

The computational prediction of pharmaceutical crystal structures and polymorphism

Sarah L. Price*

*Centre for Theoretical and Computational Chemistry, Department of Chemistry, University College London,
20 Gordon Street, London WC1H 0AJ, UK*

Received 21 May 2003; accepted 6 October 2003

Abstract

A computational method of predicting all the polymorphs of an organic molecule would be a valuable complement to polymorph screening in the developmental phase. Such a computational method is in its early stages of development, and the current methodologies, which are based on searches for the most stable lattice structure, are critically reviewed. This crude thermodynamic approach generally overestimates the propensity for polymorphism, at least for most of the molecules studied so far, showing the need to model kinetic effects as well as to refine the thermodynamic models.

Although the ultimate goal of these studies is still far off, computational predictions of crystal structures have proved useful in aiding the characterisation of polymorphs from powder X-ray data, and in providing insights into the range of types of packing that may be adopted by a given molecule. Thus, computational studies already have the potential to be a valuable tool in pharmaceutical solid state science.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Computational prediction; Crystal structure prediction; Lattice energy; Force-fields; Polymorphism

Contents

1. Introduction	302
2. The zeroth order model: finding the energetically feasible crystal structures by lattice energy minimisation	302
2.1. The molecular model	303
2.2. Intermolecular forces	304
2.3. The search procedure	307
3. Overview of results from lattice energy searches	308
3.1. The CCDC blind tests of crystal structure prediction	308
3.2. Survey of published crystal structure prediction papers	309
4. Kinetics and dynamics: the limitations on polymorph prediction	311
5. Examples of crystal structure and polymorph predictions on pharmaceuticals	313
5.1. Investigations with some consideration of kinetic factors	313
5.1.1. Acetaminophen (paracetamol)	313

* Tel.: +44-20-7679-4622; fax: +44-20-7679-7463.

E-mail address: s.l.price@ucl.ac.uk (S.L. Price).

5.2.	Use in conjunction with X-ray powder data	313
5.2.1.	Glucocorticoid	314
5.2.2.	Primidone	314
5.2.3.	Progesterone	314
5.2.4.	Estrone.	314
5.3.	Use in searching for new polymorphs	315
5.3.1.	Aspirin.	315
5.3.2.	Diflunisal.	315
5.3.3.	Allopurinol	315
6.	Conclusions and future prospects	315
	Acknowledgements	317
	References	317

1. Introduction

Can we predict the polymorphs of an organic molecule by computer modelling? Certainly not today, at least at the level that would be desirable for the pharmaceutical industry. If we were able to predict all the polymorphs of a pharmaceutical under development, this would be a considerable aid to the polymorph screening stage. It would provide considerable reassurance that a new polymorph is unlikely to appear during the manufacturing process or in rival's laboratories, thus avoiding the considerable problems exemplified [1] by the cases of ritonavir and ranitidine hydrochloride. However, although such a computational prediction scheme is not yet available, there is considerable academic work aimed at a quantitative understanding of polymorphism that is already providing results that could be of use in pharmaceutical physical form development. Indeed, the pharmaceutical industry can provide considerable assistance into developing such a computational tool.

If we ask the closely related question "Are crystal structures predictable?", then the one word answer in a review of this problem in 1994 was "No" [2]. This has progressed to "still "No", although at certain levels of discussion a "Maybe", or even a conditional "Yes", may be entertained as possible responses" [3]. This is because there are many factors [4] that make predicting crystal structures of organic molecular materials, such as pharmaceuticals, a very different challenge from modelling traditional engineering materials. The more closely defined question of whether it is possible to predict the known crystal structure of an organic molecule, just given the chemical diagram, has been partially answered by the blind tests organised by the

Cambridge Crystallographic Data Centre (CCDC). In these tests [5,6], there have been some successes, as discussed in Section 3.1. These blind tests, alongside the considerable number of independently published studies developing and applying crystal structure prediction techniques, automatically raise the question of polymorphism. Could some of the "wrong" predictions actually correspond to unknown polymorphs?

This review will look at the current state of crystal structure prediction, but with emphasis on the developments required before this could evolve into a method of polymorph prediction, and also the generality of the approach. It is already clear that some crystal structures are more "predictable" than others [7]. So assuming that a method that can predict the crystal structure of naphthalene could predict the polymorphs of your developmental compound is rather like assuming that "if it happens in *E. coli*, then it happens in elephants". An appreciation of the scientific inputs into the crystal structure prediction process, and their current limitations and developmental opportunities, is essential to understanding the potential uses of the methodology.

2. The zeroth order model: finding the energetically feasible crystal structures by lattice energy minimisation

Most methods of crystal structure prediction start by seeking the crystal structure that corresponds to the global minimum in the lattice energy. This assumes that thermodynamics controls crystallisation, and that all temperature effects and zero-point energies can be neglected. Thus, the method seeks the static 0 K

crystal structure that gives the most energetically favourable packing. If there are local minima in the lattice energy that are within the energy range of possible polymorphism, then these are energetically feasible as polymorphs.

Before we can assess the validity of this basic assumption from the results (Section 3), we should examine the necessary inputs into the search, namely a model for the molecule, a model for the forces between the molecules, and the method of searching for the low energy structures. These define the limitations on the types of molecules that can currently be studied, and also the interpretation of the successes of this approach.

2.1. *The molecular model*

For many simple molecules, the forces between the molecules are so much weaker than the forces that define their covalent bonds, that we expect the molecular structure to be the same in the gas phase and in all solid forms. In this case we can determine the molecular structure by an *ab initio* calculation on the isolated molecule, and assume that this will be the molecular structure in the crystalline forms. Many crystal structure prediction programs are restricted to the use of rigid molecules. The use of an *ab initio* or a molecular mechanics optimised structure is necessary for any genuine prediction from the chemical diagram. Furthermore, if there were any deviations in the molecular structure caused by the packing forces in one polymorph, then using this structure would bias the predictions of other polymorphs.

The problem with this assumption is that it is an approximation that varies in its accuracy. For example, 2-amino-5-nitropyridine is a molecule that most experimentalists would describe as fairly rigid. It has three known polymorphs [8]. In two of the polymorphs, the molecule is effectively planar, with small deviations arising from pyramidalisation of the amino group. These crystal structures could be reproduced when the molecular model was taken from the corresponding crystal structure, but minimisations starting from the two experimental structures using the planar, *ab initio* optimised, molecular model converged to the same minimum. Hence, although the crystal structure prediction calculations [8] with the *ab initio* model did find this type of sheet structure to be the most

favourable, it was intrinsically unable to predict that it experimentally corresponded to two structurally similar polymorphs. In the third polymorph, the nitro group was twisted out of the plane of the aromatic ring by 16°. This proved essential to the interdigitation of the hydrogen bonded ribbons in the third observed polymorph, and so the calculations with the planar molecule predicted a significantly less dense and less energetically favourable structure. However, this may be an exceptionally demanding case, as there are many examples of successful predictions with rigid molecules. A counter-example is provided by acetaminophen (paracetamol), where accurate neutron crystal structures over a range of temperatures show minor changes in the molecular structure, including an out of plane tilt of the hydroxyl group by 17–22° which appears to stabilise the structure [9]. Despite the small differences in the lattice energy minima with the different molecular structures taken from the experimental crystals of both polymorphs [9], the planar *ab initio* optimised molecular structure gave a successful crystal structure prediction [10].

For molecules that are obviously conformationally flexible, where there are low energy conformers with drastically different molecular shapes, then one approach would be to use each low energy gas-phase conformer separately as a rigid molecule in a crystal structure prediction calculation. The lattice energies would then be adjusted by the difference in the intramolecular conformational energy. An alternative, used in the commercial program Polymorph Predictor [11,12], is to use a force-field that includes intramolecular energy terms describing the energy changes with bond distortions and torsional rotations. The positions of every atom in the crystal structure under the influence of both the intermolecular and intramolecular forces are then optimised. The success of this approach depends fundamentally on how accurately the force-field represents the balance between the two sets of forces. This may be perfectly adequate for molecules containing functional groups where there are carefully developed force-fields that have been calibrated for the solid state. However, it is notable that the final detailed work [13] on the crystal structure prediction of glycol and glycerol used a method where the intramolecular energy was calculated *ab initio* for each step in the lattice energy minimisation. This computationally intensive approach provided much

more convincing relative total energies than standard force-fields [13], where the interactions between molecules are generally modelled in the same way as non-bonded interactions within the molecules.

2.2. Intermolecular forces

Determining which crystal structure has the lowest lattice energy depends critically on the accuracy of the model for the forces between the molecules. Both the forces and lattice energy are calculated from the model intermolecular potential, and the lattice energy is the sum of the energies of the intermolecular interactions (i.e. the intermolecular potentials) between all the molecules in the crystal. It is the relative accuracy that is important, as the calculations may compare crystal structures with different hydrogen bonds and degrees of π – π stacking, and certainly are comparing different atom–atom contacts. Thus, crystal structure prediction calculations are restricted to molecules where there is a sufficiently accurate intermolecular potential, if the molecule is held rigid, and to those where the non-bonded terms in the force-field are sufficiently well parameterised when an all-atom optimisation is being performed.

