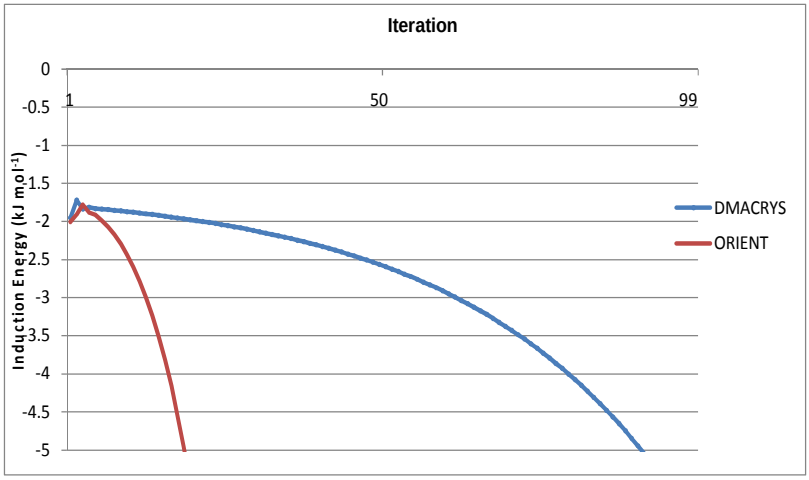


Table 21: First order induced moments of XIII system ab43, with damping, calculated with DMACRYS and ORIENT. A right-handed axis system was used, where the x axis is defined along the C(F)-C(Cl) direction and the xy-plane by this axis and C(Br).

	DMACRYS				ORIENT		
	Q10/ e a_0	Q11s/ e a_0	11c/ e a_0		Q10/ e a_0	Q11s/ e a_0	Q11c/ e a_0
Br	0.00604	0.03266	0.01141		0.00576	0.03104	0.00866
Br	0.02971	0.02837	-0.04715		0.02982	0.02855	-0.04346
Cl	0.00644	-0.00596	-0.02509		0.00595	-0.00791	-0.02694
F	-0.00203	-0.00735	0.00064		-0.00207	-0.00770	0.00059
C(Br)	-0.00079	0.00272	-0.02565		-0.00171	0.00286	-0.02668
C(Cl)	0.00204	0.00451	-0.01621		0.00185	0.00521	-0.01909
C(Br)	0.00888	0.00626	0.01730		0.00926	0.00683	0.02091
C(H)	0.00540	0.00960	0.03974		0.00531	0.00914	0.03834
C(F)	0.00084	0.00391	0.02770		0.00088	0.00252	0.02898
C(H)	-0.00275	0.00121	-0.01967		-0.00241	0.00049	-0.01524
H	0.00835	0.00999	0.00404		0.00856	0.01049	0.00378
H	-0.00893	-0.00404	-0.00284		-0.00928	-0.00407	-0.00263

Figure 28: Induction energy for XIII minima af62, which diverges. The energy of the ORIENT system is not strictly Eind, which would require an additional calculation at each step. It does exhibit the same convergence or divergence for systems as Eind calculated by DMACRYS.



The reason for the induction energy rising to infinity at a fixed geometry is because the damping of the interaction is not sufficient. The damping model was developed using small molecules containing C, H, N, O, but for XIII we are applying it to a larger system and to some very different elements. Bromine, especially, is a large atom with significant anisotropy in the shape of the charge density. As shown in Table 19 it is the most polarizable atom listed, and the molecule as a whole has a large polarizability. Examination of the induced moments in the ORIENT and DMACRYS calculations shows that it is the bromine atoms that begin to be over-polarized; the exaggerated induced moments produce stronger fields which eventually cause all sites to become over-polarized and the induction energy approaches negative infinity, as in Figure 28. The Tang-Toennies damping function has been used with a single molecule-based parameter, which was shown in Chapter 4 to be effective for those system. Unlike other molecules I have studied, XIII has atoms from four different rows of the periodic table, and it is plausible to propose that the damping should depend upon the interacting atom types.

6.5. Conclusions

The prediction of the blind crystal structure was successful, using this entirely *ab initio* derived potential. As predicted, the induction energy was not crucial to predict the correct structure, even though the molecule is quite polarizable, because of the weak electrostatic forces between molecules. Although the interaction potential modelled the global minimum as most stable by only 0.6 kJ mol^{-1}, the experimental structure, af395, also appears to probably be the global minimum when induction energy is also included. It is interesting that the correct structure has much larger induction energy than the other low-energy minima; this may be largely due to the greater density. However, it is so much

greater than any of the other structures that it seems likely to be exaggerated due to a mild form of the insufficient damping problem.

