The potential of computed crystal energy landscapes to aid solid form development

Sarah L. Price, Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK

Susan M. Reutzel-Edens, Small Molecule Design and Development, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, USA

Corresponding author: Price S.L. (s.l.price@ucl.ac.uk, +44 (0)20 7679 4622)

Keywords: Solid forms; Crystal Structures; Drug development; Computational Prediction; Polymorphism; Crystal Structure Prediction

Teaser Could computational crystal structure prediction accelerate solid form development?

Highlight Crystal Structure Prediction studies carried out with the pharmaceutical industry.

Bios:

Sarah (Sally) Price is a theoretical and computational chemist by training, who received her PhD for work on modelling intermolecular forces from the University of Cambridge in 1981. After postdoctoral work at University of Chicago and back in Cambridge, she moved to University College London, where she is now a Professor of Physical Chemistry. In 2015 she received a Royal Society of Chemistry Interdisciplinary Prize.

Susan M. Reutzel-Edens is a Senior Research Advisor in Small Molecule Design & Development at Eli Lilly and Company. She obtained her BS degree in chemistry from Winona State University (1987), then earned her PhD in organic chemistry at the University of Minnesota (1991). Susan brought her experience in hydrogen-bond directed co-crystallization and interest in crystal polymorphism to Eli Lilly, where she developed Lilly's solid form design program and for two decades led a team of cross-functional scientists charged with finding commercially-viable crystalline forms for small-molecule drug products.

Sarah Price

Susan Reutzel-Edens

Abstract

Solid form screening to identify all solid forms of an API has become increasingly important in ensuring Quality by Design of pharmaceutical products and their manufacturing processes. However, despite considerable enlargement of the range of techniques that have been shown capable of producing novel solid forms, it is possible that practically important forms may not be found in the short timescales currently allowed for solid form screening. Here, we report on the state-of-the-art use of computed crystal energy landscapes to complement pharmaceutical solid form screening. We illustrate how crystal energy landscapes can help establish a molecular level understanding of the crystallization behavior of APIs and enhance the ability of solid form screening to facilitate pharmaceutical development.

Key Resources

Books: [1] Bernstein, J. (2002) *Polymorphism in Molecular Crystals*, Clarendon Press

[2] Hilfiker, R. (2006) *Polymorphism in the Pharmaceutical Industry*, Wiley-VCH

Recent Reviews: [3]Facts and Fictions about Polymorphism

[4] Predicting crystal structures of organic compounds (a tutorial review)

[5] Why don't we find more polymorphs?

[6] Report on the 6th Blind test of crystal structure prediction

Glossary

API: Active pharmaceutical ingredient

Polymorphs: different crystal structures with the same chemical composition with the molecule(s) having a defined covalent bonding and stereochemistry. Like many solid state forms which can form a continuum [7], the definition can vary, but herein racemic crystals

containing the API and its mirror image, and conglomerates of enantiopure structures containing just one hand, are considered as polymorphs. *Conformational polymorphs* arise when the molecular conformation in each form differs significantly, such that the closest isolated molecule conformational energy minima are different [8]. ~~*Packing polymorphs* have the same molecular conformation packed in different ways.~~

Crystal form: general term for a single or multi-component crystalline solid. Crystal forms of a parent compound, salt or co-crystal that include the solvent of crystallization in the crystal structure are referred to as solvates and in the specific case where water is present, hydrates. Neat or non-solvated polymorphs refer to different crystal structures containing just the API with defined covalent bonding. The term *solid forms* includes both crystal and amorphous forms.

Polymorph screen: a survey of crystallization conditions designed to identify solid forms of an API. Solvent-based screens are oftentimes tailored to the solubility properties of the API. A screen may encompass hundreds or even thousands of crystallization experiments.

Concomitant crystallization: when two or more polymorphs are crystallized in the same experiment.

Packing coefficient: the fraction of space in the crystal structure occupied by molecules or ions as defined by their van der Waals surfaces. The packing coefficient for spheres expressed as a percentage is 74%.

Lattice energy: the energy required to separate the molecules in an infinite static perfect crystal lattice so that they are non-interacting and in their lowest energy conformation. This can be calculated at the electronic level, usually by periodic dispersion corrected density functional methods (DFT-D), or by splitting into inter and intramolecular contributions and using ab initio calculations on the molecule for the conformational energy and an accurate

model for the charge density which is used in modelling the intermolecular interactions in the crystal. The lattice energy can also be estimated by transferable force-fields, but this is rarely sufficiently accurate ~~enough~~.

Crystal Structure Prediction (CSP) methods were originally designed to predict the crystal structure as the most stable possibility. The prevalence of polymorphism means that these methods are now used to generate a (computed) *crystal energy landscape*, the set of structures which are thermodynamically plausible as polymorphs.

One of the first steps taken in developing a solid oral dosage form is to identify crystalline forms of the drug molecule, generally through some form of crystallization screening. Crystallization provides a means to purify and recover the drug substance coming out of the final step of the synthesis and to isolate the drug in a crystalline form that is suitable for downstream processing. Except in cases where an amorphous drug product is anticipated, crystallization is also used to define the material properties (e.g. stability, solubility) of the drug substance that will ensure consistency in the safety and efficacy profile of the drug product throughout its shelf life. Some compounds crystallize easily in solid forms that can be readily commercialized; however, all compounds are different and for some, the effort spent trying to crystallize the compound the first time, or in a form with suitable properties, is enormous. Other molecules crystallize in many different forms (including salts and co-crystals), creating opportunities to design properties into the drug substance, but at the same time introducing challenges in deciding which form to develop, selectively crystallizing the preferred form at large scale and establishing control strategies around crystal form in the drug product.

There is presently no way of knowing if a molecule will crystallize, let alone in how many forms or if any of those forms will be useful for ultimately delivering the drug from the dosage form to its site of action. Because solid form screening continues to be an exercise of trial and error, with no clear end point, the process is oftentimes extended well beyond the point of diminishing returns in order to provide reassurance that the relevant forms (polymorphs, hydrates) have been found. Under immense pressure to shorten development timelines and reduce costs, the pharmaceutical industry is keen on right-sizing the time and effort spent on finding commercially-viable solid forms. On the other hand, faced with the potential for a product failure should a new, more stable (less soluble) crystal form suddenly

appear in a marketed product, finding the most stable crystal form is non-negotiable. Crystal structure prediction (CSP), the process by which thermodynamically competitive crystal structures are calculated from a chemical diagram of the molecule [4,9], has the potential to change the game for the pharmaceutical industry.

Solid form screening in the pharmaceutical industry

Few active pharmaceutical ingredients (APIs) have only one crystalline phase with up to three in four compounds which are intensively screened displaying polymorphism [3]. Most are capable of forming multi-component forms such as salts, co-crystals and solvates, greatly expanding the potential solid form options for drug products [3,7,10]. The aim of experimental 'polymorph' screening is to produce the solid form landscape [11] relating all known solid forms of a parent compound, salt or co-crystal, and providing routes for their crystallization and interconversion as exemplified by the summaries in Figure 1. Although the implications of not finding the stable (generally preferred) form prior to product launch are well documented for ritonavir [12] and rotigotine [13], the conditions needed to nucleate any crystal form for the first time are not known *a priori*. Pharma has therefore largely embraced the trial and error approach to finding forms most notably with the introduction of high throughput screening in the early 2000's. The idea was to expose compounds to hundreds, even thousands, of conditions in the hope of yielding new forms[14]. Initial investment in automated screening emphasized variations in solvent, supersaturation, cooling rates and other factors that would usually be controlled in a crystallization process. However, despite the huge research effort into developing experimental screening protocols, there are no "cookbook recipes" to find all relevant solid forms. The complexities of experimental screens are very dependent on the specific

molecule [15] and such automated systems working on small amounts of material have to be supplemented by bespoke manual screening [3,16].

