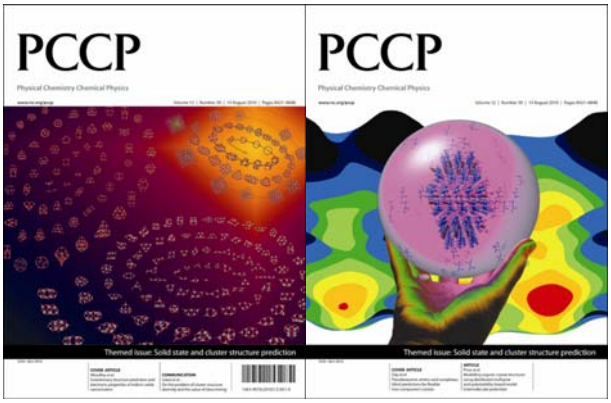
This paper is published as part of a *PCCP* themed issue on solid state and cluster structure prediction

Guest Editors: Richard Catlow and Scott M. Woodley



Editorial

Solid state and cluster prediction
Scott M. Woodley and Richard Catlow, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c0cp90058c

Communication

On the problem of cluster structure diversity and the value of data mining
Alexey A. Sokol, C. Richard A. Catlow, Martina Miskufova, Stephen A. Shevlin, Abdullah A. Al-Sunaidi, Aron Walsh and Scott M. Woodley, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c0cp00068j

Papers

Evolutionary structure prediction and electronic properties of indium oxide nanoclusters
Aron Walsh and Scott M. Woodley, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c0cp00056f

Exploration of multiple energy landscapes for zirconia nanoclusters
Scott M. Woodley, Said Hamad and C. Richard A. Catlow, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c0cp00057d

Pseudoracemic amino acid complexes: blind predictions for flexible two-component crystals
Carl Henrik Görbitz, Bjørn Dalhus and Graeme M. Day, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004055j

Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials
Sarah L. Price, Maurice Leslie, Gareth W. A. Welch, Matthew Habgood, Louise S. Price, Panagiotis G. Karamertzanis and Graeme M. Day, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004164e

Ab initio prediction of low-temperature phase diagrams in the Al–Ga–In–As system, MAs–M´As (M, M´= Al, Ga or In) and AlAs–GaAs–InAs, *via* the global study of energy landscapes
Ilya V. Pentin, J. Christian Schön and Martin Jansen, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004040c

Importance of London dispersion effects for the packing of molecular crystals: a case study for intramolecular stacking in a bis-thiophene derivative
Jonas Moellmann and Stefan Grimme, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003432k

An extensive theoretical survey of low-density allotropy in silicon
Martijn A. Zwijnenburg, Kim E. Jelfs and Stefan T. Bromley, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004375c

Predicting transition pressures for obtaining nanoporous semiconductor polymorphs: oxides and chalcogenides of Zn, Cd and Mg
Winyoo Sangthong, Jumras Limtrakul, Francesc Illas and Stefan T. Bromley, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c0cp00002g

Databases of virtual inorganic crystal structures and their applications
Armel Le Bail, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003907c

Flexibility of ideal zeolite frameworks
V. Kapko, C. Dawson, M. M. J. Treacy and M. F. Thorpe, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003977b

Constant pressure molecular dynamics simulations for ellipsoidal, cylindrical and cuboidal nano-objects based on inertia tensor information
Clive Bealing, Giorgia Fugallo, Roman Martoňák and Carla Molteni, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004053c

Appearance of bulk-like motifs in Si, Ge, and Al clusters
Wen-Cai Lu, C. Z. Wang, Li-Zhen Zhao, Wei Zhang, Wei Qin and K. M. Ho, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004059b

Small germanium sulfide clusters: mass spectrometry and density functional calculations
Joseph J. BelBruno and Andrei Burnin, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003704d

Prediction of the structures of free and oxide-supported nanoparticles by means of atomistic approaches: the benchmark case of nickel clusters
Giulia Rossi, Luca Anghinolfi, Riccardo Ferrando, Florin Nita, Giovanni Barcaro and Alessandro Fortunelli, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003949g

Crystal structure prediction and isostructurality of three small organic halogen compounds
Aldi Asmadi, John Kendrick and Frank J. J. Leusen, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003971c

Aspects of crystal structure prediction: some successes and some difficulties
Michael O'Keeffe, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004039h

Nanopolycrystalline materials; a general atomistic model for simulation
Dean C. Sayle, Benoît C. Mangili, David W. Price and Thi X. Sayle, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/b918990d

Isomorphism between ice and silica
Gareth A. Tribello, Ben Slater, Martijn A. Zwijnenburg and Robert G. Bell, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/b916367k

Investigation of the structures and chemical ordering of small Pd–Au clusters as a function of composition and potential parameterisation
Ramli Ismail and Roy L. Johnston, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004044d

Predicting crystal structures *ab initio*: group 14 nitrides and phosphides
Judy N. Hart, Neil L. Allan and Frederik Claeyssens, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004151c

Zeolitic polyoxometalates metal organic frameworks (Z-POMOF) with imidazole ligands and ε-Keggin ions as building blocks; computational evaluation of hypothetical polymorphs and a synthesis approach
L. Marleny Rodriguez Albelo, A. Rabdel Ruiz-Salvador, Dewi L. Lewis, Ariel Gómez, Pierre Mialane, Jérome Marrot, Anne Dolbecq, Alvaro Sampieri and Caroline Mellot-Draznieks, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004234j

Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials†

Sarah L. Price,*[a] Maurice Leslie,[ab] Gareth W. A. Welch,[a] Matthew Habgood,[a] Louise S. Price,[a] Panagiotis G. Karamertzanis[c] and Graeme M. Day[d]

Received 11th March 2010, Accepted 22nd June 2010
First published as an Advance Article on the web 7th July 2010
DOI: 10.1039/c004164e

Crystal structure prediction for organic molecules requires both the fast assessment of thousands to millions of crystal structures and the greatest possible accuracy in their relative energies. We describe a crystal lattice simulation program, DMACRYS, emphasizing the features that make it suitable for use in crystal structure prediction for pharmaceutical molecules using accurate anisotropic atom–atom model intermolecular potentials based on the theory of intermolecular forces. DMACRYS can optimize the lattice energy of a crystal, calculate the second derivative properties, and reduce the symmetry of the spacegroup to move away from a transition state. The calculated terahertz frequency $k = 0$ rigid-body lattice modes and elastic tensor can be used to estimate free energies. The program uses a distributed multipole electrostatic model (Q_t^a, $t = 00,\ldots,44s$) for the electrostatic fields, and can use anisotropic atom–atom repulsion models, damped isotropic dispersion up to R^{-10}, as well as a range of empirically fitted isotropic exp-6 atom–atom models with different definitions of atomic types. A new feature is that an accurate model for the induction energy contribution to the lattice energy has been implemented that uses atomic anisotropic dipole polarizability models (α_t^a, $t = (10,10)\ldots(11c,11s)$) to evaluate the changes in the molecular charge density induced by the electrostatic field within the crystal. It is demonstrated, using the four polymorphs of the pharmaceutical carbamazepine $C_{15}H_{12}N_2O$, that whilst reproducing crystal structures is relatively easy, calculating the polymorphic energy differences to the accuracy of a few kJ mol^{-1} required for applications is very demanding of assumptions made in the modelling. Thus DMACRYS enables the comparison of both known and hypothetical crystal structures as an aid to the development of pharmaceuticals and other speciality organic materials, and provides a tool to develop the modelling of the intermolecular forces involved in molecular recognition processes.

1. Introduction

The organic solid state is an area of significant academic and industrial interest because of the possibility of designing molecules so that the crystal has desired physical properties, and the need for high quality control in the manufacture of molecular materials. The incidence of polymorphism and multiple solid forms (*e.g.* solvates) for pharmaceuticals is very high, recently estimated at 50% and 90% respectively from the experience within a commercial solid form screening company.[1] However, determining the range of solid forms experimentally is difficult, with many polymorphs being discovered by a range of novel crystallization processes[2] such as crystallization from solvent under pressure.[3] Hence computational modelling has been emerging as an important complementary tool in interdisciplinary research in the organic solid state.[4–6] Methods of organic crystal structure prediction, to allow the design of molecular solids prior to synthesis, can be successful as demonstrated in the blind tests.[7] Using such methods to determine the set of thermodynamically feasible crystal structures, hereafter referred to as the crystal energy landscape, can help guide the search for and characterization of new polymorphs, rationalize disorder and generally provide insights into the range of alternatives to the currently known structures.[8] We have been developing an organic crystal structure modelling program (DMAREL)[9] for almost two decades. The success of crystal modelling studies on fairly small organic molecules has provided the impetus to extend the modelling to a wider range of molecules with more diverse functional groups and conformational flexibility. Experimental

[a] *Department of Chemistry, 20 Gordon Street, London WC1H 0AJ, UK. E-mail: s.l.price@ucl.ac.uk, maurice.leslie@ucl.ac.uk, m.habgood@ucl.ac.uk; Fax: +44 (0)20 7679 7463; Tel: +44 (0)20 7679 4622*
[b] *Formerly of STFC Daresbury Laboratory, Warrington, WA4 4AD*
[c] *Centre for Process Systems Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. E-mail: p.karamertzanis@imperial.ac.uk; Fax: +44 (0)20 7594 6606; Tel: +44 (0)20 7594 6646*
[d] *Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: gmd27@cam.ac.uk; Fax: +44 (0)1223 336362; Tel: +44 (0)1223 336390*

† Electronic supplementary information (ESI) available: An example of distributed polarizabilities and induced moments. The FIT and W99 empirical "repulsion-dispersion" potentials. See DOI: 10.1039/c004164e/

studies seeking to define the diversity of the solid state for a given molecule show that the modelling needs to be able to consider both high symmetry space-groups, and low symmetry triclinic systems with multiple independent molecules in the asymmetric unit cell. In this paper we report a new version, DMACRYS, with capabilities for studying both larger and more complex organic crystals than DMAREL. We describe some features common to both programs that have been used in published crystal structure prediction studies but need explaining in context.

