Towards the *a Priori* Prediction of Molecular Crystal Structures

David Stuart Coombes

ProQuest Number: 10017677
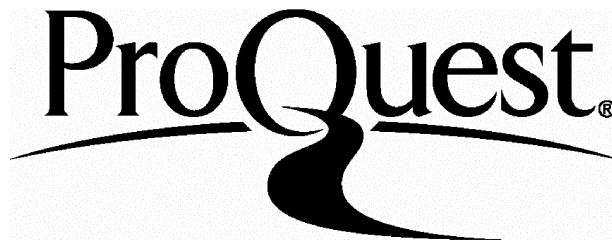
ProQuest 10017677

Previous work on modelling crystal structures of polar organic and hydrogen bonded molecules used an isotropic repulsion-dispersion potential together with a point charge electrostatic model. However, this model does not describe accurately the electrostatic forces arising from the non spherical features of the molecular charge distribution such as lone pair and π electron density. The electrostatic forces can be described far more accurately using a Distributed Multipole Analysis (DMA) of an *ab initio* wavefunction, which describes the molecular charge distribution as sets of multipoles (charge, dipole, quadrupole etc.) on every atomic site. This thesis investigates the ability of the DMA model, together with an empirical repulsion-dispersion potential, to reproduce the crystal structures of a range of organic molecules, from hydrocarbons to those which contain mixed functional groups and hydrogen bonds.

The optimisation of the empirical repulsion-dispersion potential parameters is also attempted, in order to optimise them for use with the DMA electrostatic model. However, little improvement in the already reasonable predicted structures is obtained using these potentials, suggesting the use of a static minimisation technique is the main limitation to improving the crystal structure predictions.

This potential is then used together with a powerful crystal structure prediction program to predict possible polymorphs of alloxan. This generates the observed structure and a number of alternative low energy crystal packings. Isolated dimer structures are also predicted. This is used to account for the unusual crystal packing and lack of hydrogen bonds in the observed crystal structure of alloxan.

This work shows that the combination of a powerful crystal structure prediction program together with a realistic model for the intermolecular forces represents a major advance in our ability to predict crystal structures, although other factors need to be taken into account before genuine *ab initio* crystal structure prediction becomes possible.

Acknowledgements

Table of Contents

Chapter 1   Overview

Many properties of molecular crystals, such as their non linear optical activity and effectiveness as explosives and the stability and bioavailability of drugs, depend on the crystal packing. For example, only molecules which pack without a centre of symmetry will have high non linear optical coefficients and so exhibit the property of Second Harmonic Generation (the frequency doubling of electromagnetic radiation). Hence, a method of crystal structure prediction is a useful tool in the design of new molecular materials and will also allow us to predict polymorphism (the formation of different crystal packings of the same molecule). In this thesis we show how the introduction of a distributed multipole based model potential for the electrostatic forces has revolutionised our ability to model crystal structures of a wide range of polar and hydrogen bonded molecules. These potentials can be combined with a powerful existing crystal structure generation program to produce a promising approach to crystal structure prediction.

In chapter 2 we discuss the main contributions to the intermolecular potential and show how accurate anisotropic model potentials for molecules can be obtained from calculations on the *ab inito* wavefunctions of the isolated molecules. The contribution that can be most readily obtained (even for polypeptides) is the electrostatic one, via a Distributed Multipole Analysis (DMA) of an *ab initio* wavefunction. The DMA model automatically represents the effects of the non spherical features of the charge distribution on the electrostatic forces. It was found necessary to predict the structures of van der Waals complexes[1], by minimising the electrostatic energy within accessible orientations.

Current model potentials for crystal structures commonly use a *6-exp* repulsion dispersion model together with a point charge electrostatic model and are discussed in chapter 2. These potentials suffer from a lack of transferability between different molecules and the results generally reinforce the view of Desiraju that this simple model of the electrostatic interaction[2] "gives an approximate description but does not stress adequately many structure defining interactions." The lack of transferability of current potentials and the evidence from modelling van der Waals complexes led us to study the effect of introducing distributed multipole models on the predicted structures of a range of molecular crystal structures.

The use of anisotropic intermolecular potentials required the development of a new static lattice energy minimisation program DMAREL[3]. The mathematics and physics behind this program (written by Dr. D.J. Willock and Dr. M. Leslie) is described in chapter 3. Also outlined is the program NEIGHBOURS, which is used to generate a DMAREL input file from crystal structure data. The work described in this thesis is the first major use of this suite of programs and details of the experience gained in using DMAREL (minimisation defaults, lattice energy summation limits etc.) are also given.

In chapter 4 we investigate the ability of a distributed multipole based model potential to reproduce the crystal structures of a wide range of polar organic molecules ranging from hydrocarbons to those of importance in pharmaceutics and non-linear optics. This is also compared with structures obtained using a point charge model, together with the same repulsion-dispersion potential. Attempts to improve the repulsion dispersion potentials by empirical fitting are also described. Further possible improvements to the model potential and methods used within DMAREL are outlined in chapter 7.

In chapter 5 the phenomenon of polymorphism is discussed and the available methods for crystal structure prediction are reviewed. This requires a method that will generate sufficient crude hypothetical structures that the observed structure and all possible polymorphs are likely to be found, plus a method of simulating the structure and a model for the intermolecular forces to refine these crude structures. Most work to date has relied on empirical repulsion-dispersion potentials for the final refinement. However, the use of a theoretically based potential will provide more confidence in the predicted hypothetical structures.

Hence in chapter 6 we illustrate the ability of a systematic method to generate possible hypothetical structures together with the distributed multipole based potential model defined in chapter 4 to provide a robust and powerful method for crystal structure prediction. This is used to obtain the observed structure plus possible polymorphs of alloxan. We also use the same distributed multipole based model potential to predict isolated dimer structures. This information is used to account for the unusual packing in the observed crystal structure which contains no conventional hydrogen bonds despite the alloxan molecule being made up solely of four C=O

groups and two N-H groups.

The main conclusions from the work presented in this thesis are summarised in chapter 7. It is shown that the combination of an existing method for generating hypothetical crystal structures together with a realistic distributed multipole based intermolecular potential for refinement of the crude structures provides a promising approach to molecular crystal structure prediction. Suggestions for improvements and further work are also given.

References for Chapter 1

1.   Buckingham, A.D.; Fowler, P.W. *Can. J. Chem.* **1985**, *63*, 2018.

2.   Desiraju, G.R. *Crystal Engineering: The Design of Organic Solids*, Elsevier, Amsterdam, 1989, p56.

3.   Willock, D.J.; Price, S.L.; Leslie, M.; Catlow, C.R.A. *J. Comput. Chem.* **1995**, *16*, 628.

Chapter 2   Theory of Intermolecular Forces

In this chapter we discuss the origin of the main contributions to the intermolecular forces between molecules. Models for the anisotropy in the intermolecular potential can be obtained from *ab initio* calculations of molecular wavefunctions. It is shown that for molecular crystal structures the only contribution that can be routinely derived in this way is the electrostatic one, the other contributions to the intermolecular potential being represented by an isotropic atom-atom model.

2.1 Introduction

The intermolecular energy of N rigid molecules (or lattice energy for a crystal structure as N becomes very large) can be written as

$$U = \sum_{i<j}^{N} U_{ij}(R_{ij},\Omega_{ij}) + \sum_{i<j<k}^{N} U_{ijk}(R_{ij},R_{ik},R_{jk},\Omega_{ijk}) + ... \tag{1}$$

The first term is the two body (pair) interaction potential for the interaction of two molecules of separation R_{ij} and orientation Ω_{ij}. The next term (the three body term) is the difference between the energy of three molecules and the sum of three pair potential terms. These non additive terms can only be routinely calculated for a few simple systems. For example, an intermolecular pair potential for argon[1] needed only the addition of a three body term to account for the liquid and solid state properties. In most crystal structure modelling work, use is made of the pairwise additive approximation by using only the first term in equation (1).

2.2 The Intermolecular Pair Potential at Long Range

When the molecules are well separated, the charge distributions can be considered to be associated with each molecule. The long range contributions to the intermolecular potential can be derived using Rayleigh Schrödinger Perturbation Theory. Consider two molecules A and B, whose ground state wavefunctions are O^A and O^B (with associated energies E_0^A and E_0^B) and excited state wavefunctions are N^A and N^B (with associated energies E_N^A and E_N^B). The perturbation operator for the system, H, can be written as

$$H = \frac{1}{4\pi\varepsilon_0}\sum_{ij}\frac{e_i^A e_j^B}{r_{ij}} \qquad (2)$$

where e_i^A refers to the ith electron associated with molecule A and r_{ij} is the inter-electron separation.

The long range pair potential may then be written, to second order,

$$
\begin{aligned}
U_{lr} = &\langle O^A O^B | H | O^A O^B\rangle \\
&- \sum_{N^B \neq O^B}\frac{|\langle O^A O^B | H | O^A N^B\rangle|^2}{E_N^B - E_O^B} - \sum_{N^A \neq O^A}\frac{|\langle O^A O^B | H | N^A O^B\rangle|^2}{E_N^A - E_O^A} \\
&- \sum_{N^A \neq O^A}\sum_{N^B \neq O^B}\frac{|\langle O^A O^B | H | N^A N^B\rangle|^2}{E_N^A - E_O^A + E_N^B - E_O^B} + \cdots .
\end{aligned} \qquad (3)
$$

At short range, where there is molecular overlap, this analytical treatment cannot be used. Although various perturbation methods of estimating the short range contributions to the intermolecular potential are being developed, the computational requirements are so great that it is only possible to evaluate these for a few points on the potential surface.

2.2.1 Electrostatic Energy

The first term in equation (3) is the electrostatic energy. This can be considered as the classical interaction of the undistorted charge distributions ρ^a and ρ^b around molecules A and B respectively and can in theory be calculated by integration over the *ab initio* charge distributions of the molecules A and B at each relative orientation.

$$U_{elec} = \frac{1}{4\pi\varepsilon_0}\int\frac{\rho^a(r_a)\rho^b(r_b)}{|r_a - r_b|}d^3r_a d^3r_b . \qquad (4)$$

It is strictly pairwise additive and can be either attractive or repulsive.

2.2.1.1 Point Charge Models

The simplest model for the electrostatic interactions is the point charge model, which assumes the charge is a superposition of spherical charge distributions. A number of methods are available to obtain atomic point charges and these can be

15

broadly divided into two categories: (a) those which can be obtained directly from the wavefunction such as Mulliken Population Analysis[2] or Weinhold-Reed Natural Population Analysis[3]; or (b) those which can be derived from the charge density distribution such as potential derived charges, or alternatively as the first term from an atomic multipole expansion. Atomic charges derived by Mulliken analysis are very basis set dependent and were not designed for use in calculating intermolecular interactions. Hence the most reliable method of calculating point charges is by fitting to the electrostatic potential around a molecule which can be obtained directly from the wavefunction[4]. However, Wiberg and Rablen have demonstrated[5] that none of the methods of obtaining atomic charges are capable of representing accurately molecular dipole moments and the electrostatic potential around a wide range of molecules.

The failure of the simple atomic point charge model is due to the inherent assumption that the charge distribution can be approximated by a superposition of spherical point charges. The electrostatic forces arise from the distribution of the valence electrons. On bonding, the charge is redistributed between the atoms so that the atomic charge densities remain spherical. The assumption that the charge densities remain spherical is a gross approximation, particularly as the valence electron density includes non-spherical features such as lone pair and π electron density. Williams and Weller[6] found that the errors in the representation of the charge distribution have a significant effect on the predicted crystal structures of the azabenzenes. In order to give satisfactory minimum energy crystal structures, they found it necessary to include additional lone pair sites off the nitrogen atoms. This also significantly improved the fit to the electrostatic potential around the molecules.

2.2.1.2 Central Multipole Models

The electrostatic energy can be obtained by using an expansion of r_{ij}^{-1} in terms of the centre of mass separation of the molecules R_{AB} and the distances r_i^A and r_j^B from each electron to the centre of mass of the molecule to which it belongs. Substituting the resulting expression for H into equation (4) gives

$$U_{elec} = \frac{1}{4\pi\varepsilon_o} \sum_{l_a,l_b,k_a,k_b} \binom{l_a+l_b}{l_a} R_{AB}^{-(l_a+l_b+1)} Q^A_{l_a,k_a} Q^B_{l_b,k_b} \overline{S}^{k_a,k_b}_{l_a,l_b,l_a+l_b}(\Omega). \qquad (5)$$

Here $\binom{l_a+l_b}{l_b}$ is the binomial coefficient, the Q_{lk} are the multipole moments and the

$\overline{S}$ functions describe the relative orientations of the molecules (they can also be used to describe the orientation dependence of other terms in the intermolecular potential[7]). This describes the electrostatic interaction energy in terms of the individual molecule multipole moments (ie. the charge ($l=0$), dipole ($l=1$), quadrupole ($l=2$), octopole ($l=3$) and hexadecapole ($l=4$) etc.). The interaction energy between multipoles l_a on molecule A and l_b on molecule B varies as R^{-n} where $n=l_a+l_b+1$.

This model cannot reliably be used in crystal structure modelling for a number of reasons. The central multipole expansion would have to be taken to a large number of terms to give even an approximate representation of the true charge distribution of many organic molecules, as it is only slowly convergent. Convergence may be slow for certain relative orientations of even fairly small molecules, as the convergence spheres, which contain all the charge distribution, are too close together[8]. The model is also not valid where the convergence spheres overlap, as is the case in the solid state[9].

2.2.1.3 Distributed Multipole Models

This suggests that the most accurate description of the charge distribution would be a combination of the above two models, that is a multipole series on every atom. The barrier to be overcome with this method is how to divide the contributions up between atomic sites. A number of methods of dividing up the charge distribution into contributions from each atom are available. One of the earliest methods is that due to Hirshfeld[10,11]. Here, the total charge density is calculated for a promolecule as a superposition of spherical neutral atoms and the charge density due to each atom is obtained by dividing up the total in proportion to the contribution each atom provides to the total charge density of the promolecule. One of the most rigorous methods is that of Atoms in Molecules (AIMs) due to Bader[12]. In this method, the charge density is partitioned by assuming that the gradient of the charge density is zero between the

atoms. However, multipoles obtained by this method are difficult to calculate and the divisions between the atoms often involve sharp edges which require high order multipoles to describe the charge distribution accurately. Other methods can also be used, in which partitioning of the charge density is defined by the basis set, such as those of Pullman and Perahia[13], Vigné-Maeder and Claverie[14] and the Cumulative Atomic Multipole Moments (CAMM) of Sokalski and Sarawyn[15]. The method used in this work is the Distributed Multipole Analysis (DMA) method of Stone[16,17], which was designed for optimum convergence and ease of calculation (see 2.2.1.3.1). All these methods should be able to describe the electrostatic potential outside the molecule to the same accuracy provided that sufficient expansion sites and order of multipoles are included. Spackman[18] demonstrated this by showing that, for several different multipolar expansions for $(HF_2)_2$, the predicted electrostatic energies were in excellent agreement provided that terms up to quadrupole were included. Alternatively, atomic multipole moments can be obtained from low temperature X-ray charge density analysis[19], but the charge distributions obtained by this method are influenced by the crystal environment.

2.2.1.3.1 Distributed Multipole Analysis

The DMA method[16,17] is based on the density matrix $\rho(r)$ of the molecule expressed in terms of Gaussian primitives ϕ which make up the basis set:

$$\rho(r) = \sum \rho_{AB}\phi_A\phi_B \tag{6}$$

The product of two Gaussians is another Gaussian at an overlap centre determined by the centres and exponents of the two original Gaussians. Hence, each term in (6) can be represented by a point multipole series at the overlap centre as in equation (5), which varies as R^{-n} where $n=l_A+l_B+1$ and l_A and l_B are the angular momentum quantum numbers of ϕ_A and ϕ_B respectively. Thus, for example, a charge ($l=0$) arises from two s orbitals and the terms arising from two p orbitals are represented by a charge, dipole and quadrupole at the overlap centre. If the two Gaussians are on the same atom, then the multipole series is also on that atom. However, if they are on different atoms, then the multipoles are 'shifted' to the nearest atom from the overlap centre using the relationship between a point multipole at one site and an infinite series at another point.

For example, in the case of linear molecules, the formula for the `shifting' process is

$$Q_{L0(z_0)} = \sum_{l=0}^{L} \binom{L}{l} (z_1 - z_0)^{L-l} Q_{l0} .$$ (7)

This gives the multipole moments Q_{L0} generated at the point z_0 on the molecular axis which represent the set of multipoles Q_{l0} at the overlap centre z_1.

Usually, the multipoles are `shifted' to the atomic sites, although this can be supplemented by extra sites, for example at the centre of C=0 bonds, for higher accuracy. Alternatively, sites on hydrogen atoms can be omitted or truncated at the charge (l=0) level for speed of calculation.

The process of `shifting' the multipoles to the nearest atomic site optimises the convergence of the DMA method and the use of distributed sites also ensures that the expansion is valid for almost all relative orientations. A large number of parameters are required to describe the charge distribution around an atom with no symmetry, although some of the components will be zero if the charge distribution has elements of symmetry. An atomic charge distribution with no symmetry will be described by a point charge, three components of dipole, five components of quadrupole, seven components of octopole and nine components of hexadecapole.

The electrostatic interaction energies can be evaluated using the multipole expansion formulae for interactions between pairs of sites. These have been given for terms varying in up to R^{-5} by Price et al.[20] The multipole series should be truncated at a given power of R rather than order of multipole. Hence, interactions up to R^{-5} should include hexadecapole-charge and octopole-dipole terms.

2.2.1.3.2 The Importance of the Electrostatic Contribution:
The Buckingham-Fowler Model

Considerable insights into the importance of anisotropy in the electrostatic contribution to the intermolecular potential were obtained by Buckingham and Fowler[21]. They were able to predict the structures of over two dozen van der Waals complexes by optimising the electrostatic interaction of the two molecules within the constraints of a hard sphere repulsion model for all but the hydrogen atoms (i.e. the closest approach of the two molecules was determined by the van der Waals radii of

the atoms), providing a realistic distributed electrostatic model was used. The structures of these complexes are also qualitatively predicted by the Legon-Millen rule[22], which states that hydrogen bonds will tend to form to regions of high electron density in the proton acceptor such as lone pairs and π bonds. These results were confirmed by Intermolecular Perturbation Theory (IMPT) calculations on some van der Waals complexes[23] which showed that the orientation dependence of the electrostatic contribution generally mirrored that of the total potential, as the orientation dependence of the other terms tended to cancel. The failures in the Buckingham-Fowler model could be attributed either to cases where the total potential surface is very flat (as in the case of $H_2O\cdots HF$ for example) or where the molecular shapes are not well described by the hard-sphere repulsion model, producing the wrong relative energies (for example, F-Cl$\cdots$F-H is incorrectly predicted to be more stable than Cl-F$\cdots$H-F).

The Buckingham-Fowler model has also been used to predict and account for the observed structures of van der Waals complexes involving aromatic molecules[24]. A simple isotropic model would predict binding above the centre of the aromatic ring for many of these molecules, giving a parallel plate sandwich structure. However, use of an accurate distributed multipole electrostatic model predicts a T shaped or displaced parallel benzene dimer (with the edge of one ring pointing into the π electron density of the other) as the global minimum in the potential energy surface.

2.2.2 Induction or Polarisation Energy

The second and third terms in equation (3) are expressions for the polarisation or induction energies of molecules A and B respectively. They describe the distortion of the charge density of one molecule due to the interaction with the undistorted charge density of the other.

The central multipole expansion can be used to give an expression for the induction energy in terms of the permanent multipoles and induced polarisabilities of the two molecules. The polarisabilities describe the ease with which the molecular charge distribution is distorted. The induction energy is always attractive, as the distortions in the charge density can only lower the pair interaction energy. It is extremely non additive. This is because the induced moments are determined by the electric field experienced by a molecule. For example, a charge placed next to an

atom will lead to an induced dipole. If another charge is placed symmetrically on the other side of the molecule, then there will be no induced dipole and the leading term is due to a quadrupolar polarisability.

Using the central multipole expansion to calculate the induction energy suffers from the same problems of convergence and validity as for the central multipole expansion of the electrostatic energy. In order to get a realistic description for molecules in van der Waals contact, a distributed polarisability model must be used.

One method of distributing the polarisability within the molecule is the Distributed Polarisability Analysis of Stone[25]. This model also allows for the shift of charge from one atom to another in the non-uniform electric field that arises from the proximity of other molecules. Studies using this model confirm that distributed polarisabilities are necessary for accurately calculating induced multipole moments, even for small molecules. This model is too elaborate to be used in modelling anything but simple systems, and models which transform the full description into a simpler one[26] lack transferability and accuracy. The accurate calculation of distributed polarisabilities is also basis set dependent. Hence, *ab initio* based distributed polarisability models cannot be routinely used for large molecules. The usual practice is to ignore these contributions, since they are highly non-additive and difficult to include in simulations.

2.2.3 Dispersion Energy

The last term in equation (3) is the dispersion energy. This is a purely quantum mechanical effect and represents the stabilisation resulting from the correlation of the charge fluctuations in the molecular charge distributions. Application of the multipolar expansion gives the dispersion energy between two spherical atoms (or molecules) as

$$U_{disp} = -\frac{C_6}{R_{ab}^6} - \frac{C_8}{R_{ab}^8} - \frac{C_{10}}{R_{ab}^{10}} - \dots, \tag{8}$$

where the C_n terms can be expressed as integrals over the polarisabilities at imaginary frequencies of the charge distributions of the isolated molecules. Odd powers of $n \geq 7$ arise for lower symmetry molecules. This term is additive to second order perturbation theory. In a few cases (notably for argon[1]) the three body term dispersion

term (the Axilrod-Teller term[27]) has been used to describe the interaction more accurately.

The C_n can be obtained experimentally from spectral data or from *ab initio* calculations, but these are only available for small molecules. A distributed model is needed to avoid the problems of the central multipole expansion, although a rigorous model is complicated as it has to describe charge-flow from one site to another[28]. Accurate values from *ab initio* data require very high quality wavefunctions[29] and are dependent on the method used to obtain values of the atomic polarisability[30].

2.3 Intermolecular Potential at Short Range

At the intermolecular distances found in the solid state it is no longer possible to assume that the molecules do not overlap and the charge densities cannot be assigned to one molecule or the other. The antisymmetry of electron exchange and fact that the wavefunctions are no longer orthogonal also has to be taken into account. Although there are no rigorous analytical theories for the forms of the short range terms, perturbation methods, such as that due to Hayes and Stone[31], are able to provide expressions that can estimate the various contributions, both long and short range.

2.3.1 Exchange-Repulsion

The main effects of molecular overlap are a repulsive force arising from the Pauli exclusion principle, which states that two electrons cannot occupy the same region of space if they have the same spin, and a weaker attractive force arising from electron exchange between the molecules. The resulting net repulsive force is usually represented as

$$f(R)\exp(-\alpha R) \qquad (9)$$

where $f(R)$ is a slowly varying function of the separation R. The exchange-repulsion energy is approximately additive in the case of van der Waals complexes and molecular crystals.

This term can be evaluated using supermolecule calculations or perturbation theory calculations and fitted by model potentials. The main error in these models appears to come from the assumption that the repulsive wall around each molecule is spherical. The non spherical features of the charge distribution also lead to anisotropy in the repulsive part of the potential with the functional form depending on the

22

orientation of the non spherical features of the charge distribution. Fits to *ab initio* potential energy surfaces for $(N_2)_2$[32] and $(HF)_2$[33] were found to be far more accurate when the shape of the atoms was included, by making the minimum energy separation depend on the orientation of the two atoms, than when it was isotropic.

Wheatley and Price[34] have developed a method of determining the form and parameters for exchange repulsion interaction by assuming that it is proportional to the overlap of the undistorted charge distributions. This method was tested by comparing the results with IMPT calculations on repulsion surfaces for N_2, F_2 and Cl_2. Fitting only the proportionality constant and minor adjustments of other parameters gave very good reproductions of the IMPT results. More recently Stone and Tong[35] have developed a method of investigating anisotropy in the repulsion potential using a test particle approach. In this method the repulsion potential between two molecules is determined as follows. The repulsion potential between the test particle and each molecule at a number of different configurations is calculated and the repulsion energies are fitted to a suitable potential function. Finally, the potential between the two molecules is obtained by using suitable combining rules.

2.3.2 Charge Transfer

This term arises at second order in perturbation theory and results from the transfer of charge from the occupied orbitals of one molecule to the unoccupied orbitals of another. The charge transfer energy can be obtained from perturbation theory calculations[36] and has an approximately exponential decay with separation. This term is very non additive and its effect is difficult to estimate as it is very basis set dependent. Hence this term is usually ignored and assumed to be absorbed by the exponential terms in model potentials.

2.3.3 Damping of Long Range Terms

Molecular overlap also modifies the long range electrostatic, induction and dispersion contributions. They can no longer be described by the R^{-n} form of the multipole expansion. This avoids the singularity in the separation dependence of the energy as $R \to 0$. The effects of long range damping can be significant at the distances found in the solid state and van der Waals complexes. For example, Wheatley and Mitchell[37] demonstrated that the penetration energy for the formamide-formaldehyde complex (which is the difference between the total electrostatic energy and converged

DMA energy) is about 20% of the total energy. The forms of various damping functions for the long range contributions to the intermolecular potential have been discussed[38].

2.4 Development of Model Potentials

Table 1 summarises the methods that are available to calculate the contributions to the intermolecular pair potential using the properties of the *ab initio* molecular wavefunctions of isolated molecules. It is immediately obvious that the contribution that can be most easily obtained by *ab initio* methods is the electrostatic one. It is then possible to construct intermolecular potentials by summing the atom-atom models for the different contributions. However, the need for a high quality wavefunction and the fact that minor effects such as long range damping terms will not be absorbed when using this approach (compared with empirically fitted potentials) means that the usual practice is to empirically fit model potentials to *ab initio* data, often a complicated and 'trial and error' process. Ling and Rigby[39] developed a potential for N_2, by fitting to experimental data, which was able to account for the bulk and crystal properties, but found that the resulting potential was sensitive to the damping functions. Wheatley and Price[40] developed a potential for $(Cl_2)_2$ by combining the distributed multipole and dispersion models plus the overlap model for repulsion (in which some of the parameters were empirically adjusted), which was able to simulate the thermodynamic properties and structure over a wide range of temperatures. It is likely that the empirical fitting of the parameters for the overlap model absorbed the long range damping effects due to the overlap of the charge distribution. A model potential for the water dimer $(H_2O)_2$ developed by Millot and Stone[41] comprised a distributed multipole electrostatic model, polarisabilities up to quadrupole, anisotropic atom-atom fits to intermolecular perturbation theory calculations on the exchange-repulsion and penetration interactions and a one site anisotropic dispersion term. This potential has been shown[42] to reproduce accurately the structure of the dimer.

2.4.1 Model Potentials for Crystal Structures

Most crystal structure modelling work has been concentrated on the crystal structures of non-polar molecules, particularly hydrocarbons, where optimising the close packing is usually a reliable guide to predicting the correct crystal structure[43].

The simplest realistic model for the intermolecular pair potential that can be applied to crystal structure modelling is the so called *6-exp* potential,

$$U = \sum_{ik} U_{ik} = \sum_{ik} A_{\iota\kappa} \exp(-B_{\iota\kappa} R_{ik}) - C_{\iota\kappa}/R_{ik}^6 \qquad (10)$$

where atom *i* in one molecule and atom *k* in the other are of types ι and κ respectively. The parameters $A_{\iota\kappa}$, $B_{\iota\kappa}$ and $C_{\iota\kappa}$ are required to describe the interaction between each pair of atoms. They are usually derived by empirical fitting to a set of crystal structures and heats of sublimation. Since a large number of observables would be required to fit parameters for all interactions, it is usual to calculate heteroatom parameters using the following combining rules

$$A_{\iota\kappa} = (A_{\iota\iota} A_{\kappa\kappa})^{1/2} \ ; \ B_{\iota\kappa} = \frac{(B_{\iota\iota} + B_{\kappa\kappa})}{2} \ ; \ C_{\iota\kappa} = (C_{\iota\iota} C_{\kappa\kappa})^{1/2} \ . \qquad (11)$$

Alternatively, this can be expressed as a function of the minimum energy separation $R_{\iota\kappa}^0$, well depth $\varepsilon_{\iota\kappa}$ and steepness parameter $\lambda_{\iota\kappa}$ of each individual atom-atom interaction (as shown in Figure 1).

Filippini and Gavezzotti[44] have recently derived parameters for C, H, N, O, S and Cl by empirically fitting to a database of 217 crystal structures and heats of sublimation (available for 122 compounds). The database of crystal structures included hydrocarbons, oxahydrocarbons, azahydrocarbons, chlorohydrocarbons, sulfohydrocarbons, nitro compounds, sulfones and sulfoxides, but excluded hydrogen bonded compounds. Values of the $R_{\iota\kappa}^0$ were obtained by studying the distribution of intermolecular contacts in the crystal structures in the database and $\varepsilon_{\iota\kappa}$ were empirically fitted without using combining rules due to the large amount of data available. The "steepness" parameter $\lambda_{\iota\kappa}$ was set at an average value for all interactions and a value of $\lambda=13.5$ was chosen to give the best match between observed and calculated lattice frequencies. This method was later extended to hydrogen bonded compounds[45].

These potentials are remarkably good at reproducing the experimental heats of sublimation. The calculated lattice energies are generally within about 10% of the

experimental heat of sublimation. However, the structures are predicted less well, where the cell edges are "frequently off by as much as 5%". This potential model scheme does not explicitly include the electrostatic contribution to the electrostatic forces. This has been partially absorbed into the empirical parameters. The inclusion of the electrostatic contribution in the empirical potentials is most obvious from the very large well depth for interactions involving polar hydrogen atoms, but is also seen in the (sometimes large) deviations from the traditional combining rules for interactions involving other atoms.

