International Journal of Pharmaceutics 418 (2011) 168–178

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

journal homepage: www.elsevier.com/locate/ijpharm

Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction

Andrei V. Kazantsev[a], Panagiotis G. Karamertzanis[a], Claire S. Adjiman[a], Constantinos C. Pantelides[a,*], Sarah L. Price[b], Peter T.A. Galek[c], Graeme M. Day[d,**], Aurora J. Cruz-Cabeza[c]

[a] Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK
[b] Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK
[c] Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK
[d] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

ARTICLE INFO

Article history:
Received 13 December 2010
Received in revised form 10 March 2011
Accepted 24 March 2011
Available online 8 April 2011

Keywords:
Crystal structure prediction
Solid form screening
Blind tests
Polymorphism
Conformation
Global optimization

ABSTRACT

The range of target structures in the fifth international blind test of crystal structure prediction was extended to include a highly flexible molecule, (benzyl-(4-(4-methyl-5-(p-tolylsulfonyl)-1,3-thiazol-2-yl)phenyl)carbamate, as a challenge representative of modern pharmaceuticals. Two of the groups participating in the blind test independently predicted the correct structure. The methods they used are described and contrasted, and the implications of the capability to tackle molecules of this complexity are discussed.

1. Introduction

Are crystal structures predictable? This, like the closely related problem of predicting protein folding (Dunitz and Scheraga, 2004), is periodically tested by communal experiments, where a previously determined crystal structure is only disclosed once participants have submitted their predictions. The Cambridge Crystallographic Data Centre (CCDC) has run a series of such blind tests of crystal structure prediction, starting in 1999 (Lommerse et al., 2000) and showing only occasional success in 2001 (Motherwell et al., 2002) and 2004 (Day et al., 2005), until there was a significant level of success with small molecules in 2007 (Day et al., 2009). The main method that has been applied to crystal structure prediction is global lattice energy minimization: structure searching methods are used to generate the possible ways of packing the molecule into a crystal structure, which are ranked according to their calculated energies. This set of structures, and their relative energies, are key features of the crystal energy landscape and the lowest energy structures on this landscape are assumed to be the most likely to be observed experimentally. The results of such calculations in the blind tests, and in many other independent crystal structure prediction studies, demonstrate that a wide range of different crystal structures are available to most molecules and that these structures are usually sufficiently close in energy that calculated relative crystal energies need to be accurate to a fraction of a kJ mol^{-1} for a confident ranking. This has been most frequently achieved in the blind tests by computationally expensive methods involving anisotropic atom–atom intermolecular potentials, sometimes derived purely from quantum mechanical calculations on the isolated molecule (Price, 2009), or from quantum mechanical electronic structure calculations applied directly to the crystal structures (Neumann et al., 2008).

Abbreviations: Molecule XX, (benzyl-(4-(4-methyl-5-(p-tolylsulfonyl)-1,3-thiazol-2-yl)phenyl)carbamate; FCC, Flexible CrystalPredictor–CrystalOptimizer Method; RCM, Rigid CrystalPredictor–Molecular Mechanics Method; CCDC, Cambridge Crystallographic Data Centre; CSD, Cambridge Structural Database; DFT, density functional theory i.e. electronic structure calculations; MM, molecular mechanics i.e. atomistic modeling using force-fields; DMA, distributed multipole analysis for generating atomic multipoles; PCM, polarizable continuum model; $rmsd_{15}$, root mean square deviation in the 15-molecule coordination sphere excluding hydrogen atoms; $rmsd_1$, root mean square deviation in the 1-molecule coordination sphere (i.e. molecular conformation) excluding hydrogen atoms; LHP, logit hydrogen-bonding propensity.

* Corresponding author. Tel.: +44 0 20 7594 5622; fax: +44 0 20 7594 6606.
** Corresponding author. Tel.: +44 01223 336390; fax: +44 0 1223 336362.
E-mail addresses: c.pantelides@imperial.ac.uk (C.C. Pantelides), gmd27@cam.ac.uk (G.M. Day).

0378-5173/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2011.03.058

Concurrently with the blind tests, it has become clear that the methods used in crystal structure prediction can be used to complement experimental solid form screenings (Braun et al., 2011) and hence inform the choice of solid form for drug development. The crystal energy landscape may provide additional reassurance that all likely long-lived polymorphs are already known or, where the calculations suggest thermodynamically feasible alternative crystal packings, allow the design of crystallization conditions to produce such potential polymorphs (Lancaster et al., 2006). Such computationally inspired polymorph discovery was recently demonstrated by growing the predicted catemeric polymorph (form V) of carbamazepine from the vapor onto an isomorphous crystal template (Arlin et al., submitted for publication). An alternative application of computed crystal structures is to help characterize structures where good single crystals cannot be grown, in conjunction with, for example, unindexable powder diffraction patterns (Cruz-Cabeza et al., 2010; Tremayne et al., 2004), terahertz spectra (Parrott et al., 2009) or solid-state NMR chemical shifts (Salager et al., 2010). Careful analysis of the crystal energy landscape can also point towards more complex behavior: if the crystal energy landscape has related structures that are close in energy, this may suggest a tendency to certain forms of disorder that can complicate spectra (Li et al., 2010), and may hinder the growth of single crystals and the development of a robust production process, as demonstrated for eniluracil (Copley et al., 2008). The prediction that the hydrogen bonded layers within aspirin could stack in two different ways (Ouvrard and Price, 2004) explained the later discovery of new forms with different properties and illuminated the debate over whether this was a case of polymorphism, polymorphic domains or degrees of stacking disorder (Bond et al., 2007). Hence the importance of the prediction of crystal structures in the blind tests is to verify that the crystal energy landscape is sufficiently realistic to be worth considering whether thermodynamically competitive structures may be possible polymorphs.

In the 2010 blind test, the challenge was extended to include several more complex targets, including a new category of crystal structures consisting of a flexible molecule with 50–60 atoms, 4–8 internal degrees of freedom, in any space group and with one or two independent molecules in the asymmetric unit. From the crystal structures provided in confidence to the CCDC, that of (benzyl-(4-(4-methyl-5-(p-tolylsulfonyl)-1,3-thiazol-2-yl)phenyl)carbamate was chosen for this category, noting that it was far more typical of modern pharmaceuticals than any other target. It became the 20th target in the series (and hence denoted molecule XX). In November 2009, the participants were given the molecular diagram (i.e. the covalent bonding) shown in Fig. 1 and were informed that the crystal was obtained by slow evaporation from an ethyl acetate solution. Each participant was required to submit three predictions of the crystal structure together with an extended list of low energy crystal structures on their crystal energy landscape by the 20 August 2010. A paper is jointly being prepared by all participants to describe the different approaches of the 15 contributing groups (10 of which submitted an entry for molecule XX), their performance for the 6 diverse targets, and the conclusions of the blind test meeting (Bardwell et al., in preparation). No method was successful for all targets, which were chosen to represent the challenges of different types of crystal structures, including a salt and a hydrate. The methods used in crystal structure prediction are often tailored specifically for each target molecule, and this is particularly important for molecule XX, a large molecule with greater flexibility than other targets. Here, we discuss and contrast the methods used for molecule XX by the two groups who successfully predicted the observed crystal structure of this blind test target, which was predicted as the most stable computed crystal structure by both groups. The computational models described in this paper were successful in predicting the structure of the molecule XX crystal, but this does not imply these same models are suitable for other molecules. Success depends to a large extent on finding the right combination of model accuracy and computational cost.

2. Methods

The crystal structure of molecule XX was successfully predicted in the Blind Test by two methods, referred to as Flexible CrystalPredictor–CrystalOptimizer (FCC) method and the Rigid CrystalPredictor–Molecular Mechanics method (RCM). Despite their differences, these two methodologies have noticeable similarities. In both approaches, a four-step procedure, outlined in Table 1, is followed. The key stages include a conformational analysis, an extensive crystal structure search, lattice energy minimizations using elaborate models for the intramolecular energy and the electrostatic interactions and finally, the examination of the lowest energy structure and other energetically feasible crystal structures. In both methods, the intramolecular energy and charge density are computed by *ab initio* methods assuming that the molecule in the crystalline phase can be approximated by the molecule in a vacuum or in a dielectric continuum. Thermodynamic stability is determined by the calculation of the lattice energy:

$$E_{\text{latt}} = U_{\text{inter}} + \Delta E_{\text{intra}} \quad (1)$$

where U_{inter} is the intermolecular energy contribution and ΔE_{intra} is the energy of the molecular conformation in that crystal structure, relative to its most stable conformation. In both methods, density functional theory (DFT) electronic structure calculations were performed on single molecules using GAUSSIAN (Frisch et al., 2003; Frisch et al., 2009), to obtain ΔE_{intra} and the molecular charge density, which was subjected to a distributed multipole analysis (DMA) (Stone and Alderton, 1985) using GDMA (Stone, 1999; Stone, 2005). The resulting atomic multipoles, along with an empirically derived repulsion–dispersion potential, were used to model each crystal structure using DMACRYS (Price et al., 2010) to optimize U_{inter} with the anisotropic atom–atom model intermolecular potential that represents the electrostatic effects of lone pairs and π electrons on the directionality of the hydrogen bonding and π–π stacking interactions. A general description of both methods follows, with particular emphasis on the deviations from published methods required by the novel challenge of the type of flexibility in molecule XX.

