

Towards crystal structure prediction of complex organic compounds – a report on the fifth blind test

David A. Bardwell,^{a*} Claire S. Adjiman,^b Yelena A. Arnautova,^c Ekaterina Bartashevich,^d Stephan X. M. Boerrigter,^e Doris E. Braun,^f Aurora J. Cruz-Cabeza,^{a,g,h} Graeme M. Day,ⁱ Raffaele G. Della Valle,^j Gautam R. Desiraju,^k Bouke P. van Eijck,^l Julio C. Facelli,^{m,n} Marta B. Ferraro,^o Damian Grillo,^o Matthew Habgood,^f Detlef W. M. Hofmann,^{p,q} Fridolin Hofmann,^{q,r} K. V. Jovan Jose,^s Panagiotis G. Karamertzanis,^b Andrei V. Kazantsev,^b John Kendrick,^t Liudmila N. Kuleshova,^p Frank J. J. Leusen,^t Andrey V. Maleev,^u Alston J. Misquitta,^v Sharmarke Mohamed,^f Richard J. Needs,^v Marcus A. Neumann,^w Denis Nikylov,^d Anita M. Orendt,^m Rumpa Pal,^k Constantinos C. Pantelides,^b Chris J. Pickard,^x Louise S. Price,^f Sarah L. Price,^f Harold A. Scheraga,^c Jacco van de Streek,^w Tejender S. Thakur,^k Siddharth Tiwari,^k Elisabetta Venuti^j and Iliia K. Zhitkov^u

^aCambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England, ^bImperial College London, England, ^cCornell University, USA, ^dSouth Ural State University, Russian Federation, ^eSSCI, An Aptuit Company, USA, ^fDepartment of Chemistry, University College London, England, ^gThe Pfizer Institute for Pharmaceutical Materials Science, University Chemical Laboratory, University of Cambridge, England, ^hUniversity of Amsterdam, The Netherlands, ⁱDepartment of Chemistry, University of Cambridge, England, ^jUniversità di Bologna, Italy, ^kIndian Institute of Science, India, ^lUniversity of Utrecht, The Netherlands, ^mCenter for High Performance Computing, University of Utah, USA, ⁿDepartment of Biomedical Informatics, University of Utah, USA, ^oUniversidad de Buenos Aires, Argentina, ^pParco Scientifico e Tecnologico, Italy, ^qFlexCryst, Germany, ^rUniversity Erlangen–Nürnberg, Germany, ^sRuhr-Universität Bochum, Germany, ^tUniversity of Bradford, England, ^uVladimir State Humanitarian University, Russian Federation, ^vCavendish Laboratory, England, ^wAvant-garde Materials Simulation, Germany, and ^xDepartment of Physics and Astronomy, University College London, England

Correspondence e-mail: bardwell@ccdc.cam.ac.uk

Following on from the success of the previous crystal structure prediction blind tests (CSP1999, CSP2001, CSP2004 and CSP2007), a fifth such collaborative project (CSP2010) was organized at the Cambridge Crystallographic Data Centre. A range of methodologies was used by the participating groups in order to evaluate the ability of the current computational methods to predict the crystal structures of the six organic molecules chosen as targets for this blind test. The first four targets, two rigid molecules, one semi-flexible molecule and a 1:1 salt, matched the criteria for the targets from CSP2007, while the last two targets belonged to two new challenging categories – a larger, much more flexible molecule and a hydrate with more than one polymorph. Each group submitted three predictions for each target it attempted. There was at least one successful prediction for each target, and two groups were able to successfully predict the structure of the large flexible molecule as their first place submission. The results show that while not as many groups successfully predicted the structures of the three smallest molecules as in CSP2007, there is now evidence that methodologies such as dispersion-corrected density functional theory (DFT-D) are able to reliably do so. The results also highlight the many challenges posed by more complex systems and show that there are still issues to be overcome.

1. Introduction

This paper reports on the results of the fifth blind test of crystal structure prediction (CSP), an international test hosted periodically by the Cambridge Crystallographic Data Centre (CCDC). We refer to this fifth blind test as CSP2010.

Over the last several decades there has been much research in the field of crystal structure prediction. The grand aim is to develop the ability to reliably predict, by computational methods, how a molecule will crystallize in the solid state, with only the chemical diagram and the crystallization conditions known. This would allow for the prediction of solid-state properties before the molecule or molecules in question had even been synthesized, and could also help determine the likelihood that different polymorphic forms, or as yet unseen polymorphs of currently known structures, exist. This application is of particular importance in the pharmaceutical industry where the presence of different polymorphs can lead to very different and potentially undesirable physical properties of new drugs.

For the last decade the CCDC has held periodic blind tests to assess the current reliability and capabilities of the techniques available in the field. Four blind tests, starting in 1999 and every 2 or 3 years thereafter, have previously been held. Each

Received 1 August 2011
Accepted 16 October 2011

has required the identification of a set of molecules with known but previously unpublished crystal structures to use as targets for the participants to predict using the various techniques they have developed. This approach is similar to that adopted to monitor and test advances in other areas of predictive modelling, such as protein structure prediction (Moult *et al.*, 2007). Recently there has also been a blind test for search methods for the crystal structure prediction of purely inorganic systems (Oganov, 2010). Repeating the blind test periodically helps to evaluate advances that have been made in methodologies since the last test, as well as establish the reliability of the techniques which have been successful in previous tests for a given category of target; the small number of targets in any one blind test introduces the possibility of a slightly easier or harder molecule (whose difficulty cannot be easily judged prior to commencement of the test) influencing the results.

This fifth blind test was therefore held to assess the reproducibility of the good results (Neumann *et al.*, 2008; Day *et al.*, 2009) from the previous blind test, CSP2007, and also to assess the developments in methodologies when applied to more challenging targets than the relatively simple rigid molecules mostly studied thus far. These additional targets better represent cases that would be more likely to be encountered in the pharmaceutical industry.

2. Organization and approach

The organization for this latest blind test, CSP2010, was similar to that used for the previous four evaluations of the field, the results of which have been previously published: CSP1999 (Lommerse *et al.*, 2000), CSP2001 (Motherwell *et al.*, 2002), CSP2004 (Day *et al.*, 2005) and CSP2007 (Day *et al.*, 2009). Invitations to participate were sent to 24 research groups known to be active in the field. The test was also advertised through various websites and meetings.

The previous blind test puts forward targets for prediction in the following four categories:

(1) Small, rigid molecules; only the elements C, H, N and O; $Z' = 1$ in any space group; up to 25 atoms.

(2) Rigid molecules; unusual functional groups or elements such as halogens, S, P and B; $Z' = 1$ in any space group; up to 30 atoms.

(3) Moderately flexible molecule with 2–4 internal degrees of freedom; $Z' = 1$ in any space group; up to 40 atoms.

(4) Multiple independent rigid molecules, *e.g.* solvates, co-crystals, salts or $Z' = 2$ structures; any space group; up to 30 atoms.

These four categories were left the same as those used in CSP2007 so as to facilitate comparison of results. In addition, it was decided to add two new categories that would provide greater challenges:

(5) Molecule with 4–8 internal degrees of freedom; $Z' \leq 2$ in any space group; 50–60 atoms.

(6) Molecule for which more than one polymorph is known, and which roughly falls into one of the first four categories.

The new fifth category presents a much greater challenge in terms of flexibility than previously encountered in earlier blind tests, with a large flexible molecule intended to represent those often associated with modern pharmaceuticals. The new sixth category gives an opportunity to study the challenging effects of polymorphism by introducing a molecule for which more than one polymorph is known.

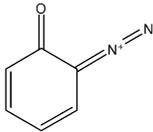
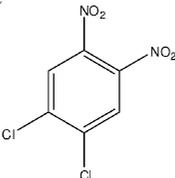
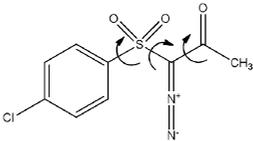
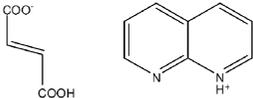
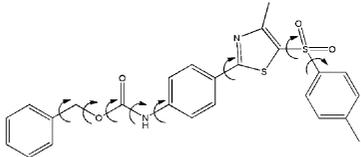
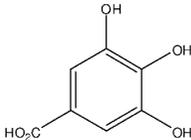
Crystallographers were contacted in August 2009 with a request for unpublished crystal structures that matched one or more of the six categories for the fifth blind test. Crystal structures were collected at the CCDC and assessed for the possibility of inclusion in one of the six possible categories. To be suitable, a crystal structure had to be of high quality and have all atoms located with no disorder. The crystal structure had to be unpublished and the donor crystallographer had to agree to postpone any publication for the duration of the blind test. Collection of suitable candidates for all six categories proved exceptionally difficult, especially for category 1, where the target molecule is very small with a very restricted set of constituent elements, and also for category 6 where few suitable candidates were available that were not of sufficient interest to be withheld from publication for the duration of this test. Almost 30 submitted crystal structures had to be rejected either due to not conforming to any of the six categories, or the presence of refinement issues such as disorder.

After considerable effort, one candidate was collected for category 1, four for category 2, eight for category 3, three for category 4, three for category 5 and one for category 6. For those categories where there was more than one candidate, the final target choice was made randomly.

For category 6, the one candidate that was submitted was gallic acid monohydrate, for which two new polymorphs had been found. These complemented the two previously published polymorphs for gallic acid monohydrate, which are located in the Cambridge Structural Database (CSD; Allen, 2002) under the KONTIQ CSD reference code family. For the purposes of this blind test, these known forms are referred to as forms (1) and (2). Of the two new forms submitted as candidates for prediction, one [form (4), as recently published by Clarke *et al.*, 2011] had one formula unit in the asymmetric unit (*i.e.* one gallic acid and one water molecule). The other, form (3), was originally solved with two formula units per asymmetric unit. However, analysis after the blind test submissions showed that this solution contained a disordered hydrogen-bonding network and the crystal structure could also be described with an ordered hydrogen-bonding network by doubling the unit cell, as now published (Clarke *et al.*, 2011). For the purposes of this blind test, form (3) was therefore deemed inappropriate as a target crystal structure. The main aim for this category, then, was to predict form (4), whose structure has been recently independently published (Demirtaş *et al.*, 2011) and see where (if at all) forms (1) and (2) appeared in the ranked list of predictions.

The molecular diagrams and crystallization conditions were sent by e-mail to 15 participant groups on 16 November 2009. Immediately after circulation of the target crystal structures we were made aware that the crystal structure of the molecule

Table 1
Diagrams and crystallization conditions for the targets of CSP2010.

Target	Crystallization conditions
(XVI)  2-Diazo-3,5-cyclohexadiene-1-one	Slow evaporation from ethanol
(XVII)  1,2-Dichloro-4,5-dinitrobenzene	Slow evaporation from ethanol/ acetone
(XVIII)  (1-((4-Chlorophenyl)sulfonyl)-2-oxopropylidene)diazenium	Slow evaporation from EtOAc
(XIX)  1,8-Naphthyridinium fumarate	Slow evaporation from methanol
(XX)  Benzyl-(4-(4-methyl-5-(<i>p</i> -tolylsulfonyl)-1,3-thiazol-2-yl)phenyl)carbamate	Slow evaporation from EtOAc
(XXI)  Gallic acid monohydrate	Form (3): Slow evaporation from water Form (4): Slow evaporation from methanol in the presence of sarcosine

selected for category 1 (4-ethynylbenzonitrile) had been solved, was undergoing publication and so would soon be in the public domain. The decision was therefore made to remove this candidate for category 1 and attempt to locate a suitable replacement. Thankfully a suitable candidate was quickly provided and the revised list of target molecules, as detailed in Table 1, was distributed to participants on 23 November 2009. Following the numbering used in the

previous blind tests we refer to these molecules by the Roman numerals (XVI)–(XXI).