Certainly, a necessary, though not sufficient, condition for a successful crystal structure prediction is that the model intermolecular potential gives a minimum in the lattice energy reasonably close to the known experimental structures. This minimum is the closest to the experimental structure that can possibly be found in a search for the global minimum, and so intrinsically limits the accuracy of the predictions. However, even with the definitively accurate model intermolecular potential, there would be a discrepancy between the minimum in the lattice energy and the experimental structure, caused by the molecular motions. Thus, the accuracy of the predictions is limited to the order of thermal expansion effects. For organic molecular crystals, this can be up to a few percent of the room temperature lattice parameters, but the effect can be very anisotropic. Although the volume of the crystal expands on heating, the molecular reorientation sometimes leads to a lattice parameter shrinking slightly with a rise in temperature. Thus, although many model intermolecular potentials are parameterised to room temperature crystal structures, and therefore have absorbed some averaged form of

thermal effects, this cannot be rigorous. Therefore, at our current crude stage of organic crystal structure modelling, we have a reasonably realistic model if the lattice energy minimisation reproduces the known crystal structures to within a few percent in the lattice parameters (with similar small deviations in molecular orientations and positions) [9]. Deviations of more than 5% give serious cause for concern about the adequacy of the model.

A discussion of intermolecular potentials and force-fields would be a major review in itself. Despite considerable fundamental research into this area [14], model intermolecular potentials that are sufficiently accurate to predict a wide variety of properties of the molecule in the solid, liquid and gaseous states within experimental accuracy are only available for the simplest systems such as argon. Although considerable progress is being made towards such a goal for water [15], the complexity of such model potentials shows that we cannot expect the model intermolecular potentials for organic molecules to be reliable. Like most computer modelling applications in the pharmaceutical industry, it is a case of trying to ensure that the model potentials available will be adequate for the purpose of the computational study. For hydrocarbons and the functional groups found in proteins and nucleic acids, there is often a choice of different force-fields. For other functional groups, the choice may be very limited, or the parameters non-existent, requiring considerable work on extending the force-field for the particular family of pharmaceuticals. However, work on organic crystal structure modelling [16] suggests that standard force-fields, designed for liquid or biomolecular docking simulations, are often not good enough, and worthwhile crystal structure prediction demands advances in the representation of intermolecular forces.

All modelling of the forces between organic molecules [17] is based on the atom–atom approach, where the interaction between two molecules is the sum of atom–atom interaction energies. In most widely used force-fields, this interaction only depends on the separation of the two atoms, i.e. assumes that the molecule interacts as if it were a superposition of spherical atomic charge densities. The electrostatic model is usually derived specifically from the charge distribution of the isolated molecule, obtained by an *ab initio* calculation. This is usually an atomic charge model,

derived by fitting the atomic charges to reproduce the electrostatic potential in the region around the molecule. However, for greater confidence in the accuracy of the electrostatic forces, the charge distribution around each atom can be represented by a multipole series (charge, dipole, quadrupole, octupole and hexadecapole), obtained by a distributed multipole analysis (DMA) [18] of the charge distribution. Such models have been found necessary, and remarkably successful, in predicting the directionality of hydrogen bonds in van der Waals complexes [19]. Because the atomic multipoles represent the electrostatic effects of the lone pair and π electron distribution, they can model their effects on hydrogen bonding, π – π stacking, etc. However, the price to be paid for being able to evaluate the electrostatic contribution to the lattice energy to essentially the accuracy of the *ab initio* wavefunction, is considerable computational complexity. The programmes DMAREL [20], ORIENT [21] and TINKER [22] can handle the non-central forces and torques which arise from such anisotropic atom–atom potentials.

In most modelling, all the other intermolecular forces are represented by an atom–atom repulsion–dispersion model, with an R^{-6} model for the long-range dispersion forces, and an exponential (or less accurately a power model in the Lennard Jones form) for the repulsion. The repulsion model predominantly determines the separation of the molecules, whereas the electrostatic term is vital in determining their relative orientation (via formation of hydrogen bonds and attraction of electropositive atoms to electronegative ones). The dispersion term is a universal attractive force, which makes a significant contribution to the lattice energy, keeping the crystal as dense as the steric (repulsion) forces allow. The repulsion–dispersion parameters have usually been empirically fitted to a range of crystal structures and sublimation energies, and are considered transferable for the atomic type. In the latest parameterisations from Williams [23–25], the intermolecular parameters are different for different hybridisation states of C, N and O, as well as distinguishing the repulsion–dispersion forces from hydrogen atoms bonded to C from those bonded to hydrogen atom donors. In this intermolecular potential parameterisation, the interaction sites for the hydrogen atoms are shifted in 0.1 Å from the nuclear position to allow for the hydrogen electron density not being

centred on the proton. This density shift results in a systematic error in the location of hydrogen atoms by X-ray diffraction, which are often subject to significant experimental uncertainty. Thus, the positioning of hydrogen nuclei and interaction sites requires considerable care in crystal structure modelling. Unfortunately, many (but not all) crystal structures can be quite sensitive to the assumed hydrogen atom positions and interactions.

There are many other sets of transferable potential parameters available in the literature, from the earliest days of organic crystal structure modelling [26]. In the search for a model force-field for a pharmaceutical with uncommon functional groups, it is necessary to be very cautious of mixing parameters from different force-fields. The empirical parameters have absorbed different errors in the approximations in the force-field and the parameters for different atom types are often correlated. As an extreme example, the UNI force-field [27,28] was developed using only the *exp-6* potential and no electrostatic model, without using the usual constraints relating heteroatomic interactions to the two homoatomic intermolecular interactions. Thus, for example, the $N \cdots H_N$ (H_N is a hydrogen bonded to N) interaction has a much deeper potential well than the $N \cdots N$ or $H_N \cdots H_N$ potentials, because it absorbs the electrostatic attraction of the hydrogen bond. Whilst it would clearly be nonsense to use the UNI parameters and an explicit electrostatic model, as this is double counting the electrostatic energy, other forms of mixing parameters may also give unsatisfactory results for analogous, but less obvious reasons. Another pitfall is the assumption of transferability. For example, many potentials that can reproduce the crystal structures of many carboxylic acids, fail to reproduce the crystal structures of the two polymorphs of oxalic acid. The lack of any hydrocarbon moiety changes the charge distribution of the carboxylic acid group, and also means that different atom–atom contacts are sampled in the oxalic acid polymorphs than in the generality of carboxylic acid crystal structures. However, a specific intermolecular potential for oxalic acid, derived from its own *ab initio* charge distribution, is able to model both polymorphs [29].

Crystal structure prediction studies have provided considerable impetus towards improving the model potentials for organic molecules. The basic assumption implies that the intermolecular potential model has to

be accurate enough to give the known crystal structure as the global minimum in the lattice energy. Indeed, a new methodology for parameterising force-fields uses the condition that the known structure should be more stable than alternative structures found in a crystal structure prediction calculation, as a constraint [30]. Certainly, pioneering work on simple sugars and alcohols showed that improving the force-field by using distributed multipoles [31], explicit modelling of polarisation [32,33], and using ab initio estimates of the intramolecular energy [13,34], increased the number of known structures that were found at the global minimum (or at least decreased the energy gap and improved the relative ranking). For chlorothalonil, a specific potential [35] derived from the ab initio charge density and representing non-spherical repulsion arising from the Cl and C $\equiv$ N lone pair density was used to refine the structures and relative energies in a crystal structure prediction search. The results aided the inter-

pretation of the two new polymorphs that were found in the collaborative experimental investigation, while the specific potential predicted would have very similar lattice energies to the previously known polymorph [35]. Indeed, the question of whether the intermolecular potential is adequate for crystal structure prediction by comparing relative lattice energies [36] has led to the development of a method of evaluating the lattice energy from the ab initio molecular charge distribution [37] as represented by tens of thousands of pixel sites. This model is too computationally expensive to be used in lattice energy minimisation, let alone a full search, at the present time, but it is providing an alternative model for estimating the relative lattice energies of structures found with atomistic model potentials. Thus, although Fig. 1, in illustrating some of the molecules that have been subject to published crystal structure prediction studies, gives an idea of the range of molecules where a suitable force-field is available, the conclusions of

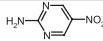
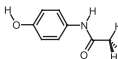
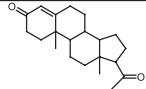
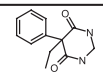
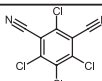
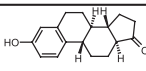
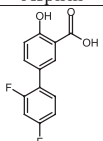
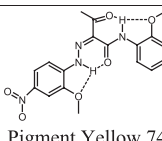
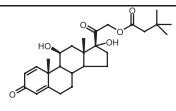
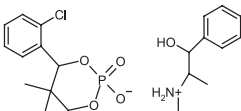
 Acetic acid	 Glycerol	 2-amino-5-nitropyrimidine
 5-azauracil	 Acetaminophen (paracetamol)	 Progesterone
 Parabanic acid	 Allopurinol	 Primidone
 Aspirin	 Chlorothalonil	 Estrone
 Diflunisal	 Pigment Yellow 74	 Prednisolone t-butylacetate
 Salt of ephedrine chlocyphos		

Fig. 1. Some of the molecules mentioned in text. The molecules in the top section (along with those in Fig. 2), illustrate the range of molecules currently considered in crystal structure prediction studies. The bottom salt represents by far the most complex system studied to date, requiring considerable computational and scientific resources, and so it is likely to be some years before such systems can be routinely studied.

many of these studies are limited by the confidence in the assumed force-field.

