From crystal structure prediction to polymorph prediction: interpreting the crystal energy landscape

Sarah L. Price*

Received 14th December 2007, Accepted 15th January 2008
First published as an Advance Article on the web 19th February 2008
DOI: 10.1039/b719351c

Many organic molecules are emerging as having many crystalline forms, including polymorphs and solvates, as more techniques are being used to generate and characterise the organic solid state. The fundamental scientific and industrial interest in controlling crystallisation is inspiring the development of computational methods of predicting which crystal structures are thermodynamically feasible. Sometimes, computing this crystal energy landscape will reveal that a molecule has one way of packing with itself that is sufficiently more favourable than any other so only this crystal structure will be observed. More frequently, there will be many energy minima that are energetically feasible, showing approximately equi-energetic compromises between the various intermolecular interactions allowed by the conformational flexibility. Such cases generally lead to multiple solid forms. At the moment, we usually calculate the lattice energy landscape, an approximation to the real crystal energy landscape at 0 K. Despite its limitations, many studies show that this is a valuable complement to solid form screening, which helps in discovering new structures as well as rationalising the solid forms that are found in experimental searches. The range of factors that can determine which of the thermodynamically feasible crystal structures are observed polymorphs, shows the many further challenges in developing crystal energy landscapes as a tool for control of the organic solid state.

1. Why computational crystal structure prediction?

The original aim of crystal structure prediction (CSP) was to find a computational method of going from the chemical diagram to predict the crystal structure of an organic molecule. The practical motivation for a method of predicting the crystal structure prior to the synthesis of the molecule was to guide the design of new organic materials where the physical property of interest was very sensitive to the crystal structure. For example, for non-linear optical materials, the molecule must pack in a non-centrosymmetric space group for the crystal to be active, or in the design of energetic materials where the molecule must pack with sufficient density for an effective explosive. Progress and interest in CSP became sufficiently widespread over the first decade in which computer power allowed such a vision, that the Cambridge Crystallographic Data Centre has been organising blind tests of whether crystal structures can be predicted from the molecular diagram. The results (Fig. 1) of the blind tests in 1999,[1] 2001[2] and 2004[3] showed that it was sometimes possible, and quickly advanced the understanding of the challenge. The most successful and widely used hypothesis was that the crystal structure would be the most thermodynamically stable one, and considerable effort went into increasing the accuracy of the relative energies of the crystal structures that were minima in the lattice (or more recently free) energy, as well as developing various strategies for searching through the range of possible crystal structures. The success rate in the most recent blind test,[4] shows that these four target crystal structures (XII–XV Fig. 1) can be predicted provided that you combine a sufficiently wide search with either state-of-the-art models for the inter- and intra-molecular forces, or specifically adapted electronic structure methods,[5] to compute the relative lattice energies.

The qualification to "these target crystal structures" in the statement above reflects the simultaneous advance in our experimental understanding of the organic solid state. There has been a growing awareness of the possibility that a molecule can adopt more than one crystal structure, and exhibit polymorphism.[6] Most of the successes in the first blind test in

Department of Chemistry, University College London, 20 Gordon Street, London, UK WC1H 0AJ. E-mail: s.l.price@ucl.ac.uk; Fax: +44 (0)20 7679 7463; Tel: +44 (0)20 7679 4622

Sarah (Sally) Price studied Natural Sciences at the University of Cambridge, obtaining a PhD in Theoretical Chemistry in 1981. After a postdoctoral year at the University of Chicago, she did further postdoctoral work at the University of Cambridge, held a Royal Society University Research fellowship, and moved to University College London as a lecturer in 1989. There she has been promoted to Professor and is applying the theory of intermolecular forces to the computational modelling of the organic solid state.

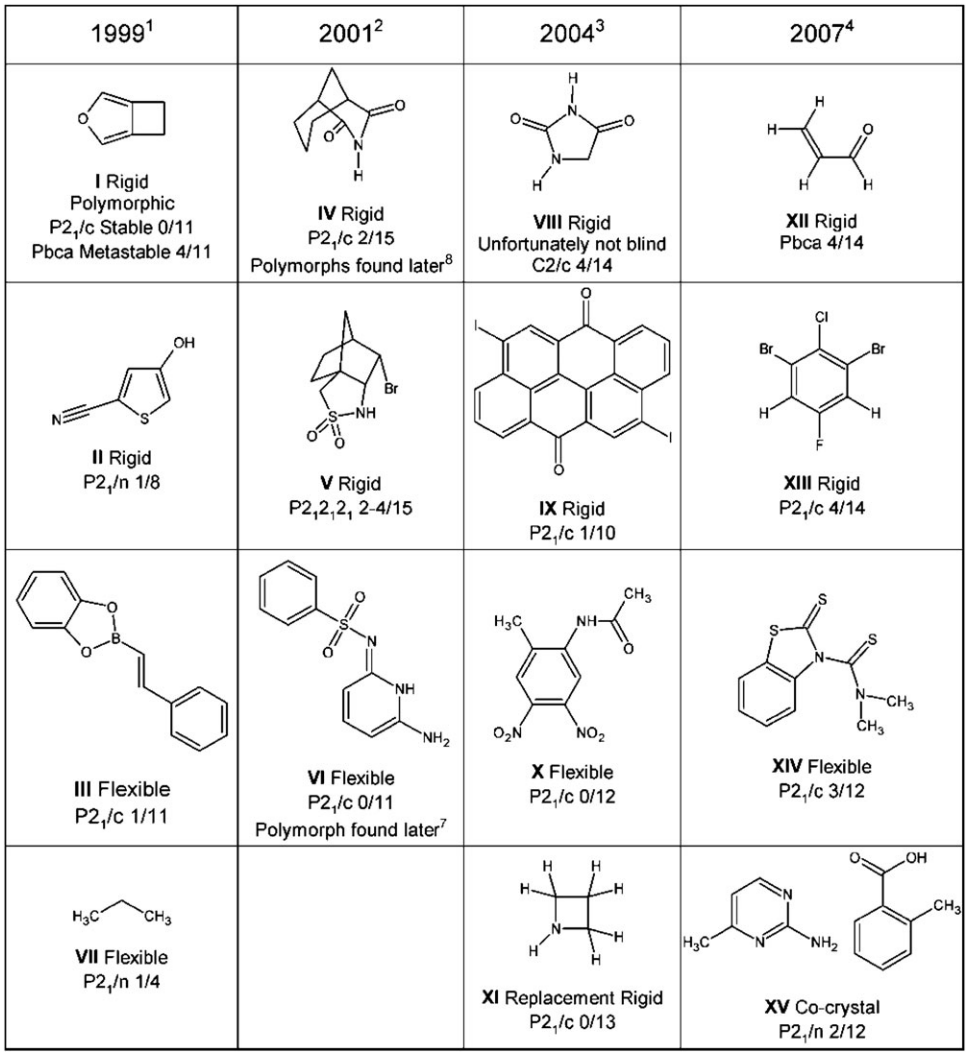

Fig. 1 The success rates for predicting the crystal structures of these molecules from the chemical diagram in the Cambridge Crystallographic Data Centre's blind tests of crystal structure prediction. *x*/*y* indicates that there were *x* correct predictions out of *y* groups submitting three predictions for the molecule.

1999 were for the metastable polymorph of I, with no successful predictions for the alternative structure found in subsequent attempts to regrow this crystal. Following the 2001 blind test, in which most participants were predicting structures with alternative hydrogen bonding motifs to the observed structures for IV and VI, polymorphs of both have been found,[7,8] albeit only one with the anticipated different hydrogen bonding motif.[7] If a structure is polymorphic, only one of the polymorphs can be the most thermodynamically stable crystal structure at a given temperature and pressure, and this is not necessarily the one whose crystal structure is the first one solved by X-ray diffraction. This is clearly the case when polymorphs "disappear",[9] where it suddenly proves exceptionally difficult to recrystallise a given crystal structure once a more stable polymorph is found. Indeed Ostwald's rule of stages[10] proposes that the first solid crystallised from the melt or solution would be the least stable form. Although this is not a universal law, its rationalisation[11] emphasises the subtleties of crystallisation and the tendency for metastable polymorphs to be the first found.

Polymorphism[6] has become a huge interdisciplinary research area, because of its importance in the quality control of all organic materials industries. The pharmaceutical industry[12] is concerned with polymorphism because different polymorphs may have different dissolution rates and hence bioavailability. Pharmaceuticals are licensed in a specific physical form and hence polymorphism has been at the centre of expensive patent litigation. The most publicised polymorphism problem was the urgent reformulation of the anti-HIV drug ritonavir when, almost two years into manufacture, the process suddenly produced a new, more stable polymorph.[13] Since then, one of the emerging specialist solid form discovery companies has reported finding a novel metastable polymorph, a hydrate phase and a formamide solvate of ritonavir.[14] The industrial requirement to know all the potentially marketable solid forms of their active drug molecule has seen a reappraisal of the prevalence of polymorphism. An analysis[15] of a series of 245 polymorph screens performed by another solid form screening company reported that about 50% exhibited polymorphism and 90% multiple solid forms, *i.e.*

polymorphs, hydrates and other solvates and noncrystalline forms. A similar analysis of the industrial pigments, whose insoluble molecules are generally chemically different from pharmaceuticals, indicated[16] about 80% are polymorphic. These estimates of the incidence of polymorphism are far more realistic than those derived from solved crystal structures that have been deposited in the Cambridge Structural Database (CSD).[17] Indeed, less than 1% of compounds with a high quality crystal structure in the CSD have a similar standard of structural data for a polymorph.[18] This reflects the great difficulty in growing suitable single crystals for diffraction studies for many polymorphs, as well as the historical bias towards just solving the structure of the first suitable crystal that could be obtained. Thus, as the informal debate has switched from "What molecules are polymorphic?" to "Which molecules are we confident cannot show polymorphism?," the aim of CSP has switched from trying to predict "the" crystal structure, to predicting the possible polymorphs of a given organic molecule.