The format of this blind test was kept broadly the same as the last blind test, with the exception that a greater length of time was allowed before submission of results. Participants were requested to forward their three 'official' predictions for each target molecule to the CCDC, where the experimentally determined crystal structures were held for the duration of the test. As well as these three main predictions, participants were urged to submit an extended list of the crystal structures they generated in order to help post-analysis and to provide insight into the performance of the various methods. The deadline for submissions was 20 August 2010. The experimentally determined crystal structures for all six categories were then circulated to all participants on 23 August 2010 to allow post-analysis of their predictions. Lastly, a workshop was held at the CCDC mid-September 2010 to discuss the results.

We present here results from the 14 participating groups that agreed to publish their results. Details of these 14 participating groups, together with a summary of which targets they attempted and if a match with the experimental structure was observed in their submission, are presented in Table 2(a).

3. Methodologies

Methodologies for the participating research groups vary significantly. A summary of the techniques used by each of the groups is presented in Table 2(b), together with key references for most of the methods used. More detailed descriptions are also provided in the associated supplementary material.¹

In general, each of the methods employed involved three general steps:

- building a three-dimensional molecular structure from the supplied two-dimensional chemical diagrams;
- searching for plausible crystal packing arrangements of the molecule;
- ranking the generated crystal structures in order of likelihood of formation.

3.1. Methods of generating the molecular structure

There are two main approaches that can be used for treating the molecular structure during crystal structure prediction. Firstly, the molecule can be treated as rigid throughout the

¹ Supplementary data for this paper are available from the IUCr electronic archives (Reference: BK5106). Services for accessing these data are described at the back of the journal.

Table 2

(a) List of participants with summary of the targets attempted and rank of any successful predictions.

Entries with a name and no number denote an attempt at prediction without success, a dash indicates no attempt at prediction.

Group	(XVI)	(XVII)	(XVIII)	(XIX)	(XX)	(XXI)†
1	Boerrigter 302	Boerrigter 121	Boerrigter	Boerrigter 12	Boerrigter	Boerrigter
2	Day 8	Day 4	Day	Day, Cruz-Cabeza 27	Day, Cruz-Cabeza 1	Day 12 (61)
3	Della Valle, Venuti	Della Valle, Venuti				
4	Desiraju, Thakur, Tiwari, Pal	Desiraju, Thakur, Tiwari, Pal 65	Desiraju, Thakur, Tiwari, Pal 13	Desiraju, Thakur, Tiwari, Pal	–	Desiraju, Thakur, Tiwari, Pal (11)
5	van Eijck 2	van Eijck 6	van Eijck	van Eijck 2	van Eijck	van Eijck 29 (1)
6	Facelli, Grillo, Ferraro, Orendt 9	Facelli, Grillo, Ferraro, Orendt	Facelli, Grillo, Ferraro, Orendt 4	Facelli, Grillo, Ferraro, Orendt	Facelli, Grillo, Ferraro, Orendt	Facelli, Grillo, Ferraro, Orendt
7	Hofmann, Hofmann, Kuleshova	Hofmann, Hofmann, Kuleshova				
8	Jose	Jose	Jose	–	–	–
9	Maleev, Zhitkov	Maleev, Zhitkov				
10	Misquitta, Pickard, Needs	–	–	–	–	–
11	Neumann, Leusen, Kendrick, van de Streek 1	Neumann, Leusen, Kendrick, van de Streek 1	Neumann, Leusen, Kendrick, van de Streek 1	Neumann, Leusen, Kendrick, van de Streek 3	Neumann, Leusen, Kendrick, van de Streek 7	Neumann, Leusen, Kendrick, van de Streek 89 (174)
12	Nikylov, Bartashevich	–	–	–	–	–
13	Price, Misquitta 15	Price, Habgood 2	Price, Price	Price, Mohamed	Price, Kazantsev, Karamertzanis, Adjiman, Pante-lides 1	Price, Braun 117 (3)
14	Scheraga, Arnautova	Scheraga, Arnautova	Scheraga, Arnautova 29	Scheraga, Arnautova	–	–

† The numbers in parentheses denote matches ignoring H-atom placement, all other values denote exact matches.

(b) Summary of methodologies

Group	Programs	Refs	Search generation	Space groups considered	Molecular model	Lattice energy / fitness function	
						electrostatic	other
1	<i>Materials Studio</i> 5.0	(a)	Monte Carlo simulated annealing	(XVI)–(XIX), (XXI): 88 most common groups (XX): 49 most common space groups with 1 molecule and 33 most common space groups with 2 independent molecules	Flexible throughout	Atomic charges fitted to the electrostatic potential obtained from the DMol3 geometry optimized molecular model (DFT method: GGA-BLYP, double numerical basis set with polarization)†	Dreiding 2.21
2	<i>Crystal Predictor</i> , <i>Cerius2</i> , <i>DMAREL</i> , <i>DMACRY</i>	(b)	Structures generated using a low discrepancy Sobol' sequence	$P1$, $P\bar{1}$, $P2_1$, $P2_1/c$, $P2_12_12_1$, $P2_12_12_2$, $P2_1/m$, $P2/c$, $P2_13$, $P4_1$, $P4/n$, $P4_2/n$, $P4_12_12_2$, $P4_2/c$, $P3_1$, $P3_12_1$, $P\bar{3}$, $P6_1$, $P6_3/m$, $Pna2_1$, $Pca2_1$, $Pbca$, $Pbcn$, $Pmn2_1$, $Pnna$, $Pccn$, $Pbcm$, $Pnnm$, $Pmnm$, $Pnma$, Pc , $Pa3$, $C2/c$, Cc , $C2$, Cm , $C2/m$, $C22_1$, $Cmc2_1$, $Cmcm$, $Cmca$, $R\bar{3}$, $R3$, $R3c$, $R\bar{3}c$, $Aba2$, $Fdd2$, $Fddd$, $Iba2$, $Ibam$, $I\bar{4}$, $I4/m$, $I4_1/a$, $I4_2d$	(XVI) Rigid throughout, (XVII)–(XXI) partly flexible	Atomic multipoles	Specifically fitted anisotropic <i>exp-6</i>
3	<i>Xfind</i> , <i>WMIN</i> , <i>GULP</i> , <i>PLATON</i> , <i>SIESTA</i>	(c)	Structures generated using a low discrepancy Sobol' sequence	(XVI): All 230 space groups (XVII)–(XXI): $P1$, $P\bar{1}$, $P2_1$, $P2_1/c$, $C2/c$, $P2_12_12_1$, $Pna2_1$, $Pbca$, $Pnma$	Rigid for search, flexible for energy minimization	Atomic charges	Empirical <i>exp-6</i>

Table 2 (continued)

Group	Programs	Refs	Search generation	Space groups considered	Molecular model	Lattice energy / fitness function	
						electrostatic	other
4	<i>Materials Studio 5.0</i>	(d)	Monte Carlo simulated annealing	$P1, P\bar{1}, P2_1, C2, Cc, P2_1/c, C2/c, P2_12_12_1, Pca2_1, Pna2_1, Pbcn, Pbca, Pnma$	Rigid for search, flexible for energy minimization	Atomic charges	Dreiding <i>exp</i> -6 [(XVI), (XVIII)] COMPASS force field [(XVII), (XIX), (XXI)]
5	<i>UPACK, XTINKER, GAMESS-UK, MOLDEN</i>	(e)	Randomly generated starting structures	$P1, P2_1/c, P\bar{1}, P2_1, P2_12_12_1, C2/c, Pbca, Pbcn, Pna2_1, Pca2_1, Cc, C2, Pc$	Flexible throughout	Atomic charges [(XIX), (XX)] atomic multipoles (others)	Empirical <i>exp</i> -6 (inter), <i>ab-initio</i> energies (intra)
6	<i>MGAC</i>	(f)	Modified genetic algorithm	$P1, P\bar{1}, P2_1, C2, Pc, Cc, P2_1/c, C2/c, P2_12_12_1, Pca2_1, Pna2_1, Pbcn, Pbca, Pnma$	Flexible throughout	Atomic charges	GAFF 6-12
7	<i>FlexCryst</i>	(g)	Random search with calibrated cell	$P2_1/c, P\bar{1}, C2/c, P2_12_12_1, P2_1$	Rigid throughout	Trained potentials	
8	<i>GA-CG-MTA</i>	(h)	Genetic algorithm	$P1, P\bar{1}, P2_1, C2, Pc, Cc, P2_1/c, P2_12_12_1, Pca2_1, Pna2_1, Pbcn, Pbca, Pnma$	Flexible throughout	EPIC	CG-MTA
9	<i>DMM</i>	(i)	Discrete modeling method of molecular packings	All space groups with $Z \leq 4$	Rigid for search and for energy minimization	Empirical <i>exp</i> -6	
10	<i>CASTEP</i>	(j)	<i>Ab initio</i> random structure searching	All space groups with $Z \leq 4$	Rigid in searches using potential, flexible in DFT-D optimizations and AIRSS searches	Plane-wave density functional theory supplemented with an empirical dispersion correction	
11	<i>GRACE 1.5-1.6 and VASP</i>	(k)	Monte Carlo parallel tempering	All 230 space groups	Flexible throughout	Plane-wave density functional theory supplemented by an empirical C_6R^{-6}	
12	<i>OPIX</i>	(l)	Random search without molecular flexibility	$P2_1/c, P\bar{1}, P2_1, P2_12_12_1, C2/c, Pbca$	Rigid throughout	Empirical <i>exp</i> -6 without atomic charges‡	
13	<i>Crystal Predictor, DMACRYS, CrystalOptimizer</i>	(m)	Sobol sequences, flexible or rigid	(XVI)–(XIX): 59 most common groups; (XX): 12 most common groups; (XXI): 24 most common groups	(XVI) Rigid, (XVII)–(XXI) Flexible	Atomic multipoles	Empirical <i>exp</i> -6 (XVII)–(XXI) non-empirically derived anisotropic <i>exp</i> -6 (XVI)§
14	<i>CRYSTALG, PLATON</i>	(n)	Conformation-Family Monte Carlo	No symmetry information used – $P1$ with varying Z ($= 2, 4, 8$)	(XVI)–(XIX) Rigid throughout	Atomic charges	ECEPP-05 [(XVI), (XVIII), (XIX)] Empirical <i>exp</i> -6 (XVII)

References: (a) Verwer & Leusen (1998), Karfunkel *et al.* (1996); (b) Karamertzanis & Pantelides (2007), Price *et al.* (2010), Day *et al.* (2007), Cooper *et al.* (2008); (c) Busing & Matsui (1984), Gale & Rohl (2003), Soler *et al.* (2002); (d) Sarma & Desiraju (2002), Dey *et al.* (2005, 2006); (e) Mooij *et al.* (2000), van Eijck & Kroon (2000), van Eijck *et al.* (2001), van Eijck (2002); (f) Bazterra *et al.* (2007); (g) Hofmann & Lengauer (1997), Hofmann & Apostolakis (2003), Hofmann & Kuleshova (2005); (h) unpublished method – see supplementary material; (i) Maleev (1995, 2001), Maleev *et al.* (2005, 2009); (j) Clark *et al.* (2005), Pickard & Needs (2006), Pickard & Needs (2011); (k) Neumann & Perrin (2005), Neumann (2008, 2011), Kresse & Furthmüller (1996), Kresse & Hafner (1993), Kresse & Joubert (1999); (l) Gavezzotti (2003), Gavezzotti & Filippini (1997), Gavezzotti (2002); (m) Karamertzanis & Pantelides (2007), Price *et al.* (2010), Kazantsev, Karamertzanis, Adjiman & Pantelides (2011), Kazantsev *et al.* (2010); (n) Pillardy *et al.* (2001). † For (XIX) the anion–cation cluster was used as the molecular model for the DFT calculations and subsequent ESP charges. The molecules were treated independently during the sampling. ‡ Application of topological characteristics of electron density, calculated for generated structures and comparison with the structures from the CCDC with the same functional groups § Choices 2 and 3 considered properties and motif including free energy.

calculations, assuming that the packing forces are too small to significantly distort the molecular geometry. In this case the method used to determine the rigid molecular structure is vitally important, as the effect of the molecular structure on crystal energy calculations can be large (Beyer & Price, 2000).

