CHAPTER

11

Molecular Crystal Structure Prediction

Sarah L. Price, Jan Gerit Brandenburg
Department of Chemistry, University College London, London, UK

11.1 WHY PREDICT ORGANIC CRYSTAL STRUCTURES?

Noncovalent forces dictate how molecules come together to form crystals, and the physical properties of those crystals, such as the elastic tensor, morphology, phonon modes, hydroscopicity, solubility, and dissolution rates. Other molecular properties may be modified by the specific crystalline environment, such as color, spin state, and electronic conductivity. Hence, the intermolecular forces are central to the materials science tetrahedron [1], the close links between a crystal structure, its properties, and how its solid state can be processed, and the performance of the drug product. The application of modeling of noncovalent forces in molecular crystals, such as pharmaceuticals, starts from the crystal structure and then extends to predicting the properties of these crystals that are central to their industrial development [2,3].

Polymorphism is the ability of a molecule to crystallize in more than one structure [4]. As an every-day example, the mouth experience of chocolate depends on the polymorphic form of the cocoa butter. The most stable polymorph of cocoa butter (form VI) has a dull surface, soft texture and higher melting point, and chocolate containing it is "bloomed." Thus the chocolatier has to prepare the chocolate to contain the metastable form V, which has the glossy surface sheen, crisp hardness and melts in the mouth at body temperature. Fortunately, there is a significant barrier for rearranging the lipid molecules from polymorph V to VI, so the transformation is very slow unless the chocolate is heated [5].

The ability to predict the crystal structures of a molecule can be seen as a test of our fundamental understanding of the factors that determine how a molecule crystallizes, including intermolecular forces and our ability to implement them in a computer code. The initial application envisaged for crystal structure prediction codes was for the design of new molecular materials with desired properties, to avoid the effort of synthesizing molecules that would not crystallize in an appropriate structure. For example, a molecule may have a large nonlinear optical coefficient and then crystallize in an inactive centro-symmetric crystal. A molecule such as octanitrocubane will not form a good energetic material unless it packs densely in the crystal. The conductivity of molecular electronic materials is very dependent on the molecular packing. Recently, considerable synthetic time has been saved in the devel-

Non-covalent Interactions in Quantum Chemistry and Physics
DOI: 10.1016/B978-0-12-809835-6.00013-X

opment of self-assembling porous organic cages by predicting whether the cage molecules will crystallize with their windows aligned to form pores [6–8]. Crystal structure prediction is intended to be applied to materials design without any experimental input for the specific molecule.

The other application of computational crystal structure prediction (CSP) is in developing organic products and crystallization processes. The molecule has been chosen for other properties, such as biological activity, and there is at least one crystallization in the manufacturing process, possibly producing the solid form incorporated in the final product, such as a pill. If the molecule readily and reliably crystallized to give just one crystalline material with reproducible properties, CSP would just be an academic exercise. However, polymorphism is a major issue for manufacturing molecular crystals. There are highly publicized cases where the supply of a drug treatment to patients has been disrupted by the unexpected crystallization of a new polymorph within the drug product. For example, rotigotine started to crystallize within Neupro transdermal patches which are used for treating Parkinson's disease [9].

In 1996, Abbott Laboratories lost the ability to produce the licensed formulation of the anti-HIV drug ritonavir as the production units started crystallizing a product that was significantly less soluble (see Fig. 11.1) [10]. A complete reformulation was required for ritonavir and new "cold chain" procedures for the storage of Neupro patches. These cases of late-appearing polymorphs cost the pharmaceutical companies a great deal financially and in reputation. Ritonavir is an example of a disappearing polymorph [11], when it becomes difficult to crystallize a polymorph reproducibly once the first sample of a more stable polymorph has been formed. This scientific anathema is becoming better understood as the role of inadvertent seeding and impurities in nucleating new polymorphs is becoming appreciated. The difficulty in experimentally being confident that all relevant solid forms are known for a given compound has led to the use of CSP as a complement to solid form development (see Section 11.4). The majority of organic molecules exhibit polymorphism when an industrial-style search for polymorphs has been conducted [4]. CSP has its greatest application as a route to predicting polymorphs and yet polymorphism considerably complicates the challenge of computationally predicting crystal structures.

11.2 BACKGROUND ON CRYSTAL STRUCTURE PREDICTION

Crystal structure prediction (CSP) is the challenge of predicting the crystal structure of an organic molecule from its molecular diagram. It has made considerable progress in the last two decades, when the inability to predict something as simple as how a molecule would crystallize was considered one of the continuing scandals in the physical sciences [12,13]. At that time, there was little awareness of polymorphism, and it was reasonable to assume that the observed crystal structure would be the most thermodynamically stable crystal structure.

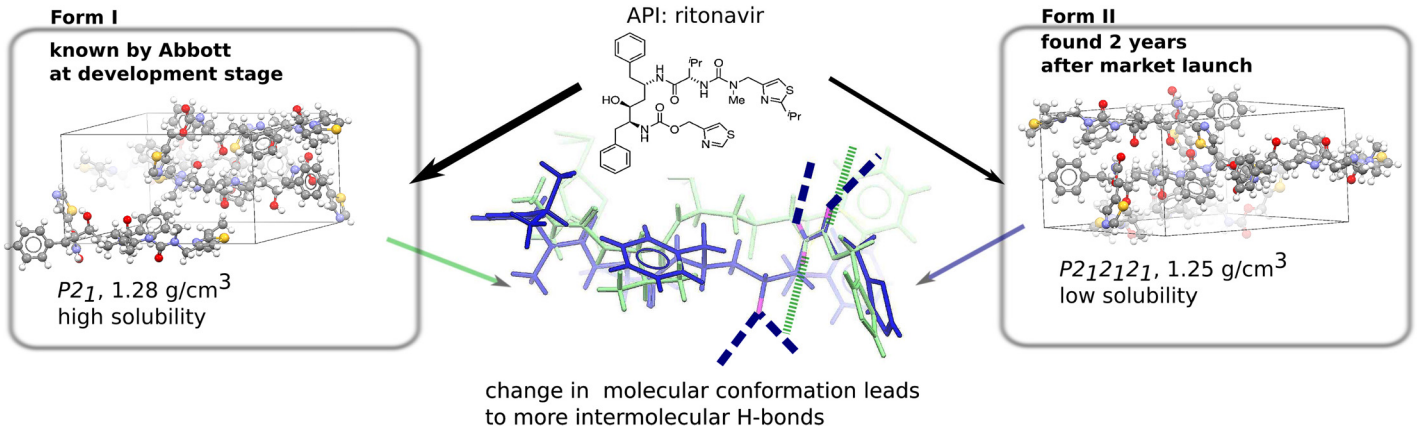

FIGURE 11.1 Problems of late appearing polymorphs: The antiviral drug ritonavir was industrially produced in form I for 2 years. It suddenly started crystallizing into the more stable form II which has a much lower solubility and so required reformulation of a new drug product to provide an effective medicine that could be given to patients. Form II has two more intermolecular hydrogen bonds (···) than form I (---) making it more stable and less soluble, but form I contains the molecule in the conformation that is dominant in solution.

11.2.1 General Algorithm

A typical work-flow of CSP is sketched in Fig. 11.2 and can be roughly broken down into four independent stages:

a. The conformational space that the single molecule could adopt in the solid state has to be determined. In many methods, it is also necessary to decide which conformational degrees of freedom should be allowed to vary in each of the following steps.
b. Plausible crystal structures have to be generated in all likely space groups (which differ in the number of molecules in the unit cell), by varying the cell parameters, relative orientations of the molecules within the unit cell, and for flexible molecules, varying the molecular conformation. Often the search is restricted to one independent molecule in the asymmetric unit cell ($Z' = 1$), with all the other molecules in the unit cell being related by the symmetry elements of the space group.
c. The generated crystal structures have to be optimized and ranked according to their lattice energy with high accuracy. The lattice energy is the energy of a hypothetical static, perfect crystal structure relative to the infinitely separated (noninteracting) molecules (or atoms) in their lowest energy conformation.
d. Further refinement of the thermodynamics of the structures on the crystal energy landscape is usually desirable, along with the calculation of the properties of interest. The crystal energy landscape is defined as the set of crystal structures that are close enough to the most stable one to be thermodynamically plausible as polymorphs (within the likely errors associated with the computational lattice energy model).

This work-flow immediately brings up the two main challenges and requirements of the computational procedure. First, a huge search space of possible crystal packings has to be sampled efficiently. In practice, this means a very fast computation of relative crystal stabilities is combined with efficient methods to scan the search space. Second, the final stability ranking of the possible polymorphs has to be done with very high accuracy as most polymorphic energy differences are well below the "chemical accuracy" of 4.2 kJ/mol. Combining the two requirements results in a need for a hierarchy of lattice-energy evaluations, discarding the less stable structures that would not be among the most stable with a more accurate energy model. This requires a good sense for the likely errors in the intermolecular force models being employed for the specific molecule, as all subtle contributions to the energy could have a significant impact on the relative stability of the structures in the crystal energy landscape, depending on how they vary among those structures.

11.2.2 Discussion of the Common Assumptions Needed in CSP Algorithms

Thermodynamics Determines the Crystal Structures

The basic assumption behind crystal structure prediction codes, that the observed crystal structure will be the most thermodynamically stable one, is complicated by polymorphism. At any given temperature and pressure, one polymorph is in principle more stable than the others, but this stability order can change under practically relevant conditions, let alone between ambient and the nominal 0 K of the lattice energy modeling. The energy differences between polymorphs are small, vary with temperature and are difficult to measure

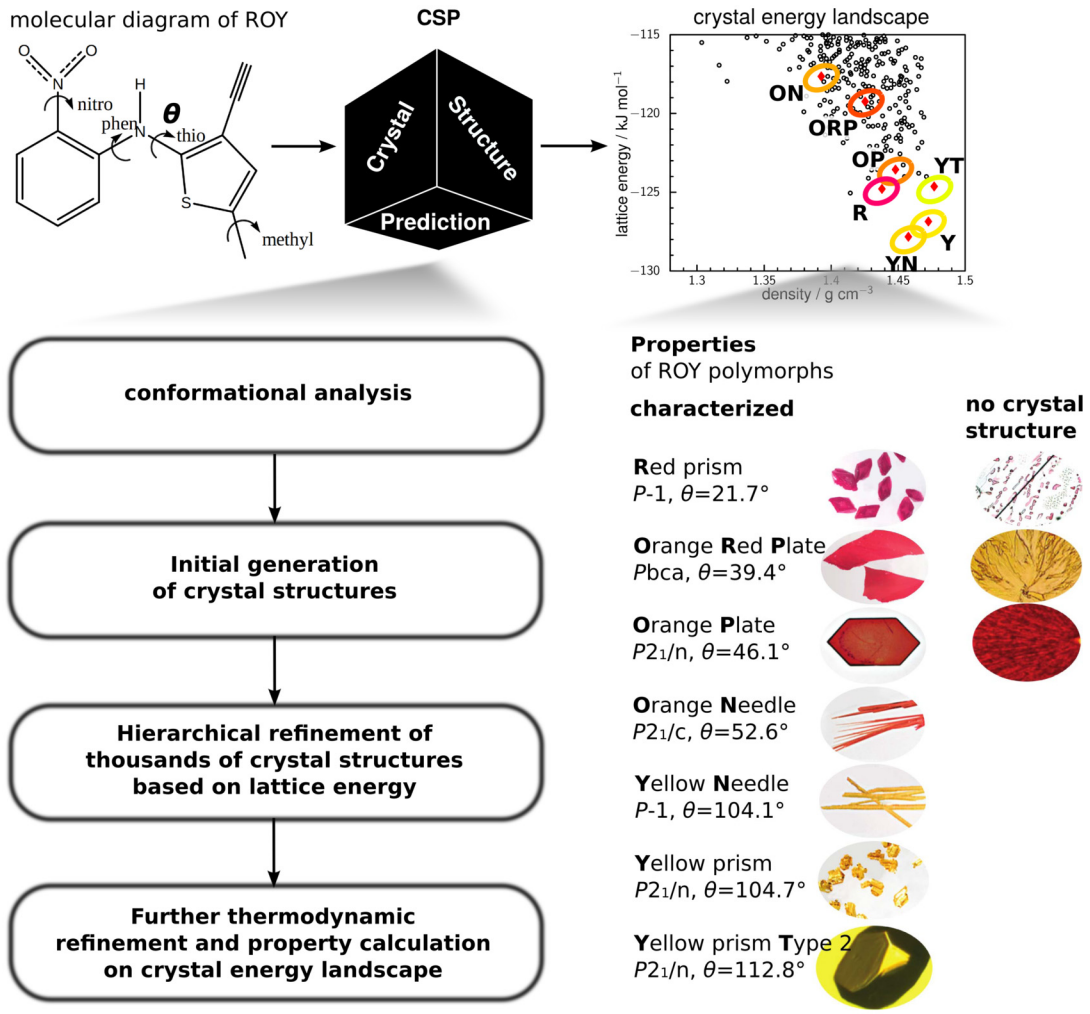

FIGURE 11.2 Typical crystal structure prediction work-flow: Starting from just the molecular diagram, a CSP method first generates and then filters millions of possible crystal packings through multiple stages of increasing accuracy of computational estimates of thermodynamic stability. The example given is for the molecule ROY (5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile) whose molecular flexibility leads to its polymorphs varying in color as well as morphology [14]. The computed crystal energy landscape, the set of structures which are thermodynamically plausible as polymorphs, is summarized by each point corresponding to a crystal structure [15]. Some correspond to the polymorphs whose crystal structures are known as labeled by the initials of the polymorphs. Some of the open black symbols may correspond to the samples where the crystals are not of quality or size to allow their structures to be derived from experimental data. *The crystal photographs are reproduced with permission from Ref. [14]. Copyright 2010 American Chemical Society.*