The energy differences between different polymorphs of organic molecules are typically of the order of a few kJ mol^{-1}, and the relative stability can frequently change with temperature, although transitions in the organic solid state are generally first order[10,11] and therefore the metastable phase can persist for a very long time, depending on the sample. Thus, organic crystal structure modelling needs to consider not only the lattice energy (the energy of a static crystal structure at 0 K relative to the infinitely separated molecules in their lowest energy conformation), but also the relative free energies. However, to a good first approximation the relative energies of crystal structures are dominated by the lattice energy differences.[12] The need for very high accuracy in the relative energies of different possible crystal structures has been apparent throughout the development of crystal structure prediction methods.[7] Systematic studies of the crystal energy landscape of glycol and glycerol[13–15] clearly show that the energy of the experimental structures relative to the most stable computer-generated structures improves with the theoretical basis for the crystal energies. The crystal energy is a particular challenge to computational chemistry,[16] as the bonding within the organic crystal ranges from the strong covalent bonds, which are almost identical in all phases, to the weak dispersion forces, which are currently challenging to electronic structure methods.[17] An adaptation of periodic density functional theory electronic structure calculations to include a damped dispersion model which has been empirically fitted to organic crystal structures is proving promising for crystal structure prediction,[18] but is a computationally expensive alternative to atom–atom models. At the other extreme, the traditional force-fields for organic systems are fast, but the simple molecular mechanics description of covalent bonding and the assumption that the intramolecular and intermolecular forces can be modelled by the same atomic charges can lead to significant distortions of the molecular conformation within the crystal[19] or insufficient accuracy in the relative energies.[20,21] An alternative approach is to concentrate on evaluating the intermolecular contribution to the lattice energy, U_{inter}, as accurately as possible using the theory of intermolecular forces to determine both the functional form and parameters of the model intermolecular potential.[22,23] For example, the electrostatic contribution to the lattice energy is calculated from a distributed multipole representation of the molecular charge distribution,[24] resulting in an anisotropic atom–atom intermolecular potential which represents the long-range electrostatic potential arising from lone pair and π electron density.[25] A new development in DMACRYS is to enable the investigation of lifting the approximation of using the gas phase molecular charge density in the crystal lattice by allowing the electron charge density to be polarized by the surrounding molecules. It has been shown that this induction contribution to the energy can be significant, both by using large clusters to model the organic crystal[26] and by computing the intermolecular energies in the crystal from the discretized electron density of the isolated molecules.[27]

Although DMACRYS only models rigid molecules, it can be coupled with *ab initio* single molecule calculations to model the distortion of the molecular conformation by the intermolecular packing forces.[28,29] Hence, the lattice energy, the energy of the static crystal relative to the infinitely separated molecules, $E_{latt} = U_{inter} + \Delta E_{intra}$, can be calculated for conformational polymorphs as a sum of the intermolecular lattice energy, U_{inter}, obtained from DMACRYS and the conformational energy penalty, ΔE_{intra}, obtained from an *ab initio* program. This can be used to refine the energies of the most plausible crystal structures generated in a multi-stage search process,[30] which relies on the ability of DMACRYS to efficiently lattice energy minimize hundreds of thousands of trial crystal structures generated by search programs such as Crystal Predictor.[31,32]

The distinctive feature of DMACRYS is that it uses anisotropic atom–atom model intermolecular potentials, and can therefore accurately reflect the effect of anisotropy in the valence electron distribution on the intermolecular forces. The parameters for the long range forces, such as distributed multipoles, polarizabilities and dispersion coefficients, can be derived from the charge distribution of the isolated molecule.[23] These parameters are defined relative to a molecule-fixed axis system so each atom–atom contribution to the intermolecular pair potential depends not only on the separation of the atoms, but also their relative orientation. The energy, non-central force and torque from each site–site interaction can be effectively expressed[9,22] in terms of the direction cosines between unit atom–atom and molecule-fixed axis vectors. The use of a distributed multipole model for the electrostatic contribution to the lattice energy has been well established using DMAREL, and shown to give better reproduction than atomic partial charge models of the crystal structures of many rigid molecules and the relative energies of their polymorphs or predicted crystal structures.[19,33] The non-empirical parameterization of the other contributions to the intermolecular potential is less straightforward, but considerable progress has been made in recent years,[23] and it is now possible to obtain realistic distributed polarizabilities and dispersion coefficients from reasonable quality monomer wavefunctions for small organic molecules. Methods of deriving anisotropic site–site models for the short range forces of pairs of molecules are also being developed, using the overlap model and a limited number of Symmetry Adapted Perturbation Theory[34] evaluations of these energies. Such work has lead to a completely non-empirical intermolecular potential[35] for $C_6Br_2ClH_2F$ successfully predicting its crystal structure in the last blind test.[7] The anisotropy in the repulsive wall around halogen atoms had already been shown to improve the modelling of the chlorobenzenes' crystal structures and properties.[36] The most recent development is the ability to obtain distributed anisotropic polarizability tensors[37–39] from

ab initio monomer calculations. These can be used to determine the long range dispersion[40] in atom–atom form. The atomic polarizabilities can also be used to model the induced atomic multipole moments arising from the electrostatic field of neighbouring molecules. The implementation of atomic dipolar polarizability tensors to model the induction energy contribution to the lattice energy is therefore required in order to be able to use completely non-empirical model intermolecular potentials in crystal structure modelling.

This paper describes the implementation of anisotropic atom–atom model intermolecular potentials to model crystals of rigid polyatomic molecules within DMACRYS, with particular emphasis on aspects of organic polymorph modelling that may not be familiar to those with expertise in the theory of intermolecular forces and *vice versa*. The most significant difference from the most commonly used bio-organic force-fields is the anisotropy of the atom–atom interactions. The methods of parameterization of the anisotropic atom–atom potential have evolved from those used to define the potential energy surface for pairs of small polyatomic molecules.[22,23] Hence, application to organic molecules differs in the lower molecular symmetry and greater conformational flexibility. Thus, for DMACRYS the intermolecular potential parameters are defined relative to molecule-fixed axes defined by the covalent bonding. The paper first describes the construction of the model crystal and defines the models for the atom–atom intermolecular potential contribution. The evaluation of the lattice energy is described, again with the emphasis on the additional challenges in applying these model intermolecular potentials with anisotropic long-range interactions to infinite perfect crystals. This is followed by notes on the optimization of the lattice energy, the use of symmetry and second derivative properties to ensure a stable minimum, and application of pressure and modelling of flexible molecules. Some of the capabilities of DMACRYS are illustrated by an application to the four polymorphs of the anti-epileptic carbamazepine, showing the sensitivity to various approximations in modelling both inter- and intra-molecular forces. The symmetry reduction capabilities are illustrated by an example taken from crystal structure prediction calculations on adenine. Finally, some initial results for the induction energy are provided, as a first stage toward the use of non-empirical potentials in crystal structure modelling. Overall, this paper describes the development of the DMACRYS modelling suite for two closely related types of research: the testing and development of more theoretically based and parameterized model intermolecular potentials for solid state properties, and the application of the most appropriate models for calculating the crystal energy landscapes and comparing polymorphs of pharmaceutical or other speciality organic molecules.

2. Method

2.1 Input description

The utility program NEIGHCRYS (an updated version of NEIGHBOURS, that was used with DMAREL) is used to construct a DMACRYS input file. The crystal structure, defined by cell parameters, symmetry operations and fractional atomic coordinates, is input into NEIGHCRYS,‡ and the cell is converted into an orthonormal Cartesian axis system using the same convention used to define the elastic tensor.[41] In most cases, this directs the unit length Z axis along the *c* axis, the X axis parallel to the reciprocal *a* axis (**a***), and the Y axis defined to give a right handed orthogonal set of axes (Fig. 1a). This global axis system is internal to the program, so the user only needs to be aware of it as the definition of the axes for the elastic tensor.

A set of maximum atom–atom covalent bond distances is also provided as input to calculate the bond connectivity that defines the rigid molecules. Atomic types that determine repulsion-dispersion parameters are automatically determined by NEIGHCRYS and can be defined solely by element, as in the simplest isotropic repulsion-dispersion potentials (*e.g.* non-hydrogen atom parameters in the FIT parameter set)[42,43] or by element and hybridization, as in the atomic typing of the most recent Williams potential (W99).[44] Hydrogen atom typing must almost always consider bonding environment, such that the parameters describing hydrogen bond donor hydrogen atoms are of different types to those bonded to carbon. The W99 potential also defines the interaction centre for hydrogen atoms to be shifted away from the nuclear site and NEIGHCRYS automates the required adjustment of the hydrogen atom interaction sites to a position 0.1 Å closer to X along the X–H bond. The prior adjustment of the hydrogen atom positions to the average values determined by neutron diffraction[45] is an option which is required for all X-ray structures to correct the systematic error in location of hydrogen atoms. It is also possible to input the user's choice of atomic types by an input file linking each atom label to its atomic type. For example, a non-empirical potential specific to pyridine could have up to three types each for carbon and hydrogen atoms and one for nitrogen, whereas modelling an organic hydrate may require different potential parameters for the hydrogen atoms in water, alcohol and carboxylic acid functional groups.

The Cartesian molecule-fixed axes (x_M, y_M, z_M in Fig. 1) are defined by three atoms in the molecule, two defining an axis direction and the third defining a plane, using the right hand coordinate system. These atoms are usually chosen to reflect any molecular symmetry and so that the axes are (approximately) aligned with the axes of inertia (*cf.* Fig. 1b). The origin of the molecular axis system is at the centre of mass. The molecular coordinates in this molecule-fixed axis system are used to calculate the molecular charge density and hence define the distributed multipole moments and polarizabilities. Since many organic molecules have no symmetry, resulting in no exact relationships between the distributed multipoles on different atoms, the distributed multipoles and polarizabilities can be viewed as atom specific parameters. Single atoms, such as chloride ions (Cl^-), can also be handled.

‡ The crystal structure is read by NEIGHCRYS in Shellx (.res) file format. This file can be derived from the .cif file format using various crystallographic programs such as Mercury,[95] and these should be used to produce a .res file containing the coordinates for entire molecules. NEIGHCRYS needs input for every atom within the molecule and so does not handle all inputs where there is less than one molecule in the asymmetric unit cell, *i.e.* $Z' < 1$.

