

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

3
4 Sarah L. Price, Department of Chemistry, University College London, 20 Gordon Street,
5 London WC1H 0AJ, UK

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

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

9

10 Keywords: Solid forms; Crystal Structures; Drug development; Computational Prediction;

11 Polymorphism; Crystal Structure Prediction

12

13 **Teaser** Could computational crystal structure prediction accelerate solid form development?

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

1 **Bios:**

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

7

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



16

17 Sarah Price



18 Susan Reutzel-Edens

18

1 **Abstract**

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

12 **Key Resources**

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

14 [2] Hilfiker, R. (2006) *Polymorphism in the Pharmaceutical Industry*, Wiley-VCH

15 Recent Reviews: [3] Facts and Fictions about Polymorphism

16 [4] Predicting crystal structures of organic compounds (a tutorial review)

17 [5] Why don't we find more polymorphs?

18 [6] Report on the 6th Blind test of crystal structure prediction

19

20 **Glossary**

21 **API:** Active pharmaceutical ingredient

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

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

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

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

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

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

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

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

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

8

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

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

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

5 *Solid form screening in the pharmaceutical industry*

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

1 molecule [15] and such automated systems working on small amounts of material have to
2 be supplemented by bespoke manual screening [3,16].
3 It has been suggested by McCrone [17] that the number of solid forms of a compound is
4 proportional to the time and money spent investigating its crystallization. Industry whilst
5 performing massive experimentation, with growth of contract research organizations to do
6 screening work, needs to ~~conclude on shorter timescales and~~ direct experimentation
7 efficiently. The plethora of possible experiments, which today extends beyond conventional
8 methods to include heteronuclear screening using a bank of polymers [18], as well as
9 crystallization in electric [19] and ultrasound [20] fields, in laser beams [21,22] and under
10 confinement [23], makes it difficult to invoke a Quality by Design argument on a purely
11 empirical approach. This has led to sustained interest in whether computational modeling
12 can be used to reliably predict polymorphs, their properties and the experiments needed to
13 produce the first sample [24].

14 *The development of computational Crystal Structure Prediction (CSP)*

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

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

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

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

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

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

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

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

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

8 **How computed crystal energy landscapes help define pharmaceutical solid form** 9 **landscapes.**

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

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

1 [excellent](#) opportunity to look at crystallization issues of pharmaceuticals, how they play out
2 in industry and what CSP can contribute. Given that there is no statistical evidence that
3 increasing molecular size and flexibility significantly changes the tendency to polymorphism
4 [3], what has the combination of experimental and computational (in silica and in silico)
5 solid form screening of pharmaceuticals taught us beyond what has already been learned
6 from small molecule model systems? Why should the pharmaceutical industry care? The
7 following examples on some molecules towards the smaller range of those being developed
8 in the pharmaceutical industry (specified in Figure 3) illustrate some of the diversity of solid
9 state behaviors that challenge solid form development and highlight the potential of
10 computed crystal energy landscapes to augment solid form screening and selection.

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

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

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

14 ***The nature of the solid form landscape***

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

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

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

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

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

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

7 *Determination of crystal structure when single crystals cannot be grown.*

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

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

8 ***Why are some molecules prolific solvate formers?***

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

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

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

22 ***Why we sometimes miss finding more stable forms***

23 *Detecting potential problems with chiral resolution by crystallization*

1 Crystallization is often seen as an ideal process for the separation of enantiomers required
2 to satisfy regulatory demands that chiral drugs are administered in an optically pure form
3 [64]. As such, screening and development are generally directed to just the active
4 enantiomer. The dangers of this strategy are shown by LY156735, a melatonin agonist, in
5 which two polymorphs are known, but the most stable form has only been obtained for the
6 inactive S enantiomer. In fact, had crystallization studies not been performed on the inactive
7 isomer, there would be no reason to believe that a more stable crystal form existed [65].
8 CSP would, however, have alerted to this possibility, with the crystal energy landscape
9 having both enantiomorphs and the known racemic form within the top 9 structures, all
10 within 1 kcal/mol of the unobserved most stable form (Box 1) [65].

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

21 *Is the lack of observed polymorphs in a screen reliable?*

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

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

16 Crizotinib was developed by Pfizer for the treatment of forms of lung cancer, and extensive
17 polymorph and hydrate screening similarly found only one crystalline form. A simple CSP
18 search, based on just four rigid, carefully selected conformers and the five most common
19 chiral space groups, showed that the known structure was significantly more stable than any
20 other generated, rationalizing the lack of polymorphs [24]. That the known structure not
21 only had the lowest energy conformation but also optimal intermolecular interactions was
22 confirmed by a CCDC solid form informatics “healthcheck” [70]. It is unusual that there are
23 no signs of alternative crystal forms in the screening and so the computational confirmation
24 that there is no compromise between conformation and intermolecular packing in the