The blind test molecule $C_6Br_2ClFH_2$, contains atoms from four different rows in the periodic table. This is sufficiently different from the small systems used to develop the single-parameter damping model that the approximation of a single damping parameter for a molecule is inadequate. Half of the low energy structures tested failed to converge the induction energy, even though they were less dense than the global minimum and did not appear to have any anomalous interactions. This exposes the weakness in the damping model, however comparison of the WSM model with the SCERP in chapter 4 shows that it is a reasonable approximation for the small organic molecules for which it was derived. Hence, although this chapter confirms the validity of the programming of the induction energy, the choice of XIII revealed problems with the damping model.

Reference List

1. Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooij, W. T. M.; Price, S. L.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. A test of crystal structure prediction of small organic molecules. *Acta Crystallogr. , Sect. B* **2000,** *56*, 697-714.

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

3. Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.;

Schmidt, M. U.; Verwer, P. A third blind test of crystal structure prediction. *Acta Crystallogr. , Sect. B* **2005,** *61* (5), 511-527.

4. Day, G. M.; Cooper, T. G.; Cruz Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A.; Schweizer, B. Significant progress in predicting the crystal structures of small organic molecules - a report on the fourth blind test. *Acta Crystallogr. , Sect. B* **2009,** *65* (2), 107-125.

5. Copley, R. C. B.; Barnett, S. A.; Karamertzanis, P. G.; Harris, K. D. M.; Kariuki, B. M.; Xu, M. C.; Nickels, E. A.; Lancaster, R. W.; Price, S. L. Predictable disorder versus polymorphism in the rationalization of structural diversity: A multidisciplinary study of eniluracil. *Cryst. Growth Des.* **2008,** *8* (9), 3474-3481.

6. Barnett, S. A.; Hulme, A. T.; Issa, N.; Lewis, T. C.; Price, L. S.; Tocher, D. A.; Price, S. L. The Observed and Energetically Feasible Crystal Structures of 5-substituted uracils. *New J. Chem.* **2008,** *32* (10), 1761-1775.

7. Barnett, S. A.; Johnson, A.; Florence, A. J.; Price, S. L.; Tocher, D. A. A systematic experimental and theoretical study of the crystalline state of six chloronitrobenzenes. *Cryst. Growth Des.* **2008,** *8* (1), 24-36.

8. Desiraju, G. R.; Parthasarathy, R. The Nature of Halogen ... Halogen Interactions - Are Short Halogen Contacts Due to Specific Attractive Forces or Due to Close Packing of Nonspherical Atoms? *J. Am. Chem. Soc.* **1989,** *111* (23), 8725-8726.

9. Price, S. L.; Stone, A. J.; Lucas, J.; Rowland, R. S.; Thornley, A. E. The Nature of -Cl-Center-Dot-Center-Dot-Center-Dot-Cl- Intermolecular Interactions. *J. Am. Chem. Soc.* **1994,** *116* (11), 4910-4918.

10. Day, G. M.; Price, S. L. A nonempirical anisotropic atom-atom model potential for chlorobenzene crystals. *J. Am. Chem. Soc.* **2003,** *125* (52), 16434-16443.

11. *CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form,* University of Cambridge (2007), http://www-stone.ch.cam.ac.uk/programs.html#CamCASP: 2007

12. Misquitta, A. J.; Welch, G. W. A.; Stone, A. J.; Price, S. L. A first principles solution of the crystal structure of C6Br2ClFH2. *Chem. Phys. Lett.* **2008,** *456* (1-3), 105-109.

13. Holden, J. R.; Du, Z. Y.; Ammon, H. L. Prediction of Possible Crystal-Structures For C-, H-, N-, O- and F-Containing Organic Compounds. *J. Comput. Chem.* **1993,** *14* (4), 422-437.

14. Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction - I. Rigid molecules. *J. Comput. Chem.* **2005,** *26* (3), 304-324.

15. Stone, A. J.; Misquitta, A. J. Atom-atom potentials from ab initio calculations. *Int. Rev. Phys. Chem.* **2007,** *26* (1), 193-222.

16. *Gaussian 03,* Gaussian Inc.: Wallingford CT, 2003

17. Allen, M. J.; Tozer, D. J. Kohn-Sham calculations using hybrid exchange-correlation functionals with asymptotically corrected potentials. *Journal of Chemical Physics* **2000,** *113* (13), 5185-5192.

18. Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *Journal of Chemical Physics* **1999,** *110* (13), 6158-6170.

19. Sadlej, A. J. Medium-Size Polarized Basis-Sets for High-Level Correlated Calculations of Molecular Electric Properties. *Collect. Czech. Chem. C.* **1988,** *53* (9), 1995-2016.