The neglect of the electrostatic contribution is useful as it is computationally difficult to calculate, but this is not theoretically rigorous and so provides little confidence that the model potential will transfer to other structures, let alone hypothetical structures. However, the electrostatic contribution is important where shape alone does not determine the structure. The simplest way of including an electrostatic contribution into the model for the electrostatic forces is by placing point charges at each nuclear site. Potential derived charges are usually used as they give the most accurate model of the electrostatic potential.

The inclusion of point charges is important in determining the minimum energy crystal structure even for aromatic hydrocarbons. Williams and Starr[46] used empirical charges of $q_H=-q_c=0.153$ e, which, as well as being an improvement over a model which included no point charges, gave an excellent reproduction of the quadrupole moments of the aromatic hydrocarbons studied[47]. However the model used by Williams and Starr[46] does not contain an electrostatic contribution from the carbon atoms which are not bonded to hydrogen, whereas a realistic DMA electrostatic model shows that these atoms contribute a significant quadrupolar term that arises from the π electrons. Point charge models have also been used for modelling the crystal structures of the perchlorohydrocarbons[48], oxohydrocarbons[49], amides[50] and carboxylic acids[51].

2.5 Conclusions

These simple models for the intermolecular potential suffer from a number of drawbacks, namely lack of transferability between molecules and sensitivity to the model potential and model for the electrostatic contribution for the intermolecular potential. For example, the O···O potential[49] fitted to β-oxygen, carbon-dioxide,

trioxane, tetroxocane and succinic anhydride, which included potential derived charges, transferred reasonably well to the structures of pentoxecane, 1,4-cyclohexanedione and diglycolic anhydride. The potential did not transfer adequately to p-benzoquinone and furan where the addition of a weak hydrogen bonding potential for a C-H⋯O hydrogen bond improved the fit to these structures. Interestingly, a potential for azahydrocarbons[52], which included additional lone pair sites off the aromatic nitrogens, transferred successfully to a number of structures which were not included in the derivation of the potentials, but a number of predictions were outside the accuracy required. Similarly, the potential for amides[50] could be extended to carboxylic acids[51] by the addition of a net charge on a hydroxyl hydrogen atom.

Evidence that the electrostatic effects of non-spherical features in the charge distribution can be important in determining crystal structures is also apparent from studying the homonuclear diatomics. English and Venables[53] found that an isotropic repulsion-dispersion model plus central quadrupole moment was unable to account satisfactorily for their structures. The lowest lattice energy crystal structure can depend on whether the same total quadrupole moment is described by atomic dipole moments (approximating axial lone pair density) or atomic quadrupole moments (approximating equatorial lone pair density)[54] for the same repulsion-dispersion potential. The crystal structures of some explosive molecular materials[55] were also found to be sensitive to the detailed electrostatic model used. Mulliken charges, potential derived charges and an accurate atomic multipole description of the electrostatic potential were used in conjunction with the same isotropic repulsion-dispersion potential to model their crystal structures. Mulliken charges gave the worst structures, but both potential derived charges and atomic multipole models also gave significantly different structures.

Thus, in order to be confident that the potential used is transferable both between known and unknown structures and that the electrostatic description is realistic, better more theoretically based potentials are required.

The following chapter describes a crystal structure minimisation program which has been developed to use a distributed multipole model for the electrostatic interaction together with an isotropic repulsion-dispersion potential.

Table 1 Summary of Methods of Deriving and Applications of Main Contributions to Intermolecular Forces

Contribution	Theory / Method	Additive	Sign	Example of Application
Long Range	Perturbation Theory			
Electrostatic	DMA[16,17] CAMM[14], X-ray data[19]	Yes	$\mp$	Up to polypeptides
Dispersion	Distributed Dispersion[28]	To 2nd order	-	$(N_2)_2$, $(Cl_2)_2$, $(C_2H_2)_2$
Induction	Distributed Polarisabilities[26]	No	-	Alkanes (C_nH_{2n+2}) Fully conjugated polyenes $(C_{2n}H_{2n+2})$
Short Range	No rigorous theory, but Hayes-Stone IMPT[31] gives guidance			
Exchange-Repulsion	Overlap Model[34], Test particle model[35]	No	+	$(N_2)_2$, $(F_2)_2$, $(Cl_2)_2$ $(N_2)_2$, $(H_2S)_2$, $(C_2H_2)_2$
Penetration	Gaussian Multipoles[37]	Yes	- (except very short range)	formamide/formaldehyde guanidinium/benzene
Damping	Dispersion Damping[38]	Approx	+	H···H, He···He, Li···Li
Charge-Transfer	Stone[36]	No	-	H_2O dimer, H-bonded complexes
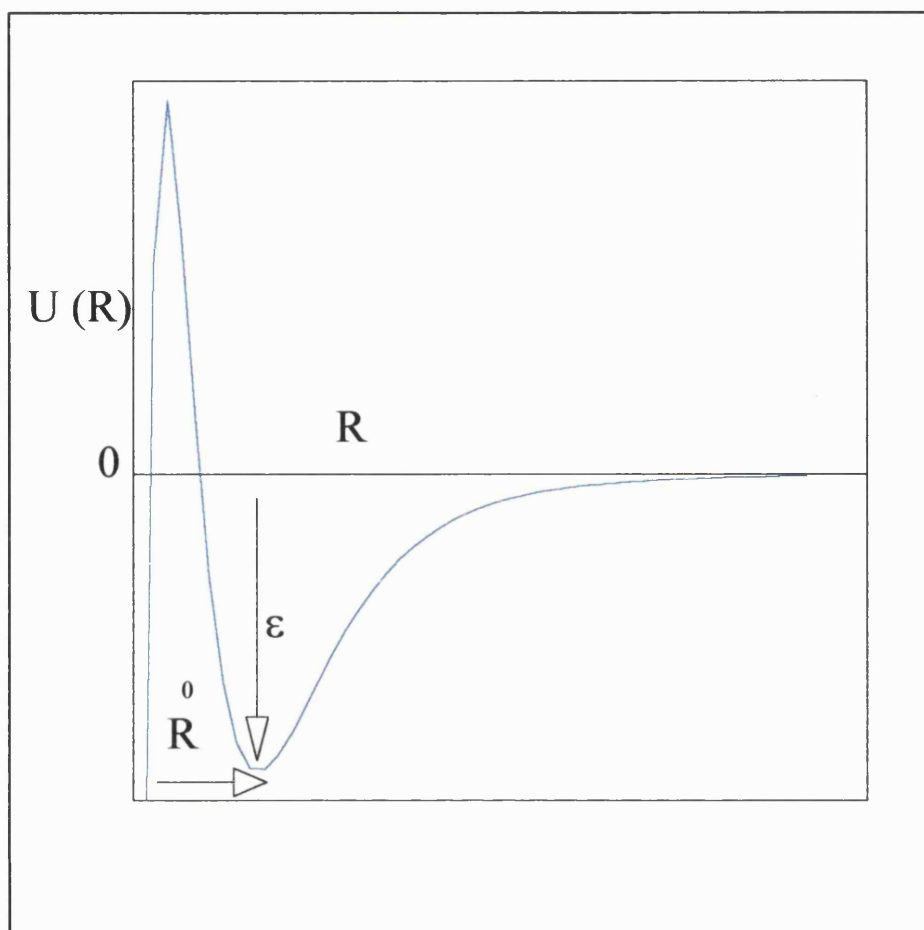
Figure 1   Schematic Representation of *6-exp* Potential Showing Well
Depth ε and Minimum Energy Separation R^0.

References For Chapter 2

1. Barker, J.A.; Fisher, R.A.; Watts, R.O. *Mol. Phys.* **1971**, *21*, 657.

2. Mulliken, R.S. *J. Chem. Phys.* **1955**, *23*, 1833.

3. Reed, A.E.; Weinstock, R.B.; Weinhold, F. *J. Chem. Phys.*, **1985**, *83*, 735.

4. Cox, S.R.; Williams, D.E. *J. Comput. Chem.* **1981**, *2*, 304.

5. Wiberg, K.B.; Rablen, P.R. *J. Comput. Chem.* **1993**, *14*, 1504.

6. Williams, D.E.; Weller, R.R. *J. Am. Chem. Soc.* **1983**, *105*, 4143.

7. Stone, A.J.; Tough, R.J.A. *Chem. Phys. Lett.* **1984**, *110*, 123.

8. Stone, A.J.; Price, S.L. *J. Phys. Chem.* **1988**, *92*, 3325.

9. Mulder, F.; Huiszoon, C. *Mol. Phys.* **1977**, *34*, 1215.

10. Hirshfeld, F.L. *Theor. Chim. Acta*, **1977**, *44*, 129.

11. Ritchie, J.P. *J. Am. Chem. Soc.* **1985**, *107*, 1829.

12. Bader, R.F.W. *Acc. Chem. Res.*, **1985**, *18*, 9.

13. Pullman, A.; Perahia, D. *Theor. Chim. Acta.*, **1978**, *48*, 29.

14. Vigné-Maeder, F.; Claverie, P. *J. Chem. Phys.* **1988**, *88*, 4934.

15. Sokalski, W.A.; Sarawyn, A. *J. Phys. Chem.* **1987**, *87*, 526.

16. Stone, A.J., *Chem. Phys. Lett.* **1981**, *83*, 233.

17. Stone, A.J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047.

18. Spackman, M.A. *J. Chem. Phys.* **1986**, *85*, 6587.

19. Ruble, J.R.; Galvao, A. *Acta Cryst.* **1995**, *B51*, 835.

20. Price, S.L.; Stone, A.J.; Alderton, M. *Mol. Phys.* **1984**, *52*, 987.

21. Buckingham, A.D.; Fowler, P.W. *Can. J. Chem.* **1985**, *63*, 2018.

22. Legon, A.C.; Millen, D.J. *Acc. Chem. Res.* **1987**, *20*, 39.

23. Hurst, G.J.B.; Fowler, P.W.; Stone, A.J.; Buckingham, A.D. *Int. J. Quantum Chem.* **1986**, *29*, 1223.

24. Price, S.L.; Stone, A.J. *J. Chem. Phys.* **1987**, *86*, 2859.

25. Stone, A.J., *Mol. Phys.* **1985**, *56*, 1065.

26. Le Sueur, C.R.; Stone, A.J. *Mol. Phys.* **1994**, *83*, 293.

27. Axilrod, B.M.; Teller, E. *J. Chem. Phys.*, **1943**, *11*, 299.

28. Stone, A.J.; Tong, C.-S. *Chem. Phys.* **1989**, *137*, 121.

29. Spackman, M.A. *J. Phys. Chem.* **1989**, *93*, 7594.

30. No, K.T.; Cho, K.H.; Jhon, M.S.; Scheraga, H.A. *J. Am. Chem. Soc.* **1993**, *115*, 2005.

31. Hayes, I.C.; Stone, A.J. *Mol. Phys.* **1984**, *53*, 83.

32. Price, S.L. *Mol. Phys.* **1986**, *58*, 651.

33. Brobjer, J.T.; Murrell, J.N. *Mol. Phys.* **1983**, *50*, 885.

34. Wheatley, R.J.; Price, S.L. *Mol. Phys.* **1990**, *69*, 507.

35. Stone, A.J.; Tong, C.-S. *J. Comput. Chem.* **1994**, *15*, 1377.

36. Stone, A.J. *Chem. Phys. Lett.* **1993**, *211*, 101.

37. Wheatley, R.J.; Mitchell, J.B.O. *J. Comput. Chem.* **1994**, *15*, 1187.

38. Wheatley, R.J.; Meath, W.J. *Mol. Phys.* **1993**, *80*, 25.

39. Ling, M.S.H.; Rigby, M. *Mol. Phys.* **1984**, *51*, 855.

40. Wheatley, R.J.; Price, S.L. *Mol. Phys.* **1990**, *71*, 1381.

41. Millot, C.; Stone, A.J. *Mol. Phys.* **1992**, *77*, 439.

42. Gregory, J.K.; Clary, D.C. *J. Phys. Chem.* **1996**, *100*, 18014.

43. Pertsin, A.J.; Kitaigorodsky, A.I. *The Atom-Atom Potential Method,* Springer-Verlag, Berlin, 1987.

44. Filippini, G.; Gavezzotti, A. *Acta Cryst.*, **1993**, *B49*, 868.

45. Gavezzotti, A.; Filippini, G. *J. Phys. Chem.* **1994**, *98*, 4831.

46. Williams, D.E.; Starr, T.L. *Comput. Chem.* **1977**, *1*, 173.

47. Price, S.L. *Chem. Phys. Lett.* **1985**, *114*, 359.

48. Hsu, L.-Y., Williams, D.E. *Acta Cryst.* **1980**, *A36*, 277.

49. Cox, S.R.; Hsu, L.-Y.; Williams, D.E. *Acta Cryst,* **1981**, *A37*, 293.

50.   Hagler, A.T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319.

51.   Lifson, S.; Hagler, A.T.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5111.

52.   Williams, D.E.; Cox, S.R. *Acta Cryst.* **1984**, *B40*, 404.

53.   English, C.A.; Venables, J.A. *Proc. Roy. Soc.* **1974**, *A340*, 57.

54.   Price, S.L. *Mol. Phys.* **1987**, *62*, 45.

55.   Ritchie, J.P.; Kober, E.M.; Copenhaver, A.S. *Proceedings of the 10th International Symposium on Detonation*, Office of Naval Research, 1993.

Chapter 3   Description of DMAREL

In this chapter we describe the methodology behind the new crystal structure minimisation program DMAREL[1] developed by Dr. D.J. Willock and Dr. M. Leslie. Also included are comments on the efficiency and working of the program. A preprocessor to DMAREL has been written as part of the work for this thesis (in conjunction with Dr. D.J. Willock) to automate the process of setting up input files for most common space groups. An instruction manual for this program (NEIGHBOURS) is given in the Appendix.

3.1 Model for the Crystal Structure

The initial geometry of the crystal structure can be taken either from the experimental X-ray or neutron diffraction data (usually from the Cambridge Structural Database[2]) or generated by a crystal structure prediction procedure. The program NEIGHBOURS is used to generate the orthogonal Cartesian coordinates of all the atoms in the unit cell and can be used to standardise the position of hydrogen atoms[3] since their location by X-ray or neutron diffraction is difficult.

The atomic multipoles (up to hexadecapole) are obtained from a DMA analysis of an *ab initio* wavefunction of the isolated molecule using an option within the program CADPAC[4]. A local axis system is generated for each molecule using DMAREL, which defines the atomic positions relative to the centre of mass of the molecule to which they belong. Within DMAREL the local axis system for each molecule must always be right handed. Thus, if a symmetry operation (such as an inversion centre) produces a left handed axis system, then the z axis is reversed to force the axis system to be right handed. NEIGHBOURS can then be used to assign the multipoles to each atomic site and change the sign of the z components of the multipoles if the z axis is reversed.

Since DMAREL uses orthogonal Cartesian coordinates, the minimisation is not constrained to a particular space group. In addition, molecules are assumed to be rigid to remove the reliance on ill-defined intramolecular potentials and reduce the number of independent variables.

The crystal structure is described in terms of a general vector $\underline{r}$, which consists of three variables describing the centres of mass of each molecule in the unit cell relative to the global axis system (R_{CMti}), three variables describing the orientation of

each molecule relative to the global axis system (θ_n) and six strain variables (E_i) describing the change in energy with size and shape of the unit cell . Thus there are 6 coordinates per molecule in the unit cell and six strain elements describing the unit cell so that, for a unit cell containing N molecules, there are $6N+6$ variables.

3.2 Minimisation Technique

To find minima in the phase space described by $\underline{r}$, the lattice energy, $U(\underline{r})$ is expanded in a Taylor expansion to second order

$$U(\underline{r}+\delta \underline{r})=U(\underline{r})+\underline{G}.\delta \underline{r}+\frac{1}{2}\delta \underline{r}.\underline{\underline{W}}.\delta \underline{r}, \tag{1}$$

where $\underline{G}$ is a vector of first derivatives and $\underline{\underline{W}}$ is a matrix of second derivatives.

At the minimum in the lattice energy the first derivative of U with respect to $\underline{r}$ is equal to zero, i.e.

$$\frac{\partial U}{\partial \delta \underline{r}}=\underline{G}+\frac{1}{2}\underline{\underline{W}}.\delta \underline{r}=0. \tag{2}$$

The step taken by the minimiser is then given by

$$\delta \underline{r}=-2\underline{\underline{W}}^{-1}.\underline{G} \tag{3}$$

subject to the constraint that the net translation of the unit cell is zero.

Equation (3) is only strictly correct for purely quadratic surfaces. However, we can assume that the potential surface is locally quadratic and so this equation can be used as part of an iterative procedure to find the nearest minimum in the lattice energy.

3.2.1 Iterative Search Methods

This iterative process involves choosing a search direction and then estimating the best distance (α) to move in that direction, so that the actual value of δr_i is given by

$$\delta r_i=\alpha \delta r_i \quad \textit{for all i.} \tag{4}$$

Hence the iterative process involves estimating the value of α by a series of line searches.

Once the search direction has been determined the best distance to move along the potential energy surface must be calculated. The maximum step size is user

controlled and the actual step taken will usually be that given by equation (3) within these limits. To ensure that the best position is found along the search direction a number of line search methods (linear interpolation or extrapolation) are employed before another search direction is chosen.

The energy drop (GD) on taking a step in the search direction can be estimated from the gradient vector (G) and step vector (D)

$$GD_n = \sum_i \left[\frac{\partial U}{\partial x_i} \right]_n \alpha_n \delta r_i \qquad (5)$$

where x_i is a structural variable and the gradient vector has been written explicitly. The subscript n indicates that the gradient and search distance has been evaluated at the nth value of α (α_1=value of α after update, α_2=value of α after first line search etc.). The initial energy drop is therefore GD_1. After a move suggested by the minimiser, the gradient is G_2 and hence a new estimated energy drop (GD_2) can be calculated for a further step in the search direction. The search moves are made relative to the starting point. After this move, a number of possibilities for the values of GD_1 and GD_2 arise.

If GD_1 and GD_2 are both negative and $|GD_1| > |GD_2|$ the step in the search direction has approached the minimum possible value but not reached it. A new search direction is chosen subject to the criterion

$$|GD_2| < \frac{1}{4}|GD_1|. \qquad (6)$$

If this criterion is not met then a further step is made in the search direction.

If $GD_2 > 0$ the step has passed over the minimum in the search direction and a linear interpolation is performed by scaling the initial step by a factor

$$\alpha = \frac{-GD_1}{GD_2 - GD_1} - 1. \qquad (7)$$

As GD_1 is negative and GD_2 is positive $0 < \alpha < 1$ and this process gives a position with

a gradient of smaller magnitude.

If $GD_2<GD_1<0$ then the gradient is becoming more negative and there is no close minimum. In this case the program will continue to search the potential surface for a position of lower energy but the search will be abandoned if no suitable point can be found. This situation, known as persistent negative curvature, may occur for a number of reasons. The nearest energy minimum may be with all atoms removed to infinity (i.e. the gas phase). In this case the final cell volume after the program fails to find a valid minimum will be very large. The other situation in which the minimiser will fall into an area of negative curvature is when a pair of atoms have come into close contact and the repulsion-dispersion potential has become unphysical. In this case the potential parameters should be carefully checked for errors and the maximum step size reduced.

Further searches are conducted if the criterion of equation (6) is not met. The old G_2 and α_2 are renamed G_3 and α_3. Hence, after one line search, three points are known whose α values are 0, α_2 and 1. A further interpolation or extrapolation can then be made using these three points (a quadratic function). The minimum position of the quadratic is then assigned to GD_2 before the line search criterion test (equation (6)) is repeated.

In practice a reasonably large maximum step size (minimum 0.5 Å for translations and 0.5° for rotations) allows satisfactory convergence. Only in one of the minimisations included in this thesis was a more frequent update of the second derivative matrix required to give a satisfactory minimum, due to the nature of the potential surface of the molecules concerned.

3.2.2 Calculation of First and Second Derivatives

In the following we will only consider the interaction of two molecules (A and B), although the extension to larger systems is straightforward and has been described in detail[1].

To obtain the minimum energy structure we thus need to obtain expressions for the first and second derivatives. Use is made of an algorithm to update the second derivative matrix based on the history of the first derivatives[5,6,7] and so not all the terms in the second derivative matrix are required. The initial second derivative matrix in the current version of DMAREL contains expressions for the position-

position, orientation-orientation and strain-strain elements (i.e. the like-like terms), with the other elements (i.e. the unlike-unlike terms) set initially to zero. Popelier and Stone have also derived expressions for the first and second derivatives with respect to molecular translation and rotation for rigid molecules[8].

We recall that the electrostatic interaction of a pair of multipole sites belonging to different molecules can be described by the equation

$$U = \frac{1}{4\pi\varepsilon_o} \sum_{l_a, l_b, k_a, k_b} \binom{l_a + l_b}{l_a} R_{AB}^{-(l_a + l_b + 1)} Q_{l_a, k_a}^A Q_{l_b, k_b}^B \bar{S}_{l_a, l_b, l_a + l_b}^{-k_a, k_b}(\Omega). \qquad (8)$$

This can be written as a function of dot products

$$U = f((R.R), (\omega_1.R), (\omega_2.R), (\omega_1.\omega_2)), \qquad (9)$$

where ω_1 and ω_2 are the local axis vectors, which allow the multipoles to be defined relative to the centre of mass of each molecule and R is the intersite vector (as shown in Figure 1) defined by

$$R = R_{CMB} - R_{CMA} + b - a. \qquad (10)$$

The first derivatives of the lattice energy with respect to the independent variables are then given by the chain rule.

So, for the derivatives with respect to R_{CMAi} and R_{CMBi} we have

$$\frac{\partial U}{\partial R_{CMAi}} = \frac{\partial U}{\partial(R.R)} \frac{\partial(R.R)}{\partial R_{CMAi}} + \sum_{\omega_1} \frac{\partial U}{\partial(\omega_1.R)} \frac{\partial(\omega_1.R)}{\partial R_{CMAi}} + \sum_{\omega_2} \frac{\partial U}{\partial(\omega_2.R)} \frac{\partial(\omega_2.R)}{\partial R_{CMAi}} \qquad (11)$$

and

$$\frac{\partial U}{\partial R_{CMBi}} = \frac{\partial U}{\partial(R.R)} \frac{\partial(R.R)}{\partial R_{CMBi}} + \sum_{\omega_1} \frac{\partial U}{\partial(\omega_1.R)} \frac{\partial(\omega_1.R)}{\partial R_{CMBi}} + \sum_{\omega_2} \frac{\partial U}{\partial(\omega_2.R)} \frac{\partial(\omega_2.R)}{\partial R_{CMBi}}. \qquad (12)$$

Similarly for the derivatives with respect to θ_{Ai} and θ_{Bi} the chain rule gives

$$\frac{\partial U}{\partial\theta_{Ai}} = \frac{\partial U}{\partial(R.R)} \frac{\partial(R.R)}{\partial\theta_{Ai}} + \sum_{\omega_1} \frac{\partial U}{\partial(\omega_1.R)} \frac{\partial(\omega_1.R)}{\partial\theta_{Ai}} + \sum_{\omega_1, \omega_2} \frac{\partial U}{\partial(\omega_1.\omega_2)} \frac{\partial(\omega_1.\omega_2)}{\partial\theta_{Ai}} \qquad (13)$$

37

and

$$\frac{\partial U}{\partial \theta_{Bi}} = \frac{\partial U}{\partial (R.R)} \frac{\partial (R.R)}{\partial \theta_{Bi}} + \sum_{\omega_2} \frac{\partial U}{\partial (\omega_2.R)} \frac{\partial (\omega_2.R)}{\partial \theta_{Bi}} + \sum_{\omega_1,\omega_2} \frac{\partial U}{\partial (\omega_1.\omega_2)} \frac{\partial (\omega_1.\omega_2)}{\partial \theta_{Bi}}. \quad (14)$$

Finally the derivative with respect to E_i is given by

$$\frac{\partial U}{\partial E_i} = \frac{\partial U}{\partial (R.R)} \frac{\partial (R.R)}{\partial E_i} + \sum_{\omega_1} \frac{\partial U}{\partial (\omega_1.R)} \frac{\partial (\omega_1.R)}{\partial E_i} + \sum_{\omega_2} \frac{\partial U}{\partial (\omega_2.R)} \frac{\partial (\omega_2.R)}{\partial E_i}. \quad (15)$$

The derivatives of U with respect to the dot products can be easily obtained from the work of Price et al^9. Here, we concentrate on calculating the dot products with respect to the independent variables.

3.2.2.1 Position Derivatives

Firstly, the derivative of $(R.R)$ with respect to R_{CMAi} is given by

$$\frac{\partial (R.R)}{\partial R_{CMAi}} = \frac{\partial}{\partial R_{CMAi}} \left(\sum_j (R_{CMB} - R_{CMA} + b - a)_j^2 \right)$$

$$= -\sum_j 2(R_{CMB} - R_{CMA} + b - a)_j \frac{\partial R_{CMAj}}{\partial R_{CMAi}} \quad (16)$$

$$= -2R_j \delta_{ij}.$$

$$\therefore \frac{\partial (R.R)}{\partial R_{CMAi}} = -2R_i.$$

Similarly the derivative of $(R.R)$ with respect to R_{CMBi} is given by

$$\frac{\partial (R.R)}{\partial R_{CMBi}} = 2R_i. \quad (17)$$

Likewise the derivatives of $(\omega_1.R)$ and $(\omega_2.R)$ with respect to R_{CMAi} and R_{CMBi} are given by

$$\frac{\partial (\omega_1.R)}{\partial R_{CMAi}} = -\omega_{1i} \quad , \quad \frac{\partial (\omega_2.R)}{\partial R_{CMAi}} = -\omega_{2i} \quad (18)$$

and

$$\frac{\partial(\omega_1.R)}{\partial R_{CMBi}} = \omega_{1i} \quad , \quad \frac{\partial(\omega_2.R)}{\partial R_{CMBi}} = \omega_{2i}. \tag{19}$$

3.2.2.2 Orientation Derivatives

Here we wish to calculate the classical torque about the centre of mass of molecule A. Consider a local charge distribution within the molecule whose centre is a distance a from the centre of mass of the molecule. We then consider an infinitesimal part m of the charge distribution ρ a distance r_m from the local centre. The torque at the centre of mass is then given by

$$T_{CMA} = \sum_{m \in \rho} R_m \times f_m, \tag{20}$$

where f_m is the force on the infinitesimal part m of the charge distribution.
Now

$$R_m = a + r_m. \tag{21}$$

Thus

$$
\begin{aligned}
T_{CMA} &= \sum_{m \in \rho} (a + r_m) \times f_m \\
&= \sum_{m \in \rho} r_m \times f_m + a \times \sum_{m \in \rho} f_m.
\end{aligned}
\tag{22}
$$

The first term in equation (22) is the torque generated at the centre of mass of the local charge distribution (local torque) and the second is the torque generated at the centre of mass of the molecule due to the net force acting on the local charge distribution (the local force), i.e.

$$T_{CMA} = T_{local} + a \times F_{local}. \tag{23}$$

So, the first derivative of U with respect to θ_{Ai} is given by

$$\frac{\partial U}{\partial \theta_{Ai}} = \left(a \times \frac{\partial U}{\partial R_{CMA}}\right)_i + \left(\frac{\partial U}{\partial \theta_{Ai}}\right)_{a=b=0} \tag{24}$$

and similarly that for U with respect to θ_{Bi} is given by

$$\frac{\partial U}{\partial \theta_{Bi}} = \left(b \times \frac{\partial U}{\partial R_{CMB}}\right)_i + \left(\frac{\partial U}{\partial \theta_{Bi}}\right)_{a=b=0}. \tag{25}$$

Since the program minimises on the centre of mass positions, there are also additional terms due to the force on the sites from the isotropic potentials that produce a torque on the centre of mass. These additional terms are

$$\left(\frac{\partial U_{SR}}{\partial \theta}\right)_{Ai} = \left(\frac{\partial U}{\partial R_{CMA}} \times a\right)_i, \tag{26}$$

for molecule A and

$$\left(\frac{\partial U_{SR}}{\partial \theta}\right)_{Bi} = \left(\frac{\partial U_{SR}}{\partial R_{CMB}} \times b\right)_i \tag{27}$$

for molecule B. The derivatives for the contributions from the isotropic potentials can be obtained from the equations for the position derivatives.

Hence, considering equation (24), the first term can be obtained using the position derivatives. The second term is equation (13) with $a=b=0$, where

$$\frac{\partial(\omega_1 . R)}{\partial \theta_{Ai}} = R. \frac{\partial \omega_1}{\partial \theta_{Ai}} = (\omega_1 \times R)_i \tag{28}$$

and

$$\frac{\partial(\omega_1 . \omega_2)}{\partial \theta_{Ai}} = (\omega_1 \times \omega_2)_i. \tag{29}$$

3.2.2.3 Strain Derivatives

To calculate the change in size and shape of the unit cell we use a strain matrix. Under the action of a strain (described by the strain matrix $\underline{E}$) a vector R is

40

transformed to R'

$$R' = R + \begin{pmatrix} E_1 & \frac{1}{2}E_6 & \frac{1}{2}E_5 \\ \frac{1}{2}E_6 & E_2 & \frac{1}{2}E_4 \\ \frac{1}{2}E_5 & \frac{1}{2}E_4 & E_3 \end{pmatrix} R = R + \underline{\underline{E}}R. \tag{30}$$

Since the molecules are rigid the vectors a and b in equation (10) remain fixed and we can thus write

$$R_{CM} = R_{CMB} - R_{CMA}. \tag{31}$$

Hence, for a general vector $\underline{u}$ equation (30) gives

$$\frac{\partial R_{CM}}{\partial E_i} = R_{CMi} u_i \quad for \ i=1,2,3 \tag{32}$$

and

$$\frac{\partial (R_{CM})}{\partial E_i} = \frac{1}{2}(R_{CMj} u_k + R_{CMk} u_j) \quad \begin{array}{l} for \ i=4,5,6 \\ with \ j=2,3,1. \\ k=3,1,2 \end{array} \tag{33}$$

The dot product $(R.R)$ is given by

$$(R.R) = \sum_j (R_{CM} + b - a)_j^2$$
$$= \sum_j (R_{CMj}^2 + 2R_{CMj}(b-a)_j + (b-a)_j^2). \tag{34}$$

The derivative of $(R.R)$ with respect to E_i is then

$$\frac{\partial (R.R)}{\partial E_i} = \sum_j 2R_{CMj} \frac{\partial R_{CMj}}{\partial E_i} + \sum_j 2(b-a)_j \frac{\partial R_{CMj}}{\partial E_i}. \tag{35}$$

For the diagonal terms this gives

41

$$\frac{\partial(R.R)}{\partial E_i} = 2(R_{CMi}^2 + R_{CMi}(b_i - a_i)) \quad \text{for } i=1,2,3 \tag{36}$$

and for the off diagonal terms

$$\frac{\partial(R.R)}{\partial E_i} = 2R_{CMj}R_{CMk} + R_{CMj}(b_k - a_k) + R_{CMk}(b_j - a_j) \quad \begin{matrix} \text{for } i=4,5,6 \\ \text{with } j=2,3,1. \\ k=3,1,2 \end{matrix} \tag{37}$$

In terms of the intersite vector R equation (36) gives

$$\frac{\partial(R.R)}{\partial E_i} = 2(R_i^2 - R_i(b_i - a_i)) \quad \text{for } i=1,2,3 \tag{38}$$

and similarly for equation (37)

$$\frac{\partial(R.R)}{\partial E_i} = 2R_jR_k - R_j(b_k - a_k) - R_k(b_j - a_j) \quad \begin{matrix} \text{for } i=4,5,6 \\ \text{with } j=2,3,1. \\ k=3,1,2 \end{matrix} \tag{39}$$

To calculate the derivatives of $(\omega_1.R)$ and $(\omega_2.R)$ with respect to E_i, let z be a general local axis vector (in this case either ω_1 or ω_2).

