Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Significant progress in predicting the crystal structures of small organic molecules – a report on the fourth blind test

Graeme M. Day,[a]* Timothy G. Cooper,[a] Aurora J. Cruz-Cabeza,[a] Katarzyna E. Hejczyk,[a] Herman L. Ammon,[b] Stephan X. M. Boerrigter,[c] Jeffrey S. Tan,[c]‡ Raffaele G. Della Valle,[d] Elisabetta Venuti,[d] Jovan Jose,[e] Shridhar R. Gadre,[e] Gautam R. Desiraju,[f] Tejender S. Thakur,[f] Bouke P. van Eijck,[g]§ Julio C. Facelli,[h] Victor E. Bazterra,[i] Marta B. Ferraro,[j] Detlef W. M. Hofmann,[k] Marcus A. Neumann,[l] Frank J. J. Leusen,[m] John Kendrick,[m] Sarah L. Price,[n] Alston J. Misquitta,[n,o] Panagiotis G. Karamertzanis,[n] Gareth W. A. Welch,[n] Harold A. Scheraga,[p] Yelena A. Arnautova,[p] Martin U. Schmidt,[q] Jacco van de Streek,[q]¶ Alexandra K. Wolf[q] and Bernd Schweizer[r]

[a]The Pfizer Institute for Pharmaceutical Materials Science, University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, England, [b]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742-2021, USA, [c]School of Pharmacy and Pharmaceutical Sciences, Purdue University, West Lafayette, Indiana, USA, [d]Dipartimento di Chimica Fisica e Inorganica and INSTM-UdR, Università di Bologna, Viale Risorgimento 4, I-40136 Bologna, Italy, [e]University of Pune, Ganeshkhind, Pune 411007, India, [f]School of Chemistry, University of Hyderabad, Hyderabad 500 046, India, [g]Department of Crystal and Structural Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands, [h]Department of Biomedical Informatics and Center for High Performance Computing, University of Utah, 155 South 1452 East Rm 405, Salt Lake City, UT 84112-0190, USA, [i]Center for High Performance Computing, University of Utah, 155 South 1452 East Rm 405, Salt Lake City, UT 84112-0190, USA, [j]Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pab. I (1428), Buenos Aires, Argentina, [k]CRS4, Edificio 1, Loc. Piscinamanna, 09010 Pula (CA), Italy, [l]Avant-garde Materials Simulation Deutschland GmbH, Merzhauser Strasse 177, D-79100, Germany, [m]Institute of Pharmaceutical Innovation, University of Bradford, Bradford BD7 1DP, England, [n]Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, England, [o]University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, England, [p]Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA, [q]Institute of Inorganic and Analytical Chemistry, University of Frankfurt, Max-von-Laue-Strasse 7, D-60438 Frankfurt am Main, Germany, and [r]Organic Chemical Laboratory, ETH-Zurich, CH-8093 Zurich, Switzerland

Correspondence e-mail: gmd27@cam.ac.uk

© 2009 International Union of Crystallography
Printed in Singapore – all rights reserved

Received 12 December 2008
Accepted 3 February 2009

We report on the organization and outcome of the fourth blind test of crystal structure prediction, an international collaborative project organized to evaluate the present state in computational methods of predicting the crystal structures of small organic molecules. There were 14 research groups which took part, using a variety of methods to generate and rank the most likely crystal structures for four target systems: three single-component crystal structures and a 1:1 cocrystal. Participants were challenged to predict the crystal structures of the four systems, given only their molecular diagrams, while the recently determined but as-yet unpublished crystal structures were withheld by an independent referee. Three predictions were allowed for each system. The results demonstrate a dramatic improvement in rates of success over previous blind tests; in total, there were 13 successful predictions and, for each of the four targets, at least two groups correctly predicted the observed crystal structure. The successes include one participating group who correctly predicted all four crystal structures as their first ranked choice, albeit at a considerable computational expense. The results reflect important improvements in modelling methods and suggest that, at least for the small and fairly rigid types of molecules included in this blind test, such calculations can be constructively applied to help understand crystallization and polymorphism of organic molecules.

1. Introduction

This paper reports the results of the fourth blind test of crystal structure prediction, an international test of current methods hosted by the Cambridge Crystallographic Data Centre (CCDC) and which we will refer to as CSP2007.

Crystal structure prediction (CSP) has been a long-standing goal of computational materials chemistry. The grand aim is the ability to predict, by computational methods, how a molecule will crystallize (*i.e.* unit cell, space group and all atomic positions), given only its chemical diagram and, perhaps, crystallization conditions. With the chemical diagram as the main input, such methods could be used even prior to the synthesis of the given molecule, leading to the possibility of the computationally led design of molecules that will

‡ Current address: Preformulation, Product Research and Development, Eli Lilly and Company, Indianapolis, IN 46285, USA.

§ Retired.

¶ Current address: Avant-garde Materials Simulation Deutschland GmbH, Merzhauser Strasse 177, D-79100, Germany.

crystallize with desired structural and physical properties. Alternatively, for a molecule with a known crystal structure, CSP could help assess the likelihood of as-yet undiscovered polymorphism. The latter application is the main motivation from the industrial sector (*e.g.* pharmaceuticals or pigments), where the unanticipated appearance of a new polymorph, with its different physical properties, can be very undesirable.

Over the past couple of decades, many methods have been developed for the purpose of CSP (Beyer *et al.*, 2001; Verwer & Leusen, 1998) and, starting in 1999, the CCDC has organized periodic blind tests to assess the reliability of individual methods and to provide an objective picture of the status of the field. These blind tests involve a set of molecules being set as targets, with participating research groups challenged to predict their crystal structures, which were unknown to the predictors prior to the test. This approach allows a side-by-side comparison of the range of methods on the same set of molecules. This type of blind test is increasingly being used to monitor advances in several areas of predictive modelling, such as protein folding (Moult *et al.*, 2007), ligand–protein binding, the prediction of solvation energies (Nicholls *et al.*, 2008), solubilities (Llinàs *et al.*, 2008), and physical properties of fluids (Case *et al.*, 2007). Developments in these areas are necessarily usually tested by retrospective prediction (or 'postdiction') of known properties or structures, whereas blind tests require prospective prediction of unknown data; successful prediction in such a setting is therefore more convincing.

The aims and methods used to approach CSP have the most in common with protein-structure prediction (PSP), which has also been the subject of blind assessments in the world-wide community, in the 'Critical Assessment of Techniques for Protein Structure Prediction', or CASP, exercises (Moult *et al.*, 2007). Both CSP and PSP are usually approached as problems in global energy minimization, assuming that the resulting structure is determined solely by energy. Computation of kinetics of crystallization is largely absent from current CSP methods, with only simple models of crystal growth occasionally being used to assess putative structures (Anghel *et al.*, 2002; Day & Price, 2003; Coombes *et al.*, 2005). Protein-folding kinetics have also been addressed in PSP (Khalili *et al.*, 2006), but the main focus in both communities has been on locating the lowest-energy structures on the complex energy surface. This energy-based approach requires a high quality potential function and a good global optimization procedure. In recent years, PSP has started to emphasize free energy (Brooks III *et al.*, 2001; Liwo *et al.*, 2007) and the calculation of both structure and thermodynamic properties, whereas CSP has largely focused on structure determination based on potential energy. Lattice dynamics (Anghel *et al.*, 2002; van Eijck, 2001; Day *et al.*, 2005) and molecular dynamics (Karamertzanis *et al.*, 2008; Raiteri *et al.*, 2005) simulations are only occasionally used to evaluate free energies in CSP. Computational efforts in both communities also make use of some experimental information to guide the modelling: PSP often makes use of the structures of homologous proteins, whereas CSP calculations are frequently guided by space-group statistics from the Cambridge Structural Database (Allen, 2002). Occasionally, the results of an energy-based search are biased using a synthon approach, where re-ranking of the low-energy structures is based on the absence or presence of commonly occurring structural motifs in the crystal structures of similar molecules (Dey *et al.*, 2005, 2006). It can also be tempting to introduce a subjective assessment of structural features in the ranking of putative crystal structures (Day & Motherwell, 2006). There is clearly much room for variation in methods and, therefore, the need to compare them side-by-side as developments are made and new approaches are tested.

Both CSP (in the CCDC blind test exercises) and PSP (CASP exercises) have been carried out as blind prediction tests in the world-wide community at regular intervals – every 2–3 years, in the case of crystal structure prediction. Based on the results of these tests, the progress in structure prediction in PSP (Oldziej *et al.*, 2005; Borreguero & Skolnick, 2007) has improved considerably in recent years, more so than in the first three tests of CSP (Lommerse *et al.*, 2000; Motherwell *et al.*, 2002; Day *et al.*, 2005). This paper reports on the results of the fourth crystal structure prediction blind test.

2. Organization and approach

The organization of this latest blind test, CSP2007, was in most aspects the same as the first three such evaluations of the field, which have been published: CSP1999 (Lommerse *et al.*, 2000); CSP2001 (Motherwell *et al.*, 2002); CSP2004 (Day *et al.*, 2005). Invitations to participate were sent to 23 research groups known to be active in the field. This year, it was felt that, with a growing community working towards crystal structure prediction, the blind test should be open to participation from anyone making developments in the field. Therefore, the test was advertised through the newsletters and websites of crystallographic associations so that interested groups could contact the organisers and take part. In the end, 14 research groups participated.

The previous blind tests put forward targets for prediction in the following three categories:

(1) small, rigid molecules; only the elements C, H, N and O; less than *ca* 25 atoms;

(2) rigid molecules, containing elements or functional groups that present a challenge for modelling methods, and are allowed to be up to *ca* 30–40 atoms;

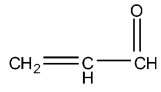
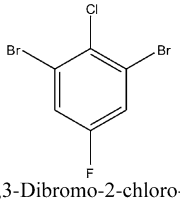
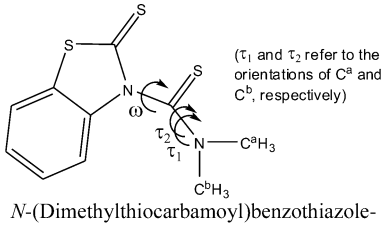
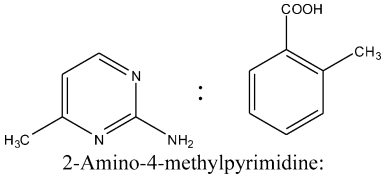
(3) molecules with several degrees of conformational flexibility, usually the rotation about exocyclic single bonds.

Molecules fitting these three categories have been included in CSP2007. Furthermore, with increasing interest within the crystal engineering community in the structures of multi-component crystals – salts, solvates and cocrystals – an additional category was added to the current test:

(4) a two-component crystal of rigid molecules.

This new fourth category specifically tests methods for sampling packing space with more than one independent molecule, which was introduced in the CSP2004 by allowing the possibility of $Z' > 1$ crystal structures in categories 1–3. In fact, the inability of many search methods to consider more

Table 1
Diagrams and crystallization conditions for the molecules of CSP2007.

Molecule		Crystallization conditions
(XII)	 2-Propenal	Grown from the melt by laser heating methods, T = 178 K
(XIII)	 1,3-Dibromo-2-chloro-5-fluorobenzene	Crystallized from acetonitrile
(XIV)	 *N*-(Dimethylthiocarbamoyl)benzothiazole-2-thione	Crystallized overnight by diethyl ether/hexane diffusion
(XV)	 2-Amino-4-methylpyrimidine: 2-methylbenzoic acid	1:1 cocrystal, crystallized by slow evaporation from ethanol

than one independent molecule contributed to the lack of prediction success for the small rigid molecule (XI) in CSP2004 (Day *et al.*, 2005). With the new category specifically testing methods for multiple independent molecules in this blind test, restrictions were reintroduced for categories 1–3: the crystal structures could be in any space group, but must have only one independent molecule ($Z' \leq 1$).

Crystallographers were contacted with a request for unpublished structures and suitable candidates were sent to an independent referee (Professor A. L. Spek, Laboratory of Crystal and Structural Chemistry, Utrecht University) who checked that they conformed to our criteria. To be suitable, a crystal structure had to be of high quality with all atoms located. After considerable effort, we collected one candidate for category 1, three for category 2, four for category 3 and three for category 4. Chemical diagrams of all candidates were then presented to an independent colleague (Dr Sijbren Otto, University Chemical Laboratory, Cambridge), who agreed to choose one target from each of the categories. The molecular diagrams and crystallization conditions, as shown in Table 1, were sent by email to all participants on 16 January 2007. Following the numbering used in the previous blind tests, we refer to these molecules by the Roman numerals (XII)–(XV).