It has been suggested by McCrone [17] that the number of solid forms of a compound is proportional to the time and money spent investigating its crystallization. Industry whilst performing massive experimentation, with growth of contract research organizations to do screening work, needs to ~~conclude on shorter timescales and~~ direct experimentation efficiently. The plethora of possible experiments, which today extends beyond conventional methods to include heteronuclear screening using a bank of polymers [18], as well as crystallization in electric [19] and ultrasound [20] fields, in laser beams [21,22] and under confinement [23], makes it difficult to invoke a Quality by Design argument on a purely empirical approach. This has led to sustained interest in whether computational modeling can be used to reliably predict polymorphs, their properties and the experiments needed to produce the first sample [24].

The development of computational Crystal Structure Prediction (CSP)

The main application of the ability to predict crystal structures from a chemical diagram, i.e. prior to synthesis, has historically been seen as the design of new materials with desired physical properties, such as energetic materials or organic semi-conductors. A computer code that could predict the crystal structures of an organic molecule has to encapsulate and quantify the fundamental theory of what determines the crystal structures. The theoretical assumption in most programs is that the crystal structure is the most thermodynamically stable possibility. This, however, is too simplistic a theory to also predict metastable forms, i.e. polymorphism [25]. Polymorph prediction would require (Box 1) the generation of a set of putative crystal structures for the molecule, and scoring them for their probability of being observed. To date, successful CSP methods have ranked the crystal structures on their

relative thermodynamic stability as approximated by the relative lattice energy effectively corresponding to 0K.

The ability to predict structures without experimental information has been periodically tested since 1999 [26-28] in Blind Tests sponsored by the Cambridge Crystallographic Data Centre, where those involved in developing CSP methods are invited to submit their predictions of unpublished structures. The papers described in the ensuing discussion show the development of the methods. Despite the limitations on the crystal structures used in the Blind Tests, they do reveal the progress and problems in the ability to predict organic crystal structures. One of the targets for the recent 6th Blind Test (Figure 2), molecule XXIII, has been shown to have five polymorphs and XXVI has also undergone commercial solid form screening. This industrial involvement is particularly valuable in reducing discussion as to whether some predicted structures are undiscovered polymorphs [29]. All of the experimentally known structures of the target systems were in fact generated by submissions in the 6th Blind Test, and many were ranked as being one of the most stable structures by one or more methods. However, no one method had all of the experimental structures as being more stable than unobserved alternatives. This is not surprising, as there are many challenges to computational chemistry involved in a CSP run [4,9] (see Box 1). The search space to be covered is immense, particularly for flexible molecules, where considering millions of putative structures may not be sufficient. For the known polymorphs of XXIII with two different conformations in the unit cell, the search space is so vast that most groups did not attempt to generate the Z'=2 structures. The Blind Test targets were restricted to those that could be reasonably approximated by the infinite perfect crystal used in the modeling, though disordered crystals are inevitable for some molecules [30].

The biggest challenge in CSP is to be able to rank the thermodynamic stability of the possible crystal structures accurately enough, given that the energy difference between known polymorphs is typically around a kJ/mol, though it can be up to 8 kJ/mol for conformational polymorphs [8]. The tiny energy differences between polymorphs can sometimes be measured by the heat of transformation or by the difference in their heats of fusion. The energy evaluation challenge is closely related to the problem of sufficient accuracy in force-fields or electronic structure calculations for predicting protein structures [31] or ligand binding energies in drug design [31,32]. Since the lattice energy of benzene has only recently been calculated to an absolute accuracy of less than 1 kJ/mol [33], CSP relies on effective cancellation of errors in the evaluation of the crystal energies. The more successful methods in the Blind Test either used electronic structure calculations on the isolated molecule conformations or on many hundreds of crystal structures. The Blind Tests have shown that worthwhile crystal energy evaluations have to be based on the electronic structure of the molecule and so scale very badly with the size of the molecule, the range of conformations that could appear in the solid state, and the number of independent molecules in the crystal structure (Box 1). The computational expense of the successful Blind Test submissions for the larger systems was measured in hundreds of thousands of CPU hours, corresponding to weeks to months of dedicated use of high performance computing clusters [6].

To date, crystal structure prediction studies have been performed on many model drug compounds, and in cases such as aspirin, paracetamol, carbamazepine and 5-fluorouracil, have anticipated the discovery of new polymorphs [34]. After the 2nd Blind Test, two additional polymorphs of the sulfonamide (VI) were found, and the more accurate ranking by DFT-D showed that the known forms were the three lowest energy structures found [35]

in a search by the GRACE code [36]. Further examples of combined CSP and experimental studies leading to the few lowest energy structures being observed polymorphs include creatine [37] and 4-aminoquinaldine [38] and its most stable form, a polymorph of 4-aminoquinaldine monohydrate, which proved very difficult to access experimentally for kinetic reasons [39]. This leads to the current research debate as to whether and why some structures may never be found, despite their apparent computed stability, or alternatively, whether an experiment can be devised to cause the first nucleation of any thermodynamically plausible structure [5,40].

Academic work contrasting the known polymorphism of small organic molecules with the output of a CSP study has led to the concept of a computed crystal energy landscape [4]. This is the set of computer-generated crystal structures which are thermodynamically plausible as polymorphs. However, in most cases, there are considerably more structures on the crystal energy landscape than known polymorphs [5]. Some of these structures are artifacts of the approximations currently used in calculating the crystal energy landscape, particularly the neglect of temperature and hence the molecular motion within the crystals, which means that not all structures are free energy minima. Would an accurate crystal energy landscape at ambient temperature and pressure define the possible polymorphs relevant to industry [4,5]? The experience from small organic molecules shows that the computed crystal energy landscape has the potential for:

- Confirming that the most stable crystal structure is known [37]
- Predicting the structures of potential polymorphs, and from these structures suggesting possible crystallization strategies to find them [39]

- Helping structurally characterize polymorphs from powder diffraction data[41], solid state NMR [42,43], or electron diffraction [44], in cases where a single crystal suitable for structure determination by crystallography cannot be obtained
- Anticipating disorder in crystal structures [45], and hence allowing more confident separation of varying disorder from mixtures of polymorphs
- Understanding or rationalizing crystallization behaviors [4,5], e.g. monomorphism v. polymorphism, solvate formation, etc.

How computed crystal energy landscapes help define pharmaceutical solid form landscapes.

Industrial pharmaceutical scientists have closely monitored the development of CSP methods and in particular, the CCDC Blind Tests. Reliable CSP would determine whether their screening had missed the most stable solid form and further screening was warranted. Alternatively, if the practically important forms had been found, resources would be saved by not unnecessarily prolonging the search for forms. When XX in the 2010 Blind Test (Figure 2) inspired two groups to algorithmic developments that successfully predicted its structure [46], it showed that CSP methods were starting to be able to tackle molecules approaching the size and complexity of some molecules in development. We review the subsequent published examples of CSP that have been done in conjunction with industrial polymorph screening to show that the value of adding this computational technique extends well beyond right-sizing solid form screens.

As this review will show, the uses of crystal energy landscapes established from the experience with small organic molecules can equally well be applied to pharmaceutical molecules. The evaluation of CSP is particularly useful on pharmaceuticals because of the extent of polymorph screening that has usually been done. It also presents a ~~great~~n

excellent opportunity to look at crystallization issues of pharmaceuticals, how they play out in industry and what CSP can contribute. Given that there is no statistical evidence that increasing molecular size and flexibility significantly changes the tendency to polymorphism [3], what has the combination of experimental and computational (in silica and in silico) solid form screening of pharmaceuticals taught us beyond what has already been learned from small molecule model systems? Why should the pharmaceutical industry care? The following examples on some molecules towards the smaller range of those being developed in the pharmaceutical industry (specified in Figure 3) illustrate some of the diversity of solid state behaviors that challenge solid form development and highlight the potential of computed crystal energy landscapes to augment solid form screening and selection.