A method of predicting "all" the polymorphs and the conditions under which they would crystallise would be a very desirable aid to solid form development for the molecular materials industries. It would also provide the intellectual satisfaction of knowing that the factors that control crystallisation are incorporated in the computational model. It is not possible to develop a computational method of polymorph prediction without being able to contrast the results with the definitive crystal structures of all the polymorphs found in an extensive and systematic experimental search. I've been privileged to be able to work on developing a computational method of polymorph prediction in collaboration with colleagues developing an automated medium throughput solvent crystallisation system[19] and a variety of other techniques specifically aimed at providing insight into polymorphism. Contrasting our computed lattice energy landscape with the range of crystal structures found in an experimental search for polymorphs, over a range of molecules, provides this perspective into the possibilities for control and prediction of the organic solid state.

2. The ideal crystal energy landscape—what it is?

In a future ideal world, the first step in polymorph prediction would be to calculate the real crystal energy landscape, the energy hypersurface as a function of all coordinates defining the possible crystal structures. This needs to be the appropriate free energy surface which should also be a function of the thermodynamic variables within the conditions of interest (*e.g.* practically accessible range of pressures and temperatures). Only the lower energy regions of the crystal energy landscape are of interest. The minima on this surface within the energy range of plausible polymorphism are the thermodynamically feasible structures. The barriers to transformations between these minima are likely to play an important role in the second step of determining which minima will actually be observed polymorphs. In section 5, we discuss the scientific challenges that would be involved in computing this crystal energy surface, and then in interpreting it to predict the possible polymorphs. However, before that we discuss the potential predictive value of crystal energy landscapes, based on results using currently feasible crude approximations to the crystal energy landscape: the set of lattice energy minima within a plausible energy range of the global minimum. This is a partial approximate lattice energy landscape, which neglects the effects of temperature and pressure, given in an incomplete search (restricted by space group and number of molecules in the asymmetric unit), using an approximate method of energy calculation. Thus, this perspective contrasts the ability to predict structures and the insights into polymorphism which have been obtained from just a partial approximate *lattice energy* landscape with those that could be obtained as we move towards being able to calculate the real *crystal energy* landscape.

The structures and relative energies of the crystal structures computed to be within the energy range of polymorphism are the most important features of the crystal energy landscape. In practice, the observation that for most crystals there are many lattice energy minima within a few kJ mol^{-1} of the global minimum has both been a major cause for doubt as to whether crystal structures are predictable,[20–24] and a challenge to account for why there are not more polymorphs observed. One explanation for this is that approximating the crystal energy landscape (which will depend on pressure and temperature) by the static 0 K lattice energy landscape is producing too many minima. Thermal motion averages over sets of lattice energy minima that are separated by small barriers to give far fewer minima in the free energy surface. Certainly, this appears to be true for benzene, which has dozens of plausible lattice energy minima, but a metadynamics study[25] to explore the free energy minima found only seven, corresponding to the observed phases. In contrast, only a quarter of the 60 structures within 8 kJ mol^{-1} of the global minimum in the lattice energy of 5-fluorouracil proved to be thermally unstable at room temperature and atmospheric pressure.[26] The free energy minima corresponded to a variety of packings of the distinct hydrogen bonded motifs observed in the polymorphs, solvate and cocrystals of 5-fluorouracil and closely related molecules. Thus, the extent to which there are many fewer minima on the crystal energy than lattice energy landscape is clearly very dependent on the strength and directionality of the intermolecular forces providing large barriers on the crystal energy surface relative to thermal energies. Crystal structures with different hydrogen-bond motifs are more likely to result in different free energy minima at ambient conditions than structures related by sterically undemanding rearrangements of molecules bound by less directional dispersion forces.

The type of inter- and intramolecular bonding of the specific molecule will also determine the energy range of possible polymorphism, and thus the range of the crystal energy landscape that needs to be calculated and considered.

Considering the minima that lie within around 10 kJ mol^{-1} of the global minimum is probably a generous allowance for the thermodynamic driving force,[6] based on rather limited experimental evidence of polymorphic energy differences. In practice, it is the barriers to transformation between polymorphic structures that will determine the degree of thermodynamic metastability possible for long-lived polymorphs. Thus higher energy differences between polymorphs are likely

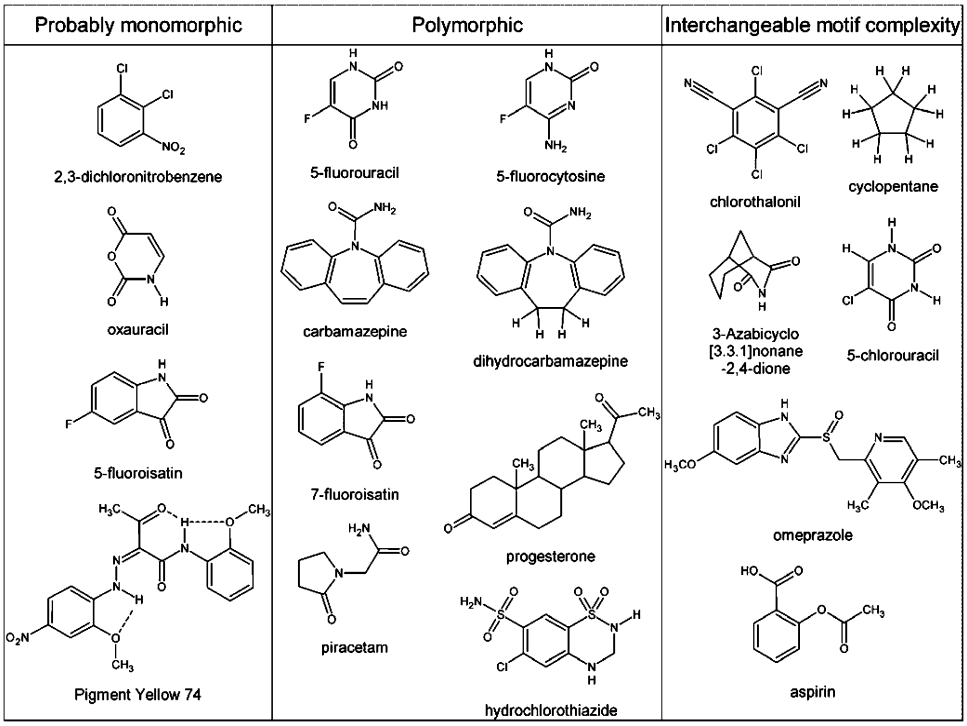

Fig. 2 Molecular diagrams of molecules mentioned in the text. These are tentatively classified according to their type of crystal energy landscape, on the basis of their known crystal structures and lattice energy landscapes.

for conformational polymorphs,[27] when the barriers for changing the molecular conformation in the solid or solution are larger.

Hence, we will proceed by considering conceptually different types of crystal energy landscape, mainly working with the simplified pictorial representation of just the relative energies of the minima and their densities, and considering the structural variations within this set of energy minimised structures. The implication for polymorphism of different extreme types of crystal energy landscape will be considered in the following sections, assuming that the qualitative type does not change with thermodynamic conditions. (Extreme counter-examples will be discussed, but not implicit phase changes or the implications of many intermediate and mixed types of landscapes.) The extreme types of crystal energy landscapes will be illustrated by deductions from the lattice energy landscapes of various molecules (Fig. 2). This comparison is necessarily limited by the knowledge that many lattice energy minima will not be free energy minima, as well as the uncertainty in the relative energies of the different crystal structures.

3. Interpretation of crystal energy landscapes

3.1 Monomorphic crystal energy landscapes

The crystal energy landscape shown in Fig. 3a has the known crystal structure at the global minimum in energy, and no others within the energy range of possible polymorphism. This is a clear prediction that there is only one crystal structure. It is the type of energy landscape that was initially expected to be common, but in practice appears rare, at least for the simple molecules that have been the subject of CSP studies. One example, where the computed energy gap is 12 kJ mol^{-1} is for Pigment Yellow 74, which indeed is monomorphic.[28] The molecule is held rigid by internal hydrogen bonds in a planar shape with distinct bumps and hollows, which happen to pack nicely into each other and give a dense close-packed sheet. The polar atoms produce strong electrostatic forces between the sheets, and the crystal energy landscape confirms that there is one uniquely favourable way of stacking these sheets. This example illustrates that to have one crystal structure that is significantly more stable than any other possibilities, it is necessary for the molecule to be able to close pack to give a dense structure, *and* have strongly directional intermolecular interactions defining the structure *in all three dimensions*. This

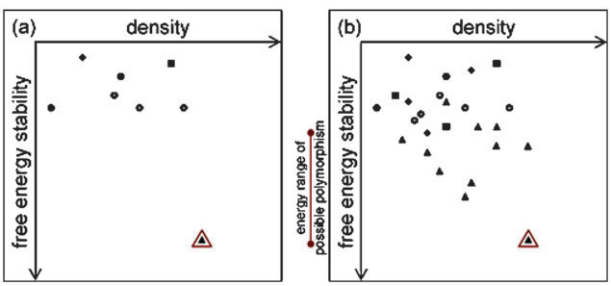

Fig. 3 Idealised monomorphic crystal energy landscapes: (a) clearly monomorphic, as only one structure within the energy range of possible polymorphism, and (b) probably monomorphic. Open symbols around a point indicate that the structure is experimentally known. Structures with the same symbol are closely related, in the sense that it is likely to be practically impossible to prevent transformation to the most stable structure with that symbol during the process of nucleation and growth.

may be relatively rare: when crystal engineers[29–31] design molecular crystals with strong interactions in all three dimensions then the resulting network often includes solvent to provide the dense packing.