For the diagonal terms of E_i

$$\frac{\partial(z.R)}{\partial E_i} = z_i R_{CMi} \quad \text{for } i=1,2,3 \tag{40}$$

and for the off diagonal terms

$$\frac{\partial(z.R)}{\partial E_i} = \frac{1}{2}(z_j R_{CMk} + z_k R_{CMj}) \quad \begin{matrix} \text{for } i=4,5,6 \\ \text{with } j=2,3,1. \\ k=3,1,2 \end{matrix} \tag{41}$$

3.2.2.4 Second Derivative Matrix

The second derivative matrix for the system is a 6N+6 by 6N+6 matrix as shown in Figure 2.

The program at present includes terms expressions for the like-like terms (i.e. the second derivatives with respect to translation-translation, orientation-orientation and strain-strain. These are updated periodically using an updating algorithm based

42

on the history of the first derivatives. This can also be used to calculate the missing terms.

Calculation of the second derivatives requires taking derivatives of the first derivatives in section 3.2.2. For example, consider the first derivative of U with respect to R_{CMAi} (equation (11))

$$\frac{\partial U}{\partial R_{CMAi}} = \frac{\partial U}{\partial (R.R)} \frac{\partial (R.R)}{\partial R_{CMAi}} + \sum_{m=1}^{2} \sum_{\omega_m} \frac{\partial U}{\partial (\omega_m.R)} \frac{\partial (\omega_m.R)}{\partial R_{CMAi}}. \tag{42}$$

Taking the derivative of equation (42) with respect to R_{CMBj} gives

$$\frac{\partial^2 U}{\partial R_{CMBj} \partial R_{CMAi}} = \frac{\partial U}{\partial (R.R)} \frac{\partial^2 (R.R)}{\partial R_{CMBj} \partial R_{CMAi}} + \frac{\partial (R.R)}{\partial R_{CMBj}} \frac{\partial^2 U}{\partial^2 (R.R)} \frac{\partial (R.R)}{\partial R_{CMAi}}$$

$$+ \sum_{m=1}^{2} \sum_{\omega_m} \frac{\partial U}{\partial (\omega_m.R)} \frac{\partial^2 (\omega_m.R)}{\partial R_{CMBj} \partial R_{CMAi}}$$

$$+ \sum_{n=1}^{2} \sum_{\omega_n} \sum_{m=1}^{2} \sum_{\omega_m} \frac{\partial (\omega_n.R)}{\partial R_{CMBj}} \frac{\partial^2 U}{\partial (\omega_m.R)\partial(\omega_n.R)} \frac{\partial (\omega_m.R)}{\partial R_{CMAi}} \tag{43}$$

$$+ \sum_{m=1}^{2} \sum_{\omega_m} \frac{\partial (R.R)}{\partial R_{CMBj}} \frac{\partial^2 U}{\partial (R.R)\partial(\omega_m.R)} \frac{\partial (\omega_m.R)}{\partial R_{CMAi}}$$

$$+ \sum_{m=1}^{2} \sum_{\omega_m} \frac{\partial (\omega_m.R)}{\partial R_{CMBj}} \frac{\partial^2 U}{\partial (R.R)\partial(\omega_m.R)} \frac{\partial (R.R)}{\partial R_{CMAi}}.$$

So the new terms that need to be calculated are

$$\frac{\partial^2 (R.R)}{\partial R_{CMBj} \partial R_{CMAi}} \tag{44}$$

and

$$\frac{\partial^2 (\omega_m.R)}{\partial R_{CMBj} \partial R_{CMAi}}. \tag{45}$$

From equation (44) we obtain

$$\frac{\partial}{\partial R_{CMBj}}\left(\frac{\partial}{\partial R_{CMAi}}\sum_k R_k^2\right) = \frac{\partial}{\partial R_{CMBj}}(-2R_{CMAi}) = 0. \qquad (46)$$

However the two sites may be images of each other. If $R_{CMBj} = R_{CMAj}$ then equation (44) gives

$$\frac{\partial}{\partial R_{CMAj}}\left(\frac{\partial}{\partial R_{CMAi}}\sum_k R_k^2\right) = \frac{\partial}{\partial R_{CMAi}}(-2R_{CMAi}) = -2\delta_{ij}. \qquad (47)$$

Alternatively, if $R_{CMAi} = R_{CMBi}$ then we obtain

$$\frac{\partial}{\partial R_{CMBj}}\left(\frac{\partial}{\partial R_{CMBi}}\sum_k R_k^2\right) = \frac{\partial}{\partial R_{CMBi}}(2R_{CMBi}) = 2\delta_{ij}. \qquad (48)$$

In all cases equation (45) gives

$$\frac{\partial^2(\omega_m.R)}{\partial R_{CMBj}\partial R_{CMAi}} = \frac{\partial}{\partial R_{CMBj}}\left(\frac{\partial}{\partial R_{CMAi}}\sum_k \omega_{mk}R_k\right) = \frac{\partial}{\partial R_{CMBj}}(-\omega_{mi}) = 0. \qquad (49)$$

The equations presented above for the first and second derivatives suggest that there are relationships between the derivatives for the two molecules (A and B). This is due to the translational symmetry of a crystal lattice. Hence, it would be a useful and time-saving exercise to consider only the interactions of a site A with a site B and obtain expressions for the interactions of site B with site A using a set of rules, both when site A and B are different and when site B is a translation of site A. This is discussed in section 3.3.1.

3.3 Calculation of Lattice Sums

Within DMAREL, the energy of the crystal lattice is calculated using an atom-atom pairwise additive approximation of site-site interactions. During calculation of the lattice energy a site I interacting with a site J will have an energy of interaction u_{IJ}. However there will be an equivalent interaction, energy u_{JI}, for site J interacting with site I. Hence we can use the relationships between two sites to reduce the number of lattice sums to be calculated for both the lattice energy and derivatives as

44

described in section 3.3.1.

The lattice sums can theoretically be calculated for all interactions in the infinite lattice. However, contributions involving charge-charge, charge-dipole and dipole-dipole interactions are evaluated using an Ewald summation technique described in section 3.3.2. The higher multipole interactions are calculated using a whole molecule cutoff. Interactions are calculated for all molecules whose centres of mass are less than this cutoff. This often results in cancellation of the interactions from multipoles on different sites. For interactions involving repulsion-dispersion potentials a site-site cutoff is used. For most of the work in this thesis a whole molecule and atom-atom cutoff of 20 Å was empirically determined to ensure convergence of the lattice energy to within 1 kJ mol^{-1}. However, for the crystal structure prediction work described in Chapter 6 cutoffs of 15 Å were found to be adequate.

3.3.1 Half Summation Scheme

When calculating a lattice sum each site interacts with all other sites and their translations (the same site in other cells generated by translations of the lattice) that are within the cutoff radii defined by the pairwise additive atom-atom potential. Each site also interacts with translations of itself. The symmetry of the interaction arises from interchanging the order of sites within a given pair interaction so that site B interacting with site A is the reverse of site A interacting with site B. This symmetry in the summation suggests that it would be a time saving exercise to consider only site A interacting with site B and fill in the terms for site B interacting with site A according to the following set of rules.

Firstly consider a site A interacting with a different site B and its translations:

a) Energy terms-the energy of site B interacting with site A is equal to the energy of site A interacting with site B. Thus these terms can be filled directly.

b) Position derivatives-the position derivatives for a pair of sites are equal and opposite and so these terms should be equal and of opposite sign.

c) Orientation derivatives-there is no simple relationship between the terms for the orientation derivatives and so these terms are calculated explicitly.

d) Strain derivatives-again no relationship exists and so these terms are calculated explicitly.

For a site which interacts with its own image in a different unit cell the

following rules apply:

a) Energy terms-here we double count and so these terms are divided by two.

b) Position derivatives-these terms are zero since equal and opposite forces on the same site can only be zero.

c) Orientation derivatives-here we consider the equations for the orientation derivatives when $a=b$ and $\omega_1=\omega_2$. Most terms sum to zero and are "force like". However the term

$$\frac{\partial(\omega_1.R)}{\partial(\theta_{Ai})}=(\omega_1\times R)_i \tag{50}$$

is "energy like" since it does not change sign. Thus, this term must be divided by two to avoid double counting.

d) Strain derivatives: again these terms are "energy like" and so must be divided by two to avoid double counting.

3.3.2 Ewald Summation

As the electrostatic contributions to the intermolecular potential (and their associated derivatives), particularly charge-charge, charge-dipole and dipole-dipole terms, are slow to converge, these are calculated using an Ewald summation technique. Since DMAREL uses rigid bodies in which only the centres of mass are minimised this affects the strain derivatives from the Ewald summed energies. In the following sections the need for an Ewald summation technique is discussed and its application to calculating the point charge interaction energy is described.

3.3.2.1 Problem of Direct Summation

Consider a neutral sphere of radius R, which contains point charges ρ_- and ρ_+. If the radius of the sphere is increased by an amount δR the change in volume (ΔVol) is given by

$$\Delta Vol=\frac{4}{3}\pi((R+\delta R)^3-R^3). \tag{51}$$

The new number of charges after increasing the volume by ΔVol will be

$$\Delta\rho=\Delta Vol(\rho_-+\rho_+). \tag{52}$$

46

Now, the charge-charge energy of two charges q_1 and q_2 a distance R apart is given by

$$U_{00} = \frac{1}{4\pi\varepsilon_0} \frac{q_1 q_2}{R}.$$

(53)

So for an additional charge q_i on the surface of the sphere the energy contribution due to the increase in volume is given by

$$\Delta U = \frac{1}{4\pi\varepsilon_0} q_i \frac{4\pi}{3R}((R+\delta R)^3 - R^3)(-\rho_- + \rho_+)$$
$$= \frac{1}{4\pi\varepsilon_0} q_i \frac{4\pi}{3R}(R^3 + 3R^2\delta R - R^3)(-\rho_- + \rho_+).$$

(54)

To first order in δR the change in energy is given by

$$\Delta U = \frac{1}{4\pi\varepsilon_0} q_i 4\pi R \delta R(\rho_+ - \rho_-).$$

(55)

The total energy is then given by integration over the whole volume

$$\Delta U_{tot} = \frac{1}{4\pi\varepsilon_0} \int_0^R q_i 4\pi R(\rho_+ - \rho_-) dR$$
$$= \frac{1}{4\pi\varepsilon_0} q_i 2\pi R^2 (\rho_+ - \rho_-).$$

(56)

If the surface of the sphere is not neutral then the energy diverges. However, for a finite system $(\rho_+ - \rho_-)$ converges to zero only slowly and so direct summation cannot be used.

The following section briefly outlines the application of the Ewald summation method to calculating the charge-charge interaction energy.

3.3.2.2 Method of Ewald Summation-Application to Point Charges

The method of Ewald Summation splits the potential of uniform point charges on the atomic sites into an expression which consists of two rapidly converging series. The first of these components consists of the true charge distribution together with an opposing set of Gaussian functions at the atomic sites. The second component

consists of a set of Gaussian functions at the atomic sites which cancel out the opposing Gaussian functions of the first component.

The required potential is then given by a Fourier representation of the potential generated by the first component combined with the potential of the second component using appropriate values of the half widths of the Gaussian functions allowing both series to converge rapidly. The first component must converge rapidly in real space because, as the radius of the sphere containing the charges increases, the Gaussian functions approach delta functions. The second component is periodic and smooth allowing good convergence for its Fourier transform.

The charge-charge energy within the crystal is given by

$$
\begin{aligned}
U_{CHG} &= \frac{1}{4\pi\varepsilon_0} \frac{1}{2} \sum_{n,n'} \frac{q_n q_{n'}}{|R_{nn'}|} \\
&= \frac{1}{4\pi\varepsilon_0} \frac{1}{2} \frac{4\pi}{NV_c} \sum_K \frac{\exp(-K^2/(4\eta^2))}{(K.K)} \sum_{n,n'} q_n q_{n'} \exp(-iK.R_{nn'}) \\
&\quad + \frac{1}{4\pi\varepsilon_0} \frac{1}{2} \sum_{n,n'} q_n q_{n'} \frac{erfc(\eta|R_{nn'}|)}{|R_{nn'}|},
\end{aligned}
\tag{57}
$$

where the summation is over pairs of charges (denoted n and n') separated by a distance R (defined in equation (10)), N is the number of unit cells in the crystal, V_c is the volume of a unit cell and K is a reciprocal lattice vector. The last term in equation (57) involves an error function to improve the rate of convergence of the term and η is chosen to make the contribution from the number of terms in each part of the Ewald summation the same.

The technique is also applied to the charge-dipole and dipole-dipole interactions as well as the associated first and second derivatives.

3.4 Details of Program Timing

The time taken for a minimisation using DMAREL with a full DMA electrostatic model depends on the number and size of the molecules in the unit cell, the distance used for the cutoffs and the number of iterations required to reach a minimum. The version of DMAREL (which had no code optimisation) used for the work in this thesis took an average of 30 minutes CPU time on an R8000 processor for the minimisations given in chapter 4.

3.5 Conclusions

This chapter has discussed the method used by DMAREL to calculate minimum energy crystal structures and outlined the mathematics behind the program. In the next chapter we use DMAREL to demonstrate the ability of a realistic distributed multipole based potential model to reproduce the crystal structure of a wide range of polar organic and hydrogen bonded molecules. Further improvements to the potential by empirical fitting of the repulsion-dispersion potentials are also investigated.
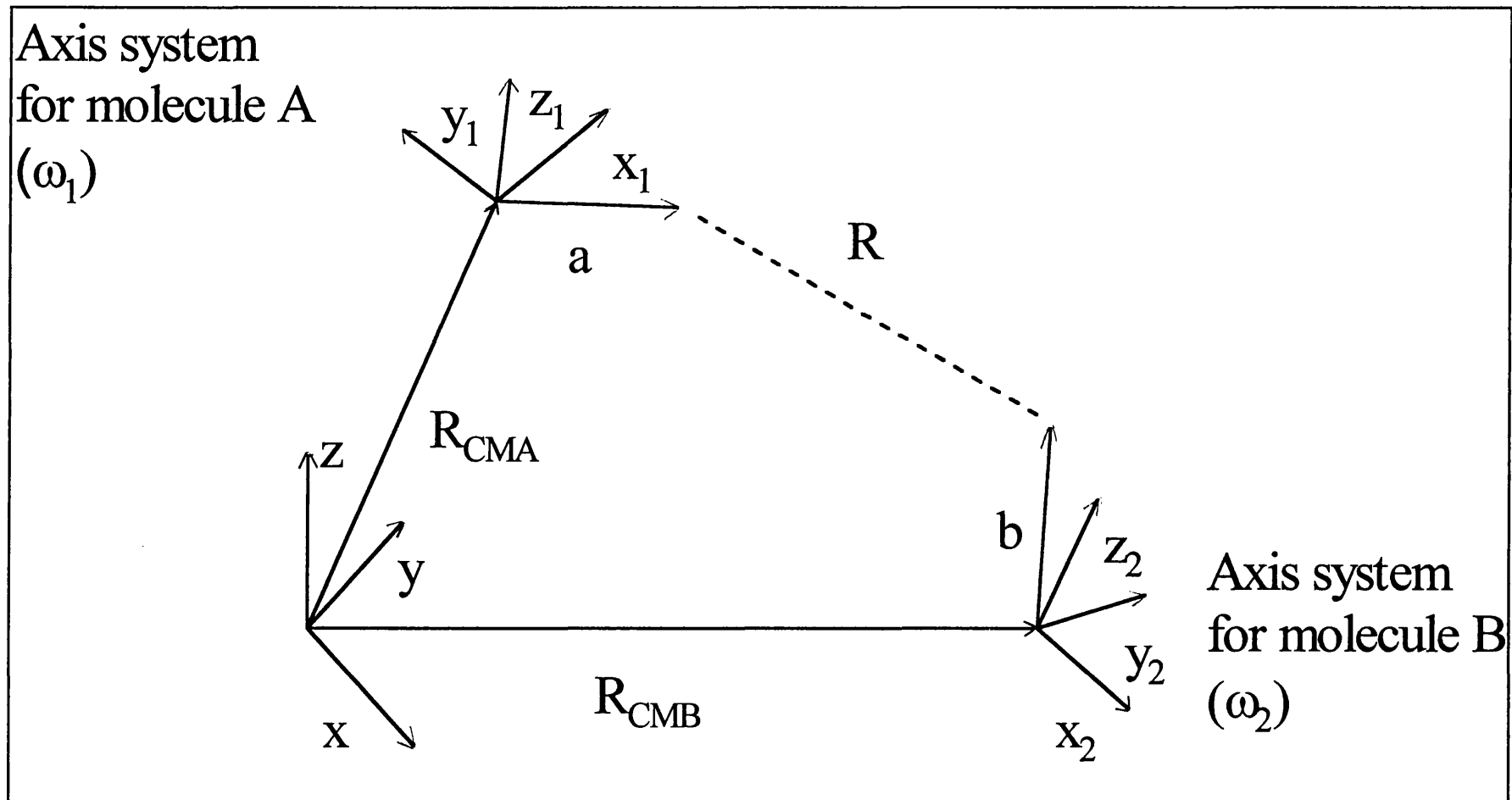
Figure 1 Interaction of Multipole Site *a* on Molecule A with Multipole Site *b* on Molecule B.

$$\begin{array}{cc}
6N \ rows & \left(\dfrac{\partial^2 U}{\partial^2 R_{CM}} \quad \dfrac{\partial^2 U}{\partial R_{CM} \partial \theta} \quad \dfrac{\partial^2 U}{\partial R_{CM} \partial E_i} \right. \\[2em]
6N \ rows & \dfrac{\partial^2 U}{\partial \theta \partial R_{CM}} \quad \dfrac{\partial^2 U}{\partial^2 \theta} \quad \dfrac{\partial^2 U}{\partial \theta \partial E_i} \\[2em]
\begin{array}{c} 6 \ rows \\ (i=1\rightarrow6) \end{array} & \left. \dfrac{\partial^2 U}{\partial E_i \partial R_{CM}} \quad \dfrac{\partial^2 U}{\partial E_i \partial \theta} \quad \dfrac{\partial^2 U}{\partial^2 E_i} \right)
\end{array}$$

Figure 2 Assignment of Elements in the Second Derivative Matrix, Which is a 6N+6 by 6N+6 Square Matrix.

References for Chapter 3

1. Willock, D.J.; Price, S.L.; Leslie, M.; Catlow, C.R.A. *J. Comput. Chem.* **1995,** *16,* 628.

2. Allen, F.H.; Kennard, O. *Chem. Design Autom. News* **1993,** *8,* 31.

3. Allen, F.H.; Kennard, O.; Watson, D.G.; Brammer, L.; Orpen, A.G.; Taylor, R. *J. Chem. Soc., Perkin Trans. 2* **1987,** S1.

4. CADPAC5: *The Cambridge Analytic Derivatives Package Issue 5,* A suite of quantum chemistry programs developed by R.D. Amos with contributions from I.L. Alberts, J.S. Andrews, S.M. Colwell, N.C. Handy, D. Jayatilaka, P.J. Knowles, R. Kobayashi, N. Koga, K.E. Laidig, P.E. Maslen, C.W. Murray, J.E. Rice, J. Sanz, E.D. Simandiras, A.J. Stone and M.-D. Su, Cambridge, U.K., 1992.

5. Fletcher, R.; Powell, M.J.D. *Computer J.* **1963,** *6,* 163.

6. Broyden, C.G. *J. Inst. Maths. Applics.* **1970,** *6,* 76 and 222.

7. Fletcher, R. *Computer J.* **1970,** *13,* 317

8. Popelier, P.L.A.; Stone, A.J. *Mol. Phys.* **1994,** *82,* 411.

9. Price, S.L.; Stone, A.J.; Alderton, M. *Mol. Phys.* **1984,** *52,* 987.

Chapter 4   The Need For a Realistic Description of the Electrostatic Forces in Determining Molecular Crystal Structures

In this chapter we demonstrate that a distributed multipole model for the electrostatic forces together with a literature based isotropic repulsion-dispersion potential is able to reproduce accurately the crystal structures of a wide range of polar organic and hydrogen bonded molecules. Attempts to optimise the repulsion-dispersion potential for use with the DMA model are also described.

4.1 Introduction

The first step in attempting to model crystal structures is to use a reasonable model for the intermolecular forces. The simplest and most widely used is the *6-exp* potential or isotropic atom-atom potential. This describes the intermolecular potential U as a sum of isotropic atom-atom repulsion and dispersion terms i.e.

$$U = \sum_{ik} U_{ik} = \sum_{ik} A_{\iota\kappa} \exp(-B_{\iota\kappa} R_{ik}) - \frac{C_{\iota\kappa}}{R_{ik}^{6}} \tag{1}$$

where atoms *i* and *k* are of types ι and κ respectively (see chapter 2).

A large number of parameter sets for interactions involving hydrocarbons, sulphur, selenium, oxygen and nitrogen atoms have been reported, and these have been summarised in a book by Pertsin and Kitaigorodsky[1]. However, these parameter sets all have shortcomings in describing the crystal structures or other physical properties of organic molecules as discussed in chapter 2. For example, Filippini and Gavezzotti have recently obtained a set of *6-exp* parameters for C, H, N, O, S and Cl atoms able to describe the crystal structures and sublimation energies of non-polar[2] and mono-functional hydrogen bonding compounds[3]. The potentials obtained are remarkably successful at predicting the lattice energies of such crystals, but the inability to reproduce accurately the experimental structures was partially ascribed to the neglect of electrostatic anisotropy.

The electrostatic forces are far larger for molecules containing heteroatoms than for saturated hydrocarbons, particularly if hydrogen bonding also occurs within the crystal. The simplest way to represent the electrostatic forces is to include the interactions between point charges on the atoms. Such a so called *6-exp-1* model is widely used in modelling for its simplicity and studies of small datasets of molecules

with either only one heteroatom or type of functional group have been carried out with varying degrees of success. However, the results obtained generally reinforce the conclusion[4] that this simple model of the intermolecular interactions "gives an approximate description but does not stress adequately many structure defining interactions". Approximating the charge distribution as a superposition of spherical atoms is incapable of giving an exact representation of the electrostatic forces outside a wide range of molecules[5]. Williams and Weller[6] found that the errors in the atomic charge model have a significant effect on the calculated crystal structures of the azabenzenes. They found it necessary to introduce additional lone pair sites off the nitrogen atoms in order to give a satisfactory minimum energy crystal structure. This also significantly improved the fit to the electrostatic potential around the molecules.

The electrostatic forces between molecules can be modelled far more accurately by representing the molecular charge distribution by sets of point multipoles (charge, dipole, quadrupole etc.) on every atomic site. Such distributed multipole electrostatic models automatically include the effects of lone pair and π electron density on the intermolecular forces. They are being increasingly widely used to model gas phase interactions, and thereby understand the structures found in van der Waals complexes of polar[7] and aromatic molecules[8], and for predicting the structures of protein side chains[9] and nucleic acid base pairs[10]. This appears to be successful since the electrostatic forces generally dominate the anisotropy of the intermolecular interactions[11]. Since the use of distributed multipole, or other accurate, electrostatic models has provided a major qualitative improvement in our ability to model van der Waals complexes, we expect the use of such models to improve likewise our ability to model crystal structures, which also contain molecules in van der Waals contact. Indeed, atomic multipole moments derived from X-ray data have already been used to account for the crystal packing characteristics of amides[12] and carboxylic acids[13].

In this chapter, the ability of a realistic electrostatic model together with a crude isotropic atom-atom model for the repulsion-dispersion potential, to model the crystal structures of a wide range of polar and hydrogen bonded molecular crystals is examined. This potential scheme was found to be able to model most of these crystal structures to a high degree of accuracy. The calculated structures are compared with those using a more approximate model, where the electrostatic interactions are only

accounted for by placing point charges on the atoms.

Attempts to improve the repulsion-dispersion potential used in these calculations by methods of least-squares fitting are then described and critically examined. These calculations are compared with other published ones and the effect of making alterations to the potential scheme for the structures which are modelled least well are investigated. Finally, the main shortcomings of the present method are discussed, with a view to stating the steps that need to be taken to improve further the modelling of such crystal structures.

4.2 Methods

4.2.1 Choice of Molecules

In this study we have considered polar organic molecules, ranging from nucleic acid bases to nitroanilines, covering examples of pharmaceutical, non-linear optic and explosive materials (see Figure 1). The molecules were chosen with some degree of arbitrariness from the Cambridge Structural Database[14], but were restricted to approximately room temperature structures which could be reasonably modelled as rigid and whose shapes were unlikely to be the main factor in determining the crystal packing. They include quasi-layered structures, but exclude most well characterized polymorphs. The molecular structures were taken directly from the experimental crystal structures with hydrogen atoms positioned to give a standard bond length[15] of 1.08 Å for C-H and 1.01 Å for N-H along the experimental bond direction.

4.2.2 Electrostatic Model

The DMA electrostatic model was obtained from a 6-31G** SCF wavefunction[16] of each isolated molecule (whose geometry is fixed) using an option within the CADPAC[17] suite of programs. This gave sets of multipoles up to hexadecapole on every atomic site. Thus, up to 25 independent multipole moments, depending on the molecular symmetry, were used to represent the electrostatic charge distribution around each atom. The electrostatic contribution to the lattice energy was calculated using all terms up to R^{-5} in the multipole expansion[18], using Ewald summation methods for the charge-charge, charge-dipole and dipole-dipole contributions to the lattice energy[19] and direct summation over all atoms in molecules whose centres are within 20 Å for the rest of the electrostatic contribution to the lattice energy. Apart from the effect of penetration into the molecular charge

distributions, the main error in calculating the electrostatic contribution to the lattice energy is the accuracy of the *ab initio* wavefunction. Comparison with the calculated dipole moments of some isolated molecules and studies of the dependence of the calculated dipole moments on the quality of the wavefunction suggests the scaling factor of 0.9 which is widely used in molecular modelling.

In this work we investigate the ability of three electrostatic models, together with an empirical isotropic atom-atom repulsion dispersion potential, to reproduce accurately the crystal structures of our database of forty molecules. The electrostatic models tested were:(a) the full DMA electrostatic model (DMA); (b) the full set of multipoles multiplied by 0.9 (0.9DMA) and (c) the charge component of the DMA (Q). Of these, 0.9DMA should be the most realistic in terms of modelling the electrostatic forces.

4.2.3 Repulsion-Dispersion Potential

Although methods of deriving repulsion and dispersion potentials from the molecular wavefunction are being developed, they cannot yet be routinely applied to organic molecules[20]. Thus, the sum of all the other contributions to the intermolecular potential was assumed to have the *6-exp* form (1), depending only on the separation of the atoms in different molecules and the types of atoms concerned. The atomic types considered were carbon (C), nitrogen (N), oxygen (O), hydrogen bonded to carbon (H_c) and hydrogen bonded to nitrogen (H_p). The distinction between polar hydrogen atoms (H_p) and non polar hydrogen atoms (H_c) is necessary to model the smaller effective van der Waals radius of the protons involved in hydrogen bonding and reflects the major difference in the electron density associated with the two types of hydrogen atoms.

The empirical repulsion dispersion parameters for C, H_c and N were taken from the work of Williams and Cox on azahydrocarbons[21], with the parameters for O from work on oxohydrocarbons[22]. The potentials for C, H_c, N and O were not designed for hydrogen bonded crystals and initial results have shown that these potentials are not able to predict accurately the crystal structures of such molecules[23]. Hence, the polar hydrogen parameters were taken from the O···H_p potential fitted to intermolecular perturbation theory calculations of the exchange-repulsion, penetration and dispersion interaction between formamide and formaldehyde in the N-H···O=C hydrogen bonding

region[24].

The parameters for heteroatom interactions were fixed using the traditional combining rules

$$A_{\iota\kappa}=\left(A_{\iota\iota}A_{\kappa\kappa}\right)^{\frac{1}{2}} \ , \ B_{\iota\kappa}=\frac{1}{2}\left(B_{\iota\iota}+B_{\kappa\kappa}\right) \ , \ C_{\iota\kappa}=\left(C_{\iota\iota}C_{\kappa\kappa}\right)^{\frac{1}{2}}. \qquad (2)$$

Although the combining rules for the $C_{\iota\kappa}$ dispersion coefficients have some physical justification at the level of the Unsöld approximation to the R^{-6} dispersion coefficients, the other rules, though widely used are poorly justified and known to be limited in their accuracy[25]. This empirical repulsion-dispersion potential will be denoted EST to emphasise that it has been derived from amongst the most appropriate literature sources and not specifically derived for the types of molecules on which it is being tested. Methods of empirical fitting to try and improve this set of repulsion-dispersion potentials are explained later.