Why do molecules from the same drug discovery program have very different crystallization behavior?

It may seem counterintuitive that molecules similar enough to act at the same protein receptor can have dramatically different crystallization behavior. However, crystal structures are very sensitive to all types of intermolecular contacts surrounding the molecule and not just the functional groups that bind to the receptor [47]. Hence, molecules from the same drug discovery program can pose very different challenges when it comes to progressing solid forms in drug development. A CSP study highlighting the different challenges used two 5-HT_{2a} agonists, LY2806920 (B5) and LY2624803 (DB7), both of which were under development for sleep disorders. B5 readily and reliably crystallizes into just one solid form containing the neutral molecule, whereas three neat polymorphs, two hydrates, three alcohol solvates and an amorphous phase are known for DB7 (Figure 1) [48]. The screening effort was roughly equivalent for B5 and DB7 with respect to the number of experiments, but since B5 crystallized so readily, it was very difficult to avoid ~~memory of the~~

~~input material~~minute molecular clusters (nuclei) of the starting form being carried over into the~~in~~ crystallization experiments, whereas amorphous DB7 could be used. The reason for the difference in number of solid forms of B5 and DB7 was not a difference in the energy spectrum of the crystal energy landscape (Figure 1), but rather in the density and nature of the low energy structures. B5 could pack densely with itself, with most low energy structures having an internal hydrogen bond. In contrast, DB7 had no good way of packing densely with itself, with the packing coefficients of predicted low energy structures lying in the lower half of the commonly observed range (65-75%). The low energy structures for DB7 mainly had intermolecular hydrogen bonds forming a range of motifs, showing that crystal packing can have a major effect on the potential hydrogen bonding. These examples follow the general rule observed for small molecules, that both the crystal structure and energy landscapes are very specific to the individual molecule, and even the smallest changes in the API will usually result in different crystal structures and probably crystallization behavior.

The nature of the solid form landscape

Solid form diversity, disorder and why mixtures of polymorphs sometimes cannot be avoided

Despite the increasingly comprehensive exploration of crystallization conditions and development of advanced analytical tools to detect and characterize solid-state forms, two challenges in generating solid form landscapes remain: finding experimental conditions which will induce a crystal form to nucleate for the first time and identifying bonafide crystal forms amidst materials that are frequently poorly crystalline, disordered or mixtures of phases. In the CSP study of 5-HT_{2a} agonist DB7, a metastable polymorph, Form III, could only be obtained in polycrystalline form by desolvating the zwitterionic dehydrate [49] (Figure 1). The question was raised as to whether sample to sample variability in the properties of Form III was due to concomitant crystallization of two closely related polymorphs. The

computed crystal energy landscape found a match to its powder diffraction pattern in two structures, differing only in the propionic acid conformation, allowing characterization of this solid form as a single polymorph with variable sidechain disorder) [48]. In this instance, the combined use of experimental screening and CSP showed a disordered structure for Form III was inevitable, helping to clarify the number of forms produced by the solid form screening) [48].

In contrast, olanzapine, marketed for the treatment of schizophrenia, generated a lot of work contesting patent claims of “novel” polymorphs by generic companies prior to its coming off patent in the USA in 2011. In many cases, the purported forms were known mixtures of the concomitantly crystallizing metastable polymorphs, Forms II and III. The structure of Form II was only determined in 2011, when a single crystal suitable for X-ray could be picked out from a sample of olanzapine that had failed to co-crystallize with nicotinamide [50]. A single crystal of Form III was not identified for X-ray structure analysis; however, the crystal energy landscape included a structure that was a sufficient match to the Form III powder pattern to show that it was a different stacking of the same molecular layers as Form II [51]. This structural model rationalizes why Forms II and III crystallize concomitantly, with it being practically impossible to generate phase pure samples.

The problems of differentiation between different polymorphs and degrees of disorder, and the consequences for the quality control of crystal properties, is further exemplified by the case of tazofelone. The original screening of racemic tazofelone had produced two polymorphs [52], which were based on the same layer structure with different stackings in the third dimension. Revisiting this compound to obtain good thermodynamic data for calibrating the CSP study [53] unexpectedly produced an alternative stacking as a third polymorph. Of particular concern for ensuring quality control over material properties was

that the large single crystals of each polymorph varied in melting point. This triggered an unusually detailed examination of the raw diffraction data to reveal evidence of significant disorder. The crystal energy landscape had the most stable form as the global minimum, but showed that there were other ways of stacking the layers that were so close in energy, that stacking errors or different polymorphic domains (polytypes) even within single crystals were probably unavoidable.

Determination of crystal structure when single crystals cannot be grown.

The examples surveyed thus far show how crystal structures are sometimes needed in order to clarify experimental solid form landscapes. Structural information is also used to rationalize and understand crystallization behaviors, and as interest in emerging in silico approaches to drug product design continues to grow, structures will undoubtedly be necessary inputs to the modeling and prediction of solid-state properties well into the future. A further example of using CSP, this time in conjunction with a range of experimental techniques to determine the crystal structure, is illustrated for AZD8329, an 11β-HSD1 inhibitor investigated for use in the treatment of type 2 diabetes. AZD8329 ~~showed significant polymorphism~~crystallizes in at least four ~~neat~~ polymorphs, with Form 4, one of two forms considered to have superior properties for development, not having a crystal structure determination. The structure was proposed [54] by comparing the experimental proton solid state NMR spectrum with those calculated from CSP generated structures (Box 1) with both the *cis* and *trans* amide conformations. The proposed structure of Form 4 with a *cis* amide was in excellent agreement with that independently determined from powder X-ray diffraction, with the advantage of NMR spectroscopy having located the carboxylic acid proton position.

A related study attempted to use CSP to find structures for the anhydrous forms of the antibiotic levofloxacin, using the commercial program Polymorph Predictor [55]. In the end, a plausible model was proposed after using the crystal structures of six carboxylic acid salt and hydrate forms to choose six likely π…π stacked dimer structures, which were optimized by electronic structure methods and held rigid during the CSP search. This illustrates the use of a limited approximate form of CSP to extrapolate from a marked structural preference ("structural synthon") seen in other solid state forms of the API.

Why are some molecules prolific solvate formers?

Many pharmaceuticals are prolific solvate formers, with sulphathiazole having over 100 solvates reported [56]. This considerably complicates the solid form screening output. Pharmaceutical solvates with solvents that would never be allowed anywhere in a production process because of their toxicity, have to be considered in screening because desolvation [57] is a sufficiently productive method of finding new forms and may be the only route to a new polymorph. Conversely, all possible hydrates have to be extensively studied and characterized because of the impossibility of rigorously excluding water from the production processes. Solvates can include multiple solvents, sometimes in variable ratios, and the distinction between surface absorbed water, stoichiometric and non-stoichiometric hydrates is both critical for process design and difficult to establish [49]. Labile solvates, where the solvent readily leaves the crystal when it is removed from the crystallizing solution are common. CSP generated structures with void spaces can be stabilized by the inclusion of solvent molecules [58]. The nature of this stabilization, the ease of removing the solvent, and ability of the molecule to rearrange into a dense unsolvated form is extremely dependent on the specific structure, but could suggest a route to a desirable high solubility, low density, and kinetically stable form. The extension of the

principles of CSP on inclusion compound frameworks and porous molecular cages [59] to pharmaceutical solids and solvates is yet in its infancy, as a quicker, more general route to predicting solvate formation [60,61]. For example, we need to progress beyond just attributing the formation of solvates and hydrates of DB7 and none for B5 to the inability of only DB7 to pack densely by itself (Figure 1).