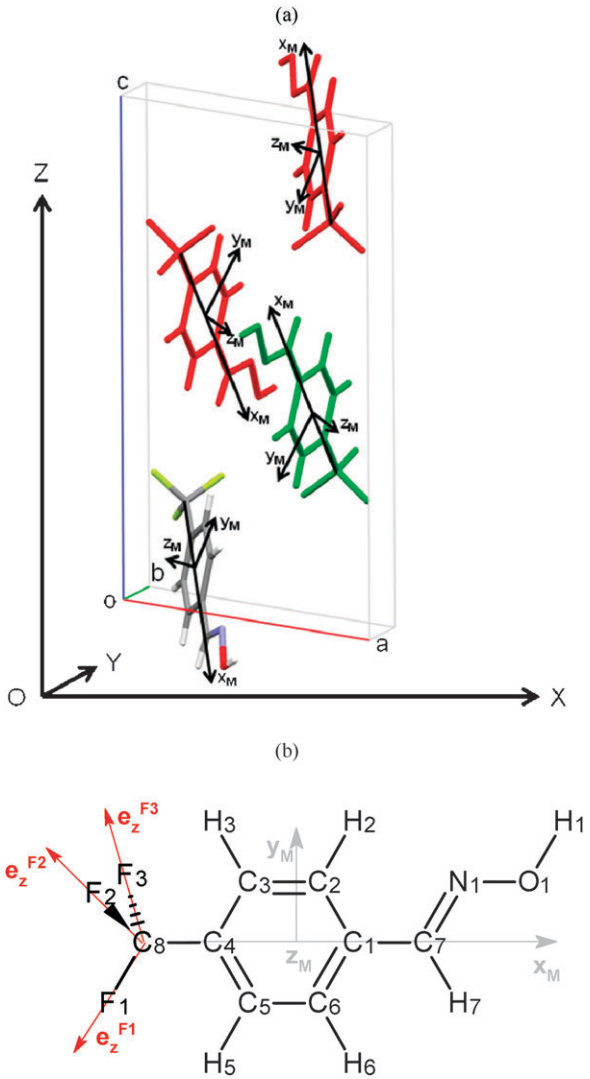


Fig. 1 (a) Illustration of the various axis systems for the $P2_1/c$ structure of (E)-4-(Trifluoromethyl)benzaldehyde oxime.[46] The Cartesian global axis system is related to the crystallographic axis system so **Z** corresponds to **c**, **X** is parallel to **bxc** (and is not along **a** as $\beta = 99.3°$) and **Y** completes a right handed orthonormal axis system (for this monoclinic cell, **Y** is parallel to **b**). The molecule in the input asymmetric unit cell is coloured by element, the molecule related by a 2-fold screw axis with identical multipole moments in dark green and the molecules related by an inversion centre or glide plane which have symmetry-related multipole moments in red. (b) The atomic numbering diagram showing the right-handed set of molecule-fixed axes ($\boldsymbol{x}_M$, $\boldsymbol{y}_M$, $\boldsymbol{z}_M$) defined by C4, C1 and C2, and, only for the three fluorine atoms, the local atomic axis vectors $\mathbf{e}_z$ along the C–F bonds which are defined to allow transferability of the anisotropic atom–atom repulsion model.

2.2 Defining the model intermolecular potential

An initial run of NEIGHCRYS is necessary to obtain the molecular structure in the molecular axis system, to be used to calculate the distributed multipoles and polarizabilities. The multipole moments at any site, a, are assumed to conform to the spherical tensor definition,[22] and can be input up to hexadecapole, *i.e.* generally for each atom there is a set Q_t^a of 25 multipole moment components, $t = 00$ for charge, $t = 10, 11c, 11s$ for the dipole, $t = 20, 21c, 21s, 22c, 22s$ for quadrupole, $t = 30, 31c, 31s, 32c, 32s, 33c, 33s$ for octopole, $t = 40, 41c, 41s, 42c, 42s, 43c, 43s, 44c, 44s$ for hexadecapole. In this paper, all sets of multipole moments are derived by the form of distributed multipole analysis appropriate to larger basis sets,[47] where the contribution to atomic multipoles from more diffuse basis functions is determined by integration,§ using the program GDMA2[48] to analyze the density matrix (.Fchk) file produced by GAUSSIAN[49] for the molecule in the molecular axis system. Other methods of deriving the atomic multipole moments from any charge density can be used, provided that they conform to these spherical tensor conventions.[22] Similarly, DMACRYS handles localized atomic dipolar polarizabilities α_t^a with components $t = (10,10), (11c,11c), (11s,11s), (10,11c), (10,11s), (11c,11s)$, defined using spherical tensors. The distributed polarizabilities are calculated by the WSM method[37] as implemented in CamCASP.[50] An analysis of the importance of induction in organic crystal structures,[26] based on large cluster representations of the crystals, suggested that modelling just the induced dipoles would be a reasonable starting point for a more thorough investigation of the effects of explicitly modelling the polarization within organic crystal structures.

The majority of organic crystal structures contain molecules related by symmetry elements in addition to translation. Mirror planes, glide planes and inversion centres will generate half the molecules in the unit cell with a molecular structure that is inverted relative to the input molecule (although rotation axes and screw axes will not). These are racemic crystal structures for molecules with chiral centres.[51] Since a right handed axis system is required in the evaluation of the torques *etc.*, NEIGHCRYS restores the molecular axis systems of the inverted molecules (Fig. 1) to the right handed set. This is done by separately labelling the inverted molecules (I) and providing the correct atomic multipoles and polarizabilities, by changing the sign of the multipoles Q_t^a and polarizabilities α_t^a whose definition has odd powers of z in the spherical tensor operator (See ESI Table 1†). To emphasize that atoms that are related by crystal symmetry elements have sets of multipole moments and polarizabilities that are symmetry-related but not necessarily identical, the description of these anisotropic terms in the potential will continue to use the superscripts a and b to refer to the particular (atomic) interaction sites on molecules A and B, allowing ease of comparison with the theory of intermolecular forces.[22,23]

The electrostatic energy is then given by

$$E_{\text{electrostatic}} = \frac{1}{2}\sum_A \sum_{B\neq A} Q_t^a T_{tu}^{ab} Q_u^b \qquad (1)$$

where the summation over molecules is specified explicitly, whilst summation over all multipoles t on atomic sites a in molecule A, and all multipoles u on all atoms b in molecule B follows the repeated index convention. The interaction functions T_{tu}^{ab} have been tabulated in terms of the direction cosines of the local axis vectors.[22] The induction energy can be similarly

§ The default atomic radii values used for the integration in GDMA2 are 0.325 Å (0.623 au) for H and 0.650 Å (1.228 au) for all other atoms. The revised radius for hydrogen has been shown to be an improvement for modelling hydrogen bonding from using equal radii.[47]

evaluated in terms of the induced multipole moments ΔQ_t^a and original multipole moments Q_t^a as

$$E_{\text{ind}} = \frac{1}{2}\sum_A \sum_{B\neq A} \Delta Q_t^a T_{tu}^{ab} Q_u^b \qquad (2)$$

once the induced moments have been determined, as described in section 2.3.

Most modelling that has been done with DMACRYS for pharmaceuticals and organic molecules assumes a transferable isotropic atom–atom model for the non-electrostatic interaction between each pair of molecules, which, when applied in crystal structure modelling, is usually expressed in the form:

$$U_{\text{rep-disp}}^{MN} = \sum_{i\in M,k\in N} A_{\iota\kappa}\exp(-B_{\iota\kappa}R_{ik}) - C_{\iota\kappa}/R_{ik}^6 \qquad (3)$$

Here the repulsion and dispersion interactions are between atom i of type ι in molecule M and atom k of type κ in molecule N, which are separated by a distance R_{ik}. (This notation is used for the atoms and parameters for all transferable sets of parameters.¶) The two sets of parameters, FIT[42,43] and W99,[44] which are frequently used in DMACRYS studies as they have been empirically fitted to organic crystal structures, are discussed further in the ESI.†

In contrast to the usual assumption (eqn (3)) that the atom–atom repulsion is isotropic, DMACRYS can model anisotropy in the repulsion terms to represent polar flattening.[52] A local z axis unit vector, $\mathbf{e}_z$, is defined at each atomic site (Fig. 1b) to allow atom–atom anisotropic repulsion potentials to be transferred directly to atoms of the same atomic type, avoiding the need for analytic rotation to the molecular axis system. The short range repulsive potential that DMACRYS can currently use is of the form:

$$U_{\text{rep}}^{MN} = \sum_{i\in M,k\in N} G\exp[-B_{\iota\kappa}(R_{ik} - \rho^{\iota\kappa}(\Omega_{ik}))] \qquad (4)$$

where G is a constant energy unit, so that the repulsion between atoms i and k is equal to G when $R_{ik} = \rho^{\iota k}(\Omega_{ik})$. In addition, it is assumed that the anisotropy associated with a given atom (i) can be modelled with cylindrical symmetry about the local z axis of the atom which is usually along a covalent bond (*cf.* C–F bonds in Fig. 1b), and does not depend on the type (κ) of the other atom (k) *i.e.* the constant repulsion contour can be represented by

$$\rho^{\iota\kappa}(\Omega_{ik}) = \rho_0^{\iota\kappa} + \rho_1^{\iota}(\mathbf{e}_\mathbf{z}^i\cdot\mathbf{e}_{ik}) + \rho_1^{\kappa}(-\mathbf{e}_\mathbf{z}^k\cdot\mathbf{e}_{ik}) + \rho_2^{\iota}(3[\mathbf{e}_\mathbf{z}^i\cdot\mathbf{e}_{ik}]^2 - 1)/2 + \rho_2^{\kappa}(3[\mathbf{e}_\mathbf{z}^k\cdot\mathbf{e}_{ik}]^2 - 1)/2 \qquad (5)$$

where the vector $R_{ik}\mathbf{e}_{ik}$ is from atom i to atom k. Various studies have shown that this model is sufficient to model the anisotropy of repulsion in cyanuric chloride,[52] chlorothalonil[53] and $C_6Br_2H_2ClF$,[35] and provide a transferable scheme for the chlorobenzenes[36] and polycyclic aromatic hydrocarbons.[54] Since the range of orientations around each atom that can be in repulsive intermolecular contact is fairly restricted for organic molecules, compared with, for example, a diatomic molecule, more complex variations from spherical symmetry are unlikely to be required.