Work on improving model intermolecular potentials on the assumption that the known structure should be the global minimum in the lattice energy, relies on the known structure being the thermodynamically most stable structure at 0K. The phenomenon of polymorphism shows the dangers of that assumption—even if only one polymorph is known, this does not guarantee that there is not another structure which is thermodynamically more stable at low temperatures, but has not been experimentally observed. However, if the force-field does not reproduce the known crystal structure(s) satisfactorily, then it definitely needs improvement. (When a rigid molecule is used, minimising the lattice energy with the intermolecular potential and both the experimental and gas phase rigid molecular structure will reveal whether any problems with the crystal structure reproduction lie in the assumed intermolecular potential or molecular structure.) However, if the model is satisfactory in respect of crystal structure reproduction, but gives a lattice energy that is within the small energy differences associated with polymorphism above the global minimum, then the interpretation becomes difficult. The more theoretically well based the model potential, the greater the suspicion that the observed structure is metastable at low temperatures.

2.3. The search procedure

Even if the model for the intra- and intermolecular forces is good enough for the stable crystal structure to correspond to the global minimum in the lattice energy, it is the search procedure that will determine whether it is found in the crystal structure prediction study. Thus, many methods for searching the huge multi-dimensional energy hypersurface of possible crystal structures have been proposed [5,6] and more will come from collaborations with the many other fields of biological and engineering research that involve global optimisation problems.

Most methods are either explicitly, or in practice, restricted to crystal structures in the most common space groups for organic molecules, such as $P2_1/c$, $P2_12_12_1$ and $P\bar{1}$. Furthermore, most searches are restricted to one molecule in the asymmetric unit. Again, most crystal structures in the Cambridge Structural

Database (CSD) [38] fall into this category, with relatively few molecules adopting crystal structures with $Z > 1$ [39]. However, even with these restrictions, the number of variables that need to be optimised will be the cell parameters (up to six for the low symmetry $P1$ and $P\bar{1}$ space groups) and the relative orientation and position of the one rigid molecule (or all atoms within one molecule if an intramolecular force-field is used) in the asymmetric unit cell. This, in itself, is computationally demanding. However, having more than one independent molecular unit in the asymmetric unit cell, as for salts, hydrates and solvates, increases the dimensionality of the hypersurface that needs to be considered, and hence the computational requirements. There are a few pioneering studies where two independent molecules in the asymmetric unit have been considered, either to increase the range of possible crystal structures considered [40], or to study monohydrates of pyranoses and polyalcohols [41], or the racemate resolution of ephedrine with the chiral acid chlocyphos [42]. However, all these studies required considerable computer time and raised issues about the completeness of the search. Increasing computer power will undoubtedly enable crystal structure predictions on salts and solvates, as well as more comprehensive searches for single molecules, to become more feasible and reliable over the next few years.

The only commercial program in this field, Polymorph Predictor [11,12], is one of the group that attempts a search that is only restricted by the space group, in this case, by a simulated annealing Monte-Carlo method. Since this is a non-deterministic method, seeking to sample the hypersurface through approximating a high temperature melt, the simulation in each space group is run a few times until further searches produce no more low energy minima. The MPA program [43] uses molecular dynamics to seek the low energy conformations as an expanded unit cell containing a fixed number of molecules is allowed to contract. The systematic UPACK [44] search considers the crystal variables on a grid, though a new search method was developed for the case of more than one molecule in the asymmetric unit cell [41]. CRYSTALG [45] developed Monte-Carlo techniques used in protein structure prediction programs to get over the local minima and search for the global minimum in the lattice energy. These programs thus consider millions of structures, but most are discarded very rapidly, for

example on the basis of an approximate model potential, which is refined in stages, or by clustering those which seem likely to be going towards same minimum. Here the danger is that minima that are very similar, but do actually correspond to distinct polymorphs, may be missed.

The other class of approach uses some scientific insight into crystal packing to determine a number of approximate structures that are then energy minimised. One example of such a program is PROMET [46], which looks for low energy structures of pairs and then clusters of molecules, built up systematically from the space group. Another is MOLPAK [47], which performs a systematic grid search with a pseudo-hard sphere molecule model in almost 30 common packing types established by analysis of the CSD. The 50 or so densest structures of the few thousand systematically generated for most packing types are then energy minimised. This approach was developed for studying energetic materials, where density is a key parameter, but has also proved successful for other hydrogen bonded rigid molecules.

All these computational approaches to crystal structure prediction by searching for the global minimum in the lattice energy have been used in the CCDC blind tests of crystal structure prediction, along with other methods based on statistical analysis of known crystal structures. Remembering that the success of a crystal structure prediction study depends on the model assumed for the intra- and intermolecular forces, as well as the search procedure, let us consider the results of such studies to date.

3. Overview of results from lattice energy searches

The hope behind the assumption that crystal structures could be predicted by searching for the global minimum in the lattice energy was that there would be a clear energy gap between the known polymorphs and the other hypothetical crystal structures. One example of this is provided by an irregular, but rigid molecule, Pigment Yellow 74, where the experimental structure was found to be 12 kJ/mol (~ 3 kcal/mol) more stable than any other hypothetical structure found in the search [48]. Since this is larger than the likely energy difference of less than 2 kcal/mol between observable polymorphs [1], this result predicts that no further

polymorphs are likely. This appears to be the case, as pigment Yellow 74 has been extensively studied [48] in Clariant's laboratories, with a series of alternative synthetic routes, with no detectable signs of another polymorph.

Unfortunately, such clear-cut predictions are rare. From the earliest studies it became clear that the search procedures usually generated far more crystal structures within the energy range of possible polymorphism, than known polymorphs. Work with increasingly accurate model potentials, including the two independent studies of acetic acid [49,50], made it clear that the existence of many low energy structures was not an artefact of the model intermolecular potential used. (Naturally, different model potentials rearranged the relative energies and thus rankings of the low energy structures, with an overall tendency for better results being given by the potentials that were theoretically more justified for the molecule in question).

3.1. The CCDC blind tests of crystal structure prediction

This plurality of low energy structures has a major influence on how we interpret the results of the CCDC's blind tests of crystal structure prediction. In 1999 [5] and again in 2001 [6], workers in the field were sent the chemical diagrams of three molecules, whose crystal structures fell within the capabilities of at least some of the programs, and challenged to submit three "predictions" for each crystal structure. The molecules and the success rate are shown in Fig. 2. After the predictions had been submitted, the experimental crystal structures were released, and the participants met to learn from the experience. Given the experimental structure, it is possible to determine why a prediction was unsuccessful. Either (a) the model force-field did not give a minimum either close enough to the experimental crystal structure or close enough to the global minimum in the lattice energy, (b) a satisfactory minimum exists, but the search had failed to locate it, or (c) there were so many crystal structures within the energy range of possible polymorphism that the participant had been "unlucky" in the three that had been selected.

The straightforward interpretation of the results in Fig. 2 is that crystal structures can be predicted, but

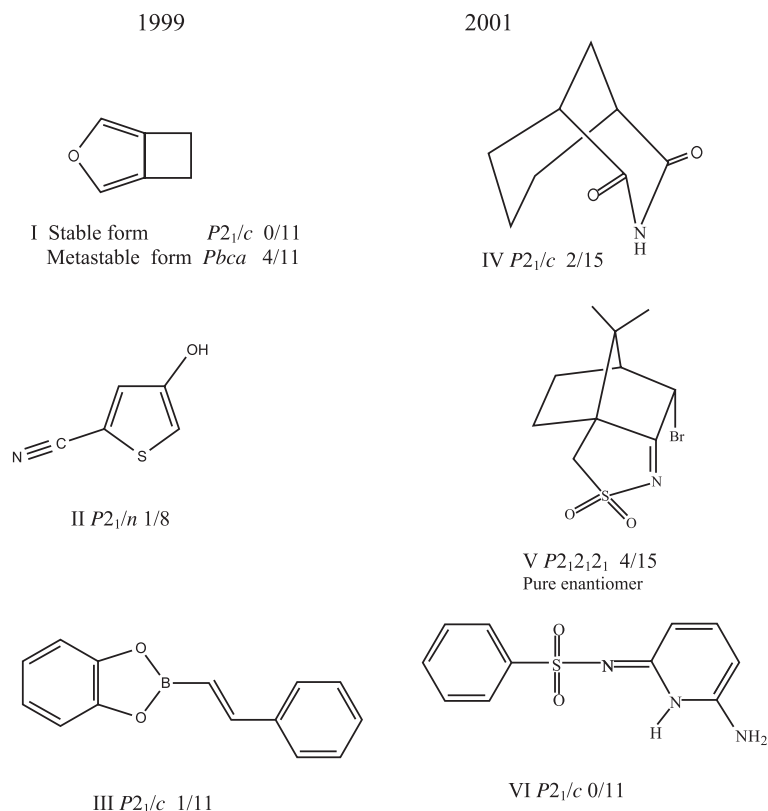


Fig. 2. The molecules used in the blind tests of crystal structure prediction, organised by the CCDC. For each crystal structure, the success rate is given as x/y , where x is the number of successful predictions, and y is the number of groups that submitted (usually) three guesses for the crystal structure.