Olanzapine also illustrates how the inability of a molecule to pack well with itself can give rise to a multitude of solid forms, with over 60 being found in the screen [51]. Many of these had differing solvent mixtures between layers of olanzapine dimers, and the crystal energy landscape showed that these layers do not stack particularly well to form an unsolvated crystal. The separation of solvate motifs and polymorphs is more challenging for the Pfizer oncolytic axitinib (AG013736), which has 71 forms, including 5 neat polymorphs [57]. A very limited CSP study, using the commercial force-field based method Polymorph Predictor [62] concluded that only a carefully modified force-field was able to find the polymorphs within the top 500 structures generated with the appropriate conformation. A fuller search [63] using ab initio methods to find molecular conformations and energies and CrystalPredictor did find all the polymorphs (Box 1), but also showed that there are many alternative structures that were thermodynamically competitive. Considerable efforts had already gone into developing targeted screens [57] to circumvent the solvation issues associated with conventional screening methods for axitinib, so that the expense of further computational work would only be justified if a clear pathway to crystallizing further polymorphs could be proposed.

Why we sometimes miss finding more stable forms

Detecting potential problems with chiral resolution by crystallization

Crystallization is often seen as an ideal process for the separation of enantiomers required to satisfy regulatory demands that chiral drugs are administered in an optically pure form [64]. As such, screening and development are generally directed to just the active enantiomer. The dangers of this strategy are shown by LY156735, a melatonin agonist, in which two polymorphs are known, but the most stable form has only been obtained for the inactive S enantiomer. In fact, had crystallization studies not been performed on the inactive isomer, there would be no reason to believe that a more stable crystal form existed [65]. CSP would, however, have alerted to this possibility, with the crystal energy landscape having both enantiomorphs and the known racemic form within the top 9 structures, all within 1 kcal/mol of the unobserved most stable form (Box 1) [65].

Chiral resolution by crystallization is, of course, only possible when the opposite enantiomer is rejected during crystal growth. For tazofelone, an unusual experiment of seeding a racemic melt with the enantiopure (R or S) crystal instead results in an isostructural solid solution [66], i.e. the enantiopure crystal structure can include a variable proportion of molecules of the other hand. This phenomenon can be explained by the computed crystal energy landscape [53] including isostructural racemic structures that are more stable than the enantiomorph, albeit metastable relative to the other known racemic polymorphs. Since a change in conformation of the molecule effectively has the same packing ability as a change in chirality, recrystallization of predominantly enantiopure tazofelone will absorb rather than exclude chiral impurities.

Is the lack of observed polymorphs in a screen reliable?

The failure of experimental solid form screens to produce more than one crystal form may be due to one form being much more stable or crystallizing much more rapidly than all others. In such cases, CSP can uniquely show whether monomorphism is a product of

thermodynamics or crystallization kinetics. An example where alternative crystal structures were calculated [67] to be only slightly less stable than the only readily crystallized form is GSK269984B. In this case, the hypothetical polymorphs had intermolecular hydrogen bonding compensating for adopting grossly different, higher energy conformations than the observed more stable, internally hydrogen bonded conformation [67]. Further screening, concentrating on solvents that would be likely to hydrogen bond to the API, produced some metastable solvates with the expected intermolecular hydrogen bonding, but the same gross conformation as in the neat form. Thus the question arises as to whether the fast crystallization of GSK269984B into its most stable form could be relied upon to prevent the crystallization of the alternative computer generated structures [67], given that solution NMR showed that a range of other conformations could exist in solution. In ritonavir, it was the small solution population of the higher energy conformation that was found in the most stable polymorph that rationalized its disastrous later appearance [68,69]. The key difference for GSK269984B is that the higher energy conformers are calculated to give metastable polymorphs.

Crizotinib was developed by Pfizer for the treatment of forms of lung cancer, and extensive polymorph and hydrate screening similarly found only one crystalline form. 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, rationalizing the lack of polymorphs [24]. That the known structure not only had the lowest energy conformation but also optimal intermolecular interactions was confirmed by a CCDC solid form informatics “healthcheck” [70]. It is unusual that there are no signs of alternative crystal forms in the screening and so the computational confirmation that there is no compromise between conformation and intermolecular packing in the

structure, and that it has a uniquely favourable packing defining all three dimensions, provides valuable reassurance.

Suggesting experiments to find new polymorphs

A polymorph that had been missed in extensive experimental screening of Roche's CETP inhibitor, Dalcetrapib, has recently been found by crystallisation under pressure, an experiment suggested by the CSP study [71]. The crystal energy landscape had two structures very close in energy, but denser than the known stable form; which was the most stable structure of all those generated in the search depending on the DFT-D method used. The unknown structures were calculated to become 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. This matched the predicted polymorph, except for disorder in the hydrocarbon tail, which could have been anticipated from the crystal energy landscape. The new form was metastable at ambient pressure, converting to the stable form over a few hours. This shows how CSP can suggest a route to a form missed by conventional screening on the basis of the properties of the computer generated structures. The new form and the disorder in the side chains may have implications for the mechanical behaviour of crystals of the most stable form if subjected to pressure during processing.

Discussion

What are the advantages of leveraging CSP in pharmaceutical development?

In the few years in which we have been able to calculate realistic crystal energy landscapes for pharmaceuticals it has been established that Crystal Structure Prediction studies can significantly complement industrial solid form screening. Enhancing the experimentally known crystal structures with computer generated ones helps make the links between the

structures of the neat forms, solvates, co-crystals, salts and their polymorphs. Establishing whether there are alternative, very similar structures of comparable energy further gives an entrée into understanding the extent of the API's solid state continuum [7] between perfect crystals and possible varying degrees of disorder that may be defined by the crystallization conditions. When used in conjunction with the cutting-edge experimental infrastructure that industry needs for less empirical formulation and process design [72], this can help define the diversity of the solid state, particularly when the boundaries are not clear cut, such as with non-stoichiometric solvates or where the proton position blurs the salt/co-crystal boundary [73].

The main advantage of calculating an extensive crystal energy landscape is to show what types of crystal structures are thermodynamically competitive with the crystal forms that have already been found in screening. Are there structures which are sufficiently different from the known forms that serious consideration must be given to whether the range of screening experiments so far is sufficient? Or should additional experiments be undertaken, such as crystallization under pressure to target denser polymorphs as illustrated by Dalcetrapib [71]? Combining the experimental data with the crystal energy landscape develops a molecular level understanding of the crystallization behavior of a specific API. This allows the experimentalist to focus efforts on the polymorphs that matter, significantly reducing the experimental effort.

What are the challenges in developing CSP as a complement to solid form screening?

As the results of the 6th Blind Test show, there is a long way to go until realistic crystal energy landscapes can be routinely calculated for pharmaceutical molecules. The successes for the larger flexible molecules were limited to methods where specialists used hundreds of thousands of hours of computer time. Incorporating CSP into solid form screening early

enough to assist the search for solid forms will require that the crystal energy landscape can be calculated far more readily and quickly. We cannot just rely on the increase in computer speeds and the increasing availability of powerful computer clusters in the pharmaceutical industry. Algorithmic developments are on-going particularly to deal with flexibility in the generation of putative crystal structures, but this requires validation data [74] which the pharmaceutical industry is uniquely suited to provide as shown by the value of BMS-488043 [75] for the development of CrystalPredictor [76]. The evaluation of the relative energies of the crystal structures by a hierarchy of increasingly more accurate models on a smaller number of structures has to push at the current boundaries of our ability to model lattice energies and describe the effects of temperature. Some state of the art methods were used in the 6th Blind Test, showing that we cannot yet usefully predict the energy differences between known polymorphs for pharmaceuticals. However, as shown by tazofelone, experimental investigations of the thermodynamics of polymorphs can reveal surprises. Real crystalline materials at ambient conditions can be dynamically or statically disordered and very different from the static perfect infinite crystals in the calculations. Insights into the differences can inform understanding the transformations and apparent stability or metastability of the forms. This helps tie together and rationalise the measurements of key solid state properties for development that may appear to contradict expected correlations in behaviour, as shown by the hydrates of DB7 [49].