2.1. FCC (Flexible CrystalPredictor–CrystalOptimizer) method

A conformational analysis is first performed to restrict the search space to energetically meaningful regions. In the Flexible CrystalPredictor–CrystalOptimizer method (FCC), this is done by quantum mechanical (B3LYP/6-31G(d,p)) scans (Frisch et al., 2009), supported by an analysis of the Cambridge Structural Database (CSD) (Allen, 2002) on fragments of molecule XX. This was used to identify the feasible ranges for the main torsional angles in the molecule (Table 2). The CSD was also searched to determine the statistically expected number of molecules in the asymmetric unit and the space groups of crystals containing molecules of similar size to molecule XX.

Based on this analysis, eight separate flexible CrystalPredictor (Karamertzanis and Pantelides, 2007) searches were carried out for $Z' = 1$ structures in the 12 most common space groups ($P2_1/c$, $P2_12_12_1$, $P\bar{1}$, $P2_1$, $C2/c$, *Pbca*, $Pna2_1$, *C2*, *P1*, *Cc*, $Pca2_1$, $P2_12_12$) in the crystal structure generation step (step 2 of Table 1). During the search, only 7 major torsional angles were allowed to change (Ph-CH_2, $PhCH_2$-OCO, CONH-Ph, R(6)-R(5), R(5)-SO_2, SO_2-Ph and CO-NH) with the amide group (CO-NH) in either the *trans* or *cis*

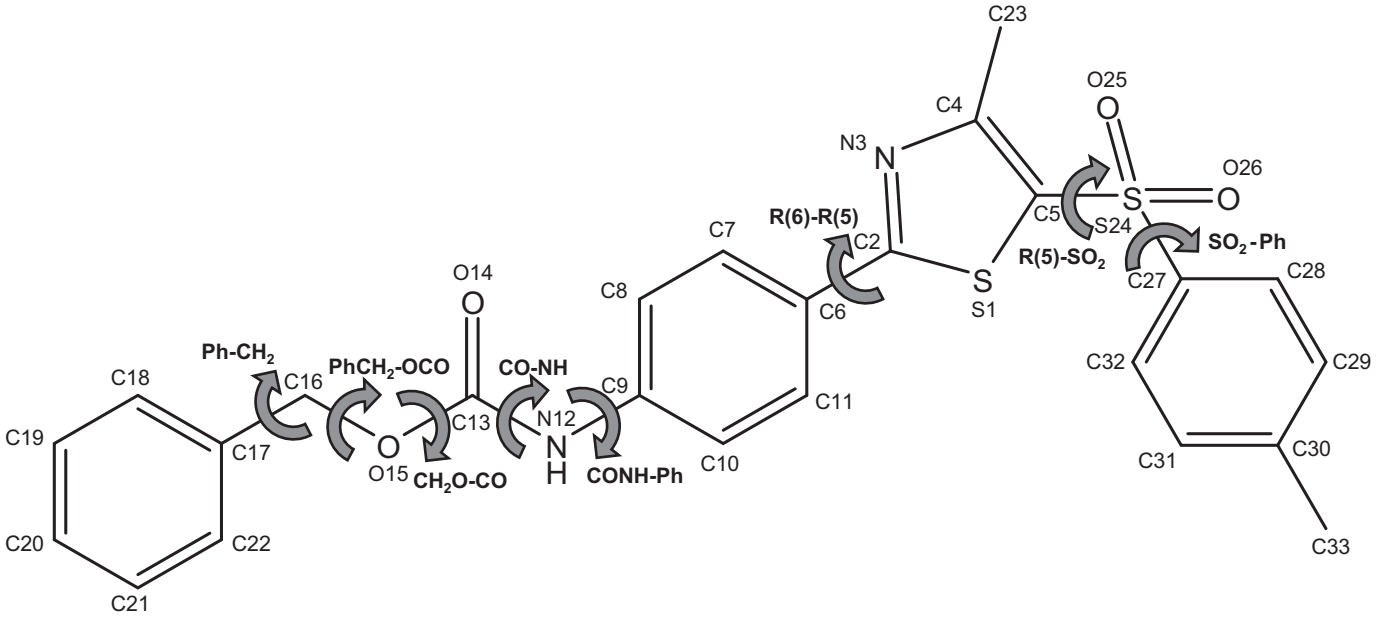


Fig. 1. Molecular structure of target XX with the definitions of key torsion angles and atom labels.

planar conformation. To reduce the computational cost, interpolation was used in the evaluation of the intramolecular energy. For this purpose, two sub-molecules were derived from molecule XX that both include the central phenyl ring: one consisted of atoms 6–22 and the other of atoms 1–11 and 23–33 (defined in Fig. 1), and a hydrogen atom was added to each of these fragments to ensure there were no free bonds. A grid of ΔE_{intra} values for the flexible torsions considered in the search was then derived by scanning three torsion angles on each sub-molecule. The deformation energy for the first sub-molecule was computed on two $7 \times 10 \times 13$ grids (one each for *trans* and *cis* amide conformations) and for the second sub-molecule, on four $7 \times 7 \times 7$ grids. At each grid point the deformation energy was calculated (Schmidt et al., 1993) at the B3LYP/6-31G(d,p) level of theory with the flexible torsions fixed

Table 1
Outline of the FCC and RCM methods for crystal structure prediction.

	FCC Method	RCM Method
Step 1. Conformational Analysis	Cambridge Structural Database (CSD) and DFT Specify 8 molecular models based on conformational regions, defined by ranges for the main torsions	Specify 48 molecular models, defined by distinct rigid conformations
Step 2. Crystal Structure Generation	CrystalPredictor and Clustering Flexible-molecule search in each conformational region, 12 common space groups, interpolated intramolecular energy, FIT+ESP charges for intermolecular forces 2,800,000 lattice energy minimizations 800,000 distinct structures generated 2000 CPU hours for intramolecular grid generation 16,000 CPU hours for search	Rigid-body search for each distinct conformation in 21 common space groups DFT(PCM, $\varepsilon = 3$)/W99 + ESP 9,600,000 lattice energy minimizations 93,000 distinct structures generated 580 CPU hours for DFT geometry optimization and ESP(PCM, $\varepsilon = 3$) charges 41,500 CPU hours for search
Step 3. Refinement of Energy Models and Clustering	CrystalOptimizer Re-minimization of stable structures using DFT-accuracy intramolecular energy and conformation-dependent DMA. 14 torsions and 5 bond angles explicitly optimized during lattice energy minimization; the rest of intramolecular degrees of freedom optimized within isolated-molecule wavefunction calculations. 1500 structures 100,000 CPU hours	MM and DMACRYS Re-minimization of stable structures using two different MM intramolecular energy models and Gasteiger-derived charges 10 torsions explicitly optimized 1: DREIDING 2: COMPASS Rigid-body minimization using DMA model Intramolecular energy obtained from DFT calculations using Polarizable Continuum Model ($\varepsilon = 3$) DFT(PCM, $\varepsilon = 3$)/W99+DMA 1500 structures 12,000 CPU hours
Step 4. Identification of 3 Structures for Submission for Blind Test	The submission structures were the two lowest in energy. The third submission was the lowest energy structure with a *cis*-amide.	The three lowest structures were chosen for submission. The ranking of the submitted structures was determined from their energies and a hydrogen bond propensity model.
Total CPU requirements	~120,000 CPU hours	~54,000 CPU hours
Further Computational Resource Details	Intel Xeon 5150 2.66 GHz processor, 1500 MB of memory DFT—GAUSSIAN 09 (Frisch et al., 2009) DMA—GDMA (Stone, 2005) Repulsion–dispersion potential—FIT (Coombes et al., 1996)	AMD Opteron 285 2.60 GHz processor, 256 MB of memory DFT—GAUSSIAN 03 (Frisch et al., 2003) DMA—GDMA (Stone, 2005) with (Stone, 1999) solver. Repulsion–dispersion potential—W99 (Williams, 2001)

Table 2
Torsion angles varied or fixed to discrete values in the search for the FCC and the RCM methods. The torsion values in the experimental conformation are also given. All combinations of the ranges/values are used to generate the search space in both methods.

#	Φ name	Φ definition	FCC Method Φ Ranges or Values (°)	RCM Method Φ Values (°)	Experimental Conformation (°)
1	Ph-CH_2	C22-C17-C16-O15	[−180.0, +180.0]	+0.0, +50.0, +90.0, +130.0	+82.2
2	$PhCH_2$-OCO	C17-C16-O15-C13	[+40.0, −50.0][a]	+90.0, +180.0, −90.0	−105.8
3	CH_2O-CO	C16-O15-C13-O14	0.0	0.0	+6.3
4	CO-NH	C9-N12-C13-O15	0.0 (*cis*) 180.0 (*trans*)	– 180.0 (*trans*)	– +176.4
5	CONH-Ph	C8-C9-N12-C13	[−60.0, +60.0]	0.0	+1.1
6	R(6)-R(5)	C7-C6-C2-S1	[−60.0, +60.0] [+120.0, −120.0]	0.0 180.0	−11.8 –
7	R(5)-SO_2	S1-C5-S24-C27	[+50.0, +170.0] [−170.0, −50.0]	+90.0 −90.0	+104.7 –
8	SO_2-Ph	C28-C27-S24-C5	[+30.0, +150.0]	+90.0	+107.0
Number of Conformational Ranges/Discrete conformations after Φ-combinations			8 conformational regions	48 conformations	–

[a] Note that this range spans 270° and corresponds to [+40.0, +310.0].

and the rest of the molecule optimized using the semi-empricial AM1 level of theory (Dewar et al., 1985). The intramolecular energy of molecule XX was then approximated as the sum of the deformation energy of the two sub-molecules, assuming that there are no significant interactions between these two parts. For the search in each conformational region, the intermolecular electrostatic interactions were modeled using the atomic charges that were derived from the B3LYP/6-31G(d,p) electrostatic potential of the B3LYP/6-31G(d,p) conformational minimum of the whole molecule in this region. All other intermolecular energy terms were derived from an empirical exp-6 potential, FIT (Coombes et al., 1996).