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

3 *Suggesting experiments to find new polymorphs*

4 A polymorph that had been missed in extensive experimental screening of Roche's CETP
5 inhibitor, Dalcetrapib, has recently been found by crystallisation under pressure, an
6 experiment suggested by the CSP study [71]. The crystal energy landscape had two
7 structures very close in energy, but denser than the known stable form; which was the most
8 stable structure of all those generated in the search depending on the DFT-D method used.
9 The unknown structures were calculated to become more stable than the observed
10 polymorph with a modest increase in pressure. Experiments recrystallizing Dalcetrapib,
11 either from solution or the melt in a diamond anvil cell under modest pressure formed a
12 new polymorph. This matched the predicted polymorph, except for disorder in the
13 hydrocarbon tail, which could have been anticipated from the crystal energy landscape. The
14 new form was metastable at ambient pressure, converting to the stable form over a few
15 hours. This shows how CSP can suggest a route to a form missed by conventional screening
16 on the basis of the properties of the computer generated structures. The new form and the
17 disorder in the side chains may have implications for the mechanical behaviour of crystals of
18 the most stable form if subjected to pressure during processing.

19 **Discussion**

20 *What are the advantages of leveraging CSP in pharmaceutical development?*

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

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

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

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

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

1 enough to assist the search for solid forms will require that the crystal energy landscape can
2 be calculated far more readily and quickly. We cannot just rely on the increase in computer
3 speeds and the increasing availability of powerful computer clusters in the pharmaceutical
4 industry. Algorithmic developments are on-going particularly to deal with flexibility in the
5 generation of putative crystal structures, but this requires validation data [74] which the
6 pharmaceutical industry is uniquely suited to provide as shown by the value of BMS-488043
7 [75] for the development of CrystalPredictor [76]. The evaluation of the relative energies of
8 the crystal structures by a hierarchy of increasingly more accurate models on a smaller
9 number of structures has to push at the current boundaries of our ability to model lattice
10 energies and describe the effects of temperature. Some state of the art methods were used
11 in the 6th Blind Test, showing that we cannot yet usefully predict the energy differences
12 between known polymorphs for pharmaceuticals. However, as shown by tazofelone,
13 experimental investigations of the thermodynamics of polymorphs can reveal surprises. Real
14 crystalline materials at ambient conditions can be dynamically or statically disordered and
15 very different from the static perfect infinite crystals in the calculations. Insights into the
16 differences can inform understanding the transformations and apparent stability or
17 metastability of the forms. This helps tie together and rationalise the measurements of key
18 solid state properties for development that may appear to contradict expected correlations
19 in behaviour, as shown by the hydrates of DB7 [49].

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

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

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

5 Pharmaceuticals (c.f. Figure 3) cannot be assumed to have the same crystallisation
6 behaviour as small, rigid molecules which can readily rearrange to the most stable form. The
7 molecule may not be able to pack densely with itself, and may not be able to change
8 conformation or hydrogen bonding sufficiently during crystallisation to readily achieve the
9 most stable structure. Polymorphs formed as desolvated solvates, or by chemical
10 transformation (such as proton exchange) once the molecules are already highly aggregated
11 may be very long-lived but far more metastable than small molecule polymorphs. Can the
12 most stable structure on the crystal energy landscape always be crystallised? There are
13 molecules that prove extremely difficult to crystallise at all, and so in these cases the first
14 form is likely to be the fastest to nucleate and grow, which is not necessarily the most
15 stable. Recent work on both a crystallographic dataset of diastereomers and chiral-racemic
16 pairs [78], and many real and CSP generated crystal structures of mandelic acids [40] both
17 point to statistics of nucleation and growth causing the rarity of spontaneous resolution of
18 enantiomers by crystallization, and raise the possibility that the most stable computer
19 generated structure may be very unlikely to ever form. Some intensively studied, highly
20 polymorphic molecules, such as the precursor of olanzapine known as ROY [79] for the red-
21 orange-yellow spectrum of its many polymorphs, or axitinib, have a crystal energy landscape
22 [80], where there are a large number of many thermodynamically competitive but
23 unobserved polymorphs. In other cases, there are a few specific structures to be targeted
24 [65], such as the melatonin agonist, or classes to be considered, such as whether it is

1 possible to crystallise olanzapine without first forming the dimer found in all known forms
2 [51], or GSK269984B in a totally different conformation [67].

3 The art of going from a computationally predicted polymorph to designing a method of
4 producing it is a real challenge to our developing understanding of organic nucleation and
5 growth[81], and an active research area. There are general approaches, such as the
6 application of pressure for denser structures, c.f. Dalcetrapib[71], or adapting solvent range
7 to observe undiscovered polymorphs with very different hydrogen-bonding motifs [82].