Distributed dispersion coefficients can also be generated using CamCASP, but following investigations of various anisotropic dispersion models[40] it seemed that the added complexity of anisotropy in the atom–atom dispersion was unlikely to be important for organic molecules. Hence the dispersion contribution is modelled as a damped isotropic C_6, C_8, C_{10} dispersion model

$$U_{\text{disp}}^{MN} = \sum_{i\in M,k\in N} -f_6(\beta R_{ik})\frac{C_6^{\iota\kappa}}{R_{ik}^6} - f_8(\beta R_{ik})\frac{C_8^{\iota\kappa}}{R_{ik}^8} - f_{10}(\beta R_{ik})\frac{C_{10}^{\iota\kappa}}{R_{ik}^{10}} \qquad (6)$$

where $f_n = 1$ is the default undamped dispersion, and the alternative Tang-Toennies damping function[38] is given by

$$f_n(\beta R_{ik}) = 1 - \left(\sum_{m=0}^{2n}\frac{(\beta R_{ik})^m}{m!}\right)\exp(-\beta R_{ik}) \qquad (7)$$

Currently, this damping function, if applied, is used for all atom–atom dispersion terms, with one parameter β defined for the molecule. This can be estimated by $\beta = 2\sqrt{2I_X}$, where I_X is the first vertical ionization potential of the molecule in atomic units,[37] as for small polyatomics.[38] However, it is likely that atom-dependent damping parameters may be needed for molecules containing atoms other than C, N, O and H, and recent work suggests that even for small molecules the vertical ionization energy may not give the optimal value of β for damping the induction.[55]

The new atom naming system implemented in NEIGHCRYS/DMACRYS is a marked improvement on that in NEIGHBOURS/DMAREL as it allows the easy implementation of either standard (FIT or W99) or user defined atomic typing, as well as allowing up to 99 999 atoms in the unit cell. DMAREL was limited to 62 atoms of an atomic type in the asymmetric unit, a limit that is too small for the number of carbon and hydrogen atoms in some organic crystals. It also allows supercells to be generated for defect and impurity calculations.

2.3 Lattice energy calculation

The method of calculating and summing the site–site intermolecular energies, and transforming the associated forces and torques to the net force and torque on the rigid molecule, is as described[9] for DMAREL. The repulsion and dispersion terms are calculated from eqn (3) or from eqn (4)–(6), and summed over all intermolecular atom–atom distances up to a defined cutoff. This is typically 15 Å for small organic molecules for fast calculations and 30 Å when highly smooth numerical derivatives are required or for particularly large molecules.

The electrostatic energy, and corresponding rigid body forces and torques, is evaluated from the molecule-fixed and site–site intermolecular vectors[9] using eqn (1). The charge–charge (R_{ab}^{-1}), charge–dipole (R_{ab}^{-2}) and dipole–dipole (R_{ab}^{-3}) terms in the electrostatic and induction contributions

¶ Unfortunately the notation for sites a and b separated by R_{ab}, traditionally used to describe an intermolecular pair potential, leads to confusion with cell axes and, if α and β are used for types, with polarizability tensors.

to the lattice energy are summed by Ewald summation.‖ All the other multipole terms up to $R_{ab}{}^{-5}$ are summed in direct space, using all atom–atom distances between molecules in a list, defined initially and updated as necessary as those whose centres of mass lie within a predefined cutoff. This is typically 15 Å but needs to be greater for very large molecules or high accuracy.

The induced multipoles can be evaluated by using the same interaction functions T_{tu}^{ab} as the field at the nucleus of atom a in molecule A is given by

$$V_t^a = \sum_{B \neq A} T_{tu}^{ab} Q_u^b \tag{8}$$

The hydrogen bonds and other close contacts in organic crystals involve sufficient overlap of the molecular charge distribution that the field arising from the point multipole description will generally need to be damped.[26] Damping by the Tang-Toennies form (eqn (7)) is easily applied to the fields that give rise to the induced dipoles, and the sum becomes

$$V_t^a = \sum_{B \neq A} f_{u+1}(\beta R_{ab}) T_{tu}^{ab} Q_u^b \tag{9}$$

to provide an array of each damped field component t for which there is a dipolar polarizability tensor for each atom a. The field is calculated by direct summation as the use of Ewald summation for fields would be unnecessarily cumbersome. It has been established that the induced dipole moments differ insignificantly whether the field is calculated using Ewald summation techniques or by direct summation to 15 Å in the absence of damping. The induced moment is the product of the electric field components with their corresponding polarizabilities

$$\Delta Q_t^a = \alpha_t^a V_t^a \tag{10}$$

The induced dipole moment components are multiplied by the original electrostatic field components, and summed to give the first order induction energy

$$E_{\text{ind}}^A = \frac{1}{2} \sum_{a \in A} \sum_t \Delta Q_t^a V_t^a \tag{11}$$

Thus in this field the original multipole moment Q_t^a changes by ΔQ_t^a, as determined by the competition between the positive energy required to distort the molecule's charge distribution from its equilibrium in zero field and the lowering of the energy of interaction with the field.[22] The induced dipole moments are then added to the original static dipole moments and used to calculate the new field.

$$\Delta Q_t^a = \alpha_{tv}^a \sum_{B \neq A} f_{u+1}(\beta R_{ab}) T_{vu}^{ab} (Q_u^b + \Delta Q_u^b) \tag{12}$$

Eqn (12) needs to be solved iteratively, reflecting that the polarizing field depends on the polarized moments of the other molecules. The process is iterated to consistency between the total moments on the molecules within the crystal, defined by the change in the induction energy between two cycles being less than a set tolerance δ_{induct}. Induction calculations are currently being used to investigate the polarization within fixed crystal structures and to estimate the induction energy: the demands on convergence for the numerical derivatives of the induction energy means that refining crystal structures including the induction energy would only be performed as a final refinement of the crystal structure and lattice energy.** Hence, the contribution to the forces, torques and second derivatives from the induction energy is omitted from the description of the lattice energy minimization and second derivative analysis described in the next section.

2.4 Lattice energy minimization and Hessian calculation for properties

The non-central forces, torques and second derivatives arising from the anisotropic atom–atom intermolecular potential are transferred to the centre of mass of each molecule and used to determine the strains on the rigid molecules within the crystal lattice.[9] In general, the initial crystal structure used as a starting point for lattice energy minimization is not a crystal in internal equilibrium, and is modelled as a crystal subject to a homogenous deformation described in terms of bulk and internal strains. The bulk strain for the deformation of the cell is a symmetric second rank tensor, $\boldsymbol{\varepsilon}$. The internal strains, which do not change the volume of the crystal, are expressed in terms of the centre of mass vector for each molecule and a vector description of the rotation of the molecule about the centre of mass, with the direction defined by a right hand screw.[9] Hence, changes in the crystal structure can be expressed as a vector $\boldsymbol{\delta}$ whose components are the three translation and three rotation vector components per molecule in the unit cell and six strain matrix elements. The intermolecular lattice energy as a function of a small change in structure ($\mathbf{r}$) can be written concisely as a power series

$$U_{\text{inter}}(\mathbf{r}') = \mathbf{U}_{\text{inter}}(\mathbf{r}) + \boldsymbol{\delta}^{\mathbf{T}} \cdot \mathbf{g} + \frac{1}{2} \boldsymbol{\delta}^T \cdot \mathbf{W} \cdot \boldsymbol{\delta} \tag{13}$$

where $\mathbf{g}$ is a vector of first derivatives, and $\mathbf{W}$ a matrix of second derivatives.[56]

There are various relationships between the second derivatives which mean that the majority can be evaluated analytically,[9] though only the analytic form for all first derivatives and those second derivatives involving two translation, two rotation or two strain variables, have been programmed into DMACRYS.

For rapid minimizations, all of the gradients and the block diagonal second derivatives are calculated analytically[9] and the elements of $\mathbf{W}$ for which there are no analytical formulae (the cross terms between the translation, rotation and strain variables) are initially set equal to zero. From the calculated $\mathbf{W}$ and $\mathbf{g}$, the displacement from equilibrium can be estimated by

‖ For polar crystals, the electrostatic energy in principle depends on the external shape of the crystal. DMACRYS provides the net dipole per unit cell so that such corrections can be estimated for different morphologies,[96] although this contribution is usually omitted on the grounds that external charges will accumulate on the surface of the crystal to annul the surface charge.

** The iterative induction energy evaluation requires all the derivatives of the induction energy with structural changes to be evaluated numerically. This requires the tolerance δ_{induct} to be around 10^{-8} kJ mol^{-1}. The option of spline smoothing is being implemented to provide smooth numerical derivatives in minimizations, which would then allow minimizations which include the induction energy.

$\boldsymbol{\delta} = -\mathbf{W}^{-1}\cdot\mathbf{g}$, allowing minimization of the lattice energy by a modified Newton-Raphson scheme which updates the matrix $\mathbf{W}^{-1}$, estimating the missing second derivatives. Hence the expensive calculation of $\mathbf{W}$ is carried out only once in the course of a typical energy minimization; recomputation of $\mathbf{W}$ is sometimes required when difficulties are encountered in locating a stationary point. Once the lattice is at equilibrium ($\mathbf{g} = 0$), the exact Hessian matrix, (strictly[9] $(\mathbf{W} + \mathbf{W}^{\mathbf{T}})/2$) with the second derivatives calculated numerically from the analytical gradients, can be used to characterize the nature of the stationary point and calculate the second derivative properties.

A significant advance on the original version of DMAREL is the partial adaptation to use group theory to exploit the symmetry relationship between the molecules in the unit cell. Standard projection operator techniques are used[56] to construct a matrix $\mathbf{P}$ whose columns transform according to the representations of the space group. A column of $\mathbf{P}$ has components for all molecules which are equivalent under the operation of one of the space group operators and defines how the translational and rotational motions are related to each other according to the particular representation of the group. $\mathbf{P}$ also has components which define how the strain tensor transforms. A symmetry adapted Hessian matrix and gradient vector are then used to calculate a symmetry adapted displacement vector (which usually contains strain components describing the change in the lattice vectors). For the relaxation phase of the calculation only the totally symmetric representation of the group needs to be considered, although for calculating the second derivative properties of the relaxed structure all representations are needed. The eigenvalues of the final symmetry adapted Hessian matrix $\mathbf{P}^{T}\cdot\mathbf{W}^{-1}\cdot\mathbf{P}$ are calculated for all representations of the group to look for any negative eigenvalues which would indicate that the structure has converged to a saddle point between lower symmetry structures. This can be done when the Newton-Raphson procedure has been used to provide an updated $\mathbf{W}^{-1}$, as this is a sufficiently good approximation that we have very rarely found cases where a negative eigenvalue becomes positive when the matrix is calculated exactly. By removing the symmetry element corresponding to the negative eigenvalue, and reminimizing within the sub-group of the original space-group, a lower energy structure which is a genuine lattice energy minimum will be found. This can be done by specifying which irreducible representation is associated with the negative eigenvalue in a further run of NEIGHCRYS, which then generates the new input files for minimization within the required subgroup, *i.e.* sets up the files including the independent labels for the second molecule in the asymmetric unit cell, for the FIT and W99 potentials and isotropic custom potentials. The input file also includes a directive to shift the structure off the saddle point by performing a line search in the direction of the eigenvector associated with the negative eigenvalue. Lattice energy minimization is then restarted from the energy minimum along this line search.