20. Sadlej, A. J. Medium-Size Polarized Basis-Sets for High-Level-Correlated Calculations of Molecular Electric Properties .2. 2Nd-Row Atoms - Si Through Cl. *Theor. Chim. Acta* **1991,** *79* (2), 123-140.

21. Misquitta, A. J.; Stone, A. J. Dispersion energies for small organic molecules: first row atoms. *Mol. Phys.* **2008,** *106* (12-13), 1631-1643.

22. Misquitta, A. J.; Szalewicz, K. Intermolecular forces from asymptotically corrected density functional description of monomers. *Chem. Phys. Lett.* **2002,** *357* (3-4), 301-306.

23. Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. Dispersion energy from density-functional theory description of monomers. *Phys. Rev. Lett.* **2003,** *91* (3), art. no.-033201.

24. Misquitta, A. J.; Szalewicz, K. Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers. *J. Chem. Phys.* **2005,** *122* (21), art-214103.

25. Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *J. Chem. Phys.* **2005,** *123* (21), 214103.

Chapter 7. Conclusions and Future Work

7.1. Towards Modelling the Polarization of Organic Molecules within Crystal Structures

The charge distribution of a molecule changes when other molecules are close by. The development of the WSM polarizability model by Misquitta and co-workers[1-5] for modelling the changes in the charge density by calculating induced multipole moments, is a significant advance in our ability to quantify the stabilising induction energy. I have started to apply this model for the polarization to organic crystals by testing them in a range of structures (chapter 4), implementing it into a lattice energy minimization code (chapter 5) and then used my implementation for further research into the induction energy of a system where there was a suitable model potential (chapter 6). The polarization of the molecular charge density in the crystal may be significant even for molecules such as naphthalene, and our model qualitatively reproduces the experimentally determined polarization effect (section 4.3.2). This effect has previously been ignored in modelling organic crystals structures, with the isolated molecular charge density being used as the model for the electrostatic forces, and the effects of induction absorbed into empirically fitted repulsion-dispersion potentials. Using large cluster calculations, I have shown that the induction effect can be significant and that WSM models are effective for estimating the induction energy of a crystal structure. The rank 1 WSM model has been coded into DMACRYS, as results presented in chapter 4 suggest that the dipolar model is effective at resolving the relative difference in induction between polymorphs.

Applications of the induction calculation are limited by the availability of non-empirical potentials. As long as the induction energy is absorbed into an empirical model potential there will be an unknown amount of double-counting that cannot be removed. This prevents the accurate relative lattice energies from being determined, and also prevents us from relaxing any crystal structures because of the extra attractive force from the double counting. Before DMACRYS can be used to minimize crystal structures with induction there are several areas that require research and development. Non-empirical potentials need to be developed to evaluate lattice energies accurately. However, my research (chapter 6) has shown that improvements to the damping model, which represents the effect of the overlapping charge distributions on the induction, need to be investigated to improve non-empirical potentials, and this is discussed below.

The investigations in this thesis have been for single lattice points only. Some preliminary work has been done to relax structures with induction (section 7.2, below), which highlights a problem with obtaining numerical derivatives of the induction energy to the accuracy needed for crystal structure optimisation.

7.1.1. Inadequate Damping Model

It was shown in Chapter 6 that several structures of $C_6Br_2ClFH_2$ (XIII) near the global minimum of the search could not have their induction energies converged. In these cases the induction energy would rapidly become large within a few dozen iterations, or slowly accelerate towards infinite energy over 350 iterations until the program terminates. This behaviour was attributed to the damping between bromine atoms being insufficient, which causes the induced moments to increase rapidly in some of the crystals. A further example of poorly converging induction was seen for the ε-oxalyl dihydrazide system in

Chapter 4.5.2, where the energy did converge to give anomalously large energies relative to the other polymorphs and the SCERP results. In this case the problem was identified as inadequate damping for the particularly short intermolecular contacts between hydrogen-bonded NH groups. This behaviour has been consistent in both the ORIENT clusters and DMACRYS lattice calculations, concluding that it is caused by ineffective damping in those systems.

It is the damping that is the least rigorous part of the whole model for the induction energy. The little research that has been done into damping induction suggests that the Tang-Toennies function, more commonly used to damp dispersion interactions, is a suitable approximation. Misquitta and Stone[1] note that it is unrealistic to expect a universal damping function to account for the effects of both penetration and the divergence of the multipole expansion, and that it will also depend on the sites involved, as well as the possibility of 'anti-damping' at intermediate distances, where the truncated series may need to be enhanced. There is also evidence that damping would need to be to anisotropic.[6] However, the Tang-Toennies function appeared to be a good choice in generating accurate potentials from small dimers.[7] It appeared to be suitable when using dimers of small molecules to develop the WSM model, and has also been somewhat effective for many of the crystal structures presented in this thesis. The halogenated benzene compound, XIII from Chapter 6, is a notable exception. It is sufficiently different from molecules used to develop the model that many of the low energy predicted crystal structures did not have an induction energy that converges, due to ineffective damping. The rest of the model potential is atom-atom based using distributed properties, yet the damping has a single parameter based on the molecule as a whole. It would be reasonable to expect the bromine-bromine interactions to require a different level of damping from CH-CH interactions.