Alternatively, the structure can be considered as flexible with intramolecular bond stretching, angle bending and torsional terms allowed to vary during the search as well as the final energy minimizations. For extremely flexible molecules such as target (XX) the conformational distributions can be reduced to a more manageable level *via* methods such as

analysis of conformational preferences using software such as *Mogul* (Bruno *et al.*, 2004).

3.2. Generating trial crystal structures

There are many diverse methods for generating crystal packing arrangements in order to achieve a variety of plausible packing arrangements. Most participants in this blind test opted to generate large numbers of crystal structures with random or quasi-random variables such as unit-cell parameters and positions and orientations of the molecules. Several groups also elected to use a low-discrepancy Sobol' sequence (Sobol', 1967; Press *et al.*, 1992). This helps ensure a more uniform and thus efficient sampling and avoids the problems of gaps and clusters that purely random sampling can exhibit. Other groups used Monte Carlo types of search, genetic algorithms, grid-based systematic searches or first-principles *ab initio* random structure searching which allows the possibility of a change in covalent bonding (Pickard & Needs, 2006, 2011).

For the majority of these methods, space-group symmetry is used. These methods search each space group and Z' separately and so in order to help reduce the computing time required, many groups chose to restrict their search to only the

most commonly adopted space groups. This blind test saw two groups electing to search all 230 space groups for some or all of their predictions. Other groups used the alternative approach of generating $P1$ crystal structures with varying numbers of independent molecules (up to 8) in the unit cell. Space-group symmetry was then identified in the resulting crystal structures, after energy minimization, using packages such as *PLATON* (Spek, 2009).

3.3. Ranking of crystal structures

The final ranking of the crystal structures is still almost exclusively based on the calculated lattice energies of the structures generated by the crystal structure search. Often tens, if not hundreds, of possible structures can exist within a few kJ mol^{-1} of the calculated global minimum (Day *et al.*, 2004) and therefore extreme accuracy is needed. One successful approach to generating these lattice energies is the DFT-D method, which can give more accurate lattice energies (Neumann & Perrin, 2005) or re-minimization of the structures with more sophisticated force fields such as distributed multipoles (Stone, 2005) and additional flexibility (Kazantsev, Karamertzanis, Adjiman & Pantelides, 2011; Day & Cooper, 2010; Görbitz *et al.*, 2010). Moreover, additional or alternative

criteria may be used to discriminate between likely and unlikely crystal structures. Such approaches include lattice dynamic contributions (van Eijck, 2001; Anghel *et al.*, 2002) or comparisons to known crystal structures in the CSD (Dey *et al.*, 2006), exploiting any isostructurality relationships (Asmadi *et al.*, 2010a,b).

4. Results

This paper is accompanied by a large amount of supplementary material: the coordinates of the experimental crystal structures, lists of predicted crystal structures by each participant, as well as detailed descriptions of methodology, results and post-analysis by most of the participating research groups. Before discussing the results of the predictions, the crystal packings in the X-ray determined crystal structures of the six categories are described.

4.1. Experimental crystal structures

4.1.1. Molecule (XVI). 2-Diazo-3,5-cyclohexadiene-1-one (C_6H_4

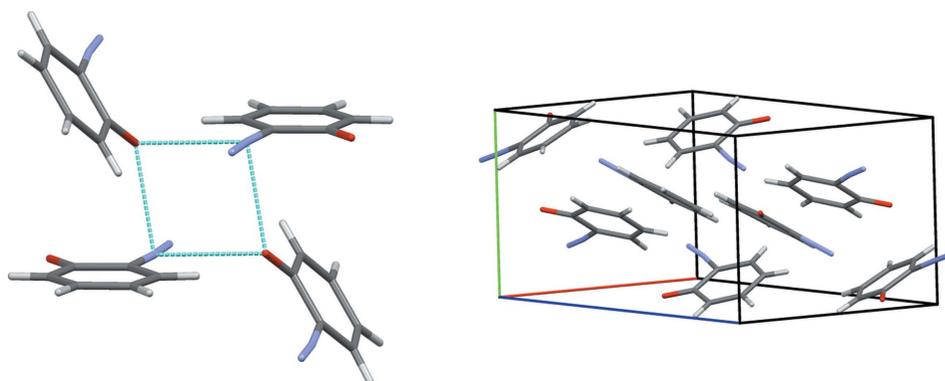


Figure 1
Packing diagram of the crystal structure of molecule (XVI). Grey = carbon, white = hydrogen, red = oxygen and blue = nitrogen. Contacts shorter than the sum of van der Waals radii are shown as blue lines.

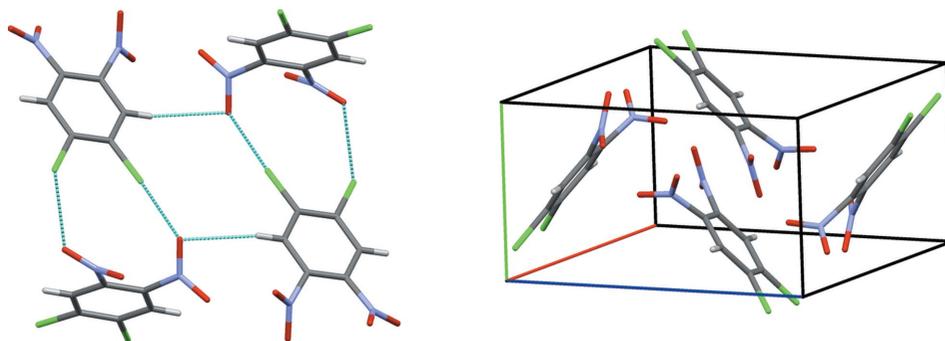


Figure 2
Packing diagram of the crystal structure of molecule (XVII). Grey = carbon, white = hydrogen, red = oxygen, blue = nitrogen and green = chlorine. Contacts shorter than the sum of van der Waals radii are shown as blue lines.

N_2O) was chosen as the blind test target for category 1 after the initial target, 4-ethynylbenzonitrile, was found to have been previously solved. Molecule (XVI) was crystallized by slow evaporation from ethanol and the crystal structure was solved from X-ray diffraction data collected at 174 K (Britton, 2010). The molecule crystallizes with $Z' = 1$ in the orthorhombic space group $Pbca$. The crystal packing shows diazide-carbonyl and $\text{CH}\cdots\text{O}=\text{C}$ interactions (Fig. 1).

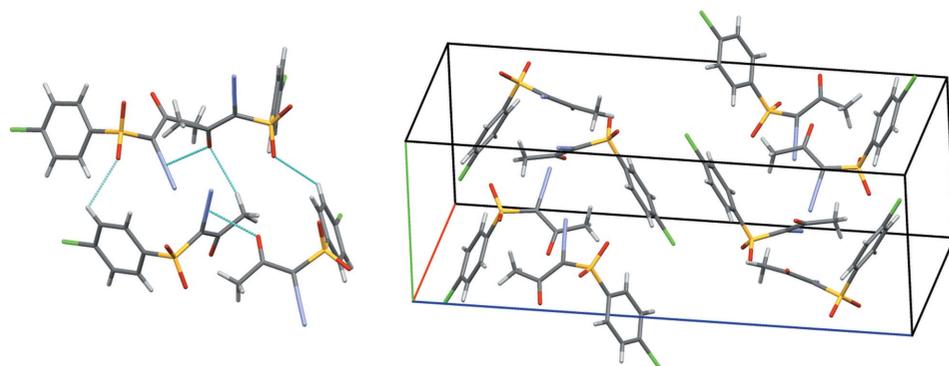


Figure 3
Packing diagram of the crystal structure of molecule (XVIII). Grey = carbon, white = hydrogen, red = oxygen, blue = nitrogen, green = chlorine and yellow = sulfur. Contacts shorter than the sum of van der Waals radii are shown as blue lines.

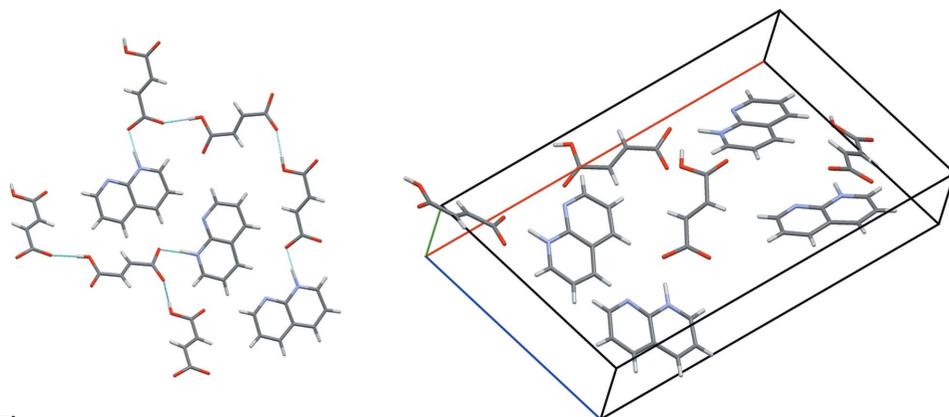


Figure 4
Packing diagram of the crystal structure of molecular salt (XIX). Grey = carbon, white = hydrogen, red = oxygen and blue = nitrogen. Hydrogen bonds are shown as blue lines.