In the refinement step (step 3 of Table 1), the 1500 stable structures within 10 kJ mol^{-1} of the global minimum were re-minimized using CrystalOptimizer (Kazantsev et al., 2010, in press) with 19 flexible degrees of freedom (14 non-aromatic torsions and 5 selected chain bond angles). Local approximate models (LAMs) were constructed on-the-fly for the conformational variations of the intramolecular energy and the distributed multipole moments (Stone, 2005) at the PBE0/6-31G(d,p) level of theory and stored for reuse for similar conformations in subsequent lattice energy minimizations. The minimized structures were then clustered based on their root mean square deviation in the 15-molecule coordination sphere, $rmsd_{15}$ (Chisholm and Motherwell, 2005). Two structures were considered to be crystallographically similar if their $rmsd_{15}$ was below 0.25 Å. Even if these are distinct minima mathematically, they are likely to interconvert to each other under thermal motion.

The structures selected for the FCC Blind Test submission were the two lowest in energy (with a 0.78 kJ mol^{-1} gap). They differed significantly in that the second ranked structure was less dense but had a conventional N–H···N hydrogen bond (N12···N3 2.9 Å). The third submission was the lowest energy structure with a *cis* amide conformation (11.43 kJ mol^{-1} above the global minimum), in case this isomer of molecule XX had been synthesized.

2.2. RCM (Rigid CrystalPredictor–Molecular Mechanics) method

2.2.1. Crystal structure prediction methodology

In the Rigid CrystalPredictor–Molecular Mechanics method (RCM), the investigation also begins with a conformational analysis. The CSD was used to analyze the conformational preferences by comparing fragments of molecule XX to flexible molecules with similar functionalities whose crystal structures are present in the crystal structure database. The CSD provides many tools (e.g. Mogul, ConQuest and Vista applications) (Bruno et al., 2002, 2004) which allow the user to extract the expected ranges of values of the flexible degrees of freedom for molecules with similar functionalities or fragments. In the case of the torsion R(5)-SO_2, the statistical data from the CSD was insufficient to define the angle distribution, so a DFT (B3LYP/6-31G(d,p)) constrained geometry scan was performed instead. As a result of the CSD analysis and DFT calculations, a total of 48 distinct conformations were obtained, with the main torsion angles summarized in Table 2. The geometry, intramolecular energy (ΔE_{intra}) and ESP atomic charges of each conformation were obtained from a constrained geometry optimization at the B3LYP/6-31G(d,p) level of theory (with the flexible torsion angles fixed at the expected CSD values and all other degrees of freedom optimized). Molecular DFT calculations at this stage, and in the final energy evaluation (described below), were performed within a continuum dielectric to approximate the molecule's environment in the solid state, and the resulting polarization of the charge density. The Polarizable Continuum Model (PCM) (Tomasi et al., 2005) was used for these calculations, with the dielectric constant fixed at a typical value for molecular organic crystals, $\varepsilon = 3$. This approximate model of introducing polarization effects has been shown to have an important influence on relative conformational energies and electrostatic interactions in crystal structure prediction of polar, flexible molecules (Cooper et al., 2008).

A separate CrystalPredictor (Karamertzanis and Pantelides, 2005) search was performed for each conformation to generate crystal structures in the 21 most common space groups ($P2_1/c$, $P2_12_12_1$, $P\bar{1}$, $P2_1$, $C2/c$, $Pbca$, $Pna2_1$, $C2$, $P1$, Cc, $Pca2_1$, $P2_12_12$, $Pbcn$, $Pnma$, $Pccn$, Pc, $P2_1/m$, $P2/c$, $C2/m$, $R3$, $R\bar{3}$), each with $Z' = 1$. The molecular geometry was treated as rigid during the RCM crystal structure generation process and relative energies of the resulting crystal structures were assessed from Eq. (1), with U_{inter} calculated from the ESP atomic charges and an empirical exp-6 repulsion–dispersion potential, W99 (Williams, 2001).

The 1500 most stable crystal structures resulting from these rigid-molecule searches were then refined in two steps, based on the procedure described in previous publications (Day et al., 2007; Day and Cooper, 2010). First, the lattice energy of the crystal structures was minimized to allow the molecular geometry to relax within each crystal structure, using a molecular mechanics (MM) description of energies associated with changes to the torsion angles that were treated as flexible. The method trusts the MM force field to provide the correct molecular geometries, but discards the MM energy, which is not of sufficient accuracy for the final ranking of crystal structures. These intermediate molecular mechanics lattice energy minimizations were performed with 11 rigid units

within the molecule constrained to their DFT optimized geometries. These calculations optimize the 8 flexible torsion angles defined in Fig. 1 plus the two methyl group rotations within each crystal structure. Bond angles between the rigid groups were also optimized, while the bond lengths between the rigid units were restrained to the DFT optimized values during this intermediate energy minimization. To avoid relying on one force field, each structure was MM energy minimized twice: once using the COMPASS force field (Sun, 1998) with its own atomic charges, and once using the DREIDING force field (Mayo et al., 1990) with Gasteiger derived charges (Gasteiger and Marsili, 1980).

The resulting crystal structures were re-optimized without further changes to molecular conformations using an atomic multipole model for the evaluation of the electrostatic interactions. The remaining intermolecular terms were obtained using the empirical W99 potential (Williams, 2001). The intramolecular energy and the atomic multipole moments for each conformation considered were derived from a single point DFT calculation at the B3LYP/6-31G(d,p) level using the PCM model with $\varepsilon = 3$. At this point, the DREIDING and COMPASS structures were re-evaluated on the same energy surface and could be combined and clustered to remove duplicates; of each set of duplicates, the lowest energy structure was retained. The intramolecular energy of the 60 most stable crystal structures was further refined through a constrained DFT minimization and a subsequent energy evaluation using DFT and the PCM model; torsion angles were constrained to the MM optimized values, while all other degrees of freedom were optimized. Hence, the final structures of the RCM model are not minima on the final energy surface; only molecular positions, orientations and the unit cell are minimized with the final energy model.

2.2.2. Hydrogen bond propensity modeling

A logit hydrogen–bonding propensity (LHP) model (Galek et al., 2007, 2009) was trained to predict the most likely hydrogen bonding to be formed in the crystal structure of molecule XX. The results were used to check that the low energy predicted crystal structures from the RCM method formed probable hydrogen bonds.

LHP modeling is a knowledge-based method for assessing the most likely acceptors and donors to form hydrogen bonds, and is trained against data from crystal structures of molecules with similar functional groups taken from the CSD. Each potential donor–acceptor (D–A) interaction is treated as having a dichotomous probability, and its propensity for formation is modeled by a strict probability function. To train a hydrogen-bond propensity model for molecule XX, 494 crystal structures of similar molecules were obtained from the CSD. The query functional groups were sulfone, thiazole and carbamate and the training set contained 148, 164 and 182 instances of each functional group, respectively. The hydrogen bonds which occur in the training structures were identified using distance (r_{D-A}) and angle (θ_{DHA}) criteria: $r_{D-A} < \Sigma r_{vdW} + 0.1$ Å, and a $\theta_{DHA} > 120^\circ$ (where r_{vdW} is the atomic van der Waals radius). The model function was then trained to best reproduce each true or false observation. The method uses logistic regression to optimize the contribution of explanatory model variables describing steric accessibility, competition, cooperativity and electrostatics (Galek et al., 2007). Statistical validation techniques were employed: hydrogen-bond propensity models achieve between 80 and 90% correct prediction in blind tests (Galek et al., 2010). This model is similarly accurate, achieving 83.1% correct discrimination of true and false outcomes in the training set.

The three lowest energy crystal structures were chosen for the RCM submission, with the lowest energy crystal structure submitted as the first prediction. Two strong candidates for an intermolecular hydrogen bond emerge from the LHP model: carbamate-NH···O=C carbamate and carbamate-NH···O–S sulfone. Either is likely to form with very little difference in probability. Of the three lowest energy calculated structures, the 1st and 3rd contain N-H···O=S hydrogen bonds, although the donor–acceptor separation is outside of the geometric criteria used in the LHP model in the lowest energy structure. The 2nd lowest energy predicted structure contains a very long, non-linear (DHA angle = 129°) N–H···O=S interaction. Therefore, the 3rd lowest energy structure was submitted as the 2nd most likely prediction, and the 2nd lowest energy as 3rd prediction.