On the other hand, as the examples of levofloxacin and crizotinib show, cheaper, restricted search space CSP studies with lower accuracy energies can be very useful for specific purposes, such as suggesting or eliminating certain types of structures. However, it is important to fit the computer modelling to the question: there have been cases where CSP

studies have wrongly concluded that further polymorphs are unlikely through insufficient coverage of the possible range of crystal structures [77].

Will polymorph prediction ever be a black-box computational tool? Fundamental understanding of crystallisation for flexible molecules

Pharmaceuticals (c.f. Figure 3) cannot be assumed to have the same crystallisation behaviour as small, rigid molecules which can readily rearrange to the most stable form. The molecule may not be able to pack densely with itself, and may not be able to change conformation or hydrogen bonding sufficiently during crystallisation to readily achieve the most stable structure. Polymorphs formed as desolvated solvates, or by chemical transformation (such as proton exchange) once the molecules are already highly aggregated may be very long-lived but far more metastable than small molecule polymorphs. Can the most stable structure on the crystal energy landscape always be crystallised? There are molecules that prove extremely difficult to crystallise at all, and so in these cases the first form is likely to be the fastest to nucleate and grow, which is not necessarily the most stable. Recent work on both a crystallographic dataset of diastereomers and chiral-racemic pairs [78], and many real and CSP generated crystal structures of mandelic acids [40] both point to statistics of nucleation and growth causing the rarity of spontaneous resolution of enantiomers by crystallization, and raise the possibility that the most stable computer generated structure may be very unlikely to ever form. Some intensively studied, highly polymorphic molecules, such as the precursor of olanzapine known as ROY [79] for the red-orange-yellow spectrum of its many polymorphs, or axitinib, have a crystal energy landscape [80], where there are a large number of many thermodynamically competitive but unobserved polymorphs. In other cases, there are a few specific structures to be targeted [65], such as the melatonin agonist, or classes to be considered, such as whether it is

possible to crystallise olanzapine without first forming the dimer found in all known forms [51], or GSK269984B in a totally different conformation [67].

The art of going from a computationally predicted polymorph to designing a method of producing it is a real challenge to our developing understanding of organic nucleation and growth[81], and an active research area. There are general approaches, such as the application of pressure for denser structures, c.f. Dalcetrapib[71], or adapting solvent range to observe undiscovered polymorphs with very different hydrogen-bonding motifs [82]. More specifically, the use of isostructural seeds to find predicted structures that were missed by conventional screening has been demonstrated by success in finally crystallising a predicted co-crystal of caffeine with benzoic acid in four geographically diverse laboratories [83]. Such specific seeding to find polymorphs has also been shown by the generation [84] of polymorph V of carbamazepine by sublimation onto an isostructural crystal of dihydrocarbamazepine form II. Form IV of carbamazepine was also not found in automated solvent screening experiments despite being exposed to a huge range of crystallisation experiments in different solvent screens and in forming its dozens of co-crystals [85]. It was found by polymer heteronucleation [86]. If we could reliably predict polymorphs in the time it takes to screen for solid forms and understood crystallisation well enough to know how to go from a predicted polymorph to the experiment to find it, this would revolutionise solid form screening.

For pharma, the question is whether all polymorphs which are practically important for solid form selection and process design are known sufficiently early in the development cycle, with further investigations for IP protection following later. In cases like crizotinib and GSK269984B, which correspond to traditional expectations of easy, reliable crystallisation into only one form, limited CSP and other healthchecks can provide the confidence to stop

screening and develop with the knowledge of the competitive structures that can be avoided (or searched more extensively for when product protection is an issue). When the CSP and screening show that there are other competitive forms, then the calculations can help structurally characterise experimental forms and inform about the range of possible undiscovered polymorphs, guiding decisions on whether and where to direct further experimental work. There is still a long way to go until we understand the kinetics of nucleation and growth sufficiently to be able to proceed from calculating a crystal energy landscape to a reliable method of polymorph prediction.

Conclusion

We have reviewed a dozen case studies that have been published in the last few years since it has become possible to perform CSP studies on small drug molecules which are being screened in industry. The studies ~~which~~ that augment solid form screening show very different behavior, from reliable crystallization into one solid form through to extensive polymorphism, and the need for CSP to characterize forms and help rationalize and cope with problematic crystallization behavior such as disorder. Each case is unique, just as every polymorph screen is unique and the optimal design of each drug product is unique. The computed crystal energy landscapes can provide the early warnings of where the issues in further development of the drug product and processing will lie. Molecular-level insights will help to us to reduce the experimental burden showing us the steps to the promised goal. Computational crystal structure prediction is *en route* towards delivering on the two most important industrial promises: lead experimentalists to new crystal forms and help decide when it is safe to stop screening. As the ability to predict key properties, such as solubility or mechanical properties from the crystal structure improves, the crystal energy landscape will

show the range of property variation possible, and perhaps lead to deliberately targeting and developing metastable or multi-component forms for their improved properties.

Acknowledgements: The Cambridge Crystallographic Data Centre are thanked for their sponsorship and organisation of the Blind Tests. The authors' collaboration was funded by Eli Lilly and Company through the Lilly Research Awards Program, and the development of the computational infrastructure through the EPSRC EP/K039229/1 and EP/F03573X/1.

References

1 Bernstein, J. (2002) *Polymorphism in Molecular Crystals*, Clarendon Press

2 Hilfiker, R. (2006) *Polymorphism in the Pharmaceutical Industry*, Wiley-VCH

3 Cruz-Cabeza, A.J. et al. (2015) Facts and fictions about polymorphism. *Chemical Society Reviews* 44, 8619-8635

4 Price, S.L. (2014) Predicting crystal structures of organic compounds. *Chemical Society Reviews* 43 (7), 2098-2111

5 Price, S.L. (2013) Why don't we find more polymorphs? *Acta Crystallographica Section B - Structural Crystallography and Crystal Chemistry* 69, 313-328

6 Reilly, A.M. et al. (2016) The sixth blind test of organic crystal structure prediction. *Acta Crystallog. B* in preparation

7 Elder, D. et al. (2015) The solid-state continuum: a perspective on the interrelationships between different solid-state forms in drug substance and drug product. *Journal of Pharmacy and Pharmacology* 67 (6), 757-772

8 Cruz-Cabeza, A.J. and Bernstein, J. (2014) Conformational Polymorphism. *Chemical Reviews* 114 (4), 2170-2191

9 Day, G.M. (2011) Current approaches to predicting molecular organic crystal structures. *Crystallography Reviews* 17 (1), 3-52

10 Chen, J. et al. (2011) Pharmaceutical Crystallization. *Crystal Growth & Design* 11 (4), 887-895

11 Thakur, T. et al. (2015) Crystal Structure and Prediction. *Annual Review of Physical Chemistry, Vol 66* 66, 21-42

12 Chemburkar, S.R. et al. (2000) Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development. *Organic Process Research & Development* 4 (5), 413-417

13 Perez-Lloret, S. et al. (2013) Rotigotine transdermal patch for the treatment of Parkinson's Disease. *Fundamental & Clinical Pharmacology* 27 (1), 81-95