experimentally. A recent study [16] estimated the energy difference between all crystallographically well-determined pairs of polymorphs of small organic molecules using DFT-D lattice energy differences (a Ψ_{cry} model, see Section 11.2.2c). They found that the energy differences were less than 5.5 kJ/mol for 70% of cases [16]. A larger survey of 508 pairs of polymorphs, using distributed multipole models for the electrostatic component of the lattice energy (a Ψ_{mol} model, see Section 11.2.2c) found that over half had a lattice energy difference of less than 2 kJ/mol and only 5% had an energy difference greater than 7.2 kJ/mol. Including a rigid-molecule harmonic estimate of the free energy changed the relative stability by up to 2 kJ/mol, which changed which polymorph was the most stable at ambient conditions in 9% of cases [17]. These computational estimates seem realistic in light of the available experimental estimates [16]. The difficulty in comparing experimental and theoretical energy differences between polymorphs is explored further in Section 11.5.

It is not the thermodynamic energy difference that determines how readily a metastable polymorph transforms to the more stable, but the energy barrier for transformation. Single-crystal to single-crystal transformations between ordered polymorphs are rare even for structurally quite similar polymorphs [18]. Most polymorphic phases transform via the solution state. Indeed, long-term slurrying when two polymorphs are left in a saturated solution is a common method of determining their relative stability at that temperature. Hence, the assumption that the most stable structure can always be accessed in an experimental screening study is questionable [19]. This is particularly true for larger, flexible molecules that may have difficulty crystallizing at all, or where all the solvent-free polymorphs are found by desolvating solvates. Indeed, it is an important use of CSP to help experimentally target previously unobserved polymorphs, and CSP has inspired the finding of the most thermodynamically stable, but kinetically disfavored, form [20]. Thus, although the basic assumption that the thermodynamically stable form must be capable of existing can be questioned, it is a role for increasingly accurate methods of evaluating the relative thermodynamic stability to test this. Nonetheless, for smaller molecules, the ability to predict that the observed crystalline structures are at least very close in energy to the most stable computer-generated structures is a good test of whether evaluation of the noncovalent interactions is suitable for modeling the molecule in other phases. For instance, a model that can describe both the gas and the solid phase accurately can be expected to perform reasonably well in the liquid phase, where reliable reference data are less accessible.

a) Molecular Bonding and Conformational Analysis

A CSP search to generate putative crystal structures must make some assumptions about the range of possible crystal structures. All methods need to assume the covalent bonding and stoichiometry, e.g., the molecule defined by the chemical diagram, and for a solvate or cocrystal, the ratio of the two components. In an ideal CSP method, facile chemical transformations, such as proton rearrangements to a different tautomer or from a salt to a cocrystal, would be modeled. However, since many molecules of interest are not chemically stable, for example, explosive molecules are significantly less stable than nitrogen, carbon dioxide, and water, CSP is not a search for absolute thermodynamic stability.

Another molecule-dependent assumption is the extent to which the molecular flexibility has to be explicitly considered in the search space to ensure that a structure is generated near all local minima in the crystal energy. Conformational polymorphs are defined as the molecu-

lar conformations in the crystals minimizing to different low energy structures of the isolated molecule [16]. However, many conformations in crystals differ sufficiently from their corresponding isolated-molecule conformations, in shape though not in energy, that they could not be found by optimization of crystal structures generated with just conformational energy minima. Indeed, there is a tendency for larger molecules to adopt more extended conformations in their crystals than in isolation [21].

b) Crystallographic Space Covered by Initial Generation of Crystal Structures

An assumption needs to be made about the crystallographic space covered. Most organic molecules crystallize in monoclinic space groups. The percentage of entries in the Cambridge Structural Database (CSD) with the most popular space group ($P2_1/c$) varied little, from 36% to 34.6%, between 1983 and 2015. In the same time frame, the number of organic crystal structures reported in the CSD database increased from 50 000 [22] to 807 000 [23]. However, the range of space groups varies considerably with the type of molecule: an enantiopure chiral molecule can only crystallize in space groups without inversion symmetry, and are most commonly found to crystallize in $P2_12_12_1$ (see Section 11.5). Although a racemic compound or nonchiral molecule could in principle crystallize in any of the 230 space groups, a search over the 60 most common space groups is likely to be adequate unless the molecule has high symmetry.

A second choice that needs to be made is the number of independent molecules in the asymmetric unit cell (Z') that the symmetry operations of the space group use to create the unit cell of Z molecules. Most CSP searches are restricted to $Z' = 1$, to avoid having to cover the variables defining the relative positions of the two molecules, something that cannot be avoided for multicomponent systems such as salts, cocrystals, or solvates. However, with time, the proportion of $Z' > 1$ crystal structures has been increasing [24]. Whilst some high Z' structures approximate $Z' = 1$ structures, there are cases where the independent molecules have very different conformations or use different hydrogen bonding donors and acceptors. Unfortunately, the incidence of $Z' > 1$ structures is higher for industrially screened polymorphic systems (approximately 20%) than the bulk of the CSD [4]. In principle it is possible to not use crystallographic symmetry in the search but this would need at least 16 independent molecules in the unit cell, and still exclude space groups with three-fold or higher symmetry axis, which are not uncommon for some classes of molecules. The search space also increases drastically with the amount of molecular flexibility that needs to be considered explicitly at the search stage. Typically, we generate a million putative structures for a search in the 60 most popular space groups for a small achiral pharmaceutical molecule, and often find that this is insufficient to ensure with confidence that all $Z' = 1$ structures within these space groups have been located.

c) Hierarchical Refinement of Lattice Energies

The sheer number of crystal structures that need to be considered and the accuracy needed for the final relative energies leads to a range of lattice energy models being employed in CSP studies. We roughly categorize them in three classes.

1. Classical analytical atomistic force fields (FF) for both the intramolecular and intermolecular interactions.

2. *Atomistic* hybrid models that assume the identity of the organic molecule within the crystal structure. The theory of intermolecular forces is used to generate an analytical intermolecular potential (cf. Chapter 1), in an anisotropic atom–atom pairwise form. The intramolecular (conformational) energy and the parameters of the intermolecular potential are derived from the electronic structure of the molecule, and so this general approach will be denoted Ψ_{mol}.
3. *Periodic electronic* structure methods that directly describe the interactions of the electrons with the nuclei in evaluating the crystal lattice energy. They typically involve a wavefunction of the periodic crystal and so are denoted Ψ_{cry}.

Classical force fields are readily available for a large class of biologically relevant molecules. These classical potential functions which are based on interatomic distance and angle dependent terms have been used for decades to approximate molecular potential energy surfaces [25,26]. Conventional FFs assume transferability of the potential parameters between different molecules within a class of chemically similar compounds and are available in standard packages like GROMACS, AMBER, CHARMM, or TINKER. Much cruder force fields designed to find reasonably close packed crystal structures without steric clashes can also be used. However, molecule specific FFs, whose parameters have been fitted to a general functional form, are often more appropriate and effective for the initial crystal structure generation stage [27].

The atomistic hybrid models emerged as a response to the accuracy requirements in CSP at a time when method development concentrated on rigid molecules. It was clear that a molecule-specific parameterization of the intermolecular forces was required. The analytical intermolecular potential can be derived using theory of intermolecular forces by parameterizing the long-range forces from computed molecular properties and the short-range terms by, for instance, fitting to DFT-SAPT calculations as described in Chapters 1 and 2. In this approach, the lattice energy is calculated by summing the analytical intermolecular potential over all pairs of molecules in the crystal, to give the dominant intermolecular contribution U_{inter}. For flexible molecules, it is necessary to add the small energy penalty for any change in conformation from that of the isolated molecule ($\Delta E_{\mathrm{intra}}$), i.e., assuming

$$E_{\mathrm{latt}} = U_{\mathrm{inter}} + \Delta E_{\mathrm{intra}}\,. \tag{11.1}$$

The ab initio calculation on the isolated molecule gives the molecular structure and also readily provides a molecule-specific electrostatic model. It has been found that using distributed multipole analysis (DMA, see Chapter 1) to parameterize an anisotropic atom–atom electrostatic model makes a significant difference to the success of CSP studies [28]. The atomic dipoles and quadrupoles represent the anisotropic electrostatic forces arising from the lone pairs and π electron density that are important in representing the hydrogen and halogen bonding and π–π stacking interactions. These interactions compete with each other and the repulsion–dispersion interactions in defining the optimum, densely packed structure. The effect of the surrounding molecules in the crystal can be estimated by calculating the molecular charge density in a polarizable continuum of typical dielectric constant $\epsilon = 3$ [29].

For many CSP applications, all the intermolecular forces apart from the distributed multipole electrostatic contribution are usually modeled by an isotropic atom–atom exp-6

repulsion–dispersion potential. These exp-6 parameters have been explicitly fitted to crystal structures and heats of sublimation and so partially absorb some errors in the assumed functional form of the intermolecular potential and the neglect of thermal expansion and others [30]. This approach using the FIT potential [31] has recently been shown to be as accurate as many first-generation DFT-D methods for the X23 benchmark set of small molecular crystals [32].

The model for the intermolecular forces can be systematically improved, not only by improvements in the molecular charge distribution and in the method of deriving the electrostatic model, but also by using the molecule's atomic polarizabilities to model the induction contribution [31] and the atomic dispersion coefficients for the dispersion contribution (see Chapter 1). Anisotropic atom–atom models for the short-range interactions can be derived using the overlap model and fitting to dimer interaction energies such as those obtained using DFT-SAPT. Such nonempirical model potentials have been very successful in predicting the structures and properties of rigid-molecule crystals such as the chlorobenzenes [18,33]. Indeed, the ability to reproduce the known crystal structures as the more stable of those generated by CSP is a severe test of the quality of the intermolecular potential model. For flexible molecules, the molecular wavefunction has to be recalculated for sufficient conformations to cover the conformational variation of ΔE_{intra} and the distributed multipoles within the crystal structures. The use of databases of the properties derived from the molecular wavefunction calculations [34] means that the optimization of the first crystal structure (Eq. (11.1)) requires many single molecule ab initio calculations on the different conformations, but the computational cost per crystal structure optimized rapidly decreases. This Ψ_{mol} approach with cruder grids for intramolecular energy can also be used in the crystal generation stage for highly flexible molecules, as in the versions designed for studying the pharmaceutical molecules of the CrystalPredictor code [35,36].