We kept the format the same as in previous blind tests, allowing each participating group to submit three predictions for each system. Participants were asked to send their predictions for each molecule to Professor Spek, who held the experimentally determined crystal structures throughout the test. As well as the three 'official' predictions, analysis of extended lists of the crystal structures generated by each group can provide useful insight into the performance of the methods (van Eijck, 2005). Therefore, participants were encouraged to submit longer lists of their predicted structures, separately from their 'official' three, but preferably in ranked order. The deadline for submissions was 20 July 2007 and the experimentally determined crystal structures of all four systems were circulated to each participant on 23 July, for post-analysis of their predictions. A workshop was held at the Cambridge Crystallographic Data Centre in September 2007 to discuss the results.

3. Methodologies

Details of the methods used by the 14 participating research groups vary significantly, although most do involve three general steps:

(i) calculating three-dimensional molecular structures from the chemical diagrams;

(ii) searching the crystal packing phase space for the possible crystal packings;

(iii) assessing the generated structures to rank them in order of likelihood of formation.

Dividing the methods into these steps is mainly to aid discussion, as the steps do overlap in some methods. For example, the structure generation step often involves calculating and locally minimizing lattice energies, with the final energies used to rank the structures; in this case, steps (ii) and (iii) are not independent.

A brief discussion of the methods used in the latest blind test is provided here in the main body of the paper, and a summary of some key details for each participant is provided in Table 2, along with key references for most of the methods. For more detailed methodological descriptions, which were provided by many of the participants, the reader should refer to the supplementary material and references provided in the footnotes to Table 2.[1]

3.1. Methods of generating the molecular structure

The molecular structure that is used as the building block in the crystal structure search is usually derived from a force field or quantum-mechanics electronic structure calculation and there has been little focus since the previous blind tests on refining methods used here. For rigid molecules there have rarely been failures in crystal structure prediction that are due to a poor choice of starting molecular structure. Many of the methods treat the resulting molecular structure as rigid throughout the remainder of the calculations, assuming that crystal packing forces are too small to significantly distort the

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: BK5081). Services for accessing these data are described at the back of the journal.

molecular geometry. Other methods allow intramolecular degrees of freedom to vary during the search and/or final energy minimizations.

3.2. Generating trial crystal structures

Many approaches have been proposed to search the energy landscape for the lowest-energy crystal structures. Amongst the participants in this blind test, the most popular method was to generate large numbers of structures with random or quasi-random values for crystal structure variables (unit-cell parameters, positions and orientations of the molecules). Variations on the random search were used by six of the 14 groups. The others applied a variety of methods: Monte Carlo types of search (three groups); genetic algorithms (two groups); systematic grid-based searches (two groups) and Gavezzotti's PROM approach (1 group), which involves the stepwise construction of crystal structures from the most promising dimers, chains and layers.

Many groups made use of space-group symmetry to guide their search, with most focusing on a set (ranging from 4 to

Table 2
Summary of methodologies.

Contributor	Molecules attempted	Programs	Refs	Search generation	Space groups considered
(*a*)					
Ammon	(XII)–(XV)	*MOLPAK*, *DMAREL*	(*a*)	Grid-based systematic	*P*1, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, $P2_12_12$, *C*2/*c*, *Pbca*, *Pbcn*, $Pna2_1$, $Pca2_1$, *C*2, *Cc*
Boerrigter, Tan	(XII)–(XV)	*CERIUS*2 Polymorph Predictor	(*b*)	Monte Carlo simulated annealing	$P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $P2_1$, *C*2/*c*, *Pbca*, *Pnma*, *Pbcn*, $Pna2_1$, *P*1, *Cc*
Day, Cooper, Cruz Cabeza, Hejczyk	(XII)–(XV)	*Crystal Predictor*, *CERIUS*2 OFF, *DMAREL*	(*c*)	Structures generated using a low discrepancy Sobol' sequence	(XII)–(XIV) *P*1, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, *C*2/*c*, *Pbca*, *Pbcn*, $Pna2_1$, $Pca2_1$, *Pnma*, $P2_12_12$, *Cc*, *C*2, *C*2/*m*, *Pc*, $P2_1/m$, *P*2/*c*, *Pccn*, $R\bar{3}$ (XV) *P*1, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, *C*2/*c*, *Pbca*, *Pbcn*, $Pna2_1$, *Pnma*, *Cc*, *C*2/*m*, *Pc*
Della Valle, Venuti	(XII), (XIII), (XV)	*Xfind*, *WMIN*, *IONIC*, *PLATON*	(*d*)	Structures generated using a low-discrepancy Sobol' sequence	$Z' = 1$: *P*1, $P\bar{1}$, $P2_1$, $P2_1/c$, *C*2/*c*, $P2_12_12_1$, $Pna2_1$, *Pbca*, *Pnma* $Z' = 2$: *P*1
Desiraju, Thakur	(XIII)–(XV)	*CERIUS*2 Polymorph Predictor	(*e*)	Monte Carlo simulated annealing	*P*1, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, $P2_12_12$, *C*2/*c*, *Pbca*, *Pbcn*, $Pna2_1$, $Pca2_1$, *C*2, *Cc*, *C*2/*m*, $P2_1/m$, *P*2/*c*, *Pnma*
van Eijck	(XII)–(XV)	*UPACK*, *XTINKER*	(*f*)	Randomly generated starting structures	*P*1, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, *C*2/*c*, *Pbca*, *Pbcn*, $Pna2_1$, $Pca2_1$, *Cc*, *C*2, *Pc*
Facelli, Bazterra, Ferraro	(XII)–(XV)	*MGAC*	(*g*)	Modified genetic algorithm	*P*1, $P\bar{1}$, $P2_1$, *C*2, *Pc*, *Cc*, $P2_1/c$, *C*2/*c*, $P2_12_12_1$, $Pca2_1$, $Pna2_1$, *Pbcn*, *Pbca*, *Pnma*
Hofmann	(XII)–(XIV)	*FlexCryst*	(*h*)	Random search with calibrated cell	$P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, *C*2/*c*, *Pbca*, *Pbcn*, $Pna2_1$, $Pca2_1$, *Cc*, *C*2, $P2_1/m$
Jose, Gadre	(XII)–(XV)	GA-CG-MTA algorithm for crystal structure prediction	(*i*)	Genetic algorithm	*P*1, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, *Pbca*, *Pbcn*, *Pnma*, $Pna2_1$, $Pca2_1$, *Cc*, *C*2, *Pc*
Neumann, Leusen, Kendrick	(XII)–(XV)	*GRACE*1.0 and *VASP*	(*j*)	Random search with molecular flexibility	All 230 space groups searched
Price, Misquitta, Karamertzanis, Welch	(XII)–(XV)	*MOLPAK* or *Crystal Predictor*, *DMAREL*, *DMAflex*, *CamCASP*	(*k*)	Grid-based systematic or using Sobol sequence	*P*1, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, $P2_12_12$, *C*2/*c*, *Pbca*, *Pbcn*, $Pna2_1$, $Pca2_1$, *C*2, *Cc*
Scheraga, Arnautova	(XII)–(XIV)	*CRYSTALG*	(*l*)	Conformation-family Monte Carlo (CFMC)	No symmetry information used – *P*1 with varying *Z* (= 2, 4, 8)
Schmidt, van de Streek, Wolf	(XII)–(XV)	*CRYSCA*	(*m*)	Randomly generated starting structures	(XII): $P2_1/c$, *C*2/*c*, $P2_1$, $P\bar{1}$, $P2_12_12_1$, *Pbca* (XIII): $P2_1/c$, *C*2/*c*, $P2_1$, $P\bar{1}$, $P2_12_12_1$, *Pbca*, *P*1, $Pna2_1$, $Pca2_1$ (XIV): $P2_1/c$, *C*2/*c*, $P2_1$, $P\bar{1}$, *Pbca* (XV): $P2_1/c$, *C*2/*c*, $P2_1$, $P\bar{1}$, $P2_12_12_1$, *Pbca*, *P*1, $Pna2_1$, $Pca2_1$, *Cc*
Schweizer	(XII), (XIII), (XV)	*ZIP-PROMET*, *PIXEL*	(*n*)	Stepwise construction of dimers and layers	$P\bar{1}$, $P2_1$, $P2_1/c$, *C*2/*c*, $P2_12_12_1$, *Pbca*

Table 2 (continued)

Contributor	Molecular model	Lattice energy/fitness function		Other criteria used to select submissions
		Electrostatic	Other	
(*b*)				
Ammon	Rigid throughout	Atomic multipoles	Empirical exp-6	
Boerrigter, Tan	Rigid throughout	Electrostatic potential derived core-shell model using the CS-RQ method	Dreiding exp-6 force field	
Day, Cooper, Cruz Cabeza, Hejczyk	(XII), (XIII), (XV) rigid throughout (XIV) partly flexible during energy minimization	Atomic multipoles	Empirical exp-6 [(XII), (XIV), (XV)], specifically fitted anisotropic exp-6 (XIII)	Free energy [(XII), (XIII)]
Della Valle, Venuti	Rigid throughout	Atomic charges	Empirical exp-6	Free energy
Desiraju, Thakur	Rigid for search, flexible for energy minimization	Atomic charges	COMPASS force field	Assessment of packing (XIV) and synthon-based re-ranking (XV)
van Eijck	Flexible throughout	Atomic multipoles [for (XIII) charges only]	Empirical exp-6	
Facelli, Bazterra, Ferraro	Flexible throughout	Atomic charges	GAFF 6-12	
Hofmann	Rigid throughout	Trained potentials		
Jose, Gadre	Flexible throughout	CG–MTA *ab initio* energy HF/STO-3G		
Neumann, Leusen, Kendrick	Flexible throughout	Plane-wave density functional theory supplemented by an empirical C_6R^{-6}		
Price, Misquitta, Karamertzanis, Welch	Rigid for search, some flexibility during energy minimization for (XIV)	Atomic multipoles	Empirical exp-6 [(XII), (XIV), (XV)] non-empirically derived anisotropic exp-6 (XIII)	Choice 2 and 3 considered properties and motif
Scheraga, Arnautova	Rigid throughout	Atomic charges	Empirical exp-6 W99 (XII) ECEPP-05 (XIII) with specifically fitted halogen parameters ECEPP-05 (XIV)	
Schmidt, van de Streek, Wolf	(XII), (XIII) rigid, (XIV), (XV) some flexibility throughout	Atomic charges	Empirical exp-6	
Schweizer	Rigid throughout	Energy minimization with exp-6 UNI potential and energy calculations with pixel-based method		

References: (*a*) Holden *et al.* (1993), Busing (1981); (*b*) Verwer & Leusen (1998), Tan *et al.* (2009); (*c*) Karamertzanis & Pantelides (2005), Day, Motherwell & Jones (2005), Cooper *et al.* (2007), Cruz Cabeza *et al.* (2006), Day *et al.* (2007); (*d*) Della Valle *et al.* (2008), Brillante *et al.* (2008), Busing (1981), Signorini *et al.* (1991); Spek (2003); (*e*) Sarma & Desiraju (2002), Dey *et al.* (2005, 2006); (*f*) Mooij *et al.* (1999), van Eijck & Kroon (2000), van Eijck (2001, 2002); (*g*) Bazterra *et al.* (2007); (*h*) Hofmann & Lengauer (1997), Hofmann & Apostolakis (2003), Hofmann & Kuleshova (2005); (*i*) unpublished method – see supplementary material; (*j*) Neumann & Perrin (2005), Neumann *et al.* (2008), Neumann (2008); (*k*) Holden *et al.* (1993), Willock *et al.* (1995), Karamertzanis & Pantelides (2005), Karamertzanis & Price (2006), Karamertzanis & Pantelides (2007), Misquitta & Stone (2007), Misquitta *et al.* (2008), Price (2008); (*l*) Pillardy *et al.* (2001); (*m*) Schmidt & Englert (1996), Schmidt & Kalkhof (1997); (*n*) Gavezzotti (1999–2000), Gavezzotti (2004).

about 20) of the most commonly adopted space groups for organic molecules with $Z' = 1$. Only one group (Neumann, Leusen and Kendrick) considered all 230 space groups. The alternative approach is to generate $P1$ unit cells with varying numbers of total molecules in the unit cell. This strategy was employed by two groups: Scheraga and Arnautova generated $P1$ structures with two, four and eight molecules in the unit cell, locating space-group symmetry in the resulting structures after energy minimization using the *CRYCOM* program (Dzyabchenko, 1994). Della Valle and Venuti performed $P1$ searches with 1 and 2 independent molecules as well as $Z' = 1$ searches in the common space groups.