14 Morissette, S.L. et al. (2004) High-throughput crystallization: polymorphs, salts, co-crystals and solvates of pharmaceutical solids. *Advanced Drug Delivery Reviews* 56, 275-300

15 Lee, A.Y. et al. (2011) Crystal Polymorphism in Chemical Process Development. *Annual Review of Chemical and Biomolecular Engineering, Vol 2* 2, 259-280

16 Newman, A. (2013) Specialized Solid Form Screening Techniques. *Organic Process Research & Development* 17 (3), 457-471

17 Haleblian, J. and McCrone, W.C. (1969) Pharmaceutical Applications of Polymorphism. *Journal of Pharmaceutical Sciences* 58 (8), 911-929

18 Pfund, L. and Matzger, A. (2014) Towards Exhaustive and Automated High-Throughput Screening for Crystalline Polymorphs. *ACS Combinatorial Science* 16 (7), 309-313

19 Taleb, M. et al. (1999) Crystallization of proteins under an external electric field. *Journal of Crystal Growth* 200 (3-4), 575-582

20 Ruecroft, G. et al. (2005) Sonocrystallization: The use of ultrasound for improved industrial crystallization. *Organic Process Research & Development* 9 (6), 923-932

21 Sun, X.Y. et al. (2006) Supersaturation and polarization dependence of polymorph control in the nonphotochemical laser-induced nucleation (NPLIN) of aqueous glycine solutions. *Crystal Growth & Design* 6 (3), 684-689

22 Ikni, A. et al. (2014) Experimental Demonstration of the Carbamazepine Crystallization from Non-photochemical Laser-Induced Nucleation in Acetonitrile and Methanol. *Crystal Growth & Design* 14 (7), 3286-3299

23 Lee, A.Y. et al. (2005) Crystallization on confined engineered surfaces: A method to control crystal size and generate different polymorphs. *Journal of the American Chemical Society* 127 (43), 14982-14983

24 Abramov, Y. (2013) Current Computational Approaches to Support Pharmaceutical Solid Form Selection. *Organic Process Research & Development* 17 (3), 472-485

25 Braun, D.E. et al. (2013) Absorbing a little water: the structural, thermodynamic and kinetic relationship between pyrogallol and its tetarto-hydrate. *Crystal Growth & Design* 13 (9), 4071-4083

26 Lommerse, J.P.M. et al. (2000) A test of crystal structure prediction of small organic molecules. *Acta Crystallographica Section B - Structural Science* 56, 697-714

27 Day, G.M. et al. (2009) Significant progress in predicting the crystal structures of small organic molecules - a report on the fourth blind test. *Acta Crystallographica Section B - Structural Science* 65 (2), 107-125

28 Bardwell, D.A. et al. (2011) Towards crystal structure prediction of complex organic compounds - a report on the fifth blind test. *Acta Crystallographica Section B - Structural Science* 67 (6), 535-551

29 Braun, D.E. et al. (2013) Complex Polymorphic System of Gallic Acid-Five Monohydrates, Three Anhydrates, and over 20 Solvates. *Crystal Growth & Design* 13 (1), 19-23

30 Habgood, M. (2011) Form II Caffeine: A Case Study for Confirming and Predicting Disorder in Organic Crystals. *Crystal Growth & Design* 11 (8), 3600-3608

31 Dunitz, J.D. and Scheraga, H.A. (2004) Exercises in prognostication: Crystal structures and protein folding. *Proceedings of the National Academy of Sciences of the United States of America* 101 (40), 14309-14311

32 Stouch, T.R. (2012) The errors of our ways: taking account of error in computer-aided drug design to build confidence intervals for our next 25 years. *Journal of Computer-Aided Molecular Design* 26 (1), 125-134

33 Yang, J. et al. (2014) Ab initio determination of the lattice energy in crystalline benzene to sub-kilojoule per mole accuracy. *Science* 345 (6197), 640-643

34 Price, S.L. (2009) Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism. *Accounts of Chemical Research* 42 (1), 117-126

35 Chan, H.C.S. et al. (2011) Molecule VI, a Benchmark Crystal-Structure-Prediction Sulfonimide: Are Its Polymorphs Predictable? *Angewandte Chemie-International Edition* 50 (13), 2979-2981

36 Kendrick, J. et al. (2011) Progress in Crystal Structure Prediction. *Chemistry - A European Journal* 17 (38), 10735-10743

37 Braun, D.E. et al. (2014) Creatine: Polymorphs Predicted and Found. *Crystal Growth & Design* 14 (10), 4895-4900

38 Braun, D.E. et al. (2015) Solid state forms of 4-aminoquinaldine - from void structures with and without solvent inclusion to close packing. *CrystEngComm* 17 (12), 2504-2516
39 Braun, D.E. et al. (2016) 4-Aminoquinaldine monohydrate polymorphism: Prediction and impurity aided discovery of a difficult to access stable form. *CrystEngComm* DOI:10.1039/C5CE01758K.
40 Hylton, R. et al. (2015) Are the Crystal Structures of Enantiopure and Racemic Mandelic Acids Determined by Kinetics or Thermodynamics? *Journal of the American Chemical Society* 137 (34), 11095-11104
41 Habermehl, S. et al. (2014) Structure determination from powder data without prior indexing, using a similarity measure based on cross-correlation functions. *Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials* 70, 347-359
42 Baias, M. et al. (2013) Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state H-1 NMR spectroscopy. *Physical Chemistry Chemical Physics* 15 (21), 8069-8080
43 Salager, E. et al. (2010) Powder Crystallography by Combined Crystal Structure Prediction and High-Resolution H-1 Solid-State NMR Spectroscopy. *Journal of the American Chemical Society* 132 (8), 2564-2566
44 Eddleston, M.D. et al. (2013) Determination of the Crystal Structure of a New Polymorph of Theophylline. *Chemistry-A European Journal* 19 (24), 7883-7888
45 Copley, R.C.B. et al. (2008) Predictable disorder versus polymorphism in the rationalization of structural diversity: A multidisciplinary study of eniluracil. *Crystal Growth & Design* 8 (9), 3474-3481
46 Kazantsev, A.V. et al. (2011) Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. *International Journal of Pharmaceutics* 418 (2), 168-178
47 Dunitz, J. and Gavezzotti, A. (2012) Supramolecular Synthons: Validation and Ranking of Intermolecular Interaction Energies. *Crystal Growth & Design* 12 (12), 5873-5877
48 Braun, D.E. et al. (2014) Contrasting Polymorphism of Related Small Molecule Drugs Correlated and Guided by the Computed Crystal Energy Landscape. *Crystal Growth & Design* 14 (4), 2056-2072
49 Braun, D.E. et al. (2015) Navigating the Waters of Unconventional Crystalline Hydrates. *Molecular Pharmaceutics* 12 (8), 3069-3088
50 Thakuria, R. and Nangia, A. (2011) Polymorphic form IV of olanzapine. *Acta Crystallographica Section C-Crystal Structure Communications* 67, O461-O463
51 Bhardwaj, R.M. et al. (2013) Exploring the Experimental and Computed Crystal Energy Landscape of Olanzapine. *Crystal Growth & Design* 13 (4), 1602-1617
52 Reutzel-Edens, S.M. et al. (2000) Molecular basis for the stability relationships between homochiral and racemic crystals of tazofelone: a spectroscopic, crystallographic, and thermodynamic investigation. *Journal of the Chemical Society-Perkin Transactions 2* 5, 913-924
53 Price, L.S. et al. (2014) A molecular picture of the problems in ensuring structural purity of tazofelone. *Journal of Molecular Structure* 1078, 26-42
54 Baias, M. et al. (2013) De Novo Determination of the Crystal Structure of a Large Drug Molecule by Crystal Structure Prediction-Based Powder NMR Crystallography. *Journal of the American Chemical Society* 135 (46), 17501-17507
55 Singh, S. and Thakur, T. (2014) New crystalline salt forms of levofloxacin: conformational analysis and attempts towards the crystal structure prediction of the anhydrous form. *Crystengcomm* 16 (20), 4215-4230
56 Bingham, A.L. et al. (2001) Over one hundred solvates of sulfathiazole. *Chemical Communications* (7), 603-604