The most prominent method from the Ψ_{cry} class are periodic density functional approximations. As highlighted in several of the preceding chapters, most DFT methods do not describe the important London dispersion interaction, the universal attractive part of the van der Waals forces. Dispersion interactions are very important in most organic crystal structures. For instance, a molecule can crystallize in structures where dispersion determines the separation between layers of hydrogen bonded molecules. Even in crystals where strong hydrogen bonds span all three unit cell directions, the long-range nature of dispersion means that lattice summation over this contribution converges slowly to a significant energy. Thus, all semilocal DFT-based Ψ_{cry} methods have to be corrected for the missing dispersion and we refer to this group of methods as DFT-D (see Fig. 11.6).

d) Further Thermodynamic Refinement and Property Calculation

At the lattice-energy refinement stage, each crystal is modeled as a static infinite perfect lattice. This ignores the energies associated with the surface, disorder or defects in real crystals. For example, the surface dipole correction term, which arguably should be applied to crystals in polar space groups and depends on the area of the polar surfaces, is ignored [37]. More importantly the calculation of the lattice energy ignores the effect of temperature on

the relative thermodynamic stability of the crystals. There are additional contributions to the enthalpy:

$$H(T) = \overbrace{E_{\text{latt}} + PV + \underbrace{E_{\text{ZPVE}}}_{\text{crystal at 0 K}}}^{\text{crystal at 0 K}} + \underbrace{\int_0^T \mathrm{d}T'\, C_V(T')}_{H_{\text{vib}}}, \tag{11.2}$$

the Helmholtz free energy:

$$F(T) = \overbrace{E_{\text{latt}} + \underbrace{E_{\text{ZPVE}}}_{\text{crystal at 0 K}}}^{\text{crystal at 0 K}} + \underbrace{\int_0^T \mathrm{d}T'\, C_V(T') - TS}_{F_{\text{vib}}}, \tag{11.3}$$

or the Gibbs free energy:

$$G(T) = \overbrace{E_{\text{latt}} + PV + \underbrace{E_{\text{ZPVE}}}_{\text{crystal at 0 K}}}^{\text{crystal at 0 K}} + \underbrace{\int_0^T \mathrm{d}T'\, C_V(T') - TS}_{G_{\text{vib}}}. \tag{11.4}$$

The additional contributions, e.g., F_{vib}, depend on the phonon spectrum of the crystal as introduced in Chapter 10. Here, we relate the vibrational contributions to the isochoric heat capacity C_V and entropy S of a crystal. At 0 K the zero-point vibrational energy E_{ZPVE} has to be considered and at finite temperatures additional thermal contributions are important. While the zero-point energy is dominated by the high-energy (mostly molecular covalent) frequency modes, the thermal contributions are most sensitive to the low energy modes which are predominantly intermolecular with some coupling to the low frequency modes of flexible molecules. The phonon mode frequencies are typically computed using the harmonic approximation. This is a reasonable simplification for all high frequency molecular modes, but is less reliable for the low energy modes that correspond to extremely flat intermolecular potential energy surfaces. All organic crystals show significant thermal expansion, which can be very anisotropic, showing the limitations of the harmonic approximation [38]. Even within the harmonic approximation, the accurate evaluation of such contributions to F_{vib} requires considerable care as well as a reliable potential energy surface (see Chapter 10) [32]. Many CSP applications therefore just use the relative lattice energies and neglect the differences in the zero-point and thermal energy.

Temperature has a significant effect on molecular crystals, as functional organic crystals are being used at temperatures that are far closer to their melting (or decomposition) temperatures than 0 K. For example, the polymorphs of ROY (Fig. 11.2) melt between 97 and 115°C. Because of molecular motions, many lattice energy minima may correspond to only one minimum in the free energy, significantly reducing the number of distinct crystal structures at practically relevant ambient temperatures. The proportion of structures that are artifacts of the neglect of temperature depends on the molecule and the crystal packing: there is a drastic reduction (approximately one order of magnitude) in the case of benzene whereas, for

5-fluorouracil, 25% of lattice energy minima were not free energy minima [39,40]. This reflects the degree of motion and ease of transformation between benzene polymorphs, in contrast to the 5-fluorouracil polymorphs, which have different hydrogen bonding patterns and do not interconvert without going through a liquid phase. Observed crystal structures can be dynamical averages over many static CSP idealized structures, ranging from rotationally disordered phases, to just having the methyl groups rotating freely. It is important to confirm that crystal structures optimized within space-group constraints are really mechanically stable minima. When the symmetry-constrained optimization has produced a transition state, removing the symmetry constraints can give such a small energy lowering that the observed structure is a dynamical average in the original space group.

Thus once the CSP computer algorithm has generated a set of lattice energy minima calculated with the most accurate model for the lattice energy that can be afforded or is justified, there needs to be an evaluation of how many structures fall within the crystal energy landscape, that is, the set of structures which are thermodynamically plausible as polymorphs [38]. This will include both the likely energy differences between polymorphs and a quantitative or qualitative estimate of the likely temperature effects and all other approximations in the estimates of the relative free energies of the crystal structures [38]. Once the crystal energy landscape is chosen, various properties can be calculated, whose accuracy will again depend on the quality of the model for the intermolecular forces and approximations made. For example, the powder X-ray diffraction pattern calculated from the lattice energy minima, even with the perfect model for the lattice energy, will be in error by the thermal and zero-point energy expansion. This stage of the CSP study will depend on the purpose of the study but also on the size of the energy gaps between the low energy crystal structure and the relationships between the crystal structures.

Comparison of Crystal Structures

During a CSP run it is essential to remove duplicate structures. This procedure needs to go beyond just removing the many mathematically equivalent sets of cell parameters. The definition of how much two crystal structures have to differ before they correspond to polymorphs (rather than sample, temperature or pressure dependent variations) is a practical issue with intellectual property and quality control implications. In practice, the ability to overlay the coordination sphere (usually a 15 molecule cluster, larger for multicomponent systems) with a reasonable root mean square difference is usually taken as a method of establishing that structures are effectively the same. The similarity between simulated powder X-ray diffraction patterns is often used, but this measure still has a gray area, and can be insensitive to hydrogen atoms or certain atomic exchanges [41]. In the CSP algorithms, there is the need to ensure that no minima that would correspond to distinct thermodynamically plausible polymorph structures in the most accurate energy model, are eliminated at an earlier stage.

A variety of tools are used to determine the relationships between the computer generated and known crystal structures. The hydrogen bonding pattern in each crystal can be described by graph sets to identify crystals with the same hydrogen-bonding motifs. Other forms of similarity in the packing, such as common dimers, stacks, or layers of molecules, can be found by pairwise comparison of the structures using XPac [42,43]. This type of analysis can be used to estimate whether structures are sufficiently distinct to be possible polymorphs. For exam-

ple, once structures with distinct hydrogen bonds are formed in the nucleation process there is likely to be a significant barrier to breaking the hydrogen bonds, thus the two structures may well be polymorphs. Structures which are similar could just crystallize in the same stable form or the observed crystal structure may incorporate the other structure in its dynamic or static disorder or as growth defects [44].

CSP methodology is still emerging as it is applied to an increasing range of molecular systems. The sheer number of different ways of packing organic molecules that need to be considered in a search means that all CSP methods with an extensive enough search involve a hierarchy of energy evaluation methods, with clustering to remove structures that are expected to correspond to the same eventual minimum. Of course, a limited search reduces the computational cost, but this runs the danger of concluding that no polymorphs are expected when a more extensive computational and experimental search finds them [45]. The final energy evaluations are likely to be the most expensive and accurate that can be afforded. However, the lattice energy of benzene has only recently been evaluated to a 1 kJ/mol accuracy, requiring the use of experimental data that is rarely available for less studied systems [46]. Hence, the success and impact of CSP methods comes from considerable cancellation of errors in the relative energies of the computer generated structures. Assessing the likely extent of this cancellation requires knowledge of the types of intermolecular forces operating for the given system as well as how much these interactions differ in the low energy crystal structures.

11.2.3 Example of CSP in Practice

As a practical example of the work-flow in Fig. 11.2, we describe a CSP study conducted on the nonsteroidal antiinflammatory drug naproxen [47]. The molecule is chiral (chiral center C11, see Fig. 11.3) and marketed in the homochiral form, which has to be separated from racemic mixtures that contain both hands of the molecule. The study was undertaken because there was no crystal structure for racemic naproxen in the CSD, and concurrent experimental studies could not grow a single crystal suitable for X-ray diffraction.

Conformational Analysis

The molecular diagram of naproxen (Fig. 11.3) shows three flexible torsion angles, with torsion Φ_1 being independent from the Φ_2 and Φ_3 torsion angles, which are coupled. A one dimensional potential energy surface scan for Φ_1 shows that one conformation is significantly lower in energy than all others, but that another more planar conformation is sufficiently low in energy that it might form crystal structures. This is confirmed by looking at structures in the CSD as there are two structures with the less favorable planar conformation (Fig. 11.3B). (This analysis also shows that if C_6 or C_8 had substituents other than hydrogen, a much larger conformational search space would have been needed.) The two local conformational minima (Fig. 11.3A) are sufficiently steep that the angle needs not be considered as a search variable in the crystal structure generation stage. A subsequent two-dimensional scan of Φ_2 and Φ_3 leads to the conclusion that 8 ab initio optimized conformations of the isolated molecule should provide a reasonable approximation to any crystalline conformers, and hence a rigid-molecule search method could be used. These calculations of the molecular wavefunctions

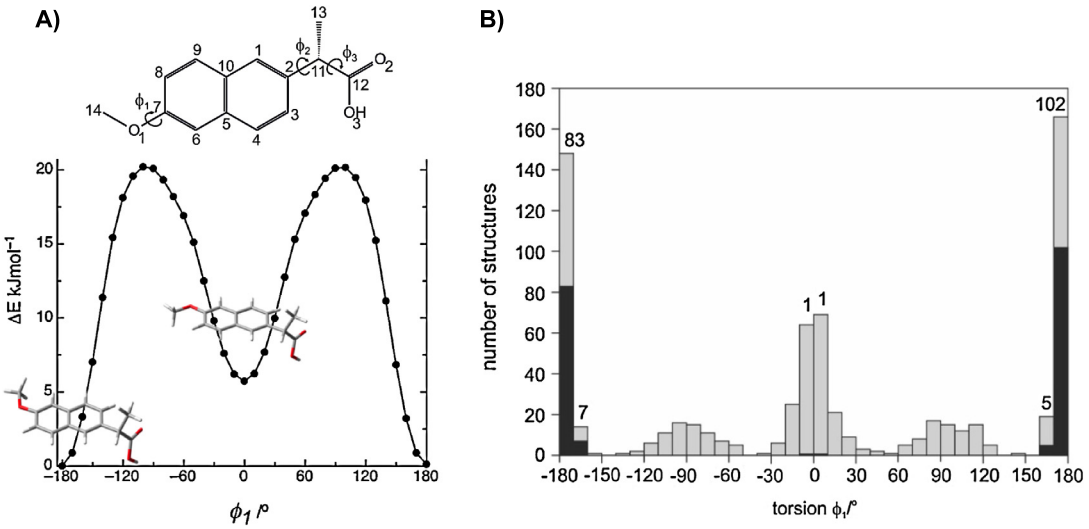

FIGURE 11.3 Molecular conformation of naproxen (see Ref. [47]). (A) The three flexible torsion angles are defined with an explicit one dimensional potential energy surface scan (PBE functional) with respect to Φ_1. (B) Summary of the statistical evaluation of the torsion angle distribution of crystal structures containing β-methoxynaphthalene moieties (CSD database v5.32). Black (and numbers above bars) are structures where the ortho-substituent are hydrogen atoms (C_8H, C_6H, as in naproxen) and gray are structures where C_6 or C_8 are substituted by larger functional groups.

also provided the relative conformational energies ΔE_{intra}, and distributed multipoles for lattice energy calculations using each conformer.

Initial Generation

The 8 rigid conformers were each used to generate crystal structures using MOLPAK [48] which performs a systematic grid search on the orientation for the rigid central molecule in 39 common coordination geometries of organic molecules (covering 19 common space groups), with one molecule in the asymmetric unit cell ($Z' = 1$). Approximately 3000 of the most closely packed structures found with MOLPAK's pseudo-hard sphere potential were carried on to the refinement stage. When this initial CSP study and experimental evidence defined the space group of the racemic compound, a more extensive CrystalPredictor search was carried out in this space group.

Refinement with Ψ_{mol} Approach

An atomistic Ψ_{mol} hybrid model was used in both steps with distributed multipoles from the charge density of the isolated molecules partitioned into atomic contributions. First a rigid body minimization is conducted, which keeps the molecular structure fixed and relaxes all intermolecular degrees of freedom. The most stable of those structures are then reoptimized allowing for small changes in Φ_1, Φ_2 and Φ_3 and other bend and torsion angles, so that the bend in the naphthalene ring in the enantiopure crystal structure was reproduced.