When the eigenvalues of the Hessian characterize the stress-free structure as a true minimum on the rigid molecule crystal potential energy surface, the forces, rigid molecule torques and strain derivatives of the unit cell are evaluated analytically, and the elastic constants of the crystal derived from the numerical second derivatives.[41,57,58] The second derivative matrix can also be used to calculate the Γ-point ($\mathbf{k} = 0$) rigid molecule phonon frequencies[59] for the crystal, for comparison with the Raman, far-infra-red, terahertz[60,61] or inelastic neutron scattering spectra. This requires the rotational components of the dynamical matrix to be referred to the principal inertial axes. Hence, the molecule-fixed axes and corresponding analytical gradients are transformed into the principal inertial axis system before numerical differencing to give the second derivatives appearing in the dynamical matrix (splitting of the transverse and longitudinal zone-centre modes is ignored). The eigenvectors corresponding to the harmonic frequencies are also calculated and can be used to visualize the molecular motions associated with each vibrational mode.[62] Computation of lattice modes at general $\mathbf{k}$-points does not seem worthwhile given the intrinsic approximations of perfect crystal rigid-molecule harmonic modes in comparison with the anharmonic modes resulting from a Molecular Dynamics simulation,[63] let alone the mixing of inter- and intramolecular dynamics in a real organic crystal. The use of Ewald summation for charge–dipole and dipole–dipole interactions means that calculating the dynamical matrix for $\mathbf{k} \neq 0$ would be not be a trivial extension.

2.5 Pressure

DMACRYS also allows the study of the effect of pressure on organic crystal structures, by adding the isotropic pressure term PV and optimizing the lattice enthalpy $H_{\mathrm{latt}} = U_{\mathrm{inter}} + \mathrm{PV}$. The first derivatives of the volume with respect to the symmetric strain tensor are calculated analytically. Pressures of a few GPa, as now being used in experimental searches for new polymorphs,[3] can have a significant effect on the crystal energy landscape, reordering the stability of the structures and adding some new structures which are only distinct lattice energy minima at finite pressure.

2.6 Extension of DMACRYS to flexible molecules

Most organic molecules exhibit some degree of conformational flexibility, so a procedure has been developed[28,29] for coupling DMACRYS with GAUSSIAN and GDMA analysis for refining rigid body lattice energy minima to allow small conformational distortions. This enables the minimization of $E_{\mathrm{latt}} = U_{\mathrm{inter}} + \Delta E_{\mathrm{intra}}$, which allows conformational changes in specified intra-molecular degrees of freedom, such as torsion angles, to improve hydrogen bonding geometries. This approach is effective since the conformation observed in most crystal structures is fairly close to a gas phase conformational minimum[64] with $\Delta E_{\mathrm{intra}}$ rarely greater than about 5 kJ mol^{-1}. (A notable exception is when the gas phase conformation has an internal hydrogen bond, which is replaced by an intermolecular hydrogen bond in the gas phase[65]). The computational cost of the *ab initio* intramolecular energies and forces is orders of magnitude larger than the DMACRYS lattice energy optimizations for each conformation. The CrystalOptimizer[22,29] suite of programs allows these calculations to performed efficiently, for example, by building up databases of multipoles and intramolecular energies, gradients and Hessians as a

function of conformation that can be carried over through a series of crystal structures of the same molecule.

3. Results: example calculations

3.1 Relative stability of carbamazepine polymorphs

As an illustration of the utility of DMACRYS for polymorphic pharmaceuticals, we contrast various models for the lattice energy of four polymorphs of the anti-epileptic carbamazepine ($C_{15}H_{12}N_2O$) (Fig. 2). Form I has four independent molecules in the asymmetric unit cell of $P\bar{1}$ symmetry, and is crystallized from the melt. The only single crystal determination is at 158 K[66] (refcode CBMZPN11 on the Cambridge Structural Database[73]). Form II is a high symmetry form, with 18 molecules in the $R\bar{3}$ unit cell, which was recently shown to sometimes contain solvent in the channels.[74] In order to use a reasonable quality crystal structure in which the hydrogen atoms have been located,[67] we have taken the carbamazepine (0.1) tetrahydrofuran solvate, MIMQIJ, with the solvent removed, as the starting point for Form II calculations. Form III is the most thermodynamically stable form and adopts the more typical symmetry $P2_1/n$, $Z' = 1$ and we used the 295 K single crystal structure,[75] CBMZPN10. Finally, polymorph IV is also $Z' = 1$ but in a centred monoclinic $C2/c$ spacegroup[70] (CBMZPN12) and has been crystallized in the presence of a polymer template or from the melt, but has not been obtained by solvent crystallization.[76] This system was chosen as having a large number of well-studied polymorphs, with a range of unit cell sizes and symmetries.

In Table 1, the cell parameters and energies for the lattice energy minima for the four polymorphs of carbamazepine are compared, for variations in the intermolecular potential and the treatment of molecular flexibility. The relative stabilities are compared in Fig. 3. The basic model used the experimental molecular conformation with the bonds to hydrogen atoms elongated to neutron values, the FIT intermolecular potential and distributed multipoles derived from the MP2/6-31G(d,p) charge density. Changing the repulsion-dispersion potential to the W99 parameters gives a marginally better fit to the experimental structures, as judged by the minimum root mean square overlay of all the non-hydrogen atoms in a 15 molecule coordination cluster,[77] $RMSD_{15}$. However, the differences are marginal if you consider the percentage errors in the cell dimensions compared with the likely thermal expansion. Similarly, replacing the molecule with the *ab initio* optimized structure makes quite a difference to the lattice energy minima, as the change to a more pyramidal NH_2 group results in some molecular reorientation to improve the hydrogen bond geometries. The optimization of the conformation within the crystals, using CrystalOptimizer, returns the amides to a planar conformation, as shown by the overlay of the modelled and experimental crystal structures in Fig. 2. This sensitivity to the modelling of NH_2 flexibility is a major problem in organic crystal modelling, and has been explicitly investigated for carbamazepine.[78]

Relative lattice energies are much more sensitive than the crystal structure to the computational model used, as shown in Fig. 3. The experimental order of stability from differential scanning calorimetry measurements,[66]

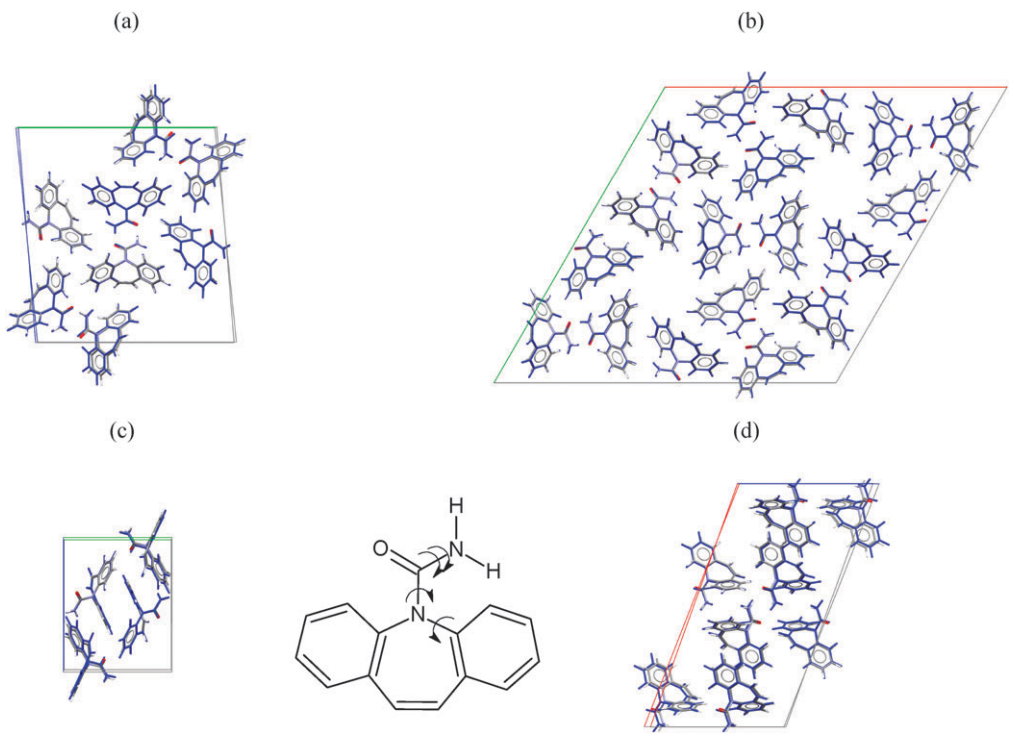


Fig. 2 The four polymorphs of carbamazepine, showing an overlay of the unit cell of each experimental crystal structure (molecules coloured by element) of carbamazepine with the model produced by relaxing the marked torsion angles within the lattice energy minimization of the crystal structure (blue, FIT relax model defined in Table 1) (a) form I, (b) form II, (c) form III, (d) form IV. The **a**-axis is in red, the **b**-axis in green and the **c**-axis in blue.