8 More specifically, the use of isostructural seeds to find predicted structures that were
9 missed by conventional screening has been demonstrated by success in finally crystallising a
10 predicted co-crystal of caffeine with benzoic acid in four geographically diverse laboratories
11 [83]. Such specific seeding to find polymorphs has also been shown by the generation [84]
12 of polymorph V of carbamazepine by sublimation onto an isostructural crystal of
13 dihydrocarbamazepine form II. Form IV of carbamazepine was also not found in automated
14 solvent screening experiments despite being exposed to a huge range of crystallisation
15 experiments in different solvent screens and in forming its dozens of co-crystals [85]. It was
16 found by polymer heteronucleation [86]. If we could reliably predict polymorphs in the time
17 it takes to screen for solid forms and understood crystallisation well enough to know how to
18 go from a predicted polymorph to the experiment to find it, this would revolutionise solid
19 form screening.

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

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

9 **Conclusion**

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

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

3

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

8

9 **References**

- 10 **1** Bernstein, J. (2002) *Polymorphism in Molecular Crystals*, Clarendon Press
11 **2** Hilfiker, R. (2006) *Polymorphism in the Pharmaceutical Industry*, Wiley-VCH
12 **3** Cruz-Cabeza, A.J. et al. (2015) Facts and fictions about polymorphism. *Chemical Society*
13 *Reviews* 44, 8619-8635
14 **4** Price, S.L. (2014) Predicting crystal structures of organic compounds. *Chemical Society*
15 *Reviews* 43 (7), 2098-2111
16 **5** Price, S.L. (2013) Why don't we find more polymorphs? *Acta Crystallographica Section B -*
17 *Structural Crystallography and Crystal Chemistry* 69, 313-328
18 **6** Reilly, A.M. et al. (2016) The sixth blind test of organic crystal structure prediction. *Acta*
19 *Crystallog. B* in preparation
20 **7** Elder, D. et al. (2015) The solid-state continuum: a perspective on the interrelationships
21 between different solid-state forms in drug substance and drug product. *Journal of*
22 *Pharmacy and Pharmacology* 67 (6), 757-772
23 **8** Cruz-Cabeza, A.J. and Bernstein, J. (2014) Conformational Polymorphism. *Chemical Reviews*
24 114 (4), 2170-2191
25 **9** Day, G.M. (2011) Current approaches to predicting molecular organic crystal structures.
26 *Crystallography Reviews* 17 (1), 3-52
27 **10** Chen, J. et al. (2011) Pharmaceutical Crystallization. *Crystal Growth & Design* 11 (4), 887-895
28 **11** Thakur, T. et al. (2015) Crystal Structure and Prediction. *Annual Review of Physical*
29 *Chemistry, Vol 66* 66, 21-42
30 **12** Chemburkar, S.R. et al. (2000) Dealing with the impact of ritonavir polymorphs on the late
31 stages of bulk drug process development. *Organic Process Research & Development* 4 (5),
32 413-417
33 **13** Perez-Lloret, S. et al. (2013) Rotigotine transdermal patch for the treatment of Parkinson's
34 Disease. *Fundamental & Clinical Pharmacology* 27 (1), 81-95
35 **14** Morissette, S.L. et al. (2004) High-throughput crystallization: polymorphs, salts, co-crystals
36 and solvates of pharmaceutical solids. *Advanced Drug Delivery Reviews* 56, 275-300
37 **15** Lee, A.Y. et al. (2011) Crystal Polymorphism in Chemical Process Development. *Annual*
38 *Review of Chemical and Biomolecular Engineering, Vol 2* 2, 259-280
39 **16** Newman, A. (2013) Specialized Solid Form Screening Techniques. *Organic Process Research*
40 *& Development* 17 (3), 457-471