3.3. Ranking of structures

The final ranking of structures was usually based on calculated lattice energies of the structures generated by the crystal structure search. Therefore, most of the variability in the ranking of structures results from different choices of model for the crystal energies (Table 2). Two groups went beyond the static lattice-energy approach and included lattice-dynamics contributions to the free-energy differences between structures. New methods of evaluating energies have been introduced in this blind test, including atom–atom potentials derived purely from molecular quantum-mechanical calculations and the direct applicaton of quantum-mechanical electronic structure calculations to the crystal structures.

Aside from methods based purely on potential or free energies, there were attempts to include other criteria in the ranking of crystal structures. The cocrystal (XV) was the most attractive target for non-energetic assessment, as it is the only molecule in this blind test with the possibility of strong hydrogen bonding. Hydrogen-bond analysis of the structures of cocrystals of similar molecules could therefore be used to assess the hydrogen bonding in predicted crystal structures.

Table 3
Summary of conformational treatment of molecule (XIV) and approach taken for generating cocrystal structures for (XV).

Contributor	Conformational treatment of (XIV)	Treatment of intramolecular energy in (XIV)	Search strategy used for target (XV)
Ammon	The gas phase minimum was used and kept rigid throughout	None	Searched separately with all four dimers as building blocks
Boerrigter, Tan	The gas phase (B3LYP/cc-PVDZ) minimum was used	Dreiding/X6 force field	Both molecules treated independently. Both conformations of the acid were considered
Day, Cooper, Cruz Cabeza, Hejczyk	Searches were carried out with 12 starting conformations, varying ω ($\angle$CNC) and τ_1 ($\angle$CNCN). These two angles were allowed to optimize during lattice energy minimization	Taken from a separate B3LYP/6-31G** calculation on the final optimized conformer in each crystal structure	Both molecules treated independently. Both conformations of the acid were considered
Della Valle, Venuti	(XIV) was not attempted	(XIV) was not attempted	Both molecules treated independently and these searches were supplemented by searches using dimers as building blocks
Desiraju, Thakur	Searches were carried out with seven rigid conformations, varying ω ($\angle$ CNC)	Dreiding force field	Searched separately with the two lowest-energy dimers as building blocks
van Eijck	Standard starting geometry for optimized values of ω, τ_1 and τ_2; random values for the two methyl dihedrals	Specifically 6-31G* derived torsional potentials	Both molecules treated independently. Only one starting conformation for the acid group was considered
Facelli, Bazterra, Ferraro	ω, τ_1 and τ_2 were searched within the genetic algorithm, along with cell parameters and molecular positions	GAFF (Generic Amber Force Field)	Both molecules treated independently
Hofmann	The gas phase (DMol3 pwc/dnp) minimum was used and kept rigid throughout	Only one conformation was considered	(XV) was not attempted
Jose, Gadre	Torsion angles were varied within the search	The Hartree–Fock energies include both inter- and intramolecular energies	Searched separately with the three lowest-energy dimers as building blocks
Neumann, Leusen, Kendrick	Conformational freedom is searched automatically within the crystal structure generation	The intramolecular energy is part of the total DFT energy	Both molecules treated independently
Price, Misquitta, Karamertzanis, Welch	Searches were carried out with ten starting conformations, varying ω ($\angle$SCNC) and τ_1 ($\angle$CNCN). These angles were later allowed to optimize during DMAflex lattice energy minimization	Calculated from MP2/6-31G** calculations on the conformation in the DMAflex minimized crystal structure	Searched separately with all four dimers as building blocks (both molecules were treated independently in a *Crystal Predictor* search which was not completed by the blind test deadline, see supplementary material)
Scheraga, Arnautova	Searches were carried out with nine rigid conformations, varying ω ($\angle$SCNC) in the range 60–85° with the remaining torsional angles optimized for the isolated molecule	Taken from DFT/6-31G** calculated energy of the relevant conformation	(XV) was not attempted
Schmidt, van de Streek, Wolf	Starting geometry taken from a HF/6-31G** optimization. Torsion angles were allowed to vary during the search and minimization	Six-term cosine series fitted to the HF/6-31G** calculated energy surface	Searched separately with two dimer structures as building blocks, with flexibility in the relative orientations and conformation of acid
Schweizer	(XIV) was not attempted	(XIV) was not attempted	Searched separately with two of the dimer structures as building blocks

3.4. Treatments of the molecular flexibility in (XIV) and the independent molecules in (XV)

We summarize the various methods of treating the flexible molecule (XIV) in Table 3, adopting the nomenclature shown in Table 1 for the torsion angles. The search strategies for the cocrystal (XV) are also summarized.

4. Results

This paper is accompanied by a large amount of supplementary material: the coordinates of the experimental structures, lists of predicted 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 packing in the X-ray determined crystal structures of the four systems is described.

4.1. Experimental structures

4.1.1. Molecule (XII). Acrolein (C_3H_4O), or 2-propenal, was chosen as the blind test target for category (1). Acrolein melts at ~186 K (Timmermans, 1922), so crystal growth was performed at 178 K *in situ* by laser-assisted zone refinement

(Boese & Nussbaumer, 1994). The crystal structure was solved from X-ray diffraction data at 150 K and the molecule was found to crystallize in the orthorhombic space group *Pbca* with $Z' = 1$ (Forster *et al.*, 2007), with the molecule adopting the energetically favourable s-*trans* conformation. The lack of conventional hydrogen-bond donor groups means that the crystal structure is determined by weak interactions and each molecule is found to make six short C—H···O contacts with neighbouring molecules (Fig. 1), three as a C—H donor and three around the acceptor O atom. These contacts form a three-dimensional network through the crystal.

4.1.2. Molecule (XIII). A halogenated molecule, 1,3-dibromo-2-chloro-5-fluorobenzene ($C_6H_2Br_2ClF$), was selected for category (2), as a test of challenging atom types for simulations. Molecule (XIII) was crystallized from acetonitrile and the structure was solved from X-ray diffraction data at T = 173 K (Britton, 2008). The molecule crystallizes with $Z' = 1$ in the space group $P2_1/c$. With three different halogens in the molecule, there are many possible types of halogen–halogen contacts, which are expected to be crucial in determining the crystal structure. There are both Br···Br and F···F close contacts in the observed structure (Fig. 2), while the closest intermolecular contact with the chlorine involves a H atom with an uninteresting Cl···H distance of 3.04 Å. Br atoms interact in quartets, with each C—Br bond pointing to the side of one other bromine atom (Fig. 2, left). There are two independent Br···Br close contact distances of 3.55 and 3.63 Å; the C—Br···Br angles are 101.7 and 175.5° around the shorter contact, and 87.4 and 169.8° around the longer contact. F atoms form nearly head-to-head close contacts between coplanar molecules with an F···F distance of 2.87 Å. There is offset face-to-face stacking of the aromatic molecules along the crystallographic direction **a**, while the molecules make tilted edge-to-face and edge-to-edge contacts in the **b** and **c** directions.

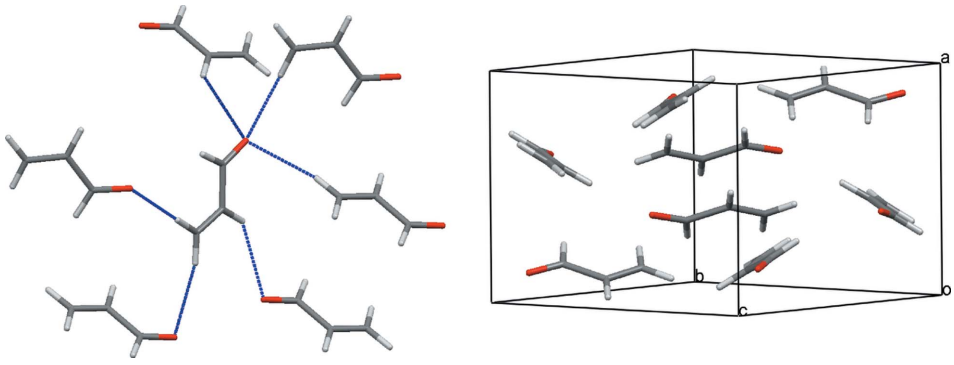


Figure 1
Packing diagrams of the crystal structure of molecule (XII). Grey = carbon, white = hydrogen, red = oxygen. C—H···O contacts (with $R_{O\cdots H}$ shorter than the sum of van der Waals radii) are indicated as blue lines.

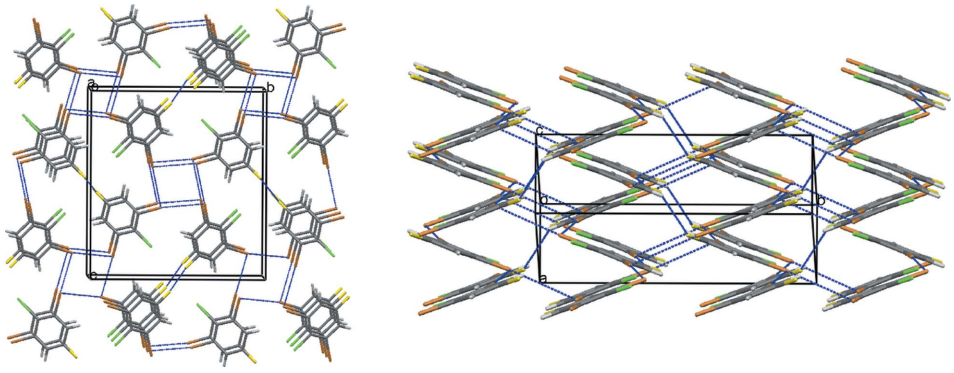


Figure 2
Packing diagrams of the crystal structure of molecule (XIII). Grey = carbon, white = hydrogen, yellow = fluorine, green = chlorine, brown = bromine. Short atom–atom contacts (with interatomic separation shorter than the sum of van der Waals radii) are indicated as blue lines.

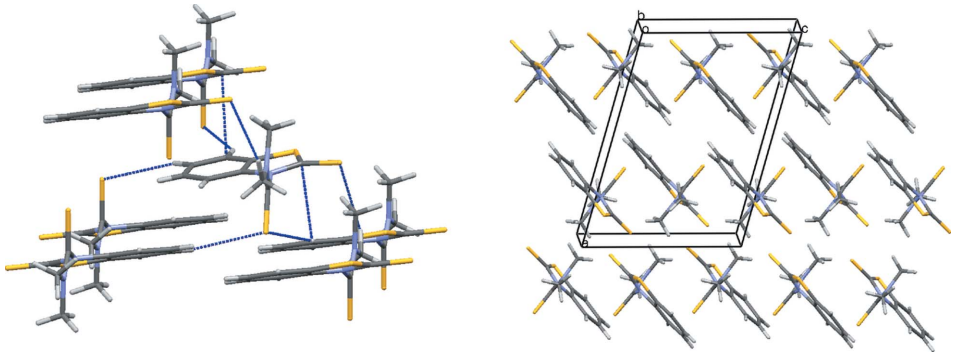


Figure 3
Crystal structure of molecule (XIV). Grey = carbon, white = hydrogen, blue = nitrogen, yellow = sulfur. Short contacts (with interatomic separation shorter than the sum of van der Waals radii) are indicated as blue lines.

4.1.3. Molecule (XIV). *N*-(Dimethylthiocarbamoyl)benzothiazole-2-thione ($C_{10}N_{10}N_2S_3$) was crystallized by diethyl ether/hexane diffusion and the crystal structure was determined from X-ray diffraction data at T = 150 K (Blake *et al.*, 2007). The molecules pack with $P2_1/c$ space-group symmetry.

The conformational flexibility can be defined by three exocyclic

torsion angles (Table 1), as well as rotation of the methyl groups, whose orientations are unlikely to be important in the crystal packing. One of these torsions (ω in Table 1) defines the angle of the thioformamide group out of the plane of the rings and the other two (τ_1 and τ_2 in Table 1) describe the orientation and planarity around the N atom; if the dimethylamine group is assumed to be planar, then the conformational flexibility can be reduced to two torsion angles. In the observed structure, the five heavy atoms of the thioformamide group ($SCNC_2$) are almost perfectly planar and nearly perpendicular to the benzothiazole plane (Fig. 3). The angle between mean planes of the thioformamide ($SCNC_2$) and benzothiazole (C_7NS) is 79.2°. There is a lack of hydrogen-bond donors in the molecule and almost all close intermolecular atom–atom contacts (*i.e.* shorter than the sum of van der Waals radii) are between S and H atoms.