Possibly the most common crystal energy landscape is shown in Fig. 3b. The known form is the structure at the global minimum in free energy, but the energy gap is too small to rule out the possibility of further polymorphs, though they would be metastable. In this case, looking at the low energy structures in terms of their intermolecular interactions can be used to make a qualitative assessment of whether it is likely that these polymorphs could be trapped and practicably significant. This would confirm that a polymorph screen that had not found any polymorphs was likely to be complete. For example the 4 kJ mol^{-1} energy gap in the lattice energy landscape of 3-oxauracil[32] meant that it would have been a failure of the CSP method if the global minimum had not been the observed structure. However, all the structures within 6 kJ mol^{-1} of the global minimum were based on the same doubly hydrogen bonded dimer, and so it seemed plausible that these dimers would have the opportunity of moving into the most stable structure during the nucleation process, even if the crystal structures corresponded to free energy minima at ambient conditions. Hence, although neither the calculation of the energy landscape nor the limited crystallisation screen that was carried out would in themselves be sufficient for confidence that 3-oxauracil was not polymorphic, the combination gives considerable confidence that practically significant, long-lived metastable polymorphs are unlikely.

3.2 "Predictive" crystal energy landscape

The energy landscape which confidently predicts that there is an alternative crystal structure that is more thermodynamically stable than the known form (Fig. 4) is perhaps the most useful as showing that the practically most important polymorph has yet to be found. It must be different in either conformation or intermolecular interactions (*i.e.* separated from the known form by significant barriers on the crystal energy landscape) or it would have been found in thermal analysis, slow crystallisation or slurrying experiments. The predicted structure could be used to design appropriate crystallisation strategies[33] to target obtaining it. This sort of warning could help avoid a ritonavir-type[13] quality control and developmental crisis.

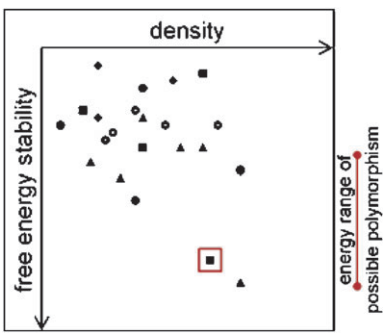

Fig. 4 Predictive crystal energy landscape. The open symbol indicates the only experimentally known structure. Structures with different symbols are sufficiently distinct that there will be a significant activation energy barrier for a solid state conversion.

5-Fluorouracil[34] provides an example of this type of lattice energy landscape leading to the discovery of the polymorph predicted as the most stable structure. The only previously known structure of this anti-cancer agent had an unusual hydrogen bonding motif, in which the most striking feature was close F···F contacts. The crystal energy landscape had many different structures based on two different hydrogen bonded ribbons (Fig. 5a). An extensive manual screen eventually crystallised the predicted structure,[34] but form 2 was only found from dry nitromethane. The difficulty in forming this polymorph was later rationalised by molecular dynamics simulations.[35] Water so strongly hydrates the N–H and C=O groups of 5-fluorouracil, that the initial aggregation of two molecules in water is usually through a close contact between the hydrophobic F atoms, as seen in form 1. Even when one N–H···O=C bond is formed between two 5-fluorouracil molecules, the hydrating water is not readily displaced to allow the formation of the second hydrogen bond, though this would readily form in the gas phase or in nitromethane giving the doubly hydrogen bonded ribbon seen in form 2. This is an example of the growing awareness that solution chemistry can be vital in determining polymorph formation. Some FTIR[36] and NMR[37] studies have demonstrated that initial aggregation in solution correlates with certain polymorphic forms, though in other cases the solute–solute interactions detected in solution have a more limited relationship to the resulting crystal structure.[38] The crystal energy landscape can contribute by showing whether thermodynamics allows the possibility of crystal structures with very different hydrogen-bonded or conformational motifs, and hence suggest further investigations targeting different initial association modes.[33]

Two contrasting case studies, where the lattice energy landscape led to the targeting of predicted structures with alternative hydrogen bonding motifs (Fig. 5), illustrate the current potential and problems arising from apparently predictive crystal energy landscapes (*cf.* Fig. 4) where solvents might be expected to influence the hydrogen-bonding motif. 3-Azabicyclo[3.3.1]nonane-2,4-dione (IV Fig. 1) was a target in the 2001 blind test of CSP.[2] The majority of participants predicted structures based on the doubly hydrogen bonded dimer (Fig. 5b, right), and even the participants who had the correct catemeric structure within their 3 guesses, predicted a dimer based structure to be more stable. A wide range of crystallisation methods was used to target a dimer based polymorph,[8] including an automated screen which produced a metastable catemeric polymorph and acetic acid and methylnaphthalene solvates. This was consistent with FTIR solution studies showing no evidence that the molecule formed the hydrogen bonded dimer, even in the non-polar solvents. The discovery that this molecule has a plastic phase above 135 °C, clearly showed that we were wrong in assuming that the N–H···O=C hydrogen bonds produced a significant barrier to changing the hydrogen bonding motif. This was confirmed by gas-phase simulations showing that a third molecule approaching a doubly hydrogen-bonded dimer could result in a trimer with the catemer motif. Thus, it appears[8] that the weakness of the

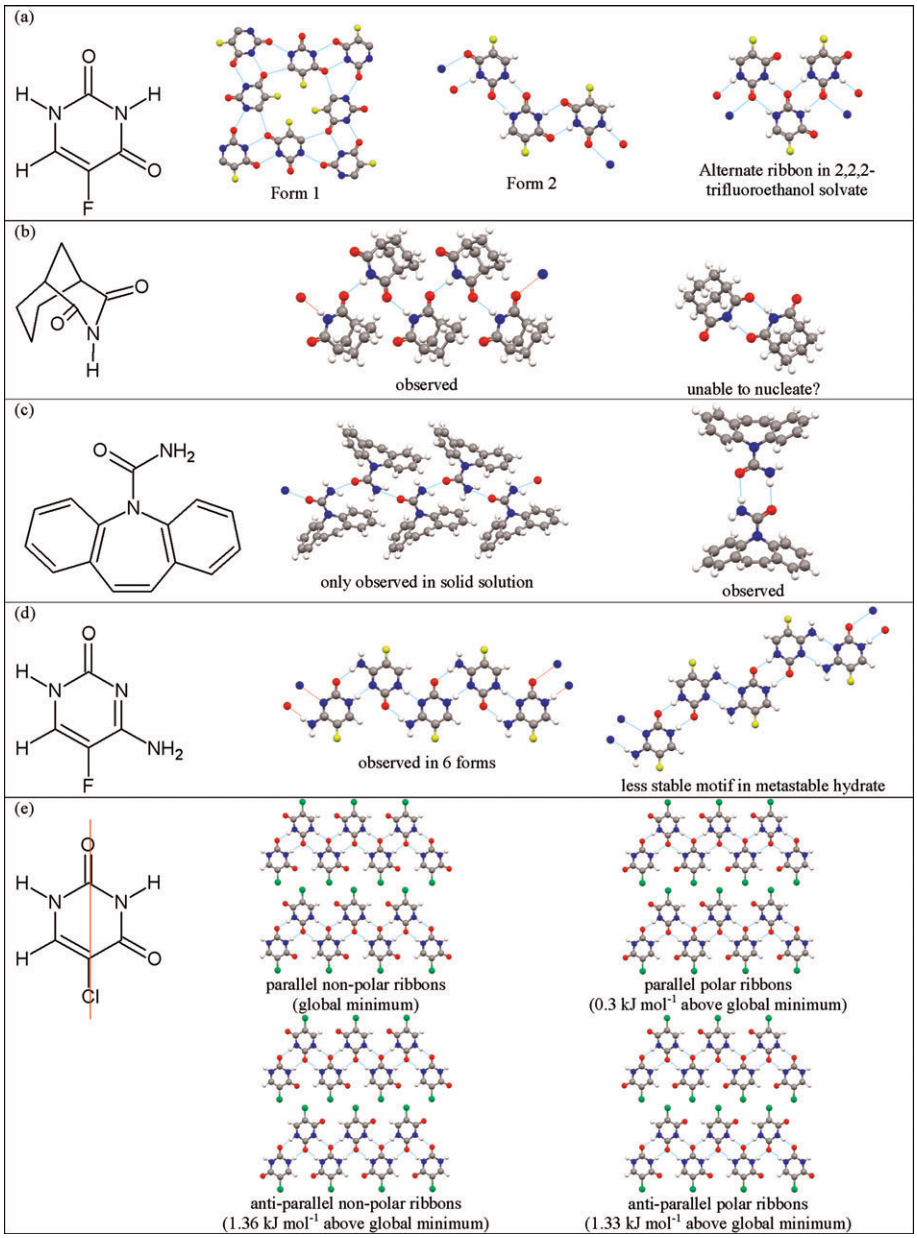

Fig. 5 Competitive hydrogen bonding motifs found in the crystal energy landscapes of (a) 5-fluorouracil, where the lattice energy range[34] of these motifs is ~6 kJ mol^{-1} (b) 3-azabicyclo[3.3.1]nonane-2,4-dione where dimer based structures are close in energy to the observed catemer polymorphs[8] (c) carbamazepine, where the chain structure is competitive[39] with the dimer-based most stable known form III (d) 5-fluorocytosine, where all structures within 8 kJ mol^{-1} had the left-hand motif (e) chlorouracil, where all four ribbons shown can interdigitate and stack to form structures that would be identical if the molecule was symmetric about the marked axis, within 1.4 kJ mol^{-1}.

N–H donor within the imide group, coupled with the globular shape of the molecule, meant that the barriers to the nucleating cluster rearranging to the most stable catemeric structure were too small to allow the trapping of a dimer based polymorph.