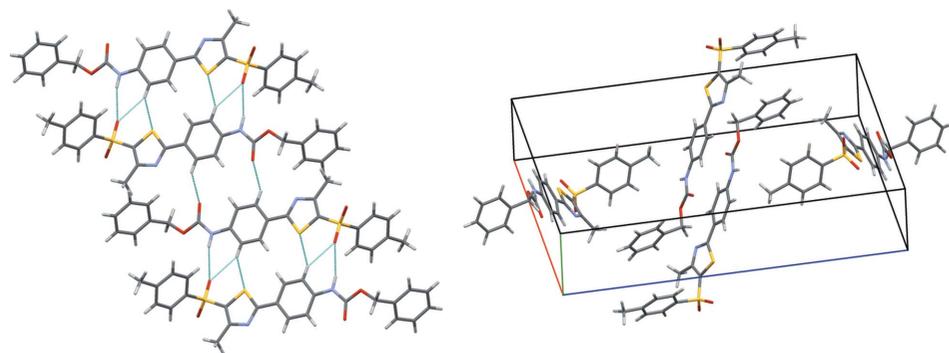


Figure 5
Packing diagram of the crystal structure of molecule (XX). Grey = carbon, white = hydrogen, red = oxygen, blue = nitrogen, green = chlorine and yellow = sulfur. Contacts shorter than the sum of van der Waals radii are shown as blue lines.

4.1.2. Molecule (XVII). 1,2-Dichloro-4,5-dinitrobenzene ($\text{C}_6\text{H}_2\text{Cl}_2\text{N}_2\text{O}_4$) was chosen as the blind test target for category 2, although it deviates somewhat from the criteria for this category as the molecule is not truly rigid; the nitro groups allow for some degree of rotational freedom. Crystals were obtained by slow evaporation of methanol and X-ray diffraction data were collected at 174 K (Britton, 2010). The molecule crystallizes in the monoclinic space group $P2_1/c$ with $Z' = 1$ (Fig. 2).

4.1.3. Molecule (XVIII). (1-((4-Chlorophenyl)sulfonyl)-2-oxo-propylidene)diazenium ($\text{C}_9\text{H}_7\text{ClN}_2\text{O}_3\text{S}$) was the target for category 3. Molecule (XVIII) was crystallized by slow evaporation from ethyl acetate (EtOAc) and the crystal structure was solved from X-ray diffraction data collected at 150 K (Blake, 2010). The crystal structure was solved in the orthorhombic space group $Pbca$ with $Z' = 1$. The conformational flexibility can be described by three exocyclic torsion angles, as shown in Table 1. The CN_2CO moiety adopts a mostly planar *trans* configuration (Fig. 3).

4.1.4. Molecular salt (XIX). 1,8-Naphthyridinium fumarate ($\text{C}_8\text{H}_7\text{N}_2$, $\text{C}_4\text{H}_3\text{O}_4$) was chosen as the target for category 4. This 1:1 salt was formed by slow evaporation from methanol and the crystal structure was solved in the orthorhombic space group $Pca2_1$ from data collected at 200 K (MacGillivray, 2010) with $Z' = 1$. The packing in this crystal structure is dominated by hydrogen bonds, with linear chains of fumarate and naphthylpyridinium ions forming alternating connections to these chains (Fig. 4). The crystal structure is isostructural with the entry RABYID in the CSD (Shan *et al.*, 2003) where quinolinium is substituted for 1,8-naphthyridinium (*i.e.* one nitrogen is replaced by a C—H group).

4.1.5. Molecule (XX). Benzyl-(4-(4-methyl-5-(*p*-tolylsulfonyl)-1,3-thiazol-2-yl)phenyl)carbamate ($\text{C}_{25}\text{H}_{22}\text{N}_2\text{O}_4\text{S}_2$) was chosen as the target for the new category 5. Molecule (XX) was crystallized by slow evaporation from EtOAc and the crystal structure solved in

Table 3

Lattice parameter deviations (predicted – experimental), ΔE and RMSD for the experimental and predicted structures of molecule (XVI).

$\alpha = \beta = \gamma = 90^\circ$ in all structures.

	Rank	ΔE^\dagger (kJ mol ⁻¹)	Density (g cm ⁻³)	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	RMSD ₁₅ ‡ (Å)
Expt. (<i>T</i> = 174 K)	–	–	1.385	9.645 (2)	7.381 (1)	16.185 (3)	–
Predicted amongst first three							
Neumann, Leusen, Kendrick, van de Streek	1	–0.70§	–0.9%	–1.6%	+1.7%	+0.7%	0.157
van Eijck	2	+0.06	–3.7%	+5.3%	–0.6%	–0.8%	0.247
Present in list, outside of first three predictions							
Day	8	+1.16	–4.8%	+6.4%	–1.2%	–0.1%	0.273
Orendt, Grillo, Ferraro, Facelli	9	+2.45	–2.1%	+6.6%	–3.2%	–1.1%	0.306
Price, Misquitta	15	+5.74	–4.9%	+14.5%	–7.1%	–1.1%	0.633
Boerigter	302	+3.38	–5.3%	+4.7%	+0.5%	+0.3%	0.190

† ΔE is calculated with respect to the lowest energy structure predicted by the same research group. ‡ RMSD₁₅ is calculated using a 15 molecule comparison in the Materials Module of *Mercury*, ignoring H atoms. § ΔE for the global minimum is calculated with respect to the second lowest energy structure.

the monoclinic space group $P2_1/n$ with $Z' = 1$ (Blake, 2010). The conformational flexibility can be described with eight exocyclic torsion angles (Table 1). The molecule adopts an elongated S shape, with the central part of the molecule mostly planar, the greatest deviation from planarity being between the phenyl and thiazol groups with an angle of 13° . The mostly planar mid-section of the molecule forms stacks *via* a series of weak interactions with CH and NH...OS as well as CH...OC atom–atom contacts (*i.e.* shorter than the sum of van der Waals radii), as shown in Fig. 5.

4.1.6. Polymorphic hydrate (XXI). Gallic acid monohydrate ($C_7H_6O_5 \cdot H_2O$) was chosen as the target for the new category 6. Gallic acid monohydrate had two previously known forms, (1) (Jiang *et al.*, 2000) and (2) (Okabe *et al.*, 2001). Form (4) of hydrate (XXI) was observed from crystals grown by slow evaporation from methanol in the presence of sarcosine and crystallized in the monoclinic space group $P2_1/c$ with $Z' = 1$ (Clarke *et al.*, 2011). The crystal structure is dominated by an extensive hydrogen-bonding network. Unlike forms (1) and (3), no carboxylic acid dimer units are formed, with forms (2) and (4) instead having hydrogen bonds from the carboxylic acid to both water and adjacent gallic acid molecules (Fig. 6).

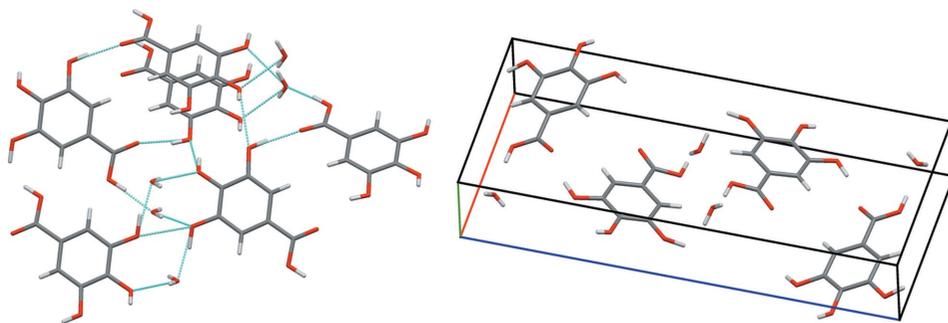


Figure 6

Packing diagram of the crystal structure of hydrate (XXI). Grey = carbon, white = hydrogen and red = oxygen. Hydrogen bonds are shown as blue lines.

4.2. Comparison of the predictions with the experimental crystal structures

The submitted predictions were compared with each experimentally determined crystal structure using the ‘Crystal Structure Similarity’ feature of the Materials Module of *Mercury* (Macrae *et al.*, 2008). The algorithm used by this feature allows comparison of the molecular packing environment between two or more crystal structures. The reference crystal structure, in this case the experimentally determined crystal structure, is analysed and represented by a reference molecule and a coordination shell of its 14 closest neighbours. This set of distances is then searched for in the predicted crystal structures and if they match to within the default geometric tolerances (distances within 20% and angles within 20°) then the coordination shells are overlaid and a root-mean-squared deviation (RMSD₁₅) of the atomic positions is calculated for all matching molecules. As with previous blind tests, this search was configured to ignore H atoms due to the uncertainty of their positions in X-ray determined crystal structures. If all 15 molecules of the reference and predicted crystal structure matched within the standard tolerances, the crystal structure was determined as having been successfully predicted.

For hydrate (XXI) it became apparent that some predictions matched all non-H atoms but not the H-atom positions as located in the target crystal structure. For this molecule we therefore re-ran the crystal structure comparison, but this time elected to include H atoms in the calculation in order to determine if an exact match was present.

Overlays of the X-ray determined crystal structure with some of the predicted structures for targets (XVI)–(XX) can be found in the supplementary material.

Table 4Lattice parameter deviations (predicted – experimental), ΔE and RMSD₁₅ for the experimental and predicted structures of molecule (XVII). $\alpha = \gamma = 90^\circ$ in all structures.

	Rank	ΔE^\dagger (kJ mol ⁻¹)	Density (g cm ⁻³)	a (Å)	b (Å)	c (Å)	β (°)	RMSD ₁₅ ‡ (Å)
Expt. ($T = 174$ K)	–	–	1.837	12.639 (1)	5.979 (1)	11.422 (1)	96.807 (1)	–
Predicted amongst first three								
Neumann, Leusen, Kendrick, van de Streek	1	–1.64§	+0.7%	0.0%	–1.0%	+0.5%	–0.1%	0.045
Price, Habgood	2	+1.05	–0.3%	+0.2%	–2.0%	+1.6%	–0.4%	0.130
Present in list, outside of first three predictions								
Day	4	+3.24	–0.2%	–0.2%	–2.6%	+2.7%	–2.0%	0.191
van Eijck	6	+3.67	–1.5%	+1.0%	–0.8%	+1.2%	–0.4%	0.102
Desiraju, Thakur, Tiwari, Pal	65	+5.00	+5.3%	+1.4%	–2.3%	–4.2%	–0.1%	0.264
Boerigter	121	+6.39	–0.9%	+2.8%	–4.3%	+2.9%	+1.0%	0.270

† ΔE is calculated with respect to the lowest energy structure predicted by the same research group. ‡ RMSD₁₅ is calculated using a 15 molecule comparison in the Materials Module of *Mercury*, ignoring H atoms. § ΔE for the global minimum is calculated with respect to the second lowest energy structure.

Table 5Lattice parameter deviations (predicted – experimental), ΔE and RMSD for the experimental and predicted structures of molecule (XVIII). $\alpha = \beta = \gamma = 90^\circ$ in all structures.

	Rank	ΔE^\dagger (kJ mol ⁻¹)	Density (g cm ⁻³)	a (Å)	b (Å)	c (Å)	RMSD ₁₅ ‡ (Å)
Expt. ($T = 174$ K)	–	–	1.566	9.889 (1)	8.887 (1)	24.969 (3)	–
Predicted amongst first three							
Neumann, Leusen, Kendrick, van de Streek	1	–1.30§	–1.2%	+0.4%	–1.0%	+1.9%	0.122
Present in list, outside of first three predictions							
Orendt, Grillo, Ferraro, Facelli	4	+2.53	+3.9%	+1.0%	–1.5%	–3.3%	0.252
Desiraju, Thakur, Tiwari, Pal	13	+5.92	–7.7%	+4.4%	+0.6%	+0.7%	0.362
Scheraga, Arnautova	29	+8.21	–5.2%	–0.1%	+6.2%	–0.6%	0.390

† ΔE is calculated with respect to the lowest energy structure predicted by the same research group. ‡ RMSD₁₅ is calculated using a 15 molecule comparison in the Materials Module of *Mercury*, ignoring H atoms. § ΔE for the global minimum is calculated with respect to the second lowest energy structure.