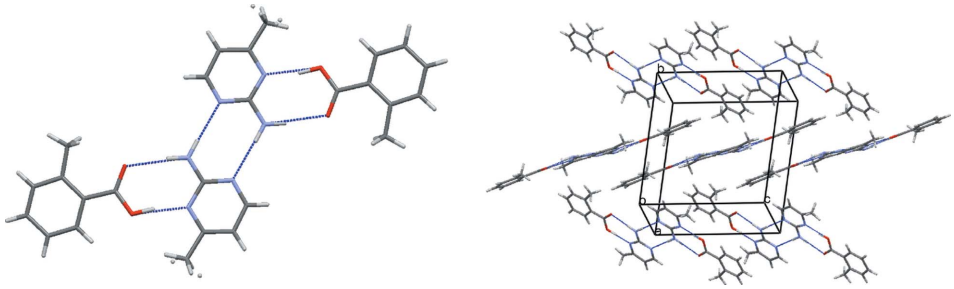


Figure 4
Crystal structure of target (XV). Grey = carbon, white = hydrogen, red = oxygen, blue = nitrogen. Hydrogen bonds indicated by thin blue lines. The unbonded H atoms in the left figure indicate the two-site disorder in the 2-amino-4-methylpyrimidine methyl-group orientation.

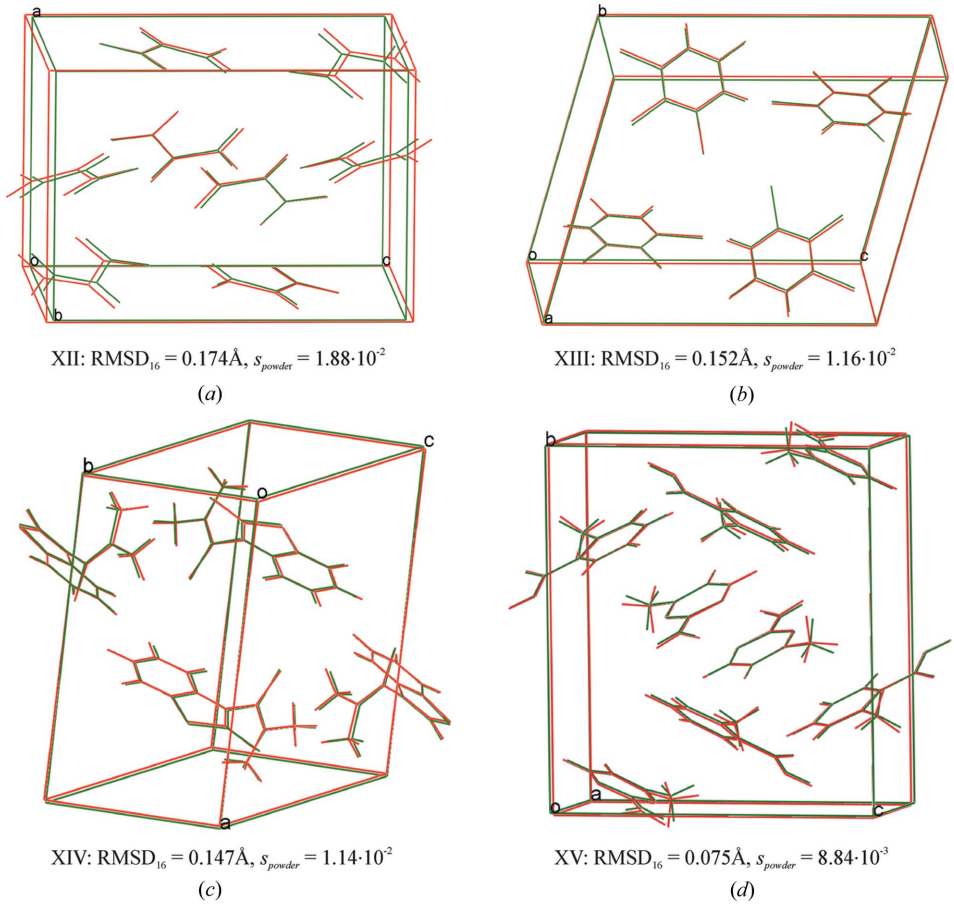


Figure 5
Overlays of the unit-cell contents of the four observed crystal structures, (XII)–(XV), and an example of one of the successful predictions for each. Observed structures are given in green, and the predicted structures in red. (*a*) Crystal structure of (XII) (green) and Ammon.XII.2 (red); (*b*) crystal structure of (XIII) (green) and Price.XIII.1 (red); (*c*) crystal structure of (XIV) (green) and vanEijck.XIV.1 (red); (*d*) crystal structure of (XV) (green) and NeumannLeusenKendrick.XV.1 (red).

4.1.4. Target (XV). The new category for this blind test was defined as a two-component crystal and the chosen target was the cocrystal formed between 2-amino-4-methylpyrimidine and 2-methylbenzoic acid. A 1:1 cocrystal was formed by slow evaporation of an ethanol solution and the crystal structure was solved from X-ray diffraction data at 203 K (Aakeröy, 2007). The prediction of what stoichiometry would form between a given pair of molecules was left as a future challenge (Cruz-Cabeza *et al.*, 2008) and, for this blind test, participants were given the observed stoichiometry as the starting information. Participants were also told that the molecules crystallize as a cocrystal, not a salt (although it is worthy of note that many pairs of similar molecules do crystallize as salts, with proton transfer from the carboxylic acid to the pyrimidine; Aakeröy *et al.*, 2003).

The molecules form nearly linear S-shaped hydrogen-bonded tetramers in the crystal structure (Fig. 4, left), with double hydrogen bonds between acid groups and the aminopyrimidine moiety. These acid–pyrimidine pairs are linked by two N—H···N hydrogen bonds between aminopyrimidines, which form over crystallographic centers of inversion in the $P2_1/n$ structure. These tetramers are nearly planar, with a 0.74 Å offset between root-mean-square (r.m.s.) planes of pyrimidine rings, and a 4.6° angle between mean planes of

Table 4
Lattice parameters, ΔE, $RMSD_{16}$ and s_{powder} for the experimental and predicted structures of molecule (XII).

$\alpha = \beta = \gamma = 90°$ in all structures.

	Rank	ΔE† (kJ mol^{-1})	Density (g cm^{-3})	a (Å)	b (Å)	c (Å)	$RMSD_{16}$‡ (Å)	$s_{powder} \times 10^2$§
Expt. (T = 150 K)	–	–	1.152	6.970 (3)	9.514 (5)	9.752 (5)	–	–
Predicted amongst first three								
Boerrigter, Tan	1	−0.15¶	1.117 (−3.0%)	6.879 (−1.3%)	9.697 (+1.9%)	9.994 (+2.5%)	0.156	1.46
Neumann, Leusen, Kendrick	1	−1.19¶	1.129 (−2.0%)	6.969 (−0..01%)	9.487 (−0.3%)	9.976 (+2.3%)	0.127	0.68
Ammon	2	+0.01	1.069 (−7.2%)	7.040 (+1.05)	9.746 (+2.4%)	10.150 (+4.1%)	0.174	1.88
Schweizer	2	+0.30	1.187 (+3.0%)	6.808 (−2.3%)	9.618 (+1.1%)	9.581 (−1.8%)	0.183	2.31
Present in the submitted extended list, outside of the first three predictions								
Price, Karamertzanis, Misquitta, Welch	2††	+0.29	1.064 (−7.6%)	7.000 (+0.4%)	9.864 (+3.7%)	10.139 (+4.0%)	0.200	–
van Eijck	6	+0.63	1.079 (−6.3%)	6.976 (+0.1%)	9.791 (+2.9%)	10.107 (+3.6%)	0.180	–
Della Valle, Venuti	13	+1.11	1.161 (+0.8%)	6.720 (−3.6%)	9.898 (+4.05)	9.644 (−1.1%)	0.268	–
Facelli, Bazterra, Ferraro	51	+2.38	1.170 (+1.6%)	6.765 (−2.9%)	9.867 (+3.75)	9.536 (−2.2%)	0.246	–
Not submitted, but located in post-analysis of predictions‡‡								
Schmidt, van de Streek, Wolf	6‡‡	+0.90	1.133 (−1.6%)	6.817 (−2.2%)	9.645 (+1.45)	10.000 (+2.5%)	0.179	–
Day, Hejczyk	117‡‡	+5.65	1.112 (−3.5%)	6.735 (−3.4%)	9.936 (+4.4%)	10.007 (+2.6%)	0.308	–
Not located in search, but energy minimized in post-analysis								
Hofmann	–	+1.68	1.062 (−7.8%)	7.311 (+4.9%)	9.708 (+2.0%)	9.883 (+1.3%)	0.190	–
Jose, Gadre	–	+2.14	1.079 (−6.3%)	7.110 (+2.0%)	9.720 (+2.2%)	9.990 (+2.4%)	0.319	–
Scheraga, Arnautova	–	+1.87	1.108 (−3.9%)	6.702 (−3.9%)	9.923 (+4.3%)	10.123 (+3.8%)	0.248	–

† ΔE is calculated with respect to the lowest-energy structure predicted by the same research group. ‡ RMSD is calculated using a 16 molecule comparison in *COMPACK*, ignoring H atoms. § s_{powder} is the normalized dissimilarity index calculated from simulated powder diffraction patterns. ¶ ΔE for the global minimum is calculated with respect to the second lowest-energy structure. †† The experimentally observed crystal structure was found as the second lowest in lattice energy, but not submitted as one of the three predictions, which were chosen from amongst the five lowest lattice-energy structures based on visual assessment and additional calculated properties. ‡‡ Structures reported in this category were submitted after the experimentally determined crystal structures were revealed, so cannot be considered blind predictions. They are included here to allow further analysis of the search and ranking methodologies, not as successful blind test predictions.

the pyrimidine and benzoic acids. The tetramers pack in a herringbone motif, with both face-to-face and tilted edge-to-face arene-arene interactions (Fig. 4, right).

There is two-site disorder of the pyrimidine methyl H atoms in the observed crystal structure, which was judged as unimportant in the selection of this target for the blind test. The observed disorder indicates that the methyl-group orientation has little effect on the energy of the crystal and H-atom positions are ignored in our comparison of predicted and observed structures, described below.

4.2. Comparison of the predictions with the experimental structures

We compared the submitted predictions with the experimentally determined crystal structures using the *COMPACK* algorithm (Chisholm & Motherwell, 2005; although the default in *COMPACK* is a 15-molecule cluster, we use a 16-molecule cluster here, to be consistent with the comparisons made in CSP2004), which compares the molecular packing environment in crystal structures. The experimentally determined crystal structure is represented by the interatomic distances between a molecule and its coordination shell of closest neighbouring molecules – here we choose 15 – and this set of distances is searched for in the predicted structures. If the distances match to within specified tolerances, then the coordination spheres are overlaid and a root-mean-squared deviation ($RMSD_{16}$) in atomic positions is calculated for all 16 molecules. We ignore H-atom positions in this comparison, because of the uncertainty in their positions in X-ray determined crystal structures.

To confirm the matches, a second screening was performed of the three predictions from each group against the experimental crystal structures. This second comparison used a measure of dissimilarity, s_{powder}, based on the calculated powder diffraction patterns of the two structures being compared. The measure amounts to an area between integrated patterns (Hofmann & Kuleshova, 2005)

$$s^{ij}_{powder} = \frac{1}{\vartheta_{max} - \vartheta_{min}} \times \int_{\vartheta_o=\vartheta_{min}}^{\vartheta_{max}} \left| \frac{1}{N_i} \int_{\vartheta=\vartheta_{min}}^{\vartheta_o} I_i(\vartheta) d\vartheta - \frac{1}{N_j} \int_{\vartheta=\vartheta_{min}}^{\vartheta_o} I_j(\vartheta) d\vartheta \right| d\vartheta_o. \quad (1)$$

Table 5
Lattice parameters, ΔE, $RMSD_{16}$ and s_{powder} for the experimental and predicted structures of molecule (XIII).