For carbamazepine, the computed crystal energy landscape[39] predicted that a catemer based structure was slightly more stable than the most stable of the four known polymorphs, all of which are based on the dimer motif (Fig. 5c). (Form II has recently been shown to contain solvent molecules[40,41] accounting for some anomalies with this polymorph.) An extensive crystallisation screen[39] did not find a catemeric polymorph, although it did find three new solvates. A retrospective statistical analysis of the database correlating the crystallisation conditions with solid form produced in this screen led to the discovery of an additional three solvates.[42] Finding 6 novel solvates in addition to the 6 solvates previously known for this heavily studied antiepileptic, confirms that the lattice energy landscape is showing that carbamazepine has problems in packing with itself which can be relieved by incorporating a range of other molecules. A 50 : 50 solid solution of dihydrocarbamazepine with carbamazepine[43] is isomorphous with a low energy predicted chain structure of carbamazepine,[39] implying that carbamazepine can form a

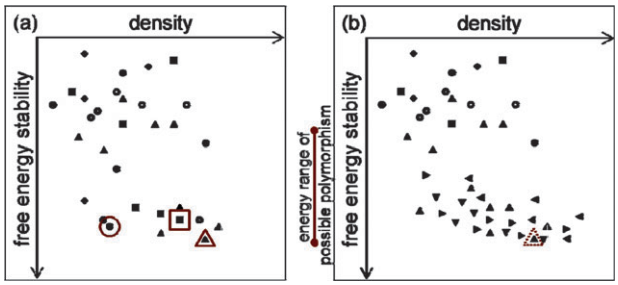

Fig. 6 Polymorphic crystal energy landscapes: (a) simple polymorphism with very distinct polymorphs, and (b) interchangeable polymorphic landscape. Open symbols indicate the (dominant component of) experimentally known polymorphs. Structures with the different symbols are sufficiently distinct that there will be a significant activation energy barrier for a solid state conversion, but those whose symbols are related by rotation have a common interchangeable component, such as a hydrogen-bonded sheet or ribbon.

hydrogen bonded catemer. Thus, whilst it is clear[44] that catemer formation is far more difficult than dimer formation for carbamazepine, the possibility of a catemeric polymorph being formed under as yet undetermined conditions sometime in the future cannot be excluded.

Whilst the prediction of a more stable structure with an alternative hydrogen bonding motif for 5-fluoruracil, 3-azabicyclo[3.3.1]nonane-2,4-dione and carbamazepine have all led to the discovery of new forms, the experimental evidence[8,34] or more accurate calculations[45,46] suggest that all the targeted structures are probably slightly less stable than the previously known forms. Hence in retrospect, their real crystal energy landscapes should have been ''polymorphic'' as depicted in Fig. 6. However, a crystal energy landscape of the ''predictive'' type should be turned into a ''polymorphic'' type with the discovery of the predicted form, unless the range of crystallisation methods that can be practically applied makes the targeted structure unobtainable, for example if the compound is highly insoluble, reactive or heat sensitive.

3.3 "Polymorphic" crystal energy landscapes

When there are many crystal structures that are within a small energy range (Fig. 6a), then the molecule clearly has a range of different ways of packing that are effectively equally poor compromises between the different intermolecular interactions. In this case, looking at the thermodynamically competitive structures can be illuminating, particularly since such landscapes generally correlate with the occurrence of polymorphs and solvates, as has already been found for carbamazepine, 5-fluorouracil and 3-azabicyclo[3.3.1]nonane-2,4-dione. One simple example is the energy landscape of 5-fluorocytosine, which showed many different packings of the same hydrogen bonded ribbon (Fig. 5d left),[47] demonstrating a strongly preferred two-dimensional motif that did not have a uniquely good way of packing with itself. The subsequent manual crystallisation screen found two polymorphs, two novel hydrates, and two solvates in addition to the previously known monohydrate. These all had the predicted 5-fluorocytosine ribbon structure, except the metastable hydrate which had the alternative ribbon structure (Fig. 5d right) found higher on the crystal energy landscape. In contrast, the analysis of the hydrogen bonding motifs in the crystal energy landscape of hydrochlorothiazide showed 11 different doubly hydrogen-bonded dimer structures, many of which were observed[48] in the 2 polymorphs and 7 new solvates found in an automated crystallisation screen with 67 solvents. The problems of packing this diuretic molecule with itself and satisfying the hydrogen bonding are further illustrated by two solvates having all hydrochlorothiazide N–H groups hydrogen bonded to solvent. The crystal energy landscape did show that some plausible doubly hydrogen-bonded motifs could not form low energy crystal structures, and these motifs were not observed in the solid forms. Hence the range of hydrogen bonding seen in the lattice energy landscape rationalises the diversity of solvates formed.[48]

This emphasis on hydrogen bonding motif partly comes from most typical hydrogen-bonds producing large barriers to rearrangement on both the crystal energy surface and during the aggregation and nucleation process. However, it also reflects our limited ability to analyse other similarities in dozens of structures on the crystal energy landscape. Such analyses are only readily done, even by the most enthusiastic crystallographer, using programs such as Mercury[49] to detect and classify the hydrogen bonds by Graph Set analysis.[50] Other types of motif may well dominate the crystallisation of some systems, and produce significant barriers to rearrangement on the free energy surface, such as different steric packings of the hydrocarbon butterfly in carbamazepine.[51] The development of more general methods of easily detecting the similarities and differences in sets of dozens of computed crystal structures and observed crystalline solid forms (co-crystals, solvates, polymorphs *etc.*) of related molecules is essential to developing the potential of crystal energy landscapes and crystallisation screens to understand the causes of solid form diversity. Several promising ideas are being pursued.[52,53] We need to be able to detect degrees of similarity in lattice energy minima that are likely to result in only one being a free energy minimum, and similarities in free energy minima to distinguish when it is likely to be practically impossible to prevent transformation to the most stable structure of a group during the process of nucleation and growth. We also need such analysis methods to detect an important subset of polymorphic crystal energy landscapes; those where the low energy structures have interchangeable motifs (Fig. 6b), *i.e.* there is a common supramolecular motif which is packed in different ways in the computed perfect crystal structures and could be interchanged within an imperfect or large unit cell crystal structure.

One type of interchangeable motif is a hydrogen-bonded sheet, which is stacked in different ways in crystal structures which are very close in energy. This implies that there is very little energy penalty for a stacking error, and that the more complete the search, the more different structures based on the same sheets with different sequences of the equi-energetic stacking arrangements would be on the crystal energy landscape. An example of interchangeable sheet motifs is provided by aspirin, where the lattice energy landscape[54] had two stackings of the same sheet almost equi-energetic at the global minimum, though one corresponded to the known structure

and the other was estimated to be rather susceptible to shear forces. The latter structure was later found to be a good match to the structure of a second, metastable, polymorph of aspirin (form II) discovered by serendipity in a failed co-crystallisation experiment.[55] This X-ray structure was disputed,[56] and then a careful X-ray study of a single crystal[57] was published showing "polymorphic domains" *i.e.* regions of form I intergrown with regions of form II. The boundary between stacking disorder from multiple stacking faults and polytypism, and also whether form II should be considered a polymorph if it cannot be obtained in 100% purity, are issues reminiscent of the crystallographic problems illustrated by many mineral systems. The more interesting outcome of this flurry of debate around a system that was once considered monomorphic, is that the prediction of the two equi-energetic structures provides a rationalisation and warning that a complex solid state involving the stacking of the sheets is possible.

Aspirin is not the first example of this. Form 2 of chlorothalonil is a disordered sheet structure[58] that is far more plausibly a stacking or domain disorder of the ordered sheet that the crystal energy landscape shows has two equi-energetic ways of stacking, than disorder in the positions of the Cl and CN groups within the sheets. The $Z' = 3$ structure of form 3 of chlorothalonil is a mixture[58] of the ribbon structures found in two low energy $Z' = 1$ structures in the search. So, although the most stable form of chlorothalonil was found as the global minimum in the lattice energy, the lattice energy landscapes point to the possibility of there being quite a range of other crystal forms which have not proved amenable to isolation and characterisation. An even clearer example of interchangeable ribbons occurs for 5-chlorouracil, where the lattice energy landscape has many structures based on four hydrogen bonded ribbons (Fig. 5e). The polar and non-polar ribbons can interdigitate either in a parallel or anti-parallel mode to form the sheets in a range of almost equi-energetic crystal structures that are very similar, apart from the significant distinction between C═O and C–H. The numerous possibilities for growth errors in either interdigitation or stacking rationalises the disorder in the crystal structure of 5-chlorouracil.[59] These examples strongly suggest that a crystal energy landscape that has a series of structures with interchangeable motifs within a small energy range is providing warning that the solid state may show forms of disorder or crystallographic complexity that may be barely detectable in routine X-ray powder patterns. However, this propensity to growth errors may produce problems in designing a robust crystallisation process that provides crystalline samples with reproducible properties. Since a degree of interchangeability can lead to a continuum of possible crystal structures, this also produces problems in the definition of polymorphism similar to those raised by the continuous composition range in solid solution of two tautomers of the anti-ulcer drug omeprezole.[60] Certainly such ranges of disordered, defective or modulated crystal structures need careful distinction from truly polymorphic structures[61] and the distinction between the two types of polymorphic crystal energy landscapes (Fig. 6) should help clarify the discussion.

We need to be careful to distinguish between the real crystal energy landscape and a lattice energy landscape with a plethora of minima. In the case of chlorouracil, it seems as unlikely that the different hydrogen bonded ribbons (Fig. 5e) would not correspond to different free energy minima as it is that the disorder corresponds to the molecules rotating about their axis breaking two pairs of hydrogen bonds. In contrast, there are many distinct lattice energy minima only fractions of a kJ mol^{-1} above the global minimum which corresponds to the known ordered structure of cyclopentane.[62] Since Molecular Dynamics studies reproduce the transition from the ordered structure to a low symmetry and then a high symmetry rotationally disordered phase,[62] we can assume that the real crystal energy landscapes at the appropriate temperatures would reflect these thermodynamically driven transitions. The more probable relative orientations of the molecules differ significantly between simulated intermediate phase II and the high symmetry, high temperature phase. Thus the lattice energy landscape having an increasing density of structures with small increases in energy rationalises the formation and problems[62] in characterising the intermediate phase II of cyclopentane.