I have implemented the single-parameter damping model, but changing this to have either different atom-atom damping parameters or changing the functional form is possible with some further modification of the code. A recent study[8] into polarization using clusters of water and glycine has shown that the current method of determining the damping parameter using the first ionization energy is not optimal. By varying the parameter they were able to reduce the errors in their calculations for a wide range of clusters. For small molecules containing only CHNO atoms the present model appears to be a reasonable approximation, however it is clear that determining what the damping model should be will require a significant amount of research and development.

7.2. Minimizing Lattice Energies Including Induction using Numerical Gradients

DMACRYS already has code to determine the derivatives of the lattice energy by numerical difference, although for normal procedures it is done more efficiently using analytical derivatives. I have attempted to use the numerical routines to allow crystal structures to relax under the influence of induction.

The crystal phase space is defined by 6N + 6 lattice variables, where N is the number of molecules in the asymmetric unit cell, consisting of 3 translational and rotational variables for each molecule within the cell plus three lattice vectors and angles of the cell itself. Normal lattice minimization uses analytical gradients and some analytical second derivatives to construct the Hessian, which is then periodically updated[9,10]. An analytic form for the derivatives of the induced moments and the fields is non-trivial because of the iterative method. For this reason the derivatives are calculated numerically, which is

handled by the `frcdif` subroutine that cycles over the lattice variables and calculates the energy as each is perturbed about the starting point. During this cycle, the current variable is perturbed by a small amount, Δ, and this change is applied to the lattice. `Iterate` is called to solve the induction energy for the new structure, then `frcnst` for the electrostatic energy as described in chapter 5.4. A further change by Δ is applied and the procedure is repeated, until four structures close to the starting point have been evaluated for the lattice energy. Using these four values the first derivatives are approximated using the lattice energies of the four structures with the formula

$$\frac{\partial U}{\partial x} = \frac{8 \cdot (E1 - E4) - E2 + E3}{12 \cdot \Delta}. \quad (72)$$

The second derivatives are calculated in a similar way, by perturbing two lattice variables at once. For each variable, all other variables up to the current one are passed over, completing the triangle of the matrix for +1+1, +1-1, -1+1, -1-1, +2+2, +2-2, -2+2, -2-2 times the Δ along each of the two variables, and which are the energies E1-E8, respectively. The second derivates can then be approximated from these using the formula

$$\frac{\partial^2 U}{\partial x \partial y} = \frac{16 \cdot (E1 + E4 - (E2 + E3)) - (E5 + E8 - (E6 + E7))}{48 \cdot \Delta^2}. \quad (73)$$

Once the triangle has been completed, it is copied to complete the square matrix, giving the full Hessian. From this point onwards the minimization can continue by updating the Hessian using the techniques described by Fletcher.[9,10]

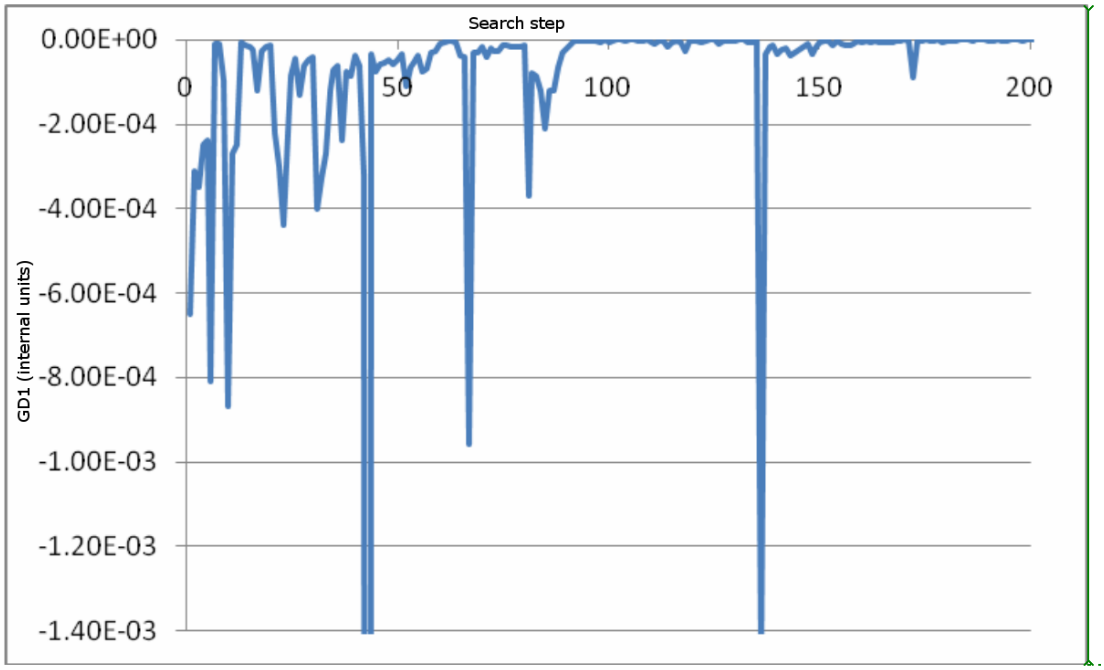
7.2.1. Numerical Noise in the Potential Energy Surface