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

	Rank	ΔE† (kJ mol^{-1})	Density (g cm^{-3})	a (Å)	b (Å)	c (Å)	β (°)	$RMSD_{16}$‡ (Å)	$s_{powder} \times 10^2$§
Expt. (T = 173 K)	–	–	2.528	3.8943 (5)	13.5109 (17)	14.4296 (17)	93.636 (2)	–	–
Predicted amongst first three									
Ammon	1	−2.01¶	2.413 (−4.5%)	3.968 (+1.9%)	13.986 (+3.55)	14.309 (−0.8%)	91.78	0.385	1.59
Day	1	−0.68¶ (−0.29)††	2.506 (−0.9%)	3.880 (−0.4%)	13.683 (+1.3%)	14.403 (−0.2%)	92.01	0.159	0.72
Neumann, Leusen, Kendrick	1	−1.34¶	2.548 (+0.8%)	3.875 (−0.5%)	13.456 (−0.4%)	14.473 (+0.3%)	94.97	0.082	0.89
Price, Karamertzanis, Misquitta, Welch	1	−0.70¶	2.517 (−0.4%)	3.805 (−2.3%)	13.791 (+2.1%)	14.531 (+0.7%)	93.78	0.152	1.16
Present in list, outside of first three predictions									
Desiraju, Thakur‡‡	14	+3.56	2.577 (+1.9%)	3.868 (−0.7%)	15.093 (+11.7%)	12.731 (−11.8%)	90.15	1.768‡‡	–
van Eijck	16	+2.47	2.344 (−7.3%)	3.959 (+1.7%)	14.189 (+5.0%)	14.547 (+0.8%)	91.38	0.410	–
Della Valle, Venuti	84	+5.18	2.297 (−9.1%)	4.096 (+5.2%)	14.138 (+4.6%)	14.419 (−0.1%)	93.08	0.500	–
Not submitted, but located in post-analysis of predictions§§									
Boerrigter, Tan	4§§	+0.78	2.632 (+4.1%)	3.707 (−4.8%)	13.604 (+0.7%)	14.475 (+0.3%)	94.45	0.285	–
Schmidt, van de Streek, Wolf	10§§	+2.43	2.601 (+2.9%)	3.649 (−6.3%)	13.649 (+1.0%)	14.834 (+2.8%)	94.69	1.059	–
Not located in search, but energy minimized in post-analysis									
Hofmann	–	+12.11	2.247 (−11.1%)	4.018 (+3.2%)	14.461 (+7.0%)	14.687 (+1.8%)	92.76	0.373	–
Facelli, Bazterra, Ferraro	–	+6.20	2.289 (−9.5%)	4.138 (+6.3%)	14.107 (+4.4%)	14.366 (−0.4%)	86.22	0.546	–
Scheraga, Arnautova	–	+2.77	2.608 (+3.2%)	3.761 (−3.4%)	13.585 (+0.6%)	14.412 (−0.1%)	94.2	0.315	–

† ΔE is calculated with respect to the lowest-energy structure predicted by the same research group. ‡ $RMSD_{16}$ is calculated using a 16 molecule comparison in *COMPACK*, ignoring H atoms. § s_{powder} is the normalized dissimilarity index calculated from simulated powder diffraction patterns. ¶ ΔE for the global minimum is calculated with respect to the second lowest-energy structure. †† Quasi-harmonic free energy with respect to the second lowest-energy structure. ‡‡ Reported as a match to the experimental structure, but with extreme deviations. §§ Structures reported in this category were submitted after the experimentally determined crystal structures were revealed, so cannot be considered blind predictions. They are included here to allow further analysis of the search and ranking methodologies, not as successful blind test predictions.

The index becomes zero for identical structures and the normalization factor

$$N_i = \int_{\vartheta=\vartheta_{\min}}^{\vartheta_{\max}} I_i(\vartheta)\mathrm{d}\vartheta \qquad (2)$$

ensures that s_{powder} has a maximum value of 1. Structures are deemed to be the same when s_{powder} is below a certain threshold.

The two comparisons, one working in direct space and one in reciprocal space, gave the same list of matching structures. Matched structures, amongst the three 'official' predictions, and the extended lists of computer-generated crystal structures, are listed in Tables 4–7 (under the headings 'predicted amongst first three' and 'present in the submitted extended list, outside of the first three predictions'). Overlays of the unit-cell contents in matches for each target are shown in Fig. 5, along with the measured value for $RMSD_{16}$ and s_{powder}. There are two other sections in some of these tables: where groups located the experimentally observed crystal structure amongst their predictions, but outside of the lists they had submitted before the prediction deadline ('not submitted, but located in post-analysis'), and where the group had not located the correct crystal structure in their search, but energy minimized the X-ray structure in post-analysis to test the performance of their energy model ('not located in search, but energy minimized in post-analysis of predictions'). It must be emphasized that structures listed in both of these final categories fall outside of the 'blind' part of the exercise and are included here as extra information that is useful in assessing the methods in detail.

4.3. Predictions results

4.3.1. Molecule (XII). 13 of the 14 participating research groups attempted predictions for molecule (XII), four of whom predicted the observed structure within their three predictions (Table 4). Two of these successes (Neumann, Leusen & Kendrick; Boerrigter & Tan) were submitted as the

Table 6
Lattice parameters, ΔE, $RMSD_{16}$ and s_{powder} for the experimental and predicted structures of molecule (XIV).

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

	Rank	ΔE† (kJ mol^{-1})	Density (g cm^{-3})	a (Å)	b (Å)	c (Å)	β (°)	$RMSD_{16}$‡ (Å)	$s_{powder} \times 10^2$§
Expt. (T = 150 K)	–	–	1.479	13.060 (3)	9.738 (2)	9.335 (2)	105.800 (3)	–	–
Predicted amongst first three									
van Eijck	1	−0.24¶	1.497 (+1.2%)	12.853 (−1.6%)	9.803 (+0.7%)	9.341 (+0.1%)	106.52	0.147	1.14
Neumann, Leusen, Kendrick	1	−1.98¶	1.450 (−2.0%)	13.242 (+1.4%)	9.821 (+0.9%)	9.314 (−0.2%)	105.80	0.130	0.80
Price, Karamertzanis, Misquitta, Welch	1	−4.19¶	1.466 (−0.9%)	12.882 (−1.4%)	9.765 (+0.3%)	9.612 (+3.0%)	107.62	0.222	1.18
Present in list, outside of first three predictions									
Facelli, Bazterra, Ferraro	6	+4.65	1.543 (+4.3%)	14.046 (+7.5%)	9.612 (−1.3%)	8.263 (−11.5%)	100.95	0.830	–
Day, Cooper	8	+6.52	1.432 (−3.2%)	12.472 (−4.5%)	9.894 (+1.6%)	10.078 (+8.0%)	108.39	0.536	–
Not submitted, but located in post-analysis of predictions††									
Scheraga, Arnautova	2†††‡‡	+4.73	1.451 (−1.9%)	13.032 (−0.2%)	9.692 (−0.5%)	9.638 (+3.3%)	106.9	0.221	–
Ammon	4††	+1.99	1.401 (−5.3%)	13.438 (+2.9%)	9.684 (−0.6%)	9.690 (+3.8%)	106.96	0.264	–
Schmidt, van de Streek, Wolf	4††	+0.67	1.513 (+2.3%)	12.957 (−0.8%)	9.569 (−1.7%)	9.366 (+0.3%)	105.97	0.127	–
Boerrigter, Tan	9††	+2.12	1.462 (−1.2%)	12.957 (−0.8%)	9.626 (−1.2%)	9.596 (+2.8%)	105.05	0.195	–
Not located in search, but energy minimized in post-analysis									
Hofmann	–	+3.09	1.454 (−1.7%)	13.041 (−0.1%)	9.821 (+0.9%)	9.443 (+1.2%)	106.12	0.081	–
Jose, Gadre	–	+103.48	1.372 (−7.2%)	13.047 (−0.1%)	10.100 (+3.7%)	9.480 (+1.6%)	99.71	0.522	–

† ΔE is calculated with respect to the lowest-energy structure predicted by the same research group. ‡ $RMSD_{16}$ is calculated using a 16 molecule comparison in *COMPACK*, ignoring H atoms. § s_{powder} is the normalized dissimilarity index calculated from simulated powder diffraction patterns. ¶ ΔE for the global minimum is calculated with respect to the second lowest-energy structure. †† Structures reported in this category were submitted after the experimentally determined crystal structures were revealed, so cannot be considered blind predictions. They are included here to allow further analysis of the search and ranking methodologies, not as successful blind test predictions. ‡‡ The correct structure was ranked second on energy, but not submitted because the predicted structure was slightly out of symmetry, with $Z' = 2$.

group's first prediction, while the other two (Ammon; Schweizer) were submitted as the participant's second prediction. All four of these correct predictions gave $RMSD_{16}$ deviations from the experimentally determined structure of less than 0.2 Å and root-mean-squared errors in the unit-cell lengths (a, b, c) of less than 3%. An overlay of one of these predictions with the X-ray determined structure is shown in Fig. 5.

Outside of the official three predictions, the observed crystal structure was present in the extended lists of six other research groups and only three of the 13 groups reported not finding the structure in their list of computer-generated crystal structures. The success rates here are a moderate improvement over the previous blind tests, whose category 1 molecules and success rates are shown in Fig. 6. Only molecule (I) from CSP1999 had as high a success rate (four of 11 groups with successful predictions), but only for one of its known polymorphs – there were no successful predictions of the other polymorph.

4.3.2. Molecule (XIII). All 14 participants attempted predictions for molecule (XIII), four of whom (Ammon; Day; Neumann, Leusen & Kendrick; Price, Karamertzanis, Misquitta & Welch) predicted the observed crystal structure (Table 5). All of these successes were found as the first predicted structure from that participant and all gave an $RMSD_{16}$ deviation from the observed structure of less than 0.4 Å, with root-mean-squared errors in the unit-cell lengths (a, b, c) of better than 2.4%. An overlay of one of the four successful predictions is shown in Fig. 5. The success rates here are about the same as in this category in CSP2001, and higher than in the other two previous blind tests (Fig. 7).

The observed crystal structure was generated by the search method used by five other research groups, outside of their top three predictions and between 0.8 and 5.2 kJ mol^{-1} above their global minimum. These generally had greater geometric deviations from the experimental structure than seen in the predictions where the structure was ranked first in energy. The other five groups reported not finding the structure in their list of computer-generated crystal structures, indicating a failure of the search method. The slightly higher rate of search method failure here than for molecule (XII) might reflect difficulties in modelling the halogen atoms, as many of the methods do involve lattice-energy calculations and crystal structure optimizations during the search procedure. Therefore, poor modelling of the interactions can lead to a failed search.

Table 7
Lattice parameters, ΔE, $RMSD_{16}$ and s_{powder} for the experimental and predicted structures of molecule (XV).

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

	Rank	ΔE† (kJ mol^{-1})	Density (g cm^{-3})	a (Å)	b (Å)	c (Å)	β (°)	$RMSD_{16}$‡ (Å)	$s_{powder} \times 10^2$§
Expt. (T = 203 K)	–	–	1.301	7.2795 (10)	13.6699 (18)	12.6695 (16)	96.646 (3)	–	–
Predicted amongst first three									
Neumann, Leusen, Kendrick¶	1††	−2.08	1.307 (+0.5%)	7.264 (−0.2%)	13.818 (+1.1%)	12.520 (−1.2%)	97.44	0.075	0.88
van Eijck¶	3	+1.36	1.303 (+0.2%)	7.336 (+0.8%)	13.556 (−0.8%)	12.674 (+0.0%)	97.34	0.294	1.32
Present in list, outside of first three predictions									
Day, Cruz Cabeza¶	4	+3.47	1.272 (−2.2%)	7.201 (−1.1%)	13.943 (+2.0%)	12.884 (+1.7%)	98.07	0.242	–
Not submitted, but located in post-analysis of predictions‡‡									
Boerrigter, Tan	26‡‡	+5.80	1.240 (−4.7%)	7.698 (+5.7%)	13.834 (+1.2%)	12.451 (−1.7%)	97.88	0.536	–
Schmidt, van de Streek, Wolf	> 100‡‡	+11.32	1.264 (−2.8%)	7.312 (+0.4%)	13.701 (+0.2%)	13.989 (+10.4%)	113.11	0.385	–
Not located in search, but energy minimized in post-analysis‡‡									
Della Valle, Venuti	–	+1.81	1.301 (−0.0%)	7.633 (+4.8%)	12.693 (−7.1%)	12.968 (+2.4%)	94.80	0.473	–
Facelli, Bazterra, Ferraro	–	+0.91	1.350 (+3.8%)	7.156 (−1.7%)	12.798 (−6.4%)	13.215 (+4.3%)	94.53	0.491	–
Hofmann	–	+47	1.271 (−2.3%)	7.486 (+2.8%)	13.512 (−1.2%)	12.737 (+0.5%)	95.71	0.157	–
Price, Karamertzanis, Misquitta, Welch	1§§	0§§	1.301 (+0.0%)	7.250 (−0.4%)	13.774 (+0.8%)	13.625 (+7.5%)	113.07	0.203	–

† ΔE is calculated with respect to the lowest-energy structure predicted by the same research group. ‡ $RMSD_{16}$ is calculated using a 16 molecule comparison in *COMPACK*, ignoring H atoms, with the 2-methylbenzoic acid as the central molecule in the cluster. § s_{powder} is the normalized dissimilarity index calculated from simulated powder diffraction patterns. ¶ All three predictions were submitted in $P2_1/c$, which have been converted to the $P2_1/n$ setting for comparison with the experimentally determined structure. †† ΔE for the global minimum is calculated with respect to the second lowest-energy structure. ‡‡ Structures reported in this category were submitted after the experimentally determined crystal structures were revealed, so cannot be considered blind predictions. They are included here to allow further analysis of the search and ranking methodologies, not as successful blind test predictions. §§ Result from post-analysis completion of *Crystal Predictor* search, see supplementary material.