4.3. Predictions results

4.3.1. Molecule (XVI). All of the participating research groups attempted predictions for molecule (XVI), two of whom predicted the observed crystal structure within their three predictions (Table 3). One of these successes (Neumann, Leusen, Kendrick and van de Streek) was submitted as the group's first prediction, while the other (van Eijck) was submitted as the participant's second prediction. Both of these successful predictions gave RMSD₁₅ deviations from the experimentally determined crystal structure of less than 0.25 Å.

Outside of the three official predictions, the observed crystal structure was present in the extended lists of five other research groups. The success rates here are comparable to the first three blind tests, while not quite as high as the results observed in the fourth blind test. This may be attributed to some methods having difficulties with many structures close in energy. The very small ΔE in Table 3, even when the observed structure is found outside of the first three predictions, shows how closely spaced the energies are for this molecule, and the accuracy in lattice energy required for a successful prediction.

4.3.2. Molecule (XVII). 13 of the participating research groups attempted predictions for molecule (XVII), two of which predicted the observed crystal structure within their three official predictions (Table 4). As with molecule (XVI), one of these successes (Neumann, Leusen, Kendrick and van de Streek) was submitted as the group's first prediction, while the other (Price and Habgood) was submitted as that group's second prediction. Both of these successful predictions gave RMSD₁₅ values of less than 0.13 Å.

Four other research groups submitted the observed crystal structure in their extended list of solutions, with energies between 3.2 and 6.4 kJ mol⁻¹ above their global minimum. The slightly lower rate of success for this category than for the last blind test may be attributed to the fact that molecule (XVII) is not truly rigid, with flexibility in the nitro groups having to be taken into consideration. Despite these additional challenges, the observed crystal structure was still successfully predicted.

4.3.3. Molecule (XVIII). 13 research groups attempted predictions for the category 3 target, molecule (XVIII), with one group (Neumann, Leusen, Kendrick and van de Streek) successfully predicting the observed crystal structure within their three predictions (Table 5). Once again, this solution was

Table 6

Lattice parameter deviations (predicted – experimental), ΔE and RMSD for the experimental and predicted structures of molecular salt (XIX).

$\alpha = \beta = \gamma = 90^\circ$ in all structures.

	Rank	ΔE^\dagger (kJ mol ⁻¹)	Density (g cm ⁻³)	a (Å)	b (Å)	c (Å)	RMSD ₁₅ ‡ (Å)
Expt. ($T = 200$ K)	–	–	1.481	23.501 (3)	3.714 (1)	12.654 (1)	–
Predicted amongst first three							
van Eijck	2	+0.83	–2.2%	+1.9%	–0.4%	+0.7%	0.220
Neumann, Leusen, Kendrick, van de Streek	3	+6.73	+0.5%	+0.6%	+1.1%	–2.2%	0.151
Present in list, outside of first three predictions							
Boerrigter	12	+2.47	–8.2%	+4.0%	–0.1%	+4.8%	0.367
Day, Cruz-Cabeza	27	+12.62	–1.6%	+4.4%	+2.0%	+1.6%	0.209

† ΔE is calculated with respect to the lowest energy structure predicted by the same research group. ‡ RMSD₁₅ is calculated using a 15 molecule comparison in the Materials Module of *Mercury*, ignoring H atoms.

Table 7

Lattice parameter deviations (predicted – experimental), ΔE and RMSD₁₅ for the experimental and predicted structures of molecule (XX).

$\alpha = \gamma = 90^\circ$ in all structures.

	Rank	ΔE^\dagger (kJ mol ⁻¹)	Density (g cm ⁻³)	a (Å)	b (Å)	c (Å)	β (°)	RMSD ₁₅ ‡ (Å)
Expt. ($T = 150$ K)	–	–	1.411	14.078 (1)	6.356 (1)	25.310 (2)	96.063 (2)	–
Predicted amongst first three								
Day, Cruz-Cabeza	1	–0.53§	–2.4%	+0.3%	–1.8%	+3.9%	–0.4%	0.429
Price, Kazantsev, Karamertzanis, Adjiman, Pantelides	1	–0.78§	–0.6%	+1.3%	–0.6%	+0.2%	+1.3%	0.178
Present in list, outside of first three predictions								
Neumann, Leusen, Kendrick, van de Streek	7	+1.90	+0.1%	+0.6%	–0.9%	+0.2%	–0.7%	0.113

† ΔE is calculated with respect to the lowest energy structure predicted by the same research group. ‡ RMSD₁₅ is calculated using a 15 molecule comparison in the Materials Module of *Mercury*, ignoring H atoms. § ΔE for the global minimum is calculated with respect to the second lowest energy structure.

submitted as this group's first submission, with an RMSD₁₅ from the observed crystal structure of just 0.12 Å.

Three other groups also reported the correct crystal structure in their extended lists of solutions, with one group (Orendt, Grillo, Ferraro and Facelli) close to having a successful prediction as their number 4 structure is a close match to the experimental structure with an RMSD₁₅ value of 0.252 Å.

4.3.4. Molecular salt (XIX). 11 participants attempted predictions for the molecular salt (XIX) and two of these predicted the observed crystal structure within the three official predictions (Table 6): van Eijck as the second prediction and Neumann, Leusen, Kendrick and van de Streek as the third prediction, with RMSD₁₅ values of 0.15 and 0.22 Å. Two other participants located the crystal structure within their extended lists of submissions.

The rate of success in searching for structures with two independent molecules in the asymmetric unit is broadly comparable with that of the last blind test. However, the energetic ranking of the salt structures provided a greater challenge than was experienced with the cocrystal used in 2007. The most successful prediction relied on the use of a supramolecular dimer owing to difficulties with modelling individual ions. Comparison with predictions and the known

crystal structure of the similar compound present in CSD entry RABYID also helped to weight some predictions, including the third placed submission made by Neumann, Leusen, Kendrick and van de Streek, which would have been ranked at position 20 by energy alone.

4.3.5. Molecule (XX). Ten participants attempted predictions for molecule (XX) and two of these predicted the observed crystal structure as their top submission (Day and Cruz-Cabeza; Price, Kazantsev, Karamertzanis, Adjiman and Pantelides). One other group (Neumann, Leusen, Kendrick and van de Streek) also located the observed crystal structure in its extended list of solutions (Table 7) at rank 7.

This category was introduced in this blind test as a new challenge and so there are no results from any previous blind tests with which to compare. However, this does appear to be the first case of a molecule of this complexity having been successfully predicted under blind test conditions and then detailed in a refereed publication. The key dependence was on the conformation of the molecule and with eight internal degrees of freedom the problem became one of completeness of the search. One team resolved this by taking into account CSD observations for each of the flexible components to reduce the search to a more manageable size.

Table 8

(a) Lattice parameter deviations (predicted – experimental), ΔE and RMSD₁₅ for the experimental and predicted structures of hydrate (XXI) (with matching hydrogen placement).

$\alpha = \gamma = 90^\circ$ in all structures.

	Rank	ΔE^\dagger (kJ mol ⁻¹)	Density (g cm ⁻³)	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	β (°)	RMSD ₁₅ ‡ (Å)
Expt. (<i>T</i> = 150 K)	–	–	1.639	9.790 (7)	3.609 (3)	21.583 (16)	91.462 (14)	–
Present in list, outside of first three predictions								
Day	12	+2.96	+2.4%	–2.9%	–0.4%	+1.1%	+0.3%	0.159
van Eijck	29	+12.47	+10.1%	–4.9%	–2.4%	–2.5%	+0.8%	0.208
Neumann, Leusen, Kendrick, van de Streek	81	+8.85	+2.0%	–2.8%	+1.2%	–0.4%	–0.5%	0.228
Price, Braun	117	+12.69	+3.7%	–1.5%	–1.9%	–0.1%	+0.8%	0.108

† ΔE is calculated with respect to the lowest energy structure predicted by the same research group. ‡ RMSD₁₅ is calculated using a 15 molecule comparison in the Materials Module of *Mercury*, ignoring H atoms.

(b) Lattice parameter deviations (predicted – experimental), ΔE and RMSD₁₅ for the predicted structures of hydrate (XXI) with alternative H-atom placement to the experimental structure. $\alpha = \gamma = 90^\circ$ in all structures.

	Rank	ΔE^\dagger (kJ mol ⁻¹)	Density (g cm ⁻³)	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	β (°)	RMSD ₁₅ ‡ (Å)
Expt. (<i>T</i> = 150 K)	–	–	1.639	9.790 (7)	3.609 (3)	21.583 (16)	91.462 (14)	–
Predicted amongst first three								
Van Eijck	1	–2.43§	+13.1%	–5.4%	–3.8%	–2.8%	+1.2%	0.232
Price, Braun	3	+1.08	+5.8%	–1.0%	–4.0%	–0.6%	+0.1%	0.224
Present in list, outside of first three predictions								
Desiraju, Thakur, Tiwari, Pal	11	+0.19	+6.2%	+0.6%	–3.3%	–3.2%	–0.4%	0.642
Day	61	+6.96	+4.5%	–1.0%	–1.4%	–2.1%	–1.6%	0.218
Neumann, Leusen, Kendrick, van de Streek	174	+11.30	+2.9%	–2.7%	–0.2%	+0.1%	–0.4%	0.192

† ΔE is calculated with respect to the lowest energy structure predicted by the same research group. ‡ RMSD₁₅ is calculated using a 15 molecule comparison in the Materials Module of *Mercury*, ignoring H atoms. § ΔE for the global minimum is calculated with respect to the second lowest energy structure.

(c) Lattice parameter deviations (predicted – experimental), ΔE and RMSD₁₅ for the experimental and predicted structures of KONTIQ [form (1)]. $\alpha = \gamma = 90^\circ$ in all structures.

	Rank	ΔE^\dagger (kJ mol ⁻¹)	Density (g cm ⁻³)	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	β (°)	RMSD ₁₅ ‡ (Å)
Expt. (<i>T</i> = 150 K)	–	–	1.599	5.794 (4)	4.719 (5)	28.688 (5)	95.080 (30)	–
Present in list								
Neumann, Leusen, Kendrick, van de Streek	19	+0.22	+3.6%	+1.3%	–5.4%	+0.8%	+0.6%	0.211
Day	48	+6.61	+3.1%	+4.3%	–6.3%	+0.2%	–0.6%	0.295
van Eijck	74	+18.80	+10.8%	+6.1%	–17.0%	+2.2%	–5.0%	0.690
Desiraju, Thakur, Tiwari, Pal	143	+12.11	+7.6%	–4.4%	–3.3%	+1.1%	+2.9%	0.442
Boerrigter	282	+13.18	–2.6%	–8.42%	+10.2%	+2.8%	+4.7%	0.631
Price, Braun	338	+11.87	+1.9%	+11.0%	–14.6%	+5.0%	+6.8%	0.683

† ΔE is calculated with respect to the lowest energy structure predicted by the same research group. ‡ RMSD₁₅ is calculated using a 15 molecule comparison in the Materials Module of *Mercury*, ignoring H atoms.