Table 1 Lattice energy minima for the polymorphs of carbamazepine for variations in the intermolecular potential and treatment of molecular flexibility

Structure	Cell parameters a/Å	b/Å	c/Å	α (°)	β (°)	γ (°)	Density/g cm^{-3}	F^a	RMSD$_{15}$/Å	U_{inter}/ kJ mol^{-1}	ΔE_{intra}/ kJ mol^{-1}	E_{latt}/ kJ mol^{-1}	Thermal energy^b/ kJ mol^{-1}	ZPE^b/ kJ mol^{-1}	Free energy^c/ kJ mol^{-1}
Form I	5.171	20.574	22.245	84.12	88.01	85.19	1.34	—	—	—	—	—	—	—	—
CBMZPN11,[66] $P\bar{1}$, $Z' = 4$, $Z = 8$, $T = 158$ K, $R = 5.06\%$															
FIT Exp[d]	5.223	20.633	22.390	84.51	87.45	85.26	1.31	4.7	0.2	—	—	−134.68	−20.65	2.42	−152.91
WILL Exp[e]	5.237	20.666	22.326	84.45	87.63	85.25	1.31	4.1	0.1	—	—	−126.94	−21.21	2.30	−145.85
FIT Opt[f]	5.030	21.248	23.066	85.03	86.56	89.40	1.28	110.2	0.4	−117.54	0	−117.54	−22.13	2.18	−137.49
FIT relax[g]	5.262	20.517	22.365	85.16	86.36	85.93	1.31	n/a	0.2	−127.44	2.64	−124.80	−21.11	2.34	−143.57
Form II	35.243	—	5.185	90	90	120	1.27	—	—	—	—	—	—	—	—
MIMQIJ[67] $R\bar{3}$, $Z' = 1$, $Z = 18$, $T = 100$ K, $R = 5.12\%$ CBMZPN03,[68] room temperature, $R = 6.9\%$ $a = 35.454$ Å $c = 5.253$ Å															
FIT Exp[d]	35.571	—	5.241	90	90	120	1.23	6.3	0.1	—	—	−127.97	−20.46	2.39	−146.04
WILL Exp[e]	35.533	—	5.259	90	90	120	1.23	5.1	0.2	—	—	−119.24	−21.26	2.24	−138.26
FIT Opt[f]	36.524	—	4.996	90	90	120	1.22	62.6	0.4	−115.31	0	−115.31	−21.54	2.22	−134.63
FIT relax[g]	35.264	—	5.272	90	90	120	1.24	n/a	0.1	−125.34	3.37	−121.97	−20.57	2.35	−140.19
Form III	7.537	11.156	13.912	90	92.86	90	1.34	—	—	—	—	—	—	—	—
CBMZPN10,[69] $P2_1/n$, $Z' = 1$, $Z = 4$, room temperature, $R = 3.9\%$															
FIT Exp[d]	7.695	11.034	13.623	90	92.01	90	1.36	16.4	0.2	—	—	−135.33	−18.28	2.87	−150.74
WILL Exp[e]	7.719	11.065	13.627	90	92.29	90	1.35	16.5	0.2	—	—	−126.18	−18.84	2.72	−142.30
FIT Opt[f]	7.867	11.044	13.649	90	92.30	90	1.32	32.7	0.3	−124.07	0	−124.07	−19.10	2.69	−140.48
FIT relax[g]	7.885	11.018	13.427	90	92.29	90	1.35	n/a	0.3	−130.27	1.51	−128.76	−18.74	2.76	−144.74
Form IV	26.609	6.9269	13.957	90	109.70	90	1.30	—	—	—	—	—	—	—	—
CBMZPN12,[70] $C2/c$, $Z' = 1$, $Z = 8$, $T = 158$ K, $R = 3.57\%$															
FIT Exp[d]	26.676	6.889	14.305	90	110.37	90	1.27	8.1	0.1	—	—	−130.90	−20.27	2.54	−148.63
WILL Exp[e]	26.722	6.891	14.187	90	110.00	90	1.28	3.8	0.1	—	—	−123.78	−20.30	2.50	−141.58
FIT Opt[f]	26.435	7.157	14.547	90	110.79	90	1.22	43.4	0.3	−116.72	0	−116.72	−21.98	2.31	−136.38
FIT relax[g]	26.856	6.916	14.413	90	111.39	90	1.26	n/a	0.2	−124.98	1.73	−123.25	−20.49	2.48	−141.26

[a] $F = (\Delta\theta/2)^2 + (10\Delta x)^2 + (100\Delta a/a)^2 + (100\Delta b/b)^2 + (100\Delta c/c)^2 + \Delta\alpha^2 + \Delta\beta^2 + \Delta\gamma^2$ ("Figure of Shame" defined for rigid molecule minimizations).[71] [b] The vibrational contribution to the Helmholtz free energy[72] is calculated using the Einstein approximation of no **k**-dependence of optical mode frequencies and the Debye approximation for the acoustic modes, and is given by $F_{vib} = \frac{1}{2}\sum\hbar\omega_E^i + \frac{9}{8}\hbar\omega_D + kT\sum\ln[1-\exp(-\hbar\omega_E^i/kT)] + 3kT\ln[1-\exp(-\hbar\omega_D/kT)] - kTD\left(\frac{\hbar\omega_D}{kT}\right)$ where ω_E^i are the $\mathbf{k} = 0$ optical (angular) frequencies, ω_D is an average estimated acoustic phonon mode (angular) frequency at the zone boundary, and $D(x)$ is the Debye function. This is subdivided into the vibrational zero point energy, ZPE (first 2 terms), and thermal energy (final 3 terms). [c] Helmholtz free energy at 298 K is the sum of E_{latt}, the thermal energy and the ZPE. [d] FIT Exp denotes the use of the molecular conformation held rigid at the experimental geometry, with the bonds to hydrogen corrected to standard neutron lengths, and the FIT intermolecular potential. [e] WILL Exp denotes the use of the molecular conformation held rigid at the experimental geometry, with the bonds to hydrogen corrected to standard neutron lengths, and the W99 intermolecular potential. [f] FIT Opt denotes the use of the isolated molecule conformation optimized at the MP2 level of theory with the 6-31G(d,p) basis set, and the FIT intermolecular potential. [g] FIT relax denotes the CrystalOptimizer refinement of FIT Opt with the conformational variability modelled at the HF level of theory with the 6-31G(d,p) basis set. The conformational variables were the rotation and tilt of the amide group and the rotation of the NH_2 group with both hydrogen positions optimized independently as shown in Fig. 2.

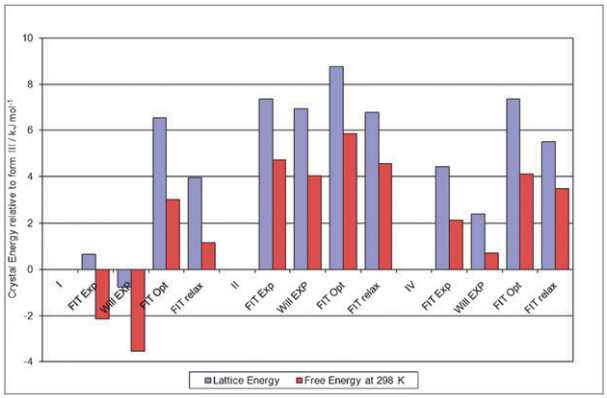


Fig. 3 Crystal energies of carbamazepine polymorphs relative to form III (the experimentally most stable form) as a function of computational model, as defined in Table 1. Lattice energies are shown in blue and free energies at 298 K in red.

form III > form I > form IV > form II, is maintained by the lattice energies of all computational models, other than the W99 potential. However, the relative stability at 0 K still varies significantly. The enthalpies of melting of the four forms[66] all lie within 3 kJ mol^{-1}, and so the relative lattice energies over-estimate the energy range by a factor of about two, if we make the crude assumption that the relative enthalpy does not change with temperature and ignore the heat capacity difference between solid and liquid. The error in ΔE_{intra} is likely to be a major contributor, given the sensitivity of biomolecular conformational energies, including even the barrier to planarity of NH_2 functional groups,[79] to wavefunction quality.[80] There will also be errors in U_{inter} as this can change with the wavefunction and type of representation used to calculate the electrostatic and induction energies.[26]

The intermolecular vibrational modes calculated for carbamazepine have proved useful in assigning the terahertz spectra.[60] The net zero point energy from these modes varies little between polymorphs (Table 1). However, if we estimate†† the free energy at 298 K using these modes and estimating the acoustic modes from the elastic constants,[72] the total thermal energy is also sensitive to the potential and the molecular conformation. As is generally the case,[81] the relative ambient free energy differences are smaller than the nominally 0 K lattice energy differences.

3.2 Induction

The dipole moments induced by the fields within the crystal structure of trifluoromethyl benzaldehyde oxime (Fig. 1a) are generally small (See ESI Table 1†), with the largest being on the nitrogen atom which is involved in hydrogen bonding. The undamped induction energy is estimated to be 7 kJ mol^{-1} with 1 kJ mol^{-1} coming from iterating the induced dipoles to consistency. If damping is applied, then the induction energy is reduced by about 1 kJ mol^{-1}, mainly due to a reduction in the induced moments on the hydrogen-bonding proton (H1) and on the hydrogen-bond acceptor (N1). The analysis of the polarizabilities and induced dipoles of the different functional groups (ESI†) confirms that intermolecular polarization is not negligible and will be sensitive to the relative orientations of different functional groups within the crystal.

Carbamazepine illustrates the importance of the induction energy for discriminating between polymorphs.[26] A display of electrostatic potential from the induced dipole moments for forms III and IV (Fig. 4) shows that the main intermolecular polarization occurs in the hydrogen bonding region. However, since both forms contain the same $R_2^2(8)$ hydrogen bonded dimer, the differences in the induced dipoles have quite a small effect, mainly reflecting the closer CH···O interactions that join the dimers in form IV.[66] The difference in polarization of the hydrocarbon region is subtle because both forms stack the hydrocarbon ring system in the inverted cup motif.[83] Despite

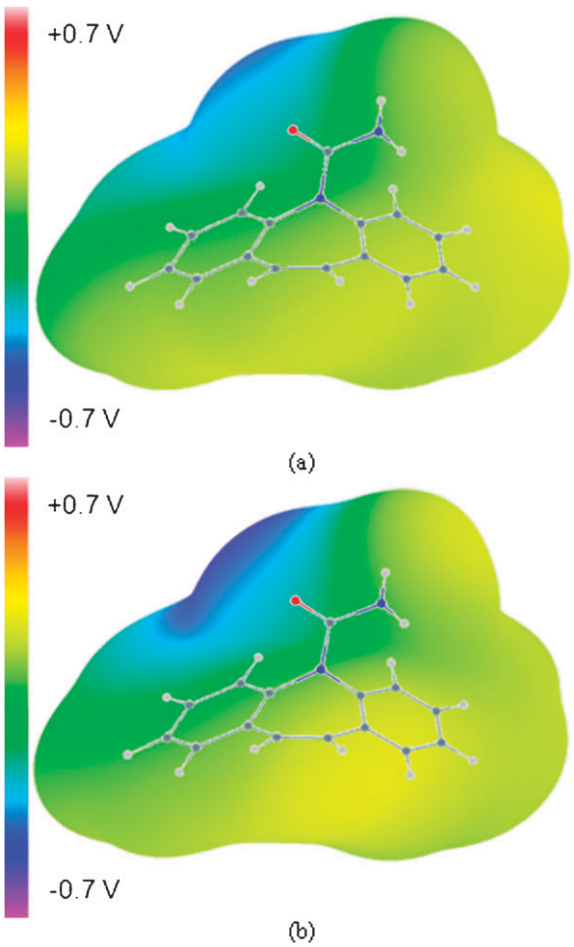


Fig. 4 A display of electrostatic potential on the Bondi van der Waals surface[82] plus 1.2 Å, arising from the induced dipole moments of carbamazepine within (a) form III and (b) form IV. The potential maxima and minima of the surfaces are 0.234 V (corresponding to 22.6 kJ mol^{-1} for a charge of +e at this position) and −0.438 V (−42.3 kJ mol^{-1}) for form III, and 0.241 V (23.3 kJ mol^{-1}) and −0.523 V (−50.5 kJ mol^{-1}) for form IV. The induced moments have been calculated for the optimized molecular structure in the experimental crystal structure with both polarizabilities and multipole moments calculated as in ref. 26

†† This comparison ignores the different cell sizes of the polymorphs, introducing an inconsistency in the partitioning between acoustic and optical modes in the Debye-Einstein model used here. However, the variation in the free energy of form III calculated using different supercells is of the order of tenths of kJ mol^{-1}.

the similarities between the polymorphs, the DMACRYS estimates for the induction energies using the corresponding optimized molecular conformation in the experimental unit cell are -16 kJ mol^{-1} for form III and -23 kJ mol^{-1} for form IV, with 3.5 and 6.5 kJ mol^{-1}, respectively, of this coming from iterating the damped induction to consistency.