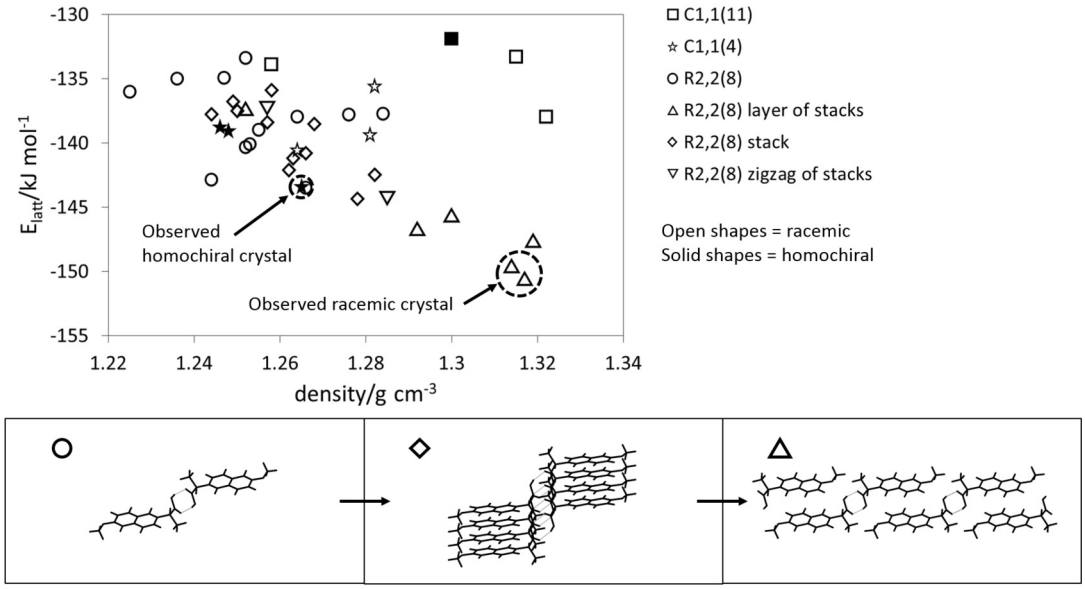

FIGURE 11.4 Crystal energy landscape of naproxen [47]. Each symbol represents a stable crystal structure, classified by chirality, hydrogen-bonding graph set and common packing motifs. The XPac-derived packing motif hierarchy is illustrated below, with the layer observed in the racemic structure (△) being one of the ways of combining a stack of hydrogen bonded dimers (◇) (another is a zigzag of stacks) with there being other unique structures containing this stack, and other structures that contain the dimer (◯), but not the stack. The two structures corresponding to the racemic crystal structure are the $Z' = 1$ transition state found in the search and the very similar $Z' = 2$ lattice energy minimum [47].

Property Calculation

The lattice energies and densities of the structures on the crystal energy landscape are shown in Fig. 11.4 with each crystal structure labeled by its packing type. These labels were derived by considering whether the crystal was racemic or enantiopure (from the space group), its hydrogen bonding graph set and information from the XPac analysis. This showed (Fig. 11.4) that many low energy structures contained the same stack of hydrogen-bonded dimers, and some had the same layers in common.

The known enantiopure structure was the lowest energy structure in the chiral space groups. However, there were various lattice energy minima for racemic naproxen based on essentially the same layer containing the inversion-related carboxylic acids dimers. The lowest energy racemic structure in the $Z' = 1$ search was actually mechanically unstable, and removing a symmetry constraint gave a $Z' = 2$ lattice energy minimum. This was only 1 kJ/mol lower in energy, which is less than the estimated zero-point energy of the structure. Using these structures as a starting point allowed the solution of the experimental structure from powder X-ray diffraction data, but solid-state NMR had to be used to show that the structure was definitely a $Z' = 1$ average over the lattice energy minima. From the lattice energies, or even rigid-body harmonic free energy calculations (see Eq. (11.3)), it is hard to conclude that the other layer structures would not correspond to free energy minima,

TABLE 11.1 Evolution of the CSP blind challenges

Year	Success[a]	Attempts[b]	Atoms	Flexible bonds	Elements	Z''_{max}[c]	Multicomponent
1999 [53]	3/4	6–11	11–28	0, 1, 2, 2	H, C, N, O	1	—
2001 [54]	2/3	11–15	18–28	0, 2, 2	H, B, C, N, O	1	—
2004 [55]	2/4	15–18	11–35	0, 0, 6, 0	H, B, C, N, O, S, Br	2	—
2007 [56]	4/4	12–14	8–32	1, 0, 4, 4	H, C, N, O, I	2	cocrystal
2010 [57]	6/6	10–14	13–55	0, 2, 4, 2, 10, 4	H, C, N, O, F, S, Cl, Br	2	hydrate, salt
2014 [58]	8/9[d]	8–24	15–62	0, 7, 5, 6, 8	H, C, N, O, S, Cl	3	hydrated salt

[a] *X/Y gives the number of targets structures (Y) and the number of targets with at least one successful submission (X). In the first five blind tests a success means that the experimental structure matches one out of three structures submitted by each group. In the sixth blind test, a success means the structure is found in any of the two lists of 100 structures that could be submitted.*

[b] *Minimum to maximum number of groups trying any of the Y targets.*

[c] *Largest number of independent molecules in the asymmetric unit cell.*

[d] *One target was polymorphic with 5 target structures.*

capable of crystallizing separately as polymorphs. Kinetically, they are likely to appear as growth faults, possibly accounting for why it was not possible to grow single crystals of racemic naproxen. The determination of the racemic structure was the starting point to testing whether the energy differences between the racemic and enantiopure crystal structures could be evaluated sufficiently accurately for designing chiral separation by crystallization processes (Section 11.4)

11.3 BLIND TEST OF CRYSTAL STRUCTURE PREDICTION

Knowing the industrial interest in polymorph prediction (see Section 11.4), the Cambridge Crystallographic Data Centre (CCDC) hosts CSP blind test challenges every few years, to provide an unbiased assessment of the state-of-the-art ability to predict organic crystal structures. With the assumption that the target crystal structure is the most thermodynamically stable, this is an assessment of methods for modeling the thermodynamics of organic crystal structures (see Chapter 10 and Refs. [49,50]). The CSP blind tests assess the ability to calculate relative, rather than absolute, thermodynamic stabilities in contrast to the Assessment of Modeling Proteins and Ligands (SAMPL) [51,52]. In particular, it also tests the initial generation of crystal structures and all the other assumptions involved in a CSP protocol (see Section 11.2.2), stimulating advances in the field.

Table 11.1 summarizes the progress in CSP since 1999, both in terms of the size and flexibility of the molecules involved, and the number of groups with different codes and protocols participating and the success rate for the leading codes. Hence, the blind test papers provide a snapshot of the state of development of the field [53–58].

The difficulty of the CSP challenges has been increasing, and yet there is still a long way to go before CSP can be routinely used on most molecules of industrial interest. First, the sampling space that has to be covered has grown dramatically from small rigid molecules to a

molecule with eight flexible torsion angles, and an increase in the number of molecules in the asymmetric unit cell. Second, the challenge of describing the intra- and intermolecular forces grew with the introduction of heavier elements, hydrated crystals, and salts. The strong intermolecular forces involved in salt hydrates, for example, are challenging to the accuracy of many approximate density functionals (especially of the GGA level, Ψ_{cry}) as well as methods based on the molecular charge density (Ψ_{mol}). Thus, the CSP blind test challenges are testing many different aspects of computational modeling.

11.3.1 Evolution of CSP Methods

The first few blind tests showed that classical transferable intra- and intermolecular force fields (FFs), of the type used in many commercial modeling packages, including Polymorph-Predictor, are not accurate enough for reliable CSP. This can occur because the balance between the inter- and intramolecular forces is not sufficiently realistic to give the correct conformation in the crystal structure. An error in a torsional potential can have a major effect on the shape of the molecule and hence its packing. This, and problems in relative energy ranking may be attributed to using the same atomic charges for both the inter- and intramolecular forces [59]. Overall the first blind tests had limited successes, with most being from methods that use a hybrid atomistic Ψ_{mol} model with the electrostatic Coulomb interaction from distributed multipole analysis of the molecular wavefunction.

In the fourth blind test there was the first completely nonempirical model potential, using the Ψ_{mol} approach, with anisotropic repulsion as well as electrostatic terms derived using monomer long-range properties and SAPT calculations of dimer energies (see Chapter 2) [60]. This was achieved for $C_6Br_2ClFH_2$, where the polarization of the molecular charge distribution by the crystalline environment could be ignored and the molecule treated as rigid. This test also saw the company Avante-Garde Simulation introduce their CSP method using a periodic DFT (PBE functional) plus dispersion correction method that had been explicitly derived for modeling organic crystal structures [61]. This dispersion correction is similar in spirit to the D1 scheme [62]. The use of periodic electronic structure calculations in the final optimization and lattice energy evaluation has the advantage of automatically optimizing the molecular conformation and including the polarization energy. Their success was also based on fitting a molecule-specific tailored force-field to DFT-D data and using this to generate the structures using the code GRACE [63]. The use of the more accurate force field reduces the number of structures that need refinement by periodic DFT-D. This step-change in methodology enabled them to correctly predict all the target structures.

However, some of the limitations of this specific Ψ_{cry} approach became apparent in the fifth blind test with a salt and a hydrate system, whose intermolecular interactions are not well modeled by the PBE functional. The hydrate was particularly challenging, partly because many structures were essentially equivalent apart from the proton positions and very similar in energy. The calculations led to a reinterpretation of the diffraction data and a revised structure for form IV of gallic acid monohydrate, and to the discovery of a further polymorph and 22 solvates, many containing water [64]. The fifth blind test was the first to contain a molecule of the size and flexibility of the smaller pharmaceuticals in development, which was correctly predicted by two Ψ_{mol} methods [65]. This success inspired industrial

pharmaceutical scientists to collaborate on developing the application of CSP for use in the pharmaceutical industry (see Section 11.4).

The blind tests have also seen an evolution in the methods of generating plausible starting structures, as well as the number of variables that need to be considered explicitly in the generation stage. Multicomponent systems (or $Z' = 2$ searches) need to consider the relative orientation of the molecules in the asymmetric unit cell, and each flexible torsion angle that can make a significant difference to the shape of the molecule adds another dimension to the search space. Different systematic methods of sampling are possible and can include (quasi)random searches, Monte-Carlo techniques, and evolutionary or particle swarm algorithms. For example, GRACE uses a modified simulated annealing search [66], and CrystalPredictor [36] is based on Sobol sequences. Both codes use fairly good lattice energy models for optimizing millions of crystal structures: either a tailor-made force field or an atomic charge plus exp-6 potential with a grid-based estimate of ΔE_{intra}. Crucially, these codes have tests to determine the completeness of the search within the defined search space, in terms of the number of times the most promising structures are found. Aiming for a complete search (within the defined search space, see Section 11.2.2) can make the computational cost of the search vary considerably for similar molecules with the same flexibility. Though an essential step in the CSP task, we will not go into deeper discussion of these crystal structure generation techniques and concentrate on the modeling of noncovalent interactions.

11.3.2 State-of-the-Art CSP as a Test of Modeling Noncovalent Interactions

The sixth blind test had five target systems, shown in Fig. 11.5, and since target XXIII had uniquely been screened for polymorphs, allowed the submission of two lists of 100 structures. There was a considerable enlargement in the number of participating groups, with the 91 researchers from 14 countries introducing many new methods, described in the account of the test [58]. Only a few of the groups were able to attempt the $Z' = 2$ polymorphs C and E of XXIII because of its size and flexibility (Fig. 11.5). The successes in finding the structures were dominated by the experienced groups: Avante-Garde Simulation was particularly successful, finding all but one structure (XXIII E, $Z' = 2$) and academic groups using the Ψ_{mol} approach to CSP found all but the salt hydrate.

Considerably more was learned about the state of current periodic electronic structure (Ψ_{cry}) methods in the post submission analysis. Many of the dispersion-corrected density-functional approximations discussed in this book were used, i.e., variants of the D2 scheme, D3, TS, MBD, XDM, and vdW-DF, mainly combined with the generalized gradient approximation PBE functional. This reflects the prohibitive computational cost of using a hybrid functional such as PBE0 on so many crystal structures. In the ranking of different structures of the same molecule, it could be hoped that many approximations in the treatment of intermolecular forces would cancel out, provided sufficient care had been taken to eliminate numerical errors in comparing very different sized and shaped unit cells (most CSP generated structures were $Z' = 1$ with 2–8 molecules in the unit cell).