3. Results and discussion

The lowest energy structures found using both computational methodologies are almost identical and overlay with the experimental structure to within the accuracy required for the Blind Test. Overlays of these predictions with the experimental structure are given in Fig. 2 and Table 3, along with the predicted and observed unit cell parameters. Both predictions overlay the 15-molecule coordination sphere of the experimental structure well, with a value of $rmsd_{15}$ of 0.178 Å in the FCC method and 0.429 Å in the RCM method: this level of accuracy was considered a good agreement in previous blind tests on smaller molecules ($rmsd_{15} < 0.5$ Å). The experimental structure contains the elongated hydrogen bond, N–H···O=S of the sulfone, predicted by the hydrogen bond propensity model, with a N12···O26 distance of 3.377 Å, 0.3 Å longer than the sum of van der Waals radii. This distance is reproduced very well in the RCM predicted structure (3.328 Å), but is longer (3.626 Å) in the FCC predicted structure.

Additional differences and similarities in each stage of the two approaches are analyzed in more detail in this section.

3.1. Step 1: Conformational analysis

Identification of the accessible conformational space of molecule XX by the FCC and RCM methods was achieved using two different approaches, but with some overlap. Both effectively split the molecule into smaller fragments, assuming that the fragments can be chosen in such a way that the conformational changes in one fragment do not significantly affect the conformation of the other fragments. Molecule XX was particularly suited for calculating the intramolecular energy grid from two sub-molecules (FCC method) and being likely to conform to the statistical analysis of experimental crystal structures deposited in the CSD. Studies of conformational preferences in small molecule crystal structures (Brameld et al., 2008) have shown that similar fragments in different molecules often adopt similar conformations. Furthermore, where the conformational preferences are statistically significant, they are very rarely found to have comparatively high gas-phase energies (Allen et al., 1996). The close relationship between the two types of conformational analysis, suggested by Table 2, is illustrated in Fig. 3 for the torsion $PhCH_2$-OCO. The DFT conformational energy and the CSD observations are generally in good agreement. At values for which the conformational energy penalty is high, few or no experimental occurrences are found in the CSD. The FCC method used relevant conformational ranges [+40.0, −50.0] from this analysis while the RCM adopted discrete values of the torsion angles using the maxima of the CSD distribution (+90.0, +180.0, −90.0). The experimental conformation takes the absolute value of 105.8° for this torsion, 25° from the closest local energy minimum and 16° from one of the maxima in the CSD statistical distribution (Fig. 3).

The successful prediction of crystal structures of a wide range of small molecules using local torsional energy minima and rigid-molecule searches suggests the crystal conformation may be reasonably close to a gas phase conformation that corresponds to a low energy minimum. This assumption does not hold for flexible molecules such as molecule XX. A more appropriate assumption

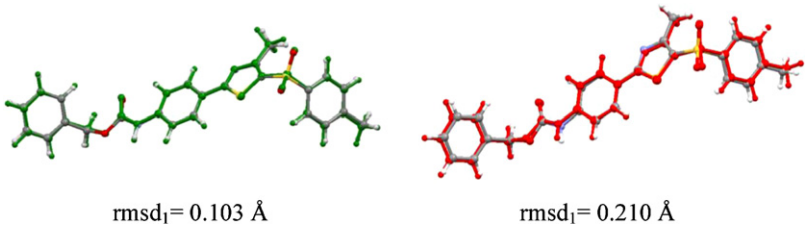
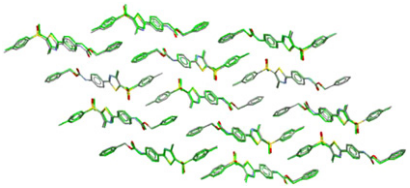


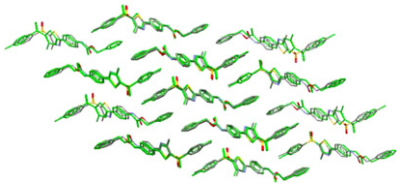
Fig. 2. Overlay of the experimental conformation (grey; colored by element in the online version) with the final conformations obtained using the FCC (dark grey, left; green in the online version) and RCM (dark grey, right; red in the online version) methods respectively.

Table 3
Overlay of the experimental structure (grey; colored by element in the online version) and lowest energy structures (black; green in the online version) obtained by the two computational approaches (excluding hydrogen atoms). The unit cell parameters and structure similarity are also given.

FCC Method



RCM Method



Method	Density (g cm^{-3})	Space Group	Unit Cell Dimensions				rmsd$_{15}$ (Å)[a]
			a (Å)	*b* (Å)	*c* (Å)	*β* (°)	
Experimental	1.410	$P2_1/n$	14.08	6.36	25.31	96.1	–
FCC	1.402	$P2_1/n$	14.26	6.32	25.36	97.3	0.178
RCM	1.375	$P2_1/n$	14.13	6.24	26.31	95.6	0.429

[a] Root mean square deviation in atomic positions (excluding hydrogen atoms) of the 15-molecule coordination spheres compared to the experimental structure.

is that the crystal conformation usually has a gas-phase energy that is close to the minimum gas-phase energy (i.e. within a few kJ mol^{-1}). This does not preclude significant geometrical differences between the crystal conformation and the minimum energy gas phase conformation, when large differences in torsions angle incur a small energy penalty. Fig. 4a shows that torsion SO_2-Ph, and especially torsion $PhCH_2$-OCO, in the crystal structure of molecule XX deviate significantly from the closest local minimum conformation from DFT gas phase calculations. The torsion angle $PhCH_2$-OCO can change by more than 100° with an energy penalty of less than 2 kJ mol^{-1} (Fig. 3). Thus, it is appropriate to select a large torsion range for the $PhCH_2$-OCO fragment, and in general it is important to include all low energy conformational regions (Table 2). The success of the two approaches is shown in Fig. 4: despite significant differences in the crystal conformation and the closest local gas phase energy minimum (Fig. 4a), the RCM method selected one initial

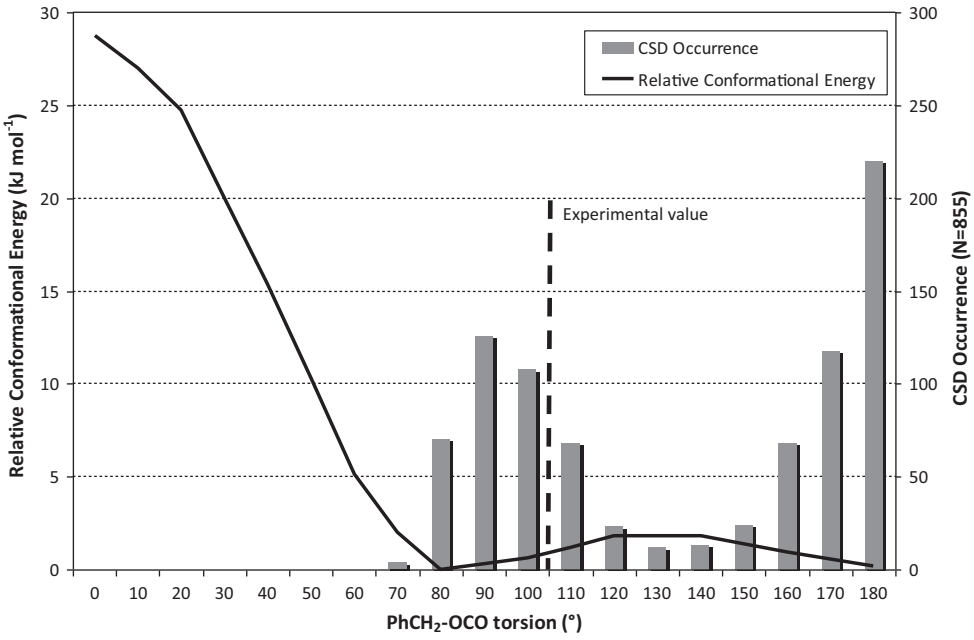


Fig. 3. Comparison of the CSD MOGUL geometry analysis (bars, right axis) with gas-phase intramolecular energy scan at the B3LYP/6-31G(d,p) level of theory (black line, left axis) for the torsion $PhCH_2$-OCO in molecule XX. *N* is the total number of similar fragments obtained from the CSD. The dashed line indicates the experimental value as observed in the crystal structure of molecule XX.

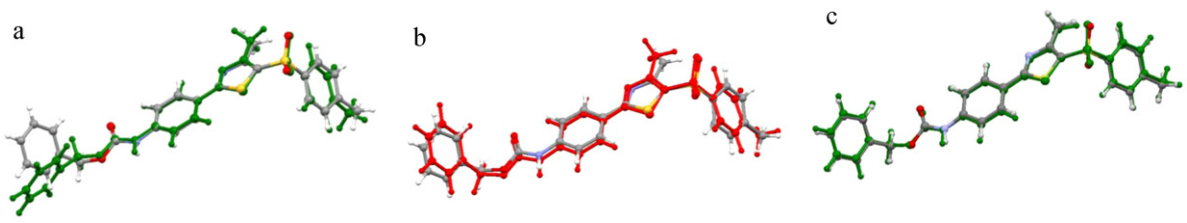


Fig. 4. Overlay of the experimental conformation (grey; colored by element in the online version) with a) the closest local gas-phase minimum conformation obtained with B3LYP/6-31G(d,p) (black; green in the online version) (b) the closest conformation generated using the CSD statistical data on torsions (with remaining intramolecular degrees of freedom at the B3LYP/6-31G(d,p) minimum) in the RCM method (black; red in the online version) (step 1 in Table 1), and (c) the conformation obtained from the flexible search in the FCC method (black; green in the online version) after the flexible search (step 2 in Table 1).

conformation which is very similar to the conformation later found in the experimental crystal structure ($rmsd_1$ = 0.368 Å, Fig. 4b), and the FCC method's flexible Crystal Predictor search found a good approximation to the final conformation ($rmsd_1$ = 0.167 Å, Fig. 4c).