The induction energy cannot simply be added to the lattice energy calculated with empirically based model potentials (Table 1, Fig. 3), because these potentials have already absorbed the induction energy in the fitting. Indeed, the similarity between the additional electrostatic field arising from the polarization within the crystal structures of the two polymorphs (Fig. 4) is consistent with the success with which induction has been absorbed into empirically fitted potentials. This double counting also means that minimizing a structure is not yet meaningful. However, using DMACRYS to study the polarization of organic molecules within their crystal structures will help develop a better quantitative understanding of how this affects crystal properties.

3.3 Symmetry reduction

Symmetry reduction, as described in 2.4, is important in crystal structure prediction, as a strain-free structure that is constrained by space group symmetry may be a transition state between lower symmetry structures. The symmetry reduction process produces new low symmetry structures and, for most molecules, introduces new structures in the low energy region of the energy landscape.[84] For example, a search for low energy structures of planar adenine found a *Cmca* $Z' = 1$, $Z = 8$ strain free structure with a negative eigenvalue ($U_{\mathrm{inter}} = -119.7$ kJ mol^{-1}) which corresponded to stacked hydrogen bonded sheets of molecules. On removal of the inversion symmetry element in between the sheets, the structure relaxed to a *Pnma* $Z' = 2$ ($Z = 8$) structure with $U_{\mathrm{inter}} = -131.5$ kJ mol^{-1}. This, too, had a negative eigenvalue, and the $Z' = 4$ structure generated by NEIGHCRYS on removal of the screw axis ($-x$, $+y + 1/2$, $-z + 1/2$) led to a slight buckling of the sheets giving a structure of lattice energy -134.1 kJ mol^{-1}. Using Platon,[85] this was determined to be equivalent to a $Z' = 2$ structure of $P2_1/c$ symmetry. Thus, symmetry reduction can generate genuine $Z' = 2$ structures from a $Z' = 1$ search, although it is unlikely to find very different structures, such as the most stable form II of fluoroisatin,[86] where the two independent molecules have different hydrogen-bonded neighbours. It can also show the limitations of concentrating on lattice energy. A notorious example of this is azetidine, C_3NH_7, a molecule whose $Z' = 2$ $P2_1/c$ crystal structure was the target in the third blind test of crystal structure prediction.[87] Four groups reported that the experimental structure was a saddle point on the potential energy surface, with symmetry reduction to a $Z' = 4$ $P\bar{1}$ structure lowering the energy by between 0.08 and 0.4 kJ mol^{-1}. In this case the zero-point motion would average over the related lattice energy minima to give the higher symmetry observed space group.

Symmetry reduction is much more computationally efficient than doing a primitive Molecular Dynamics "shake up" at each low energy minimum, which can also have the effect of significantly reducing the number of low energy structures that are apparently minima.[88] The number of lattice energy minima that are artefacts of the neglect of thermal motion is very dependent on the specific system: a Molecular Dynamics study[89] of 5-fluorouracil showed that $\sim$75% of the most stable lattice energy minima were free energy minima at ambient temperature.

4. Discussion and conclusions

The program suite DMACRYS has been developed to enable the modelling of a wider range of organic crystals, allowing both larger molecules and more molecules in the asymmetric unit cell, and the use of more theoretically based models for the intermolecular forces. The program uses a distributed multipole representation of the molecular charge density at a particular conformation to parameterize the anisotropic atom–atom model for the electrostatic contribution to the lattice energy. The other terms in the intermolecular potential can be modelled by one of the isotropic atom–atom potentials that have been empirically fitted to organic crystal structures, or a range of more sophisticated anisotropic atom–atom potentials. The corresponding lattice energy can be evaluated sufficiently rapidly to be able to refine the structures and energies of hundreds of thousands of trial crystal structures generated by an effective search program such as Crystal Predictor on a massively parallel cluster, or the few thousand structures generated by MOLPAK[90] on a single processor, on a reasonable timescale. The symmetry reduction process ensures that this results in genuine lattice energy minima. The resulting energies are usually sufficiently accurate that the number of crystal structures that need their relative energies refined by more computationally expensive models for the crystal energy, appropriate to the molecule, is reasonable.

The example calculations on carbamazepine emphasize how demanding the calculation of the relative energies of different organic crystal structures can be. There is significant sensitivity to the model for the intermolecular forces, including the induction contribution, the molecular conformation, and the inclusion of an estimate of the thermal effects. Nonetheless, the results in Fig. 3 are considerably better than can be provided by isotropic force-field modelling.[21] A recent comparison for eleven polymorphic energy differences, five of which were for the carbamazepine system,[21] found that even the best force-field model agreed with experimental results within 1 kcal mol^{-1} (4.2 kJ mol^{-1}) for only nine of the energy differences. The success of the distributed multipole plus empirical force-field approach is also clear in that it usually locates the known structures of rigid neutral organic molecules within the energy range of plausible polymorphism of a few kcal mol^{-1}.[8,19,33] However, once the crystal structure prediction search has generated the types of crystal structure that are thermodynamically plausible, it is often desirable to improve the estimates of the polymorphic energy differences.

A major disadvantage of the use of empirically fitted force-fields is that they have absorbed thermal effects as well as the errors in the model intermolecular potential. The ability to deconvolute these effects would help the development of methods for reliably estimating thermodynamic polymorphic

energy differences and investigating polymorphic phase changes by Molecular Dynamics simulations. (The program DL_MUTLI[91] can be used to perform Molecular Dynamics simulations using distributed multipole models). The new ability of DMACRYS to investigate the polarization of the molecule within the crystal is a prerequisite for the use of accurate molecule-specific anisotropic atom–atom potentials for all contributions to the intermolecular lattice energy, as can be parameterized using CamCASP.[50] This has already been shown to provide an impressive improvement in the ability to predict crystal structures[36,53] for molecules where the induction energy contribution to the relative lattice energies is negligible. Explicitly modelling the induction is likely to improve the prediction of molecular diastereomeric salts[92] and other systems with larger variations in electrostatic field between structures.

The intermolecular induction and molecular flexibility can be effectively modelled by periodic electronic structure methods, once these overcome the problems associated with modelling dispersion and low barrier conformational changes.[65,93] However, electronic level modelling will remain prohibitively expensive for the lattice energies of large pharmaceutical crystal structures and many other types of properties which require Molecular Dynamics based methods. A recent review[94] of the progress in validating the AMOEBA biomolecular simulation force-field, which incorporates polarization and uses distributed multipole electrostatics, notes how the increasing use of computer simulations to complement experimental work demands this generational transition in force-field assumptions. DMACRYS has many capabilities necessary for enabling this type of development for understanding and predicting polymorphism.

Acknowledgements

This program development has been funded by EPSRC since it commenced in 1989, currently by Basic Technology Translation Funding EP/F03573X/1. Further details, including licensing information are available on www.cposs.org.uk. Drs David Coombes and David Willock are also thanked for their input into the original NEIGHBOURS/DMAREL program along with the many users who have assisted the development of DMACRYS through their research.