Four groups had entered reranking results, having been provided with large sets of structures by other groups with structure generation (search) capabilities. They were able to use these and the experimental structures to see the sensitivity to different models. This highlighted how much the differences in the types of intermolecular forces in the targets affected

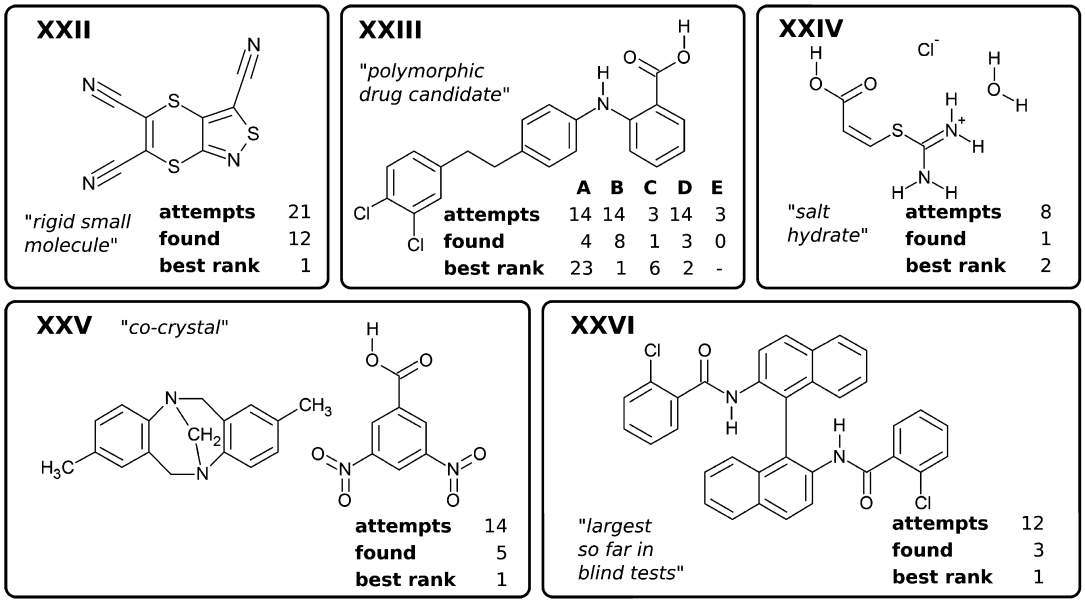

FIGURE 11.5 Target molecules and success rates in the sixth CSP blind test. For each target crystal structure, the number of attempts to predict the structure is given, along with the number of groups where the correct structure was found within their lists of 100 crystal structures. The best rank is the lowest rank in all submitted lists, e.g., a best rank of 1 means that it was found as the global energy minimum by at least one group.

which methods were necessary. The salt hydrate, with the relative positions of the three components in the asymmetric unit cell as additional search variables, was a major challenge to the search programs. The strength of the polarization around the ion means that Ψ_{mol} methods would not be expected to be able to evaluate relative lattice energies reliably. However, the Ψ_{cry} methods are able to cope with polarization and fine adjustments of the cation conformation. Moreover, the energy difference between the experimental and many other structures was sufficiently large that any reasonable Ψ_{cry} method could find the observed structure to be the most stable [67].

In contrast, the energy differences between plausible structures of the other molecules were much smaller, and the dispersion contribution was generally dominant. The relative energies of the polymorphs of XXIII were contrasted for several variants of both Ψ_{mol} and Ψ_{cry} methods, both as lattice energy differences and including various harmonic estimates of the free energy. The results show that most lattice energy models have all polymorphs within a plausible 10 kJ/mol in energy, but with major changes in the relative stability order [58]. In many cases, the free energy contribution changes the stability order of the polymorphs. Denser structures are generally stabilized by the dispersion contribution to the lattice energy and destabilized by the free energy contribution, and the density of structures of flexible molecules like XXIII is affected by how the intermolecular forces change the molecular conformation.

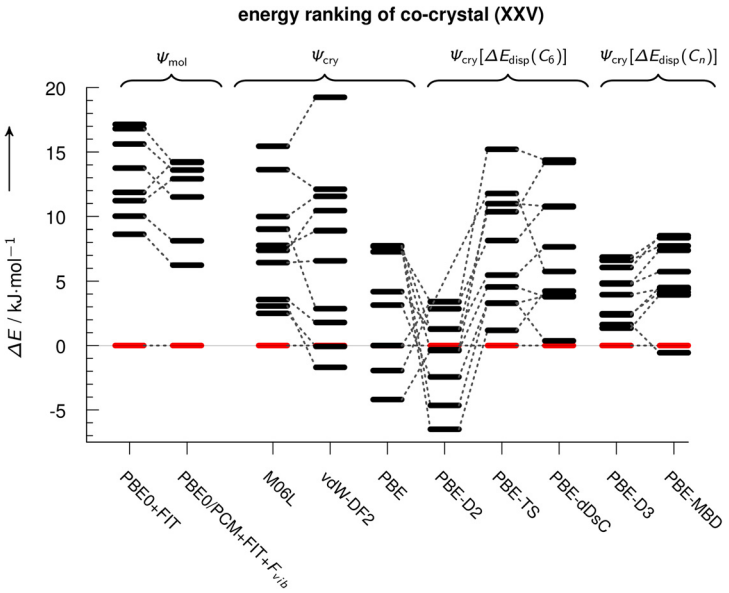

FIGURE 11.6 Lattice energies of 10 hypothetical low energy structures for cocrystal XXV relative to the experimental structure, computed with various types of methods Ψ_{mol} or Ψ_{cry}. Ψ_{mol} shows the difference of the relative free energies using a rigid-body estimate for F_{vib} (see Eq. (11.3)) and molecular charge densities calculated in a polarizable continuum. For Ψ_{cry} we discriminate between dispersion corrections that describe only the leading order dispersion coefficient (C_6) and schemes that include higher order contributions in both the many-body and the multipole sense (C_n).

The effect of the dispersion correction for Ψ_{cry} methods is more clearly seen for XXV, a cocrystal of two fairly rigid molecules (Fig. 11.6). The Ψ_{mol} methods are able to predict the known structure as the most stable, independent of whether the molecular charge density is calculated in a polarizable continuum and a free energy difference is included. Many dispersion-corrected density functionals also find the known structure as the most stable, but only by a margin that raises the question as to whether free energy differences could change this.

Hence, more accurate methods for both energy and free energy contributions are needed to confidently identify the thermodynamically most stable crystal structure. Working with the assumption that the observed structure should be the most stable, the comparison of real and hypothetical structures is clearly still a challenge to whether the method is able to balance all the different contributions to the inter and intramolecular forces adequately (see Section 11.2.2). Most significantly, the performance of a periodic DFT-D method on the relative crystal stabilities in the sixth blind test correlates well with the accuracy for absolute sublimation enthalpies as tested with the X23 (or C21) set of crystal structures of small molecules [67].

The size of the crystallographic search space means that CSP has to involve a hierarchy of methods, so that the most expensive is only used on a computationally affordable number of structures. At each stage in a hierarchy, it is necessary to cover sufficient structures to be sure

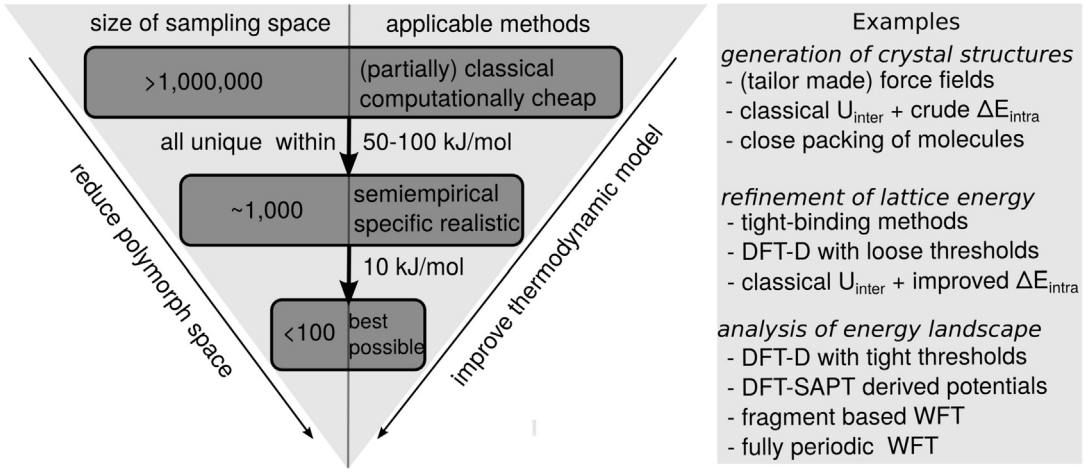

FIGURE 11.7 Sketch of possible hierarchies of methods to evaluate the crystal energies during a CSP study of a moderately sized flexible neutral molecule, with order of magnitude estimates of numbers of structures and energy differences that may be involved, which will depend on the specific molecule as well as method.

that the experimental structures are not eliminated. An illustration of possible hierarchical schemes for moderately sized molecules is shown in Fig. 11.7, looking forward to the wide application of the methods described within this book. The resources required for a thorough CSP study are determined by the specific molecule as the number of structures that are on the crystal energy landscape (i.e., within the energy range of be thermodynamically plausible polymorphs) varies enormously: monomorphic isocaffeine can be readily predicted with any reasonable CSP method, whereas there are about 500 low energy ordered crystal structures corresponding to the statically disordered low temperature crystal structure of caffeine, with energy differences that are so low that configurational entropy stabilizes this structure [68].

Before CSP can be routinely used in industry, the issue of computational cost and speed relative to the experimental timescales will have to be addressed. The reliability of different methods depends on the functional groups involved. The strong intermolecular forces involved in salt hydrates, for example, are challenging to the accuracy of many approximate density functionals (especially of the GGA level) as well as methods based on the molecular charge density. The Ψ_{mol} methods can use a far more accurate model for the charge density, and hence the electrostatic and other long-range forces, than Ψ_{cry} methods, provided that the atomistic intermolecular potential only has to be derived for a limited number of conformations.

Thus, the blind tests demonstrate whether CSP can predict the crystal structure of a molecule, prior to synthesis, and simultaneously provide a test bed for methods of predicting relative energies of known and competitive computer-generated structures. The target molecules are chosen from those where there is a well-determined crystal structure, within specified constraints or Z' and no disorder. The target structures are likely to be the polymorphs that crystallize most readily. Indeed, in some cases, further experiments have found polymorphs of target molecules after the blind tests. Whilst the blind tests are invaluable in

developing CSP methodology, the value of CSP is in what it can reveal about polymorphs and the control of crystallization processes in the manufacture of specialty organic chemicals.

11.4 CSP AS AN AID IN PHARMACEUTICAL DEVELOPMENT

The practical applications of CSP are not to predict easily crystallized structures, but to help find more polymorphs as a complement to experimental polymorph screening [3,77]. Such screening usually exposes the compounds to hundreds, even thousands, of crystallization conditions, often using automation to vary solvent, supersaturation, cooling rates, and other factors that would usually be controlled in a solvent crystallization process [78]. Crystallization from the melt is another standard means of looking for polymorphs and phase transitions for molecules that do not decompose. CSP is applied to determine the crystal energy landscape, the range of structures which are thermodynamically plausible as polymorphs, to see which unobserved types of crystal packing are thermodynamically competitive with the structures that are already known. Most pharmaceuticals that are in active development are rather larger, and more flexible, than the systems that have been extensively studied in the development of CSP methods such as the small generic drugs like aspirin and paracetamol. This brief survey of some of the recently published studies on the value of CSP in studying pharmaceutical materials (Fig. 11.8) aims to highlight how realistic modeling of noncovalent interactions could assist in predicting the properties of organic molecules that would be valuable in an industrial context [79,80]. Often this consists of giving a molecular overview of the crystallization behavior of a compound whose polymorphs do not all crystals of a size and quality suitable for experimental structure determination.

11.4.1 The Diversity of the Crystal Energy Landscape

The crystal energy landscape is very dependent on the individual molecule. For example, strychnine has one way of crystallizing with itself which is estimated to be 10 kJ/mol more stable than any other [79]. Hence, the fact that strychnine does not have polymorphs could have been confidently predicted by a far less computationally expensive method than the density functional calculations (PBE+TS) used in the CSP study. This is unusual, and the energy gaps, and hence the accuracy needed for ranking of very different types of crystal packings can differ considerably between closely related molecules.