Thus there are a range of complex organic solid state behaviours which are being uncovered by in-depth experimental investigations, driven by the interest in polymorphism. Disordered and modulated structures (such as incommensurate structures and polytypes) will have different implications for the quality control of the crystalline product, depending on the cause. The intriguing observation is that this complexity may be rationalised by interchangeable motifs on the polymorphic type of crystal energy landscape.

4. Other uses of lattice energy landscapes

The above discussion illustrates how lattice energy landscapes can provide worthwhile predictions, whose nature is very dependent on the energy separation between the structures (Fig. 3, 4 and 6) and provide a useful complement to experimental polymorph screening. However, there are uses for any reasonably realistic set of predicted low energy crystal structures, provided that it includes the experimental structure and those that are energetically competitive. It shows the range of possible crystal structures, and can distinguish what types of packing motifs are thermodynamically plausible and which are not, and so prevent the over-interpretation of individual experimental crystal structures. For example, if an experimental crystal structure lacks an expected hydrogen bond, then an examination of the alternative structures on the crystal energy landscape may rationalise this in terms of close packing requirements[63] rather than implying anything unusual about the nature of the hydrogen bond acceptor or donor. Indeed, crystal energy landscapes can demonstrate the strengths and limitations of crystal engineering concepts that link the crystal structure to the functional groups.[64] For example, the energy landscapes of the isomers of dichloronitrobenzene[65] (Fig. 7) all contain crystal structures with near linear C–Cl···O═N interactions rather than Cl···Cl close contacts, clearly confirming the greater strength of the heteroatomic interaction. More importantly, Fig. 7 demonstrates that the crystal energy landscape is very much specific to a given molecule. 2,3-Dichloronitrobenzene has an energy landscape that is probably

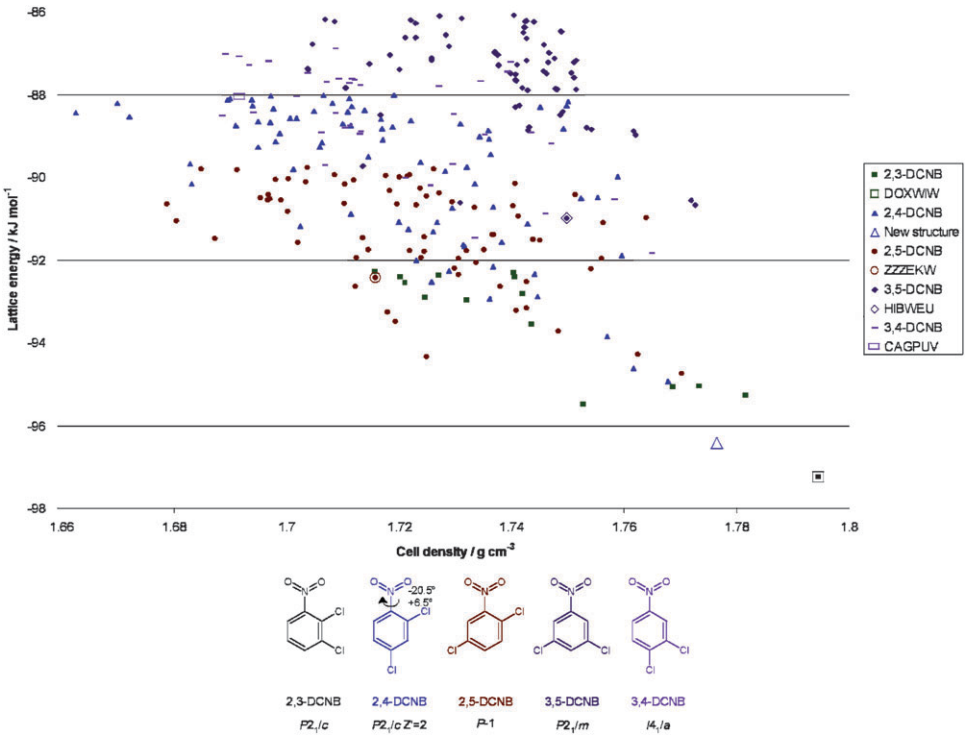

Fig. 7 Lattice energy landscapes for isomers of dichloronitrobenzene[65] calculated with the molecule in its *ab initio* optimised "gas phase" conformation, distinguished by molecular diagram and space group colour. Only structures within 10 kJ mol^{-1} of the global minimum for each molecule are included for clarity. Open symbols denote the lattice energy minima corresponding to the experimental structures, calculated with the same computational model. An exception to this is 2,4-dichloronitrobenzene, where the experimental nitro group torsion angles were used as they differed significantly, as indicated, from the *ab initio* conformation.

monomorphic, and this gives the most stable lattice energy of all the isomers. Since lattice energy generally correlates with molecular formula,[66] we can see that 2,3-dichloronitrobenzene has one way of packing that is clearly better than the others, whereas its isomers have relatively poorer compromises between the C–Cl···O=N interactions and steric requirements. This leads some of the molecules to adopt structures outside the scope of the simple rigid-body lattice energy landscape shown in Fig. 7 although it was perfectly adequate for predicting the structure of 2,3-dichloronitrobenzene. A pair of diastereomeric salts, which only differ by inversion of one chiral centre[67] also illustrate this general point: an approximate lattice energy landscape that gives a clear "probably monomorphic" structure prediction for one molecule can only give a warning of the range of thermodynamically plausible structures for a very closely related one. Until you try computing the energy landscape and see its type, and the structural motifs produced, it is hard to understand why.

A more specific use of crystal energy landscapes where the confidence in the relative energies is low, but probably covers a sufficient range that it should include the experimental structure, is in solving crystal structures from X-ray powder data. When the powder data cannot be indexed to provide a cell, because of peak overlap or low resolution, then comparison of the experimental powder pattern with those simulated from the computed structures can provide a close enough match to provide the starting point needed to solve the structure.[68]

5. Questions that need to be addressed in developing crystal energy landscapes for polymorph prediction

5.1 How accurate a computational model is needed?

The accuracy of the relative energies of the crystal energy landscape depends on how well the intermolecular and intramolecular energies and thermal effects can be estimated, which is very much a function of the type of molecule. Furthermore, as Fig. 7 demonstrated, the accuracy needed to rank the structures in order of energy is very specific to the molecule. Modest errors are permissible for a clearly monomorphic landscape (Fig. 3) and the crystal structure of Pigment Yellow 74 might be well be predicted[28] by quite a crude model, whereas an accurate ranking for an interchangeable polymorphic landscape (Fig. 6b) is probably as unobtainable theoretically as experimentally. Nevertheless, there are clear indicators of what is needed to be reasonably confident of a crystal energy landscape that will give the type of insights discussed above.

5.1.1 Extent of search. The first issue is how widely you should search through the common space groups and number of molecules in the asymmetric unit (Z') in calculating the crystal energy landscape. Many methods are intrinsically limited to $Z' = 1$ and methods that do also search for $Z' = 2$ will produce many more low energy structures for analysis. In the case of interchangeable motifs, such as different stackings of sheets, increasing the range of the search will increase

the number of variants on the structural type observed in the landscape, so the increased computational effort merely reinforces the information already given by a more restricted search. Many high Z' structures are metastable polymorphs, which have been discussed in terms of fossil relics[69] or arrested crystallisations[70] and are closely related to a more stable $Z' = 1$ structure. It will depend on the lifetime and significance of such metastable polymorphs whether it matters if such structures are missed in the search. However, the most stable polymorph of 7-fluoroisatin is a genuine $Z' = 2$ structure where the hydrogen bonds formed by the independent molecules use different donors and acceptors.[71] Hence, in this case a $Z' = 1$ search would be misleading, as it would miss both the thermally most stable structure (form 2) and form 3, a $Z' = 2$ structure which is possibly an example of "arrested crystallisation". Form 1 of 7-fluoroisatin is the global minimum in a simple $Z' = 1$ search, as is the only known form of its isomer 5-fluorisatin. Hence a lattice energy landscape restricted to $Z' = 1$ appears quite reliable for 5-fluorisatin but seriously misleading for 7-fluorisatin.

The computational expense is considerably greater for a $Z' = 2$ search, and for molecular salts,[67] monohydrates[72] and any other system with two independent molecular entities in the asymmetric unit cell, because of the additional variables defining their relative positions. Hence, whilst suitable general search methods exist,[73,74] and there is often the option of just considering likely relative hydrogen bonded geometries within the asymmetric unit,[67,72] the extent of the search will always be a compromise determined by the aim of the study and the computer science facilitation[75] of a CPU intensive process.

5.1.2 The model for the intermolecular forces. The overall picture to emerge from CSP studies is that the better the theoretical basis for the model for the intermolecular forces, the more likely the known crystal structure is to be at, or nearly at, the global minimum of the crystal energy landscape. The use of a realistic representation of the electrostatic interactions due to the lone pair and π electron density is particularly important for modelling the directionality of hydrogen bonding and π–π stacking. This was shown by a survey[76] of the crystal energy landscapes for 50 small rigid molecules which showed that replacing an atomic point charge model with an atomic multipole model, obtained by a distributed multipole analysis (DMA[77–79]) of the same *ab initio* molecular charge density, led to a significant improvement with half the known crystal structures being within 0.5 kJ mol^{-1} of the global minimum. Although the non-electrostatic contribution to the lattice energy is usually modelled by an isotropic atom–atom model potential[80] which has been empirically fitted to a set of organic crystal structures, the use of potentials where the repulsive wall around an atom is anisotropic (derived from *ab initio* charge densities of the molecules) has proved very successful for chlorothalonil[58] and for the chlorobenzenes.[81] Indeed, in the recent blind test of crystal structure prediction, a potential derived[82] solely by applying the theory of intermolecular forces to calculations on the monomer and dimer was able to predict the structure of $C_6Br_2ClFH_2$ (XIII Fig. 1) as the most stable. Thus, the calculation of crystal energy landscapes can provide such a stringent test of the model for the intermolecular forces,[80,83] that it is driving advances[84] in the modelling of organic molecule interactions. For example, the induction energy, the additional stabilisation due to the distortion of the molecular charge distribution in the field of its surrounding molecules, can significantly reorder[46] the relative energies in the lattice energy landscape of carbamazepine.