that no single method is reliable. This is within the limitations of the exercise, in which the molecules had been chosen from unpublished crystal structures in a common space group with $Z' = 1$, and fell into the three categories of challenges. The first were rigid molecules containing C, H, N and O atoms only, the second a rigid molecule with a larger range of atomic types, presenting a challenge to the intermolecular potential, and finally a molecule with limited flexibility. The results do reflect this increasing difficulty. Unfortunately, despite valuable lessons being learnt from the collaboration on the first blind test, the results in the second one were not an obvious improvement. These blind tests have been welcomed by pharmaceutical industry scientists as valuable means of assessing progress in the field. However, we should take the phenomenon of polymorphism into account in making conclusions. Most obviously, if the meta-

stable, disappearing $Pbca$ polymorph of molecule I had not been characterised in the first crystallisation attempt, then the results of the first blind test would have seemed most discouraging. Secondly, in discussions of the second blind test, it was suggested that some of the hypothetical structures that were submitted by several groups might be undiscovered polymorphs. Indeed, a later search for other polymorphs of VI has indeed found [51] another polymorph that is more stable than the one with the unusual conformation that was used to judge the crystal structure prediction methods.

3.2. Survey of published crystal structure prediction papers

Thus the tendency shown by the phenomenon of disappearing polymorphs [52], and described over a

century ago by Ostwald's rule of stages [53,54], for the first crystalline polymorph found to correspond to a metastable polymorph, complicates the development of polymorph prediction. Fig. 3 represents the results of a survey [7] of all published papers on crystal structure prediction by lattice energy minimisation found in 2000 (the picture has not yet changed substantially). From this it is clear that in approximately half the cases, the known crystal structure was found in the search as a local, not the global, minimum. The metastable polymorph should be found as a local minima, but although this survey included an unrepresentatively high proportion of molecules that are polymorphic (29/189 in contrast to the CSD having only a few hundred in nearly a quarter of a million of the molecules with more than one crystal structure [55]), this only accounts for a small proportion of these local minima. The most likely explanation for the rest is inadequacies in the assumed force-field. However, if you look at the molecules involved, and the dates at which their crystal structures were determined, it is clear that the majority represent the structure of the first crystal suitable for single crystal X-ray diffraction

that could be obtained. The chemist would have been unlikely to be actively looking for polymorphs. Hence, we have to consider the possibility that the only known crystal structure should correspond to a local minimum, and that another crystal form is the thermodynamically most stable at low temperatures.

This survey also shows that current conclusions about the success of crystal structure and polymorph prediction is based on a far too small, and very unrepresentative set of molecules, particularly from the point of view of the pharmaceutical industry. A very large proportion of the molecules studied so far are hydrocarbons, alcohols and sugars (Fig. 4), molecules with particularly weak non-directional interactions, or notoriously difficult to crystallise. This perverse studying of molecules whose crystal structures would be expected to be particularly difficult to predict, arises from many people choosing to study hydrocarbons so they could use the well-developed intermolecular potentials, and the spectacular bravery of the Utrecht group's work on sugar and polyalcohol structures. Despite the pharmaceutical industry being the potentially most significant beneficiary of a meth-

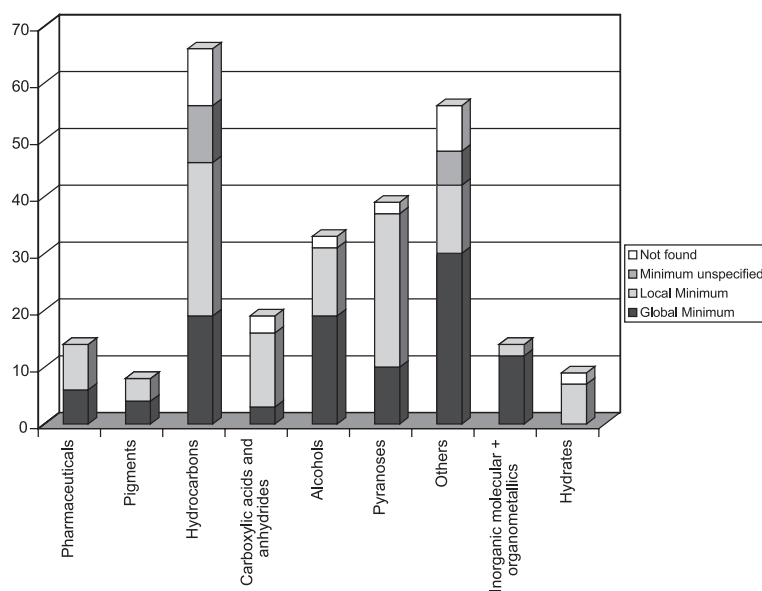


Fig. 3. The success rate of searches for a given crystal structure by lattice energy minimisation, as derived from a survey of published studies. The results are distinguished as to whether the known structure was found as the global or local minimum, or not found. The number of searches significantly exceeds the 189 molecules categorised in Fig. 4 because different groups have published a search for the same structure and because it includes searches on polymorphic systems, where there are multiple crystal structures and only one can be the global minimum.

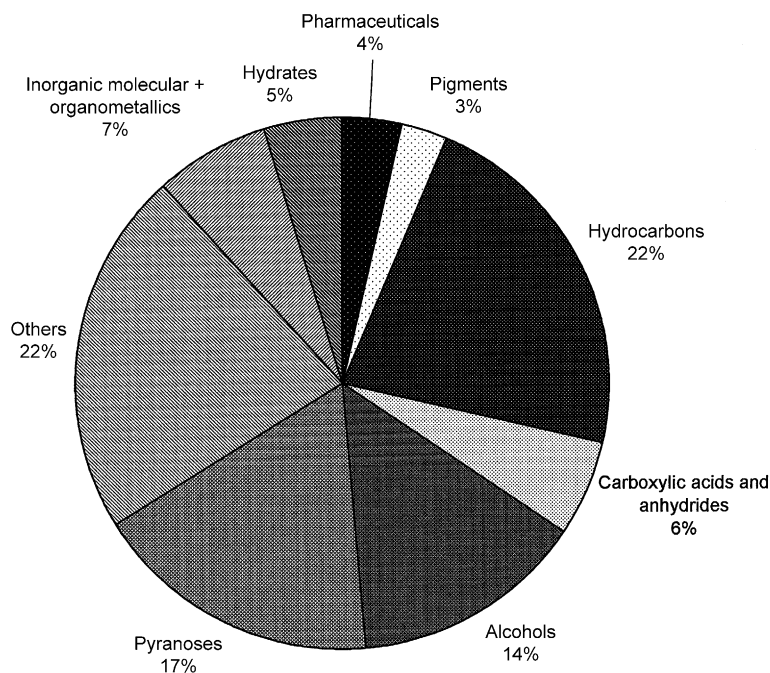


Fig. 4. The types of molecules studied in the first decade of crystal structure prediction, as derived from a survey of published lattice energy minimisation based studies. Within each category, the molecules are generally the smaller and more rigid molecules, whose published crystal structures are in common space groups with $Z' = 1$.

od of predicting crystal structures and polymorphism, there are only about seven pharmaceuticals in the sample, which will be reviewed in Section 5.

4. Kinetics and dynamics: the limitations on polymorph prediction

The existence of polymorphs demonstrates that experimental crystal structures need not correspond to the global minimum in the lattice energy. It is necessary that their energies are not much higher than the global minimum, and determining which structures are energetically feasible is certainly the first stage in a prediction. Entropic factors reflecting the increasing dynamical motions of the molecules can obviously switch the relative stability of two structures with increasing temperature. Structures which are always metastable (i.e. the lower melting structure of a monotropically related pair) also exist, but are only likely to be persistent enough to be of practical interest to the pharmaceutical industry if the thermodynamic driving

force is small compared to the kinetic barrier to transformation. Thus, the possibility of being an observed polymorph certainly decreases as the lattice energy increases above the global minimum, probably exponentially. It is debatable what sort of energy range should be considered [1] because of uncertainties in the limited available experimental data. (An additional margin should be allowed for the relative errors in the calculation, with a worst case example being *o*-acetamidobenzamide where a hydrogen bond goes from being intramolecular to intermolecular in the polymorphic phase change [56].) It is usually estimated that thermodynamic energy differences between polymorphs are less than about 2 kcal/mol. Structures within this limit, plus uncertainty in the relative energies, can be considered as energetically feasible polymorphs. However, if there are a multitude of lower energy structures, the probability of a structure readily transforming to a lower energy structure increases.