4.2.4 Heats of Sublimation

A further test of the validity of the intermolecular potential model is provided by its ability to reproduce the experimental heats of sublimation. Where more than one experimental value was available[26], either the most recent or largest was usually used, as heats of sublimation are more easily underestimated[27]. This comparison also assumes that the molecule has the same conformation in the solid and gaseous state. If this is not the case, the calculated lattice energy and experimental heat of sublimation will differ by an amount equal to the energy difference on going from the crystal to the gas phase structure. Where experimental heats of sublimation were available, these were included as observables in the empirical fitting.

4.2.5 Empirical Fitting of Repulsion-Dispersion Potentials

To attempt to improve the literature set of repulsion-dispersion parameters EST, the method of least squares fitting is used. This involves finding a set of potential parameters which reproduce most accurately the observed crystal properties such as lattice vectors or heat of sublimation.

For N observations and M adjustable parameters, a fitting function F can be defined thus

$$F = \sum_{i=1}^{N} \left[w_i (y_i^0 - f(z_i, a_1 \ldots a_m)) \right]^2 \qquad (3)$$

Here, y_i^0 is the value of the ith observable at the experimental structure, f is the parameterised function used to determine the value of the corresponding variable z_i calculated using the parameter set a and w_i is a weighting factor used to alter the relative importance of a given set of observables and variables in determining the value of F.

In order to improve the set of repulsion-dispersion parameters, the fitting function F is minimised with respect to the parameter set a. Numerical derivatives of F are taken for a given set of repulsion-dispersion parameters. The Newton-Raphson method is then used to search for minima in the fitting function. This is analogous to performing a crystal structure relaxation except that now the potential surface is altered to put fixed structures at minima rather than moving the structures across a fixed potential surface. Since the contributions of the electrostatics remain fixed during optimisation of F, we can calculate these once at the beginning of a fit and only optimise the parameter set a.

In this work, we have used two different functions for F, based using the first derivatives of the lattice energy (gradient fitting) and also the predicted first step of the structure minimisation (first step fitting).

4.2.5.1 Gradient Fitting

A necessary condition for a minimum in the lattice energy is that the first derivatives of the lattice energy U with respect to the structural variables (forces, torques and strains) are zero at the experimental structure. The fitting function F can then be written as

$$F = \sum_{S} \left[\sum_{i} \left(w_{S_i} \frac{\partial U}{\partial x_{S_i}} \right)^2 + \left[w_{SU} (\Delta H_S - U_S) \right]^2 \right] \qquad (4)$$

In the above equation, the experimental heat of sublimation ΔH_s is also included as an observable in the fitting. The weighting factors make the contributions from each of the first derivatives (and error in the calculated lattice energy) approximately equal. Thus we alter the repulsion-dispersion parameters to give zero

forces, torques and strains at the experimental structure.

4.2.5.2 First Step Fitting

The second method which can be used is to note that during a structural relaxation, the first step taken by the minimiser can be written

$$\delta = -\underline{W}^{-1}.\underline{G} \tag{5}$$

where $\underline{G}$ is a vector of first derivatives and $\underline{W}$ a matrix of second derivatives with respect to the structural variables.

The function F can then be written

$$F = \sum_{S} \left[\sum_{x_i} (10\delta_{S_{x_i}})^2 + \sum_{\theta_i} (0.5\delta_{S_{\theta_i}})^2 + \sum_{\epsilon_i} (10\delta_{S_{\epsilon_i}})^2 + \left(\frac{\Delta H_S - U_S}{10} \right)^2 \right] \tag{6}$$

As for gradient fitting, the experimental heat of sublimation ΔH_s is also included as an observable in the fitting. The weightings used make the contributions from the first step in different types of variable approximately equal. Thus we alter the repulsion-dispersion parameters to give zero displacement on the first step of the minimisation.

4.2.5.3 Choice of Parameters to Fit

Ideally, all the parameters for all interactions (45 in total) should be optimised. However, simultaneously fitting A, B and C parameters is impossible as the A and B parameters are highly correlated. Fitting only the parameters for important interactions may well fail if the particular interaction is not sampled. This scheme may not give good transferability to other families of molecules (for example Gavezzotti and Filippini[3] found that separate potentials were required for $H_p\cdots N$ interactions in amides ($NH_2\cdots N$) and purines ($N-H\cdots N$).

Thus, in this work we have usually followed the practice of only fitting parameters for homoatoms, with heteroatom parameters obtained using combining rules.

4.2.6 Other Variations in Model

A number of structures which have been investigated contain exocyclic NH_2 groups where the C-N bond length is longer than 1.34 Å, suggesting a significant degree of partial sp^3 hybridisation. Hence, there is further uncertainty in the direction of the bond to the mobile hydrogen atoms in these groups and so we have investigated

the effect of altering the hydrogen atom positions, by making the NH_2 groups approximately co-planar with the ring systems of the molecules concerned. For benzene and indazole we also investigate the effect of crudely representing the shift of the hydrogen electron density into the bond[28] by placing their interaction sites at 1.01 Å along the C-H bonds i.e. a distance of 0.07 Å from the hydrogen nuclear positions used in determining the wavefunctions.

4.3 Results

4.3.1 Variations in Electrostatic Model

The errors in the calculated crystal structures for the theoretically best justified model, EST+0.9DMA are given in Table 1. Half this diverse range of structures are reproduced with r.m.s. errors in the lattice parameters of less than 2% (Table 2), including examples from each type of functional group. A large proportion of those structures calculated with errors of 2% to 5% include exocyclic NH_2 groups where the C-N bond lengths suggest a significant degree of sp^3 hybridisation and thus uncertainty in the position of the hydrogen atoms, limiting the accuracy of the calculated structures. Only three molecules are calculated to have a wrong structure: benzene, indazole and s-tetrazine (these are discussed in section 4.3.4).

The effect of removing the 0.9 scaling factor i.e. removing the model for electron correlation and the further removal of the anisotropic multipoles is shown in Table 2. Although scaling the electrostatic forces by 0.81 (i.e. 0.9*0.9) has a noticeable affect, it affects the lattice vectors by less than 1% on average. Comparison of the r.m.s errors using potentials EST+0.9DMA and EST+DMA clearly shows that some structures, for example cytosine, cyanuric acid, indazole and FITXIP are very sensitive to the electrostatic model, whereas others, for example pyrazole, imidazole and pyrazine are fairly insensitive. On the other hand, removing the anisotropic multipole moments has a major effect on most of the calculated structures. Only cyanuric acid and FITXIP are calculated as having less than 2% error in the lattice parameters. Many structures are poorly reproduced, with many (for example m-dinitrobenzene, allopurinol, formamide, uric acid, uracil, o-nitroaniline and p-nitroaniline) calculated to have a qualitatively wrong structure.

These calculated structures are compared with other published results in section 4.3.4.

4.3.2 Lattice Energies

Figure 2 shows the contribution of the electrostatic energy as a percentage of the total calculated lattice energy at the experimental structure. This graph shows that the electrostatic contribution varies from approximately 20% for the hydrocarbons to about 100% for some heavily hydrogen bonded structures. Thus, in a number of cases the electrostatic energy sufficiently dominates the potential such that small changes in the repulsion-dispersion potential give little change in the calculated structure.

The calculated lattice energies at the relaxed structures using the three electrostatic models are compared in Figure 3. The full DMA calculations (using EST+DMA) always give lower lattice energies at the relaxed structure than calculations using EST+0.9DMA as the electrostatic contribution is always attractive. The 0.9 scaling factor has the effect of reducing the lattice energy by less than 5 kJ mol^{-1} for the hydrocarbons to over 20 kJ mol^{-1} for the multiply hydrogen bonded systems. Hence, the reliability of the electron correlation fudge factor is important in determining the lattice energies.

The calculations using EST+Q often predict very different lattice energies at the relaxed structure from those using the full DMA. This can be attributed to the neglect of the higher multipole moments and is particularly marked for the molecules which contain amino and nitro groups. This provides further evidence that a charge only model is unable to represent accurately the electrostatic forces around some molecules[6]. However, the effect is naturally dependent on the method used to partition the charge. Other charge only models are likely to give improved predictions.

Table 3 shows the error in the calculated lattice energies using the three electrostatic models in comparison with the experimental heats of sublimation. This comparison is difficult due to the neglect of zero point energy in the calculations and also the possible high experimental error[29]. Kitaigorodsky[1] has highlighted the problems of comparing calculated lattice energies with experimental heats of sublimation and concluded that discrepancies of up to 3-4 kcal mol^{-1} (12-17 kJ mol^{-1}) should not cause any concern when judging the quality of a potential model. Many of the lattice energies are predicted within this limit by either DMA model. The EST+0.9DMA model gives the best predictions of the lattice energies for the non hydrogen bonded compounds, generally overestimated by EST+DMA. However, the

lattice energies for the hydrogen bonded compounds are generally underestimated by EST+0.9DMA. This is because the scaling of the electrostatic interactions allows for an overestimate of the electrostatic estimate of the electrostatic energy by an SCF wavefunction, but the estimated polar hydrogen parameters do not include the attractive polarisation and charge-transfer contributions and thus the well depth associated with a hydrogen bond is underestimated.

4.3.3 Fitting

We attempted to improve the repulsion-dispersion potentials, using the methods described in section 4.2.5 as they had not been designed to model such a wide range of molecular crystal structures.

Note: In the following work only, the electrostatic model for s-tetrazine was found to correspond to an excited state after the work was completed. However, it is assumed that this will have little effect on the results as s-tetrazine does not dominate the fitting process. In any case, the fitted potentials are not generally an improvement in predicting the structures over those of EST. In the following sections, only the results from a selection of fits are discussed. For values of fitted parameters for these and other fits see Table 4.

4.3.3.1 Correlation Between Sum of Squares and r.m.s Errors in Predicted Structures

The aim of empirical fitting is to obtain a set of 6-exp parameters which either (a) give zero forces, torques and strains, or (b) zero displacement on the first step of the minimisation. Hence, there should be good correlation between the sum of squares at the beginning of the fitting process and the r.m.s. percentage error in the independent lattice vectors using the starting point potential.

Figures 4a and 4b show a comparison of the initial contribution to the sum of squares against r.m.s. percentage error in the independent lattice vectors using the starting point potential for both gradient (section 4.3.3.2.1) and first step (section 4.3.3.2.2) fits. These graphs show that there is little (if any) correlation between initial sum of squares and r.m.s. percentage error. They also show that the fitting is dominated in either case by certain types of compounds. Gradient fitting is dominated by hydrogen bonded compounds, particularly uric acid. However, the sum of squares for the first step method is dominated by o-benzoquinone (see also Tables 5a and 5b).

Table 6 compares *A* parameters determined using the two fitting methods, for homoatom interactions. These results show that both methods give plausible, yet different parameters and show the same trends for changes in the parameters except for O···O. The largest change is in the N···N parameter in both cases and this is likely to be the main source of error. More restricted fits removed the problem with the large change in N···N, but produced little improvement in the structure predictions.

Thus, it would seem that the lack of correlation between the initial sum of squares and errors in the predicted structure could explain the failure of our attempts at fitting described below. This may be for a number of reasons. Firstly, the dominance of the electrostatic anisotropy in the intermolecular potential means that the structures are insensitive to fairly small (and reasonable) changes in the parameters. This can be justified by the general observation that, small changes in parameters generally produced no improvement in prediction of the crystal structures. On the other hand, fits which gave physically unreasonable parameters (usually as a result of a poorly defined fitting problem) gave generally worse structure predictions. Secondly, the failure of the first step fitting method may result from the fact that there is also a large change in the displacement after the first iteration of the minimisation. For example, for *o*-benzoquinone the translations on the first and second iteration are 0.012 Å and 0.018 Å respectively, and the corresponding rotations are 4.160° and 0.092° respectively. Thus, using the first step taken by the minimiser may not be a reliable guide to predicting the minimum in the relaxed structure.

4.3.3.2 Summary of Fits

A large number of fits using both fitting methods and making different assumptions about the parameters to be fitted have been attempted. In the following sections, only *A* parameters are fitted, except where stated.

4.3.3.2.1 Gradient Fits

The results for this set of fits are summarised in Table 4a. In order to test the effect of the weighting scheme on the fitted parameters, a related series of fits were carried out (fit PHB-PHBd in Table 4a). Here, the *A* parameters for all homoatom interactions were varied, with the heteroatom parameters calculated using combining rules. These results show that all fits except PHBb give the same parameters, and the N···N parameter consistently becomes more repulsive. The structure predictions using

63

fit PHB (summarised in Table 7) also suggest[30] that the structures that are predicted worse than the starting point are those where the π-π interactions are important. Hence, all future fits use an electrostatic model where the DMAs are multiplied by the scaling factor 0.9. An equivalent fit to that described above, but using the reduced electrostatic model EST+0.9DMA, gave parameters (fit 5, Table 4b) which were similar to those of fit PHB (Table 4a), with the N···N parameter again becoming more repulsive.

Analysis of the initial contributions to the sum of squares (Table 5a) suggests that substances containing H_p are dominating the sum of squares. Thus, all following fits in this section exclude substances containing polar hydrogens. However, with this modification, the C···C and N···N parameters become too repulsive (fit 6, Table 4b).

Since the charge distributions around N in NO_2 groups and other nitrogens are very different we then attempted a fit where N in NO_2 and other Ns were treated as separate parameters (fit 7, Table 4b). This had the effect of making the nitro group N (labelled NO) very repulsive, which is as expected since these nitrogens are largely shielded from interaction by the oxygens.

As a final attempt at improving the fitted parameters using the gradient fit method, we attempted to fit the parameters to the interactions which are important in this set of crystal structures and which were not considered in the derivation of the initial parameters (fit 10, Table 4b). However, the parameter for C···N becomes too repulsive, suggesting that either (a) the interaction is not sampled well enough or (b) the model is too flexible (i.e. too many interactions are being fitted).

4.3.3.2.2 First Step Fits

The fits using the first step fitting method are summarised in Table 4c. Again a fit to all the A parameters for homoatom interactions was attempted. This gave an increase in the A parameters for C···C and N···N and a decrease in the parameters for O···O (fit 1, Table 4c). The differences in the parameters between this fit and fit 5 of the gradient fitting method (Table 4b), can be attributed to the use of a different fitting function. The contributions to the initial sum of squares were also found to be dominated by different substances, and so for all subsequent fits, formamide was excluded. A fit to the A parameters for all homoatom interactions, analogous to fit 1 but excluding formamide gave parameters (fit 3) which differed little from those of

fit 1, the main difference being in the parameters for H_p. This result is as expected since formamide contains a large number of contacts involving H_p, which have been excluded. A further test of the ability of the fitting method to improve the parameters is given by a fit to the A parameters for all interactions (i.e. 15 in total) (fit 5). For the interactions which are important within the structures in the fitting database, the parameters remain reasonable. The largest changes are in the parameters for polar hydrogens. This is expected as these parameters are likely to be the most approximate as they were obtained from the most inappropriate source.

The final series of fits carried out on this dataset involve interactions and parameters which were not considered in the derivation of the starting point potential. A fit to the parameters for all interactions except C···C, C···H_c and H_c···H_c (fit 7) gave reasonable parameters with C···N and C···O becoming less repulsive than the starting point value and the parameters involving polar hydrogens less repulsive. A related fit (fit 10) gave parameters very different to those at the starting point, and again no improvement in the predicted structures. The other related fits (8-12) all give parameters which are not an improvement in predicting the structures on the starting point potential.

The main deficiency in the starting point potential (EST+0.9DMA) is that the polar hydrogen parameters had not been derived in conjunction with the other empirical parameters and did not include polarisation and charge-transfer contributions to the hydrogen bonding energy. The most obvious effect of this is an underestimate of the lattice energy of the hydrogen bonding substances. Thus, in order to improve the predicted lattice energies, we optimised the H_p parameters to a dataset of thirteen hydrogen bonded structures and six lattice energies, defined in Table 1. This fit[31] (hereafter referred to as fit JPC) produced a decrease in $A_{H_pH_p}$ from 7017.3 to 5029.6 kJ mol^{-1} and an increase in $C_{H_pH_p}$ from 16.4 to 21.5 kJ mol^{-1} Å^6. The structures calculated with these new parameters (potential FIT+0.9DMA) show some sensitivity to the H_p repulsion dispersion parameters, but the overall quality is the same. The agreement between the experimental heats of sublimation and the calculated lattice energies is improved somewhat, as the hydrogen bonding potential well becomes deeper (as shown in Figure 5), but they are still generally underestimated. This figure also shows the fitted O···H_p potential obtained by Mitchell and Price[24], which includes

the polarisation contributions to the formamide/formaldehyde hydrogen bonding potential. Thus, comparing the O⋯H_p(DISP+PL) potential with O⋯H_p(FIT) suggests that fitting has absorbed some of this contribution to the potential. However, the uncertainty in the electrostatic scaling factor and the probable inability of the repulsion-dispersion potential to absorb completely the charge-transfer and polarisation contributions, combined with the experimental error in the known sublimation energies, reduce the reliance that can be put on the absolute values of the calculated lattice energies.

4.3.4 Exceptions and Comparisons With Published Work

The results presented above show that a detailed, realistic electrostatic model together with an isotropic atom-atom repulsion-dispersion potential is able to account for the crystal structures of a wide range of organic molecular crystals to within a few percent error in the lattice parameters.

The effects of altering the hydrogen atom positions for those structures containing exocyclic NH_2 groups which may be partially sp^3 hybridised are shown in Table 8. These results show that some structures are very sensitive to the assumed static position of the hydrogen atoms, while others are fairly insensitive. Thus the unknown dynamic behaviour of sp^3 amine groups within crystals limits the accuracy to which the experimental structures can be reproduced, as well as the usual errors in determination of the hydrogen atom positions.

The structures which are poorly reproduced by this potential scheme are amongst the simplest: benzene, s-tetrazine and indazole. The benzene crystal undergoes a molecular rotation of 25° on minimisation with a gain of only 1.7 kJ mol^{-1} in lattice energy. With the hydrogen interaction sites shifted by 0.07 Å towards the carbon atoms along the C-H bonds an r.m.s. error of only 3.7% was obtained (see Table 8), demonstrating the sensitivity of this structure to the anisotropy of the H atom interactions. The calculated crystal structure of benzene will also be particularly dependent on the simulation method given the large thermal expansion and dynamic behaviour. The packing of the carbon aromatic ring in indazole suggests that it may be poorly calculated for similar reasons. However, the crystal structure of indazole was found to be much more sensitive to the scaling of the electrostatic forces than the position of the hydrogen interaction sites. The crystal structure of s-tetrazine is also

probably unusually sensitive to the anisotropy of the repulsion potential, as it has unusual N···N close contacts and has been reproduced accurately only by an anisotropic repulsion potential[32].

A comparison of these calculated structures with a number of other published ones is given in Table 2. The structures calculated using FIT (or EST)+0.9DMA are very comparable for all structures except benzene and 1,3,5-triamino-2,4,6-trinitrobenzene, despite the other published results using potentials optimised to a smaller family of molecules. Not surprisingly, the calculated structures for the azabenzenes are very similar to those of Williams and Weller[6], since the two models mainly differ in the use of a full DMA rather than potential derived charges with lone pair sites to represent the electrostatic potential.

4.4 Discussion

These results show that the electrostatic forces between polar and aromatic molecules clearly play an important role in determining their crystal structures. The large quantitative and sometimes qualitative changes which occur when the anisotropic multipoles are removed demonstrate clearly that the packing of most of these molecules is sensitive to the detailed model of the electrostatic forces. This implies that different charge models could give very different results, but the demonstration by Williams and Weller[6] that the azabenzene structures could not be reproduced by even a potential derived charge model without the use of additional lone pair sites shows that no atomic charge model will be able to represent accurately the electrostatic potential in the van der Waals contact region. Simpler electrostatic models than the DMA will be able to model the crystal structures of many molecules adequately. It is clearly necessary to use an electrostatic model that represents the electrostatic forces accurately to be confident that the potential model will be transferable to a wide range of molecules and crystal structures, which this study on a large number of molecules has shown to be the case. Indeed, the sensitivity of the calculated crystal structure to the detailed form of the electrostatic model has already been demonstrated for some explosive molecular crystals[33].

Even though a distributed multipole representation of the electrostatic forces is far less sensitive to the basis set used for the wavefunction than most atomic charge models, there is still a variation resulting from the change in quality of the

wavefunction. The results presented here demonstrate that the strength of the electrostatic forces relative to the other components of the potential, and hence the quality of the wavefunction used to derive the electrostatic model does affect the calculated structure as well as the energy. The scaling factor of 0.9 which is used to counteract the neglect of electron correlation on 6-31G** wavefunctions is a crude one, as the ratio of experimental to calculated dipole moments considered by Cox and Williams[34] varied from 0.78 for NH_3 to 0.96 for CH_3CN. This factor may be even more approximate for quadrupolar molecules[35].

The vast amount of data on fitting potentials presented above suggests that none of the potentials are a real improvement on EST+0.9DMA, the most well justified fits being for interactions which were not considered in the original derivation of the potentials. Comparison of the two fitting methods (Table 6) shows that the fitted parameters do not differ very significantly, as observed by Hsu and Wiliams in their work on fitting parameters for perchlorohydrocarbons[36]. The results of our work on fitting potentials can be summarised as falling into two categories. If the fitted potentials changed little from those of the starting point, then the calculated structures using the fitted potentials were overall of the same quality as those using the starting point potentials. On the other hand, if the potentials changed a lot from the starting point values, then this generally resulted in worse calculated structures than those using the starting potentials. This second case was usually the result of a poorly defined fitting problem, but not restricted to this. For example, for the fit to the A parameters for all homoatom interactions[30] (fit PHB in Table 4a), there appeared to be a large increase in the parameters involving nitrogen interactions. A common occurrence in the fits was that the sum of squares for individual parameters was just redistributed amongst the structures in the fitting dataset. The initial and final sum of squares for fits PHB and JPC are shown in Tables 5a and 5b respectively. This shows that the sum of squares for the dominant substances decreases, while the sum of squares for other substances increases during the fit. However, the fit to the polar hydrogen parameters (JPC) showed this effect less, which may be due to the fact that only two parameters are being fitted in this fit, as opposed to five in fit PHB. Thus the effect of fitting in some cases may be to just redistribute the errors in the potential among the different parameters. This suggests that the empirical fitting technique is

unable to improve the repulsion-dispersion parameters using the initial sum of squares, which was obtained by static simulation.

With the most realistic electrostatic model (EST+0.9DMA) most of the lattice parameters for the crystal structures are being predicted to within a few percent. The calculations use a static minimisation technique which, in principle, calculates the 0K structure, but this is compared with the room temperature one. This method has previously been justified by the fact that the thermal effects are minor relative to the errors produced by the potential model and that they may be partially absorbed into the fitted repulsion-dispersion parameters.

However, the effect of zero point vibrations may be significant. For example, the unit cell dimensions of n-hexane and n-octane were found to vary from 0.06 to 0.16 Å when the effects of zero point vibrations were taken into account, comparable to the effect of thermal expansion[37]. Comparisons between the low temperature structures and room temperature structures which exist for some of the molecules in our selection[38] show that the cell edges contract usually of the order of 1-2% on cooling. However, this is extremely variable and anisotropic, with individual cell edges contracting by up to 4% (p-benzoquinone) or even expanding by nearly 2% (formamide) on cooling. Therefore, even though some thermal effects will have been absorbed in the empirical fitting of the repulsion-dispersion parameters, this cannot adequately account for the temperature effects.

The approximation of a rigid molecule will be particularly poor for the NH_2 groups which are partially *sp³* hybridised and the determination of a mean position for the mobile protons of this functional group by X-ray (or neutron) diffraction will be particularly difficult. Hence, a significant proportion of the larger errors in the calculated structures using either EST+0.9DMA or FIT+0.9DMA can be attributed to the errors in representing this group (Table 8). Of course, there is some uncertainty in the experimental hydrogen atom positions for all molecules. This situation has caused Gavezzotti to locate the hydrogen atoms by standard geometrical procedures[39,40].

4.5 Conclusions

Many of the errors remaining in these crystal structures can be attributed to the use of a static simulation technique. The potential could be improved for specific

molecules if sufficient experimental or theoretical data are available, but this would most likely produce a potential scheme which would not be so readily transferable to such a wide range of molecules. Further improvements using this approach of a simple repulsion-dispersion potential plus realistic electrostatic model could be obtained by developing a dynamical simulation method.

In the following chapter the phenomenon of polymorphism is discussed and the methods available to predict molecular crystal structures are reviewed.

Table 1. Calculated Molecular Crystal Structures, Using Distributed Multipole Based Model Potentials [continues over 5 pages]

Crystal REFCODE	Experimental Structure				Errors in Calculated Structures (Calc.- Expt.) First line EST+0.9DMA, second line FIT+0.9DMA for molecules containing polar hydrogen atoms							
	a/Å	b/Å	c/Å	β/°	Δa/Å	Δb/Å	Δc/Å	Δβ/°	Δθ/°	Δx/Å	F[b]	Notes[c]
URICAC	14.464	7.403	6.208	65.10	0.436	0.029	0.030	-1.36	2.849	0.318	23.50	
					0.284	0.127	-0.074	-1.64	1.467	0.228	16.69	
GEBTUC	7.102	9.759	10.387	58.85, 67.64, 72.00	-0.294	0.241	0.502	1.35, -3.09, -2.13	3.786	0.356	78.71	
					-0.301	0.123	0.486	1.70, -2.99, -2.02	4.100	0.362	74.65	
URACIL	11.938	12.3760	3.6552	120.90	0.226	0.2803	0.0439	3.16	2.932	0.275	29.85	
					0.164	0.1933	0.0497	3.19	1.827	0.247	23.30	
AZURAC10		17.611	5.022		0.293	-0.534	-0.066		4.592	0.341	63.93	FS
					0.318	-0.752	-0.086		4.684	0.342	80.88	
ALOPUR	3.683	14.685	10.318	97.47	0.016	0.006	0.030	-2.15	2.142	0.083	6.75	
					0.011	-0.129	0.009	-2.04	2.248	0.087	7.07	
CYTSIN01	13.044	9.496	3.814		-0.069	-0.003	0.042		2.569	0.216	7.83	
					-0.277	-0.073	0.068		3.832	0.247	18.10	
CYURAC10	7.900	6.732	11.951	130.67	0.269	0.199	0.238	2.90	2.185	0.229	39.15	AΔT

71

					0.209	0.138	0.084	2.36	2.107	0.196	22.21	
PURINE	15.553	9.374	3.664		-0.214	0.186	0.000		7.533	0.168	22.82	FS
					-0.096	0.059	0.002		6.963	0.138	14.81	
PYRZOL02	8.232	12.840	7.054		-0.336	0.451	0.061		3.363	0.335	43.77	ΔT FSL
					-0.429	0.451	0.100		3.681	0.365	58.27	
IMAZOL13	7.732	5.458	9.779	117.26	0.181	-0.180	0.191	3.90	2.381	0.264	43.79	AΔT
					0.084	-0.122	0.064	3.32	3.118	0.240	25.83	
INDAZL	7.570	5.760	7.700	119.5	-0.331	0.534	0.955	7.7	11.099	0.583	383.20	FSL
					-0.278	0.549	0.696	5.9	11.545	0.478	276.62	
FITXIP	6.257	6.129	6.129	79.508, 90.000, 90.000	0.002	-0.062	-0.062	-0.787, 0.000, 0.000	0.615	0.380	17.17	FS
					-0.010	-0.135	-0.135	-1.600, 0.000, 0.000	0.631	0.467	34.02	
DAMTRZ20	10.658	10.837	4.339	90.00, 90.00, 118.83	-0.031	0.435	-0.280	0.00, 0.00, 0.84	4.842	0.185	67.72	N sp^3? FS
					-0.134	0.474	-0.334	0.00, 0.00, 1.70	6.555	0.214	98.36	
MELAMI04	10.606	7.495	7.295	112.26	-0.362	-0.141	-0.073	1.38	1.829	0.127	20.49	N sp^3?
					-0.443	-0.231	-0.170	1.26	2.079	0.157	37.48	

TETRAZ	5.23	5.79	6.63	115.25	0.34	0.43	-0.40	8.66	26.704	0.324	398.72	
PRMDIN	11.698	9.493	3.806		-0.110	0.202	-0.154		2.254	0.141	25.03	
PYRAZI	9.316	3.815	5.911		0.254	-0.074	-0.184		2.013	0.114	23.18	
JOWWIB	14.199	4.844	14.258	105.98	0.480	-0.241	-0.074	1.30	3.140	0.266	47.64	N sp^3? FS
					0.351	-0.282	-0.134	0.39	2.709	0.254	49.33	
TRPHAM	5.027	5.355	7.165	103.02, 100.90, 92.17	0.132	-0.100	-0.029	2.56, 2.51, 0.04	4.741	0.189	32.52	FS
					0.057	-0.121	-0.067	2.11, 1.67, -1.32	2.877	0.156	20.76	
BZAMID01	5.607	5.046	22.053	90.66	-0.104	0.060	0.065	-1.99	2.111	0.202	14.14	
					-0.111	0.011	0.056	-1.74	2.044	0.179	11.26	
BICWEP	16.557	10.2453	5.9400	91.840	-0.082	-0.1304	-0.0066	-0.923	0.598	0.109	4.01	
					-0.168	-0.1500	-0.0187	-0.893	0.577	0.135	5.98	
FORMAM	3.69	9.18	6.87	98.0	0.12	-0.20	-0.09	13.2	15.144	0.404	265.32	AΔT
					0.08	-0.27	-0.19	11.6	13.783	0.382	218.41	
UREAXX09	5.645	5.645	4.704		-0.075	-0.075	0.091		0.000	0.046	7.50	
					-0.167	-0.167	0.033		0.000	0.093	18.95	
ONITAN	15.45	10.01	8.57	107.4	-0.35	0.57	-0.35	0.2	3.683	0.399	73.40	N sp^3? FSL
					-0.34	0.47	-0.33	-0.2	2.993	0.344	56.03	
MNIANL01	6.499	19.369	5.084		0.264	0.285	-0.389		4.274	0.385	96.49	N sp^3? FSL

					0.337	0.491	-0.518		6.272	0.532	175.06	
NANILI21	12.337	6.037	8.597	91.42	0.045	0.096	-0.248	0.14	2.420	0.092	13.32	N sp^3? FSL
					0.086	0.060	-0.291	-0.24	2.619	0.109	15.91	
CUZDEG	3.782	27.729	7.152	95.41	0.055	-0.204	-0.031	3.15	4.439	0.271	24.97	FS
					0.057	-0.317	-0.031	3.39	4.338	0.312	29.76	
TNIOAN	6.137	9.217	15.323	99.67	0.323	-0.709	0.319	3.24	5.766	0.450	130.35	N sp^3? FSL
					0.287	-0.684	0.309	3.38	5.587	0.442	119.86	
TATNBZ	9.010	9.028	6.812	108.58, 91.82, 119.97	-0.057	-0.077	-0.584	-15.16, -6.23, 0.09	12.835	1.171	521.58	
					-0.107	-0.123	-0.597	-15.27, -6.17, 0.08	12.964	1.154	526.35	
BAZGOY	21.822	5.867	8.386	101.01	-0.309	0.275	-0.175	0.39	6.238	0.368	51.68	N sp^3?
					-0.415	0.341	-0.290	0.27	7.006	0.444	81.35	
TNBENZ10	9.78	26.94	12.82		-0.27	0.16	-0.06		1.560	0.186	12.04	
ZZZFYW01	7.945	12.975	7.421	111.88	-0.003	-0.040	-0.359	-2.05	3.933	0.158	34.01	
DNBENZ10	13.257	14.048	3.806		-0.152	-0.143	-0.021		3.945	0.157	9.00	
DNITBZ11	11.137	5.461	5.684	92.22	-0.333	-0.117	0.164	0.24	4.771	0.135	29.43	
OBNZQU	6.3137	5.7985	6.8865		0.1064	-0.0669	0.1579		3.692	0.121	14.29	
BNZQUI	7.055	6.795	5.767	101.47	-0.057	0.006	-0.117	-2.67	6.843	0.083	24.30	ΔT
NAPHTD	8.266	5.968	8.669	122.92	0.051	-0.101	-0.026	-1.33	1.934	0.058	6.39	

ANTCEN14	8.553	6.016	11.172	124.596	0.040	-0.109	-0.062	-1.640	1.637	0.068	7.66	
PHENAN08	8.441	6.140	9.438	97.96	0.165	-0.099	-0.224	-3.14	1.957	0.080	23.50	
BENZEN02	7.46	9.66	7.03		0.40	-1.93	1.10		25.290	0.799	897.55	ΔT

Footnotes for Table 1

[a] For monoclinic systems, β and $\Delta\beta$ are given. No entry implies a tetragonal or orthorhombic crystal system. For other systems, cell angles are given in the order (α, β, γ) and cell angle errors in the order ($\Delta\alpha$, $\Delta\beta$, $\Delta\gamma$).