(d) Lattice parameter deviations (predicted – experimental), ΔE and RMSD₁₅ for the experimental and predicted structures of KONTIQ01 [form (2)]. $\alpha = \gamma = 90^\circ$ in all structures.

	Rank	ΔE^\dagger (kJ mol ⁻¹)	Density (g cm ⁻³)	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	β (°)	RMSD ₁₅ ‡ (Å)
Expt. (<i>T</i> = 150 K)	–	–	1.636	14.150 (10)	3.622 (9)	15.028 (10)	97.520 (70)	–
Present in list								
van Eijck	9	+8.44	+11.0%	–2.2%	–2.7%	–5.2%	+0.5%	0.206
Price, Braun	23	+4.83	+3.9%	+0.7%	–3.6%	–0.6%	+0.5%	0.186
Neumann, Leusen, Kendrick, van de Streek	49	+0.34	+2.0%	–1.0%	–0.1%	–2.2%	+0.1%	0.090
Day	53	+6.68	+4.7%	–1.4%	–0.6%	–2.1%	+1.7%	0.135
Desiraju, Thakur, Tiwari, Pal	126	+10.93	+4.9%	1.4%	–2.1%	–0.7%	+2.2%	0.228

† ΔE is calculated with respect to the lowest energy structure predicted by the same research group. ‡ RMSD₁₅ is calculated using a 15 molecule comparison in the Materials Module of *Mercury*, ignoring H atoms.

4.3.6. Polymorphic hydrate (XXI). Ten participants attempted predictions for the hydrate (XXI). This category featured the opportunity to find and locate both an unknown polymorph and two polymorphs whose crystal structures had previously been determined. During analysis of the results it became apparent that there is an alternative proton arrangement in the hydrogen-bonding network of form (4) involving the central OH moiety of the acid and the water molecules (see Fig. 7). Solutions with both proton conformations were generated by some groups, but no agreement was observed in which form had the lower energy.

In previous blind tests, H-atom placement has been ignored in determining if a participant's entry matches the target crystal structure, but in this case it was evident that the two groups that submitted a match within their top three submissions (Price and Braun; van Eijck) did so with the *p*-hydroxy conformation of form (4)_{alt}, not that of the target crystal structure form (4)_{expt} (Fig. 8). As the *p*-hydroxy gallic acid proton shows enlarged displacement parameters, it could be argued that some disorder is present in the structure.

Given this, we present here results for both exact matches including H-atom placement (Table 8*a*) and matches for non-H atoms only (Table 8*b*). No groups submitted an exact match in their top three solutions. Four groups (Day; van Eijck; Neumann *et al.*; Price and Braun) had exact matches within

their extended lists of submissions. For matches involving only the non-H atoms, two groups located the target crystal structure within their top three solutions (van Eijck; Price and Braun) as their first and third submissions respectively. Both of these groups also located the exact match, but at significantly higher energies of approximately 12 kJ mol⁻¹ above their global minimum. Three other groups (Desiraju *et al.*; Day; Neumann *et al.*) also located this crystal structure in their extended lists of submissions.

Tables 8(*c*) and (*d*) show successful matches for the existing polymorphs [forms (1) and (2) in this test]. Six groups located form (1) in their extended lists of submissions, and five groups located form (2). These were generally predicted at high relative energies and rankings, and with no consistency between groups on the stability order between form (1) and (2). This highlights problems in modelling the stability of hydrates.

4.4. Computational expense

Table 9 summarizes the approximate computational resources used by some of the participants. Of particular note is the disparity between some of the groups; the range of computational expense seen in CSP2010 varies from a few thousand CPU hours to almost 200 000 CPU hours (which translates to over 22 CPU years). Clearly the resources required for this blind test have increased. A large portion of the total CPU time was devoted to targets (XX) and (XXI), and is therefore clearly dependent upon the complexity of the molecule. Fortunately, the computer systems required to meet this increased need are also now more readily accessible, as shown by several groups reporting increases of computing resource of over an order of magnitude (and sometimes almost two orders of magnitude) over the resources used for their CSP2007 submissions. As computers get progressively faster and with greater numbers of computing cores per processor, the real time required for these computations is decreasing. This makes modern computers more viable for fast prediction of the simpler targets.

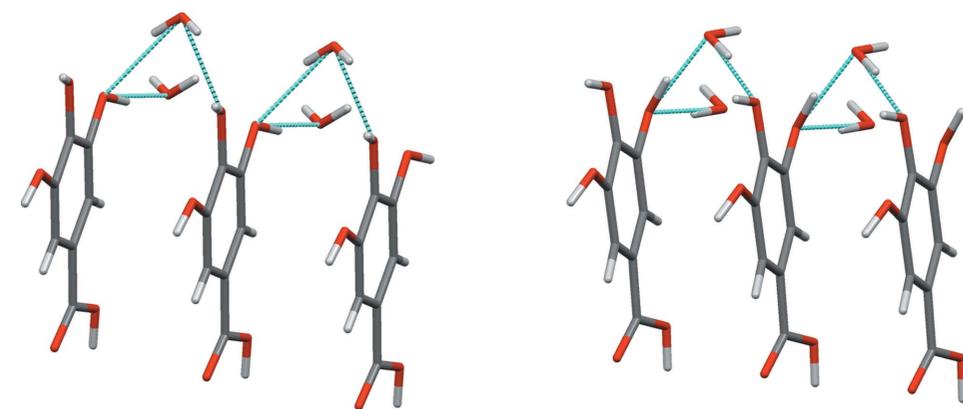


Figure 7
Alternative hydrogen-bond networks possible in (XXI). The left image shows the hydrogen bonds as defined in the crystal structure [form (4)_{expt}], the right image shows the alternative network as located by some participants [form (4)_{alt}]. Hydrogen bonds are shown as blue lines.

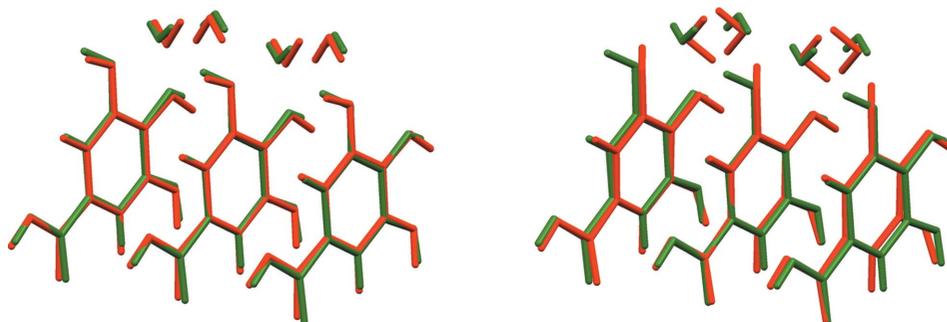


Figure 8
Overlay of the unit-cell contents of the observed crystal structure (XXI) (green) and Day *et al.* (XXI).12 (red, left image), RMSD 0.159 Å, and van Eijck (XXI).1 (red, right image), RMSD 0.219 Å

5. Discussion

5.1. Overall success rates

The success rate for previous blind tests has shown a fluctuating, but generally upward trend, with

Table 9

Summary of computational resources used by some of the participants in CSP2010.

Group	Comments on computing time used	Total computational cost, approximately normalized to 3.0 GHz CPU hours †
Boerrigter	All calculations were performed on an Intel core i7-950 (3.07 GHz) (single core). Approximate execution times: (XVI): 90 h (XVII): 100 h (XVIII): 350 h (XIX): 650 h (XX): 2105 h (XXI): 600 h	~ 3800 CPU hours
Day, Cruz-Cabeza	Most calculations were performed on AMD Opteron 280, 2.6 GHz processors, although parts of the calculations were performed on CPUs with lower performance. (XVI): 110 h (XVII): 1941 h (XVIII): 21 051 h (XIX): 6097 h (XX): 54 090 h (XXI) 22 197 h	~ 91 400 CPU hours
Desiraju, Thakur, Tiwari, Pal	Calculations were performed on four 3.2 GHz processors. (XVI): 114 h (XVII): 2303 h (XVIII): 324 h (XIX): 114 h (XXI): 1431 h	~ 4600 CPU hours
Hofmann	Calculations were performed on 3.0 GHz processors. (XVI): 2 h (XVII): 7 h (XVIII): 12 h (XIX): 694 h (XX): 670 h (XXI): 187 h‡	~ 1600 CPU hours
Neumann, Leusen, Kendrick, van de Streek	Approximately 122 000 CPU hours on 2.8 GHz processors, mostly spent on the generation of reference data for force field parameterization and the final energy ranking with the hybrid method	~ 115 000 CPU hours
Price <i>et al.</i>	(XVI): 200 h (XVII): 5000 h (XVIII): 14 000 h (XIX): 3000 h (XX): 120 000 h (XXI): 52 800 h	~ 195 000 CPU hours
Van Eijck	Calculations performed on 2.66 GHz processors. Molecular calculations: 27 h Structure generation: 4910 h energy minimization: 4526 h	~ 9500 CPU hours
Della Valle, Venuti	Approximately 4400 h on 2.2 GHz processors. Initial rigid-molecule optimizations, DFT calculations, potential fitting and final flexible-molecule optimizations consumed 40, 11, 26 and 22% of the time.	~ 3200 CPU hours
Maleev, Zhitkov	Crystal structure search + energy minimization 2 intel® core™ i5Cpu 750 at 2.67 GHz processors and each has 2 GB of memory	~ 7500 CPU hours
Misquitta, Pickard & Needs	AIRSS/DFT-D search: 130 000 core hours First search using structures obtained with FIT+Q potential: 30 000 core hours Post-Blind test analysis: < 2000 core hours	~ 162 000 CPU hours
Scheraga, Arnautova	Calculations were carried out on Intel Xeon 2.4 GHz processors (XVI): 150 h (XVII): 150 h (XVIII): 610 h (XIX): 720 h	~ 1300 CPU hours

† Note that the large difference in total CPU time presented here is in part due to the various participants electing to predict differing numbers of target molecules and also represents the use of large parallelized computing arrays. ‡ For compound (XXI) the calculation was interrupted after 187 h due to not being able to estimate the convergence in the energetic landscape by that time.

particular success shown in the fourth blind test (Day *et al.*, 2009). This fifth blind test was designed to see if the successes of the fourth test could be repeated, and also to provide more challenging targets to try to stretch the techniques that have thus far been developed. This test therefore saw the introduction of more flexible molecules, as well as hydrates and salts, significantly increasing the complexity of the challenge.

Success for these tests is a combination of two factors: Firstly the ability to generate all possible crystal structures, and secondly the ability to evaluate and rank those crystal structures. The search performance can be impacted by methods that are presently unable to search for crystal structures in space groups with higher values of *Z*, or simply through a lack of computing time and resources. This will lead to an incomplete search space, which may cause the correct solution to be missed entirely. For flexible molecules the conformation of the molecule is also of great importance. Failure to use the correct conformation or to allow for flexibility during the search will lead to failure to predict the correct crystal structure, and this problem becomes greater the more flexible the target molecule. Lastly, the crystal structures generated must be ranked, which is often complicated by the fact that most molecules tend to have many distinct crystal packing possibilities within a small energy range (Day *et al.*, 2004), so that the energy differences between crystal structures are generally very small. The identification and use of accurate energy models can often prove to be the most challenging aspect of successful crystal structure prediction. Ranking is further complicated by thermodynamic kinetic aspects, *i.e.* energies alone may not be sufficient; entropies and nucleation kinetics could also be relevant.