5.1.3 Molecular flexibility. Calculating the energy landscape for flexible molecules provides additional challenges, as it is necessary to include in the lattice energy the intramolecular energy penalty that the molecule must pay for changing its conformation from the most stable (gas phase) conformation to one that allows better intermolecular interactions. Few force-fields have been sufficiently well parameterised to balance the inter- and intra- molecular forces adequately for CSP, as is evidenced by the proportion of flexible pharmaceuticals which unrealistically change conformation on lattice energy minimisation.[85] Calculating the energy penalty by *ab initio* methods proved adequate for the prediction of the structure of a novel conformational polymorph of the nootropic drug piracetam,[86] following a challenge to do so whilst the novel structure found by high-pressure recrystallisation was being written up for publication.[87] This study used multiple searches with a large number of different rigid conformations. This approach is now being refined to require fewer conformations, by using DMAflex to optimise specified torsion angles under the influence of the intermolecular forces,[45] by combining *ab initio* calculations of the intramolecular energy penalty with the intermolecular lattice energy. DMAflex is not just being used to refine the torsion angles that traditionally define conformational polymorphism. Variations in the degree of pyramidalisation of nitrogen in amine groups, or rotation of hydroxyl groups can make a difference to the proton position and significantly affect the relative lattice energies because of the improved hydrogen bonding. The sensitivity to such details of the molecular geometry implies that ideally we would want to optimise the crystal structures with respect to all atomic positions as well as cell parameters.

To avoid making assumptions about intramolecular flexibility, CSP studies could, in principle, be done by fully optimising the crystal structures by an electronic structure method. Most commonly used solid state *ab initio* methods, based on density functional theory, are not adequate (let alone affordable) for the majority of crystal structure prediction problems,[88] because they do not describe the dispersion interaction well. This error is not systematic in most CSP studies. For example, crystal structures that are hydrogen bonded in all three dimensions will be far better described than those involving hydrogen bonded sheets, because of the importance of the dispersion in determining the sheet stacking distance. The success of empirically dispersion-corrected DFT methods[5] in the recent blind test (Fig. 1) is extremely encouraging. However, even isolated molecule conformational energies can be very sensitive to which high-level computationally demanding method is used when there is significant intramolecular dispersion, for example between different side chains.[89] This suggests that a significant increase in computer speed as well as development of solid state *ab initio* theory will be

needed before we have a method that can accurately rank relative lattice energies for predictive and polymorphic landscapes for larger, flexible organic molecules.

5.1.4 Calculation of free energy rather than lattice energy. The accurate estimation of the free energy at the temperature range of practical interest, rather than the static lattice energy, formally approximating 0 K, is even more challenging. Second-derivative (harmonic) lattice dynamics models can be used to estimate the elastic constants,[90] phonons[91,92] and hence the thermal and zero-point energies[93] for rigid polar molecules with a realistic electrostatic model. Experience with harmonic free energies show that the degree of reordering the relative energies is very dependent on the diversity of structures on the crystal energy landscape and tends to compress the relative energies.[94,95]

The most fundamental limitation of the harmonic approximation is that it cannot eliminate any lattice energy minima as not corresponding to free energy minima. This requires molecular dynamics simulations. Doing such simulations with a sufficiently realistic intermolecular potential surface is challenging even for rigid molecules,[96] let alone with the appropriate incorporation of molecular flexibility to model the mixing of inter and intramolecular modes that clearly affects the thermodynamics of many molecules. Although new methods of free energy calculation are being developed,[26] performing the numerical integrations required to the accuracy of fractions of kJ mol^{-1} needed in relative free energies in this field, let alone the number of structures that need consideration, make this a major challenge to computational chemists in both theory and exploitation of the increasing availability of computing resources.

It is already clear that the ability to simulate the effects of temperature would have a qualitative impact on polymorph prediction, by eliminating many lattice energy minima that are not free energy minima under conditions of practical interest, and demonstrating when thermodynamically driven phase transitions lead to higher symmetry structures (*cf.* the disordered phases of cyclopentane[62]). However, the real crystal energy landscape for ambient conditions could underestimate the number of practically important polymorphs observed. It is extremely unusual for the solid state transformation between polymorphs to be so facile (*i.e.* second order)[97] that there is no hysteresis and sample dependence in the transition which proceeds by nucleation and growth. For example, the transition between the two polymorphs of tetrachlorobenzene is first order, with significant sample dependent hysteresis[98] and yet the differences between the polymorphs are very subtle except in terms of crystal symmetry. The kinetic barriers to transformation to the most stable polymorph can easily be so high that the solid state transformation to the more stable structure is not observed in thermal measurements, for example in 5-fluorouracil.[34]

5.1.5 Overview of theoretical developments. Clearly, calculating the accurate crystal energy landscape envisaged in section 2 is a huge challenge which can involve the application of many areas of computational chemistry to the organic solid state[99] depending on the flexibility and types of functional groups of the molecule. It is unlikely that this could become routine. Consider the lattice energy landscapes derived using the *ab initio* optimised conformers of the dichloronitrobenzenes in Fig. 7. It is plausible that a future study could allow for the NO_2 torsional flexibility using DMAflex,[45] use an intermolecular potential with anisotropic repulsion developed specifically for this range of molecules and search for structures with $Z' = 2$ as well as 1 and in all space groups. This might lead to all the known structures being found as global minima in free energy at ambient temperature and pressure. Since such conditions are not far below the melting point of these compounds, it is plausible that many of the lattice energy minima in Fig. 7 would not be free energy minima. Nevertheless, the range of low energy structures for some isomers is such that we can no more be confident that all the accurate crystal energy landscapes would be "monomorphic", than we can be sure from the practical limitations on the experimental studies[65] that no metastable polymorphs could be found. Hence, even with foreseeable developments in the calculation of crystal energy landscapes, the more important question for polymorph prediction is to understand which of the energy minima are likely to be observed.

5.2 Which thermodynamically feasible structures are polymorphs?

There are many factors that can decide which of the energetically feasible structures are likely to be experimentally found as persistent polymorphs. To take an extreme example of a kinetic barrier, the global minimum energy structure on the crystal energy landscape of progesterone could only be produced by crystallising natural progesterone with its synthetic mirror image *ent*-progesterone.[100] The stable and "disappearing" polymorphs of natural progesterone were much higher in energy, though they were the most stable structures within the chiral space groups.[101] The inability of many chiral molecules to change between enantiomers is the basis for using the crystal energy landscapes (restricted to the chiral space groups) to predict chiral separation by crystallisation.[102] However, for other molecules, there may be a conformation that is so predominant in solution, that the probability of it not being in any homogeneously nucleating cluster is vanishingly small. This conformation is likely to be in the resulting crystal structure, even if it is metastable, provided that the conformational equilibrium and activation energy for nucleation allow it to form faster than a more stable structure with an alternative conformation.[103] Similar kinetic arguments will apply to tautomeric and bimolecular association equilibria. An early nucleating polymorph can nucleate another, faster growing polymorph,[104] and many nucleation processes involve other surfaces or impurities. Thus, whilst studies of nucleation clearly hold the key to understanding polymorphism,[105] the current state-of-the-art of computer simulation of nucleation from solution[106] or the liquid,[107,108] is generally too idealised to model laboratory or industrial crystallisation process.

Seeding plays an important role in dictating the rate of crystallisation processes, and can change the polymorphic outcome, most dramatically accounting[9] for "disappearing" polymorphs. Seeding with a more thermodynamically stable

polymorph could allow its production under conditions where it would not normally nucleate. Hence, any method of finding polymorphs that may generate the most stable form is potentially worthwhile, however impractical for mass production. Many discoveries of new polymorphs appear associated with forcing molecules into conformations and associations that might not otherwise occur and stabilising specific types of nuclei, either deliberately, as in crystallisation in the presence of polymers,[109] templating surfaces[110] or capillary confinement,[111] or accidentally though the effect of impurities[101] (from synthesis) or failed co-crystallisation experiments.[55,112] The range of variables in crystallisation experiments that could influence polymorphic outcome can be so large, that trying them all experimentally is impractical. For example, the polymorphic outcome of glycine has been shown to be affected, not only by the thermodynamic variables of pressure (with the δ and ε polymorphs being recently discovered at high pressure[113]) and temperature, but also whether the crystallisation occurs in the bulk or on the thin film on the walls of the vessel.[114] The most thermodynamically stable form at ambient conditions, γ-glycine, was not discovered until 44 years after the usually kinetically favoured α form,[115] but since then, a huge range of methods that have been shown to influence the polymorphic outcome of glycine crystallisations. These vary from the polarisation state of the laser beam in non-photochemical laser induced nucleation,[116] through ultrasound,[117] deuteration,[118] the use of patterned self-assembled metallic monolayer templates,[119] various methods of supersaturation control,[120,121] to simply adding alcohols to the aqueous solution[122] (though when spraying aqueous solutions with an ethanol and supercritical CO_2 mix,[123] the flow rate affects polymorphic outcome). In such cases, identifying the key factors that control the nucleation and growth is clearly very challenging, let alone incorporating them in a computational model!

6. Conclusion

The practical importance of polymorphism has generated many multidisciplinary studies which reveal the potential complexity of the organic solid state. The recent discovery of new polymorphs of even heavily studied systems, as well the computational chemistry challenges in calculating realistic crystal energy landscapes, means that we can only consider the crystallisation process adequately understood for molecules with the simplest "monomorphic" energy landscapes. To progress towards polymorph prediction when kinetic factors are important requires extensive data linking polymorphic outcome with highly controlled crystallisation conditions,[124] such as we are publishing for automated[125] solvent crystallisations.[8,39,48,65] It also requires generating more realistic crystal energy landscapes, though the utility of the current lattice energy landscapes justifies their storage on a database and dataportal system.[126] Whilst the ability to predict some crystal structures is already demonstrated, the increasing characterisation of multiple solid forms clearly shows that polymorph prediction will be an important scientific "grand" challenge, involving many disciplines, for decades to come.