- 1 **17** Halebian, J. and McCrone, W.C. (1969) Pharmaceutical Applications of Polymorphism.
2 *Journal of Pharmaceutical Sciences* 58 (8), 911-929
- 3 **18** Pfund, L. and Matzger, A. (2014) Towards Exhaustive and Automated High-Throughput
4 Screening for Crystalline Polymorphs. *ACS Combinatorial Science* 16 (7), 309-313
- 5 **19** Taleb, M. et al. (1999) Crystallization of proteins under an external electric field. *Journal of*
6 *Crystal Growth* 200 (3-4), 575-582
- 7 **20** Ruecroft, G. et al. (2005) Sonocrystallization: The use of ultrasound for improved industrial
8 crystallization. *Organic Process Research & Development* 9 (6), 923-932
- 9 **21** Sun, X.Y. et al. (2006) Supersaturation and polarization dependence of polymorph control in
10 the nonphotochemical laser-induced nucleation (NPLIN) of aqueous glycine solutions. *Crystal*
11 *Growth & Design* 6 (3), 684-689
- 12 **22** Ikni, A. et al. (2014) Experimental Demonstration of the Carbamazepine Crystallization from
13 Non-photochemical Laser-Induced Nucleation in Acetonitrile and Methanol. *Crystal Growth*
14 *& Design* 14 (7), 3286-3299
- 15 **23** Lee, A.Y. et al. (2005) Crystallization on confined engineered surfaces: A method to control
16 crystal size and generate different polymorphs. *Journal of the American Chemical Society*
17 127 (43), 14982-14983
- 18 **24** Abramov, Y. (2013) Current Computational Approaches to Support Pharmaceutical Solid
19 Form Selection. *Organic Process Research & Development* 17 (3), 472-485
- 20 **25** Braun, D.E. et al. (2013) Absorbing a little water: the structural, thermodynamic and kinetic
21 relationship between pyrogallol and its tetarto-hydrate. *Crystal Growth & Design* 13 (9),
22 4071-4083
- 23 **26** Lommerse, J.P.M. et al. (2000) A test of crystal structure prediction of small organic
24 molecules. *Acta Crystallographica Section B - Structural Science* 56, 697-714
- 25 **27** Day, G.M. et al. (2009) Significant progress in predicting the crystal structures of small
26 organic molecules - a report on the fourth blind test. *Acta Crystallographica Section B -*
27 *Structural Science* 65 (2), 107-125
- 28 **28** Bardwell, D.A. et al. (2011) Towards crystal structure prediction of complex organic
29 compounds - a report on the fifth blind test. *Acta Crystallographica Section B - Structural*
30 *Science* 67 (6), 535-551
- 31 **29** Braun, D.E. et al. (2013) Complex Polymorphic System of Gallic Acid-Five Monohydrates,
32 Three Anhydrates, and over 20 Solvates. *Crystal Growth & Design* 13 (1), 19-23
- 33 **30** Habgood, M. (2011) Form II Caffeine: A Case Study for Confirming and Predicting Disorder in
34 Organic Crystals. *Crystal Growth & Design* 11 (8), 3600-3608
- 35 **31** Dunitz, J.D. and Scheraga, H.A. (2004) Exercises in prognostication: Crystal structures and
36 protein folding. *Proceedings of the National Academy of Sciences of the United States of*
37 *America* 101 (40), 14309-14311
- 38 **32** Stouch, T.R. (2012) The errors of our ways: taking account of error in computer-aided drug
39 design to build confidence intervals for our next 25 years. *Journal of Computer-Aided*
40 *Molecular Design* 26 (1), 125-134
- 41 **33** Yang, J. et al. (2014) Ab initio determination of the lattice energy in crystalline benzene to
42 sub-kilojoule per mole accuracy. *Science* 345 (6197), 640-643
- 43 **34** Price, S.L. (2009) Computed crystal energy landscapes for understanding and predicting
44 organic crystal structures and polymorphism. *Accounts of Chemical Research* 42 (1), 117-126
- 45 **35** Chan, H.C.S. et al. (2011) Molecule VI, a Benchmark Crystal-Structure-Prediction
46 Sulfonimide: Are Its Polymorphs Predictable? *Angewandte Chemie-International Edition* 50
47 (13), 2979-2981
- 48 **36** Kendrick, J. et al. (2011) Progress in Crystal Structure Prediction. *Chemistry - A European*
49 *Journal* 17 (38), 10735-10743
- 50 **37** Braun, D.E. et al. (2014) Creatine: Polymorphs Predicted and Found. *Crystal Growth & Design*
51 14 (10), 4895-4900