The thermodynamic effects of temperature on the crystal structure and polymorph prediction problem are being investigated. An estimate [57] of differences

in the lattice-dynamical entropy between known polymorphs in the CSD concluded that the differences do not exceed 15 J/(K mol) (~ 3.5 kcal/(K mol)). This confirms that the polymorphic energy differences are dominated by the enthalpy (and hence by the lattice energy since polymorphs differ so little in density). However, when the lattice energy differences are negligible, then the entropy differences matter [58]. This has been investigated for glycol and glycerol [59] and shown to be significant in ranking the thermodynamic stability of structures. Even for rigid molecules, the differences in the rigid body motions can reorder the relative energies of the structures at experimentally significant temperatures [60]. Indeed, in parabanic acid, the experimental structure is only 2.5 kJ/mol (~ 0.6 kcal/mol) more stable than any hypothetical structure in lattice energy [61], but a low frequency mode increases the free energy difference to 4 kJ/mol (~ 1 kcal/mol) at room temperature. However, it has to be remembered that all of these calculations are approximate, and most significantly, are derived from the curvature of the minimum in the lattice energy hypersurface. For organic crystals near their melting point, and for low energy, large amplitude motions that eventually result in phase transformations, such approximations are very suspect.

The key factor determining which thermodynamically plausible structures will be observed is kinetics. When does a particular crystal structure have such a kinetic advantage in nucleation and growth, that it will be observed first, despite being a metastable polymorph? When will the kinetics of transformation to a more stable structure be so slow that the metastable form will persist for long enough to require careful study by the formulation pharmaceutical scientist?

These and related questions are extremely hard to answer with our current knowledge of nucleation, crystallite growth and phase transformations [4]. Nucleation is certainly a key step in understanding polymorphism [54], as changes in solvent to favour certain packing arrangements in the nucleus cannot only rationalise some polymorphism, but also such predictions can lead to the discovery of new polymorphs. The observation of concomitant polymorphism [53], and the large delays before the first crystallisation of the thermodynamically stable form in heavily investigated systems (e.g. ritonavir [62]) all point to these being extremely difficult phenomena to

understand, let alone quantify into a computational model.

One reaction to this is to assume that both thermodynamic and kinetic factors are reflected in known crystal structures in the CSD [38] and, therefore, use this as the basis for crystal structure prediction. Certain methods, based on statistical analysis of the CSD are being developed and were applied, unsuccessfully, in the blind tests. An alternative is to use the CSD structures to select the most likely of the low energy structures produced in a search for the global minimum in the lattice energy. This has been applied to blind test IV (Fig. 2) by looking for similar structures [63]. The number of related structures available in the CSD limits the considerable promise of this approach. Certainly, if all crystal structures were deposited in this database, and the information on the growth conditions and morphology were available, the opportunities to deduce factors that influence crystallisation would be considerably enhanced.

The alternative is to consider the energetically feasible hypothetical structures generated in a lattice energy search, and consider whether their properties make them unlikely to be observed, or favour their observation. This has been done at a very primitive level by considering the elastic constants and growth morphologies of the known and hypothetical low-energy crystal structures of a given molecule. The basic idea is that when a molecule can adopt several structures that are thermodynamically almost equivalent, those crystal structures that grow particularly slowly, or are particularly susceptible to mechanical deformation, are unlikely to be obtained experimentally. However, to apply this, first one needs an idea of how low a resistance to shearing forces is likely to make a crystallite unlikely to grow under competitive crystallisation conditions, or how slowly must a crystal face grow (giving an extreme plate or needle morphology) to be unlikely to be observed. Although these criteria seem to be helpful in eliminating some hypothetical crystal structures in studies of acetaminophen [10], pyridine [60], parabanic acid [61] and chlorothalonil [35], considerably more experience is required before they can be applied with confidence. The second issue is how to predict these properties, and whether simple calculations are likely to be adequate. The mechanical properties of an infinite perfect crystal at 0 K can be estimated from the second derivative matrix of the

energy with respect to distortions [64]. Such estimates do provide worthwhile predictions of both the elastic properties of single crystals, and the mechanically averaged properties of aggregates [65] for organic crystals such as pharmaceuticals. The attachment energy model can predict the morphologies of vapour-grown crystals (below their roughening temperature). This assumes that the growth rate of a face is proportional to the energy released when a growth layer of the crystal is attached, and is reasonably successful [66] at predicting the morphologies of organic crystals when solvent effects on the morphology are small. Thus, although the use of these computer predictions in assessing which hypothetical structures may be kinetically disadvantaged is still tentative, they do provide useful predictions of the properties of known polymorphs.

5. Examples of crystal structure and polymorph predictions on pharmaceuticals

Reviewing the crystal structure prediction studies that have been performed on pharmaceutical molecules, not only shows how the technique has been developing, but also shows that the calculations have practical uses, even if they predict too many structures as thermodynamically feasible polymorphs.

5.1. Investigations with some consideration of kinetic factors

5.1.1. Acetaminophen (paracetamol)

The polymorphism of acetaminophen has been widely studied both experimentally and in computational polymorph prediction studies, mainly because of the considerable published literature on two of the polymorphs. The most stable form at ambient temperatures (form I ($P2_1/c$)) is marketed, but form II ($Pbca$) has also been carefully studied [67] because its mechanical properties would allow tableting by direct compression. An early Polymorph Predictor study [12] on acetaminophen correctly found the two known structures, and after some experimentation with different force-fields, predicted that these would be the only two polymorphs. However, a third polymorph of acetaminophen had been reported as being observed by thermal microscopy, but transforming so readily on removal of the cover slide that it could not be charac-

terised [68]. A second crystal structure prediction study [10], using a rigid molecule, a distributed multipole electrostatic model and a MOLPAK search, also found form I as the global minimum, and form II only 3.6 kJ/mol (0.9 kcal/mol) higher in energy. (Since prediction studies are effectively 0 K, this may not be the correct order of stability if the two forms are enantiotropically related.) However, there were a dozen hypothetical structures within 10 kJ/mol (2.5 kcal/mol) of the global minimum in the lattice energy, some only very slightly less stable than form II. The properties of some of these structures suggested that they were unlikely to be observed. One was likely to undergo a facile transformation to the more stable form II, another was too susceptible to deformation by shearing forces, and three more were predicted to have a very slow growing face, resulting in an extreme plate-like morphology. However, this still left several candidate structures for additional polymorphs that were not significantly less stable than form II. Thus, all the low energy crystal structures were deposited as supplementary information in the publication [10], to allow comparison with any future experiments on form III.

A high throughput polymorph search for form III of acetaminophen [69], described in this issue, did obtain a powder X-ray diffraction pattern of form III, although it transformed into form II within a few hours. These data provided a sufficiently good match to one of the computed patterns for the corresponding hypothetical structure to be suggested as the structure of form III. However, this particular $P2_1/c$ structure was one of those predicted to have one particularly slow growing face, and so deemed to be unlikely to be an observed polymorph. This prediction of difficulty in growing in one direction was compatible with the powder pattern suggesting only short-range order in one lattice direction. Whether you regard the acetaminophen story as a success, failure, or learning experience for computational polymorph prediction, really depends on how much weight you attach to predicting very unstable polymorphs.

5.2. Use in conjunction with X-ray powder data

The use of predicted hypothetical crystal structures to aid the structural characterisation of crystalline solids from powder X-ray diffraction data is a potentially important application. Whilst the acetaminophen

example shows the use for very unstable crystal structures, it is very common for it to be very difficult, even impossible, to grow sufficiently large crystals for single crystal X-ray diffraction. Most current methods of solving structures from powder X-ray diffraction require that the unit cell parameters and space group can be extracted from the data. This is not always possible, because of sample quality or the overlap of peaks. In such a case, comparison of the experimental powder pattern with those corresponding to the hypothetical crystal structures should suggest a crystal structure that can then be refined by Rietveld methods. This approach has been developed in the pigment industry [48], and there are some examples of applications in pharmaceutical research.

5.2.1. Glucocorticoid

The steroid, prednisolone *tert*-butylacetate (patented as glucocorticoid by Merck & Company) had two anhydrous polymorphs, one of which did not have a known crystal structure. A Polymorph Predictor study [12], considering two of the low energy conformations of the compound in the indexed $P2_12_12_1$ space group, found a stable structure with a simulated powder pattern sufficiently similar to the experimental structure that it could be used as a starting point for a successful Rietveld refinement.

5.2.2. Primidone

The anticonvulsant primidone and the steroid progesterone were used to independently test [70] the use of the Polymorph Predictor program for pharmaceutical type molecules that had known polymorphs. Both polymorphs of primidone were found in the top 10 structures that had been generated in the known space groups. It was noted that the conformation change between the two polymorphs could be modelled, in that the crystal structure of form A was found in a search within the space group of form A, starting from the conformer for form B. However, they also noted that it would have been difficult to identify the correct packing arrangement by comparison with experimental X-ray powder data alone. The difference between the experimental powder pattern and that of the closest lattice energy minimum was sufficiently large that establishing that the search had found the “experimental crystal structure” required the powder pattern of the corresponding lattice energy minimum. However, once

this most similar lattice minimum had been recognised, it was possible to obtain a better fit to the powder pattern of form A by Rietveld refinement.