Of the groups that participated in the fifth blind test, most attempted solutions for the four targets [(XVI), (XVII), (XVIII) and (XIX)] that matched the criteria of the previous

blind test. Overall, the success rates for these four targets were a little lower than for CSP2007, but generally at least as good if not better than the results obtained for CSP1999, CSP2001 and CSP2004. What these results do show, however, is that just as in CSP2007, the method adopted by Neumann, Leusen, Kendrick and van de Streek again excelled, with this group able to successfully predict the crystal structures of the first three categories with their number 1 submission, as well as the fourth category with their number 3 submission. They were the only participants able to generate all target crystal structures within their extended list of submissions. They did so with the lowest RMSD₁₅ values for all except the hydrate crystal structure. This demonstrates the reliability of DFT-D methods to predict the crystal structures of small organic molecules (Asmadi *et al.*, 2009; Chan *et al.*, 2011). For the fourth category, complete crystal-structure prediction studies were performed for (XIX) and for model compound RABYID from the CSD. The energy landscapes of these two systems were analysed and showed significant similarities. Based on these similarities, it had to be concluded that the experimental structure of (XIX) could be isostructural to the experimental structure of RABYID, and this structure, even though it was ranked 20th by energy (22nd for RABYID), was submitted as the third candidate structure (Kendrick *et al.*, 2011).

More complex systems such as salts continue to provide some challenge, perhaps suggesting that a salt should be considered a new, more challenging category than the current 'cocrystal' definition of category four.

This test also introduced two new categories that provided much greater challenges to the participants and 11 out of the participating groups attempted at least one of the targets (XX) and (XXI). Particularly encouraging was that two groups (Price *et al.*; Day *et al.*) successfully predicted the crystal structure for the largest, most flexible molecule to be included in this series of blind tests. The hydrate target (XXI) proved to be a considerable challenge, even to methods that have been successful for hydrates of *o*-dihydroxybenzoic acids (Braun *et al.*, 2011), and highlights the many difficulties that such a system can pose to characterization as well as successful structure prediction. However, this is a system that needs to be tackled; water is one of the most complex solvents to model, yet it is also one of the most important.

This blind test has also once again highlighted that the use of generic standard force fields does not lead to good crystal structure prediction results. We have also observed that the more extensive search methods are adequate within the limitations (*Z'*, no disorder *etc.*) implicit in the blind test categories, but have to assume the covalent bonding in the chemical diagram and rely on a sufficient number of search structures being refined by the more accurate and expensive model for the lattice energy. Successful prediction of small molecule crystal structures has been shown to require both accuracy of energies and the ability to coordinate the inter- and intramolecular force field contributions. The methods that gave the greatest success were varied and were modified to take on these tougher challenges.

5.2. Challenges faced

Molecule (XVI), while the simplest of the rigid molecule targets, proved to have many crystal structures close in energy. Transferable empirical potentials had difficulty coping with diazide–carbonyl interactions with induction being a problem. Simple point-charge models used by some groups failed completely for this molecule, although van Eijck did find in post-analysis that one set of charges predicted the observed structure.

This system is the first to be tackled by an *ab initio* random search method (Misquitta, Pickard and Needs) which does not fix the chemical bonding and uses electronic structure methods during the search, although this approach results in a significant increase in the computing resources required when compared with the methods employed by the other participating groups. This method failed as the search was not extended to eight formula units in the cell. Many of the numerous minima, including the global minimum, corresponded to an isomer with the formation of a bond to give a heterocyclic ring with the two N atoms, showing the promise of this method for cases, such as tautomers, where the covalent bonding is uncertain.

Molecule (XVII) was perhaps not well selected as a target for its category as the molecule was not truly rigid; the orientation of the nitro groups may be affected by intermolecular interactions in the crystal structure. As a result participants were forced to first consider how to deal with this flexibility. The electrostatic potentials of the nitro groups also proved to be unusually challenging to model successfully, although the dispersion proved to be a very important contribution to the lattice energy. These additional issues lead to molecule (XVII) being a significantly more difficult problem than previous targets in this category. Despite these extra challenges, the success rate for this category was good compared with previous blind tests.

For molecule (XVIII) flexibility proved to be the key to successfully locating the crystal structure in the search. Some searches missed the crystal structure, with the most fundamental reason being the wrong conformation of the C(N₂)C(O) bond. The relative energies of the *cis* and *trans* configurations were sufficiently sensitive to the methods being used to cause mis-assignment.

For the salt (XIX), there were again some difficulties with flexibility with the relative orientation of the two fragments; for the acid there is considerable conformational flexibility and the calculated stability of the conformers alters between the gas and solid.

All groups encountered significant problems with developing suitable methods of evaluating the relative lattice energies of structures containing the different conformers. Plane-wave *ab initio* methods do not cope well with isolated ions in a vacuum, causing problems with ion-specific reference data calculated with DFT-D methods for force-field parameterization. Induction and charge transfer, which are stronger in molecular salts, limited the transferability of *exp-6* potentials which had been fitted to crystal structures of neutral

molecules. Successful prediction based on energy (van Eijck) was achieved by the use of a supramolecular dimer, rather than two ions as individual molecules.

Other groups noted the similarity of an existing crystal structure in the CSD, RABYID, the crystal structure of which consisted of the same fumarate ion but with a quinolinium counterion instead of 1,8-naphthyridinium (*i.e.* where the unprotonated nitrogen is instead a CH moiety). The energy landscapes generated for both salts proved similar enough to encourage the speculation that the two crystal structures could be isostructural and one group (Neumann *et al.*) submitted a successful prediction based on this approach (Kendrick *et al.*, 2011).

Molecule (XX) was the first large flexible molecule to feature in the blind tests and proved a considerable challenge. For such highly flexible molecules there is a key dependence on the conformation of the molecule and successful prediction involved succeeding at this early step. One of the main difficulties is the computing power required to make a complete search for all available space groups with a flexible molecule; when all standard orientations about the exocyclic single bonds are considered, there are over one thousand possible conformations. The two successful strategies (Day and Cruz-Cabeza; Price, Kazantsev, Karamertzanis, Adjiman and Pantelides) reduced the search space to a more manageable level, producing innovations in methodology that have been described and contrasted in detail elsewhere (Kazantsev, Karamertzanis, Adjiman, Pantelides, Price, Galek, Day & Cruz-Cabeza, 2011). Day *et al.* used geometry data for similar systems from the CSD to help limit the search further and considered a set of predefined conformations. Price *et al.* identified likely ranges of values for the flexible torsions and used an extension to the *CrystalPredictor* methodology and databases of the *ab initio* calculations on the isolated molecule to allow the crystal structures and conformations to be simultaneously refined (Kazantsev, Karamertzanis, Adjiman & Pantelides, 2011). Neumann *et al.* employed a fully flexible molecule, allowing all conformations to be explored during the crystal structure generation step. Use of multipoles and empirical potentials performed better than DFT-D in this case, with both groups using this method (Day *et al.*; Price *et al.*) successfully predicting the crystal structure in first place.

Hydrate (XXI) proved to be one of the most challenging systems in the blind test. For this molecule two known polymorphs already existed. However, the difficulty in predicting this crystal structure was not due to the availability of two already known polymorphs, but rather that the representation of water–water and water–gallic acid interactions is extremely difficult to model, making the successful prediction of even the known polymorphs a difficult task.

As a hydrate, the hydrogen-bonding network enabled by the water molecules and the various hydrogen-bond donors and acceptors in the acid proved key to successfully predicting the crystal structure, but it is also obvious that the sheer number of different possible hydrogen-bond networks make the problem a difficult one. The results obtained for form (4)

show that with the same placement of non-H atoms there is more than one set of hydrogen positions that is possible. Energetically, the OH conformation observed in the experimental structure is not the most favourable in isolation and, given the nature of X-ray diffraction, the positions of these protons cannot be deemed as unequivocally determined. Indeed, there is evidence of large displacement parameters for the protons involved in the two alternative hydrogen-bond networks. This leads us to consider that the structure is best described as disordered with respect to which network is present. This matter would only be resolved with an in-depth temperature-dependent X-ray and NMR study. A post-blind test polymorphism screen (Braun, Personal communication) showed that the ordered form (2) structure is the most stable polymorph at room temperature.

Overall, the systems that gave the most difficulty are those where the molecules can adopt very different low-energy conformations, where current methods may not accurately reflect the energy differences between the conformations in the solid state. Work on improving the estimates of polymorphic energy differences in challenging cases where the polymorphs have different numbers of inter- and intramolecular hydrogen bonds (Karamertzanis *et al.*, 2008) shows that improving the theoretical basis of the methods used to evaluate the lattice energies will lead to further progress.

6. Conclusion

This fifth blind test has built upon the successes of previous blind tests and shows that a state-of-the-art method for crystal structure prediction is able to reliably predict crystal structures of small rigid and slightly flexible molecules, and methods are emerging that are able to tackle larger more flexible molecules and complex systems such as salts and hydrates.

For each of the six target crystal structures, there was at least one successful prediction under the criteria stated for success at the start of the test [although for the hydrate (XXI) certain protons were incorrectly placed]. The number of successful predictions for each of the first four categories was broadly comparable with the first three blind tests, but slightly less than the great successes observed with CSP2007. This may be due to the difficulty of easily gauging a target's difficulty based on its molecular diagram alone – several of the targets in this blind test showed additional challenges not faced in CSP2007 even though the target molecule met the same selection criteria. One observation that is easily made, however, is that the DFT-D method continues to perform very well for these molecule types, although it does not yet supply a comprehensive solution, as observed by the inability to predict some targets [such as (XIX)] by energy methods alone. Most other successes were based on using realistic models for the intermolecular forces (Stone, 1996), which included a distributed multipole representation of the molecular charge distribution.

For the large flexible molecule [target (XX)] it is promising that two groups were able to successfully predict the crystal

structure as their first place entry. In both cases success was achieved by systematic reduction of the problem to more manageable proportions, such as through the use of CSD geometry data to determine the more likely conformations in the experimental crystal structure. More methodological and program development should allow the most thermodynamically stable crystal structures to be computed more readily for molecules of this complexity in the future. The best approach for such complex systems may well be the use of experimental data, including polymorph screening, alongside the calculations to move towards a predictive technology for the understanding and anticipation of polymorphism.

The difficulties faced in this blind test have helped push the participating teams to adopt novel approaches in an attempt to successfully predict the experimental crystal structures. While some challenges remain, such as the need to include a direct consideration of temperature (thermodynamics prescribes that relative stability is a function of temperature), the results achieved in this blind test demonstrate that crystal structure prediction can now be performed reliably for small molecules using a state-of-the-art method. Furthermore, results on the large molecule [target (XX)] as well as the salt [target (XIX)] and the hydrate [target (XXI)] provide encouragement that crystal structure prediction can move on from prediction of small rigid molecules to more complex systems, while highlighting deficiencies in current methods where key developments are still required.