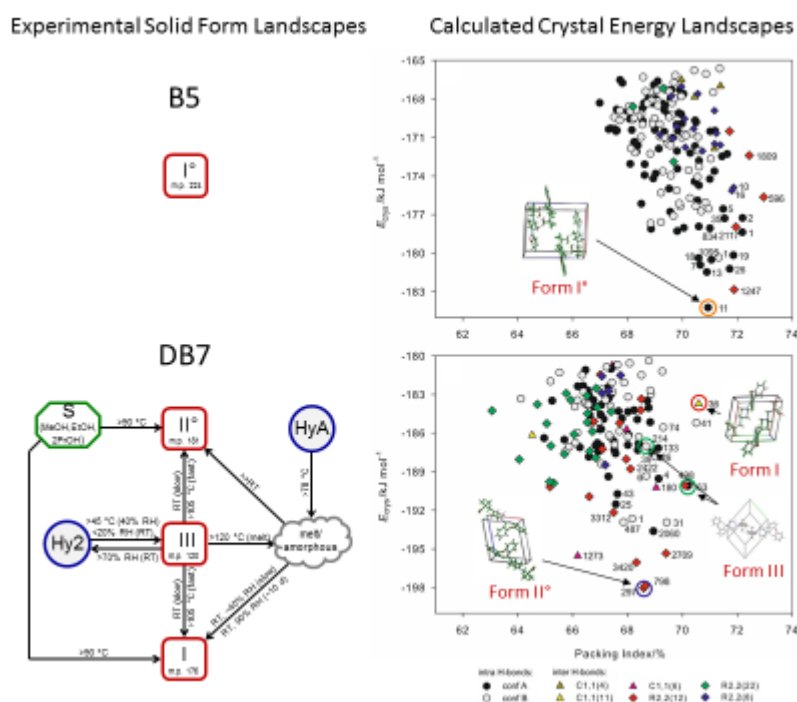
- 1 **38** Braun, D.E. et al. (2015) Solid state forms of 4-aminoquinaldine - from void structures with
2 and without solvent inclusion to close packing. *CrystEngComm* 17 (12), 2504-2516
- 3 **39** Braun, D.E. et al. (2016) 4-Aminoquinaldine monohydrate polymorphism: Prediction and
4 impurity aided discovery of a difficult to access stable form. *CrystEngComm*
5 DOI:10.1039/C5CE01758K.
- 6 **40** Hylton, R. et al. (2015) Are the Crystal Structures of Enantiopure and Racemic Mandelic Acids
7 Determined by Kinetics or Thermodynamics? *Journal of the American Chemical Society* 137
8 (34), 11095-11104
- 9 **41** Habermehl, S. et al. (2014) Structure determination from powder data without prior
10 indexing, using a similarity measure based on cross-correlation functions. *Acta*
11 *Crystallographica Section B-Structural Science Crystal Engineering and Materials* 70, 347-359
- 12 **42** Baias, M. et al. (2013) Powder crystallography of pharmaceutical materials by combined
13 crystal structure prediction and solid-state H-1 NMR spectroscopy. *Physical Chemistry*
14 *Chemical Physics* 15 (21), 8069-8080
- 15 **43** Salager, E. et al. (2010) Powder Crystallography by Combined Crystal Structure Prediction
16 and High-Resolution H-1 Solid-State NMR Spectroscopy. *Journal of the American Chemical*
17 *Society* 132 (8), 2564-2566
- 18 **44** Eddleston, M.D. et al. (2013) Determination of the Crystal Structure of a New Polymorph of
19 Theophylline. *Chemistry-A European Journal* 19 (24), 7883-7888
- 20 **45** Copley, R.C.B. et al. (2008) Predictable disorder versus polymorphism in the rationalization
21 of structural diversity: A multidisciplinary study of eniluracil. *Crystal Growth & Design* 8 (9),
22 3474-3481
- 23 **46** Kazantsev, A.V. et al. (2011) Successful prediction of a model pharmaceutical in the fifth
24 blind test of crystal structure prediction. *International Journal of Pharmaceutics* 418 (2), 168-
25 178
- 26 **47** Dunitz, J. and Gavezzotti, A. (2012) Supramolecular Synthons: Validation and Ranking of
27 Intermolecular Interaction Energies. *Crystal Growth & Design* 12 (12), 5873-5877
- 28 **48** Braun, D.E. et al. (2014) Contrasting Polymorphism of Related Small Molecule Drugs
29 Correlated and Guided by the Computed Crystal Energy Landscape. *Crystal Growth & Design*
30 14 (4), 2056-2072
- 31 **49** Braun, D.E. et al. (2015) Navigating the Waters of Unconventional Crystalline Hydrates.
32 *Molecular Pharmaceutics* 12 (8), 3069-3088
- 33 **50** Thakuria, R. and Nangia, A. (2011) Polymorphic form IV of olanzapine. *Acta Crystallographica*
34 *Section C-Crystal Structure Communications* 67, O461-O463
- 35 **51** Bhardwaj, R.M. et al. (2013) Exploring the Experimental and Computed Crystal Energy
36 Landscape of Olanzapine. *Crystal Growth & Design* 13 (4), 1602-1617
- 37 **52** Reutzel-Edens, S.M. et al. (2000) Molecular basis for the stability relationships between
38 homochiral and racemic crystals of tazofelone: a spectroscopic, crystallographic, and
39 thermodynamic investigation. *Journal of the Chemical Society-Perkin Transactions 2* 5, 913-
40 924
- 41 **53** Price, L.S. et al. (2014) A molecular picture of the problems in ensuring structural purity of
42 tazofelone. *Journal of Molecular Structure* 1078, 26-42
- 43 **54** Baias, M. et al. (2013) De Novo Determination of the Crystal Structure of a Large Drug
44 Molecule by Crystal Structure Prediction-Based Powder NMR Crystallography. *Journal of the*
45 *American Chemical Society* 135 (46), 17501-17507
- 46 **55** Singh, S. and Thakur, T. (2014) New crystalline salt forms of levofloxacin: conformational
47 analysis and attempts towards the crystal structure prediction of the anhydrous form.
48 *Crystengcomm* 16 (20), 4215-4230
- 49 **56** Bingham, A.L. et al. (2001) Over one hundred solvates of sulfathiazole. *Chemical*
50 *Communications* (7), 603-604