Molecules that are similar enough to act at the same protein receptor can have dramatically different crystallization behavior. For example, two 5-HT2a agonists, LY2806920 (B5) and LY2624803 (DB7) (Fig. 11.8), which were under development for sleep disorders, posed very different challenges in developing appropriate solid forms for drug development. B5 readily and reliably crystallizes in just one solid form of the neutral molecule, whereas screening for DB7 found three neat polymorphs, two hydrates, three alcohol solvates, and an amorphous phase. Although the numbers of experiments in the solid-form screens were roughly equivalent, the ability to use an amorphous form of DB7 gave a greater ability to avoid the memory of the input material seeding the crystallizations [69]. The reason for the difference in the

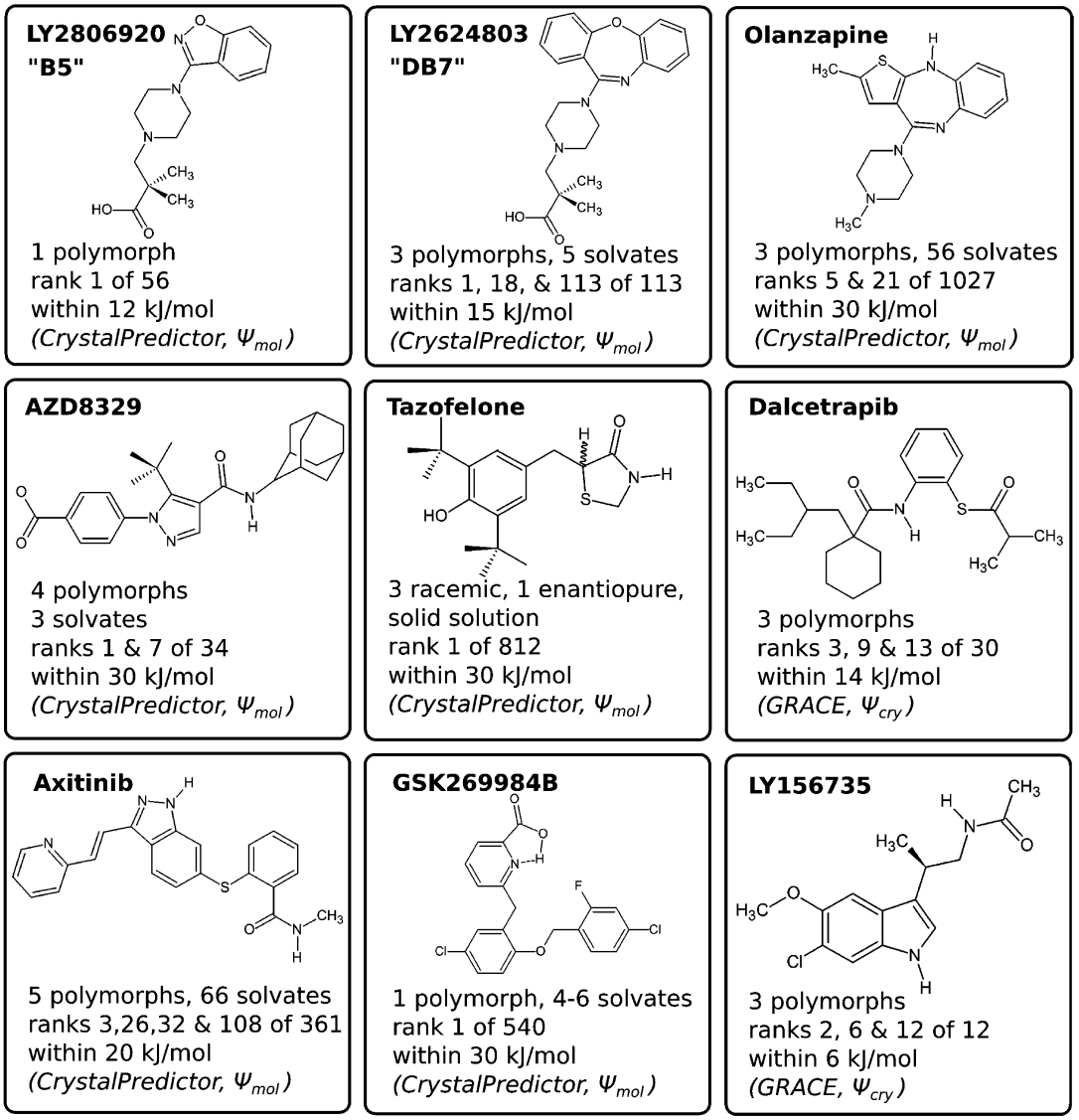

FIGURE 11.8 Pharmaceuticals discussed in text where a CSP study has been performed to complement industrial solid-form screening, with the number of experimentally determined solid forms. The number of CSP generated structures which were carefully analyzed as being within the stated range of the most stable is given. The ranking of the polymorphs is given appropriate for the molecule and study (e.g., enantiopure structures only considered in chiral space groups; $Z' = 2$ polymorphs are not found or ranked in a $Z' = 1$ search). The references to the CSP and often experimental studies are LY2806920 "B5" [69], LY2624803 "DB7" [69], olanzapine [70], AZD8329 [71], tazofelone [72], dalcetrapib [73], axitinib [74], GSK269984B [75], and LY156735 [76].

solid forms of B5 and DB7 was not in the number or spacing of the lattice energy minima, but in their nature. B5 could pack densely with itself, with most low energy structures having an internal hydrogen bond. Hence trapping a long-lived, practically important metastable polymorph of B5 in competition with the readily crystallizing structure is unlikely. In contrast DB7 had no good way of packing densely with itself, and the low energy structures for DB7 mainly had intermolecular hydrogen bonds forming a range of motifs. This promotes solvate formation and long-lived metastable polymorphs, such as form III, which could only be obtained in polycrystalline form by desolvating the zwitterionic dihydrate [81].

The question of whether a molecule is prone to solvate formation is fundamental to both the experimental screening process and the relative energies of polymorphs [4]. The small amount of molecular and conformational rearrangement that is possible as solvent is removed from a crystal structure may mean that a higher energy polymorph is kinetically trapped, and so the energy differences between conformational polymorphs can be larger than for packing polymorphs. For promiscuous solvate formers such as axitinib [74], CSP will often generate structures with voids that can be stabilized by solvents, or even form inclusion compounds, however, in other cases such as olanzapine, the majority of solvates have the solvent separating olanzapine layers that do not stack readily [70].

11.4.2 Finding the Most Stable Polymorph

It is usually preferred to use the most stable crystal form in a product to avoid the risk of transformation during storage in the range of temperatures and humidity operational throughout the world. The dangers of a late-appearing, more stable, polymorph mean that it is highly desirable to be able to confirm that the most stable crystal structure is known [45,82]. Manufacturing processes could subject the crystals to an even wider range of conditions, in particular shear pressure in a milling step which is often used to provide the required particle size distribution. Hence, the need to be able to reliably compute the relative stability of polymorphs over a wide range of conditions. There are cases of small molecules where CSP has led to the finding of the most stable form [20]. However, as already discussed (see Section 11.3), the challenge of performing a full search and accurate enough energy evaluations increases with the size of molecule or diminishing energy gaps between polymorphs. However, it may not be necessary to perform a full CSP study. For example, crizotinib, which was developed by Pfizer for the treatment of forms of lung cancer, was unusual in that extensive polymorph and hydrate screening found only one crystalline form. This structure not only had the lowest energy molecular conformation but also optimal intermolecular interactions, as confirmed by a CCDC solid form informatics "healthcheck" [83]. A simple CSP search, based on just four rigid, carefully selected conformers and the five most common chiral space groups, showed that the known structure was significantly more stable than any other generated, increasing the confidence that there was no danger of a late-appearing more stable form [84].

11.4.3 Characterizing Polymorph Structure, Disorder, and Polymorphic Purity

The structure of a polymorph is the key to understanding, and eventually computationally predicting, all the relevant properties of the crystalline form for designing the drug product

and manufacturing process [1]. Hence, there is great interest in the ability to predict the morphologies, mechanical properties [85,86], etc., of polymorphs from the crystal structure [87]. Diffraction patterns can be readily simulated from CSP-generated structures and hence the output of a search can help with the structural characterization of samples when there is insufficient data for a confident assignment [88]. Many polymorphs do not readily grow crystals suitable for a single-crystal diffraction determination of the structure. Indeed, it has only recently become routine in academic labs to check that the single crystal selected is typical of the sample by comparing its simulated powder pattern to that of the experimental powder pattern of the batch of solid material. Concomitant crystallization, where more than one polymorph appears in the same sample, is relatively common [89]. Hence, a polycrystalline sample may be a mixture of polymorphs. Alternatively, the sample crystallization conditions can have a major influence on the degree of disorder within the crystallites. The use of CSP in combination with powder diffraction and solid-state NMR was able to solve the structure of form III of DB7 as an intrinsically disordered polymorph, obtained only by dehydration [81]. Whether a microcrystalline sample is a physical mixture of polymorphs or one form with variable disorder makes a huge difference to the problems that have to be circumvented in pharmaceutical development for quality control.

The Lilly blockbuster olanzapine, marketed for the treatment of schizophrenia, generated a lot of patent claims of "novel" polymorphs by generic companies. In many cases, the purported new forms were mixtures of the concomitantly crystallizing metastable polymorphs, forms II and III. The structure of form II was only determined in the year that olanzapine came off patent, when a single crystal suitable for X-ray could be picked out from a sample of olanzapine that had failed to cocrystallize with nicotinamide. It has not yet been possible to isolate a pure form III sample, but its powder diffraction pattern can be corrected for the presence of form II. The olanzapine crystal energy landscape included a structure that was a sufficient match to the form III powder pattern to give a structural model of form III as a different stacking of the same molecular layers as form II, rationalizing why forms II and III always crystallize concomitantly [70]. If the CSP study is being done only to help propose structural models from limited experimental data, then a more limited CSP search may be appropriate. For example, to find structures for the anhydrous forms of the antibiotic levofloxacin, an analysis was made of crystal structures of six carboxylic acid salt and hydrate forms to choose six likely π stacked dimer structures, which were optimized by electronic structure methods and held rigid during the CSP search [90]. Alternatively, if the space group can be established, a more complete search can be done on that specific space group as done for naproxen [47].

An increasing variety of methods are being used to characterize organic solid state samples, and so the range of spectral properties that are being calculated from the CSP-generated structures is increasing. An extreme example of the use of the calculated spectra of the structures on a crystal energy landscape was the characterization of a new polymorph of theophylline by electron diffraction, from a single crystallite polymorphic impurity. This polymorph could not be detected by powder X-ray diffraction on the sample [91]. The simulation of the solid state NMR spectrum from the crystal structure is more challenging, requiring special functionalized DFT or molecular methods (see Chapter 10) [92]. This is being combined with CSP for structural characterization, for example, for AZD8329, an 11-HSD1 inhibitor investigated for use in the treatment of type 2 diabetes. Form 4, one of two forms considered

to have superior properties for development, had a structure proposed by comparing the experimental proton solid state NMR spectrum with those calculated from CSP generated structures with both the *cis* and *trans* amide conformations [71].

A major strength of CSP is that it can generate closely related structures that are so similar in energy that the alternative packings are likely to appear in the crystals, either as growth errors or from configurational entropy stabilization of disorder [68]. For example, a study on tazofelone unexpectedly found a crystal of form III that had the same layer structure as known forms I and II, but was stacked in a different way [72]. The crystal energy landscape found alternative stackings of this layer close in energy to form II, the most stable form. This helped rationalize why the large single crystals of each polymorph varied in melting point, as there were sufficient stacking faults to be apparent in the raw diffraction data.

11.4.4 Suggesting Experiments to Find Predicted Polymorphs

In the long-term, there should be sufficient understanding of the factors determining how a molecule crystallizes that the prediction of a polymorph would be associated with a suggestion of the experiment in which it could be found [44,93]. This is a distant goal, but there is some evidence of how solution behavior can modulate the polymorph form [94]. In the case of GSK269984B, the low energy structures on the crystal energy landscape had intermolecular hydrogen bonding compensating for adopting grossly different, higher energy conformations than in the observed most stable, internally hydrogen-bonded structure [75]. Further screening concentrated on solvents that would be likely to hydrogen bond to the active pharmaceutical ingredient (API). This produced some metastable solvates with the expected intermolecular hydrogen bonding, but the same gross conformation as in the neat form. Although solution NMR showed that a range of other conformations could exist in solution, it appeared that the fast crystallization of GSK269984B into its most stable form prevented the crystallization of metastable polymorphs with a different gross conformation as well as hydrogen bonding.

The use of pressure to favor higher density polymorphs has been successful in finding a polymorph that had been missed in extensive experimental screening of Roche's CETP inhibitor, dalcetrapib. The crystal energy landscape had two structures denser than the known stable form but very close in energy (based on the PBE-D3 level), which became more stable than the observed polymorph with a modest increase in pressure. Experiments recrystallizing dalcetrapib, either from solution or the melt, in a diamond anvil cell under modest pressure formed a new polymorph [73]. This matched the predicted polymorph, except for disorder in the hydrocarbon tail which could have been anticipated from other structures on the crystal energy landscape.

More specific targeting is becoming possible with the appreciation of the role of impurity crystals in templating the first nucleation of a crystal. The predicted cocrystallization of caffeine with benzoic acid needed an initial seeding with an isostructural cocrystal of caffeine with a fluorobenzoic acid [11]. New CSP-predicted polymorphs of carbamazepine (form V) and cyheptamide (form III) were found by sublimation onto an isostructural template of dihydrocarbamazepine form II. These polymorphs had not been found by extensive traditional polymorph screening [95]. Since synthetic impurities often have a marked effect on

the crystallization behavior, there is clearly considerable potential for templating different polymorphs, if we understood nucleation and growth better.