3.2. Step 2: Crystal structure generation

The main purpose of the search algorithm (step 2 of Table 1) is to generate good initial starting crystal structures for subsequent lattice energy minimization. This search is a complex multidimensional problem in which it is necessary to consider many optimization variables such as the space group, the size and shape of the unit cell as well as the relevant molecular conformations. Both crystal structure prediction methodologies outlined in this paper use the same search algorithm (Karamertzanis and Pantelides, 2005, 2007) which systematically and uniformly samples different space groups, torsion values, unit cell dimensions and position and orientation of molecules within the crystal. In both methodologies, a limited range of space groups was considered and it was assumed there would be only one molecule in the asymmetric unit in order to concentrate computer resources on the most probable types of structures. The main difference between the FCC and RCM methods in step 2 is the treatment of molecular flexibility during the search (Table 1).

In the FCC method, the only assumptions about the molecular geometry were the choice of the 7 main torsion angles to be treated as flexible, and their specified ranges (as outlined in Table 2). The explicit treatment of flexibility during the search led to the identification of a conformation similar to the experimental one ($rmsd_1$ = 0.167 Å, Fig. 4c) and of a crystal structure within $rmsd_{15}$ = 0.311 Å of the experimental structure, by the end of step 2.

In the RCM method, the initial conformational analysis in step 1 is fundamental. This analysis led to the choice of 48 discrete molecular conformations generated using the experimental observations in the CSD (Table 2) for the series of rigid-body searches and preliminary minimizations. The main assumption in this approach is that the error introduced in the lattice arrangement by imposing a rigid body can be recovered by subsequent refinement of low energy structures when molecular flexibility is allowed during lattice energy minimization (step 3). One of the conformations chosen in the search for molecule XX was very close to the experimental geometry ($rmsd_1$ = 0.368 Å, Fig. 4b), which ultimately led to the generation of the observed crystal structure. The RCM method would have had a lower chance of locating the experimental crystal structure if energy minimized geometries of the isolated molecule had been chosen as the starting conformations (e.g. the closest DFT gas phase local minimum to the experimental conformation shown in Fig. 4a), as this would have relied on large conformational changes during the subsequent flexible-molecule lattice energy minimization.

From a practical point of view, rigid-molecule searches are simpler to implement than flexible searches, and preliminary results can be obtained almost immediately. However, due to the high degree of flexibility of molecule XX, 48 different conformations were considered in the RCM method, and an extensive search (200,000 minimizations) was completed for every conformation. As a result, the structure generation step required 42,000 CPU hours for input file creation and search. In the case of the flexible CrystalPredictor search used in the FCC approach, it was necessary to create an intramolecular energy grid from relatively expensive DFT calculations before any structure generation could take place. Furthermore, because more variables were sampled by the algorithm than in a rigid-body search, more minimizations were performed (350,000 minimizations in each of the 8 distinct regions). This was sufficient to capture most of the effects of the molecular flexibility exhibited by molecule XX within a total of 18,000 CPU hours for input file creation and structure generation. In comparing the CPU times for both methods, it should be noted that 12 space groups were considered in the FCC method and 21 in the RCM method, and that the *cis* conformation of the carboxamide group was additionally considered in the FCC method but not in the RCM method.

The correct prediction of the crystal structure of molecule XX, as seen in Table 3, shows that both methods were successful in identifying crystal structures that were sufficiently good initial points to lead to the determination of the experimental form in the subsequent lattice energy minimizations.

3.3. Steps 3 & 4: Crystal structure refinement and relative lattice energies

The ability to search the very large space of possible crystal structures in step 2 relies on simplified, relatively inexpensive models for the intermolecular and intramolecular energy contributions. Therefore, the unit cell parameters, molecular conformations, and the lattice energies require improvement using more realistic models for the lattice energy before a final assessment of the predicted crystal structures can be made. This is apparent from the observation that the experimental structure was ranked 223rd and 427th after step 2 in the FCC and RCM methods respectively. In the FCC method, the molecular geometry and the lattice parameters were simultaneously re-optimized using the CrystalOptimizer software (Kazantsev et al., 2010, in press) using the PBE0/6-31G(d,p) level for the molecular geometries and charge density; in the RCM method, a sequential optimization approach was used: the molecular geometry was refined using intermediate MM lattice energy minimizations before applying the final energy model to the refinement of lattice parameters. Although both methods ultimately used DFT estimates of ΔE_{intra} and evaluated U_{inter} from a distributed multipole electrostatic model, isolated molecule charge densities were used in the FCC model, whereas charge densities were derived in the RCM model by using a dielectric continuum to approximate the crystal environment (using the PCM model).

Table 4
Lattice parameters and energy ranking of the 10 most stable structures obtained using the RCM method.

Structure	E_{latt} (kJ mol^{-1})	Density (g cm^{-3})	Space Group	Unit Cell Dimensions						[a] E_{latt} with FCC method (kJ mol^{-1})
				a	*b*	*c*	*α*	*β*	*γ*	
				(Å)			(°)			
RCM_1	−216.93	1.375	$P2_1/n$	14.13	6.24	26.31	90.0	95.6	90.0	−218.61
RCM_2	−216.41	1.417	$P2_1/c$	6.72	12.86	26.65	90.0	76.6	90.0	−195.73
RCM_3	−215.14	1.353	$P\bar{1}$	8.03	13.61	11.53	84.6	71.9	78.5	−211.67
RCM_4	−214.69	1.385	$P2_1/c$	16.28	4.89	28.97	90.0	83.3	90.0	−203.40
RCM_5	−214.06	1.374	$P2_1/c$	4.60	23.52	21.65	90.0	80.6	90.0	−208.94
RCM_6	−213.94	1.359	$P2_1/c$	20.55	5.12	32.90	90.0	137.5	90.0	−206.57
RCM_7	−213.63	1.389	$P2_1$	16.39	4.78	14.65	90.0	94.9	90.0	−204.21
RCM_8	−213.14	1.351	$P\bar{1}$	7.96	13.60	11.75	84.5	72.3	76.1	−212.31
RCM_9	−213.02	1.382	$P\bar{1}$	10.11	15.74	7.30	93.4	97.4	87.6	−208.68
RCM_10	−212.94	1.394	*Pbca*	21.28	12.19	17.55	90.0	90.0	90.0	−198.71

[a] The lattice energy, E_{latt}, of these structures when re-optimized using CrystalOptimizer and the FCC final energy model.

Table 5
Lattice parameters and energy ranking of the 10 most stable structures obtained using the FCC method.

Structure	E_{latt} (kJ mol^{-1})	Density (g cm^{-3})	Space Group	Unit Cell Dimensions						[a] E_{latt} using RCM final energy model (kJ mol^{-1})
				a	*b*	*c*	*α*	*β*	*γ*	
				(Å)			(°)			
FCC_1	−218.73	1.402	$P2_1/n$	14.26	6.32	25.36	90.0	97.3	90.0	−215.50
FCC_2	−217.95	1.352	$P2_1/c$	13.54	10.71	17.69	90.0	66.4	90.0	−227.28
FCC_3	−216.35	1.373	$I2/a$	28.04	6.66	26.58	90.0	68.9	90.0	−218.08
FCC_4	−213.14	1.401	$P\bar{1}$	5.01	10.52	21.85	84.8	95.9	96.4	−219.65
FCC_5	−212.58	1.388	$P2_1/a$	16.15	12.39	12.73	90.0	64.1	90.0	−217.51
FCC_6	−212.31	1.407	$P\bar{1}$	24.27	27.49	12.23	54.3	63.2	116.6	−215.43
FCC_7	−211.47	1.338	$P\bar{1}$	9.98	12.57	11.72	64.1	66.4	70.2	−215.59
FCC_8	−211.04	1.372	$Pca2_1$	45.06	5.13	10.02	90.0	90.0	90.0	−211.62
FCC_9	−210.88	1.381	$P2_1/n$	6.38	18.32	19.76	90.0	85.5	90.0	−218.62
FCC_10	−210.76	1.403	$P\bar{1}$	4.92	23.02	10.06	85.6	93.4	92.1	−210.93

[a] The lattice energy, E_{latt}, of these structures when re-calculated using the RCM final energy model, without allowing the molecular conformation to change.

Both methods successfully identified the experimental structure as the lowest energy structure after the final refinement of the lattice energies (structures ranked as 1, Tables 4 and 5), and both showed a small energy gap to other structures, with 10 distinct[1] crystal structures spanning 7.97 kJ mol^{-1} (less than 2 kcal mol^{-1}) and 3.99 kJ mol^{-1} (1 kcal mol^{-1}) with the FCC and RCM methods respectively (Tables 4 and 5). These structures had very diverse conformations, with different packing motifs and, in some cases, different hydrogen bonds (Table 6). Hence both calculations find that other sufficiently different crystal structures are well within the energy range to be thermodynamically feasible as polymorphs.