4.3.3. Molecule (XIV). 12 research groups attempted predictions for the category 3 target, molecule (XIV), three of whom (van Eijck; Neumann, Leusen & Kendrick; Price, Karamertzanis, Misquitta & Welch) were found to have predicted the observed crystal structure within their three predictions (Table 6). An overlay of one of the successful predictions is shown in Fig. 5. Each of these groups found the correct structure as their first ranked prediction, with $RMSD_{16}$ deviations from the observed structure of 0.22 Å or lower and root-mean-squared errors in the unit-cell lengths (a, b, c) smaller than 2%. The observed crystal structure was present in the extended lists of six other research groups, while three groups did not find the observed structure in their search.

CSP1999 (molecule I) successful predictions: 0/11 (polymorph 1) 4/11 (polymorph 2)

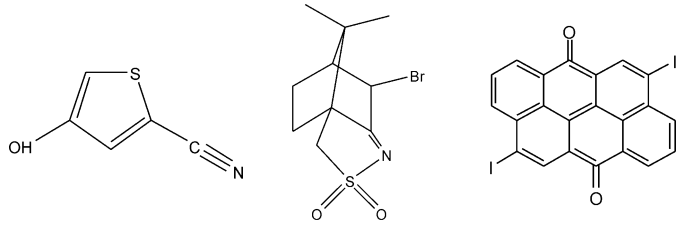
CSP1999 (molecule VII) successful predictions: 1/6

CSP2001 (molecule IV) successful predictions: 2/15

CSP2004 (molecule XI) successful predictions: 0/18

Figure 6
Previous blind test molecules in category (1) (simple rigid molecules). Success rates for these are given as number of correct predictions/number of participants.

These rates of success are similar to those seen for rigid molecules from previous blind tests and a noticeable improvement on what has previously been seen for flexible molecules (Fig. 8). There has only been one successful prediction for a flexible molecule in all three of the previous tests. This striking improvement might partly reflect advances in methods of dealing with conformational flexibility during the crystal structure search and during the ranking of struc-



CSP1999 (molecule II) successful predictions: 1/8

CSP2001 (molecule V) successful predictions: 4/15

CSP2004 (molecule IX) successful predictions: 1/15

Figure 7
Previous blind test molecules in category (2) (rigid molecules with challenging functional groups). Success rates for these are given as number of correct predictions/number of participants.

Table 8
Summary of computational resources used by some of the participants in CSP2007.

Group	Comments on computing time used	Total computational cost, approximately normalized to 2.8 GHz CPU hours
Boerrigter, Tan	(XII): 55 h, 200 MHz octane (XIII): 177 h, 200 MHz octane (XIV): 299 h, 200 MHz octane (XV): 194 h 200 MHz octane, + 280 h 3.0 GHz Pentium 4	350 CPU hours
Day, Cooper, Cruz Cabeza, Hejczyk	Crystal structure search (on 1.3 GHz Itanium processors): (XII) $\simeq$ 200 CPU hours; (XIII)–(XV) $\simeq$ 300 CPU hours each. Lattice-energy minimization and free-energy calculations (on 2.4 GHz Opteron processors) ranged from 70 (XV) to 320 CPU hours (XIV)	$\sim$ 1000 CPU hours
Della Valle, Venuti	Processor times on 2.2 GHz 64-bit processors were 12, 15 and 43 d for molecules (XII), (XIII) and (XV). About 96% of this time was spent on energy minimization	1320 CPU hours
van Eijck	147 CPU hours molecular calculations, 1611 CPU hours spent on searches, 732 CPU hours on energy minimization, time standardized to 2.8 GHz processors	2490 CPU hours
Facelli, Bazterra, Ferraro	Approximated total computer time was 200 000 CPU hours on 2.5 GHz class processors	180 000 CPU hours
Hofmann	(XII): 30 h, (XIII): 60 h, (XIV): 60 h, 3.0 GHz processor	160 CPU hours
Neumann, Leusen, Kendrick	Approximately 280 000 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	$\sim$ 280 000 CPU hours
Price, Karamertzanis, Misquitta, Welch	Each *MOLPAK* search could run overnight on the UCL Condor cluster of PCs, and a similar period was required for a simple reminimization of order of 1000 structures with *DMAREL* on one processor. The *Crystal Predictor* searches took a few days for (XIII) DMAflex refinements (XIV) took several days of CPU time for each of the ten structures Total excludes the potential development for (XIII) of $\sim$ 4000 CPU hours, and the work on (XV) using *Crystal Predictor* search which was only completed after the deadline (2 weeks CPU time for the intramolecular potential surfaces, and an equivalent amount of time for the *Crystal Predictor* search, and $\sim$ 30 DMAflex minimizations of about a week each)	$\sim$ 5000 CPU hours
Schmidt, van de Streek, Wolf	$\sim$ 8 months CPU time, 1.7 GHz AMD processors	$\sim$ 3500 CPU hours

tures. We must also consider that the molecule chosen for this category was less challenging than those in previous blind tests (Fig. 8). This is one unavoidable weakness of using the blind tests to measure progress in the field – variations in the difficulty of molecules can be as important as changes in the methods used to predict crystal structures and it is difficult to judge the difficulty associated with a molecule before performing the calculations involved in its prediction. In the case of molecule (XIV), it was felt that the conformational flexibility of the molecule was less challenging than in previous blind tests. Several groups performed quantum mechanical calculations to map out the energy of molecule (XIV) as a function of rotation about one or more of the exocyclic single bonds. While details of the methods varied, the minimum-energy conformation was generally found to have a planar geometry about the exocyclic N atom and an angle of 70–80° between the thioformamide and benzothiazole groups, *i.e.* very close to the conformation found in the crystal structure. Therefore, crystal structure searches using the gas phase minimum molecular geometry had a good chance of finding the observed structure and ranking it favourably on energy. Predictions were simplified by there only being one minimum on the conformational energy surface: for a cost of *ca* 10 kJ mol^{-1}, the out-of-plane angle of the thioformamide can distort about 30° either side of the minimum and the geometry around the exocyclic N atom can rotate by a similar amount. This much intramolecular energy could be compensated for by improved packing and intermolecular interactions, so these distortions from the gas-phase minimum geometry had to be considered during the crystal structure predictions. However, the resulting relevant conformational space was fairly restricted compared with the flexible molecules in the previous blind tests. As an example, the packing of molecule (X) in CSP2004 (Fig. 8) was found to be quite sensitive to six torsion angles, all of whose orientations had to be considered during the predictions (Day *et al.*, 2005).

4.3.4. Target (XV). 12 participants attempted predictions for the cocrystal (XV) and two of these predicted the observed cocrystal structure within the three official predictions (Neumann, Leusen & Kendrick and van Eijck), as the first and third predictions, respectively (Table 7). Both had RMS errors in the lattice constants (a, b, c) of less than 1% and $RMSD_{16}$ deviations in atomic positions better than 0.3 Å. An overlay of one of the two successful predictions is shown in Fig. 5 (where the disordered pyrimidine methyl group H atoms in the observed structure are shown in the site with highest occupancy). Three other groups had found the observed crystal structure outside of their three best predic-

tions (Table 6), while the other seven failed to locate the observed crystal structure in their search.

The cocrystal was introduced in this blind test as a new category of prediction challenge, so there are no results from previous blind tests with which to compare. The most similar example from previous blind tests is that of molecule (XI) of CSP2004, which crystallized with two independent molecules. No groups predicted the correct crystal structure for that molecule, partly because many could not or opted not to search for crystal structures with $Z' = 2$. In CSP2004 the value of Z' was not given, but $Z' > 1$ was allowed as a possibility, unlike here, where the contents of the asymmetric unit were specified.

In this blind test two different approaches were applied to searching phase space with more than one type of molecule in the crystal structure. One option was to search all of the packing space, with the positions and orientations of the two molecules treated independently. As discussed by van Eijck (van Eijck & Kroon, 2000; van Eijck, 2002), it is a considerable computational challenge to exhaustively search all of the crystal-packing space with two independent molecules, because of the six extra degrees of freedom compared with the search space when there is only one molecule in the asymmetric unit. Furthermore, there is a choice of conformation for the acid molecule which had to be considered.

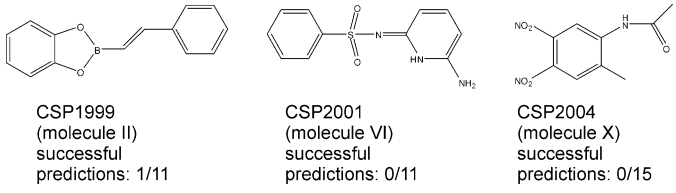


Figure 8
Previous blind test molecules in category (3) (flexible molecules). Success rates for these are given as number of correct predictions/number of participants.

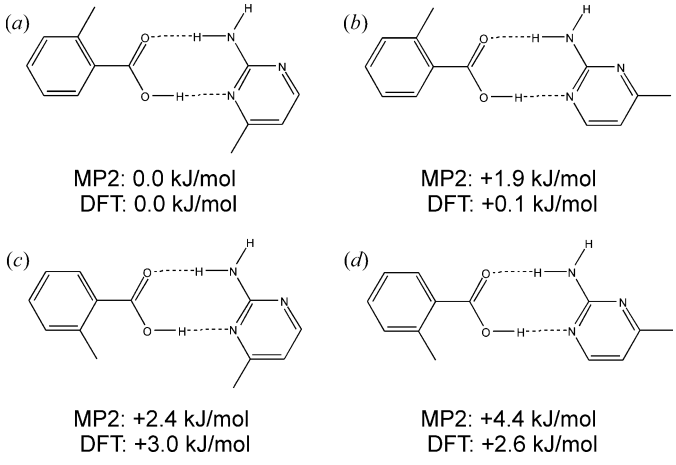


Figure 9
The four likely hydrogen-bond dimer structures formed between 2-methylbenzoic acid and 2-amino-4-methylpyrimidine. Calculated energies at MP2 (MP2/6-31G** from Ammon) and DFT (B3LYP/6-31G**, from Thakur & Desiraju, 2008) levels of theory are given, relative to the most stable dimer.

The other strategy used to generate crystal structures takes advantage of the strong interactions between the two molecules, which helps predict their relative orientation before starting to generate crystal structures. Several groups deemed that hydrogen-bond dimers were likely and used dimers as the basic unit with which crystal structures were generated, essentially reducing the problem back to that of a single-component crystal. Indeed, a survey of known structures of carboxylic acid: pyrimidine cocrystals in the CSD finds that such dimers are always formed, so this strategy was well founded in this case. For this pair of molecules, four planar dimer geometries are possible (Fig. 9), so there is a choice of which dimer geometries to consider in generating crystal structures. Some groups performed searches with all four possibilities, while others chose the most likely dimer structures from calculated energies. In this case the dimer geometry in the observed crystal structure corresponds to the lowest-energy dimer from various flavours of quantum mechanical calculation (Fig. 9*a*).

Table 3 summarizes the cocrystal search strategy used by each participant; five groups used the approach with two independent molecules, six groups used the dimer-based approach and one group used a combination of the two approaches (performing searches both with independent molecules and with dimers as starting points). Of the five groups who found the observed crystal structure either in their three official predictions or in their extended lists, two had used the dimer-based approach in the crystal structure search and three (including the two successful predictions) had used independent molecules in the search. Both methods can clearly be successful, but several groups using either search strategy also failed to produce the observed structure.

4.4. Computational expense

The range of methods being applied to crystal structure prediction come at varying costs in terms of computational time and resources, and some of the methods now being used in the blind tests have only been made possible by access to high-performance computing resources. Therefore, participants in CSP2007 were asked to keep track of the computational resources used to come up with their predictions, to give an idea of the resources required for each approach. Table 8 summarizes the resources used by some of the participants, where available and easily quantifiable.

The computing requirements can clearly be very high and would be a consideration in the choice of method for a particular problem. Most methods have required a few hundred hours (weeks) to a few thousand hours (months) CPU time on a modern processor. The computing requirement for the very successful method of Neumann, Leusen and Kendrick is several orders of magnitude higher, at 280 000 CPU hours ($\sim$ 32 CPU years) for predictions on the four targets. In the majority of methods, most of the computing time is being spent on the energy-minimization

part of the problem. The real time used for the calculations is often much shorter, because of the use of parallel or distributed computing setups. The price of powerful computing clusters is decreasing year-on-year such that even the most expensive of these methods could be brought down to a matter of weeks to a few months in real computing time at relatively low cost.