11.5 CHIRAL SEPARATION BY CRYSTALLIZATION

The separation of enantiomers by crystallization is an important industrial process, given their difference in biological activity [96]. Crystal structure prediction has always been recognized as a potentially important tool in designing such processes, as it should provide the lowest energy structure in the racemic and chiral space groups and their energy difference. However, it is clear that several complications can arise. The first problem may come with the assumption that the molecules will crystallize in the most stable crystal structure: a CSP study on 3-chloromandelic acid generated many structures that were simpler, denser, and thermodynamically competitive, if not more stable (using a range of different approaches), than the known structures [19]. Consideration of both experimental and computational evidence on the mandelic acids and other flexible chiral molecules raises questions of the role of the encounter statistics in crystallization kinetics in determining which structures can be nucleated and observed [97]. Indeed, experimental screening of LY1567365 could not find a crystal structure that was generated by CSP, but it was found in experiments using the inactive enantiomer [76].

Another problem may be the role of templating between crystal structures of large molecules where the chiral center has little influence on the packing preferences of the rest of the molecule. Tazofelone illustrates this, as seeding a melt of the two enantiomers with the enantiopure ($Z' = 2$) crystal produces an isostructural solid solution [72]. The CSP study generated an isostructural racemic form more stable than the enantiopure crystal (but less stable than the racemic polymorphs). The two conformations in the enantiopure crystal could be readily replaced by molecules of the opposite chirality, meaning that the crystallization of enantiopure tazofelone could readily incorporate impurities of the other hand.

However, despite these counterexamples, CSP has been used to successfully predict the most stable crystal structures of the enantiopure and racemic compound for many molecules. The ability to separate out just one hand of the molecule from a racemic mixture depends on the relative solubility of the two enantiopure and racemic crystals. Predicting solubility is notoriously difficult but the solvation energy in an achiral solvent of the two enantiomers will be identical if the solution is ideal [98]. Similarly, the energy of a gas is the same regardless of the ratio of the two hands of the molecules. The melts of the racemic and enantiopure crystals may not be too different, depending on the molecule.

Hence, if we could assume that the heat capacities and zero-point energies (F_{vib}, see Eq. (11.3)) are the same for both crystals, the lattice energy difference corresponds to the free energy difference between the crystals. The enantiomeric excess (ee), the excess of one enantiomer over the other that results from crystallizing a racemic mixture (ee = 0), depends exponentially on the solubility difference. The enantiomeric excess has been predicted with reasonable success for 10 peptides by using lattice energy differences [99]. However, the difference in lattice energies will only dominate the difference in relative thermodynamic

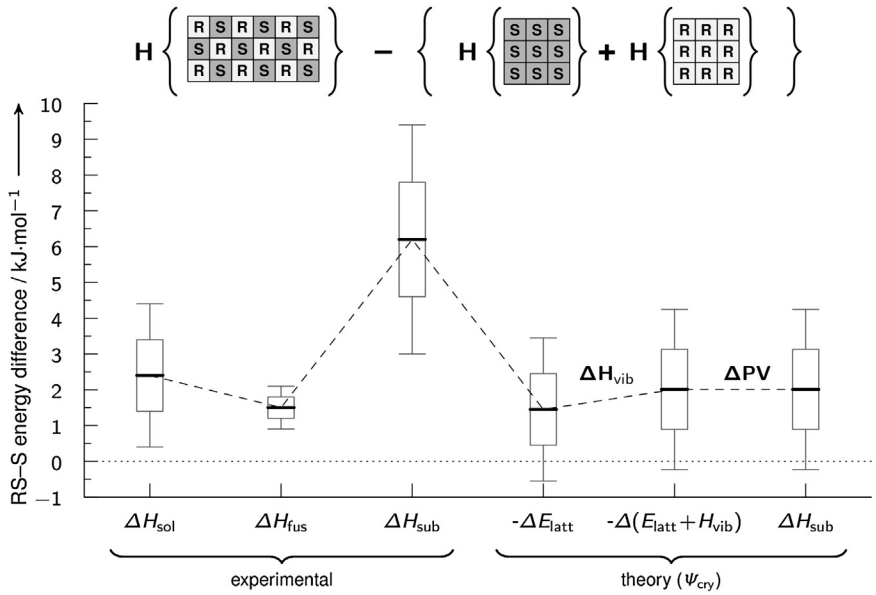

FIGURE 11.9 Illustration of the different approximations to the difference in energies of the enantiopure and the more stable racemic crystal structures of naproxen. The experimental energies are contrasted with a TPSS-D3 estimate of the lattice energy, with the addition of harmonic phonon contributions, and the pressure–volume term to yield ΔH_{sub}. The experimental energy difference derived from the solubility difference in ethanol/water [47] and the sublimation enthalpies are at normal conditions [100], while the enthalpy of fusion is measured at 427 K [47]. The boxplot gives the 68% and 95% confidence intervals either estimated from statistical experimental errors or, in the case of theory, estimated from typical errors derived from benchmark sets of organic crystals [99].

quantities for chiral systems, such as the sublimation energy or solubility difference, if this is larger than the errors in the assumptions being made [47].

This is illustrated by the antiinflammatory naproxen (Fig. 11.3), where the relative heats of solution, ΔH_{sol} [47], differences in heats of fusion, ΔH_{fus} [47], and sublimation enthalpy, ΔH_{sub} [100], between the racemic and enantiopure crystals have been measured [100]. These data (Fig. 11.9) clearly show that the assumptions above, which lead to the approximate relationships between the crystal energies,

$$\Delta H^{\text{cry}}_{\text{R+S−RS}} \approx \Delta H_{\text{sol}} \approx \Delta H_{\text{fus}} \approx \Delta H_{\text{sub}} \approx -\Delta E_{\text{latt}}, \tag{11.5}$$

do not hold, even allowing for the experimental errors. For example, the experimental errors in measuring the two heats of fusion may be larger than the difference between the racemic and enantiopure liquid state.

Computationally, the Ψ_{mol} methods used in the original study (Fig. 11.4 [47]) can be improved by a recent Ψ_{cry} study (TPSS-D3, Fig. 11.9). The variations in the relative lattice energies E_{latt} with method will be qualitatively similar to those shown for the cocrystal XXV in Fig. 11.6, and hence could be larger the experimental differences. Fig. 11.9 also examines the approximations in the calculations of ΔH_{sub}. The pressure–volume work (ΔPV) contri-

bution to ΔH_{sub} is only 0.01 kJ/mol, though this will increase when there are larger density differences between the crystals. The thermal contribution, H_{vib}, is more problematic to calculate, with an error of 10 kJ/mol when using the very different unit cells rather than sampling the Brillouin zones sufficiently, a common problem in comparing polymorphs [17]. However, the converged difference from harmonic values is 0.5 kJ/mol, which, although small, could make a difference in the eutectic composition of up to 10%. Hence, as Fig. 11.9 indicates, the computational prediction of the parameters needed for designing chiral resolution processes is still an even greater challenge to theoretical methods than CSP.

11.6 CONCLUSIONS AND FUTURE CHALLENGES

The ability to predict organic crystal structures is both a stringent test of our ability to model intermolecular forces, and an industrially and scientifically relevant application of computer modeling. The vision of being able to design a pharmaceutical product aided by computer simulation is an exciting challenge to our fundamental scientific understanding of what determines the behavior of these molecules and how this can be encapsulated in computer codes. It requires not only the prediction of all potential solid forms, but also of all their properties relevant to processing, manufacture and use.

CSP has shown up the inadequacies of simple, transferable, isotropic atom–atom FFs, of the type in most commercial modeling packages, for molecule-specific modeling of the organic solid state. The success of GRACE in using a molecule-specific tailored force-field [66] derived by fitting to DFT-D data, of a functional form similar to the classical transferable force-fields, in the first step of crystal structure generation, shows the potential and limitations of optimizing the parameterization of this functional form.

The accurate evaluation of the relative energies (and solubilities) of polymorphs remains a challenge to computational chemistry, with the adequate evaluation of the interatomic forces as a major part of the challenge. Two approaches are currently being used, often successfully for practical purposes, in CSP on large molecules, namely those which rely on the theory of intermolecular forces and require calculations based on the molecular charge density (Ψ_{mol}) and periodic electronic structure modeling (Ψ_{cry}). Both methods have different strengths and weaknesses depending on different types and sizes of the target molecules.

The recent blind test shows that methods for evaluating the crystal energies, based on the various modern approaches to evaluating intermolecular forces are converging to being able to predict relative lattice energies well within 10 kJ/mol even for quite large molecules. Unfortunately, these errors are significant compared with polymorphic energy differences. It depends on the number and relationship of the low energy CSP generated structures whether this accuracy is sufficient to be able to predict the most thermodynamically stable structure at normal conditions.

Periodic electronic structure calculations (Ψ_{cry}) require less assumptions and can be performed more automatically, but are still sensitive to the density functional and dispersion correction. Additionally, the computational costs can be high, and scale poorly with the number of crystal structures being used. The atomistic modeling of intermolecular forces, using anisotropic atom–atom models, with distributed multipoles and empirically fitted potentials,

has the advantage of absorbing at least some of the errors [101]. The model for the intermolecular interactions can be systematically improved using monomer properties and SAPT calculations of the various components such as anisotropic repulsion. Such intermolecular potentials, once derived, can be used very cheaply to model the crystal structures of rigid molecules. However, for flexible molecules, the conformation dependence of the intermolecular potential is an issue, and the need to allow the molecular conformation to adjust in response to the packing forces becomes computationally expensive. There is some hope that separating the intermolecular from intramolecular terms in the force-field may be a way to make anisotropic atom–atom force-fields more accurate, thereby by-passing the current need to do ab initio calculations on the molecule in each conformation relevant to its crystal packing [101].

Hence, molecular crystal structure prediction provides both a challenge in modeling accuracy, but also in modeling speed, so that a range of properties can be evaluated, and the calculations performed on a time and resource scale suitable for application in industry. Molecular dynamics studies will be needed for determining which CSP generated crystal structures are free energy minima, and which artifacts of the neglect of temperature, but this requires a force-field that is sufficiently realistic. Crystallization is a complex phenomenon, as shown by the phenomenon of polymorphism. However, separating out thermodynamics from kinetic aspects of nucleation and growth to allow predictive control of crystallization will remain both a fundamental and practically vital scientific question for some time to come.

Acknowledgments

This work was supported by EPSRC (EP/K039229/1). J.G.B. acknowledges support by the Alexander von Humboldt foundation within the Feodor Lynen program.

References

[1] C.C. Sun, J. Pharm. Sci. 98 (2009) 1671–1687.
[2] A. Gavezzotti, Molecular Aggregation: Structure Analysis and Molecular Simulation of Crystals and Liquids, Oxford University Press, 2006.
[3] Y.A. Abramov, Computational Pharmaceutical Solid State Chemistry, John Wiley & Sons, 2016.
[4] A.J. Cruz-Cabeza, S.M. Reutzel-Edens, J. Bernstein, Chem. Soc. Rev. 44 (2015) 8619–8635.
[5] S.T. Beckett, Science of Chocolate, RSC Paperbacks, The Royal Society of Chemistry, 2000.
[6] T. Hasell, J.L. Culshaw, S.Y. Chong, M. Schmidtmann, M.A. Little, K.E. Jelfs, E.O. Pyzer-Knapp, H. Shepherd, D.J. Adams, G.M. Day, A.I. Cooper, J. Am. Chem. Soc. 136 (2014) 1438–1448.
[7] J.T.A. Jones, T. Hasell, X. Wu, J. Bacsa, K.E. Jelfs, M. Schmidtmann, S.Y. Chong, D.J. Adams, A. Trewin, F. Schiffman, F. Cora, B. Slater, A. Steiner, G.M. Day, A.I. Cooper, Nature 474 (2011) 367–371.
[8] E.O. Pyzer-Knapp, H.P.G. Thompson, F. Schiffmann, K.E. Jelfs, S.Y. Chong, M.A. Little, A.I. Cooper, G.M. Day, Chem. Sci. 5 (2014) 2235–2245.
[9] S. Perez-Lloret, M.V. Rey, P.L. Ratti, O. Rascol, Fundam. Clin. Pharmacol. 27 (2013) 81–95.
[10] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, J. Morris, Pharm. Res. 18 (2001) 859–866.
[11] D.-K. Bučar, R.W. Lancaster, J. Bernstein, Angew. Chem. Int. Ed. 54 (2015) 6972–6993.
[12] A. Gavezzotti, Acc. Chem. Res. 27 (1994) 309–314.
[13] J. Maddox, Nature 335 (1988) 201.
[14] L. Yu, Acc. Chem. Res. 43 (2010) 1257–1266.
[15] M. Vasileiadis, A.V. Kazantsev, P.G. Karamertzanis, C.S. Adjiman, C.C. Pantelides, Acta Crystallogr. B 68 (2012) 677–685.
[16] A.J. Cruz-Cabeza, J. Bernstein, Chem. Rev. 114 (2014) 2170–2191.