For the minimizer to function, the potential energy surface must be relatively smooth and not contain any discontinuities, as this produces unphysical spikes in the numerical derivatives. This requires very tight convergence of the induction energy to far more

significant figures than physically reasonable. A convergence to a tolerance of 10^{-8} kJ mol^{-1} has been found sufficient for the induction energy to change smoothly for small changes in the lattice parameters.

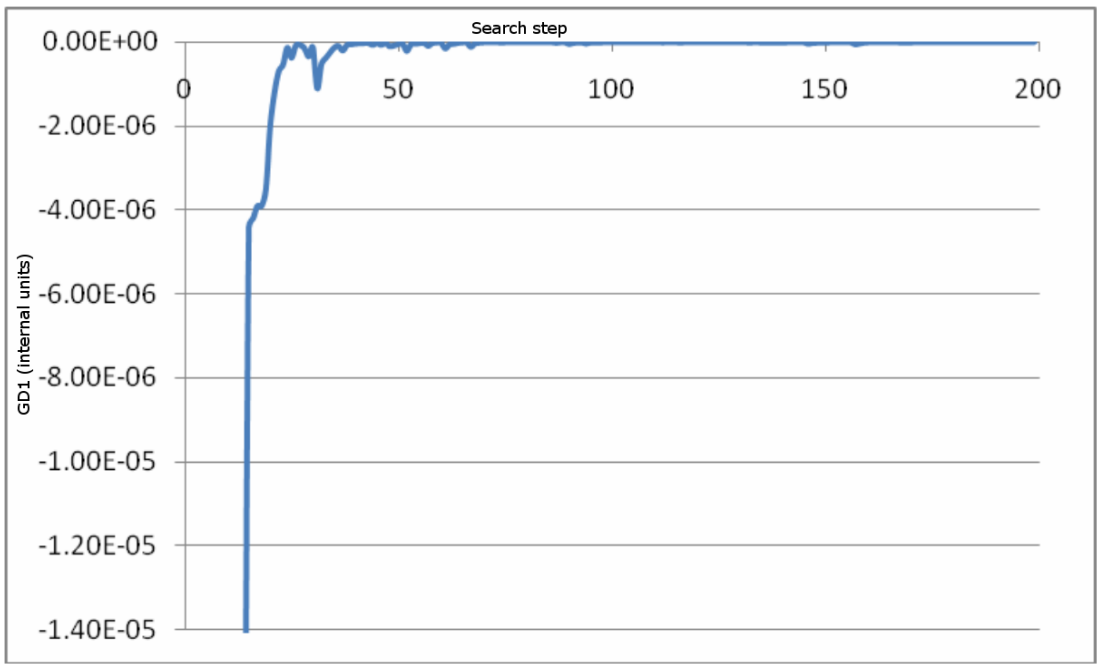
A second source of discontinuities for all contributions to the lattice energy is the cut-off used for summing the lattice energy over direct space. As the structure is relaxed, molecules move in and out of the cut-off sphere, and the larger the radius of this sphere the more molecules may lie on the boundary and move across it. Discontinuities are introduced into the potential energy surface by the induction code when there is an update of the nearest neighbour list used for the sum, where small changes to the structure mean that the tiny interaction between molecules at the edge of the cut-off are switched on and off abruptly. Figure 29 shows a plot of the first derivative (GD1) of the lattice energy of the XIII minimum from chapter 5, with respect to the lattice parameters in internal units, for the first 200 lattice perturbations. The large jumps in energy seen in Figure 29 are the result of the inconsistent neighbour list. Considering that the minimization starts from the search minimum that corresponds to the observed crystal, the derivative should not be jumping by two orders of magnitude after as many as 140 steps.

Figure 29: First derivative of the energy with respect to lattice parameters (GD1) for the XIII minimum af395. The first 200 points of a lattice minimization using induction are shown, with the nearest neighbour list updated as normal.