[b] F, the structural drift factor is defined[2],

$$F=(\Delta\theta/2)^2+(10\Delta x)^2+(100\Delta a/a)^2+(100\Delta b/b)^2+(100\Delta c/c)^2+\Delta\alpha^2+\Delta\beta^2+\Delta\gamma^2$$

where r.m.s. values for $\Delta\theta$ (the net rotation of the molecule (°)) and Δx (the net translation of the centre of mass of the molecule (Å)) are used when there is more than one independent molecule in the asymmetric unit cell.

[c] N sp^3? denotes molecules in which the NH_2 group is probably significantly sp^3 hybridised as the C-N bond to the aromatic ring is longer than 1.34Å. Thus, there is additional uncertainty in the hydrogen atom position. ΔT indicates that there is a low temperature crystal structure available which suggests that the molecule has a particularly large thermal expansion coefficient (>0.03 $Å^3K^{-1}$, >0.05 $Å^3K^{-1}$ for hydrocarbons), or AΔT indicates that a cell edge expands on cooling. This suggests that the error in using static minimisation will be particularly large for these molecules. However this assessment can only be made for a minority (12) of these molecules. FS denotes structures used in the fitting of the reported H_p parameters, with FSL denoting that the lattice energy was also used.

75

Table 2 The Sensitivity of the Calculated Crystal Structure to the Intermolecular Potential   [continues for 3 pages]

	r.m.s. % Error in the lattice parameters[a]				
Crystal	Variations in electrostatic model			Other variations in potential	
	EST+0.9DMA	EST+DMA	EST+Q	FIT+0.9DMA	Published[b]
URICAC	1.8	1.3	20.8	1.7	
GEBTUC	3.9	3.3	5.4	3.7	
URACIL	1.8	1.4	25.4	1.4	
AZURAC10	4.0	4.2	8.0	4.6	
ALOPUR	0.3	0.6	11.6	0.5	
CYTSIN01	0.7	2.2	4.4	1.7	
CYURAC10	2.8	1.7	1.0	2.0	
PURINE	1.4	0.5	5.9	0.5	
PYRZOL02	3.1	3.3	0.9(nc)	3.7	
IMAZOL13	2.6	2.5	6.7	1.5	
INDAZL	9.3	5.6	6.1	7.9	
FITXIP	0.7	1.5	1.7	1.6	
DAMTRZ20	4.4	4.4	5.8	5.2	
MELAMI04	2.3	3.2	2.5	3.3	
TETRAZ	6.7	7.9	24.5		7.4 LP[6]
PRMDIN	2.7	2.5	5.5		3.0 LP[6]

76

PYRAZI	2.6	2.8	9.5		2.6 LP[6]
JOWWIB	3.5	3.1	8.3	3.7	
TRPHAM	1.9	1.5	6.0	1.6	
BZAMID01	1.3	0.9	4.4	1.2	
BICWEP	0.8	1.1	4.1	1.0	
FORMAM	2.3	3.2	12.9	2.6	4.3 LJ[41]
UREAXX09	1.7	1.6	3.0	2.2	1.2 LJ[41]
ONITAN	4.2	3.6	13.2	3.7	
MNIANL01	5.1	6.1	5.4	6.8	
NANILI21	1.9	2.1	27.2	2.1	
CUZDEG	1.0	1.4	38.0(nc)	1.1	
TNIOAN	5.5	5.5	4.3	5.2	
TATNBZ	5.0	5.3	9.0	5.2	2.0 EH [42]
BAZGOY	3.1	4.5	2.5	4.1	
TNBENZ10	1.6	1.6	3.3		
ZZZFYW01	2.8	2.9	2.8		
DNBENZ10	0.9	1.2	40.3		
DNITBZ11	2.7	2.4	3.8		
OBNZQU	1.8	1.5	3.3		
BNZQUI	1.3	1.5	8.1		4.1[22]

NAPHTD	1.1	1.8	3.2		1.3[21]
ANTCEN14	1.1	2.0	3.9		0.9[21]
PHENAN08	2.0	2.6	5.2		1.2[21]
BENZEN02	15.0	14.6	7.9		1.7[21]

Footnotes for Table 2

[a] The r.m.s. % error is calculated as

$$3\delta^2 = (100 * \Delta a/a)^2 + (100 * \Delta b/b)^2 + (100 * \Delta c/c)^2$$

except for the higher symmetry tetragonal cell of UREAXX09 and the pseudo monoclinic cell of FITXIP, where the average is taken over the two independent cell edges.

[b] Values of δ from other published work, where available, based on isotropic *6-exp* atom-atom potentials, usually with atomic charges. Significant variations in the potential form are denoted: LP additional lone pair charge sites included, LJ 12-6 repulsion-dispersion potential, EH empirical hydrogen bond potential with no electrostatic model.

nc denotes minimisations which were not properly converged.

Table 3 Lattice Energies

Model	ΔH_{sub} / kJ mol^{-1}	Lattice energies / kJ mol^{-1}				At experimental structure	
		Error (ΔH_{sub}-U_{latt}) in calculated energy at relaxed structure					
		EST+0.9DMA	EST+DMA	EST+Q	FIT+0.9DMA	0.9DMA	EST+0.9DMA
GEBTUC	-158.1	-24.7	-1.1	31.6	-18.6	-110.5	-129.7
URACIL	-131.0	-23.9	-6.0	33.4	-19.7	-85.6	-102.7
AZURAC10	-141.0	-44.5	-29.3	21.4	-40.2	-70.5	-93.0
CYTSIN01	-155.0	-23.2	1.5	15.3	-15.8	-105.7	-128.5
CYURAC10	-133.0	-21.8	-0.5	92.3	-14.5	-103.3	-105.1
PYRZOL02	-74.0	-11.0	0.1	-26.8	-7.8	-46.4	-61.5
IMAZOL13	-83.0	-14.1	-2.5	-13.9	-11.1	-51.4	-67.2
INDAZL	-91.1	-14.1	-6.2	-23.4	-12.0	-26.3	-73.6
MELAMI04	-123.3	7.6	36.0	-10.7	21.8	-97.2	-126.7
PRMDIN	-49.9	7.3	12.1	10.3		-20.9	-56.0
PYRAZI	-56.3	-2.8	1.2	-1.6		-17.0	-52.6
BZAMID01	-101.7	-6.8	5.1	19.2	-2.5	-51.2	-94.1
FORMAM	-71.7	-5.4	9.1	38.9	-1.0	-63.1	-61.8
UREAXX09	-98.6	-20.3	-2.8	26.7	-13.8	-75.1	-77.5
ONITAN	-90.0	-8.9	-3.1	16.7	-7.4	-23.6	-79.3
MNIANL01	-108.3	-13.9	-3.6	-2.3	-11.0	-37.6	-92.6

NANILI21	-107.0	-11.7	-1.6	18.7	-9.0	-41.7	-94.6
TNIOAN	-125.3	-5.5	4.1	105.4	-3.8	-40.6	-118.5
TATNBZ	-168.2	-7.0	11.9	133.0	-0.1	-60.6	-154.0
TNBENZ10	-107.3	0.0	8.4	119.5		-35.2	-105.7
ZZZFYW01	-87.9	12.8	20.6	41.6		-32.3	-98.1
DNBENZ10	-87.0	3.1	8.3	62.6		-21.5	-88.5
DNITBZ11	-96.2	4.0	11.8	68.1		-32.3	-97.5
BNZQUI	-62.8	11.2	18.3	58.4		-30.7	-71.6
NAPHTD	-72.5	3.1	5.1	-3.3		-8.4	-74.7
ANTCEN14	-97.8	6.7	9.6	-3.0		-11.8	-102.9
PHENAN08	-90.9	10.2	12.7	1.0		-10.0	-99.0
BENZEN02	-44.4	5.0	7.4	-1.8		-6.6	-47.7

Footnotes for Table 3

Experimental heats of sublimation (ΔH_{sub}) taken from reference[26], except PRMDIN[43]. Where more than one experimental determination has been made, the most recent or largest result was usually used, since experimental heats of sublimation are more easily underestimated[27]. The lattice energy was evaluated by direct summation to 20 Å for all terms except the charge-charge, charge-dipole and dipole-dipole electrostatic contribution, which were evaluated by Ewald type summation.

Table 4a Summary of Fits Using Gradient Fitting Method for Repulsion-Dispersion Potential Together With Full DMAs

[continues over 2 pages]

	EST	2 [a]	3 [b]	PHB [a]	PHBb [c]	PHBc [d]	PHBd [e]
C....C	369743.0	277497.1	277489.0	363178.0	285015.4	364316.0	363110.0
Hc...Hc	11971.0	7824.9	7823.8	10082.0	5316.3	10111.6	10092.3
N...N	254529.1	486394.8	486434.8	432455.0	511976.0	431459.6	432250.8
O...O	230064.1	292864.5	292843.7	256650.2	279824.5	256475.2	256774.3
Hp...Hp	7017.3	2268.3	2268.4	2873.6	2942.8	2875.2	2875.2
C...Hc	66529.7	80608.7	80600.4	60510.7	38925.9	60694.4	60536.0
C...N	306774.1	367386.9	367396.7	396305.6	381996.1	396469.0	396174.9
C...O	291658.4	285077.3	285063.0	305302.6	282408.0	305676.4	305347.9
C...Hp	50937.3	25088.5	25089.2	32305.2	28961.0	32365.0	32311.4
Hc...N	55199.4	61692.7	61691.1	66030.2	52171.0	66051.0	66048.5
Hc...O	52479.6	47871.0	47866.1	50867.8	38569.8	50925.1	50906.2
Hc...Hp	9165.4	4212.9	4212.8	5382.5	3955.3	5392.0	5386.8
N...O	241987.6	377422.5	377424.7	333151.1	378501.6	332654.0	333153.0
N...Hp	42262.4	33215.5	33218.3	35252.0	38815.4	35221.4	35253.7
O...Hp	40180.0	25773.9	25774.0	27157.1	28696.0	27155.6	27171.5

Footnotes on next page

Footnotes for Table 4a

Emboldened text indicates that this parameter was included in fitting. Other parameters were obtained by combining rules. Weighting scheme:

[a]	Force: 1; Torque: 1; Strain: 1; Energy: 1
[b]	Force: 1; Torque: 1; Strain: 1; Energy :0.05
[c]	Force: 1; Torque: 1; Strain: 0.05; Energy: 1
[d]	Force: 1; Torque: 0.05; Strain: 1; Energy: 1
[e]	Force: 0.05; Torque: 1; Strain: 1; Energy: 1

Table 4b Summary of Fits Using Gradient Fitting Method for Repulsion-Dispersion Potential Together With 0.9DMA

[continues over 2 pages]

	EST	5[a]	6[b]	7[b]	8[b]	10[b]
C....C	369743.0	379096.1	433138.3	393321.0	304637.3	358990.1
Hc...Hc	11971.0	9392.5	8062.8	9918.0	14434.4	11236.2
N...N	254529.1	394540.5	335162.4	274299.4		
NO...NO	254529.1			834804.4		
O...O	230064.1	246230.1	258638.7	200711.1		
Hp...Hp	7017.3	2066.1				
C...Hc	66529.7	59671.2	59095.9	62457.8	66311.8	63511.3
C...N	306774.1	386741.2	381014.0	328462.7		642462.1
C...NO	306774.1			573015.0		
C...O	291658.4	305524.0	334703.3	280969.6		321350.1
C...Hp	50937.3	27986.7				
Hc...N	55199.4	60874.6	51984.2	52158.5		37329.1
Hc...NO	55199.4			90992.4		
Hc...O	52479.6	48090.7	45665.7	44616.8		44625.2
Hc...Hp	9165.4	4405.2				
N...NO	254529.1			478525.2		
N...O	241987.6	311685.3	294424.8	234637.9		427667.2
N...Hp	42262.4	28551.1				
NO...O	241987.6			409334.3		
O...Hp	40180.0	22555.3				

Footnotes on next page

83

Footnotes to Table 4b

Emboldened text indicates that this parameter was included in fitting. Other parameters were obtained by combining rules. Weighting scheme :

[a]   Force: 1; Torque: 1;   Strain: 10; Energy: 1

[b]   Force: 1; Torque: 1;   Strain: 1; Energy : 1

Table 4c Summary of Fits Using First Step Fitting Method for Repulsion-Dispersion Potential Together With 0.9DMA

Potential	EST	1[a]	2[a]	3[b]	4[b]	5[b]	6[b]	7[b]	8[b]	9[b]	10[b]	11[b]	12[b]
C....C	369743.0	481953.8		476639.7	450912.9	446200.3	498083.2						
Hc...Hc	11971.0	5949.1		5890.8	6818.4	8591.9	9300.0						
N...N	254529.1	445293.2		415516.1	346137.6	604137.8	565638.3	494523.5	225257.7		333392.9		
O...O	230064.1	182473.6		181150.4	167563.9	366131.5	383598.2	404148.0		143659.3	94525.8	149697.1	
Hp...Hp	7017.3	2591.9	2558.3	3580.4	3284.2	0.0	0.0	0.0			3239.3		3332.5
C...Hc	66529.7	53546.3		52988.4	55448.1	64493.3	62135.0						
C...N	306774.1	463261.0		445029.8	395067.0	269677.8	148723.8	230375.5	288595.6		257544.2		
C...O	291658.4	296553.3		293842.6	274875.8	208247.7	224979.1	220486.6		230471.3	186949.9	235264.7	
C...Hp	50937.3	35343.4	30755.6	41310.8	38482.5	38586.3	48171.0	41268.8			34607.8		35102.2
Hc...N	55199.4	51469.5		49474.3	48580.8	44209.6	51581.1	50825.1	51928.5		63174.8		
Hc...O	52479.6	32947.9		32666.7	33801.1	31596.4	30973.7	30271.4		41469.9	33638.8	42332.4	
Hc...Hp	9165.4	3926.8	5534.0	4592.6	4732.1	1.3	136.3	13941.9			6227.2		6316.1
N...O	241987.6	285051.3		274355.5	240832.2	247261.0	243198.4	235250.6	227648.2	191221.0	425105.1	195198.0	
N...Hp	42262.4	33972.6	25517.8	38571.1	33716.4	19353.5	22274.1	25406.1	39758.1	31750.6	32862.6		29124.1
O...Hp	40180.0	21747.3	24260.5	25467.6	23458.9	22840.3	23641.6	24293.5			17498.4		27689.1

Footnotes on next page

85

Footnotes for Table 4c

Emboldened text indicates that this parameter was included in fitting. Other parameters were obtained by combining rules. In addition to the crystal structures in Table 5a, s-tetrazine was also included in the fitting dataset for this series of fits.

Weighting scheme : Fits 1, 3, 4 and 5 have incorrect weighting for s-tetrazine, which makes a negligible difference to the sum of squares and hence fitted parameters.

[a]   Force: 10; Torque: 0.5; Strain: 10; Energy: 0.01
[b]   Force: 10; Torque: 0.5; Strain: 10; Energy :0.1

Table 5a Comparison of Initial and Final Sums of Squares for Fit PHB

Molecule	Initial	Final
o-dinitrobenzene	76292.64	52351.52
m-dinitrobenzene	32621.94	17432.64
p-dinitrobenzene	31038.68	15333.07
Naphthalene	8703.541	12045.18
Anthracene	34954.76	38536.13
Phenanthrene	23345.66	38253.32
Allopurinol	73762.76	172545.8
Cytosine	114605.2	52205.71
Urea	1936.961	6798.097
Formamide	20562.2	41965.97
Aniline	26823.71	87474.01
Pyrimidine	17903.45	32519.95
Imidazole	10958.9	11327.62
BICWEP	455397.9	139602.3
o-benzoquinone	57249.04	38191.68
pyrazine	4480.553	14303.25
uric Acid	1249201	134777.5
TATNBZ	283465.9	131955.3
Uracil	345167.3	54870.92
Total	2868473	1092490

87

Table 5b Comparison of Initial and Final Sums of Squares for Fit JPC

	Initial	Final
6-Azauracil	6.1	2.6
2,6-Dinitroaniline	4.5	4.6
DAMTRZ20	2.8	4.6
FITXIP	6.0	6.7
Indazole	3.8	3.2
JOWWIB	0.4	0.8
m-Nitroaniline	4.1	4.2
o-Nitroaniline	4.9	6.0
p-Nitroaniline	2.3	2.2
Purine	12.0	9.2
Pyrazole	13.8	9.5
TNIOAN	0.7	0.5
Terephthalamide	0.5	0.6
Total	61.9	55.0

Table 6 Comparison of Starting Point and Fitted A Parameters for Fitting Methods

Interaction	EST	Gradient Fit	First Step Fit
C...C	369743.0	379096.1 (2.5)	481953.8 (30.3)
Hc...Hc	11971.0	9392.5 (-21.5)	5949.1 (-50.3)
N...N	254529.0	394540.5 (55.0)	445293.2 (74.9)
O...O	230064.1	246230.1 (7.0)	182473.6 (-25.9)
Hp...Hp	7017.3	2066.1 (-70.6)	2591.9 (-63.1)

All values in kJ mol^{-1}. Percentage differences between fitted and EST parameters are shown in parentheses.

Table 7 Comparison of Predicted Crystal Structures and Lattice Energies Using EST+DMA (first line) and PHB+DMA (second line)

Crystal REFCODE	Differences Between Predicted and Experimental Structures[a] [continues over 3 pages]					Energies[b] / kJ mol⁻¹		
	$\Delta a(\%)$	$\Delta b(\%)$	$\Delta c(\%)$	$\Delta\beta(\%)$[c]	$\delta(\%)$[d]	ΔH_s	U_r	ΔH_s-U_r
ZZZFYW01	-0.344	-0.372	-5.036	-1.753	2.922	-87.9	-108.5	20.6
	-0.243	0.086	-4.628	-2.029	2.676	-87.9	-106.2	18.3
DNBENZ10	-1.388	-1.432	-0.366		1.171	-87.0	-95.3	8.3
	-1.314	-1.197	0.682		1.099	-87.0	-92.7	5.7
DNITBZ11	-2.502	-2.523	2.084	-0.030	2.378	-96.2	-108.0	11.8
	-1.552	0.157	0.380	0.442	0.927	-96.2	-103.4	7.2
NAPHTD	1.611	-2.605	-0.135	-0.764	1.770	-72.5	-77.6	5.1
	0.722	-3.339	-0.894	-0.671	2.039	-72.5	-82.1	9.6
ANTCEN14	1.504	-3.038	-0.445	-1.262	1.974	-97.8	-107.4	9.6
	0.561	-3.770	-1.045	-1.235	2.282	-97.8	-113.2	15.4
PHENAN08	2.817	-2.499	-2.575	-3.357	2.634	-92.5 (-90.9)	-103.6	11.1 (12.7)
	1.558	-2.912	-3.069	-3.239	2.603	-92.5 (-90.9)	-108.8	16.3 (17.9)
ALOPUR	0.393	-0.888	-0.240	-1.830	0.578	----	-148.3	----
	2.067	-0.651	0.900	-0.433	1.355	----	-141.8	----
CYTSIN01	-2.666	-0.654	2.691		2.219	-176.0 (-155.0)	-156.5	-19.5 (1.5)
	-4.910	-0.151	6.328		4.625	-176.0 (-155.0)	-156.3	-19.7 (1.3)
UREAXX09	-2.128	-2.128	0.772		1.601	-98.6	-95.8	-2.8
	-3.194	-3.194	0.296		2.268	-98.6	-99.8	1.2
FORMAM	3.223	-2.804	-3.404	12.328	3.154	-71.7	-80.8	9.1
	2.065	-4.821	-2.203	5.427	3.284	-71.7	-84.8	13.1

Table 7 (continued)

Crystal REFCODE	Differences Between Predicted and Experimental Structures[a]					Energies[b] / kJ mol⁻¹		
	Δa(%)	Δb(%)	Δc(%)	Δβ(%)[c]	δ(%)[d]	ΔH[ₐ]	Uᵣ	ΔHₐ-Uᵣ
BAZGOY	-2.251	6.694	-3.239	0.673	4.486	----	-73.1	----
	-2.766	5.625	-3.324	0.946	4.096	----	-74.5	----
PRMDIN	-1.417	1.386	-3.876		2.513	-49.0	-62.0	13.0
	-0.502	2.118	-2.942		2.113	-49.0	-58.7	9.7
IMAZOL13	2.251	-3.675	0.648	2.585	2.516	-75.4 (-83.0)	-80.5	5.1 (-2.5)
	4.871	-5.497	0.770	2.103	4.264	-75.4	-79.1	3.7
BICWEP	-0.840	-1.620	-0.415	-0.856	1.080	----	-186.2	----
	-0.689	-0.274	1.162	-0.625	0.796	----	-174.6	----
OBNZQU	0.985	-1.286	1.978		1.476	----	-88.9	----
	1.306	-1.996	1.545		1.641	----	-90.2	----
PYRAZI	2.080	-1.063	-4.319		2.835	-56.3	-57.5	1.2
	3.695	0.508	-4.743		3.484	-56.3	-54.0	-2.3
URICAC	1.991	0.841	-0.546	-1.684	1.287	----	-192.1	----
	0.450	8.871	-4.983	-2.818	5.880	----	-198.3	----
TATNBZ	-1.209	-1.396	-8.973	-14.433,-6.269,0.074	5.289	(-168.2)	-180.1	(11.9)
	-0.884	-1.148	2.100	2.906,-6.942,0.070	1.473	(-168.2)	-162.1	(-6.1)
URACIL	0.968	1.534	1.501	2.598	1.359	-131.0	-125.0	-6.0
	-0.249	0.693	3.051	1.962	1.812	-131.0	-128.2	-2.8

Footnotes on next page

Footnotes for Table 7

[a] Percentage errors in lattice vectors and cell angles are defined thus: (100*(predicted value-experimental value)/experimental value).

[b] Lattice energies are calculated using a 20Å cutoff, except the charge-charge, charge-dipole and dipole-dipole terms which were evaluated by Ewald summation. ΔH_s is the experimental heat of sublimation[26] (for PRMDIN from Filippini and Gavezzotti[2]). U_r is the predicted lattice energy at the relaxed structure. Values in parentheses were obtained from an updated list of heats of sublimation, but were not used in fitting. [c] For triclinic systems, errors in cell angles are given in the order (α, β, γ). No entry implies a tetragonal or orthorhombic crystal system. [d] δ is the root mean square percentage error in the independent lattice vectors.

Table 8 Effect of Hydrogen Atom Positions on Calculated Crystal Structures

Crystal REFCODE	Errors in EST+0.9 DMA for altered H atom positions								δ for X-H direction from experimental structure (%)	ΔU_{latt}[a]/kJ mol⁻¹
	$\Delta a/Å$	$\Delta b/Å$	$\Delta c/Å$	$\Delta\beta/°$	$\Delta\theta/°$	$\Delta x/Å$	F	δ (%)		
Use of idealised sp^2 NH₂ group.										
DAMTRZ20	0.014	0.720	-0.373	0.00, 0.00, 0.15	10.057	0.331	154.15	6.3	4.4	-9.9
MELAMI04	0.077	-0.277	0.018	-0.44	5.354	0.109	22.79	2.2	2.3	-17.1
JOWWIB	0.620	-0.282	0.070	2.01	5.987	0.306	75.68	4.2	3.5	-2.3
ONITAN	-0.29	0.46	-0.28	-0.4	2.789	0.313	47.33	3.4	4.2	0.3
MNIANL01	0.002	0.008	-0.125		3.655	0.052	9.65	1.4	5.1	-5.0
NANILI21	0.041	0.092	-0.234	0.20	2.508	0.082	12.16	1.8	1.9	-0.6
TNIOAN	0.320	-0.706	0.330	3.23	5.800	0.449	129.39	5.5	5.5	1.1
BAZGOY	0.170	-0.252	0.207	-0.12	3.777	0.202	32.80	2.9	3.1	-4.7
Use of foreshortened C-H bond[b].										
BENZEN02	-0.47	-0.12	0.03		18.972	0.172	134.76	3.7	15.0	1.8
INDAZL	-0.376	0.493	0.942	8.2	11.193	0.611	384.39	9.1	9.3	2.5

[a] $\Delta U_{latt}= U_{latt}$ (X-H direction from experimental structure) - U_{latt} (new H positions)