5.2.3. Progesterone

For progesterone, where the two polymorphs are in the same $P2_12_12_1$ space group and have essentially the same conformation, two initial runs with an ab initio optimised conformation failed to locate either polymorph. However, they were located as the two most energetically favourable structures within their space group when the initial conformation was obtained by force-field minimisation. Thus, the more accurate model for the gas phase conformer produced worse results than an initial molecular model that was more compatible with the simulation methodology. This study [70] of these two polymorphic pharmaceuticals is particularly useful in showing some of the many pitfalls that could lead an inexperienced, less critical, or time-pressured computational chemist or powder X-ray crystallographer to prematurely dismiss the usefulness and realism of crystal structure prediction calculations.

5.2.4. Estrone

Estrone, a female sex hormone, is another polymorphic steroid, with at least three crystal forms. A Polymorph Predictor study has been reported to have successfully predicted the two conformational polymorphs with one molecule in the asymmetric unit [71], and claimed that the simulation also showed that no further polymorphs are to be expected. This early paper on crystal structure prediction claimed that the low energy structures, representing possible polymorphs, can be used for, by comparison to experimental powder patterns, Rietveld refinement, crystal dynamic simulations to assess thermal stability, prediction of solid state properties, determination of solvent effects on the polymorphic behaviour, etc. Whilst this is still the aim of the computational chemists in the field, further developments in crystal structure and property prediction methodologies are required before this can be done reliably for a wide range of molecules and types of crystals. For example, the differences between the energy minimised and experimental crystal structure are frequently sufficient to make the comparison of their powder patterns non-trivial. Further research from diffraction scientists is required to develop the potential of crystal structure

prediction studies to aid the characterisation of polymorphs from powder X-ray data.

5.3. Use in searching for new polymorphs

5.3.1. Aspirin

The one well-characterised crystal structure of aspirin has been successfully predicted to be the most stable structure in a PROMET search [57] and as a local minimum in a Polymorph Predictor search [72]. A key difference between these studies is that the PROMET search used the experimental non-planar conformation as a rigid molecule, whereas the Polymorph Predictor search modelled the conformational flexibility of this molecule. Since a more stable crystal structure for aspirin was found with a planar conformation, it was speculated [72] that another polymorph of aspirin might be formed if crystallisation conditions could be found that favoured the less stable planar molecular structure. Since there has been considerable debate about whether certain experimental observations demonstrated the existence of a second polymorph, aspirin is likely to prove a scientific headache for both experimentalists and theoreticians for years to come.

5.3.2. Diflunisal

This leads to the use of crystal structure prediction studies in finding new polymorphs. This elegant approach consists of analysing the predicted low energy structures and then selecting solvents to promote the formation of crystals containing the most commonly predicted hydrogen bonding motifs. An example of such a study [73] was performed on diflunisal (a non-steroidal antiinflammatory drug), because it was reported to have at least four polymorphic structures on the basis of X-ray powder data, of which only one structure had been solved. Analysis of the hydrogen bonding graph sets of the low energy structures led to the suggestion that the different types of hydrogen bonding might be associated with the use of acetic acid, acetone, chloroform or aprotic solvents such as toluene, as solvents. A limited experimental screen led to the solution of four new crystal structures, two solvates and two new polymorphs.

5.3.3. Allopurinol

The phenomenon of a variety of crystal structures with different hydrogen bonding networks having

similar lattice energies is not confined to conformationally flexible molecules. An early study on allopurinol, used in the treatment of gout, had the known structure predicted to be 1.8 kJ/mol (<0.5 kcal/mol) more stable than any other structures in a lattice energy minima search using MOLPAK and an accurate distributed multipole electrostatic model [74]. The known structure is based on $\text{N}-\text{H}\cdots\text{N}$ hydrogen bonds, but alternative structures with a $\text{N}-\text{H}\cdots\text{O}$ hydrogen bond were found about 6 kJ/mol (~ 1.5 kcal/mol) less stable, i.e. as thermodynamically feasible polymorphs. Hydrogen bonds are sufficiently strong that, once formed, it is unlikely that there will be a low energy pathway for a solid state polymorphic phase transformation. Hence, if the molecules can pack in a variety of energetically feasible structures with different hydrogen bonding networks, then it is likely that any metastable crystals formed may not readily transform to more stable forms with a different hydrogen bonding network.

6. Conclusions and future prospects

We cannot yet computationally predict the possible polymorphs of any given organic molecule. However, even now, for some molecules, a crystal structure prediction run can give useful information about the likely polymorphism. This depends on the outcome of a thorough search that considers all likely conformers of the molecule performed with a realistic force-field (i.e. when all the factors discussed in Section 2 have been addressed). Some possible outcomes of such an ideally accurate search are illustrated schematically in Fig. 5. These results have different interpretations in terms of what they are predicting, or what further information is necessary.

If the known structure is at the global minimum in the lattice energy, and clearly more stable than the other hypothetical structures found in the search by a margin greater than the energy range of possible polymorphism (Fig. 5a), then this would clearly predict that no polymorphs are to be expected. This could not yet be accepted with sufficient confidence that an experimental polymorph screen was unnecessary, but it would be strong corroboration that a screen that had failed to find any new polymorphs had done so because there were not any to be found.

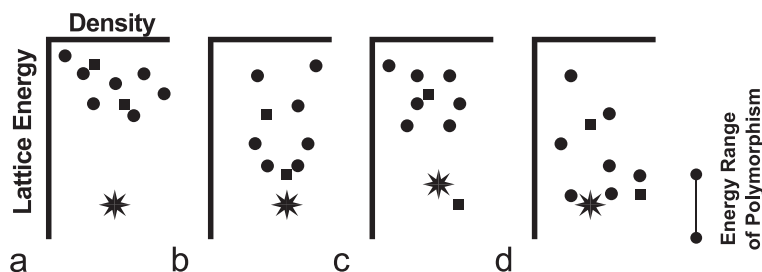


Fig. 5. Some possible distributions of lattice energy minima found in a crystal structure prediction study. * is the lattice energy minimum corresponding to the experimental crystal structure. ■ and ● are hypothetical crystal structures, where each ● denotes a distinct crystal structure, with some similarity in the stronger intermolecular interactions to the experimental structure, and each ■ denotes a distinct crystal structure which is so different from the experimental structure that interconversion is likely to be kinetically prohibited.

If the known structure is at the global minimum of the lattice energy, but only by a small gap (Fig. 5b), then it needs to be established whether the known structure remains the thermodynamically most stable structure at room temperature. Should reliable calculations predict that all the hypothetical structures would be monotonically related to the known form, this implies that all the hypothetical structures are thermodynamically metastable. Such polymorphs would be of less practical importance in pharmaceutical development, because a metastable polymorph that was strongly favoured by kinetics factors would be readily obtained experimentally. However, if we were to develop a computational method that can predict such metastable polymorphs prior to synthesis, the model would need to reflect the kinetic factors involved. Are any of the hypothetical structures likely to be kinetically favoured over the thermodynamic product, by nucleating or growing more readily, so that it could be an observed polymorph? Would such a structure have a sufficient kinetic barrier to transforming to the thermodynamically stable form that it would be sufficiently long-lived to be a practically important polymorph? The answer to these questions will surely depend on whether the hypothetical energetically-feasible structures are similar to the known, global minimum structure. If a metastable structure has the same networks (in graph set terms) of hydrogen bonds and other strong intermolecular contacts, then it is more likely to transform to the global minimum structure and, therefore, unlikely to be a polymorphism problem. Also, if the packing was very similar, the differences in the properties of any undis-

covered metastable structure are less likely to be important. For example, in anhydrous 5-azauracil, the known structure was found at the global minimum [75], and all other structures within 1.5 kcal/mol were based on the same hydrogen bonded sheet. In this situation, there are likely to be low energy pathways that would enable the metastable structures (if they ever formed) to transform to the more stable structure and so no polymorphs are expected.

From the pharmaceutical development point of view, the prediction that there is a hypothetical crystal structure that is significantly more stable than the known structure (Fig. 5c) needs to be considered most seriously. If the predictions are correct, this implies that the known product is metastable, and that if the more stable hypothetical structure could be found, this might lead to problems with producing the known structure. In this case, manufacturing the known structure without a most exhaustive polymorph screen, considering all possible contaminants that might cause nucleation of the hypothetical structure, would be a considerable risk. In the ritonavir case, the analogy was drawn between predicting when a hurricane will strike a certain community and predicting when polymorphism will strike a particular drug [76]. Following this analogy, a graph of type Fig. 5c would correspond to Florida, and Fig. 5a to Hertfordshire, Herefordshire and Hampshire (UK).

The common scenario where there are many minima close in energy (Fig. 5d) is the situation where the results are not much help in predicting polymorphism—though, as discussed above, these results could be useful in designing strategies to find new

polymorphs, and characterising them from powder data. Here, some of the structures might be as yet undiscovered polymorphs, though most would probably readily convert to a more stable structure if they ever nucleated.