On the other hand, Figure 30shows the value of GD1 when the same structures is minimized with a fixed neighbour list. The derivative quickly approaches zero as the discontinuity is removed. By fixing the neighbour list to be constant the discontinuity can be removed, but as the crystallographic cell changes the sphere of interaction will become ovoid and the minimized structure will not be correct.

Figure 30: First derivative of the energy with respect to lattice parameters (GD1) shown for the XIII minimum af395. The first 200 points of a lattice minimization using induction are shown, where the nearest neighbour list is kept constant.



The discontinuity in the energy from the use of a neighbour list has also affected other codes relying on numerical gradients, such as DMAFLEX[11]. Unfortunately the handling of the interaction list is a fundamental feature of the DMACRYS, and discussions with Dr Maurice Leslie indicate that it would require significant restructuring of the minimization routines to resolve. Several solutions have been proposed, such as implementing a spline function for a region outside of the cut-off radius so that the interactions reduce smoothly to zero; also adapting the cut-off so that it is not molecular centre of mass that is a cut-off criterion, but that any molecule with atoms falling inside the sphere should be included. These are also significant developments to the code that require thorough development and testing.

The issues of damping and lattice minimization will be the subject of further research projects. This thesis has provided the assurance that modelling the induction energy in organic crystal structures is an achievable and worthwhile research pursuit.

7.3. Conclusions

In this thesis I have demonstrated that it is possible to evaluate the induction energy contribution to the lattice energy of organic crystals, using the WSM polarizability model. This model has been developed from well founded chemically theory as part of a move towards using fully *ab initio* derived intermolecular interaction models. I have compared the results of the WSM model in clusters against the SCERP model, and found that they agree on the relative stabilisation of different crystal structures due to the induction energy. Although there is a question as to the accuracy of the absolute induction energy, there is a need to implement the WSM model to at least the dipolar level so that the effect of the induction energy on the relative stability of crystals can be evaluated. I believe the effect will be most pronounced where there are competing hydrogen-bonding motifs between low energy structures, such as in the predicted low energy crystal structures of carbamazepine and 3-azabicyclo[3,3,1]nonane-2,4-dione (chapter 4). I have also shown that competing inter and intramolecular hydrogen bonding options in oxalyl dihydrazine can have a significant effect on the magnitude of the induction energy. In cases where the molecules in low energy crystal structures have competing inter and intramolecular hydrogen bonding options it is absolutely crucial to consider the induction energy when comparing relative lattice energies. This has been confirmed by periodic *ab iniito* calculations of oxalyl dihydrazide and o-acetobenzamine.[12] Studies on co-crystals[13] involve a greater number of options for intermolecular interactions where multiple components have different functional groups and varying degrees of polarizability.

Where competing low energy predicted structures sample a broad range of intermolecular interactions, they are likely to benefit from having the induction energy calculated more than where structures have broadly similar packing.

I have also demonstrated in chapter 6, and discussed above, that further research is needed in to the induction model. Research needs to be done on how best to determine the damping parameter for a molecule, if it should be dependent on atomic types which are interacting, or if a different damping function entirely should be used. Such research would increase the accuracy of the model and the range of systems for which it can be confidently used. Since I have modified DMACRYS to calculate the induced moments, it would be a relatively trivial matter to include additional functional forms of the damping, and to implement atom-atom damping parameters, to allow this research to take place. The future for crystal structure prediction is accurate models of the intermolecular interactions that are based on our best theory. We now have the tools to explicitly calculate the induction energy. In conjunction with the use of *ab initio* derived custom potentials developed alongside my own research, we are able to continue developing and testing new interaction models to improve the scope and accuracy of organic crystal structure prediction.

It is a natural requirement to be able to minimize lattices including the induction energy. In order to do this, it is essential to have *ab initio* derived potentials that exclude contributions from induction, to use in conjunction with the WSM model. However, there are several technical problems that are fundamental within the DMACRYS code, some components of which were developed more than 15 years ago, that prevent us from minimizing crystal structures right away, even with suitable potentials. While it is

possible to do in principle, the discontinuities in the induction contribution to the lattice energy described above will require a major restructuring of the code to overcome.

These issues will be the subject of future research projects, as they require a significant amount of investigation and development. This thesis has laid the foundations of these projects, by showing how important it can be to take account of the induction energy when comparing the relative lattice energies of polymorphs, and implementing and validating the model for calculating it into the lattice energy program DMACRYS.

Reference List

1. Misquitta, A. J.; Stone, A. J. Accurate induction energies for small organic molecules: 1. Theory. *J. Chem. Theory Comput.* **2008,** *4* (1), 7-18.