A comparison of the two energy landscapes is aided by re-minimizing the two sets of low energy structures obtained from the FCC and RCM methods with each other's final model for the lattice energy (van Eijck, 2005). To explore the differences in the final energy models used in both methods, the 10 lowest energy distinct RCM structures have been fully minimized using the CrystalOptimizer approach and FCC energy model. These re-minimized RCM structures (whose energies are reported in Table 4) are directly comparable to those generated by the FCC prediction methodology. The reverse comparison has been performed by performing rigid-molecule lattice energy minimization using the RCM final energy model on the 10 lowest energy distinct FCC structures; these results are summarized in Table 5. The comparison here is less straightforward because the molecular geometries of the FCC structures are not allowed to relax to a local minimum during re-evaluation with the RCM model.

The marked differences in the relative energies (Tables 4 and 5) clearly demonstrate that the uncertainties in the relative lattice energies are large compared with the small energy differences between the possible structures. Only three of the ten low energy structures produced in the RCM search are sufficiently low in energy to be amongst the ten lowest energy structures found by the FCC method, the experimental structure (RCM_1 ~ FCC_1) and two others (RCM_3 ~ FCC_7 and RCM_8 ~ FCC_6). These differences in the relative energies are not surprising given the range of conformations and hydrogen bonds (Table 6) found within the low energy structures. The conformational energy penalty, ΔE_{intra} for such a flexible molecule is quite sensitive to the quantum mechanics method used (van Mourik et al., 2006); even greater variations can be seen if we approximate the effect of the crystalline environment on the molecular energy. The use of the PCM model stabilizes certain molecular geometries, changing the relative ΔE_{intra} values by up to 2 kJ mol^{-1}. The intermolecular lattice energy, $U_{inter,}$ also differs between the two empirically fitted repulsion–dispersion models, and is affected by the difference in molecular charge density between the isolated molecule and that in the crystalline environment. The PCM model with $\varepsilon = 3$ changed the relative electrostatic contributions to U_{inter} by up to 15 kJ mol^{-1} and led to shorter hydrogen bond donor–acceptor distances than an electrostatic model derived from unpolarized molecular calculations; these effects are sensitive to the value of the dielectric constant, ε (Cooper et al., 2008). The overall influence on the relative energies here is important: there is significant re-ranking of the FCC structures when re-calculated using the RCM final energy model, so that

[1] The FCC structures presented here include some structures which are beyond the first 10 in the extended lists submitted as part of the blind test. The clustering tolerance used in producing the extended list for the blind test has been found to be too tight. The list presented in Table 5 has been generated with a less stringent tolerance, thereby eliminating a few very similar structures.

Table 6
Predicted combinations of hydrogen bond propensities for various acceptor atoms with the N12 carbamate as donor in molecule XX. Hydrogen bonds were detected in the structures using Tables 4 and 5 using Mercury and the following cut-off values: $r_{H\text{-}A} < \Sigma r_{vdW} + 0.1$ Å, and a $\theta_{DHA} > 120°$.

Hydrogen bond acceptor	Predicted Propensity	FCC structures containing the hydrogen bond	RCM structures containing the hydrogen bond
O25/26 of sulfone	0.72	FCC_1, FCC_3, FCC_7, FCC_9	RCM_1, RCM_3, RCM_4, RCM_5, RCM_6, RCM_7, RCM_8, RCM_10
O14 of carbamate	0.68	FCC_8 most stable *cis* carbamate structures	–
N3 of thiazole	0.39	FCC_2	–
O15 of carbamate	0.03	–	–
S1 of thiazole	0.00	–	–
No clear hydrogen bond		FCC_4, FCC_5, FCC_6, FCC_10	RCM_2, RCM_9

the energies of 5 of the 10 low energy structures are lower than the energy of the experimental structure (Table 5). It is important to remember that the intramolecular degrees of freedom in these structures were optimised by different methods, so that this comparison does not show how the structures would have been ranked using a consistent optimisation strategy. A complete lattice energy minimization on the final RCM energy surface, including the influence of the polarization model on molecular geometry, would be required to determine whether these resulting crystal structures truly correspond to lower energy structures than the experimental structure.

Overall, these results confirm that the choice of computational model and of minimization strategy have an important impact on the relative stabilities of different crystal structures. These issues, which are explored in detail by the blind tests, apply to molecule XX as much as to smaller molecules. Nonetheless, the agreement between the hydrogen bonding seen in the low energy predicted structures, and the predicted hydrogen bond propensities (Table 6) shows encouraging consistency. The low energy RCM structures overwhelmingly favor the hydrogen bond with the highest predicted propensity and the FCC search has a low energy structure for each of the three hydrogen bonds with a significant propensity; none of the computed structures showed the two lowest-propensity hydrogen bonds.

The RCM and the FCC structure refinement methods differ significantly in computational cost. The use of molecular mechanics force fields for the optimization of molecular geometry within the crystal is computationally inexpensive (50 CPU hours for the minimization of 1500 structures), but relies on the force field to approximate the true equilibrium molecular geometry in each crystal structure. Therefore, the speed comes at the expense of sacrificing some accuracy. In this case, simple, general force fields were employed and the accuracy of this approach will improve as higher quality transferable force fields, or molecule-specific "tailor-made" force fields (Neumann, 2008) are developed. By far, the main cost in the RCM refinement approach (12,000 CPU hours) arises from the use of DFT to calculate the intramolecular energies and the atomic multipole moments for the final energy calculation, in which the molecular positions and unit cell parameters are optimized. The automated FCC refinement using the CrystalOptimizer algorithm is significantly more computationally expensive (100,000 CPU hours for the refinement of 1500 structures) due to the use of the results of a large number of optimization variables (molecular geometry and lattice parameters) and of DFT calculations during lattice energy minimization. Nevertheless, the computational cost was kept manageable by using local approximate models (LAMs) and LAM databases that provide DFT accuracy at a much reduced cost when CrystalOptimizer is used to refine many structures.

4. Conclusions and outlook

We have been able to successfully predict the crystal structure of a highly flexible molecule, with a complexity typical of those currently being developed in the pharmaceutical industry. By setting molecule XX (benzyl-(4-(4-methyl-5-(p-tolylsulfonyl)-1,3-thiazol-2-yl)phenyl)carbamate) as a target molecule, the Fifth Blind Test of crystal structure prediction has inspired innovations to adapt the methods previously used for small peptides (Day and Cooper, 2010; Gorbitz et al., 2010), uracils (Barnett et al., 2008) and small generic pharmaceuticals and their multicomponent crystals (Cruz-Cabeza et al., 2006; Habgood and Price, 2010; Karamertzanis et al., 2009) to molecules of such complexity.

One challenge was to account for the high degree of flexibility of molecule XX during the search for crystal structures, as several torsion angles can vary considerably with only small variations in molecular energy. Searching through the entire conformational space remains prohibitively expensive for this molecular size, so that a subset of the conformational space was considered in both methods. This was achieved either by explicitly considering the torsion angles as variables in the search (FCC), but with limited ranges, or by carefully choosing a large number of rigid conformations for the preliminary search (RCM) using experimental data available from the CSD. The question in both cases is how the conformational space that is considered in crystal structure prediction can be effectively reduced, without eliminating important conformations that will lead to low energy crystal structures. Simply identifying all low energy conformational energy minima and assuming that these will be close to any solid state conformation is clearly inadequate for molecule XX, although this strategy has been successful for some smaller molecules. It is clear that all low energy conformational regions have to be considered, and the agreement in defining these regions from an analysis of existing crystal structures and from the use of quantum mechanical energy scans is encouraging.

Another key challenge was the computational cost of dealing with such a large and flexible molecule during the more accurate and demanding calculations of the final energy refinement stage. Two approaches were investigated. In one case, the optimization problem was decomposed, so that molecular and lattice geometries were optimized sequentially, using energy functions with different costs and accuracies (MM or DFT + PCM in RCM). In the other case, the recently developed CrystalOptimizer algorithm was used to minimize the lattice energy evaluated at the final level of accuracy, with respect to 19 geometrical variables simultaneously with the lattice variables, at a higher but nevertheless accessible computational cost (FCC). While both approaches resulted in the identification of the experimental structure as that with the lowest total energy, many of the other low energy structures found by the two methods differed significantly in stability order. The differences can be attributed partly to the use of different models of lattice energy, for both the conformational and intermolecular contributions, and partly to the design of the optimization strategies.

The hope that molecules with sufficient flexibility will find one mode of packing that is significantly more stable than any others, and therefore being readily predictable, has not been realized with molecule XX. Both methods show that there are alternative structures with different conformations and intermolecular interactions

that are well within the energy range of being possible polymorphs, let alone the likely errors in the models for the relative lattice energies. These structures all satisfy several "sanity-checks", such as the likely hydrogen bonds and conformations derived from the CSD structures, as well as falling in a small density range. Hence, large flexible molecules, like most small molecules, provide a challenge to computational chemistry to develop sufficiently accurate and efficient models for the relative energies of crystal structures to be able to confidently predict the most thermodynamically stable form.

In treating molecule XX, the assumption that the target structure is either monomorphic or, if polymorphic, the most thermodynamically stable one, and that it can be identified as the global minimum in the lattice energy rather than the free energy, appears to be appropriate. However, the blind tests of crystal structure prediction only test our ability to predict the structure of a molecule that crystallizes well enough to be solved by single crystal X-ray diffraction, in a crystal with one or two molecules in the asymmetric unit cell, and without disorder.