In 1965 McCrone[127] said that "the number of forms known for a given compound is proportional to the time and energy spent in research on that compound." Since then, the development of new ways of crystallising molecules and characterising them has vastly increased the potential amount of time and money that could be spent. At least computational methods can potentially limit the maximum number of polymorphs to those which appear on the crystal energy landscape, and give a clear warning of whether the molecule is one that will challenge our ability to control its crystallisation.

Acknowledgements

This perspective owes a huge debt to all those who have worked on the "Control and Prediction of the Organic Solid" which was funded by the Basic Technology Program of the Research Councils UK. Thanks to the many scientists who have generously contributed by collaboration with CPOSS, or in the CCDC blind tests, is implicit in the references. However particular thanks go to Dr Louise S Price for her administrative support of CPOSS (http://www.cposs.org.uk), and enabling all the computed low energy structures to be stored on the National Grid Service and so be available on request.

References

1 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 and D. E. Williams, *Acta Crystallogr., Sect. B*, 2000, **56**, 697–714.
2 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 and D. E. Williams, *Acta Crystallogr., Sect. B*, 2002, **58**, 647–661.
3 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*, 2005, **61**, 511–527.
4 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*, 2008, in preparation.
5 M. A. Neumann and M. A. Perrin, *J. Phys. Chem. B*, 2005, **109**, 15531–15541.
6 J. Bernstein, *Polymorphism in Molecular Crystals*, Clarendon Press, Oxford, 2002.
7 R. K. R. Jetti, R. Boese, J. Sarma, R. S. Reddy, P. Vishweshwar and G. R. Desiraju, *Angew. Chem., Int. Ed.*, 2003, **42**, 1963–1967.
8 A. T. Hulme, A. Johnston, A. J. Florence, P. Fernandes, K. Shankland, C. T. Bedford, G. W. A. Welch, G. Sadiq, D. A. Haynes, W. D. S. Motherwell, D. A. Tocher and S. L. Price, *J. Am. Chem. Soc.*, 2007, **129**, 3649–3657.
9 J. D. Dunitz and J. Bernstein, *Acc. Chem. Res.*, 1995, **28**, 193–200.
10 W. Z. Ostwald, *Z. Phys. Chem.*, 1897, **22**, 289–330.
11 T. Threlfall, *Org. Process Res. Dev.*, 2003, **7**, 1017–1027.
12 H. G. Brittain, *Polymorphism Pharm. Solids*, 2005.

13 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. Donaubauer, B. A. Narayanan, M. Soldani, D. Riley and K. McFarland, *Org. Process Res. Dev.*, 2000, **4**, 413–417.
14 S. L. Morissette, S. Soukasene, D. Levinson, M. J. Cima and O. Almarsson, *P. Natl. Acad. Sci. USA*, 2003, **100**, 2180–2184.
15 G. P. Stahly, *Cryst. Growth Des.*, 2007, **7**, 1007–1026.
16 M. U. Schmidt, in *Crystal Engineering of organic pigments*, ed. M. U. Schmidt, 2007, International School of Crystallography, 39th Course: "Engineering of crystalline materials: state-of-the-art in modelling, design, applications", Erice/Italy.
17 F. H. Allen, *Acta Crystallogr., Sect. B*, 2002, **58**, 380–388.
18 J. van de Streek, *Acta Crystallogr., Sect. B*, 2006, **62**, 567–579.
19 A. J. Florence, A. Johnston, P. Fernandes, N. Shankland and K. Shankland, *J. Appl. Crystallogr.*, 2006, **39**, 922–924.
20 A. Gavezzotti, *Acc. Chem. Res.*, 1994, **27**, 309–314.
21 A. Gavezzotti, *CrystEngComm*, 2002, **4**, 343–347.
22 J. D. Dunitz, *Chem. Commun.*, 2003, 545–548.
23 J. D. Dunitz and H. A. Scheraga, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 14309–14311.
24 G. R. Desiraju, *Nat. Mater.*, 2002, **1**, 77–79.
25 P. Raiteri, R. Martonak and M. Parrinello, *Angew. Chem., Int. Ed.*, 2005, **44**, 3769–3773.
26 P. G. Karamertzanis, P. Raiteri, M. Parrinello, M. Leslie and S. L. Price, *J. Phys. Chem. B*, 2007, submitted.
27 M. Rafilovich, J. Bernstein, R. K. Harris, D. C. Apperley, P. G. Karamertzanis and S. L. Price, *Cryst. Growth Des.*, 2005, **5**, 2197–2209.
28 M. U. Schmidt, in *Crystal Engineering: From Molecules and Crystals to Materials*, ed. D. Braga, F. Grepioni and A. G. Orpen, Kluwer, Dordrecht, 1999, pp. 331–348.
29 G. R. Desiraju, *Crystal Engineering: the Design of Organic Solids*, Elsevier, Amsterdam, 1989.
30 C. B. Aakeroy, *Acta Crystallogr., Sect. B*, 1997, **53**, 569–586.
31 B. Moulton and M. J. Zaworotko, *Chem. Rev.*, 2001, **101**, 1629–1658.
32 R. C. B. Copley, L. S. Deprez, T. C. Lewis and S. L. Price, *CrystEngComm*, 2005, **7**, 421–428.
33 W. I. Cross, N. Blagden, R. J. Davey, R. G. Pritchard, M. A. Neumann, R. J. Roberts and R. C. Rowe, *Cryst. Growth Des.*, 2003, **3**, 151–158.
34 A. T. Hulme, S. L. Price and D. A. Tocher, *J. Am. Chem. Soc.*, 2005, **127**, 1116–1117.
35 S. Hamad, C. Moon, C. R. A. Catlow, A. T. Hulme and S. L. Price, *J. Phys. Chem. B*, 2006, **110**, 3323–3329.
36 R. J. Davey, G. Dent, R. K. Mughal and S. Parveen, *Cryst. Growth Des.*, 2006, **6**, 1788–1796.
37 C. A. Hunter, J. K. M. Sanders and A. J. Stone, *Chem. Phys.*, 1989, **133**, 395–404.
38 C. S. Towler and L. S. Taylor, *Cryst. Growth Des.*, 2007, **7**, 633–638.
39 A. J. Florence, A. Johnston, S. L. Price, H. Nowell, A. R. Kennedy and N. Shankland, *J. Pharm. Sci.*, 2006, **95**, 1918–1930.
40 F. P. A. Fabbiani, L. T. Byrne, J. J. McKinnon and M. A. Spackman, *CrystEngComm*, 2007, **9**, 728–731.
41 A. J. C. Cabeza, G. M. Day, W. D. S. Motherwell and W. Jones, *Chem. Commun.*, 2007, 1600–1602.
42 A. Johnston, B. F. Johnston, A. R. Kennedy and A. J. Florence, *CrystEngComm*, 2008, **10**, 23–25.
43 A. J. Florence, C. K. Leech, N. Shankland, K. Shankland and A. Johnston, *CrystEngComm*, 2006, **8**, 746–747.
44 A. J. C. Cabeza, G. M. Day, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2007, **7**, 100–107.
45 P. G. Karamertzanis and S. L. Price, *J. Chem. Theory Comput.*, 2006, **2**, 1184–1199.
46 G. W. A. Welch, P. G. Karamertzanis, A. J. Misquitta, A. J. Stone and S. L. Price, *J. Chem. Theory Comput.*, 2007, in press.
47 A. T. Hulme and D. A. Tocher, *Cryst. Growth Des.*, 2006, **6**, 481–487.
48 A. Johnston, A. J. Florence, N. Shankland, A. R. Kennedy, K. Shankland and S. L. Price, *Cryst. Growth Des.*, 2007, **7**, 705–712.
49 C. F. Macrae, P. R. Edgington, P. McCabe, E. Pidcock, G. P. Shields, R. Taylor, M. Towler and J. De Streek, *J. Appl. Crystallogr.*, 2006, **39**, 453–457.
50 M. C. Etter, J. C. MacDonald and J. Bernstein, *Acta Crystallogr., Sect. B*, 1990, **46**, 256–262.
51 A. J. Florence, K. Shankland, T. Gelbrich, M. B. Hursthouse, N. Shankland, A. Johnston, P. Fernandes and C. K. Leech, *CrystEngComm*, 2008, **10**, 26–28.
52 T. Gelbrich and M. B. Hursthouse, *CrystEngComm*, 2005, **7**, 324–336.
53 A. Parkin, G. Barr, W. Dong, C. G. Gilmore, D. Jayatilaka, J. J. McKinnon, M. A. Spackman and C. C. Wilson, *CrystEngComm*, 2007, **9**, 648–652.
54 C. Ouvrard and S. L. Price, *Cryst. Growth Des.*, 2004, **4**, 1119–1127.
55 P. Vishweshwar, J. A. McMahon, M. Oliveira, M. L. Peterson and M. J. Zaworotko, *J. Am. Chem. Soc.*, 2005, **127**, 16802–16803.
56 A. D. Bond, R. Boese and G. R. Desiraju, *Angew. Chem., Int. Ed.*, 2007, **46**, 615–617.
57 A. D. Bond, R. Boese and G. R. Desiraju, *Angew. Chem., Int. Ed.*, 2007, **46**, 618–622.
58 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.
59 H. Sternglanz and C. E. Bugg, *Biochim. Biophys. Acta*, 1975, **378**, 1–11.
60 P. M. Bhatt and G. R. Desiraju, *Chem. Commun.*, 2007, 2057–2059.
61 A. Gavezzotti, *J. Pharm. Sci.*, 2007, **96**, 2232–2241.
62 A. Torrisi, C. K. Leech, K. Shankland, W. I. F. David, R. M. Ibberson, J. Benet-Buchholz, R. Boese, M. Leslie, C. R. A. Catlow and S. L. Price, *J. Phys. Chem. B*, 2007, accepted.
63 T. C. Lewis, D. A. Tocher and S. L. Price, *Cryst. Growth Des.*, 2005, **5**, 983–993.
64 J. D. Dunitz and A. Gavezzotti, *Angew. Chem., Int. Ed.*, 2005, **44**, 1766–1787.
65 S. A. Barnett, A. Johnson, A. J. Florence, S. L. Price and D. A. Tocher, *Cryst. Growth Des.*, 2008, **8**, 24–36.
66 C. Ouvrard and J. B. O. Mitchell, *Acta Crystallogr., Sect. B*, 2003, **59**, 676–685.
67 P. G. Karamertzanis and S. L. Price, *J. Phys. Chem. B*, 2005, **109**, 17134–17150.
68 M. U. Schmidt, M. Ermrich and R. E. Dinnebier, *Acta Crystallogr., Sect. B*, 2005, **61**, 37–45.
69 K. M. Anderson and J. W. Steed, *CrystEngComm*, 2007, **9**, 328–330.
70 G. R. Desiraju, *CrystEngComm*, 2007, **9**, 91–92.
71 S. Mohamed, S. A. Barnett, D. A. Tocher, K. Shankland, C. K. Leech and S. L. Price, *CrystEngComm*, 2008, DOI: 10.1039/b714566g.
72 A. T. Hulme and S. L. Price, *J. Chem. Theory Comput.*, 2007, **3**, 1597–1608.
73 P. G. Karamertzanis and C. C. Pantelides, *Mol. Phys.*, 2007, **105**, 273–291.
74 V. E. Bazterra, M. Thorley, M. B. Ferraro and J. C. Facelli, *J. Chem. Theory Comput.*, 2007, **3**, 201–209.
75 W. Emmerich, B. Butchart, L. Chen, B. Wasserman and S. L. Price, *J. Grid Comput.*, 2006, DOI: 10.1007/s10723-005-9015-3.
76 G. M. Day, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2005, **5**, 1023–1033.
77 A. J. Stone and M. Alderton, *Mol. Phys.*, 1985, **56**, 1047–1064.
78 A. J. Stone, *The Theory of Intermolecular Forces*, Clarendon Press, Oxford, 1st, 1996.
79 A. J. Stone, *J. Chem. Theory Comput.*, 2005, **1**, 1128–1132.
80 S. L. Price, *CrystEngComm*, 2004, **6**, 344–353.
81 G. M. Day and S. L. Price, *J. Am. Chem. Soc.*, 2003, **125**, 16434–16443.
82 A. J. Misquitta, G. W. A. Welch, A. J. Stone and S. L. Price, *Chem. Phys. Lett.*, 2008, submitted.
83 S. L. Price and L. S. Price, in *Intermolecular Forces and Clusters I*, ed. D. J. Wales, Springer-Verlag, Berlin, Heidelberg, Germany, 2005, ch. 3, pp. 81–123.
84 A. J. Stone and A. J. Misquitta, *Int. Rev. Phys. Chem.*, 2007, **26**, 193–222.
85 S. Brodersen, S. Wilke, F. J. J. Leusen and G. Engel, *Phys. Chem. Chem. Phys.*, 2003, **5**, 4923–4931.