- 1 **57** Campeta, A.M. et al. (2010) Development of a targeted polymorph screening approach for a
2 complex polymorphic and highly solvating API. *Journal of Pharmaceutical Sciences* 99 (9),
3 3874-3886
- 4 **58** Cruz-Cabeza, A.J. et al. (2009) Predicting Inclusion Behaviour and Framework Structures in
5 Organic Crystals. *Chemistry - A European Journal* 15 (47), 13033-13040
- 6 **59** Jelfs, K.E. et al. (2013) In silico Design of Supramolecules from Their Precursors: Odd-Even
7 Effects in Cage-Forming Reactions. *Journal of the American Chemical Society* 135 (25), 9307-
8 9310
- 9 **60** Braun, D.E. et al. (2011) Which, if any, hydrates will crystallise? Predicting hydrate formation
10 of two dihydroxybenzoic acids. *Chemical Communications* 47 (19), 5443-5445
- 11 **61** Cruz-Cabeza, A.J. et al. (2010) Predicting stoichiometry and structure of solvates. *Chemical*
12 *Communications* 46 (13), 2224-2226
- 13 **62** Lupyan, D. et al. (2012) Close intramolecular sulfur-oxygen contacts: modified force field
14 parameters for improved conformation generation. *Journal of Computer-Aided Molecular*
15 *Design* 26 (11), 1195-1205
- 16 **63** Vasileiadis, M. et al. (2015) Prediction of the crystal structures of axitinib, a polymorphic
17 pharmaceutical molecule. *Chemical Engineering Science* 121, 60-76
- 18 **64** Lorenz, H. and Seidel-Morgenstern, A. (2014) Processes to separate enantiomers.
19 *Angewandte Chemie (International ed. in English)* 53 (5), 1218-1250
- 20 **65** Kendrick, J. et al. (2013) Crystal Structure Prediction of a Flexible Molecule of Pharmaceutical
21 Interest with Unusual Polymorphic Behavior. *Crystal Growth & Design* 13 (2), 581-589
- 22 **66** Huang, J. et al. (2006) Discovery of a solid solution of enantiomers in a racemate-forming
23 system by seeding. *Journal of the American Chemical Society* 128 (36), 11985-11992
- 24 **67** Ismail, S.Z. et al. (2013) Evaluating a Crystal Energy Landscape in the Context of Industrial
25 Polymorph Screening. *Crystal Growth & Design* 13 (6), 2396-2406
- 26 **68** Morissette, S.L. et al. (2003) Elucidation of crystal form diversity of the HIV protease
27 inhibitor ritonavir by high-throughput crystallisation. *Proceedings of the National Academy*
28 *of Sciences of the United States of America* 100 (5), 2180-2184
- 29 **69** Bauer, J. et al. (2001) Ritonavir: An extraordinary example of conformational polymorphism.
30 *Pharmaceutical Research* 18 (6), 859-866
- 31 **70** Feeder, N. et al. (2015) The integration of solid-form informatics into solid-form selection.
32 *Journal of Pharmacy and Pharmacology* 67 (6), 857-868
- 33 **71** Neumann, M. et al. (2015) Combined crystal structure prediction and high-pressure
34 crystallization in rational pharmaceutical polymorph screening. *Nature Communications* 6,
35 art 7793.
- 36 **72** Sun, C.C. (2009) Material Science Tetrahedron-A Useful Tool for Pharmaceutical Research
37 and Development. *Journal of Pharmaceutical Sciences* 98, 1744-1749
- 38 **73** Mohamed, S. et al. (2011) Computational prediction of salt and cocrystal structures-Does a
39 proton position matter? *International Journal of Pharmaceutics* 418 (2), 187-198
- 40 **74** Pantelides, C. et al. (2014) General Computational Algorithms for Ab Initio Crystal Structure
41 Prediction for Organic Molecules. *Topics in Current Chemistry* 345, 25-58
- 42 **75** Fakes, M. et al. (2009) Enhancement of oral bioavailability of an HIV-attachment inhibitor by
43 nanosizing and amorphous formulation approaches. *International Journal of Pharmaceutics*
44 370 (1-2), 167-174
- 45 **76** Habgood, M. et al. (2015) Efficient Handling of Molecular Flexibility in Ab Initio Generation
46 of Crystal Structures. *Journal of Chemical Theory and Computation* 11 (4), 1957-1969
- 47 **77** King, M.D. et al. (2011) Prediction of the Unknown Crystal Structure of Creatine Using Fully
48 Quantum Mechanical Methods. *Crystal Growth & Design* 11 (12), 5733-5740
- 49 **78** Gavezzotti, A. and Rizzato, S. (2014) Are Racemic Crystals Favored over Homochiral Crystals
50 by Higher Stability or by Kinetics? Insights from Comparative Studies of Crystalline
51 Stereoisomers. *Journal of Organic Chemistry* 79 (11), 4809-4816