57 Campeta, A.M. et al. (2010) Development of a targeted polymorph screening approach for a complex polymorphic and highly solvating API. *Journal of Pharmaceutical Sciences* 99 (9), 3874-3886
58 Cruz-Cabeza, A.J. et al. (2009) Predicting Inclusion Behaviour and Framework Structures in Organic Crystals. *Chemistry - A European Journal* 15 (47), 13033-13040
59 Jelfs, K.E. et al. (2013) In silico Design of Supramolecules from Their Precursors: Odd-Even Effects in Cage-Forming Reactions. *Journal of the American Chemical Society* 135 (25), 9307-9310
60 Braun, D.E. et al. (2011) Which, if any, hydrates will crystallise? Predicting hydrate formation of two dihydroxybenzoic acids. *Chemical Communications* 47 (19), 5443-5445
61 Cruz-Cabeza, A.J. et al. (2010) Predicting stoichiometry and structure of solvates. *Chemical Communications* 46 (13), 2224-2226
62 Lupyan, D. et al. (2012) Close intramolecular sulfur-oxygen contacts: modified force field parameters for improved conformation generation. *Journal of Computer-Aided Molecular Design* 26 (11), 1195-1205
63 Vasileiadis, M. et al. (2015) Prediction of the crystal structures of axitinib, a polymorphic pharmaceutical molecule. *Chemical Engineering Science* 121, 60-76
64 Lorenz, H. and Seidel-Morgenstern, A. (2014) Processes to separate enantiomers. *Angewandte Chemie (International ed.in English)* 53 (5), 1218-1250
65 Kendrick, J. et al. (2013) Crystal Structure Prediction of a Flexible Molecule of Pharmaceutical Interest with Unusual Polymorphic Behavior. *Crystal Growth & Design* 13 (2), 581-589
66 Huang, J. et al. (2006) Discovery of a solid solution of enantiomers in a racemate-forming system by seeding. *Journal of the American Chemical Society* 128 (36), 11985-11992
67 Ismail, S.Z. et al. (2013) Evaluating a Crystal Energy Landscape in the Context of Industrial Polymorph Screening. *Crystal Growth & Design* 13 (6), 2396-2406
68 Morissette, S.L. et al. (2003) Eludication of crystal form diversity of the HIV protease inhibitor ritonavir by high-throughput crystallisation. *Proceedings of the National Academy of Sciences of the United States of America* 100 (5), 2180-2184
69 Bauer, J. et al. (2001) Ritonavir: An extraordinary example of conformational polymorphism. *Pharmaceutical Research* 18 (6), 859-866
70 Feeder, N. et al. (2015) The integration of solid-form informatics into solid-form selection. *Journal of Pharmacy and Pharmacology* 67 (6), 857-868
71 Neumann, M. et al. (2015) Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nature Communications* 6, art 7793.
72 Sun, C.C. (2009) Material Science Tetrahedron-A Useful Tool for Pharmaceutical Research and Development. *Journal of Pharmaceutical Sciences* 98, 1744-1749
73 Mohamed, S. et al. (2011) Computational prediction of salt and cocrystal structures-Does a proton position matter? *International Journal of Pharmaceutics* 418 (2), 187-198
74 Pantelides, C. et al. (2014) General Computational Algorithms for Ab Initio Crystal Structure Prediction for Organic Molecules. *Topics in Current Chemistry* 345, 25-58
75 Fakes, M. et al. (2009) Enhancement of oral bioavailability of an HIV-attachment inhibitor by nanosizing and amorphous formulation approaches. *International Journal of Pharmaceutics* 370 (1-2), 167-174
76 Habgood, M. et al. (2015) Efficient Handling of Molecular Flexibility in Ab Initio Generation of Crystal Structures. *Journal of Chemical Theory and Computation* 11 (4), 1957-1969
77 King, M.D. et al. (2011) Prediction of the Unknown Crystal Structure of Creatine Using Fully Quantum Mechanical Methods. *Crystal Growth & Design* 11 (12), 5733-5740
78 Gavezzotti, A. and Rizzato, S. (2014) Are Racemic Crystals Favored over Homochiral Crystals by Higher Stability or by Kinetics? Insights from Comparative Studies of Crystalline Stereoisomers. *Journal of Organic Chemistry* 79 (11), 4809-4816

79 Yu, L. (2010) Polymorphism in Molecular Solids: An Extraordinary System of Red, Orange, and Yellow Crystals. *Accounts of Chemical Research* 43 (9), 1257-1266

80 Vasileiadis, M. et al. (2012) The polymorphs of ROY: application of a systematic crystal structure prediction technique. *Acta Crystallographica Section B-Structural Science* 68, 677-685

81 Davey, R.J. et al. (2013) Nucleation of Organic CrystalsA Molecular Perspective. *Angewandte Chemie-International Edition* 52 (8), 2166-2179

82 Cross, W.I. et al. (2003) A whole output strategy for polymorph screening: Combining crystal structure prediction, graph set analysis, and targeted crystallization experiments in the case of diflunisal. *Crystal Growth & Design* 3 (2), 151-158

83 Bucar, D.K. et al. (2013) The curious case of (caffeine).(benzoic acid): how heteronuclear seeding allowed the formation of an elusive cocrystal. *Chemical Science* 4 (12), 4417-4425

84 Arlin, J.B. et al. (2011) A strategy for producing predicted polymorphs: catemeric carbamazepine form V. *Chemical Communications* 47 (25), 7074-7076

85 Childs, S.L. et al. (2009) Analysis of 50 Crystal Structures Containing Carbamazepine Using the Materials Module of Mercury CSD. *Crystal Growth & Design* 9 (4), 1869-1888

86 Price, C.P. et al. (2005) Crystalline polymorph selection and discovery with polymer heteronuclei. *Journal of the American Chemical Society* 127 (15), 5512-5517

87 Steed, K. and Steed, J. (2015) Packing Problems: High Z' Crystal Structures and Their Relationship to Cocrystals, Inclusion Compounds, and Polymorphism. *Chemical Reviews* 115 (8), 2895-2933

88 Threlfall, T. (2003) Structural and thermodynamic explanations of Ostwald's rule. *Organic Process Research & Development* 7 (6), 1017-1027

89 Bernstein, J. et al. (1995) Patterns in Hydrogen Bonding: Functionality and graph set analysis in crystals. *Angewandte Chemie-International Edition in English* 34 (15), 1555-1573

Additional references, not in text, but in Box 1 [87] and [88]caption to Fig 1 [89]