86 H. Nowell and S. L. Price, *Acta Crystallogr., Sect. B*, 2005, **61**, 558–568.
87 F. P. A. Fabbiani, D. R. Allan, S. Parsons and C. R. Pulham, *CrystEngComm*, 2005, **7**, 179–186.
88 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.*, 2007, submitted.
89 T. van Mourik, P. G. Karamertzanis and S. L. Price, *J. Phys. Chem. A*, 2006, **110**, 8–12.
90 G. M. Day, S. L. Price and M. Leslie, *Cryst. Growth Des.*, 2001, **1**, 13–26.
91 G. M. Day, S. L. Price and M. Leslie, *J. Phys. Chem. B*, 2003, **107**, 10919–10933.
92 B. P. van Eijck, *J. Comput. Chem.*, 2001, **22**, 816–826.
93 A. T. Anghel, G. M. Day and S. L. Price, *CrystEngComm*, 2002, **4**, 348–355.
94 A. Gavezzotti and G. Filippini, *J. Am. Chem. Soc.*, 1995, **117**, 12299–12305.
95 J. D. Dunitz, G. Filippini and A. Gavezzotti, *Tetrahedron*, 2000, **56**, 6595–6601.
96 A. E. Gray, G. M. Day, M. Leslie and S. L. Price, *Mol. Phys.*, 2004, **102**, 1067–1083.
97 S. C. Tuble, J. Anwar and J. D. Gale, *J. Am. Chem. Soc.*, 2004, **126**, 396–405.
98 S. A. Barnett, C. K. Broder, K. Shankland, W. I. F. David, R. M. Ibberson and D. A. Tocher, *Acta Crystallogr., Sect. B*, 2006, **62**, 287–295.
99 A. Gavezzotti, *Molecular Aggregation: Structure Analysis and Molecular Simulation of Crystals and Liquids*, Oxford University Press, Oxford, 2007.
100 R. W. Lancaster, P. G. Karamertzanis, A. T. Hulme, D. A. Tocher, D. F. Covey and S. L. Price, *Chem. Commun.*, 2006, 4921–4923.
101 R. W. Lancaster, P. G. Karamertzanis, A. T. Hulme, D. A. Tocher, T. C. Lewis and S. L. Price, *J. Pharm. Sci.*, 2007, **96**, 3419–3431.
102 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.
103 S. Roy and A. Nangia, *Cryst. Growth Des.*, 2007, **7**, 2047–2058.
104 L. Yu, *CrystEngComm*, 2007, **9**, 847–851.
105 R. J. Davey, K. Allen, N. Blagden, W. I. Cross, H. F. Lieberman, M. J. Quayle, S. Righini, L. Seton and G. J. T. Tiddy, *CrystEngComm*, 2002, **4**, 257–264.
106 J. Anwar and P. K. Boateng, *J. Am. Chem. Soc.*, 1998, **120**, 9600–9604.
107 M. Matsumoto, S. Saito and I. Ohmine, *Nature*, 2002, **416**, 409–413.
108 F. Trudu, D. Donadio and M. Parrinello, *Phys. Rev. Lett.*, 2006, **97**, 105701–105704.
109 C. P. Price, A. L. Grzesiak and A. J. Matzger, *J. Am. Chem. Soc.*, 2005, **127**, 5512–5517.
110 C. A. Mitchell, L. Yu and M. D. Ward, *J. Am. Chem. Soc.*, 2001, **123**, 10830–10839.
111 J. L. Hilden, C. E. Reyes, M. J. Kelm, J. S. Tan, J. G. Stowell and K. R. Morris, *Cryst. Growth Des.*, 2003, **3**, 921–926.
112 M. Rafilovich and J. Bernstein, *J. Am. Chem. Soc.*, 2006, **128**, 12185–12191.
113 A. Dawson, D. R. Allan, S. A. Belmonte, S. J. Clark, W. I. F. David, P. A. McGregor, S. Parsons, C. R. Pulham and L. Sawyer, *Cryst. Growth Des.*, 2005, **5**, 1415–1427.
114 M. Xu and K. D. M. Harris, *J. Phys. Chem. B*, 2007, **111**, 8705–8707.
115 C. S. Towler, R. J. Davey, R. W. Lancaster and C. J. Price, *J. Am. Chem. Soc.*, 2004, **126**, 13347–13353.
116 X. Y. Sun, B. A. Garetz and A. S. Myerson, *Cryst. Growth Des.*, 2006, **6**, 684–689.
117 M. Louhi-Kultanen, M. Karjalainen, J. Rantanen, M. Huhtanen and J. Kallas, *Int. J. Pharm.*, 2006, **320**, 23–29.
118 C. E. Hughes, S. Hamad-Gomez, K. D. M. Harris, C. R. A. Catlow and P. C. Griffiths, *Faraday Discuss.*, 2007, **136**, 71–89.
119 A. Y. Lee, I. S. Lee and A. S. Myerson, *Chem. Eng. Technol.*, 2006, **29**, 281–285.
120 J. W. Chew, S. N. Black, P. S. Chow, R. B. H. Tan and K. J. Carpenter, *CrystEngComm*, 2007, **9**, 128–130.
121 G. Di Profio, S. Tucci, E. Curcio and E. Drioli, *Cryst. Growth Des.*, 2007, **7**, 526–530.
122 I. Weissbuch, V. Y. Torbeev, L. Leiserowitz and M. Lahav, *Angew. Chem., Int. Ed.*, 2005, **44**, 3226–3229.
123 A. Bouchard, N. Jovanovic, G. W. Hofland, E. Mendes, D. J. A. Crommelin, W. Jiskoot and G. J. Witkamp, *Cryst. Growth Des.*, 2007, **7**, 1432–1440.
124 T. Threlfall, *Org. Process Res. Dev.*, 2000, **4**, 384–390.
125 A. Florence, A. Johnston, P. Fernandes, N. Shankland and K. Shankland, *J. Appl. Crystallogr.*, 2006, **39**, 922–924.
126 L. Roberts, L. J. Blanshard, K. K. van Dam, S. L. Price, L. S. Price and I. Brown, Providing an Effective Data Infrastructure for the Simulation of Complex Materials. Fifth All Hands e-Science Meeting, 101–105. Springer, Berlin, 2006. Nottingham, National e-Science Centre.
127 W. C. McCrone, in *Physics and Chemistry of the Organic Solid State*, ed. D. Fox, M. Labes and M. M. Weissberger, Interscience, New York, 1965, pp. 726–727.