The successful prediction of the crystal structure of molecule XX in the blind test indicates that search methods and models for lattice energy are capable of tackling this type of molecule to give worthwhile results, both in terms of the range of structures considered in the search and relative energies of the structures. However the two methods disagree as to the most likely structures if polymorphs of molecule XX exist. Thus, there remains a need to further develop algorithms that are more efficient for this type of molecule so that we can increase the level of accuracy of the relative energies and extent of the search as the molecule and type of study demands (Price, 2008). In pharmaceutical development, the calculation of the crystal energy landscape can complement solid form screening beyond confirming that the most thermodynamically stable form has been found. Guiding the search for different types of crystal structure that appear to be feasible polymorphs should aid late stage polymorph screening. Polymorphs that are formed by desolvation of metastable solvates are more likely to be kinetically trapped for large molecules, which are generally much less able than approximately spherical molecules to rearrange significantly late in the crystallization process (Hulme et al., 2007). Isomorphic desolvates should be predictable, as relatively stable structures which contain voids do appear as local minima on the crystal energy landscapes of molecules which form inclusion compounds (Cruz-Cabeza et al., 2009). This paper shows that we are now at the stage where we can learn more about the crystallization of pharmaceuticals from comparing the crystal energy landscapes with the outcomes of polymorph screens.

Although we are still a long way from understanding, let alone reliably predicting all solid forms of pharmaceutical molecules, the results of the blind test of crystal structure prediction for molecule XX demonstrate a step change in the complexity of molecules for which a crystal energy landscape can be calculated. It is now possible to calculate crystal energy landscapes that can be used in conjunction with experimental polymorph screening. By improving computational models and techniques to give reliable crystal free energy landscapes and consider kinetic and other factors (such as solvent effects), we can move towards a predictive technology for the understanding and anticipation of polymorphism.

Acknowledgements

The financial support from the Engineering and Physical Sciences Research Council (EPSRC) from the Molecular Systems Engineering grant (EP/E016340) and the CPOSS Basic Technology program (www.cposs.org.uk, EP/F03573X/1) is gratefully acknowledged. Calculations using the FCC method were performed on the High Performance Computing Cluster at Imperial College London. GMD and AJCC thank the Pfizer Institute for Pharmaceutical Materials Sciences for funding. GMD also thanks the Royal Society for funding of a University Research Fellowship. The CCDC is thanked for arranging and hosting the Blind Tests of Crystal Structure Prediction.

References

Allen, F.H., 2002. The Cambridge Structural Database: a quarter of a million crystal structures and rising. Acta Crystallogr. Sect. B 58, 380–388.

Allen, F.H., Harris, S.E., Taylor, R., 1996. Comparison of conformer distributions in the crystalline state with conformational energies calculated by ab-initio techniques. J. Comput. Aided Mol. Des. 10, 247–254.

Arlin, J.B., Price, L.S., Price, S.L., Florence, A.F. A strategy for producing predicted polymorphs: catemeric carbamezapine form V. Chem. Commun., submitted for publication.

Bardwell, D.A., Adjiman, C.S., Ammon, H.L., Arnautova, E.A., Bartashevich, E., Boerrigter, S.X.M., Braun, D.E., Cruz-Cabeza, A.J., Day, G.M., Della Valle, R.G., Desiraju, G.R., van Eijck, B.P., Facelli, J.C., Ferraro, M.B., Grillo, D., Habgood, M., Hofmann, D.W.M., Hofmann, F., Jose, J., Karamertzanis, P.G., Kazantsev, A.V., Kendrick, J., Kuleshova, L.N., Leusen, F.J.J., Maleev, A., Misquitta, A.J., Mohamed, S., Needs, R.J., Neumann, M.A., Nikylov, D., Orendt, A.M., Pal, R., Pantelides, C.C., Pickard, C.J., Price, L.S., Price, S.L., Scheraga, H. A., van de Streek, J., Thakur, T.S., Tiwari, S., Venuti, E., Zhitkov, I. Towards crystal structure prediction of pharmaceutically relevant molecules—a report on the fifth blind test. Acta Crystallogr., Sect. B, in preparation.

Barnett, S.A., Hulme, A.T., Issa, N., Lewis, T.C., Price, L.S., Tocher, D.A., Price, S.L., 2008. The observed and energetically feasible crystal structures of 5-substituted uracils. New J. Chem. 32, 1761–1775.

Bond, A.D., Boese, R., Desiraju, G.R., 2007. On the polymorphism of aspir Crystalline aspirin as intergrowths of two "polymorphic" domains. Angew Chem. Int. Ed. 46, 618–622.

Brameld, K.A., Kuhn, B., Reuter, D.C., Stahl, M., 2008. Small molecule conformational preferences derived from crystal structure data. A medicinal chemistry focused analysis. J. Chem. Inform. Model. 48, 1–24.

Braun, D.E., Karamertzanis, P.G., Arlin, J.B., Florence, A.J., Kahlenberg, V., Tocher, D.A., Griesser, U.J., Price, S.L., 2011. Solid-state forms of β-resorcylic acid: how exhaustive should a polymorph screen be? Cryst. Growth Des. 11, 210–220.

Bruno, I.J., Cole, J.C., Edgington, P.R., Kessler, M., Macrae, C.F., McCabe, P., Pearson, J., Taylor, R., 2002. New software for searching the Cambridge Structural Database and visualizing crystal structures. Acta Crystallogr. Sect. B 58, 389–397.

Bruno, I.J., Cole, J.C., Kessler, M., Luo, J., Motherwell, W.D.S., Purkis, L.H., Smith, B.R., Taylor, R., Cooper, R.I., Harris, S.E., Orpen, A.G., 2004. Retrieval of crystallographically-derived molecular geometry information. J. Chem. Inform. Comp. Sci. 44, 2133–2144.

Chisholm, J.A., Motherwell, S., 2005. COMPACK: a program for identifying crystal structure similarity using distances. J. Appl. Crystallogr. 38, 228–231.

Coombes, D.S., Price, S.L., Willock, D.J., Leslie, M., 1996. Role of electrostatic interactions in determining the crystal structures of polar organic molecules. A Distributed Multipole Study. J. Phys. Chem. 100, 7352–7360.

Cooper, T.G., Hejczyk, K.E., Jones, W., Day, G.M., 2008. Molecular polarization effects on the relative energies of the real and putative crystal structures of valine. J. Chem. Theory Comput. 4, 1795–1805.

Copley, R.C.B., Barnett, S.A., Karamertzanis, P.G., Harris, K.D.M., Kariuki, B.M., Xu, M.C., Nickels, E.A., Lancaster, R.W., Price, S.L., 2008. Predictable disorder versus polymorphism in the rationalization of structural diversity: A multidisciplinary study of eniluracil. Cryst. Growth Des. 8, 3474–3481.

Cruz-Cabeza, A.J., Day, G.M., Jones, W., 2009. Predicting inclusion behaviour and framework structures in organic crystals. Chem. Eur. J. 15, 13033–13040.

Cruz-Cabeza, A.J., Day, G.M., Motherwell, W.D.S., Jones, W., 2006. Prediction and observation of isostructurality induced by solvent incorporation in multicomponent crystals. J. Am. Chem. Soc. 128, 14466–14467.

Cruz-Cabeza, A.J., Karki, S., Fabian, L., Friscic, T., Day, G.M., Jones, W., 2010. Predicting stoichiometry and structure of solvates. Chem. Commun. 46, 2224–2226.

Day, G.M., Cooper, T.G., 2010. Crystal packing preditions of the alpha-amino acids: methods assessment and structural observations. CrystEngComm 12, 2443–2453.

Day, G.M., Cooper, T.G., Cruz-Cabeza, A.J., Hejczyk, K.E., Ammon, H.L., Boerrigter, S.X.M., Tan, J.S., Della Valle, R.G., Venuti, E., Jose, J., Gadre, S.R., Desiraju, G.R., Thakur, T.S., van Eijck, B.P., Facelli, J.C., Bazterra, V.E., Ferraro, M.B., Hofmann, D.W.M., Neumann, M.A., Leusen, F.J.J., Kendrick, J., Price, S.L., Misquitta, A.J., Karamertzanis, P.G., Welch, G.W.A., Scheraga, H.A., Arnautova, Y.A., Schmidt, M.U., van de Streek, J., Wolf, A.K., Schweizer, B., 2009. Significant progress in predicting the crystal structures of small organic molecules—a report on the fourth blind test. Acta Crystallogr. Sect. B 65, 107–125.

Day, G.M., Motherwell, W.D.S., Ammon, H.L., Boerrigter, S.X.M., Della Valle, R.G., Venuti, E., Dzyabchenko, A., Dunitz, J.D., Schweizer, B., van Eijck, B.P., Erk, P., Facelli, J.C., Bazterra, V.E., Ferraro, M.B., Hofmann, D.W.M., Leusen, F.J.J., Liang, C., Pantelides, C.C., Karamertzanis, P.G., Price, S.L., Lewis, T.C., Nowell, H., Torrisi, A., Scheraga, H.A., Arnautova, Y.A., Schmidt, M.U., Verwer, P., 2005. A third blind test of crystal structure prediction. Acta Crystallogr. Sect. B 61, 511–527.