We are grateful to the crystallographers who supplied candidate structures: Professor Doyle Britton [molecules (XVI) and (XVII)], Professor Alexander Blake [molecules (XVIII) and (XX)], Professor Leonard MacGillivray [salt (XIX)] and Professor Michael Zaworotko [molecule (XXI)]. SLP group developments were funded by CPOSS Basic Technology EP/F03573X, <http://www.cposs.org.uk>. DEB was funded by the Austrian Science fund (FWF) J2897-N17. The Imperial College team (AVK, PGK, CSA and CCP) gratefully acknowledge the Engineering and Physical Sciences Research Council (EPSRC) under the Molecular Systems Engineering grant (EP/E016340) for financial support and the High Performance Computing Cluster at Imperial College London for providing resources for calculations. GMD thanks the Royal Society for funding. AJCC thanks the Pfizer Institute for Pharmaceutical Materials Science for funding. ST, RP and TST thank the Indian Institute of Science for fellowships. GRD thanks the DST for the award of a JC Bose fellowship. BvE thanks Paul Ruttink. The University of Utah team gratefully acknowledges an allocation of computer time from the Center for High Performance Computing at the University of Utah. RJN and AJM would like to acknowledge EPSRC grant EP/F032773/1 for funding.

References

Allen, F. H. (2002). *Acta Cryst.* **B58**, 380–388.
 Anghel, A. T., Day, G. M. & Price, S. L. (2002). *CrystEngComm*, **4**, 348–355.

Asmadi, A., Kendrick, J. & Leusen, F. J. J. (2010a). *Chem. Eur. J.* **16**, 12701–12709.
 Asmadi, A., Kendrick, J. & Leusen, F. J. J. (2010b). *Phys. Chem. Chem. Phys.* **12**, 8571–8579.
 Asmadi, A., Neumann, M. A., Kendrick, J., Girard, P., Perrin, M. A. & Leusen, F. J. J. (2009). *J. Phys. Chem. B*, **113**, 16303–16313.
 Bazterra, V. E., Thorley, M., Ferraro, M. B. & Facelli, J. C. (2007). *J. Chem. Theory Comput.* **3**, 201–209.
 Beyer, T. & Price, S. L. (2000). *CrystEngComm*, **2**, 183–190.
 Blake, A. (2010). Personal communication.
 Braun, D. E., Karamertzanis, P. G. & Price, S. L. (2011). *Chem. Commun.* **47**, 5443–5445.
 Britton, D. (2010). Personal communication.
 Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E. & Orpen, A. G. (2004). *J. Chem. Inf. Comput. Sci.* **44**, 2133–2144.
 Busing, W. R. & Matsui, M. (1984). *Acta Cryst.* **A40**, 532–538.
 Chan, H. C. S., Kendrick, J. & Leusen, F. J. J. (2011). *Angew. Chem. Int. Ed.* **50**, 2979–2981.
 Clark, S. J., Segall, M. D., Pickard, C. J., Hasnip, P. J., Probert, M. J., Refson, K. & Payne, M. C. (2005). *Z. Kristallogr.* **220**, 567.
 Clarke, H. D., Arora, K. K., Wojtas, L. & Zaworotko, M. J. (2011). *Cryst. Growth Des.* **11**, 964–966.
 Cooper, T. G., Hejczyk, K. E., Jones, W. & Day, G. M. (2008). *J. Chem. Theory Comput.* **4**, 1795–1805.
 Day, G. M., Chisholm, J., Shan, N., Motherwell, W. D. S. & Jones, W. (2004). *Cryst. Growth Des.* **4**, 1327–1340.
 Day, G. M. & Cooper, T. G. (2010). *CrystEngComm*, **12**, 2443–2453.
 Day, G. M. *et al.* (2005). *Acta Cryst.* **B61**, 511–527.
 Day, G. M. *et al.* (2009). *Acta Cryst.* **B65**, 107–125.
 Day, G. M., Motherwell, W. D. S. & Jones, W. (2007). *Phys. Chem. Chem. Phys.* **9**, 1693–1704.
 Demirtaş, G., Dege, N. & Büyükgüngör, O. (2011). *Acta Cryst.* **E67**, o1509–o1510.
 Dey, A., Kirchner, M. T., Vangala, V. R., Desiraju, G. R., Mondal, R. & Howard, J. A. (2005). *J. Am. Chem. Soc.* **127**, 10545–10559.
 Dey, A., Pati, N. N. & Desiraju, G. R. (2006). *CrystEngComm*, **8**, 751–755.
 Eijck, B. P. van (2001). *J. Comput. Chem.* **22**, 816–826.
 Eijck, B. P. van (2002). *J. Comput. Chem.* **23**, 456–462.
 Eijck, B. P. van & Kroon, J. (2000). *Acta Cryst.* **B56**, 535–542.
 Eijck, B. P. van, Mooji, W. T. M. & Kroon, J. (2001). *J. Comput. Chem.* **22**, 805–815.
 Gale, J. D. & Rohl, A. L. (2003). *Mol. Simul.* **29**, 291–341.
 Gavezzotti, A. (2002). *Modell. Simul. Mater. Sci. Eng.* **10**, R1.
 Gavezzotti, A. (2003). *OPiX*. University of Milano, Italy.
 Gavezzotti, A. & Filippini, G. (1997). *Energetic Aspects of Crystal Packing: Experiment and Computer Simulations in Theoretical Aspects and Computer Modeling of the Molecular Solid State*. Chichester: Wiley and Sons.
 Görbitz, C. H., Dalhus, B. & Day, G. M. (2010). *Phys. Chem. Chem. Phys.* **12**, 8466–8477.
 Hofmann, D. W. M. & Apostolakis, J. (2003). *J. Mol. Struct. THEOCHEM*, **647**, 17–39.
 Hofmann, D. W. M. & Kuleshova, L. (2005). *J. Appl. Cryst.* **38**, 861–866.
 Hofmann, D. W. M. & Lengauer, T. (1997). *Acta Cryst.* **A53**, 225–235.
 Jiang, R.-W., Ming, D.-S., But, P. P. H. & Mak, T. C. W. (2000). *Acta Cryst.* **C56**, 594–595.
 Karamertzanis, P. G., Day, G. M., Welch, G. W. A., Kendrick, J., Leusen, F. J. J., Neumann, M. A. & Price, S. L. (2008). *J. Chem. Phys.* **128**, 244708.
 Karamertzanis, P. G. & Pantelides, C. C. (2007). *Mol. Phys.* **105**, 273–291.
 Karfunkel, H. R., Wu, Z. J., Burkhard, A., Rihs, G., Sinnreich, D., Buerger, H. M. & Stanek, J. (1996). *Acta Cryst.* **B52**, 555–561.

- Kazantsev, A. V., Karamertzanis, P. G., Adjiman, C. S. & Pantelides, C. C. (2011). *J. Chem. Theory Comput.* **7**, 1998–2016.
- Kazantsev, A. V., Karamertzanis, P. G., Adjiman, C. S., Pantelides, C. C., Price, S. L., Galek, P. T. A., Day, G. M. & Cruz-Cabeza, A. J. (2011). *Int. J. Pharm.* **218**, 168–178.
- Kazantsev, A. V., Karamertzanis, P. G., Pantelides, C. C. & Adjiman, C. S. (2010). *Molecular System Engineering*, edited by C. S. Adjiman & A. Galindo, Vol. 6, pp. 1–42. Weinheim: Wiley-VCH Verlag GmbH and Co.
- Kendrick, J., Leusen, F. J. J., Neumann, M. A. & van de Streek, J. (2011). *Chem. Eur. J.* **17**, 10736–10744.
- Kresse, G. & Furthmüller, J. (1996). *Phys. Rev. B*, **54**, 11169–11186.
- Kresse, G. & Hafner, J. (1993). *Phys. Rev. B*, **47**, 558–561.
- Kresse, G. & Joubert, D. (1999). *Phys. Rev. B*, **59**, 1758–1775.
- 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). *Acta Cryst.* **B56**, 697–714.
- MacGillivray, L. R. (2010). Personal communication.
- Macrae, C. F., Bruno, I. J., Chisholm, J. A., Edgington, P. R., McCabe, P., Pidcock, E., Rodriguez-Monge, L., Taylor, R., van de Streek, J. & Wood, P. A. (2008). *J. Appl. Cryst.* **41**, 466–470.
- Maleev, A. V. (1995). *Crystallogr. Rep.* **40**, 354–356.
- Maleev, A. V. (2001). *Crystallogr. Rep.* **46**, 13–18.
- Maleev, A. V., Zhitkov, I. K. & Rau, V. G. (2005). *Crystallogr. Rep.* **50**, 727–734.
- Maleev, A. V., Zhitkov, I. K. & Rau, V. G. (2009). *J. Struct. Chem.* **50**, S1–S7.
- Mooij, W. T. M., van Eijck, B. P. & Kroon, J. (2000). *J. Am. Chem. Soc.* **122**, 3500–3505.
- Motherwell, W. D. S. *et al.* (2002). *Acta Cryst.* **B58**, 647–661.
- Moult, J., Fidelis, K., Kryshchak, A., Rost, B., Hubbard, T. & Tramontano, A. (2007). *Proteins*, **69**, 3–9.
- Neumann, M. A. (2008). *J. Phys. Chem. B*, **112**, 9810–9829.
- Neumann, M. A. (2011). *GRACE*. Avant-garde Materials Simulation GmbH, Germany, <http://www.avmatsim.eu>.
- Neumann, M. A., Leusen, F. J. J. & Kendrick, J. (2008). *Angew. Chem. Int. Ed.* **47**, 2427–2430.
- Neumann, M. A. & Perrin, M. A. (2005). *J. Phys. Chem. B*, **109**, 15531–15541.
- Oganov, A. R. (2010). Editor. *Modern Methods of Crystal Structure Prediction*. Weinheim: Wiley-VCH Verlag GmbH and Co.
- Okabe, N., Kyoyama, H. & Suzuki, M. (2001). *Acta Cryst.* **E57**, o764–o766.
- Pickard, C. J. & Needs, R. J. (2006). *Phys. Rev. Lett.* **97**, 045504.
- Pickard, C. J. & Needs, R. J. (2011). *J. Phys. Condens. Matter*, **23**, 053201.
- Pillard, J., Arnautova, Y. A., Czaplewski, C., Gibson, K. D. & Scheraga, H. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 12351.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in FORTRAN*. Cambridge University Press.
- Price, S. L., Leslie, M., Welch, G. W. A., Habgood, M., Price, L. S., Karamertzanis, P. G. & Day, G. M. (2010). *Phys. Chem. Chem. Phys.* **12**, 8478–8490.
- Sarma, J. A. R. P. & Desiraju, G. R. (2002). *Cryst. Growth Des.* **2**, 93–100.
- Shan, N., Batchelor, E. & Jones, W. (2003). *Acta Cryst.* **E59**, o397–o398.
- Sobol', I. M. (1967). *Comput. Math. Math. Phys.* **7**, 86–112.
- Soler, J. M., Artacho, E., Gale, J. D., García, A., Junquera, J., Ordejón, P. & Sánchez-Porta, D. (2002). *J. Phys. Condens. Matter*, **14**, 2745–2779.
- Spek, A. L. (2009). *Acta Cryst.* **D65**, 148–155.
- Stone, A. J. (1996). *The Theory of Intermolecular Forces*. Clarendon Press: Oxford.
- Stone, A. J. J. (2005). *J. Chem. Theory Comput.* **1**, 1128–1132.
- Verwer, P. & Leusen, F. J. J. (1998). *Rev. Comput. Chem.* **12**, 327–365.