[b] The interaction site was placed 1.01 Å along the C-H bond, a distance of 0.07 Å from the H nuclear position used in determining the wavefunction.
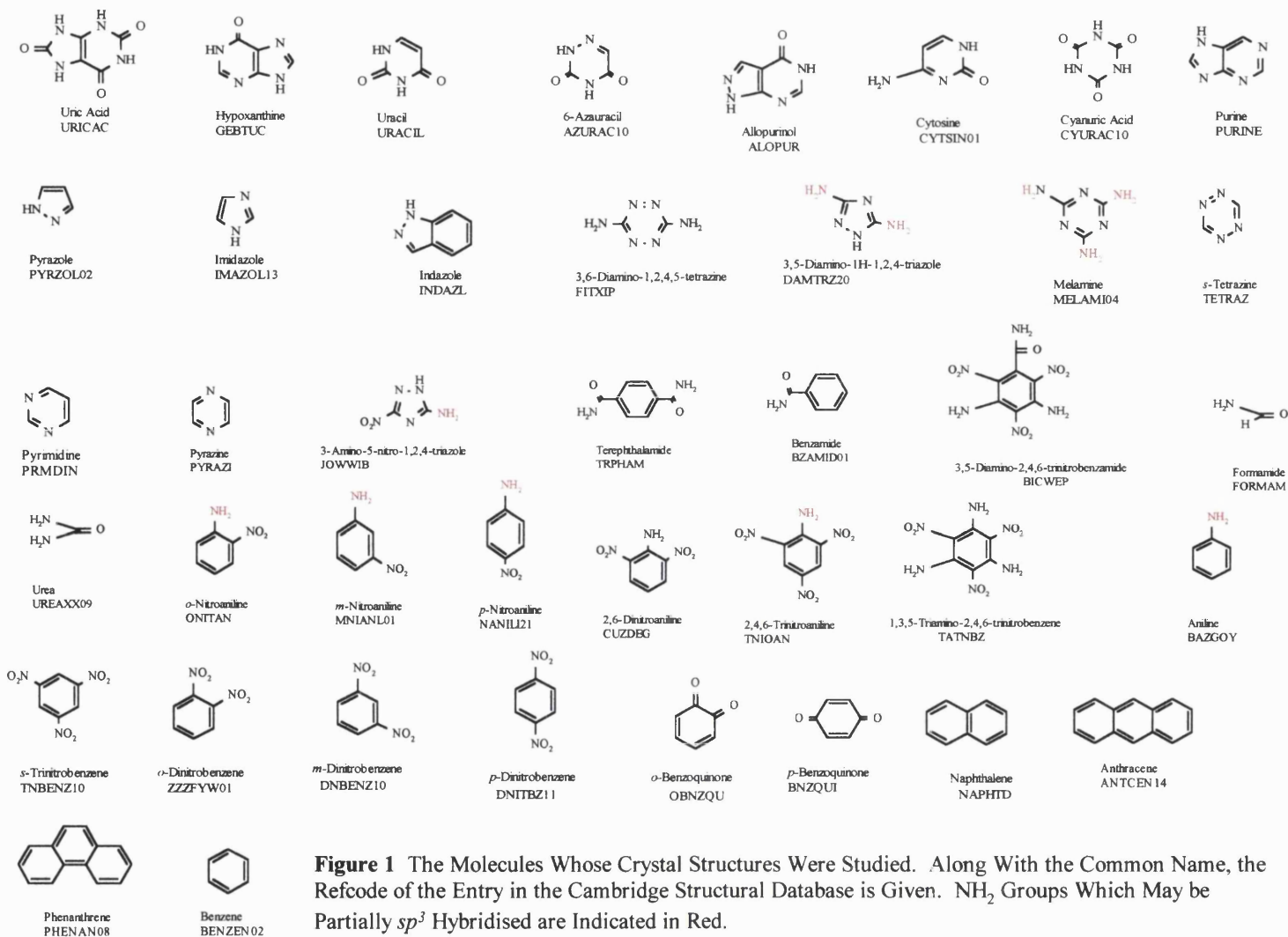
Figure 1 The Molecules Whose Crystal Structures Were Studied. Along With the Common Name, the Refcode of the Entry in the Cambridge Structural Database is Given. NH$_2$ Groups Which May be Partially sp^3 Hybridised are Indicated in Red.
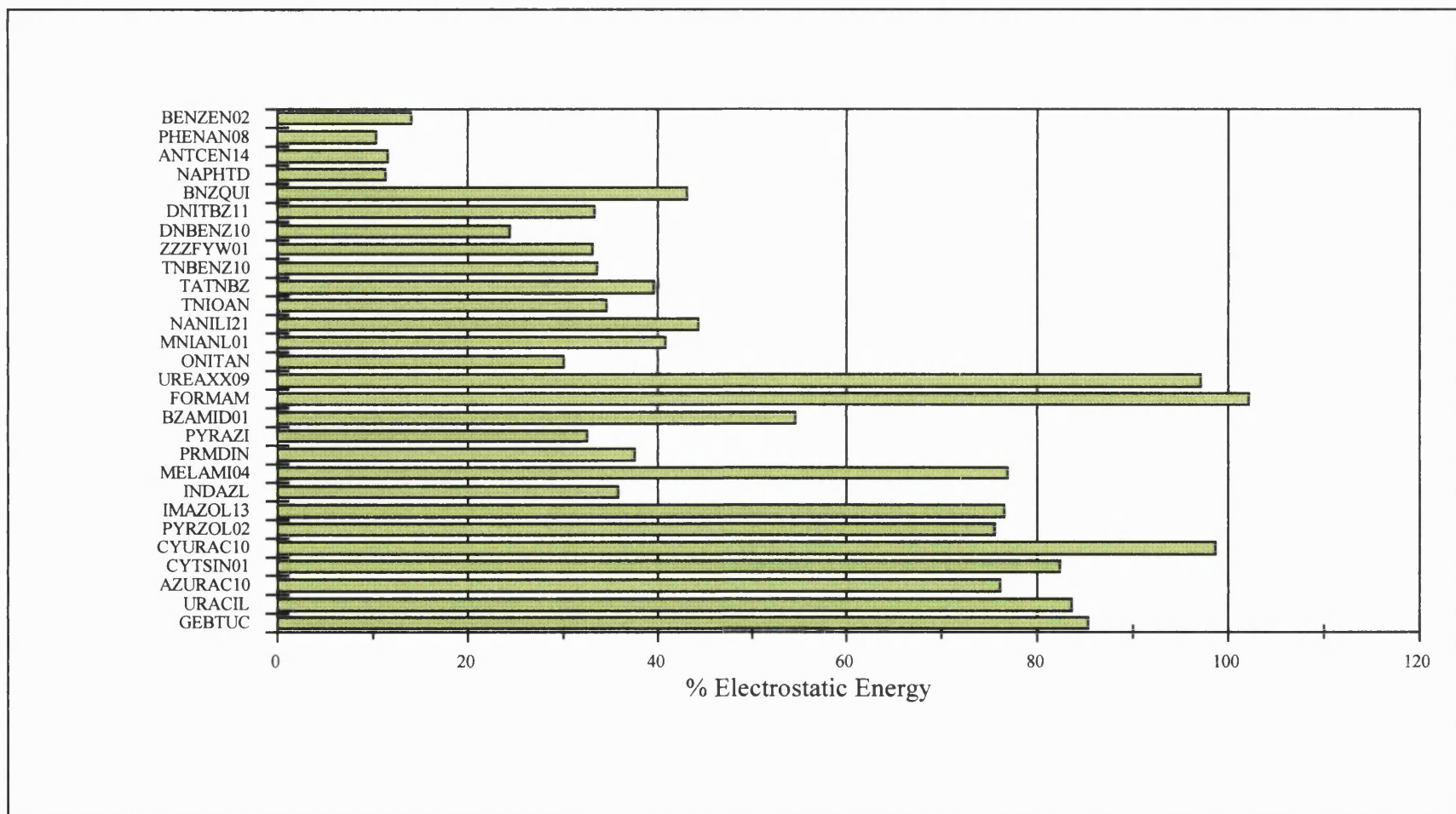
Figure 2 Graph to Show Contribution of Electrostatic Energy as a Percentage of the Total Lattice Energy at the Experimental Structure Using Potential EST+0.9DMA. The Electrostatic Contribution to the Total Lattice Energy at the Experimental Structure for FORMAM is Greater than 100% as the Repulsion and Dispersion Forces are of Approximately Equal Magnitude and Opposite Sign.
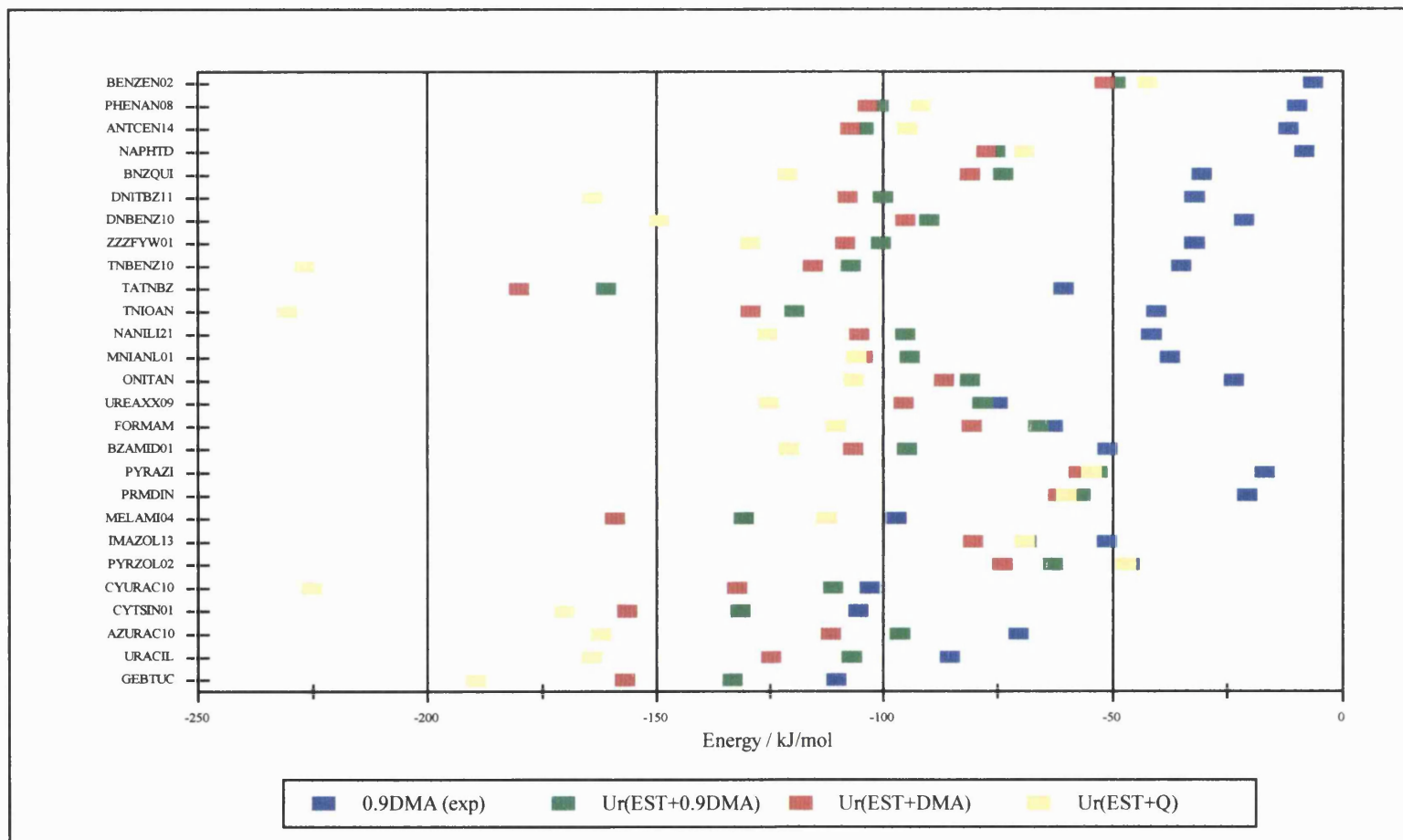
Figure 3 Calculated Lattice Energies at the Relaxed Structure Using the Three Electrostatic Models Together With the Repulsion-Dispersion Potential EST. The Electrostatic Contribution at the Experimental Structure for the Most Accurate Model, 0.9DMA, is also Shown. All Values are in kJ mol[-1].
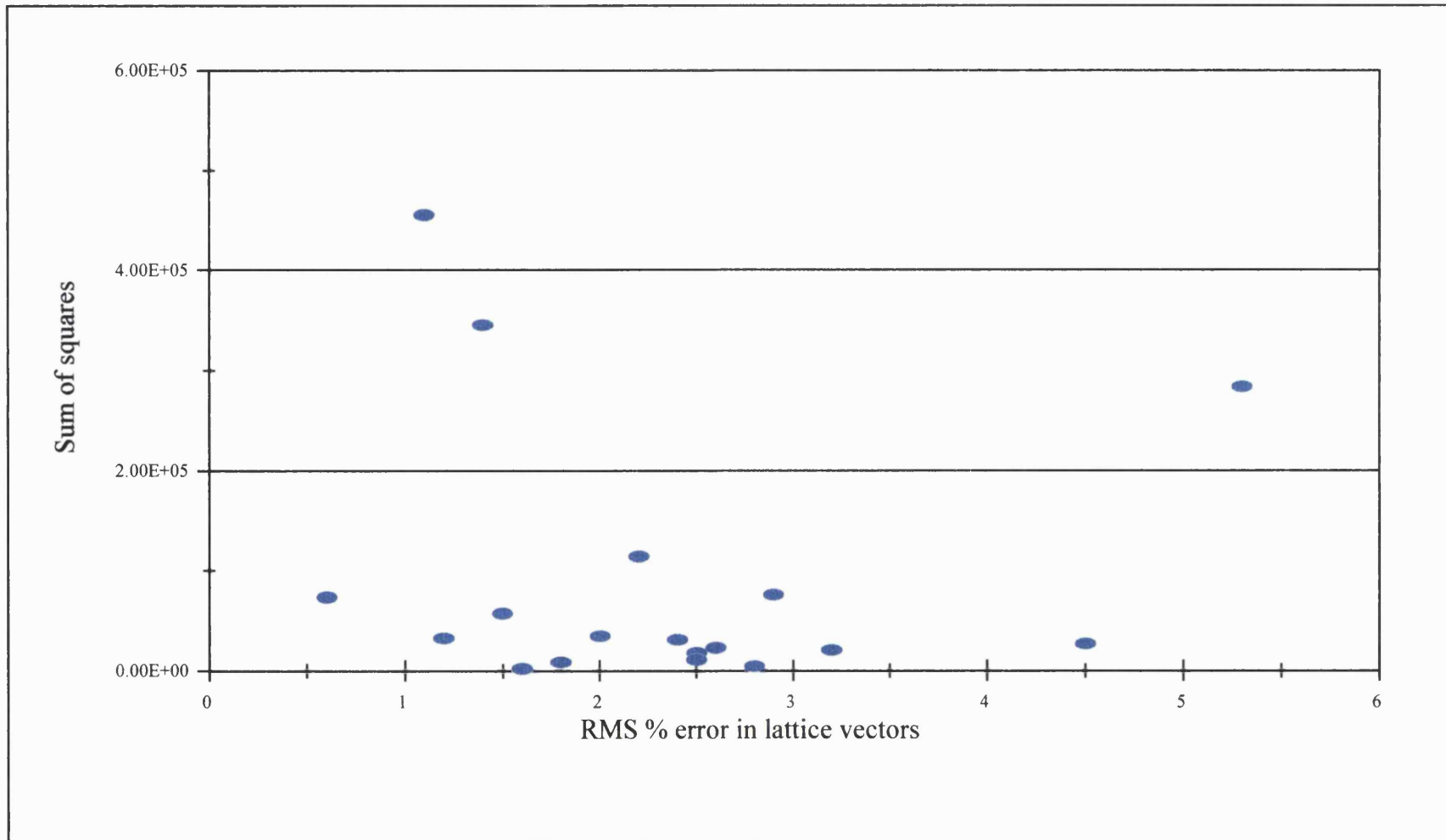
Figure 4a Graph Showing Variation of Sum of Squares With Root Mean Square Percentage Error in Independent Lattice Vectors for Gradient Fit PHB.
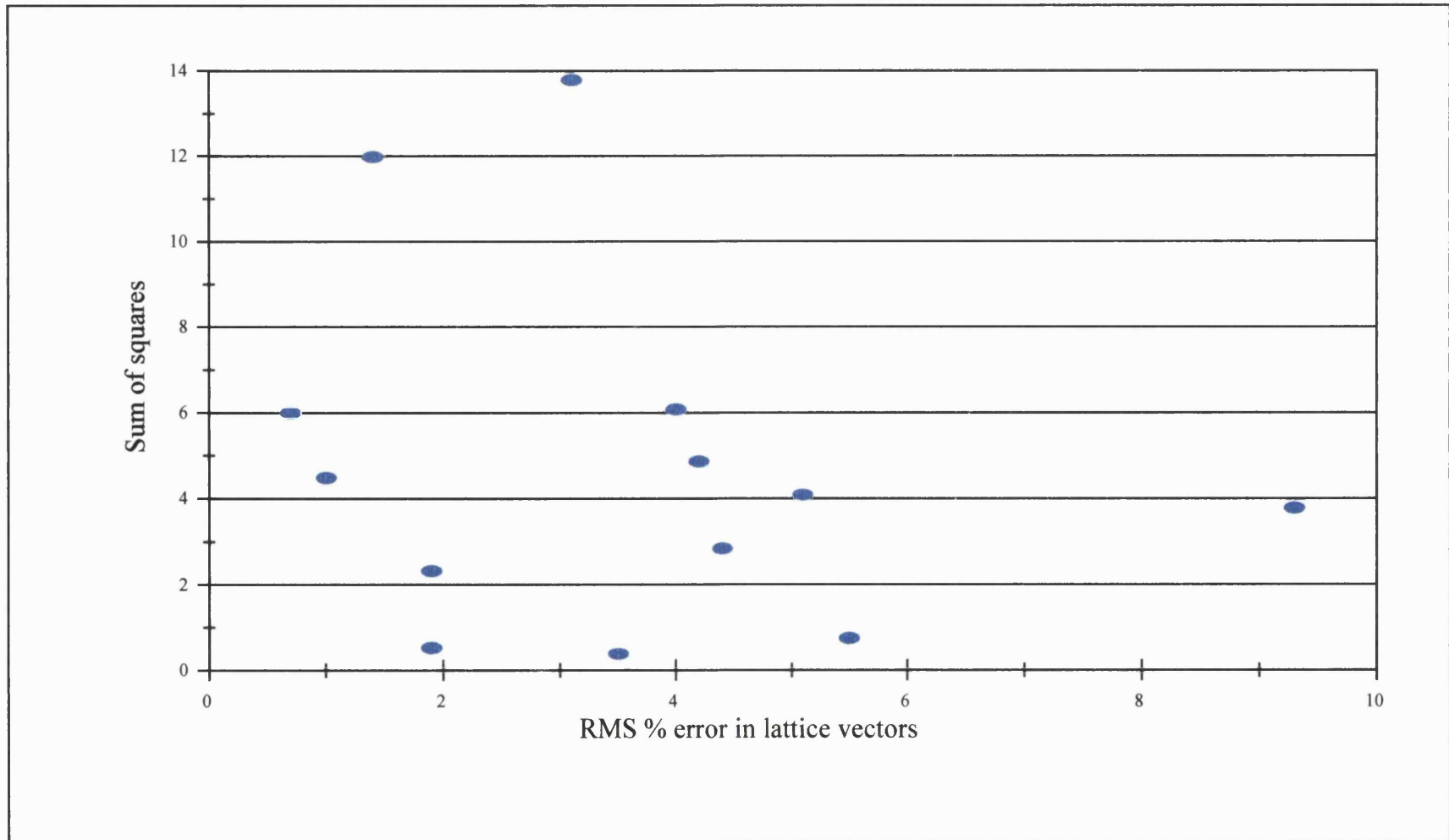
Figure 4b Graph Showing Variation of Sum of Squares With Root Mean Square Percentage Error in Independent Lattice Vectors for First Step Fit JPC.

Figure 5 Comparison of Fitted O⋯H_p Hydrogen Bonding Potentials from Fitting to IMPT Calculations of the Formamide/Formaldehyde Hydrogen-bonding Interaction and this Work.

References and Notes for Chapter 4

1. Pertsin, A.J.; Kitaigorodsky, A.I. *The Atom-Atom Potential Method*, Springer-Verlag, Berlin, 1987.

2. Filippini, G.; Gavezzotti, A. *Acta Cryst.* **1993**, *B49*, 868.

3. Gavezzotti, A.; Filippini, G. *J. Phys. Chem.* **1994**, *98*, 4831.

4. Desiraju, G.R. *Crystal Engineering: The Design of Organic Solids*, Elsevier, Amsterdam, 1989, p56.
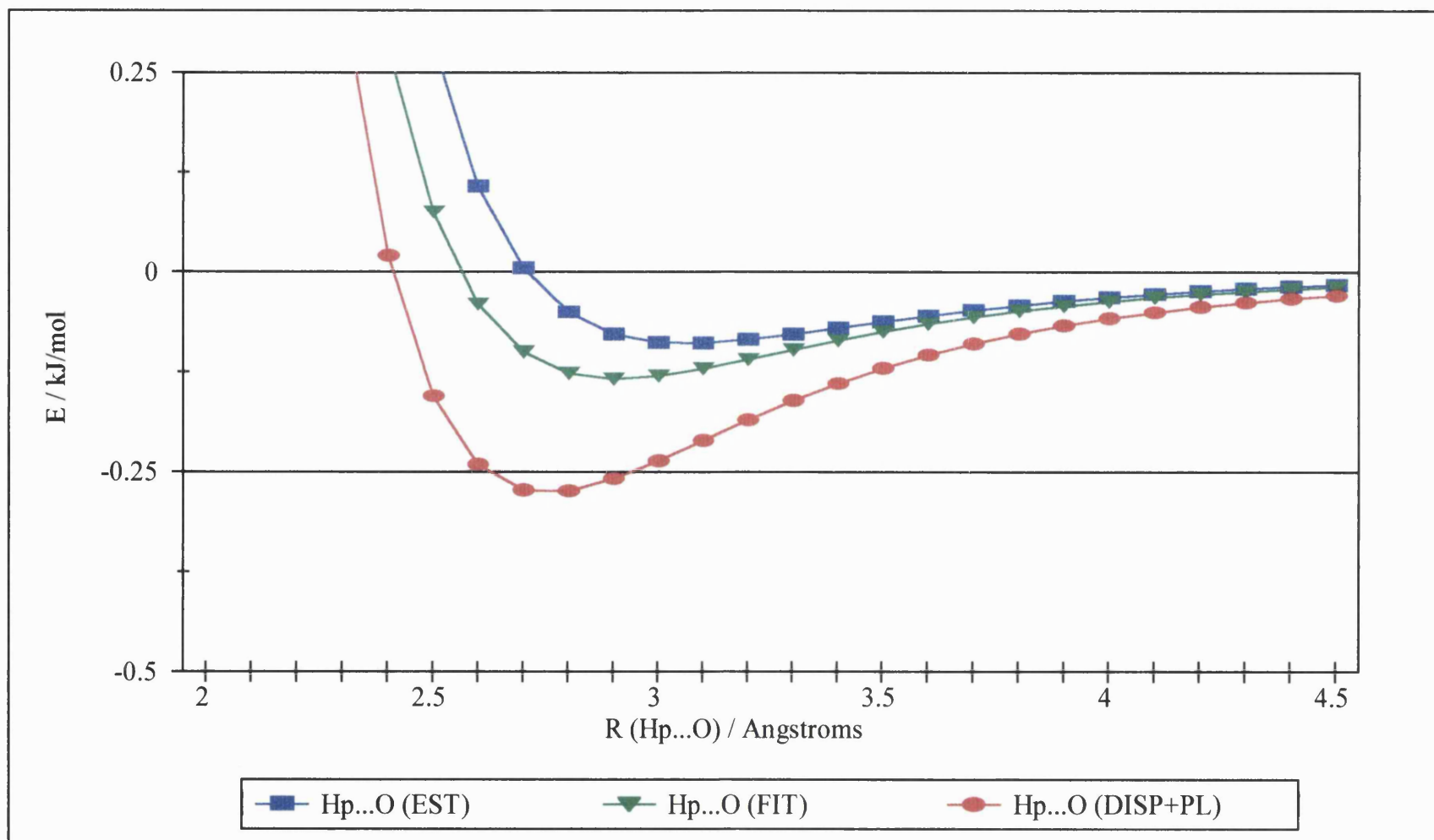
5. Wiberg, K.B.; Rablen, P.R. *J Comput. Chem.* **1993**, *14*, 1504.

6. Williams, D.E.; Weller, R.R. *J. Am. Chem. Soc.* **1983**, *105*, 4143.

7. Buckingham, A.D.; Fowler, P.W. *Can. J. Chem.* **1985**, *63*, 2018.

8. Price, S.L.; Stone, A.J. *J. Chem. Phys.* **1987**, *86*, 2859.

9. Mitchell, J.B.O.; Nandi, C.L.; Thornton, J.M.; Price, S.L.; Singh, J.; Snarey, M. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 2619.

10. Price, S.L.; Lo Celso, F.; Treichel, J.A.; Goodfellow, J.M.; Umrania, Y. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 3407.

11. Hurst, G.J.B.; Fowler, P.W.; Stone, A.J.; Buckingham, A.D. *Int. J. Quantum Chem.* **1986**, *29*, 1223.

12. Berkovitch-Yellin, Z.; Leiserowitz, L. *J. Am. Chem. Soc.* **1980**, *102*, 7677.

13. Berkovitch-Yellin, Z.; Leiserowitz, L. *J. Am. Chem. Soc.* **1982**, *104*, 4052.

14. Allen, F.H.; Kennard, O. *Chem. Design Autom. News* **1993**, *8*, 31.

15. Allen, F.H.; Kennard, O.; Watson, D.G.; Brammer, L.; Orpen, A.G.; Taylor, R. *J. Chem. Soc., Perkin Trans. II* **1987**, S1.

16. Hariharan, P.C.; Pople, J.A. *Theor. Chim. Acta* **1973**, *28*, 213.

17. CADPAC5: *The Cambridge Analytical Derivatives Package Issue 5*, A suite of quantum chemistry programs developed by R.D. Amos with contributions from I.L. Alberts, J.S. Andrews, S.M. Colwell, N.C. Handy, D. Jayatilaka, P.J. Knowles, R. Kobayashi, N. Koga, K.E. Laidig, P.E. Maslen, C.W. Murray, J.E. Rice, J. Sanz, E.D. Simandiras, A.J. Stone and M.-D. Su, Cambridge, U.K., 1992.

18. Price, S.L.; Stone, A.J.; Alderton, M. *Mol. Phys.* **1984**, *52*, 987.

19. W. Smith, *The Program MDMULP*, CCP5 Program Library, Daresbury Laboratory, 1982.

20. Price, S.L. *Phil. Mag.* **1996**, *B73*, 95.

21. Williams, D.E.; Cox, S.R. *Acta Cryst.* **1984**, *B40*, 404.

22. Cox, S.R.; Hsu, L.-Y., Williams, D.E. *Acta Cryst.* **1981**, *A37*, 293.

23. Willock, D.J.; Leslie, M.; Price, S.L.; Catlow, C.R.A. *Mol. Cryst. Liq. Cryst. Sci. Technol. Sect. A.* **1993**, *235*, 217.

24. Mitchell, J.B.O.; Price, S.L. *J. Comput. Chem.* **1990**, *11*, 1217.

25. Stone, A.J.; Tong, C.-S. *J. Comput. Chem.* **1994**, *15*, 1377.

26. Chickos, J.S. in Molecular Structure and Energetics, Vol 2, edited by J.F. Liebman and A. Greenberg, VCH, New York, 1987, p67; and personal communication of updated tables.

27. Bondi, A. *J. Chem. Eng. Data* **1963**, *8*, 371.

28. Williams, D.E.; Starr, T.L. *Comput. Chem.* **1977**, *1*, 173.

29. To illustrate this problem, a revision of the sublimation energy of cytosine by 21 kJ mol^{-1} came to our attention during the course of the work, bringing the experiment into agreement with our predictions. A lattice energy for TRPHAM is also not included in Table 3 as this seems unlikely to refer to complete separation of the molecules.

30. Coombes, D.S., *Phil. Mag.* **1996**, *B73*, 117.

31. Coombes, D.S.; Price, S.L.; Willock, D.J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.

32. Price, S.L.; Stone, A.J. *Mol. Phys.* **1984**, *51*, 569.

33. Ritchie, J.P.; Kober, E.M.; Copenhaver, A.S. *Proceedings of the 10th International Symposium on Detonation*, Office of Naval Research, 1993.

34. Cox, S.R.; Williams, D.E. *J. Comput. Chem.* **1981**, *2*, 304.

35. Spackman, M.A. *Chem. Rev.* **1992**, *92*, 1769.

36. Hsu, L.-Y.; Williams, D.E. *Acta Cryst.* **1980**, *A36*, 277.

37. Warshel, A.; Lifson, S. *J. Chem. Phys.* **1970**, *53*, 582.

38. S.L. Price, in "Computer Modelling in Inorganic Crystallography", Ed. C.R.A. Catlow, Academic Press, 1997.

39. Gavezzotti, A. *J. Am. Chem. Soc.* **1989**, *111*, 1835.

40. Gavezzotti, A. *J. Phys. Chem.* **1991**, *95*, 8948.

41. Hagler, A.T.; Lifson, S.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5122.

42. Filippini, G.; Gavezzotti, A. *Chem. Phys. Lett.* **1994**, *231*, 86.

43. Nabavian, M.; Sabbah, R.; Chestel. R.; Laffite, M. *J. Chim. Phys. Phys.-Chim. Biol.* **1977**, *74*, 115.

Chapter 5   Prediction of Molecular Crystal Structures

In this chapter we discuss the phenomenon of polymorphism. The methods available to predict possible crystal structures are described. Although these methods represent considerable progress towards the goal of crystal structure prediction, it is shown that further work is needed before genuine crystal structure prediction becomes possible. In chapter 6 we apply a crystal structure prediction technique to account for the lack of hydrogen bonds in the observed crystal structure of alloxan.

5.1 Introduction

Polymorphism is purely a solid state effect. The usual definition of a polymorph is that due to McCrone[1]: "A polymorph is a solid crystalline phase of a given compound resulting from the possibility of at least two crystalline arrangements of the molecules of that compound in the solid state". The phenomenon of polymorphism has been known for a few centuries. Although Mitscherlich is usually credited with introducing the term polymorphism during his work on isomorphous sulphates of cobalt, iron, nickel, magnesium, copper, zinc and manganese, the idea is even older than that. Klaproth in 1788 observed that calcium carbonate crystallises as calcite and aragonite and Humphrey Davy (1809) noted that diamond and graphite are both forms of carbon that only differ in the arrangement of carbon atoms in the solid.

The existence of polymorphs in organic compounds is not uncommon. Investigations of the thermodynamics and infrared spectroscopy of organic compounds has suggested that around one third of organic substances exhibit polymorphism under normal pressure conditions[2,3,4]. Gavezzotti and Filippini[5] performed lattice energy calculations on a few hundred pairs of room temperature organic polymorphs and found that polymorphism is very frequent, provided the correct temperature range is explored. Indeed, it has been suggested that the number of polymorphic forms known for a compound is dependent on the time and money spent in research on it. Hence, important compounds such as soaps, paraffin waxes and drugs often have many well characterised polymorphs. Although some polymorphs of inorganic compounds have different names (such as rutile, brookite and anatase (TiO_2)) this is much less common for organic compounds where the different polymorphs are usually designated with Greek letters or Roman numerals.

The existence of different crystalline forms can result from different conformations of the molecule in the crystal or from the existence of other solvates within the crystal structure, as well as the temperature and pressure at which the crystal is formed.

5.2 The Importance of Polymorphism

The study and prediction of polymorphism is important in a number of areas of research, notably pharmaceutics, non linear optics and explosives, as different polymorphs have different physical properties.

The area in which the study of polymorphism has received most interest is in the pharmaceutical industry due to the need to produce drugs which have the required ease of tableting and bioavailability resulting from the solubility and rate of dissolution. The solubility is dependent on the polymorph (as well as the crystal morphology) since different polymorphic forms have different energies. However, only a few products show marked bioavailability differences resulting from polymorphism. An emerging area of interest and one which has a large financial incentive, is in prolonging the life of patents of drugs such as ranitidine[6] (Zantac).