Thus, we are at the stage where the ability to predict the polymorphism of a given molecule is unknown until there has been a thorough computational search to establish whether there are energetically feasible alternative structures. From the molecules studied so far, it seems that in many cases it will be necessary to develop computational methods of assessing kinetic factors, as well as improving the thermodynamic modelling. However, we have yet to study a sufficient range of molecules to establish when the polymorphism (or lack thereof) can be predicted just from the zeroth order model of lattice energy minimisation, and to develop the models for the kinetics factors when required. Most fundamentally, we cannot tell whether the computational models are good enough, and how reliable they are, without many more computational studies on systems where we are reasonably confident that all polymorphs are known. We expect considerable progress on this fundamental problem in the future, from academic research combining theoretical and experimental work on polymorphism. However, the pharmaceutical industry can assist in providing polymorph screening data and helping to develop the interpretation of the theoretical calculations. A reliable computational approach has to reflect the factors that cause polymorphism, and so can only be as good as the understanding that emerges from experimental studies.

Acknowledgements

My past and present research group, many collaborators from the European Science Foundation (ESF) Polymorphism Network, and participants in the Cambridge Crystallographic Data Centre blind tests have contributed to developing this exciting and rapidly evolving field through valuable discussions. Some of the scientific challenges raised will be tackled in a Basic Technology Research Project, recently funded by The UK Research Councils with collaboration from the ESF network, CCDC, AstraZeneca, Avecia, and GlaxoSmithKline.

References

- [1] J. Bernstein, *Polymorphism in Molecular Crystals*, Clarendon Press, Oxford, 2002.
- [2] A. Gavezzotti, Are crystal structures predictable? *Acc. Chem. Res.* 27 (1994) 309–314.
- [3] J.D. Dunitz, Are crystal structures predictable? *Chem. Commun.* (2003) 545–548.
- [4] A. Gavezzotti, Structure and intermolecular potentials in molecular crystals, *Model. Simul. Mater. Sci. Eng.* 10 (2002) R1–R29.
- [5] J.P.M. Lommerse, W.D.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Gavezzotti, D.W.M. Hofmann, F.J.J. Leusen, W.T.M. Mooij, S.L. Price, B. Schweizer, M.U. Schmidt, B.P. van Eijck, P. Verwer, D.E. Williams, A test of crystal structure prediction of small organic molecules, *Acta Crystallogr. B* 56 (2000) 697–714.
- [6] W.D.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D.W.M. Hofmann, F.J.J. Leusen, J.P.M. Lommerse, W.T.M. Mooij, S.L. Price, H. Scheraga, B. Schweizer, M.U. Schmidt, B.P. van Eijck, P. Verwer, D.E. Williams, Crystal structure prediction of small organic molecules: a second blind test, *Acta Crystallogr. B* 58 (2002) 647–661.
- [7] T. Beyer, T. Lewis, S.L. Price, Which organic crystal structures are predictable by lattice energy minimisation? *CrystEngComm* 3 (2001) 178–212.
- [8] C.B. Aakeroy, M. Nieuwenhuyzen, S.L. Price, Three polymorphs of 2-amino-5-nitropyrimidine: experimental structures and theoretical predictions, *J. Am. Chem. Soc.* 120 (1998) 8986–8993.
- [9] T. Beyer, S.L. Price, The errors in lattice energy minimisation studies: sensitivity to experimental variations in the molecular structure of paracetamol, *CrystEngComm* 2 (2000) 183–190.
- [10] T. Beyer, G.M. Day, S.L. Price, The prediction, morphology, and mechanical properties of the polymorphs of paracetamol, *J. Am. Chem. Soc.* 123 (2001) 5086–5094.
- [11] Accelrys Inc., Cerius2 Modelling Environment, Accelrys Inc., San Diego, 1999.
- [12] P. Verwer, F.J.J. Leusen, Computer simulation to predict possible crystal polymorphs, in: K.B. Lipkowitz, D.B. Boyd (Eds.), *Reviews in Computational Chemistry*, vol. 12, Wiley-VCH, New York, 1998, pp. 327–365.
- [13] W.T.M. Mooij, B.P. van Eijck, J. Kroon, Ab initio crystal structure predictions for flexible hydrogen-bonded molecules, *J. Am. Chem. Soc.* 122 (2000) 3500–3505.
- [14] A.J. Stone, *The Theory of Intermolecular Forces*, Clarendon Press, Oxford, 1996.
- [15] C. Millot, J.C. Soetens, M. Costa, M.P. Hodges, A.J. Stone, Revised anisotropic site potentials for the water dimer and calculated properties, *J. Phys. Chem. A* 102 (1998) 754–770.
- [16] A. Gavezzotti, in: J.D. Dunitz (Ed.), *Theoretical Aspects and Computer Modeling of the Molecular Solid State*, vol. 1, Wiley, Chichester, 1997.
- [17] S.L. Price, Toward more accurate model intermolecular potentials for organic molecules, in: K.B. Lipkowitz, D.B. Boyd

- (Eds.), *Reviews in Computational Chemistry*, vol. 14, 2000 pp. 225–289.
- [18] A.J. Stone, M. Alderton, Distributed multipole analysis—methods and applications, *Mol. Phys.* 56 (1985) 1047–1064.
- [19] A.D. Buckingham, P.W. Fowler, A.J. Stone, Electrostatic predictions of shapes and properties of van der Waals molecules, *Int. Rev. Phys. Chem.* 5 (1986) 107–114.
- [20] S.L. Price, D.J. Willock, M. Leslie, G.M. Day, DMAREL 3.02 <http://www.ucl.ac.uk/~ucca17p/dmarelmanual/dmarel.html>, 2001.
- [21] A. Stone, A. Dullweber, M.P. Hodges, P.L.A. Popelier, D.J. Wales, ORIENT, University of Cambridge, 1996.
- [22] P.Y. Ren, J.W. Ponder, Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations, *J. Comput. Chem.* 23 (2002) 1497–1506.
- [23] D.E. Williams, Improved intermolecular force field for crystalline hydrocarbons containing four- or three-coordinated carbon, *J. Mol. Struct.* 486 (1999) 321–347.
- [24] D.E. Williams, Improved intermolecular force field for crystalline oxohydrocarbons including O–H···O hydrogen bonding, *J. Comput. Chem.* 22 (2001) 1–20.
- [25] D.E. Williams, Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals, *J. Comput. Chem.* 22 (2001) 1154–1166.
- [26] A.J. Pertsin, A.I. Kitaigorodsky, *The Atom–Atom Potential Method. Applications to Organic Molecular Solids*, Springer, Berlin, 1987.
- [27] G. Filippini, A. Gavezzotti, Empirical intermolecular potentials for organic-crystals—the 6-Exp approximation revisited, *Acta Crystallogr. B* 49 (1993) 868–880.
- [28] A. Gavezzotti, G. Filippini, Geometry of the intermolecular X–H···Y (X, Y=N, O) hydrogen-bond and the calibration of empirical hydrogen-bond potentials, *J. Phys. Chem.* 98 (1994) 4831–4837.
- [29] I. Nobeli, S.L. Price, A non-empirical intermolecular potential for oxalic acid crystal structures, *J. Phys. Chem. A* 103 (1999) 6448.
- [30] Y.A. Arnautova, J. Pillardy, C. Czaplewski, H.A. Scheraga, Global optimisation-based method for deriving intermolecular potential parameters for crystals, *J. Phys. Chem. B* 107 (2003) 712–723.
- [31] W.T.M. Mooij, F.J.J. Leusen, Multipoles versus charges in the 1999 crystal structure prediction test, *Phys. Chem. Chem. Phys.* 3 (2001) 5063–5066.
- [32] W.T.M. Mooij, B.P. van Eijck, J. Kroon, Transferable ab initio intermolecular potentials. 2. Validation and application to crystal structure prediction, *J. Phys. Chem. A* 103 (1999) 9883–9890.
- [33] W.T.M. Mooij, F.B. van Duijneveldt, J. van Duijneveldt-van de Rijdt, B.P. van Eijck, Transferable ab initio intermolecular potentials. 1. Derivation from methanol dimer and trimer calculations, *J. Phys. Chem. A* 103 (1999) 9872–9882.
- [34] B.P. van Eijck, W.T.M. Mooij, J. Kroon, Ab initio crystal structure predictions for flexible hydrogen-bonded molecules. Part II. Accurate energy minimisation, *J. Comput. Chem.* 22 (2001) 805–815.
- [35] M. Tremayne, L. Grice, J.C. Pyatt, C.C. Seaton, B.M. Kariuki, H.H.Y. Tsui, S.L. Price, J.C. Cherryman, Characterisation of complex new polymorphs of chlorothalonil by X-ray diffraction and computer crystal structure prediction (submitted for publication).
- [36] A. Gavezzotti, Ten years of experience in polymorph prediction: what next? *CrystEngComm* 4 (2002) 343–347.
- [37] A. Gavezzotti, Calculation of intermolecular interaction energies by direct numerical integration over electron densities. 2. An improved polarization model and the evaluation of dispersion and repulsion energies, *J. Phys. Chem. B* 107 (2003) 2344–2353.
- [38] F.H. Allen, The Cambridge Structural Database: a quarter of a million crystal structures and rising, *Acta Crystallogr. B* 58 (2002) 380–388.
- [39] T. Steiner, Frequency of Z' values in organic and organometallic crystal structures, *Acta Crystallogr. B* 56 (2000) 673–676.
- [40] B.P. Van Eijck, Crystal structure predictions using five space groups with two independent molecules. The case of small organic acids, *J. Comput. Chem.* 23 (2002) 456–462.
- [41] B.P. van Eijck, J. Kroon, Structure predictions allowing more than one molecule in the asymmetric unit, *Acta Crystallogr. B* 56 (2000) 535–542.
- [42] F.J.J. Leusen, Crystal structure prediction of diastereomeric salts: a step toward rationalization of racemate resolution, *Cryst. Growth Des.* 3 (2003) 189–192.
- [43] D.E. Williams, Ab initio molecular packing analysis, *Acta Crystallogr. A* 52 (1996) 326–328.
- [44] B.P. van Eijck, J. Kroon, UPACK program package for crystal structure prediction: force fields and crystal structure generation for small carbohydrate molecules, *J. Comput. Chem.* 20 (1999) 799–812.
- [45] J. Pillardy, Y.A. Arnautova, C. Czaplewski, K.D. Gibson, H.A. Scheraga, Conformation-family Monte-Carlo: a new method for crystal structure prediction, *Proc. Natl. Acad. Sci. USA* 98 (2001) 12351–12356.
- [46] A. Gavezzotti, Generation of possible crystal-structures from the molecular-structure for low-polarity organic-compounds, *J. Am. Chem. Soc.* 113 (1991) 4622–4629.
- [47] J.R. Holden, Z.Y. Du, H.L. Ammon, Prediction of possible crystal-structures for C-, H-, N-, O- and F-containing organic-compounds, *J. Comput. Chem.* 14 (1993) 422–437.
- [48] M.U. Schmidt, Energy minimisation as a tool for crystal structure determination of industrial pigments, in: D. Braga, F. Grepioni, A.G. Orpen (Eds.), *Crystal Engineering: From Molecules and Crystals to Materials*, vol. 538, Kluwer Academic Publishers, Dordrecht, 1999, pp. 331–348.
- [49] W.T.M. Mooij, B.P. van Eijck, S.L. Price, P. Verwer, J. Kroon, Crystal structure predictions for acetic acid, *J. Comput. Chem.* 19 (1998) 459–474.
- [50] R.S. Payne, R.J. Roberts, R.C. Rowe, R. Docherty, Generation of crystal structures of acetic acid and its halogenated analogs, *J. Comput. Chem.* 19 (1998) 1–20.
- [51] R.K.R. Jetti, R. Boese, J.A.R.P. Sarma, L.S. Reddy, P. Vishweshwar, G.R. Desiraju, Searching for a polymorph: second crystal form of 6-amino-2-phenylsulfonylimino-1,2-dihydropyridine, *Angew. Chem. Int. Ed.* 42 (2003) 1963–1967.