Figure Captions

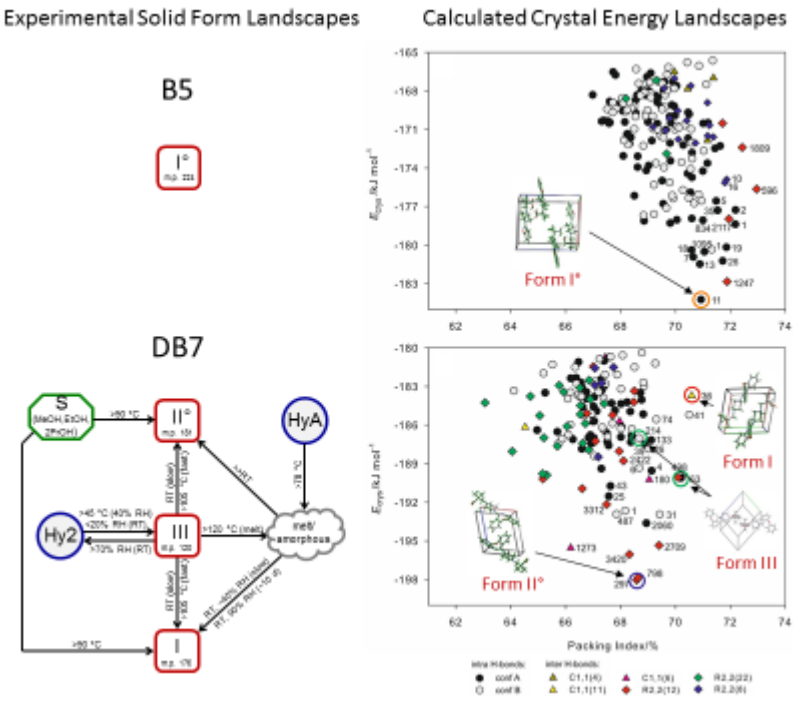
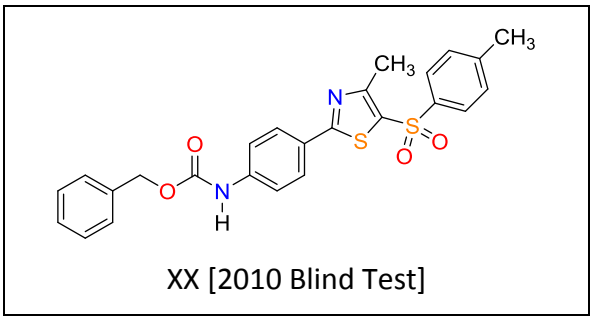
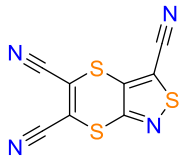


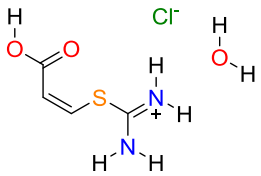
Figure 1 Contrasting summaries of experimental solid form screening and CSP results for LY2806920 “B5” and LY2624803 “DB7” [48], whose molecular structures are given in Figure 3. Left, the solid forms of the molecules and their interrelationships; right, the energies and packing efficiencies of the computer generated structures, with every symbol denoting a computer generated structure that is mechanically stable, according to the hydrogen bonding motif in graph-set notation [89].



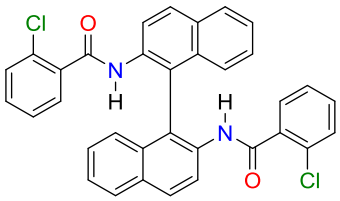
XX [2010 Blind Test]



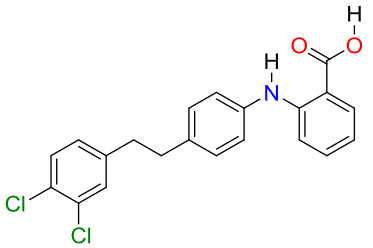
XXII



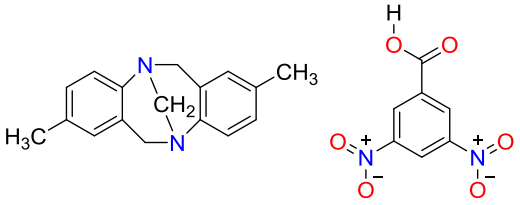
XXIV salt hydrate



XXVI largest so far in Blind Tests



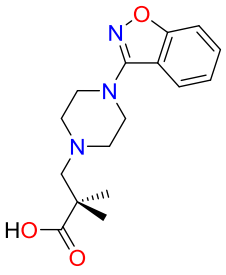
XXIII, 5 polymorphs, C and E are Z’=2



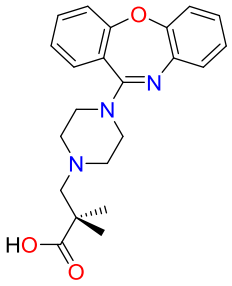
XXV co-crystal

Figure 2 The molecules whose crystal structures have been used in the Cambridge Crystallographic Data Centre’s Blind Tests of crystal structure prediction (top, molecule XX from 5th test in 2010 [28]; rest, molecules from the 6th test in 2015). Participants were

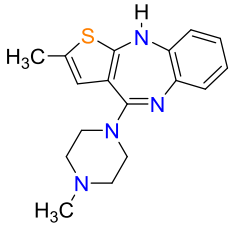
invited to submit lists of 100 structures and success was judged as having the experimental structures within these lists. Full analysis of the ranking of the structures within each submission will be published shortly [6].



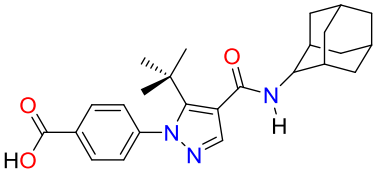
LY2806920 “B5”
1 polymorph [48]



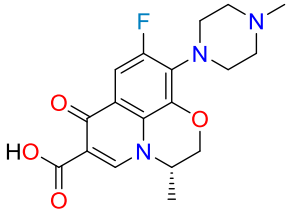
LY2624803 “DB7”
3 polymorphs, 5 solvates[48]



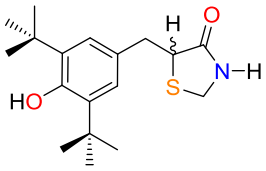
Olanzapine
3 polymorphs, 56 solvates [51]



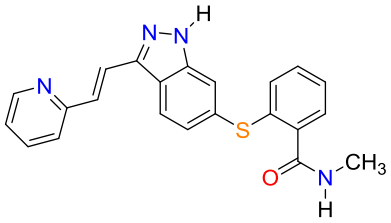
AZD8329
4 polymorphs, 3 solvates [54]



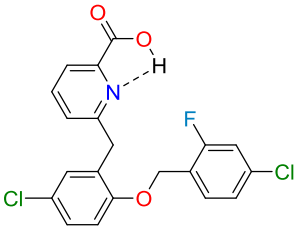
Levofloxacin
1 polymorph, 2 hydrates [55]



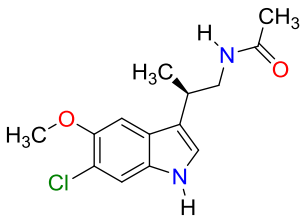
Tazofelone
3 racemic polymorphs, 1 enantiomorph, solid solution [53]



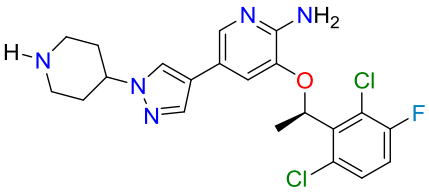
Axitinib
5 polymorphs, 66 solvates [57]



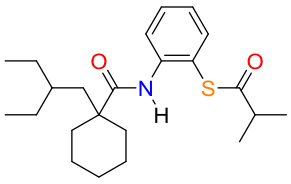
GSK269984B
1 polymorph, 4-6 solvates [67]



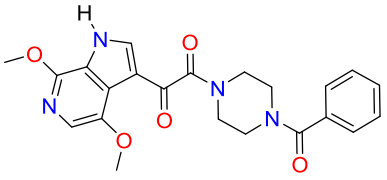
LY156735
3 polymorphs [65]



Crizotinib
1 polymorph[24]



Dalcetrapib
3 polymorphs [71]



BMS-488043
2 polymorphs [76]

Figure 3 Pharmaceuticals with combined experimental and computational solid form screening.