The study and prediction of polymorphs is also important in the area of non-linear optics and the design of electronic materials. Only compounds which crystallise in a non centrosymmetric space group will exhibit the non linear optical property of Second Harmonic Generation (S.H.G.), which is used for frequency doubling of electromagnetic radiation, since if a centre of symmetry exists within the crystal the second order susceptibility vanishes.

For an organic compound the molecular dipole moment μ induced by an electric field E is given by

$$\mu = \alpha E + \beta E^2 + \gamma E^3 + ... \tag{1}$$

Here α is the molecular polarisability, β is the first hyperpolarisability and γ is the second hyperpolarisability. For a crystal the polarisation P then may be written

$$P = \varepsilon_0[\chi_1 E + \chi_2 E^2 + \chi_3 E^3 + ...] \tag{2}$$

where ε_0 is a constant and χ_1 is the linear electric susceptibility. The odd terms χ_1, χ_3 etc. in equation (2) are present for all substances but the even coefficients χ_2, χ_4

etc. are non zero only if the crystal is non centrosymmetric. The values of the coefficients χ_2 and χ_3 are important in practical applications.

Early developments in non linear optics relied on inorganic materials such as lithium niobate ($LiNbO_3$) and potassium dihydrogen phosphate (KDP), but it was soon realised that organic materials, such as 3-methyl-4-nitroaniline, have S.H.G. efficiencies far in excess of inorganic materials. As well as the space group, the nature of the functional groups is also important in determining the effectiveness of an S.H.G. material[7]. Thus, designing S.H.G. materials may be divided into two parts (1) optimising the molecular electronic properties (molecular engineering) and (2) achieving crystallisation in a certain space group (crystal engineering). However, these two areas are not mutually exclusive, as a compound which has been designed to have a large β value by including a delocalised π system may well tend to form a centrosymmetric crystal structure due to dipole-dipole stabilisation. Other factors such as thermal stability, ease of synthesis and mechanical strength also combine to make the search for organic non linear optic materials a difficult task.

Crystallisation in certain space groups is also necessary in several other areas of solid state chemistry[8], notably pyroelectricity, piezoelectricity, triboluminescence and ferroelectricity. Pyroelectricity, which is a change in the separation of positive and negative charge on heating, requires the primitive unit cell to possess a dipole moment. Pyroelectric crystals such as *p*-bromobenzoic anhydride can be used as thermal detectors. Piezoelectricity, which is the change in the separation of positive and negative charge on compressing a crystal, is also of importance. Although inorganic materials such as quartz and sphalerite show this effect, it is influenced by covalency in the bonds and so is more widespread in organic crystals such as amines and amino acids. Triboluminescence, which is the emission of light on application of mechanical force, requires that the crystal has a permanent dipole moment and is seen in substances such as sucrose and tartaric acid. Ferroelectricity, which is a change in the direction of spontaneous polarisation on application of an electric field, is seen in substances such as Rochelle salt and sodium potassium tartrate tetrahydrate. All these require non centrosymmetric space groups. The prediction of polymorphs also has importance in the area of explosives as the more densely packed the crystal structure of an explosive material is the more effective it will be. Finally the study of

polymorphism also gives us considerable insight into understanding the importance of intermolecular interactions as well as the relationship between crystal structure and molecular structure.

5.3 Progress Towards Prediction of Molecular Crystal Structures

In 1988, Maddox[9] commented on the scandal of the fact that it was impossible to predict the structure of even simple crystal structures from a knowledge of their chemical composition. This area is still the subject of much debate[10,11] although recently considerable progress has been made in the field. However, before crystal structure prediction techniques can be applied to predicting new structures we need to be able to predict known crystal structures.

To perform genuine crystal structure prediction, two components are required: (a) a method of simulating the crystal structure and a model for the intermolecular forces that determine the molecular packing and (b) a method of generating sufficient hypothetical structures that the most stable structure is likely to be found. A molecular crystal structure is usually simulated by minimising the static lattice energy. This assumes that the simulation is carried out at 0K and that the observed structure is the global minimum in the lattice energy. However, the predicted structure is usually compared with a room temperature one and so the temperature effects (which can be of the order of a few percent error in the lattice vectors[12]) as well as the vibrational and entropic effects are neglected. The model used for the intermolecular potential must be able to find a lattice energy minimum corresponding to the observed structure. In addition there must be sufficient confidence that the model will reliably extrapolate to hypothetical structures.

A number of methods of producing hypothetical crystal structures have been proposed. These can broadly be divided into "dynamic" methods which use molecular dynamics or Monte Carlo methods to generate possible crystal structures and "static" methods where a large number of possible structures are generated and some property of the crystal structures (such as packing density) is used to assess the likelihood of them forming polymorphs.

5.3.1 "Dynamic" Methods

Karfunkel and Gdanitz[13] have developed a method for predicting polymorphs based on simulated annealing. This method involves three stages. Firstly a series of

points are generated in the search space using the Metropolis Monte Carlo algorithm[14]. Given a point A_n (a vector containing the degrees of freedom) in search space where the potential energy is E_n, a trial point $A^*=A_n+\Delta A$ is generated where ΔA is a random vector. ΔE is the energy difference between A^* and A_n. If ΔE is negative then A^* is accepted as the new point A_{n+1} of the sequence, otherwise A^* is accepted as the new point with a probability of $\exp(-\beta\Delta E)$ where β is proportional to the reciprocal "temperature". During the procedure the temperature of the simulation was first increased and then decreased until the structures were "frozen" when the simulation was complete.

Initially it was assumed that the variables that define the search space are the lattice parameters, rotations and translations of the molecules in the asymmetric unit and the main torsion angles for flexible molecules. However, it was found that the crystal structures often remained as a "gas" (i.e. one of the lattice vectors remained very large and so there was no contact between the molecules in certain directions). To avoid this problem the degrees of freedom which determine the spatial extent of the crystal (the lattice parameters and some translations) were made dependent on the angular degrees of freedom (i.e. the three lattice angles, the Euler angles describing the orientation of the molecules and the polar angles of the translation vector). Hence, each Monte Carlo step consisted of choosing a set of angular variables and then optimising the energy with respect to the variables defining the spatial extent of the crystal.

The Monte Carlo method generates a large number of possible crystal structures and so the second stage uses a method of cluster analysis to group similar structures together. In the final stage only the lowest energy crystal structures are subject to full lattice energy minimisation which includes all the degrees of freedom.

This procedure[13] was able to predict the observed structure as the lowest energy structure from a large number of hypothetical ones for a range of structures containing N and O atoms. In one case, however the observed structure was not found and it was suggested that, based on the calculations, the published structure is incorrect. This method has recently[15] been applied to predicting the crystal structures of benzene, urea, terephthalic acid (which is used to produce PET), the female sex hormone estrone (an example of a steroid prototype) and the pigment quinacridone.

Both the stable and high pressure metastable polymorphs of benzene were predicted correctly, although the lattice vectors of the high pressure form were too large as pressure was not taken into account in the simulation. The predicted structure of urea showed a small deviation from the experimental one, though this could be corrected by using Rietveld refinement[16]. Terephthalic acid is crystallised at high temperature and pressure since it is extremely insoluble in water and the two known polymorphs are very similar, both containing a one dimensional hydrogen bonding pattern. Although the polymorph prediction method was unable to distinguish between these two forms, Rietveld refinement could generate either of them. A more stable structure which contains a two dimensional hydrogen bonding network was also predicted, which may be the structure that would form at normal room temperatures and pressures. Estrone and quinacridone exist in polymorphic forms which were predicted in the correct order of stability. This program is now commercially available as a component of the crystal modelling software CERIUS2 (available from Molecular Simulations Limited). The fact that it is a 'black box' means that it cannot be evaluated or modified by the user.

Another promising method for predicting three dimensional crystal structures may be through the prediction of one or two dimensional optimum packings. Following the work of Kitaigorodsky on the relationship between molecular structures and possible packing modes[17], Perlstein defined Kitaigorodsky's Aufbau Principle as follows: "Molecular crystals are built up from 1-dimensional aggregates which are packed together to form 2-dimensional monolayers which in turn are packed together to form full 3-dimensional crystalline solids"[19]. Perlstein[18,19] used a Monte Carlo cooling technique to predict the packing of one dimensional molecular aggregates. Only four types of one dimensional aggregate commonly occur within crystal structures[20]. These comprise lines of molecules related by inversion, translation, screw and glide symmetry operations. Perlstein[19] showed that, of 60 aggregate structures randomly selected from the Cambridge Structural Database, 53 were found to lie at a local minimum for the isolated chain which was 0.1-5.7 kcal mol^{-1} above the global minimum. However, many of these structures were found to contain surface cavities generated by the screw, glide or inversion offset, which could not be filled by parts of similar molecules. These results, together with earlier work on predicting

translation aggregates[18], suggest predicting three dimensional crystal structures can be broken down into substructures, each of which is a local energy minimum according to Kitaigorodsky's Aufbau principle.

A constant pressure molecular dynamics (MD) technique has also been developed[21] to search for possible crystal structures. MD techniques involve solving the equations of motion subject to certain boundary conditions and so necessarily include an element of time. The initial velocities are set proportional to the forces acting on the structures except when the forces are zero by symmetry, in which case the initial velocities are set in a random manner. This produces a large number of candidate structures. In order to remove redundant points the following process is carried out. The structure produced at the current timestep (S) is compared with that produced at the previous timestep (S'). If the standard deviation of the intermolecular distances is greater than a preset threshold, then S' is accepted as a candidate structure and S is stored in a temporary buffer. If the standard deviation is less than the preset threshold, then S' will be updated with S only if the energy of S is lower than that of S'. This process generates a number of candidate structures, each of which are then subjected to minimisation of the free energy (which can be obtained from a partition function using the harmonic approximation) to produce a number of hypothetical crystal structures. This method reproduced successfully the crystal structure of CO_2 and both the room temperature and high pressure forms of benzene. However, for pyrimidine the MD search followed by lattice energy minimisation was unable to predict the structure minimised from experiment. The difference from the observed packing for the hypothetical structures for pyrimidine was in the orientation of some of the molecules which resulted in differences in the lattice vectors and space groups. When the molecules were rotated so that they were orientated as in the observed structure, the predicted structure after lattice energy minimisation corresponded to the global minimum in the lattice energy. This was attributed to the nature of the potential energy surface for pyrimidine, as there were found to be many local energy minima having negligible energy difference in preliminary calculations.

5.3.2 "Static" Methods

Gavezzotti[22] suggested a method (using the program PROMET) of predicting molecular crystal structures based on generating nuclei of 2-4 molecules using the

most common symmetry elements (inversion centre, screw, glide and translation). Promising clusters are selected on the basis of their cohesive energy and are built up in to full 3D crystal structures.

This produces a large number of hypothetical crystal structures. Those which fall within a certain range of packing energy and packing coefficient (based on a statistical analysis of known compounds) are then subjected to a full lattice energy minimisation first by the steepest descent method and then using the program PCK83[23]. This method has been partially successful for predicting the structures of a number of known hydrocarbons. The method has also been used to predict a possible polymorph of aspirin[5] following suggestion that the existence of different crystal forms may be due to morphology differences between crystals of the same phase[24]. PROMET has also been used to investigate crystal packing in the compound 7-dimethylaminocyclopenta[c]coumarin[25]. This exists in two polymorphic forms, one of which packs in a centrosymmetric space group (Pbca) and the other in a non-centrosymmetric space group (Pna2$_1$) resulting from the different conformations of the molecules in each case. The polymorph in space group Pbca contains molecules which are planar, while the polymorph in Pna2$_1$ contains molecules which are twisted along the flexible 5-membered ring of the molecule. Both conformers were packed in the space groups P2$_1$, P2$_1$/c, P2$_1$2$_1$2$_1$ and Pna2$_1$ and the results used to provide possible explanations for the reasons why the two conformers pack in different space groups. The formation of the centrosymmetric or non-centrosymmetric polymorph results from the molecular conformation when crystallisation occurs, since the molecules still have conformational freedom at nucleation time. If the molecule is in a planar conformation at nucleation time then centrosymmetric dimers are formed and the resulting space group is Pbca. However, if the molecular conformation is twisted at nucleation time, then the molecules are driven by electrostatic forces to form non centrosymmetric dimers and a crystal structure in space group Pna2$_1$ with herringbone packing is formed. This work also showed that there are a large number of possible alternative crystal structures with energies between the two observed polymorphs which can be considered as possible alternative crystal packings. PROMET has both advantages and disadvantages. The method of building up clusters using certain symmetry operators means that more than one space group can be studied using results

from one set of clusters. This may be most appropriate for polar molecules. However, the method is very time consuming and also requires the user to have an intimate knowledge of space group symmetry and geometrical crystallography.

Another method[26], which has been used to predict the structures of hydrocarbons, relies on the fact that close packing is only possible in 13 of the 230 space groups[27]. To create a crystal structure, a reference molecule is rotated around each of the three Cartesian axes and at each orientation a reference unit cell is constructed. The crystal structure is then subjected to energy minimisation using a crude Lennard-Jones potential (whose parameters were derived by empirical fitting to low temperature crystal structures) and only including the pairwise interactions between atoms in molecules whose centres of mass are separated from the reference molecule by less than the maximum molecular diameter plus the minimum van der Waals radius.

The energies of up to ten thousand structures produced are then refined using a Lennard-Jones potential together with an electrostatic potential calculated using molecular multipole moments. The energies of each structure are then compared with the three best structures within each space group and the ten best overall. The use of molecular multipole moments was justified by the authors as it gave reasonable agreement with more accurate models for the electrostatic interaction and was the only one found fast enough to be able to be used with every close packed structure. As pointed out by the authors, this method is inappropriate for cases where the charge distribution of one molecule penetrates that of another. Thus, it will not give a good representation of the charge distribution for most molecular crystals, particularly polar and hydrogen bonded ones, let alone certain relative orientations of smaller molecules where the charge distributions are close together.

This method found a minimum energy structure corresponding to the experimental structure for ten of the sixteen molecules studied. For the other six, the predicted structure was very similar to the experimental one and the inability to distinguish the predicted structures from the experimentally observed one was attributed to the need for more accurate and realistic intermolecular potentials.

The program MOLPAK has been developed by Holden et al.[28] to generate crude hypothetical structures by considering their packing density. The original

version of the program could handle the triclinic, Z=2 monoclinic and primitive Z=4 monoclinic and orthorhombic space groups, accounting for 70% of the known C, H, N, O and F containing crystal structures with one molecule in the asymmetric unit cell. MOLPAK constructs hypothetical crystal structures by bringing molecules into contact with a central molecule as defined by the coordination sphere corresponding to the space group and a crude repulsion potential. The minimum unit cell volume is evaluated as the central molecule is rotated around each of the Cartesian axes for each coordination sphere type, producing a three dimensional map of volume against orientation. The most dense structures are then refined by lattice energy minimisation using the program WMIN[29].

This program was able to reproduce correctly the crystal structures of a number of explosive materials in the known space group. However, structures were often found in other space groups that were only slightly less stable in energy and in some cases the lattice energy for structures in the wrong space groups was lower than that calculated from minimising the experimental structure. This was mainly attributed to inadequacies in the *exp-6* parameters that had been developed for the final lattice energy minimisation, but could also indicate unknown polymorphism. The ability of an AM1 optimised test probe to determine the correct crystal structure was also assessed. It was found that the predicted crystal structure depended on the similarity between the geometry of the molecule in the test probe and the crystal. MOLPAK has also been used to solve the crystal structures of 1,1,5,5-tetranitro-[4]peristylane[30] and 1,3,5,7-tetranitro-3,7-diazabicyclo[3.3.0]octane[31].

Thus MOLPAK provides a robust and user friendly method for predicting molecular crystal structures. The authors suggested the ability to use conformationally flexible test probes and further refinements of the potential parameters used in WMIN as further improvements to producing a useful prediction program.

MOLPAK has also recently been interfaced with DMAREL so that the final refinement stage can use an accurate, distributed multipole based model for the intermolecular forces and thus provide greater confidence that the potential model extrapolates to hypothetical structures. This was used[32] to study the crystal packing of uracil, 6-azauracil and allopurinol. In all cases, the observed structure was found, corresponding to the global minimum in the lattice energy. However, in all cases

alternative hydrogen bonded structures were found within 10 kJ mol^{-1} of the global minimum of the lattice energy, suggesting that other factors than just the static lattice energy need to be taken into account to predict experimental crystal structures.

A similar method to MOLPAK has also been used to generate possible crystal structures of the six hexapyranoses[33]. This study first involved a systematic search for possible structures with only one molecule in the asymmetric unit within the space group $P2_12_12_1$. This produced of the order of a thousand possible structures within 10 kcal mol^{-1} of the observed structure. It was suggested that the large number of possible polymorphs may be due to the notorious crystallisation propensity of the hexapyranoses. The experimental structure was always found amongst the low energy structures and for four of the structures this was either the global minimum or only slightly above it. However, in two cases the energy of the predicted structure was over 5 kcal mol^{-1} above the global minimum. Thus, the authors suggested that improvement of the force field would be necessary to improve their predictions.

5.4 Summary

Thus, there are already a number of promising methods that are able to generate possible crystal structures either using Monte Carlo or molecular dynamics techniques or by systematically searching for structures with high density, favourable nucleus interaction energies or formation of one and two dimensional aggregates. All the methods of crystal structure prediction use the close packing principle and static lattice energy minimisation at 0K for the final structure refinement. Such methods are most inappropriate when the potential energy surface is very flat and there are multiple adjacent minima[12]. However, a minimisation technique which is independent of the crystal symmetry may well allow the prediction of minima with different symmetry from the initial structure.

The procedure to generate the initial structures may well depend on the types of molecules involved. For example, MOLPAK was designed for predicting the structures of energetic materials and so for such molecules a search which is determined by packing density will be most appropriate. Optimising nucleus interaction energies may well be more suited to polar and hydrogen bonded structures, where the variation in the packing energy with orientation will be larger than it was for hydrocarbons.

5.5 Conclusions

All the crystal structure prediction work performed so far has demonstrated that a large number of structures are predicted with similar energies to the observed structures for most molecules. The use of more rigorous theoretically based model potentials will give greater confidence that the potential will reliably extrapolate to hypothetical structures, but this will only remove one uncertainty in predicting molecular crystal structures[34]. Although a more rigorous treatment of the thermodynamics may improve the calculated energies at finite temperatures Gavezzotti and Filippini suggested that including the lattice energy vibrations did not affect the relative stabilities of known polymorphs[5]. This work also demonstrated empirically that there is negligible energy difference between alternative crystal structures. Hence, in many cases, hypothetical structures are found to be similar in stability to the observed structures and further research into the factors which control polymorphism, such as including the kinetic effects as well as the effects of the crystal environment (e.g. solvent) is required before genuine *ab initio* crystal structure prediction becomes possible.

In the following chapter the program MOLPAK is combined with the crystal structure minimisation program DMAREL to produce a powerful crystal structure prediction method. This is applied to predicting polymorphs of alloxan and used to account for the lack of hydrogen bonds in the observed crystal structure.

References for Chapter 5

1. McCrone, W.C. in *Physics and Chemistry of the Organic Solid State*, edited by Fox, D.; Labes, M.M. and Weissberger, A., Interscience, New York, vol II, p 725, 1965.

2. Kuhnert-Brandstätter, M.; Riedmann, M. *Mikrochim. Acta* **1987**, *2*, 107.

3. Kuhnert-Brandstätter, M.; Riedmann, M. *Mikrochim. Acta* **1989**, *1*, 81.

4. Kuhnert-Brandstätter, M.; Riedmann, M. *Mikrochim. Acta* **1989**, *2*, 173.

5. Gavezzotti, A.; Filippini, G. *J. Am. Chem. Soc.* **1995**, *117*, 12299.

6. Cholerton, T.J.; Hunt, J.H.; Klinkert, G.; Martin-Smith, M. *J. Chem. Soc., Perkin Trans. 2* **1984**, 1761.

7. Zyss, J.; Nicoud, J.F. *Current Opinion in Solid State and Materials Science* **1996**, *1*, 533.

8. Curtin, D.Y.; Paul, I.C. *Chem. Rev.* **1981**, *81*, 525.

9. Maddox, J. *Nature* **1988**, *335*, 201.

10. Ball, P. *Nature* **1996**, *381*, 648.

11. Gavezzotti, A. *Acc. Chem. Res.* **1994**, *27*, 309.

12. Price, S.L. In *Computer Modelling in Inorganic Crystallography*, Ed. C.R.A. Catlow, Academic Press, London, 1997.

13. Karfunkel, H.R.; Gdanitz, R.J. *J. Comput. Chem.* **1992**, *13*, 1171.

14. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N. ; Teller, A.H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

15. Leusen, F.J.J. *J. Cryst. Growth* **1996**, *166*, 900.

16. Wiles, D.B.; Young, R.A. *J. Appl. Cryst.* **1981**, *14*, 149.

17. Kitaigorodsky, A.I. *Organic Chemical Crystallography*, Consultants Bureau, New York, 1961.

18. Perlstein, J. *J. Am. Chem. Soc.* **1992**, *114*, 1955.

19. Perlstein, J. *J. Am. Chem. Soc.* **1994**, *116*, 455.

20. Scaringe, R.P.; Perez, S. *J. Phys. Chem.* **1987**, *91*, 2394.

21. Tajima, N.; Tanaka, T.; Arikawa, T.; Sakurai, T.; Teramae, S.; Hirano, T. *Bull. Chem. Soc. Jpn.* **1995**, *68*, 519.

22. Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622.

23. Williams, D.E. PCK83. QCPE Program 548. Quantum Chemistry Program Exchange, Chemistry Department, Indiana University, Bloomington, Indiana, 1983.

24. Kim, Y.; Matsumoto, M.; Machida, K. *Chem. Pharm. Bull.* **1985**, *33*, 4125.

25. Gavezzotti, A. *Acta Cryst.* **1996**, *B52*, 201.

26. Chaka, A.M.; Zaniewski, R.; Youngs, W.; Tessier, C.; Klopman, G. *Acta Cryst.* **1996**, *B52*, 165.

27. Kitaigorodsky, A.I. *Molecular Crystals and Molecules*, Academic Press, New York, 1973.

28. Holden, J.R.; Du, Z.; Ammon, H.L. *J. Comput. Chem.* **1993**, *14*, 422.

29. Busing, W.R. *A Computer Program to Model Molecules and Crystals in Terms of Potential Energy Functions*, Report ORNL-5747, Oak Ridge National Laboratory, Oak Ridge, TN, 1981.

30. Ammon, H.L.; Du, Z.; Holden, J.R.; Paquette, L.A. *Acta Cryst.* **1994**, *B50*, 216.

31. Ammon, H.L.; Du, Z.; Gilardi, R.D.; Dave, P.R.; Forohar, F.; Axenrod, T. *Acta Cryst.* **1996**, *B52*, 352.

32. Price, S.L.; Wibley, K.S. *J. Phys. Chem. A.* **1997**, *101*, 2198.

33. van Eijck, B.P.; Mooij, W.T.M.; Kroon, J. *Acta. Cryst.* **1995**, *B51*, 99.

34. Price, S.L. *J. Chem. Soc., Faraday Trans.* **1996**, *92*, 3005.

Chapter 6   Application of Crystal Structure Prediction Method:

The Lack of Hydrogen Bonds in the Crystal Structure of Alloxan

In this chapter we use the crystal structure prediction program MOLPAK to generate possible hypothetical crystal structures. The most promising are then subjected to minimisation using a realistic distributed multipole based intermolecular potential. This procedure generated the observed structure as the global minimum in the lattice energy with hydrogen bonded polymorphs less stable by at least 5 kJ mol^{-1}. We also predict isolated dimer structures with close (C=O)···(C=O) interactions to have similar energies to the hydrogen bonded dimers. This is used to account for the lack of hydrogen bonds in the observed crystal structure of alloxan and illustrates the limitations of the functional group approach to predicting molecular crystal structures. This work has been published in a slightly shortened form[1].

6.1 Introduction

The crystal structure of alloxan[2] (space group P4$_1$2$_1$2) does not contain conventional hydrogen bonds, even though the alloxan molecule is made up of four C=O groups and two N-H groups (see Figure 1). This is an unusual exception to the general rule that good proton donors and acceptors are used in hydrogen bonding in molecular crystal structures[3]. The crystal structure represents a balance between nearest neighbour and more distant interactions and the requirement for translational symmetry. Thus, to understand the alloxan crystal structure, we calculate the most favourable structure of isolated dimers and compare these structures with those found in possible crystal structures.

Previous studies of alloxan suggest that the unusual molecular structure does affect the charge distribution of the N-H and C=O groups significantly, with the anisotropic charge distribution around the carbonyl group contributing significantly to the electrostatic interactions[4]. Calculations using the experimental X-ray charge distribution for the electrostatic interactions[5] showed that the hydrogen bond energy was small, even for the optimised dimers and was very weak[4,5] for two molecules in the crystal orientation corresponding to the shortest O···H contacts (which could be described as a very long bifurcated hydrogen bond to O1 and O2). Although the oxygen atoms are not conventionally hydrogen bonded to the protons, there are short (C=O)···(C=O) contacts (C3···O2=2.8 Å, C2···O2=3.0 Å). The existence of close

(C=O)···(C=O) contacts has been widely observed in crystalline oxohydrocarbons[6] and plots of the orientational distribution of the close (C=O)···(C=O) contacts in the molecular crystal structures in the Cambridge Structural Database suggest that the contacts found in the alloxan crystal are favourable[7].

The main feature of this study is the use of a realistic distributed multipole electrostatic model, obtained from a Distributed Multipole Analysis (DMA)[8] of an *ab initio* wavefunction. The use of a theoretically well-founded electrostatic model provides confidence that the intermolecular potential is transferable to hypothetical structures. The other contributions to the intermolecular potential are represented by an empirically fitted *6-exp* repulsion-dispersion potential. This model has previously been used to model the crystal structures of a wide range of forty organic molecular crystals[9], from hydrocarbons to nucleic acid bases, which it reproduces well within the limitations of the static minimisation method (see Chapter 4). The methodology used in this study was able to predict[10] successfully the observed crystal structure (as the global minimum) and other possible polymorphs for azauracil, allopurinol and uracil.

6.2 Method

The molecular structure of alloxan was taken directly from the experimental room temperature crystal structure[2]. The hydrogen atoms in the crystal structure were positioned to give a standard bond length[11] of 1.01 Å for N-H along the experimental bond direction. The model used for the intermolecular potential consists of an *ab initio* based DMA model for the electrostatic contribution together with an empirically fitted *6-exp* repulsion-dispersion potential. The electrostatic model included all terms in the atom-atom multipole series up to R^{-5}, using atomic multipoles up to hexadecapole which had been obtained by a Distributed Multipole Analysis[8] of a 6-31G** SCF wavefunction[12] of the isolated molecule, using the program suite CADPAC[13]. The atomic multipoles were scaled by the widely used fudge factor of 0.9[14] to account for the neglect of electron correlation inherent in an SCF wavefunction, which leads to an overestimate of the molecular dipole moment by 12-15%[15]. The *6-exp* repulsion-dispersion potential parameters for C, N and O were empirically fitted to azahydrocarbon[16] and oxohydrocarbon[17] crystal structures and the parameters for polar hydrogens (H_p) were derived from intermolecular perturbation theory calculations on the formamide-formaldehyde complex[18] as described in chapter

4.

This model potential was used to predict possible dimer structures, by minimisation of the energy from a range of starting orientations using the program ORIENT[19] and to predict hypothetical crystal structures using the new crystal structure minimisation program DMAREL[20], starting from the structures generated by the program MOLPAK[21].

MOLPAK was used to perform an extensive and systematic search for hypothetical crystal structures of alloxan within the most common space groups for organic molecules with one molecule in the asymmetric unit, namely $P\bar{1}$, $P2_1$, $P2_1/c$, $P2_12_12_1$, P1, $Pca2_1$, $Pna2_1$ and Pbca and within C2/c using the two-fold axis symmetry of the molecule. (The space group of the experimental structure ($P4_12_12$), with half a molecule in the asymmetric unit, is not treated explicitly by MOLPAK.) MOLPAK builds minimum volume packing arrangements for each of the 18 different coordination types found for these common space groups by moving the molecules along the three orthogonal Cartesian axes for different orientations of a central molecule until they are in van der Waals contact, as defined by a crude pseudo-repulsion potential. The central molecule was rotated from -90° to +90° in steps of 10° around each of the Cartesian axes to produce a total of 19^3=6859 structures in most coordination geometries. The most densely packed structures (usually 25) for each coordination type were then refined to within 2° and considered as possible hypothetical starting point structures.

The lattice energies of the most promising hypothetical crystal structures for each coordination type produced by MOLPAK were calculated using the program DMAREL. The 20 lowest energy structures for each coordination type corresponding to the most common space groups $P\bar{1}$, $P2_1$, $P2_1/c$ and $P2_12_12_1$ and 10 lowest energy structures for the coordination types corresponding to the less common space groups P1, $Pca2_1$, $Pna2_1$ and Pbca and the 8 for the special symmetry arrangements in C2/c, were subjected to full lattice energy minimisation with respect to the position and orientation of each molecule in the unit cell and distortions in the size and shape of the unit cell. The resulting low energy minima in the lattice energy were clustered into structures that were so similar that they would correspond to the same structure in a finite temperature dynamic simulation, by examining the matrix of intermolecular

contact distances and the lattice parameters when transformed to the conventional unit cell[22]. (The structures thus deemed to be effectively equivalent corresponded to a variation in molecular volumes of less than 0.5 Å^3, in corresponding cell axes of less than 0.05 Å and an energy range of less than 1 kJ mol[-1]). The matrix of intermolecular distances (as calculated by NIPMAT[23]) and 3-dimensional graphic displays were used to determine the hydrogen bonding or other close contacts in the hypothetical crystal structures.