2. Misquitta, A. J.; Stone, A. J.; Price, S. L. Accurate induction energies for small organic molecules. 2. Development and testing of distributed polarizability models against SAPT(DFT) energies. *J. Chem. Theory Comput.* **2008,** *4* (1), 19-32.

3. *CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form,* University of Cambridge (2007), http://www-stone.ch.cam.ac.uk/programs.html#CamCASP: 2007

4. Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *J. Chem. Phys.* **2005,** *123* (21), 214103.

5. Misquitta, A. J.; Szalewicz, K. Intermolecular forces from asymptotically corrected density functional description of monomers. *Chem. Phys. Lett.* **2002,** *357* (3-4), 301-306.

6. Sanz-Garcia, A.; Wheatley, R. J. Atomic representation of the dispersion interaction energy. *Phys. Chem. Chem. Phys.* **2003,** *5* (5), 801-807.

7. Stone, A. J.; Misquitta, A. J. Atom-atom potentials from ab initio calculations. *Int. Rev. Phys. Chem.* **2007,** *26* (1), 193-222.

8. Sebetci, A.; Beran, G. J. O. Spatially Homogeneous QM/MM for Systems of Interacting Molecules with on-the-Fly ab initio Force-Field Parametrization. *J. Chem. Theory Comput.* **2010,** *6*, 155-167.

9. Fletcher, R. A New Approach to Variable Metric Algorithms. *Computer Journal* **1970,** *13* (3), 317-&.

10. Fletcher, R.; Powell, M. J. D. A Rapidly Convergent Descent Method for Minimization. *Computer Journal* **1963,** *6* (2), 163-&.

11. Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction - I. Rigid molecules. *J. Comput. Chem.* **2005,** *26* (3), 304-324.

12. Karamertzanis, P. G.; Day, G. M.; Welch, G. W. A.; Kendrick, J.; Leusen, F. J. J.; Neumann, M. A.; Price, S. L. Modeling the interplay of inter- and intramolecular hydrogen bonding in conformational polymorphs. *J. Chem. Phys.* **2008,** *128* (24), art-244708.

13. Issa, N.; Karamertzanis, P. G.; Welch, G. W. A.; Price, S. L. Can the Formation of Pharmaceutical Cocrystals Be Computationally Predicted? I. Comparison of Lattice Energies. *Cryst. Growth Des.* **2009,** *9* (1), 442-453.

Appendix A: Acronyms and Program Names

The naming convention is that program names are listed in SMALL CAPS.

CAMCASP	Cambridge package for Calculation of Anisotropic Site Properties http://www-stone.ch.cam.ac.uk/programs.html#CamCASP
CCDC	Cambridge Crystallographic Data Centre
CRYSTAL2CLUSTER	A code written by Dr. Panos Karamertzanis, which can generate clusters of molecules built from complete unit cells, from the SHELX file. It can output atomic XYZ coordinates, as well as the molecular centre of mass translations and Euler angles used by ORIENT.
CRYSTALPREDICTOR	A code written by Dr. Panos Karamertzanis, used for crystal structure prediction. It allows flexible molecules to be modelled as a set of connected rigid fragments, and may be used to predict structures in less common space groups and with more than one molecule in the asymmetric unit. For efficiency, it uses a point-charge model and is parallelized, and was mainly used as a preliminary, but extensive, structure search. *P.G. Karamertzanis and C.C. Pantelides, "Ab initio crystal structure prediction. II. Flexible molecules," Mol.Phys., 105, 273-291 (2007)*
CSP	Crystal Structure Prediction
DMA	Distributed Multipole Analysis
DMACRYS (formerly DMAREL)	The crystal lattice energy minimization code, which assumes rigid-body molecules, and uses distributed multipoles, Ewald

	summation and repulsion-dispersion potentials to model the lattice energy. I have adapted the original code to include the induction energy for a dipole-dipole polarizability model, to use custom potential types for atoms, and model dispersion using damped C_6 C_8 C_{10} coefficients.
DMAFLEX, DMAFLEXQUICK	Two related codes written by Dr. Panos Karamertzanis, which interface with an *ab initio* package and DMACRYS to include flexibility in the CSP search. Both codes extract rigid molecules from DMACRYS, perturb pre-defined bond lengths, angles and torsions, and then use this as a rigid molecule input to DMACRYS. Both codes use first and second derivatives of the crystal energy, considering inter- and intramolecular energies, however DMAFLEX performs an *ab inito* calculation each time the molecular structure is perturbed, while DMAFLEXQUICK rotates multipoles and extrapolates E_{intra} if the changes are smaller than a defined amount, making the process more computationally efficient.
FDDS	Frequency Dependent Density Susceptibility
GAUSSIAN03	An electronic structure program that predicts energies, molecular structures and molecular properties.
GDMA	Gaussian Distributed Multipole Analysis: a code written by Prof. A. J. Stone to perform DMA on the output of a GAUSSIAN03 wavefunction calculation. http://www-stone.ch.cam.ac.uk/programs.html#GDMA
MOLPAK	The code used at the beginning of a CSP search to generate