5. Discussion

5.1. Overall success rates

The success rates in the blind test depend principally on the performance of two main elements in the prediction methodology: the generation of all possible crystal structures, followed by the evaluation and ranking of these structures. Both must be performed effectively for a successful prediction, while failure of either the search or the ranking precludes success. In this fourth blind test, we have observed improved overall rates of successful prediction over the first three blind tests, reflecting developments in the methods applied to each step in crystal structure prediction.

Of the 14 groups participating in CSP2007, most attempted predictions for all four targets and half (seven groups) had at least one successful prediction within the rules of the blind test, where three predictions are allowed for each molecule. Four of the participating groups had multiple successes and, overall, there were 13 successful predictions, ten of which were submitted as a participant's first choice prediction. The quality of these predictions is illustrated in the overlays of predicted structures with those determined from X-ray diffraction data (Fig. 5). The success rate here is an important improvement over the results from the previous blind tests: 11 of the successful predictions were for molecules in the 'original' three categories of molecules that formed the first three blind tests, while in CSP2004 there was only one successful prediction from 18 participating groups[2] and there were six successes in each of CSP1999 (Lommerse *et al.*, 2000) and CSP2001 (Motherwell *et al.*, 2002). The CSP2004 results looked discouraging at the time and they highlighted areas requiring development in methods; some of these have clearly been addressed to some extent in the three years between CSP2004 and this latest blind test. The search methods used and approaches taken for ranking of the computer-generated crystal structures are discussed in the following sections.

As well as the overall increase in successful predictions compared with previous years, it is important to note that these predictions were distributed amongst the four categories and there were successful predictions from at least two groups for each of the four crystal structures [four successes for each of molecules (XII) and (XIII), three for (XIV) and two for (XV)]. It is significant that there were three correct predictions for the flexible molecule. While molecule (XIV) might not have been as flexible as previous molecules in this category, the successes here do demonstrate that progress is being made in extending the generality of CSP methods to larger molecules. Furthermore, the new challenge of a cocrystal was not insurmountable, despite the added complexity of searching phase space with two independent molecules coupled with the issue of two possible conformations of the acid molecule.

Of course, the most impressive results from this blind test are those of Neumann, Leusen and Kendrick, who successfully predicted all four crystal structures, each as their first choice amongst their submitted predictions. Their calculations also produced the lowest RMSDs of all successful predictions for all four crystal structures, demonstrating that their method produces excellent matches to the true structures. One observation from previous blind tests was that there has not been one method which has been successful in general over the three categories of molecule. This group's results are certainly a striking improvement over what has been achieved before and indicate that generally applicable methods can be successful across the various types of molecules and crystals represented in the blind test.

5.2. The search problem

For each of the single-component systems, a few groups did not locate the observed crystal structures using their search method. The failure rate seems to have been lower for the flexible molecule (XIV) than in previous tests, which would contribute to the increased success overall for the flexible molecule. As expected, the cocrystal was the main problem – the increased search space was the main reason for including the new category of two-component crystals in this blind test and seven of the 12 groups who attempted predictions for this system did not locate the observed crystal structure in their search. The successes and failures in generating the observed structure were roughly evenly split between those who took either of the approaches described in §4.3.4 (*i.e.* treating the two molecules completely independently or starting with hydrogen-bonded dimers). There is certainly a gain in computational efficiency for the dimer approach in the crystal structure search, but the method did not prove more reliable than the full-blown independent molecule search in locating the observed structure.

It is not possible to completely analyse the rates of success of the various search methods without very long lists of the computer-generated crystal structures from each participating group. Nine of the 14 groups did submit extended lists of up to 100 predicted structures beyond their three predictions per molecule. However, in some cases these lists were not always long enough to fully assess the search method. Sometimes the energy model used to rank the structures performed poorly enough that the observed crystal structure was generated in the search, but ranked outside the best 100 structures and, so, was absent from the list. What we can say is that at least five groups (Neumann, Leusen & Kendrick; van Eijck; Day, Cooper, Cruz Cabeza & Hejczyk; Schmidt, van de Streek & Wolf; Boerrigter & Tan) located all four observed structures;

[2] Counting only the predictions that were completely blind. Limited structural information on one of the CSP2004 molecules was discovered part way through the blind test.

four of these groups used a variation on random sampling of structural variables to generate crystal structures, while the fifth employed Monte Carlo simulated annealing. Several other groups only missed one of the observed structures in their searches.

An analysis akin to that performed by van Eijck following CSP2004 (van Eijck, 2005) was performed on the extended lists of submitted structures; all participants' extended lists of structures were re-minimized in a high-quality common force field (that used by van Eijck) to remove structural differences owing to the use of different force fields. The lists from each participant were then compared. For the three single-component crystals [(XII), (XIII) and (XIV)], most of the low-energy structures were located by more than one participant. This indicates that, by combining the lists, the search space is probably sampled completely. However, individual participant's lists often missed important low-energy structures; this could indicate either incomplete sampling or that minima exist on the energy surface described by van Eijck's force field that are absent from the energy surface of other groups' energy model. For the cocrystal (XV), there was less overlap between participants' lists, suggesting that the sampling of crystal packing possibilities might not be complete, even after combining the extended lists. There are limitations to the conclusions that can be drawn from comparisons of structures generated using different energy models (van Eijck, 2005). However, these observations point to the conclusion that there is still room for improvement in methods for structure generation that will provide reliably complete sets of the low-energy crystal structures.

5.3. Ranking of the generated structures

As well as the computational and algorithmic challenge of providing an adequate sampling of crystal-packing phase space, it has also been clear from previous studies that the reliable ranking of the set of computer-generated crystal structures has been an obstacle for crystal structure prediction. This is because most molecules are found to have many distinct crystal packing possibilities within a small energy range (Day *et al.*, 2004), so that the energy differences between structures are often very small. As most methods rely on the ranking on lattice energy as a major, if not the only, part of the ranking process, the challenge of crystal structure prediction has been a driving force for the development of highly accurate methods for calculating the relative energies of crystal structures.

Until recently, developmental work has focused on improving atom–atom model potential methods, both in the functional form used and their parameterization (Price, 2008). Most of the participants in CSP2007 have based their predictions on atom–atom model potential calculations, and the nature of these models ranges from the fairly simple to quite elaborate anisotropic descriptions of atom–atom interactions. Most often, the atom–atom anisotropy is only included in the electrostatic term of the model potential and methods that have been used here are atomic multipole expansions of the charge density (four groups) and optimized core-shell models (one group). As in CSP2004, two groups treated the category (2) molecule with non-spherical atom–atom repulsion models, to account for the well known anisotropy in close contact distances for interactions involving halogen atoms (Nyburg & Faerman, 1985). Both of these resulted in the successful predictions of the observed crystal structure as the first choice prediction; Day used information from CSD distributions of atom–atom contacts to parameterize the atom–atom anisotropy, whereas Price, Misquitta, Karamertzanis and Welch's prediction employed a completely non-empirically derived model potential, where all atom–atom terms were derived from quantum-mechanics calculations on monomers and dimers (Misquitta *et al.*, 2008).

Beyond the atom–atom model potential approach, one group used Gavezzotti's semi-classical density sums (or PIXEL) method (Gavezzotti, 2003*a*,*b*), which avoids the partitioning into atom–atom terms, but discretizes the molecular electron density into a set of roughly 10^4 interaction sites. Additionally, a new class of approach that has been applied for the first time in the blind tests is the use of quantum-mechanical electronic structure based methods to treat the intermolecular as well as intramolecular energy. While the application of electronic structure calculations has been widely accessible for understanding molecular structure and properties for many years, applying these methods directly to the organic solid state presented further challenges – the computational expense and the now well known limitations of affordable methods (*i.e.* Hartree–Fock or density-functional theory) in modelling the dispersion interactions between molecules (Kristyán & Pulay, 1994; Hobza *et al.*, 1995). The computational problem is being resolved by a combination of algorithmic improvements and increased access to high-performance computational clusters that allow the application of such methods to the large numbers of crystal structures that must be considered for each molecule. Two groups used electronic structure methods in CSP2007 for energy minimization and ranking of their predicted crystal structures. The details of the methods are very different. Jose and Gadre applied the cardinality-guided molecular tailoring (CG-MTA) approach (Ganesh *et al.*, 2006), which has previously been applied to the energy minimization of large molecules. The crystal is described by a set of overlapping clusters of molecules. The total energy of the crystal is evaluated as the sum of the cluster energies minus the energies of the overlap regions, which are each evaluated at the HF/STO-3G level of theory. This method did not locate the experimental structure within the best three predictions for any of the molecules. A critical post analysis of (XII) and (XIV) as test cases showed that the method does locate crystal structures close to the experimental ones (see Tables 4 and 6). However, in the case of molecule (XIV), the energy difference from the lowest-energy generated structures clearly shows the inadequacy of the employed level of theory and basis set in accounting for the weak interactions in the crystals.

The other electronic structure based approach was that of Neumann, Leusen and Kendrick, who used periodic density-functional theory (DFT) calculations for their final energy minimizations. An important aspect of their calculations was that the DFT approach was empirically corrected to account for the poor treatment of the dispersion attraction between molecules by DFT. By supplementing the DFT energy from the *VASP* program (Kresse & Hafner, 1993; Kresse & Furtmüller, 1996; Kresse & Joubert, 1999) with an empirically parametrized atom–atom C_6R^{-6} correction (Neumann & Perrin, 2005), the most successful predictions were obtained: all four crystal structures were predicted as the global minimum on the static potential-energy surface. The approach is by far the most computationally demanding of any of the methods used for the final optimization and evaluation of lattice energies, so could only be applied to a limited number of crystal structures for each molecule. Whereas many methods involve the optimization of thousands of putative structures, only between 32 and 100 structures were optimized for each of the four targets. Therefore, another key component of the method was the use of highly accurate tailor-made force fields (Neumann, 2008) to provide an initial ranking of the structure and ensure that the observed structures were within this relatively small set.

Overall, we find that there is a rough correlation between how sophisticated a method is used for the final energy ranking and rates of success in this blind test. In fact, all successful predictions in this blind test were achieved by using more sophisticated methods for the final lattice-energy minimization than the traditional force fields with isotropic atom–atom interaction and atomic point charges.

Of course, the ranking of the stability of crystal structures should be based on calculated free energies, including contributions from lattice dynamics, rather than the potential energies of static configurations of atoms. However, the results here suggest that such dynamic contributions to the free energy are unimportant for the molecules included in CSP2007; those calculations that did include a lattice dynamical calculation of the free energy did not result in important re-rankings of structures. For the molecules studied here, the entropy differences between crystal structures were typically smaller than the lattice-energy differences.

5.4. Kinetics

As in the first three blind tests, the ranking of crystal structures has been based almost exclusively on calculated energies. Kinetic considerations are still fairly poorly understood: we know that growth conditions can be used to select polymorphs, but do not have a good understanding of how often observed crystal structures do not correspond to the lowest-energy possibility. Attempts have been made in the past to consider the kinetics of crystal growth in crystal structure prediction studies. The only major use of non-energetic criteria in this blind test was the synthon-based approach of Desiraju and Thakur, who considered hydrogen-bond motifs in the final ranking of crystal structures of the cocrystal (XV). The analysis of the arrangement of molecules in cocrystals of similar molecules in the CSD and in an in-house library of crystal structures (Thakur & Desiraju, 2008) highlighted the linear tetramer seen in the observed crystal structure (Fig. 4) as the most likely hydrogen-bond configuration for this pair of molecules. Therefore, predicted structures containing this synthon were given preference in the ranking. Statistics on the occurrence of hydrogen-bond motifs in known crystal structures will contain both energetic and kinetic information; observed crystal structures must be stable as well as kinetically accessible. Therefore, using such information in the ranking of predicted crystal structures could be a way of including some kinetic information. In this case, the approach successfully predicted the hydrogen-bonding motif. However, the observed structure was not present in the list, perhaps because of a poor choice of electrostatic model.[3] One might conclude that the analysis of crystal structures of similar molecules (or pairs of molecules, in the case of a cocrystal) can give valuable predictive information, but the requirement of a robust search method and accurate energy model remains.

The results presented in this blind test demonstrate important steps forward in the calculation of relative energies and, if we can reliably get these right, then we have a necessary starting point for considering the, perhaps less important, effects of entropy, nucleation and crystal growth on the crystal structure that results from a particular experiment. We were perhaps somewhat lucky in this blind test that the four target crystal structures all appear to be the most stable form on the potential energy surface, *i.e.* static $T = 0$ K energies.