6.3 Results

Five distinct dimer structures were found (Figure 1 and Table 1). Three of these are planar doubly hydrogen bonded structures with energies of -34.8 kJ mol[-1], -35.7 kJ mol[-1] and -37.0 kJ mol[-1] respectively. None of these packing motifs are observed in the experimental crystal structure. However, there are also two minima with (C=O)···(C=O) interactions. One (COCOTIL) is a 'tilted' structure with an energy of -36.2 kJ mol[-1] (comparable to the hydrogen bonded structures). This has close C2···O2 and C3···O2 contacts of 3.0 Å and a long O2···H contact of 2.2 Å and is similar to one of the nearest neighbour orientations in the experimental crystal structure. The second (COCOST) has the molecular planes more parallel and the C=O groups displaced from anti-parallel vertical stacking by interactions with the neighbouring C=O group.

Minimising the lattice energy starting from the experimental room temperature crystal structure (Table 2) gives a root mean square percentage error in the independent lattice vectors of 1.9% confirming that the potential is adequate for calculations on alloxan. (This error is larger, 3.0%, if the comparison is made with the 42K neutron structure[4], as might be expected since the repulsion-dispersion potentials were fitted to room temperature crystal structures. The largest error is in the c axis length, which also shows the largest thermal expansion.) This structure (denoted MIN) has a lattice energy of -120.3 kJ mol[-1]. This estimate of the (as yet) undetermined heat of sublimation should be more accurate than the -107 kJ mol[-1] estimated[24] using charges derived from a semi-empirical charge density.

The MOLPAK/DMAREL crystal structure prediction procedure generates dozens of hypothetical crystal structures with energies between -88 kJ mol[-1] and -122 kJ mol[-1], as shown in Figure 2. Table 2 shows representatives of the few clusters of

distinct minimum energy crystal structures within about 10 kJ mol^{-1} energy of MIN. The lowest energy cluster of structures (identified by orange circles in Figure 2), of which AQ24 is an example, has the molecules packed in a 3D network of interlocking spirals by (C=O)···(C=O) interactions, as in the experimental structure. The intermolecular contacts, as well as the lattice vectors, are all within 0.1 Å of those of MIN. Although some of these structures are of slightly lower energy than MIN and many have, or approximate, higher symmetry than the space group of the starting point structure, they appear to be equivalent within the limitations of a static model. Thus the lowest energy structure found by the search procedure has effectively located the experimental structure, validating the search procedure.

The next most stable crystal structures (Table 2) have energies of around -115 kJ mol^{-1} and are significantly less stable than the experimental structure, though within the energy range of possible polymorphs. They all contain hydrogen bonds and are all less dense than the experimental structure which has an unusually high density[2,4] (1.93 g cm^{-3} at 295 K). There is a range of hydrogen bonding types: from AF24 (P2$_1$, red triangles in Figure 2) with two conventional hydrogen bonds per molecule, though each molecule is hydrogen bonded to four different neighbours, to structures with a mixture of conventional and elongated hydrogen bonds, with a proton often interacting with oxygen atoms in different molecules. All of these structures are quite complex, with molecules tilted relative to each other and short (C=O)···(C=O) contacts and so are stabilised by a mixture of hydrogen bonding and (C=O)···(C=O) contacts. Structures involving two hydrogen bonds to the same neighbouring molecule occur at even less favourable lattice energies, such as the lowest energy P$\bar{1}$ sheet structure (green rectangles in Figure 2) which has planes containing the HBOND01 dimer structure cross linked by O2···H bonds with an energy of -104.4 kJ mol^{-1}.

Thus, crystal structures containing hydrogen bonds have been located within 10 kJ mol^{-1} of the global minimum and so might be energetically feasible polymorphs, but are not as stable as the experimental structure without hydrogen bonds.

6.4 Discussion

The unusual bonding environment of the C=O and N-H groups in alloxan appears to produce weak C=O···H-N hydrogen bonds and competitively strong (C=O)···(C=O) interactions. Thus although alloxan molecules are expected to form

hydrogen bonded dimers in the gas phase other dimer structures that are stabilised by strong (C=O)···(C=O) interactions are competitive. The uncertainties in the model intermolecular potential, which assumes that all but the electrostatic contribution can be represented by an empirical isotropic repulsion-dispersion model, are unlikely to be of a magnitude to change this qualitative conclusion.

The nearest neighbour (C=O)···(C=O) interactions appear to stabilise the experimental crystal structure of alloxan, as they are similar to the 'tilted' gas-phase dimer structure. This relative orientation of C=O groups was found to be quite common in a survey of structures in the 1995 release of the Cambridge Structural Database, particularly for molecules with a large amount of delocalisation[25]. Intermolecular Perturbation Theory calculations on the propanone dimer give the most favourable structure (-22 kJ mol^{-1}) for a square anti-parallel arrangement of C=O dipoles[26]. This confirms that (C=O)···(C=O) stabilisation will be significant, but the different minimum energy structures found in this study are likely to be more stable for alloxan because of the multiple (C=O)···(C=O) interactions.

The experimental structure is more stable than any others found in the search through the most common space groups. A few hydrogen bonded structures have been found which are within the energy range of possible polymorphs. Although it is possible that not all low energy crystal structures have been found, the search has been sufficiently extensive to show that the constraints of crystal packing can be satisfied with a lower energy structure without hydrogen bonds, but with favourable (C=O)···(C=O) interactions.

6.5 Conclusions

The observed crystal structure of alloxan is an exception to the general rule that hydrogen bonding dominates crystal packing and is an example of where consideration of just the functional groups fails to predict the crystal structure. The charge distribution of alloxan results in weak hydrogen bonded dimers in the gas phase, with other structures involving close (C=O)···(C=O) contacts being competitive in energy. One of these motifs appears in the experimental crystal structure. Since this crystal structure has the lowest energy found in a search for possible crystal structures, it appears the (C=O)···(C=O) stabilised structure is able to pack into a crystal structure more favourably than any hydrogen bonding motif.

We have shown in this chapter that the combination of the program MOLPAK to generate crude crystal structures with a distributed multipole based model potential for the final refinement of the crude structures is a powerful and robust method for crystal structure prediction. In the following chapter the main conclusions of the work in this thesis are summarised and suggestions for further work are given.

Table 1 Intermolecular Energies and Important Contacts in the Predicted Alloxan Dimer Structures

Structure	Energy / kJ mol⁻¹	Short O···H contacts	Distance / Å	Short C···O contacts	Distance / Å
HBOND02	-34.8	O2···H	2.03	C2···O2	3.53
HBOND01/02	-35.7	H···O1	2.00	C2···O1	3.57
		O2···H	2.03	C1···O2	3.60
HBOND01	-37.0	H···O1	2.00	C1···O1	3.64
COCOTIL	-36.2	O2···H	2.15	C3···O2	2.99
		O3···H	2.69	C2···O2	2.99
COCOST	-34.2	H···O3	2.94	C2···O2	3.04, 3.15
				C3···O2	3.10, 3.11

Table 2 Structures Corresponding to Minima in the Lattice Energy of Alloxan

Structure	Energy / (kJ mol⁻¹)	V_m / Å³	a / Å	b / Å	c / Å	α °	β °	γ °	O H distances / Å	C O distances / Å
OBSD (P4₁2₁2)	-118.9	122.1	5.886	5.886	14.100	90.00	90.00	90.00	2.37 (H O1) 2.44 (H O2)	2.79 (C3 O2) 2.97 (C2 O2)
MIN (P4₁2₁2)	-120.3	122.3	5.820	5.820	14.442	90.00	90.00	90.00	2.32 (H O2) 2.36 (H O1)	2.84 (C3 O2) 3.06 (C2 O2)
AQ24 (P2₁2₁2₁)	-121.1	122.1	5.815	5.816	14.439	90.00	90.00	90.00	2.31, 2.32 (H O2) 2.34, 2.35 (H O1)	2.84 (C3 O2) 3.03, 3.07, 3.08 (C2 O2)
AK23 (P2₁/c)	-115.4	125.8	8.139	5.946	10.771	90.00	105.11	90.00	2.04, 2.33 (H O2) 2.29 (H O1)	2.86 (C3 O1) 3.00, 3.03 (C2 O1) 3.08 (C3 O2) 3.13 (C1 O3)
CC22 (Pbca)	-114.6	125.8	5.801	8.375	20.715	90.00	90.00	90.00	2.09, 2.41 (H O2) 2.32 (H O1)	2.91 (C3 O1) 2.99 (C3 O2) 3.05, 3.09 (C2 O1) 3.09 (C1 O3)
CB21 (Pbca)	-114.6	126.0	6.759	8.303	17.959	90.00	90.00	90.00	2.13 (H O1) 2.28, 2.34 (H O2) 2.59 (H O3)	2.93 (C1 O2) 2.97 (C2 O1) 3.02 (C3 O2) 3.07 (C3 O1) 3.18, 3.21 (C2 O2)
BD15 (Pna2₁)	-114.2	127.9	5.762	6.144	14.448	90.00	90.00	90.00	2.04, 2.70 (H O2) 2.04 (H O1) 2.59 (H O3)	2.95, 3.09 (C3 O2) 3.01, 3.09, 3.18 (C2 O2)
AF24 (P2₁)	-113.5	129.9	6.758	5.911	6.860	90.00	108.60	90.00	1.98, 2.14, 2.66 (H O2)	2.99 (C1 O1) 3.02 (C2 O2) 3.08 (C2 O3) 3.10 (C2 O1) 3.16 (C3 O2)

Footnotes on next page

Footnotes for Table 2

The experimental structure (OBSD) is compared with the corresponding minimum in the lattice energy (MIN) and structures predicted starting from MOLPAK generated structures, denoted by the MOLPAK packing type/number. V_m is the unit cell volume per molecule. The lattice energy was evaluated by direct summation to 15Å for all terms except the charge-charge, charge-dipole and dipole-dipole electrostatic contribution which were evaluated by Ewald summation.
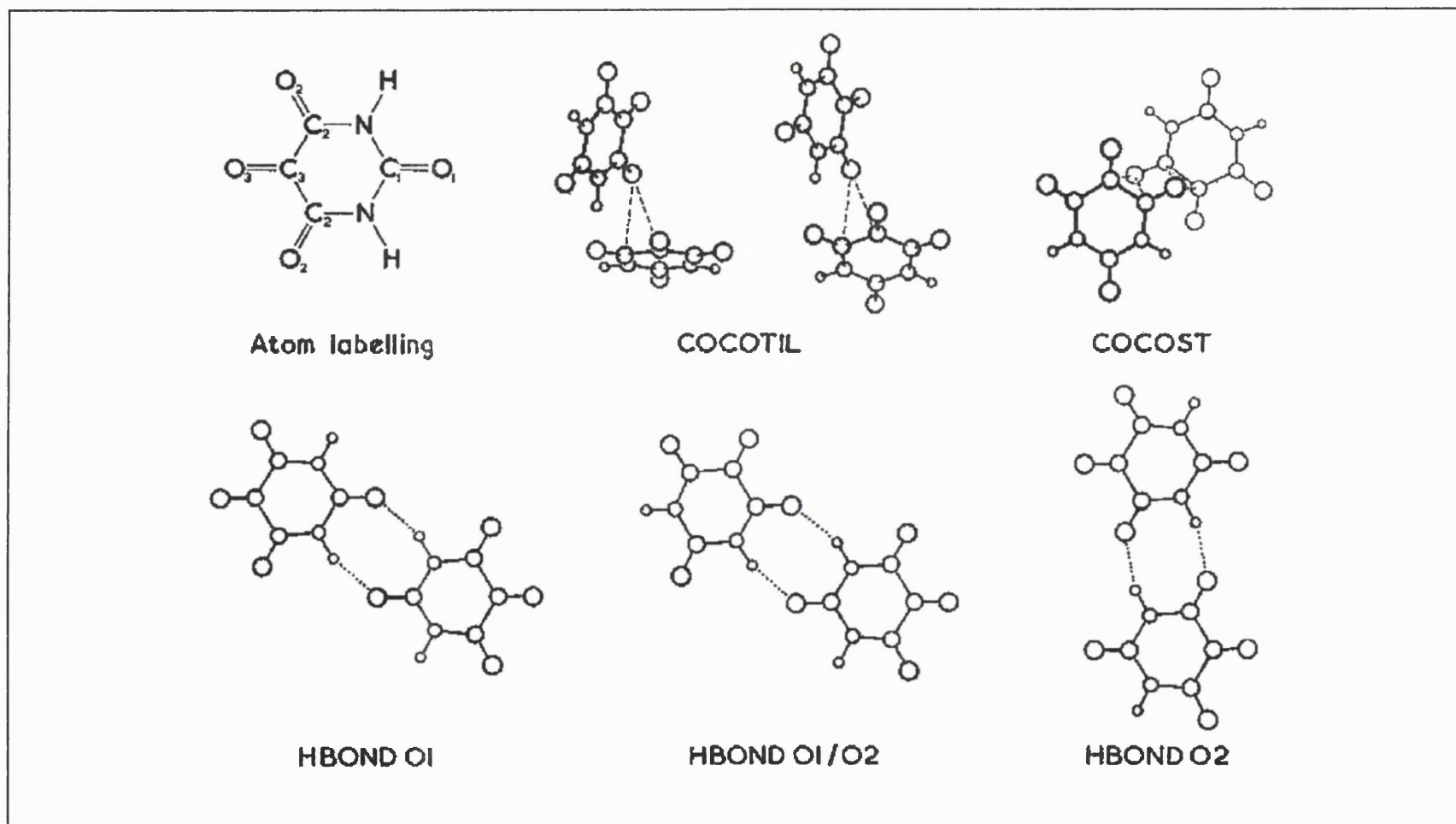
Figure 1 Atom Labelling for Alloxan and Predicted Dimer Structures (Table 1). Short Intermolecular C⋯O Contacts are Shown as Dashed Lines and Conventional Hydrogen Bonds as Dotted Lines.
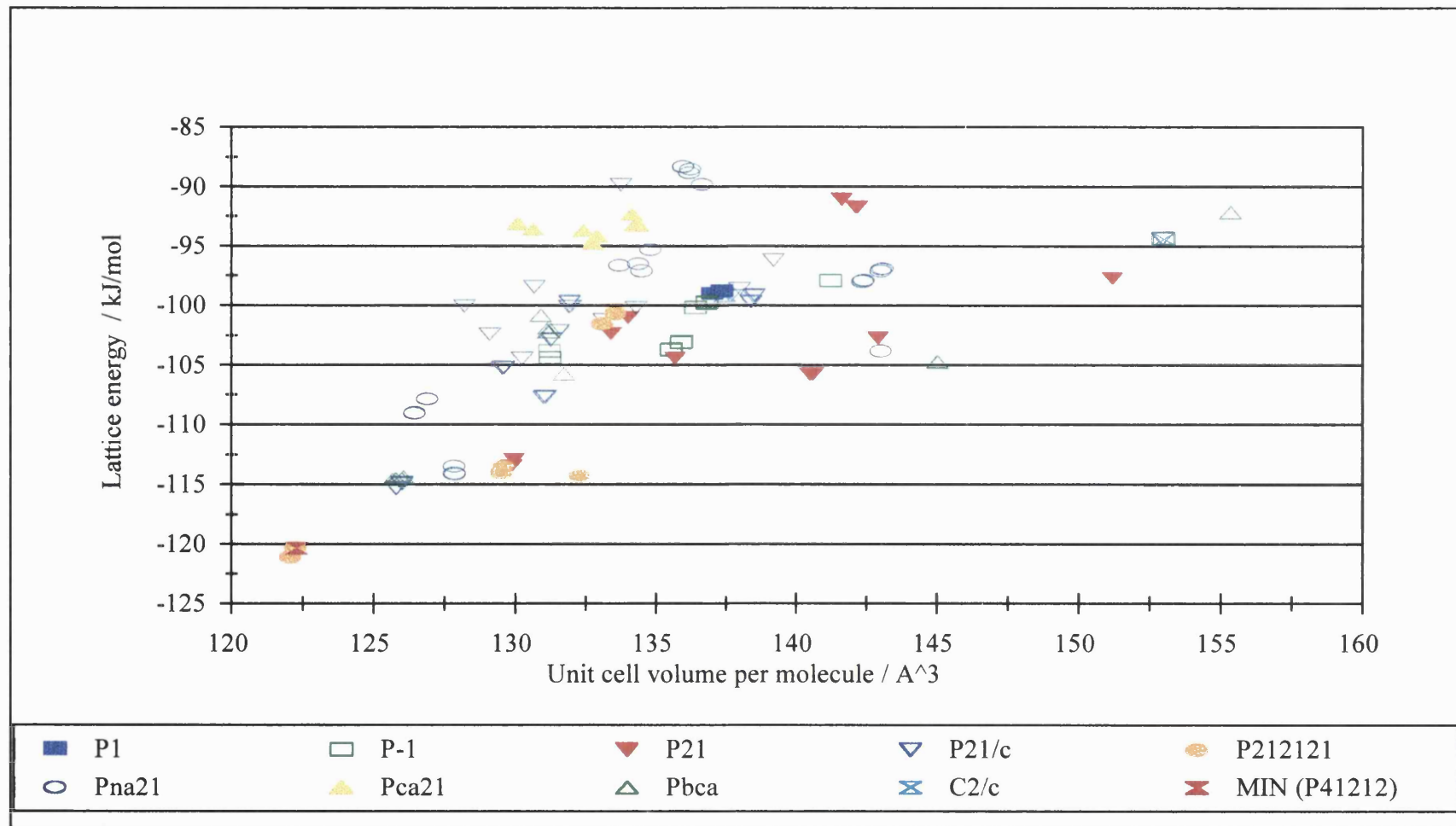
Figure 2 Minimum Energy Crystal Structures of Alloxan.

References for Chapter 6

1. Coombes, D.S; Nagi, G.K.; Price, S.L. *Chem. Phys. Lett.* **1997**, *265*, 532.

2. Bolton, W. *Acta Cryst.* **1964**, *17*, 147.

3. Etter, M.C. *Acc. Chem. Res.* **1990**, *23*, 120.

4. Swaminathan, S.; Craven, B.M.; McMullan, R.K. *Acta Cryst.* **1985**, *B41*, 113.

5. Spackman, M.A.; Weber, H.P; Craven, B.M. *J. Am. Chem. Soc.* **1988**, *110*, 775.

6. Gavezzotti, A. *J. Phys. Chem.* **1991**, *95*, 8948.

7. Taylor, R.; Mullaley, A.; Mullier, G.W. *Pestic. Sci.* **1990**, *29*, 197.

8. Stone, A.J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047.

9. Coombes, D.S.; Price, S.L.; Willock, D.J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.

10. Price, S.L.; Wibley, K.S. *J. Phys. Chem. A.* **1997**, *101*, 2198.

11. Allen, F.H.; Kennard, O.; Watson, D.G.; Brammer, L.; Orpen, A.G.; Taylor, R. *J. Chem. Soc., Perkin Trans. 2* **1987**, S1.

12. Hariharan, P.C.; Pople, J.A. *Theor. Chim. Acta* **1973**, *28*, 213.

13. CADPAC5: The Cambridge Analytical Derivatives Package Issue 5.0, A suite of quantum chemistry programs developed by R.D. Amos with contributions from I.L. Alberts, J.S. Andrews, S.M. Colwell, N.C. Handy, D. Jayatilaka, P.J. Knowles, R. Kobayashi, N. Koga, K.E. Laidig, P.E. Maslen, C.W. Murray, J.E. Rice, J. Sanz, E.D. Simandiras, A.J. Stone, and M.-D. Su, Cambridge, U.K., 1992.

14. Cox, S.R.; Williams, D.E. *J. Comput. Chem.* **1981**, *2*, 304.

15. Ryan, M.D. in Modeling the Hydrogen Bond, *ACS Symp. Ser.* **1994**, *569*, 36.

16. Williams, D.E.; Cox., S.R. *Acta Cryst.* **1984**, *B40*, 404.

17. Cox, S.R.; Hsu, L.-Y.; Williams, D.E. *Acta Cryst.* **1981**, *A37*, 293.

18. Mitchell, J.B.O.; Price, S.L. *J. Comput. Chem.* **1990**, *11*, 1217.

19. A.J. Stone, P.L.A. Popelier and D.J. Wales, ORIENT: a program for calculating electrostatic interactions, Version 3.0, University of Cambridge, 1994.

20. Willock, D.J.; Price, S.L.; Leslie, M.; Catlow, C.R.A. *J. Comput. Chem.* **1995**, *16*, 628.

21. Holden, J.R.; Du, Z.; Ammon, H.L. *J. Comput. Chem.* **1993**, *14*, 422.

22. Spek, A.L. *J. Appl. Cryst.* **1988**, *21*, 578.

23. Rowland, R.S. *Acta Cryst.*, in preparation.

24. Poltev, V.I. *Int. J. Quantum Chem.* **1979**, *16*, 863.

25. Baalham, C.A. personal communication.

26. Baalham, C.A.; Lommerse, J.P.M.; Allen, F.H.;Taylor, R. *British Crystallographic Association Annual Meeting*, University of Cambridge, poster CP13, 1996.

Chapter 7   Conclusions and Suggestions for Further Work

In this chapter we summarise the work presented in this thesis and show that we have developed a promising approach to crystal structure prediction. Such methods will be important in many areas of solid state chemistry as well as providing a useful aid to understanding the factors controlling crystal packing and the interplay between molecular structure and crystal structure.

We have used a distributed multipole plus isotropic repulsion dispersion potential to model a wide range of molecular crystal structures ranging from aromatic hydrocarbons to those of importance in pharmaceutics and non linear optics. The accurate model was compared with a charge only plus repulsion-dispersion model obtained by omitting the anisotropic components. We also studied the effect of introducing a scaling factor for the multipoles to counteract the lack of electron correlation inherent in the use of a SCF wavefunction. This theoretically realistic model provides a major improvement over current potentials used to model crystal structures. The remaining errors can be attributed either to the use of rigid molecules or to where anisotropy in other terms in the potential is important. The model is significantly less successful for the structures which contain NH_2 groups which may be partially hybridised and thus the location of the hydrogen atoms is difficult.

However, the majority of the r.m.s. errors in the independent lattice vectors are below 3%. This is comparable with the thermal expansion on going from 0K to room temperature[1] and so the main limitation to obtaining greater accuracy seems to be the use of a static simulation technique. Hence, a dynamical simulation technique taking into account temperature effects would be the next step in improving the crystal structure modelling. This would be a straightforward adaptation of DMAREL since the required expressions for the forces and torques are already programmed.

For larger molecules the ability to model conformationally flexible molecules within DMAREL will be necessary. A partially flexible molecule could be modelled as a set of rigid molecular fragments. These fragments would then be joined using the well parameterised intramolecular potentials that are available in a variety of force fields.

We also attempted to improve the repulsion-dispersion parameters by empirical fitting either to the experimental forces, torques and strains or alternatively to the first

step of the minimisation. However, little or no improvement in the predicted structures was obtained, the most justified fits being to the polar hydrogen parameters which had not been considered in the original fitting of the potentials. It appears that the use of a static simulation technique limits the ability to improve the crystal structure predictions.

The results presented in chapter 4 suggest that anisotropic repulsion potentials are necessary for crystal structures which contain N⋯N contacts such as s-tetrazine[2] and in cases where displacing the hydrogen site into the C-H bond makes a marked difference to the predicted structure. Anisotropic repulsion potentials have also been shown to be required to predict the crystal structure[3] of Cl_2 and hence may be necessary to model crystal structures containing close Cl⋯Cl contacts. Indeed, it is likely that they would be needed where there is a marked dependence of van der Waals contact distance on the angle between the intermolecular contact vector and the C-X bond[4]. Introduction of the functional form for anisotropic repulsion potentials into DMAREL is relatively straightforward. However, they cannot be parameterised using the empirical fitting methods described in this thesis.

There are a large number of methods that are able to generate hypothetical structures as discussd in chapter 5. The refinement of these crude structures is carried out using a static simulation technique. For many molecules there are a large number of possible crystal structures within 10 kJ mol^{-1} of the global minimum and thus must be considered as possible polymorphs.

Finally we have demonstrated (in chapter 6) that combining a powerful crystal structure prediction program with a theoretically rigorous distributed multipole based potential (described in chapter 4) provides a useful method for crystal structure prediction. The combination of MOLPAK to generate the crude structures and DMAREL to refine them was used to obtain possible crystal structures of alloxan. This was able to predict the observed structure, together with a large number of hydrogen bonded polymorphs that were less stable by more than 5 kJ mol^{-1}. The same potential model was also used to predict isolated dimer structures. This information was used to explain the lack of hydrogen bonds and the crystal packing in the observed crystal structure of alloxan, which is a balance between hydrogen bonding and C⋯O interactions. It also serves to illustrate that a simple consideration of the

functional groups cannot be used as a reliable guide to predicting the molecular crystal structure in this case.

Thus, this combination of a powerful method for generating hypothetical crystal structures together with an accurate model for the intermolecular forces provides a robust method of crystal structure prediction and is a useful tool for understanding the factors affecting crystal packing. Such methods are also likely to be useful in solving crystal structures where partial data is available (as has been demonstrated using the program PROMET[5]).

However, although the introduction of a theoretically well based realistic model for the electrostatic contribution to the intermolecular forces will provide confidence that the potential will extrapolate reliably to hypothetical structures, this has removed only one uncertainty in predicting molecular crystal structures. It is clear that considerably more research into other factors such as the crystal solvent and the kinetics of crystallisation need to be taken into account before we will genuinely be able to predict which structures are likely to be formed by a given molecule.

References for Chapter 7

1. Price, S.L. in *Computer Modelling in Inorganic Crystallography* Ed. C.R.A. Catlow, Academic Press, 1997.

2. Price, S.L.; Stone, A.J. *Mol. Phys.* **1984**, *51*, 569.

3. Price, S.L.; Stone, A.J. *Mol. Phys.* **1982**, *47*, 1457.

4. Nyburg, S.C.; Faerman, C.H. *Acta Cryst.* **1985**, *B41*, 274.

5. Gavezzotti, A.; Filippini, G. *J. Am. Chem. Soc.* **1996**, *18*, 7153.

Introduction

The program NEIGHBOURS is needed to convert from crystallographic coordinates to the orthogonal Cartesian coordinates used by DMAREL and is successful for most common space groups for organic molecules. In addition NEIGHBOURS can be used to standardise bond lengths to hydrogen atoms and to paste the atomic multipoles (obtained using the program CADPAC) onto the correct atoms on a subsequent run.

To set up a DMAREL input file containing full multipoles, the usual procedure is as follows:

1) Use NEIGHBOURS to convert from crystallographic coordinates to orthogonal Cartesian coordinates and define the local axis system for the molecules (described in the DMAREL manual). NEIGHBOURS also identifies molecules related by an inversion centre or other symmetry operation that requires a reversal of the z axis of the local axis system.

2) Obtain the coordinates in the local axis system for each molecule by running a MAXI 0 (no minimisation) run of DMAREL. A separate program can then be used to convert the local axis system into a form suitable for input to the *ab initio* program CADPAC. CADPAC is then used to obtain the atomic multipoles from the *ab inito* wavefunction.

3) Use NEIGHBOURS to 'paste in' the atomic multipoles to the DMAREL input file obtained in step (2) above. Then check the minimisation defaults, supply Buckingham repulsion-dispersion parameters etc. and run the minimisation.

Input for NEIGHBOURS

Note: In the following section x implies that one space should be left in the input file

a) The user must first type the name of the file containing the coordinates and the lattice vectors.

The Input for NEIGHBOURS can be of two types,

(1) an FDAT file from the Cambridge Structural Database or

(2) a DMAREL input file produced on a previous run.

Option (2) is usually only necessary when dealing with a crystal structure where molecular symmetry and crystallographic symmetry elements coincide (as described

below in (g)).

b) The name of a file which contains maximum possible bond lengths for all intramolecular bonds found in the structure must then be given. This is necessary for NEIGHBOURS to calculate the molecule to which each atom belongs and the atom connectivities.

The format of each entry in this file is

CA xx CA xx 1.32

Here the maximum possible bond length for an intramolecular carbon-carbon bond is 1.32 Å.

Note in the above that the atom labels must consist of two characters to identify the atom type and the spacing in the line is crucial!

c) The type of coordinate file must then be given, 1 for a CSD input and 2 for other. If the file is of type 'other' then the value of the c lattice vector must also be given.

d) A maximum distance (in Å) must then be given for a list of intermolecular contact distances that are written to fortran channel 15.

e) The user is then given the option to standardise the intramolecular bond lengths to hydrogen atoms, since the location of these by X-ray or neutron diffraction is difficult. The bond lengths are standardised to 1.08 Å for C-H and 1.01 Å for N-H. This list of standard bond lengths can be altered by editing the fortran source code file getdat.f and recompiling the program.

f) Bond centre sites can then be added. These can be used for extra sites in the DMA calculation. The user is given a list of all the intramolecular bonds found and asked to choose which require bond centres.

g) If a CADPAC punch file from a previous DMAREL run is available then this can be added. For many space groups, where a centre of symmetry introduces a reversal of the z axis in DMAREL, the correct multipoles are 'inverted'. The atoms where this has occurred are identified by an 'I' in the third character of the atom label. However, if the molecular and crystallographic symmetry elements coincide then no reversal of the multipoles is usually needed, although this relabelling still occurs. In this case the user is warned of this after step (d) and it is necessary to manually edit the file produced by NEIGHBOURS so that it runs through DMAREL without errors.

h) The name of the file containing the local axis definition should then be given. The

local axis system defines the coordinates of each atom relative to the centre of mass of the molecule to which it belongs. This is necessary for each unique molecule in the asymmetric unit cell (specified by the number n). Axis systems need to be defined that refer to the normal and 'inverted' molecules. It is also necessary to ensure that these axis systems are consistent (i.e. only the z axis is reversed for related molecules). However, if the input is from the CSD then only the molecules that are not affected by a centre of symmetry need to be given.

The local axis system definition should be in the following format

MOLX x n

X x LINE xxx ATM1 x CODA x ATM2 x CODA x b1

Y x PLANE xx ATM1 x CODA x ATM2 x CODA x b1 x ATM3 x CODA x b2

where b1 is the minimum number of bonds between ATM1 and ATM2 and b2 is the minimum number of bonds between ATM1 and ATM3.

The above process provides a nearly complete DMAREL input. All that is now required is to insert the correct Buckingham repulsion-dispersion parameters and set the required number of iterations, multipole scaling factor and cutoffs for lattice summations.