[17] J. Nyman, G.M. Day, CrystEngComm 17 (2015) 5154–5165.
[18] S.A. Barnett, C.K. Broder, K. Shankland, W.I.F. David, R.M. Ibberson, D.A. Tocher, Acta Crystallogr. B 62 (2006) 287–295.
[19] R.K. Hylton, G.J. Tizzard, T.L. Threlfall, A.L. Ellis, S.J. Coles, C.C. Seaton, E. Schulze, H. Lorenz, A. Seidel-Morgenstern, M. Stein, S.L. Price, J. Am. Chem. Soc. 137 (2015) 11095–11104.
[20] D.E. Braun, H. Oberacher, K. Arnhard, M. Orlova, U.J. Griesser, CrystEngComm 18 (2016) 4053–4067.
[21] H.P.G. Thompson, G.M. Day, Chem. Sci. 5 (2014) 3173–3182.
[22] A.D. Mighell, V.L. Himes, J.R. Rodgers, Acta Crystallogr. A 39 (1983) 737–740.
[23] C.R. Groom, I.J. Bruno, M.P. Lightfoot, S.C. Ward, Acta Crystallogr. B 72 (2016) 171–179.
[24] K.M. Steed, J.W. Steed, Chem. Rev. 115 (2015) 2895–2933.
[25] H.C.J. Berendsen, Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics, Cambridge University Press, 2007.
[26] W.L. Jorgensen, J. Tirado-Rives, Proc. Natl. Acad. Sci. USA 102 (2005) 6665–6670.
[27] M.A. Neumann, J. Phys. Chem. B 112 (2008) 9810–9829.
[28] G.M. Day, W.D.S. Motherwell, W. Jones, Cryst. Growth Des. 5 (2005) 1023–1033.
[29] T.G. Cooper, K.E. Hejczyk, W. Jones, G.M. Day, J. Chem. Theory Comput. 4 (2008) 1795–1805.
[30] E.O. Pyzer-Knapp, H.P.G. Thompson, G.M. Day, Acta Crystallogr. B 72 (2016) 477–487.
[31] S.L. Price, M. Leslie, G.W.A. Welch, M. Habgood, L.S. Price, P.G. Karamertzanis, G.M. Day, Phys. Chem. Chem. Phys. 12 (2010) 8478–8490.
[32] J. Nyman, O.S. Pundyke, G.M. Day, Phys. Chem. Chem. Phys. 18 (2016) 15828–15837.
[33] G.M. Day, S.L. Price, J. Am. Chem. Soc. 125 (2003) 16434–16443.
[34] A.V. Kazantsev, P.G. Karamertzanis, C.S. Adjiman, C.C. Pantelides, J. Chem. Theory Comput. 7 (2011) 1998–2016.
[35] C.C. Pantelides, C.S. Adjiman, A.V. Kazantsev, Top. Curr. Chem. 345 (2014) 25–58.
[36] M. Habgood, I.J. Sugden, A.V. Kazantsev, C.S. Adjiman, C.C. Pantelides, J. Chem. Theory Comput. 11 (2015) 1957–1969.
[37] B.P. van Eijck, J. Kroon, J. Phys. Chem. B 101 (1997) 1096–1100.
[38] M. Rossi, P. Gasparotto, M. Ceriotti, Phys. Rev. Lett. 117 (2016) 115702.
[39] P. Raiteri, R. Martoňák, M. Parrinello, Angew. Chem. Int. Ed. 44 (2005) 3769–3773.
[40] P.G. Karamertzanis, P. Raiteri, M. Parrinello, M. Leslie, S.L. Price, J. Phys. Chem. 112 (2008) 4298–4308.
[41] J. van de Streek, Acta Crystallogr. B 62 (2006) 567–579.
[42] J. Bernstein, R.E. Davis, L. Shimoni, N.-L. Chang, Angew. Chem. Int. Ed. 34 (1995) 1555–1573.
[43] T. Gelbrich, M.B. Hursthouse, CrystEngComm 7 (2005) 324–336.
[44] S.L. Price, Acta Crystallogr. B 69 (2013) 313–328.
[45] D.E. Braun, M. Orlova, U.J. Griesser, Cryst. Growth Des. 14 (2014) 4895–4900.
[46] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, G.K.-L. Chan, Science 345 (2014) 640–643.
[47] D.E. Braun, M. Ardid-Candel, E. D'Oria, P.G. Karamertzanis, J.-B. Arlin, A.J. Florence, A.G. Jones, S.L. Price, Cryst. Growth Des. 11 (2011) 5659–5669.
[48] J.R. Holden, Z. Du, H.L. Ammon, J. Comput. Chem. 14 (1993) 422–437.
[49] S. Grimme, A. Hansen, J.G. Brandenburg, C. Bannwarth, Chem. Rev. 116 (2016) 5105–5154.
[50] G.J.O. Beran, Chem. Rev. 116 (2016) 5567–5613.
[51] H.S. Muddana, C. Daniel Varnado, C.W. Bielawski, A.R. Urbach, L. Isaacs, M.T. Geballe, M.K. Gilson, J. Comput.-Aided Mol. Des. 26 (2012) 475–487.
[52] H.S. Muddana, A.T. Fenley, D.L. Mobley, M.K. Gilson, J. Comput.-Aided Mol. Des. 28 (2014) 305–317.
[53] 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, Acta Crystallogr. B 56 (2000) 697–714.
[54] W.D.S. Motherwell, et al., Acta Crystallogr. B 58 (2002) 647–661.
[55] G.M. Day, et al., Acta Crystallogr. B 61 (2005) 511–527.
[56] G.M. Day, et al., Acta Crystallogr. B 65 (2009) 107–125.
[57] D.A. Bardwell, et al., Acta Crystallogr. B 67 (2011) 535–551.
[58] A.M. Reilly, et al., Acta Crystallogr. B 72 (2016) 439–459.
[59] J. Kendrick, M.D. Gourlay, M.A. Neumann, F.J.J. Leusen, CrystEngComm 11 (2009) 2391–2399.
[60] A.J. Misquitta, G.W. Welch, A.J. Stone, S.L. Price, Chem. Phys. Lett. 456 (2008) 105–109.

[61] M.A. Neumann, M.-A. Perrin, J. Phys. Chem. B 109 (2005) 15531–15541.
[62] S. Grimme, J. Comput. Chem. 25 (2004) 1463–1473.
[63] J. Kendrick, F.J. Leusen, M.A. Neumann, J. Comput. Chem. 33 (2012) 1615–1622.
[64] D.E. Braun, R.M. Bhardwaj, A.J. Florence, D.A. Tocher, S.L. Price, Cryst. Growth Des. 13 (2013) 19–23.
[65] A.V. Kazantsev, P.G. Karamertzanis, C.S. Adjiman, C.C. Pantelides, S.L. Price, P.T. Galek, G.M. Day, A.J. Cruz-Cabeza, Int. J. Pharm. 418 (2011) 168–178.
[66] M.A. Neumann, J. Phys. Chem. B 112 (2008) 9810–9829.
[67] J.G. Brandenburg, S. Grimme, Acta Crystallogr. B 72 (2016) 502–513.
[68] M. Habgood, Cryst. Growth Des. 11 (2011) 3600–3608.
[69] D.E. Braun, J.A. McMahon, L.H. Koztecki, S.L. Price, S.M. Reutzel-Edens, Cryst. Growth Des. 14 (2014) 2056–2072.
[70] R.M. Bhardwaj, L.S. Price, S.L. Price, S.M. Reutzel-Edens, G.J. Miller, I.D.H. Oswald, B.F. Johnston, A.J. Florence, Cryst. Growth Des. 13 (2013) 1602–1617.
[71] M. Baias, J.-N. Dumez, P.H. Svensson, S. Schantz, G.M. Day, L. Emsley, J. Am. Chem. Soc. 135 (2013) 17501–17507.
[72] L.S. Price, J.A. McMahon, S.R. Lingireddy, S.-F. Lau, B.A. Diseroad, S.L. Price, S.M. Reutzel-Edens, J. Mol. Struct. 1078 (2014) 26–42.
[73] M.A. Neumann, J. van de Streek, F.P.A. Fabbiani, P. Hidber, O. Grassmann, Nat. Commun. 6 (2015) 7793.
[74] M. Vasileiadis, C.C. Pantelides, C.S. Adjiman, Chem. Eng. Sci. 121 (2015) 60–76.
[75] S.Z. Ismail, C.L. Anderton, R.C.B. Copley, L.S. Price, S.L. Price, Cryst. Growth Des. 13 (2013) 2396–2406.
[76] J. Kendrick, G.A. Stephenson, M.A. Neumann, F.J.J. Leusen, Cryst. Growth Des. 13 (2013) 581–589.
[77] S.L. Price, L.S. Price, in: Solid State Characterisation of Pharmaceuticals, John Wiley & Sons, 2011, Chapter 12.
[78] M.L. Peterson, et al., J. Am. Chem. Soc. 124 (2002) 10958–10959.
[79] S.L. Price, D.E. Braun, S.M. Reutzel-Edens, Chem. Commun. 52 (2016) 7065–7077.
[80] S.L. Price, S.M. Reutzel-Edens, Drug Discov. Today 21 (2016) 912–923.
[81] D.E. Braun, L.H. Koztecki, J.A. McMahon, S.L. Price, S.M. Reutzel-Edens, Mol. Pharm. 12 (2015) 3069–3088.
[82] D.E. Braun, H. Oberacher, K. Arnhard, M. Orlova, U.J. Griesser, CrystEngComm 18 (2016) 4053–4067.
[83] N. Feeder, E. Pidcock, A.M. Reilly, G. Sadiq, C.L. Doherty, K.R. Back, P. Meenan, R. Docherty, J. Pharm. Pharmacol. 67 (2015) 857–868.
[84] Y.A. Abramov, Org. Process Res. Dev. 17 (2013) 472–485.
[85] M.H. Shariare, F.J.J. Leusen, M. de Matas, P. York, J. Anwar, Pharm. Res. 29 (2012) 319–331.
[86] G.M. Day, S.L. Price, M. Leslie, Cryst. Growth Des. 1 (2001) 13–27.
[87] P. Dandekar, Z.B. Kuvadia, M.F. Doherty, Annu. Rev. Mater. Res. 43 (2013) 359–386.
[88] S. Habermehl, P. Mörschel, P. Eisenbrandt, S.M. Hammer, M.U. Schmidt, Acta Crystallogr. B 70 (2014) 347–359.
[89] J. Bernstein, R.J. Davey, J.-O. Henck, Angew. Chem. Int. Ed. 38 (1999) 3440–3461.
[90] S.S. Singh, T.S. Thakur, CrystEngComm 16 (2014) 4215–4230.
[91] M.D. Eddleston, K.E. Hejczyk, E.G. Bithell, G.M. Day, W. Jones, Chem. Eur. J. 19 (2013) 7883–7888.
[92] E. Salager, G.M. Day, R.S. Stein, C.J. Pickard, B. Elena, L. Emsley, J. Am. Chem. Soc. 132 (2010) 2564–2566.
[93] J. Anwar, D. Zahn, Angew. Chem. Int. Ed. 50 (2011) 1996–2013.
[94] R.J. Davey, S.L.M. Schroeder, J.H. ter Horst, Angew. Chem. Int. Ed. 52 (2013) 2166–2179.
[95] V.K. Srirambhatla, R. Guo, S.L. Price, A.J. Florence, Chem. Commun. 52 (2016) 7384–7386.
[96] H. Lorenz, A. Seidel-Morgenstern, Angew. Chem. Int. Ed. 53 (2014) 1218–1250.
[97] A. Gavezzotti, L.L. Presti, Cryst. Growth Des. 15 (2015) 3792–3803.
[98] R.E. Skyner, J.L. McDonagh, C.R. Groom, T. van Mourik, J.B.O. Mitchell, Phys. Chem. Chem. Phys. 17 (2015) 6174–6191.
[99] A. Otero-de-la-Roza, B.H. Cao, I.K. Price, J.E. Hein, E.R. Johnson, Angew. Chem. Int. Ed. 53 (2014) 7879–7882.
[100] H. Buchholz, V.N. Emelyanenko, H. Lorenz, S.P. Verevkin, J. Pharm. Sci. 105 (2016) 1676–1683.
[101] O.G. Uzoh, P.T.A. Galek, S.L. Price, Phys. Chem. Chem. Phys. 17 (2015) 7936–7948.