- [52] J. Bernstein, J.O. Henck, Disappearing and reappearing polymorphs—an anathema to crystal engineering? *Mater. Res. Bull.* 33, Suppl. 1 (1998) 119–128.
- [53] J. Bernstein, R.J. Davey, J.O. Henck, Concomitant polymorphs, *Angew. Chem. Int. Ed.* 38 (1999) 3441–3461.
- [54] R.J. Davey, K. Allen, N. Blagden, W.I. Cross, H.F. Lieberman, M.J. Quayle, S. Righini, L. Seton, G.J.T. Tiddy, Crystal engineering-nucleation, the key step, *CrystEngComm* 4 (2002) 257–264.
- [55] L. Yu, G.A. Stephenson, C.A. Mitchell, C.A. Bunnell, S.V. Snorek, J.J. Bowyer, T.B. Borchardt, J.G. Stowell, S.R. Byrn, Thermochemistry and conformational polymorphism of a hexamorphic crystal system, *J. Am. Chem. Soc.* 122 (2000) 585–591.
- [56] D. Buttar, M.H. Charlton, R. Docherty, J. Starbuck, Theoretical investigations of conformational aspects of polymorphism. Part 1: *O*-acetamidobenzamide, *J. Chem. Soc. Perkin Trans. 2* (1998) 763–772.
- [57] A. Gavezzotti, G. Filippini, Polymorphic forms of organic-crystals at room conditions—thermodynamic and structural implications, *J. Am. Chem. Soc.* 117 (1995) 12299–12305.
- [58] J.D. Dunitz, G. Filippini, A. Gavezzotti, Molecular shape and crystal packing: a study of $C_{12}H_{12}$ isomers, real and imaginary, *Helv. Chim. Acta* 83 (2000) 2317.
- [59] B.P. van Eijck, Ab initio crystal structure predictions for flexible hydrogen-bonded molecules. Part III. Effect of lattice vibrations, *J. Comput. Chem.* 22 (2001) 816–826.
- [60] A.T. Anghel, G.M. Day, S.L. Price, A study of the known and hypothetical crystal structures of pyridine: why are there four molecules in the asymmetric unit cell? *CrystEngComm* 4 (2002) 348–355.
- [61] T.C. Lewis, D.A. Tocher, G.M. Day, S.L. Price, A computational and experimental search for polymorphs of parabanic acid—a salutary tale leading to the crystal structure of oxo-ureido-acetic acid methyl ester, *CrystEngComm* 5 (2003) 3–9.
- [62] S.R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubaue, B.A. Narayanan, M. Soldani, D. Riley, K. McFarland, Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development, *Org. Process Res. Dev.* 4 (2000) 413–417.
- [63] J. Sarma, G.R. Desiraju, The supramolecular synthon approach to crystal structure prediction, *Cryst. Growth Des.* 2 (2002) 93–100.
- [64] G.M. Day, S.L. Price, M. Leslie, Elastic constant calculations for molecular organic crystals, *Cryst. Growth Des.* 1 (2001) 13–26.
- [65] G.M. Day, S.L. Price, Properties of crystalline organic molecules, in: M. Levy (Ed.), *Handbook of Elastic Properties of Solids, Liquids and Gases. Volume III: Elastic Properties of Solids: Biological and Organic Materials, Earth and Marine Sciences*, Academic Press, New York, 2001, pp. 3–49.
- [66] M. Brunsteiner, S.L. Price, Morphologies of organic crystals: sensitivity of attachment energy predictions to the model intermolecular potential, *Cryst. Growth Des.* 1 (2001) 447–453.
- [67] G. Nichols, C.S. Frampton, Physicochemical characterisation of the orthorhombic polymorph of paracetamol crystallized from solution, *J. Pharm. Sci.* 87 (1998) 684–693.
- [68] P. Di Martino, P. Conflant, M. Drache, J.-P. Huvenne, A.-M. Guyot-Herman, Preparation and physical characterisation of form II and III of paracetamol, *J. Therm. Anal.* 48 (1997) 447–458.
- [69] M.L. Peterson, S.L. Morissette, C. McNulty, A. Goldsweig, P. Shaw, M. LeQuesne, J. Monagle, N. Encina, J. Marchionna, A. Johnson, J. Gonzalez-Zigasti, A.V. Lemmo, S.J. Ellis, M.J. Cima, O. Almarsson, Iterative high-throughput polymorphism studies on acetaminophen and an experimentally derived structure for form III, *J. Am. Chem. Soc.* 124 (2002) 10958–10959.
- [70] R.S. Payne, R.J. Roberts, R.C. Rowe, R. Docherty, Examples of successful crystal structure prediction: polymorphs of primidone and progesterone, *Int. J. Pharm.* 177 (1999) 231–245.
- [71] F.J.J. Leusen, Ab initio prediction of polymorphs, *J. Cryst. Growth* 166 (1996) 900–903.
- [72] R.S. Payne, R.C. Rowe, R.J. Roberts, M.H. Charlton, R. Docherty, Potential polymorphs of aspirin, *J. Comput. Chem.* 20 (1999) 262–273.
- [73] W.I. Cross, N. Blagden, R.J. Davey, R.G. Pritchard, M.A. Neumann, R.J. Roberts, R.C. Rowe, A whole output strategy for polymorph screening: combining crystal structure prediction, graph set analysis and targeted crystallisation experiments in the case of diflunisal, *Cryst. Growth Des.* 3 (2003) 151–158.
- [74] S.L. Price, K.S. Wibley, Predictions of crystal packings for uracil, 6-azauracil, and allopurinol: the interplay between hydrogen bonding and close packing, *J. Phys. Chem. A* 101 (1997) 2198–2206.
- [75] B.S. Potter, R.A. Palmer, R. Withnall, B.Z. Chowdhry, S.L. Price, Aza analogues of nucleic acid bases: experimental determination and computational prediction of the crystal structure of anhydrous 5-azauracil, *J. Mol. Struct.* 486 (1999) 349–361.
- [76] E. Sun, Excerpt from news conference, 15.10.98. IAPAC, International Association of Physicians in AIDS Care, 225 West Washington St., Suite 2200, Chicago, IL 60606, USA, 1998 (iapac@iapac.org: <http://www.iapac.org>, http://www.iapac.org/norvir/abt_sun.html).