6. Summary

Blind tests of crystal structure prediction of organic molecules have been carried out periodically over the past decade, serving as an objective evaluation of the state-of-the-art in crystal structure prediction methods and we believe that they are useful for those monitoring progress in the field. The latest such international investigation, CSP2007, has revealed major steps forward. Given the molecular diagrams of four targets (three molecules and the two components of a cocrystal), 13 successful predictions were achieved by the 14 participating research groups; here, a success means that a very good representation of the observed crystal was included in three allowed predictions per target. Amongst these successes were three predictions of the crystal structure of the flexible molecule – a category for which there has only been one correct prediction over the three first blind tests – and two predictions of the structure of a cocrystal of small molecules.

Much of the improved success over previous blind tests can be associated with improvements in how we calculate the relative energies of putative crystal structures. All successful predictions in this blind test were achieved by going beyond standard force fields, although several very different approa-

[3] When the CSP exercise was repeated in the post deadline period with electrostatic potential-derived charges instead of the default *COMPASS* force-field charges that were used in the official submission, the experimental structure was located and favourably ranked (Thakur & Desiraju, 2008).

ches are represented in the successful methods: dispersion-corrected periodic density-functional theory (Neuman, Leusen, Kendrick), empirically and non-empirically derived anisotropic atom–atom model potentials (Ammon; van Eijck; Day, Cooper, Cruz Cabeza, Hejczyk; Price, Misquitta, Karamertzanis & Welch), core-shell electrostatic models (Boerrigter, Tan) and the semi-classical density sums (SCDS-PIXEL) approach (Schweizer). The evidence here and in the published literature strongly suggests that off-the-shelf force fields will not be generally successful for the final energy ranking in crystal structure prediction.

The most successful results were achieved by applying plane-wave density-functional theory calculations, supplemented by an empirically derived C_6-dispersion energy term, to the final energy minimization and energy ranking of the putative crystal structures. While the method comes at an appreciable computational cost, all four observed crystal structures were predicted, each as the first ranked structure; such consistently correct predictions have not previously been achieved under blind test conditions. These impressive results support the approach of pursuing increasingly accurate relative lattice energies as worthwhile for crystal structure prediction, and this method has proven to be sufficiently accurate for the challenges presented in CSP2007.

These results demonstrate that the crystal structures of small rigid organic molecules are predictable, within the restrictions of the blind test. The molecules represented in the blind test are all quite small and have very little conformational flexibility; even the molecule chosen to represent the category designed to test methods for dealing with flexibility turned out to be fairly rigid. Furthermore, the contents of the asymmetric unit are known: $Z' \leq 1$ was given, as well as the stoichiometric ratio and non-ionicity of components in the cocrystal. Therefore, the whole challenge of crystal structure prediction of organic molecules is certainly not solved and the area is open to further development. However, even within these limits of applicability tested here, the results that have been achieved by some of the methods are of a sufficient level to be constructively applied to a number of problems in materials design and to improving our understanding of crystallization and polymorphism. Furthermore, we believe that it will be possible to extend some of the current methods to the successful prediction of the crystal structures of even more complex systems, broadening the possible areas of application of such methods.

We thank the CCDC for its support in organizing the blind test, and hosting of the workshop – in particular, Dr Frank Allen and Jean Mabbs. Many thanks are due to Professor A. L. Spek for accepting the role of independent referee, holding the experimental structures during the course of the blind test, and collecting submitted predictions from each of the participants. We are grateful to the crystallographers who supplied candidate structures: Professor Simon Parsons [molecule (XII)], Professor Doyle Britton [molecule (XIII)], Professor Alexander Blake [molecule (XIV)] and Professor Christer Aakeröy [molecule (XV)]. Also, thanks to Dr Sijbren Otto for selecting the target molecules from the list of candidates. G. M. Day thanks the Royal Society for a University Research Fellowship and the Pfizer Institute for Pharmaceutical Materials Science (University of Cambridge) for funding.

References

Aakeröy, C. B. (2007). Unpublished results.
Aakeröy, C. B., Beffert, K., Desper, J. & Elisabeth, E. (2003). *Crystal Growth Des.* **3**, 837–846.
Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Anghel, A. T., Day, G. M. & Price, S. L. (2002). *CrystEngComm*, **4**, 348–355.
Bazterra, V. E., Thorley, M., Ferraro, M. B. & Facelli, J. C. (2007). *J. Chem. Theory Comput.* **3**, 201–209.
Beyer, T., Lewis, T. & Price, S. L. (2001). *CrystEngComm*, **3**, 178–212.
Blake, A. J., Clark, J. S. & Conroy, J. (2007). Unpublished results.
Boese, R. & Nussbaumer, M. (1994). Crystallographic Symposia 7, Transformations and Interactions in Organic Crystal Structure. International Union of Crystallography.
Borreguero, J. & Skolnick, J. (2007). *Proteins Struct. Funct. Bioinf.* **68**, 48–56.
Brillante, A., Bilotti, I., Della Valle, R. G., Venuti, E. & Girlando, A. (2008). *CrystEngComm*, **10**, 937–946.
Britton, D. (2008). Private communication (CCDC711859). CCDC, Cambridge, England.
Brooks III, C. L., Onuchic, J. N. & Wales, D. J. (2001). *Science*, **293**, 612–613.
Busing, W. R. (1981). *WMIN*. Report ORNL-5747. Oak Ridge National Laboratory, Tennessee, USA.
Case, F. H., Brennan, J., Chaka, A., Dobbs, K. D., Friend, D. G., Frurip, D., Gordon, P. A., Moore, J., Mountain, R. D., Olson, J., Ross, R. B., Schiller, M. & Shen, V. K. (2007). *Fluid Phase Equilib.* **260**, 153–163.
Chisholm, J. A. & Motherwell, S. (2005). *J. Appl. Cryst.* **38**, 228–231.
Coombes, D. S., Catlow, C. R. A., Gale, J. D., Rohl, A. L. & Price, S. L. (2005). *Cryst. Growth Des.* **5**, 879–885.
Cooper, T. G., Jones, W., Motherwell, W. D. S. & Day, G. M. (2007). *CrystEngComm*, **9**, 595–602.
Cruz-Cabeza, A. J., Day, G. M. & Jones, W. (2008). *Chem. Eur. J.* **14**, 8830–8836.
Cruz Cabeza, A. J., Day, G. M., Motherwell, W. D. S. & Jones, W. (2006). *J. Am. Chem. Soc.* **128**, 14466–14467.
Day, G. M., Chisholm, J., Shan, N., Motherwell, W. D. S. & Jones, W. (2004). *Cryst. Growth Des.* **4**, 1327–1340.
Day, G. M. *et al.* (2005). *Acta Cryst.* B**61**, 511–527.
Day, G. M. & Motherwell, W. D. S. (2006). *Cryst. Growth Des.* **6**, 1985–1990.
Day, G. M., Motherwell, W. D. S. & Jones, W. (2005). *Cryst. Growth Des.* **5**, 1023–1033.
Day, G. M., Motherwell, W. D. S. & Jones, W. (2007). *Phys. Chem. Chem. Phys.* **9**, 1693–1704.
Day, G. M. & Price, S. L. (2003). *J. Am. Chem. Soc.* **125**, 16434–16443.
Della Valle, R. G., Venuti, E., Brillante, A. & Girlando, A. (2008). *J. Phys. Chem. A*, **112**, 6715–6722.
Dey, A., Kirchner, M. T., Vangala, V. R., Desiraju, G. R., Mondal, R. & Howard, J. A. K. (2005). *J. Am. Chem. Soc.* **127**, 10545–10559.
Dey, A., Pati, N. N. & Desiraju, G. R. (2006). *CrystEngComm*, **8**, 751–755.
Dzyabchenko, A. V. (1994). *Acta Cryst.* B**50**, 414–425.
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 (2005). *Acta Cryst.* B**61**, 528–535.
Eijck, B. P. van & Kroon, J. (2000). *Acta Cryst.* B**56**, 535–542.

Forster, T., Oswald, I. D. H. & Parsons, S. (2007). Unpublished results.
Ganesh, V., Dongare, R. K., Balanarayan, P. & Gadre, S. R. (2006). *J. Chem. Phys.* **125**, 104109.
Gavezzotti, A. (1999–2000). *ZIP-PROMET*. University of Milano, Italy.
Gavezzotti, A. (2003*a*). *CrystEngComm*, **5**, 429–438.
Gavezzotti, A. (2003*b*). *CrystEngComm*, **5**, 439–446.
Gavezzotti, A. (2004). *OPIX*. University of Milano, Italy.
Hobza, P., Sponer, J. & Reschel, T. (1995). *J. Comput. Chem.* **16**, 1315–1325.
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.* A**53**, 225–235.
Holden, J. R., Du, Z. Y. & Ammon, H. L. (1993). *J. Comput. Chem.* **14**, 422–437.
Karamertzanis, P. G. & Pantelides, C. C. (2005). *J. Comput. Chem.* **26**, 304–324.
Karamertzanis, P. G. & Pantelides, C. C. (2007). *Mol. Phys.* **105**, 273–291.
Karamertzanis, P. G. & Price, S. L. (2006). *J. Chem. Theory Comput.* **2**, 1184–1199.
Karamertzanis, P. G., Raiteri, P., Parrinello, M., Leslie, M. & Price, S. L. (2008). *J. Phys. Chem. B*, **112**, 4298–4308.
Khalili, M., Liwo, A. & Scheraga, H. A. (2006). *J. Mol. Biol.* **355**, 536–547.
Kresse, G. & Furtmüller, J. (1996). *Phys. Rev. B*, **54**, 11169.
Kresse, G. & Hafner, J. (1993). *J. Phys. Rev. B*, **47**, 558.
Kresse, G. & Joubert, D. (1999). *Phys. Rev. B*, **59**, 1758.
Kristyán, S. & Pulay, P. (1994). *Chem. Phys. Lett.* **229**, 175–180.
Liwo, A., Khalili, M., Czaplewski, C., Kalinowski, S., Oldziej, S., Wachucik, K. & Scheraga, H. A. (2007). *J. Phys. Chem. B*, **111**, 260–285.
Llinàs, A., Glen, R. C. & Goodman, J. M. (2008). *J. Chem. Inf. Model.* **48**, 1289–1303.
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.* B**56**, 697–714.
Misquitta, A. J. & Stone, A. J. (2007). *CamCASP*. Cambridge, England.
Misquitta, A. J., Welch, G. W. A., Stone, A. J. & Price, S. L. (2008). *Chem. Phys. Lett.* **456**, 105–109.
Mooij, W. T. M., van Eijck, B. P. & Kroon, J. (1999). *J. Phys. Chem. A*, **103**, 9883–9890.
Motherwell, W. D. S. *et al.* (2002). *Acta Cryst.* B**58**, 647–661.
Moult, J., Fidelis, K., Kryshtafovych, A., Rost, B., Hubbard, T. & Tramontano, A. (2007). *Proteins Struct. Funct. Bioinf.* **69**, 3–9.
Neumann, M. A. (2008). *J. Phys. Chem. B*, **112**, 9810–9829.
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.
Nicholls, A., Mobley, D. L., Guthrie, J. P., Chodera, J. D., Bayly, C. I., Cooper, M. D. & Pande, V. S. (2008). *J. Med. Chem.* **51**, 769–779.
Nyburg, S. C. & Faerman, C. H. (1985). *Acta Cryst.* B**41**, 274–279.
Oldziej, S. *et al.* (2005). *Proc. Natl. Acad. Sci.* **102**, 7547–7552.
Pillardy, J., Arnautova, Y. A., Czaplewski, C., Gibson, K. D. & Scheraga, H. A. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 12351–12356.
Price, S. L. (2008). *Int. Rev. Phys. Chem.* **27**, 541–568.
Raiteri, P., Martonák, R. & Parrinello, M. (2005). *Angew. Chem. Int. Ed.* **44**, 3769–3773.
Sarma, J. A. R. P. & Desiraju, G. R. (2002). *Cryst. Growth Des.* **2**, 93–100.
Schmidt, M. U. & Englert, U. (1996). *J. Chem. Soc. Dalton Trans.* pp. 2077–2082.
Schmidt, M. U. & Kalkhof, H. (1997). *CRYSCA*. Clariant GmbH, Frankfurt.
Signorini, G. F., Righini, R. & Schettino, V. (1991). *Chem. Phys.* **154**, 245.
Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.
Tan, J. S., Boerrigter, S. X. M., Scaringe, R. P. & Morris, K. R. (2009). *J. Comput. Chem.* **30**, 733–742.
Thakur, T. S. & Desiraju, G. R. (2008). *Cryst. Growth Des.* **8**, 4031–4044.
Timmermans, J. (1922). *Bull. Soc. Chim. Belg.* **31**, 389.
Verwer, P. & Leusen, F. J. J. (1998). *Rev. Comput. Chem.* **12**, 327–365.
Willock, D. J., Price, S. L., Leslie, M. & Catlow, C. R. A. (1995). *J. Comput. Chem.* **16**, 628–647.