1 **79** Yu, L. (2010) Polymorphism in Molecular Solids: An Extraordinary System of Red, Orange,
2 and Yellow Crystals. *Accounts of Chemical Research* 43 (9), 1257-1266
3 **80** Vasileiadis, M. et al. (2012) The polymorphs of ROY: application of a systematic crystal
4 structure prediction technique. *Acta Crystallographica Section B-Structural Science* 68, 677-
5 685
6 **81** Davey, R.J. et al. (2013) Nucleation of Organic Crystals A Molecular Perspective. *Angewandte*
7 *Chemie-International Edition* 52 (8), 2166-2179
8 **82** Cross, W.I. et al. (2003) A whole output strategy for polymorph screening: Combining crystal
9 structure prediction, graph set analysis, and targeted crystallization experiments in the case
10 of diflunisal. *Crystal Growth & Design* 3 (2), 151-158
11 **83** Bucar, D.K. et al. (2013) The curious case of (caffeine).(benzoic acid): how heteronuclear
12 seeding allowed the formation of an elusive cocrystal. *Chemical Science* 4 (12), 4417-4425
13 **84** Arlin, J.B. et al. (2011) A strategy for producing predicted polymorphs: catemeric
14 carbamazepine form V. *Chemical Communications* 47 (25), 7074-7076
15 **85** Childs, S.L. et al. (2009) Analysis of 50 Crystal Structures Containing Carbamazepine Using
16 the Materials Module of Mercury CSD. *Crystal Growth & Design* 9 (4), 1869-1888
17 **86** Price, C.P. et al. (2005) Crystalline polymorph selection and discovery with polymer
18 heteronuclei. *Journal of the American Chemical Society* 127 (15), 5512-5517
19 **87** Steed, K. and Steed, J. (2015) Packing Problems: High Z' Crystal Structures and Their
20 Relationship to Cocrystals, Inclusion Compounds, and Polymorphism. *Chemical Reviews* 115
21 (8), 2895-2933
22 **88** Threlfall, T. (2003) Structural and thermodynamic explanations of Ostwald's rule. *Organic*
23 *Process Research & Development* 7 (6), 1017-1027
24 **89** Bernstein, J. et al. (1995) Patterns in Hydrogen Bonding: Functionality and graph set analysis
25 in crystals. *Angewandte Chemie-International Edition in English* 34 (15), 1555-1573
26
27

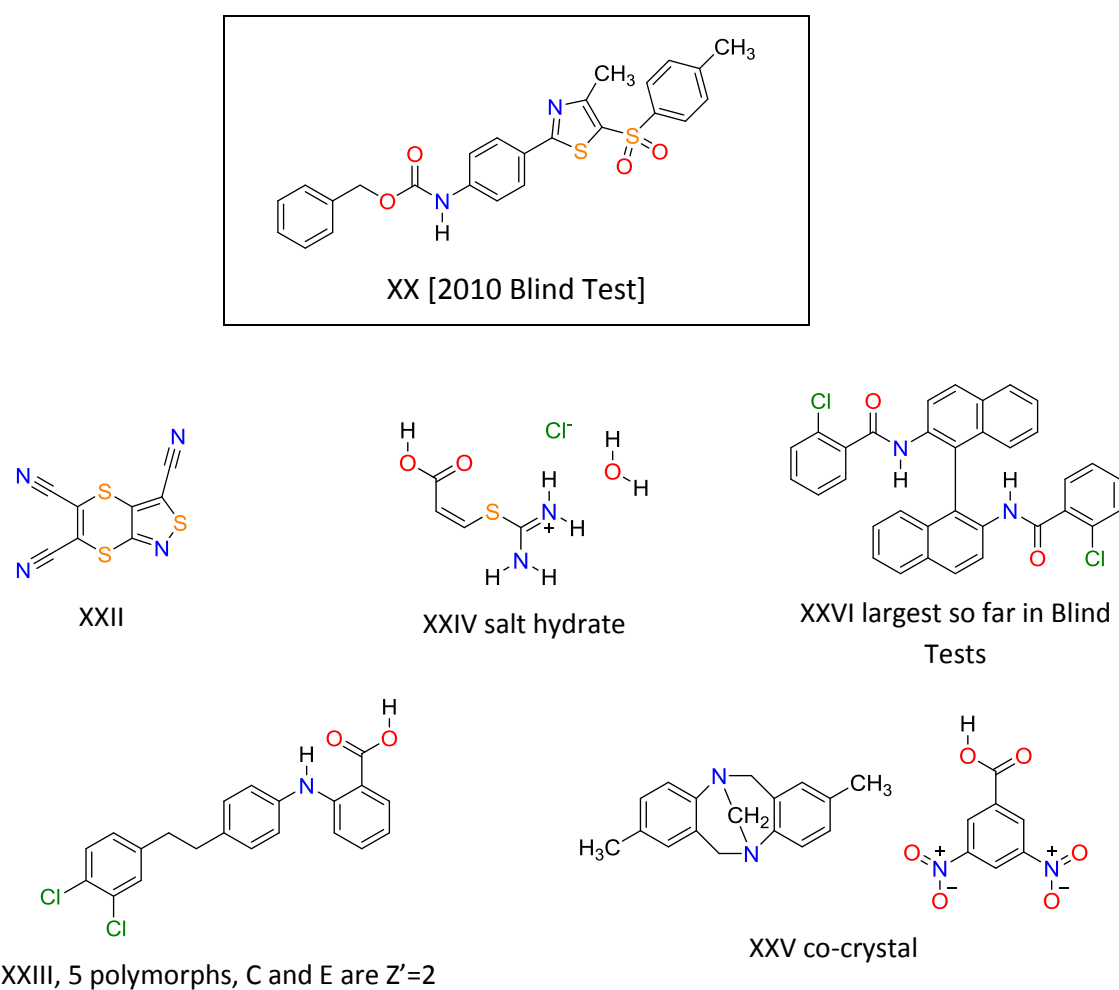
28 **Additional references, not in text, but in Box 1 [87] and [88]caption to Fig 1 [89]**