	densely packed crystal structures with one molecule in the asymmetric unit. It uses a set of geometrical rules to generate crystal structures usually in 38 space groups, and a simple point-charge and repulsion-dispersion model to efficiently minimize them prior to using more accurate models.
NEIGHCRYS (formerly DMACRYS)	A pre-processor for DMACRYS, collating the necessary input files relating to the crystallographic cell and model potentials. Its purpose is to check that the required input data are complete and consistent, and to generate the crystallographic cell from the asymmetric unit and symmetry operators.
NEIGHIND	A version of NEIGHCRYS, adapted to include the input of distributed polarizabilities.
ODH	Oxalyl Dihydrazide
ORIENT	Program for carrying out calculations of interacting molecules, using site-site potentials that include electrostatic, induction, repulsion, dispersion and charge-transfer. http://www-stone.ch.cam.ac.uk/programs.html#Orient
SAPT(DFT)	Symmetry Adapted Perturbation Theory (Density Functional Theory): an extension of SAPT that expresses the interacting molecules in terms of Kohn-Sham orbitals and orbital energies of DFT.
SCERP	Self Consistent Electronic Response to Point-charges
SHELX	A least-squared refinement program, used by crystallographers to determine structures from single-crystal diffraction. The SHELX '.res' file is a standard way to record

	the cell parameters, fractional atomic coordinates and symmetry operators that define a crystal structure.
WSM	Williams-Stone-Misquitta model of distributed molecular polarizability.
XIII	The identification code used for the 13th molecule in the CCDC blind test for the polyhalogenated benzene 1,3-dibromo-2-chloro-5-fluorobenzene

Appendix B: Programming Contributions

As well as shell scripts written to process the datasets of specific experiments, and minor changes to codes to increase the memory allocation and array sizes, to fix bugs, or to allow data to be passed between functions and subroutines, the following list describes my main modifications and codes. The naming convention is that complete programs are listed in SMALL CAPS, while subroutine and function names are listed in `fixed-width lowercase`.

`iterate`	A new subroutine in DMACRYS, which calls for the electrostatic fields to be calculated at each polarizable site, rotates the fields from the global axes to the molecular local axes, and from this, calculates the induced dipole moments due to that field, and the induction energy. It subsequently recalculates the fields due to the static and induced multipole moments, and iterates this procedure until the induction energy converges.
`frcnst`	The DMACRYS subroutine that calls the functions which return the electrostatic energy for the multipole interactions. A series of flags control whether any derivatives are also calculated. This subroutine has been modified to calculate only the energy while the induced moments are being calculated, as well as to pass the array

	that stores the fields between the energy calculation routines.
`damp`	A new subroutine in DMACRYS, which calculates the damping coefficient for a multipole interaction, based on the supplied damping parameter, distance, and the order of the multipoles. It implements only the Tang-Toennies damping function, but may be trivially expanded to include other damping functions if required, and also used to damp dispersion interactions.
`chquad, chhex, octch, chdi`	DMACRYS functions for calculating the multipole interactions, which have been modified to include reference to the induction energy calculation and calls to the damping function.
`inpalph`	A new subroutine to read the polarizability data from the input file.
`direcdamp`	A new function for DMACRYS which calculates the damped charge-charge interaction in direct space.
`pairvec`	DMACRYS function that cycles through each of the pairwise interactions and calls the required functions to calculate the interaction energies of those pairs. This has been modified to pass the stored fields between functions.
`frcdif`	DMACRYS subroutine originally written to calculate numerical gradients for accurate crystal properties. It has been heavily modified in an attempt to calculate

	numerical derivatives of the lattice energy including induction so as to allow minimization of crystal structures taking the induction energy in to account.
`getcust`	A new NEIGHCRYS subroutine, which reads a file containing custom atom labels and applies them to the molecule so as to use a custom potential.
`getpol`	A new NEIGHCRYS subroutine to read the atomic polarizabilities, and match them to the symmetry generated atoms.
GDMANEIGHCRYS	A utility code based on GDMANEIGH, that accounts for input of custom potentials and polarizabilities.
GDMA_FOR	A utility code to facilitate automatic foreshortening of covalent bonds of hydrogen, for use with the Williams potential.
POL2NEIGH	A utility code that combines the CAMCASP polarizability output with the molecular geometry to produce a formatted input file for NEIGHCRYS.
CROSSPOT	A utility code to automatically generate atom-atom coefficients for the exp-6 repulsion-dispersion potential.