Day, G.M., Motherwell, W.D.S., Jones, W., 2007. A strategy for predicting the crystal structures of flexible molecules: the polymorphism of phenobarbital. Phys. Chem. Chem. Phys. 9, 1693–1704.

Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., Stewart, J.J.P., 1985. AM1: a new general purpose quantum mechanical molecular model. J. Am. Chem. Soc. 107, 3902–3909.

Dunitz, J.D., Scheraga, H.A., 2004. Exercises in prognostication: Crystal structures and protein folding. In: Proc. Natl. Acad. Sci. U. S. A. , 101, pp. 14309–14311.

Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, J., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J.E., Hratchian, H.P., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Gonzalez, C., Pople, J.A., 2003. Gaussian 03. Gaussian Inc, Wallingford, CT.

Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J.A., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg, J.J., Dapprich, S., Daniels, A.D., Farkas, O., Foresman, J.B., Ortiz, J.V., Cioslowski, J., Fox, D.J. 2009. Gaussian 09. Gaussian, Inc, Wallingford, CT.

Galek, P.T.A., Fabian, L., Motherwell, W.D.S., Allen, F.H., Feeder, N., 2007. Knowledge-based model of hydrogen-bonding propensity in organic crystals. Acta Crystallogr. Sect. B B63, 768–782.

Galek, P.T.A., Allen, F.H., Fabian, L., Feeder, N., 2009. Knowledge-based H-bond prediction to aid experimental polymorph screening. CrystEngComm 11, 2634–2639.

Galek, P.T.A., Fabian, L., Allen, F.H., 2010. Truly prospective prediction: inter- and intramolecular hydrogen bonding. CrystEngComm 12, 2091–2099.

Gasteiger, J., Marsili, M., 1980. Iterative partial equalization of orbital negativity—a rapid access to atomic charges. Tetrahedron 36, 3219–3228.

Gorbitz, C.H., Dalhus, B., Day, G.M., 2010. Pseudoracemic amino acid complexes: blind predictions for flexible two-component crystals. Phys. Chem. Chem. Phys. 12, 8466–8477.

Habgood, M., Price, S.L., 2010. Isomers, conformers, and cocrystal stoichiometry: insights from the crystal energy landscapes of caffeine with the hydroxybenzoic acids. Cryst. Growth Des. 10, 3263–3272.

Hulme, A.T., Johnston, A., Florence, A.J., Fernandes, P., Shankland, K., Bedford, C.T., Welch, G.W.A., Sadiq, G., Haynes, D.A., Motherwell, W.D.S., Tocher, D.A., Price, S.L., 2007. Search for a predicted hydrogen bonding motif—A multidisciplinary investigation into the polymorphism of 3-azabicyclo[3.3.1]nonane-2,4-dione. J. Am. Chem. Soc. 129, 3649–3657.

Karamertzanis, P.G., Kazantsev, A.V., Issa, N., Welch, G.W.A., Adjiman, C.S., Pantelides, C.C., Price, S.L., 2009. Can the formation of pharmaceutical cocrystals be computationally predicted? II. Crystal structure prediction. J. Chem. Theory Comput. 5, 1432–1448.

Karamertzanis, P.G., Pantelides, C.C., 2005. Ab initio crystal structure prediction—I. Rigid molecules. J. Comput. Chem. 26, 304–324.

Karamertzanis, P.G., Pantelides, C.C., 2007. Ab initio crystal structure prediction II. Flexible molecules. Mol. Phys. 105, 273–291.

Kazantsev, A.V., Karamertzanis, P.G., Adjiman, C.S., Pantelides, C.C. Efficient handling of molecular flexibility in lattice energy minimization of organic crystals. J. Chem. Theory Comput., in press.

Kazantsev, A.V., Karamertzanis, P.G., Pantelides, C.C., Adjiman, C.S., 2010. An efficient algorithm for lattice energy minimization of organic crystals using isolated-molecule quantum mechanical calculations. In: Adjiman, C.S., Galindo, A. (Eds.), Process Systems Engineering: Molecular Systems Engineering, vol. 6. Wiley-VCH, Hamburg, pp. 1–42.

Lancaster, R.W., Karamertzanis, P.G., Hulme, A.T., Tocher, D.A., Covey, D.F., Price, S.L., 2006. Racemic progesterone: predicted in silico and produced in the solid state. Chem. Commun., 4921–4923.

Li, R., Zeitler, J.A., Tomerini, D., Parrott, E.P.J., Gladden, L.F., Day, G.M., 2010. A study into the effect of subtle structural details and disorder on the terahertz spectrum of crystalline benzoic acid. Phys. Chem. Chem. Phys. 12, 5329–5340.

Lommerse, J.P.M., Motherwell, W.D.S., Ammon, H.L., Dunitz, J.D., Gavezzotti, A., Hofmann, D.W.M., Leusen, F.J.J., Mooij, W.T.M., Price, S.L., Schweizer, B., Schmidt, M.U., van Eijck, B.P., Verwer, P., Williams, D.E., 2000. A test of crystal structure prediction of small organic molecules. Acta Crystallogr. Sect. B 56, 697–714.

Mayo, S.L., Olafson, B.D., Goddard, W.A., 1990. Dreiding—A generic force-field for molecular simulations. J. Phys. Chem. 94, 8897–8909.

Motherwell, W.D.S., Ammon, H.L., Dunitz, J.D., Dzyabchenko, A., Erk, P., Gavezzotti, A., Hofmann, D.W.M., Leusen, F.J.J., Lommerse, J.P.M., Mooij, W.T.M., Price, S.L., Scheraga, H., Schweizer, B., Schmidt, M.U., van Eijck, B.P., Verwer, P., Williams, D.E., 2002. Crystal structure prediction of small organic molecules: a second blind test. Acta Crystallogr. Sect. B 58, 647–661.

Neumann, M.A., 2008. Tailor-made force fields for crystal-structure prediction. J. Phys. Chem. B 112, 9810–9829.

Neumann, M.A., Leusen, F.J.J., Kendrick, J., 2008. A major advance in crystal structure prediction. Angew. Chem. Int. Ed. 47, 2427–2430.

Ouvrard, C., Price, S.L., 2004. Toward crystal structure prediction for conformationally flexible molecules: The headaches illustrated by aspirin. Cryst. Growth Des. 4, 1119–1127.

Parrott, E.P.J., Zeitler, J.A., Friscic, T., Pepper, M., Jones, W., Day, G.M., Gladden, L.F., 2009. Testing the sensitivity of terahertz spectroscopy to changes in molecular and supramolecular structure: a study of structurally similar cocrystals. Cryst. Growth Des. 9, 1452–1460.

Price, S.L., 2008. Computational prediction of organic crystal structures and polymorphism. Int. Rev. Phys. Chem. 27, 541–568.

Price, S.L., 2009. Computational methodologies: towards crystal structure and polymorph prediction. In: Brittain, H.G. (Ed.), Polymorphism in Pharmaceutical Solids. Informa Healthcare USA Inc., New York, pp. 53–76.

Price, S.L., Leslie, M., Welch, G.W.A., Habgood, M., Price, L.S., Karamertzanis, P.G., Day, G.M., 2010. Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. Phys. Chem. Chem. Phys. 12, 8478–8490.

Salager, E., Day, G.M., Stein, R.S., Pickard, C.J., Elena, B., Emsley, L., 2010. Powder crystallography by combined crystal structure prediction and high-resolution h-1 solid-state NMR spectroscopy. J. Am. Chem. Soc. 132, 2564–2566.

Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.H., Koseki, S., Matsunaga, N., Nguyen, K.A., Su, S., Windus, T.L., Dupuis, M., Montgomery, J.A., 1993. General atomic and molecular electronic structure systems. J. Comput. Chem. 14, 1347–1363.

Stone, A.J., 1999. GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System Cambridge. University of Cambridge, United Kingdom.

Stone, A.J., 2005. Distributed multipole analysis: Stability for large basis sets. J. Chem. Theory Comput. 1, 1128–1132.

Stone, A.J., Alderton, M., 1985. Distributed multipole analysis—methods and applications. Mol. Phys. 56, 1047–1064.

Sun, H., 1998. COMPASS: An ab initio force-field optimized for condensed-phase applications—Overview with details on alkane and benzene compounds. J. Phys. Chem. B 102, 7338–7364.

Tomasi, J., Mennucci, B., Cammi, R., 2005. Quantum mechanical continuum solvation models. Chem. Rev. 105, 2999–3093.

Tremayne, M., Grice, L., Pyatt, J.C., Seaton, C.C., Kariuki, B.M., Tsui, H.H.Y., Price, S.L., Cherryman, J.C., 2004. Characterization of complicated new polymorphs of chlorothalonil by X-ray diffraction and computer crystal structure prediction. J. Am. Chem. Soc. 126, 7071–7081.

van Eijck, B.P., 2005. Comparing hypothetical structures generated in the third Cambridge blind test of crystal structure prediction. Acta Crystallogr. Sect. B 61, 528–535.

van Mourik, T., Karamertzanis, P.G., Price, S.L., 2006. Molecular conformations and relative stabilities can be as demanding of the electronic structure method as intermolecular calculations. J. Phys. Chem. A 110, 8–12.

Williams, D.E., 2001. Improved intermolecular force field for molecules containing H, C N, and O atoms, with application to nucleoside and peptide crystals. J. Comput. Chem. 22, 1154–1166.