29 **Figure Captions**



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

7

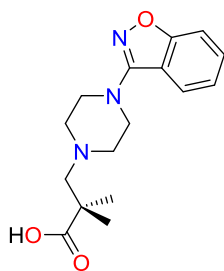


8 Figure 2 The molecules whose crystal structures have been used in the Cambridge
 9 Crystallographic Data Centre’s Blind Tests of crystal structure prediction (top, molecule XX
 10 from 5th test in 2010 [28]; rest, molecules from the 6th test in 2015). Participants were

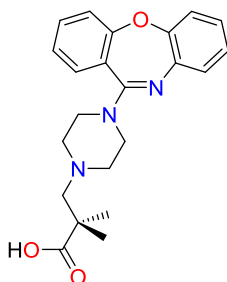
1 invited to submit lists of 100 structures and success was judged as having the experimental
2 structures within these lists. Full analysis of the ranking of the structures within each
3 submission will be published shortly [6].

4

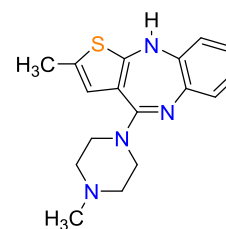
5



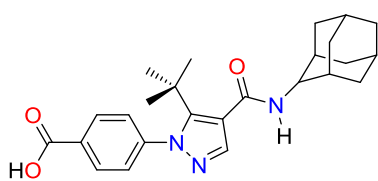
LY2806920 "B5"
1 polymorph [48]



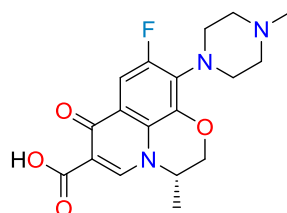
LY2624803 "DB7"
3 polymorphs, 5 solvates[48]



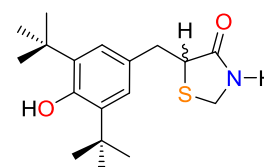
Olanzapine
3 polymorphs, 56 solvates [51]



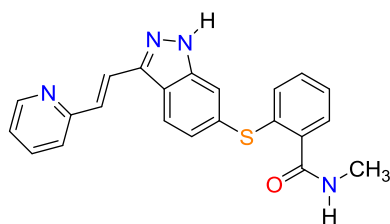
AZD8329
4 polymorphs, 3 solvates [54]



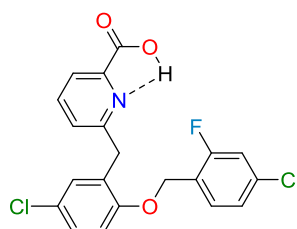
Levofloxacin
1 polymorph, 2 hydrates [55]



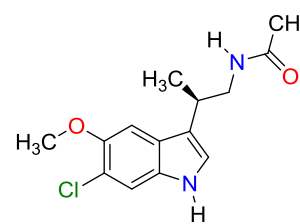
Tazofelone
3 racemic polymorphs, 1
enantiomorph, solid solution [53]



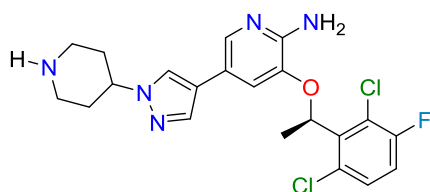
Axitinib
5 polymorphs, 66 solvates [57]



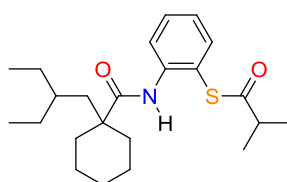
GSK269984B
1 polymorph, 4-6 solvates
[67]



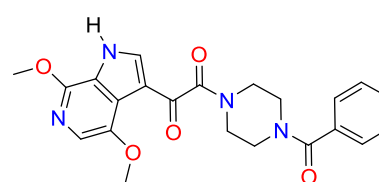
LY156735
3 polymorphs [65]



Crizotinib
1 polymorph[24]



Dalcetrapib
3 polymorphs [71]



BMS-488043
2 polymorphs [76]

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