Notes and references

1 G. P. Stahly, *Cryst. Growth Des.*, 2007, **7**, 1007–1026.
2 A. Llinas and J. M. Goodman, *Drug Discovery Today*, 2008, **13**, 198–210.
3 F. P. A. Fabbiani and C. R. Pulham, *Chem. Soc. Rev.*, 2006, **35**, 932–942.
4 J. Bernstein, *Polymorphism in Molecular Crystals*, Clarendon Press, Oxford, 2002.
5 S. L. Price, in *Polymorphism in Pharmaceutical Solids*, ed. H. G. Brittain, Informa Healthcare USA, Inc., New York, 2009, ch. 3, pp. 53–76.
6 A. Gavezzotti, *Theoretical Aspects and Computer Modeling of the Molecular Solid State*, John Wiley, Chichester, 1997.
7 G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. van de Streek, A. Wolf and B. Schweizer, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2009, **65**, 107–125.
8 S. L. Price, *Acc. Chem. Res.*, 2009, **42**, 117–126.
9 D. J. Willock, S. L. Price, M. Leslie and C. R. A. Catlow, *J. Comput. Chem.*, 1995, **16**, 628–647.
10 S. C. Tuble, J. Anwar and J. D. Gale, *J. Am. Chem. Soc.*, 2004, **126**, 396–405.
11 Y. Mnyukh, *Fundamentals of Solid-State Phase Transitions, Ferromagnetism and Ferroelectricity*, 1st Books Library, Bloomington, 2001.
12 A. Gavezzotti and G. Filippini, *J. Am. Chem. Soc.*, 1995, **117**, 12299–12305.
13 W. T. M. Mooij, B. P. van Eijck and J. Kroon, *J. Am. Chem. Soc.*, 2000, **122**, 3500–3505.
14 B. P. van Eijck, W. T. M. Mooij and J. Kroon, *J. Comput. Chem.*, 2001, **22**, 805–815.
15 B. P. van Eijck, *J. Comput. Chem.*, 2001, **22**, 816–826.
16 S. L. Price, *Int. Rev. Phys. Chem.*, 2008, **27**, 541–568.
17 Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.*, 2006, **2**, 1009–1018.
18 M. A. Neumann and M. A. Perrin, *J. Phys. Chem. B*, 2005, **109**, 15531–15541.
19 S. Brodersen, S. Wilke, F. J. J. Leusen and G. Engel, *Phys. Chem. Chem. Phys.*, 2003, **5**, 4923–4931.
20 M. D. Gourlay, J. Kendrick and F. J. J. Leusen, *Cryst. Growth Des.*, 2007, **7**, 56–63.
21 K. R. Mitchell-Koch and A. J. Matzger, *J. Pharm. Sci.*, 2008, **97**, 2121–2129.
22 A. J. Stone, *The Theory of Intermolecular Forces*, Clarendon Press, Oxford, 1st edn, 1996.
23 A. J. Stone and A. J. Misquitta, *Int. Rev. Phys. Chem.*, 2007, **26**, 193–222.
24 A. J. Stone and M. Alderton, *Mol. Phys.*, 1985, **56**, 1047–1064.
25 S. L. Price, *J. Chem. Soc., Faraday Trans.*, 1996, **92**, 2997–3008.
26 G. W. A. Welch, P. G. Karamertzanis, A. J. Misquitta, A. J. Stone and S. L. Price, *J. Chem. Theory Comput.*, 2008, **4**, 522–532.
27 A. Gavezzotti, *Z. Kristallogr.*, 2005, **220**, 499–510.
28 P. G. Karamertzanis and S. L. Price, *J. Chem. Theory Comput.*, 2006, **2**, 1184–1199.
29 A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman and C. C. Pantelides, in *Molecular System Engineering*, ed. C. S. Adjiman and A. Galindo, WILEY-VCH Verlag GmbH & Co., Weinheim, 2010, ch. 1, pp. 1–42.
30 P. G. Karamertzanis, A. V. Kazantsev, N. Issa, G. W. A. Welch, C. S. Adjiman, C. C. Pantelides and S. L. Price, *J. Chem. Theory Comput.*, 2009, **5**, 1432–1448.
31 P. G. Karamertzanis and C. C. Pantelides, *Mol. Phys.*, 2007, **105**, 273–291.
32 P. G. Karamertzanis and C. C. Pantelides, *J. Comput. Chem.*, 2005, **26**, 304–324.
33 G. M. Day, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2005, **5**, 1023–1033.
34 A. J. Misquitta, R. Podeszwa, B. Jeziorski and K. Szalewicz, *J. Chem. Phys.*, 2005, **123**, 214103.
35 A. J. Misquitta, G. W. A. Welch, A. J. Stone and S. L. Price, *Chem. Phys. Lett.*, 2008, **456**, 105–109.
36 G. M. Day and S. L. Price, *J. Am. Chem. Soc.*, 2003, **125**, 16434–16443.
37 A. J. Misquitta and A. J. Stone, *J. Chem. Theory Comput.*, 2008, **4**, 7–18.
38 A. J. Misquitta, A. J. Stone and S. L. Price, *J. Chem. Theory Comput.*, 2008, **4**, 19–32.
39 A. J. Misquitta and A. J. Stone, *J. Chem. Phys.*, 2006, **124**, 024111.
40 A. J. Misquitta and A. J. Stone, *Mol. Phys.*, 2008, **106**, 1631–1643.
41 G. M. Day, S. L. Price and M. Leslie, *Cryst. Growth Des.*, 2001, **1**, 13–27.
42 D. S. Coombes, S. L. Price, D. J. Willock and M. Leslie, *J. Phys. Chem.*, 1996, **100**, 7352–7360.
43 T. Beyer and S. L. Price, *J. Phys. Chem. B*, 2000, **104**, 2647–2655.
44 D. E. Williams, *J. Comput. Chem.*, 2001, **22**, 1154–1166.
45 F. H. Allen, O. Kennard and D. G. Watson, *J. Chem. Soc., Perkin Trans. 2*, 1987, S1–S19.

46 J. Jia, X. Z. Wang, Y. Zhang and J. W. Wang, *Acta Crystallogr., Sect. E: Struct. Rep. Online*, 2006, **62**, o2683–o2684; there is some disorder in the CF_3 group—the major component was used.
47 A. J. Stone, *J. Chem. Theory Comput.*, 2005, **1**, 1128–1132.
48 A. J. Stone, *GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System [1.0]*, University of Cambridge, Cambridge, United Kingdom, 1999.
49 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03*, Gaussian, Inc., Wallingford, CT, 2004.
50 A. J. Misquitta and A. J. Stone, *CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form University of Cambridge*, http://www-stone.ch.cam.ac.uk/programs.html#CamCASP, 2007.
51 C. P. Brock, W. B. Schweizer and J. D. Dunitz, *J. Am. Chem. Soc.*, 1991, **113**, 9811–9820.
52 J. B. O. Mitchell, S. L. Price, M. Leslie, D. Buttar and R. J. Roberts, *J. Phys. Chem. A*, 2001, **105**, 9961–9971.
53 M. Tremayne, L. Grice, J. C. Pyatt, C. C. Seaton, B. M. Kariuki, H. H. Y. Tsui, S. L. Price and J. C. Cherryman, *J. Am. Chem. Soc.*, 2004, **126**, 7071–7081.
54 T. S. Totton, A. J. Misquitta and M. Kraft, *J. Chem. Theory Comput.*, 2010, **6**, 683–695.
55 A. Sebetci and G. J. O. Beran, *J. Chem. Theory Comput.*, 2010, **6**, 155–167.
56 M. Leslie, *Solid State Ionics*, 1983, **8**, 243–246.
57 S. H. Walmsley, *J. Chem. Phys.*, 1968, **48**, 1438–1444.
58 S. H. Walmsley, in *Lattice Dynamics and intermolecular Forces*, ed. L. V. Corso, Academic Press Inc., New York, 1975, pp. 82–114.
59 G. M. Day, S. L. Price and M. Leslie, *J. Phys. Chem. B*, 2003, **107**, 10919–10933.
60 G. M. Day, J. A. Zeitler, W. Jones, T. Rades and P. F. Taday, *J. Phys. Chem. B*, 2006, **110**, 447–456.
61 R. Y. Li, J. A. Zeitler, D. Tomerini, E. P. J. Parrott, L. F. Gladden and G. M. Day, *Phys. Chem. Chem. Phys.*, 2010, **12**, 5329–5340.
62 G. M. Day, *RUDOLPh. A program for visualising phonon modes in crystals of rigid molecules [1.0]*, London, 2001.
63 A. E. Gray, G. M. Day, M. Leslie and S. L. Price, *Mol. Phys.*, 2004, **102**, 1067–1083.
64 F. H. Allen, S. E. Harris and R. Taylor, *J. Comput.-Aided Mol. Des.*, 1996, **10**, 247–254.
65 P. G. Karamertzanis, G. M. Day, G. W. A. Welch, J. Kendrick, F. J. J. Leusen, M. A. Neumann and S. L. Price, *J. Chem. Phys.*, 2008, **128**, 244708.
66 A. L. Grzesiak, M. D. Lang, K. Kim and A. J. Matzger, *J. Pharm. Sci.*, 2003, **92**, 2260–2271.
67 F. P. A. Fabbiani, L. T. Byrne, J. J. McKinnon and M. A. Spackman, *CrystEngComm*, 2007, **9**, 728–731.
68 M. M. J. Lowes, M. R. Caira, A. P. Lotter and J. G. Vanderwatt, *J. Pharm. Sci.*, 1987, **76**, 744–752.
69 V. L. Himes, A. D. Mighell and W. H. DeCamp, *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.*, 1981, **37**, 2242–2245.
70 M. D. Lang, J. W. Kampf and A. J. Matzger, *J. Pharm. Sci.*, 2002, **91**, 1186–1190.
71 G. Filippini and A. Gavezzotti, *Acta Crystallogr., Sect. B: Struct. Sci.*, 1993, **49**, 868–880.
72 A. T. Anghel, G. M. Day and S. L. Price, *CrystEngComm*, 2002, **4**, 348–355.
73 F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 380–388.
74 A. J. Cruz Cabeza, G. M. Day, W. D. S. Motherwell and W. Jones, *Chem. Commun.*, 2007, 1600–1602.
75 V. L. Himes, A. D. Mighell and W. H. DeCamp, *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.*, 1981, **37**, 2242–2245.
76 A. J. Florence, A. Johnston, S. L. Price, H. Nowell, A. R. Kennedy and N. Shankland, *J. Pharm. Sci.*, 2006, **95**, 1918–1930.
77 J. A. Chisholm and S. Motherwell, *J. Appl. Crystallogr.*, 2005, **38**, 228–231.
78 A. J. Cruz Cabeza, G. M. Day, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2006, **6**, 1858–1866.
79 S. Y. Wang and H. F. Schaefer, *J. Chem. Phys.*, 2006, **124**, 044303.
80 T. van Mourik, P. G. Karamertzanis and S. L. Price, *J. Phys. Chem. A*, 2006, **110**, 8–12.
81 J. D. Dunitz, G. Filippini and A. Gavezzotti, *Helv. Chim. Acta*, 2000, **83**, 2317–2335.
82 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441–451.
83 S. L. Childs, P. A. Wood, N. Rodriguez-Hornedo, L. S. Reddy and K. I. Hardcastle, *Cryst. Growth Des.*, 2009, **9**, 1869–1888.
84 G. M. Day, J. Chisholm, N. Shan, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2004, **4**, 1327–1340.
85 A. L. Spek, *PLATON, A Multipurpose Crystallographic Tool Utrecht University Utrecht*, The Netherlands, 2003.
86 S. Mohamed, S. A. Barnett, D. A. Tocher, K. Shankland, C. K. Leech and S. L. Price, *CrystEngComm*, 2008, **10**, 399–404.
87 G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt and P. Verwer, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2005, **61**, 511–527.
88 W. T. M. Mooij, B. P. van Eijck, S. L. Price, P. Verwer and J. Kroon, *J. Comput. Chem.*, 1998, **19**, 459–474.
89 P. G. Karamertzanis, P. Raiteri, M. Parrinello, M. Leslie and S. L. Price, *J. Phys. Chem. B*, 2008, **112**, 4298–4308.
90 J. R. Holden, Z. Y. Du and H. L. Ammon, *J. Comput. Chem.*, 1993, **14**, 422–437.
91 M. Leslie, *Mol. Phys.*, 2008, **106**, 1567–1578.
92 P. G. Karamertzanis, P. R. Anandamanoharan, P. Fernandes, P. W. Cains, M. Vickers, D. A. Tocher, A. J. Florence and S. L. Price, *J. Phys. Chem. B*, 2007, **111**, 5326–5336.
93 K. E. Riley, M. Pitonak, J. Cerny and P. Hobza, *J. Chem. Theory Comput.*, 2010, **6**, 66–80.
94 J. W. Ponder, C. J. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.
95 C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek and P. A. Wood, *J. Appl. Crystallogr.*, 2008, **41**, 466–470.
96 B. P. van Eijck and J. Kroon, *J. Phys. Chem. B*, 1997, **101**, 1096–1100.