A third blind test of crystal structure prediction

G. M. Day,[a]* W. D. S. Motherwell,[b] H. L. Ammon,[c] S. X. M. Boerrigter,[d] R. G. Della Valle,[e] E. Venuti,[e] A. Dzyabchenko,[f] J. D. Dunitz,[g] B. Schweizer,[g] B. P. van Eijck,[h] P. Erk,[i] J. C. Facelli,[j] V. E. Bazterra,[j] M. B. Ferraro,[j] D. W. M. Hofmann,[k] F. J. J. Leusen,[l] C. Liang,[m] C. C. Pantelides,[n] P. G. Karamertzanis,[n] S. L. Price,[o] T. C. Lewis,[o] H. Nowell,[o] A. Torrisi,[p] H. A. Scheraga,[q] Y. A. Arnautova,[q] M. U. Schmidt[k] and P. Verwer[r]

[a]The Pfizer Institute for Pharmaceutical Materials Science, University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, England, [b]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England, [c]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742-2021, USA, [d]School of Pharmacy and Pharmacal Sciences, Purdue University, West Lafayette, Indiana, USA, [e]Dipartimento di Chimica Fisica e Inorganica and INSTM-UdR, Universita di Bologna, Viale Risorgimento 4, I-40136 Bologna, Italy, [f]Karpov Institute of Physical Chemistry, Vorontsovo pole 10, 103064 Moscow, Russia, [g]Organic Chemical Laboratory, ETH-Zurich, CH-8093 Zurich, Switzerland, [h]Department of Crystal and Structural Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands, [i]Performance Chemicals Research, BASF AG, 67056 Ludwigshafen, Germany, [j]Center for High Performance Computing, University of Utah, Salt Lake City, Utah 84112-0190, USA, [k]Institute of Inorganic and Analytical Chemistry, University of Frankfurt, Marie Curie Str. 11, 60439 Frankfurt, Germany, [l]Institute of Pharmaceutical Innovation, University of Bradford, Bradford BD7 1DP, England, [m]Accelrys Inc., San Diego, California, USA, [n]Center for Process Systems Engineering, Department of Chemical Engineering and Chemical Technology, Imperial College London, South Kensington Campus, London SW7 2AZ, England, [o]Department of Chemistry, University College London, 20 Gordon St, London WC1H 0AJ, England, [p]The Royal Institution, 21 Albermarle St, London W1X 4BS, England, [q]Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA, and [r]Solid State Chemistry Group and CMBI, University of Nijmegen, The Netherlands

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

Following the interest generated by two previous blind tests of crystal structure prediction (CSP1999 and CSP2001), a third such collaborative project (CSP2004) was hosted by the Cambridge Crystallographic Data Centre. A range of methodologies used in searching for and ranking the likelihood of predicted crystal structures is represented amongst the 18 participating research groups, although most are based on the global minimization of the lattice energy. Initially the participants were given molecular diagrams of three molecules and asked to submit three predictions for the most likely crystal structure of each. Unlike earlier blind tests, no restriction was placed on the possible space group of the target crystal structures. Furthermore, $Z' = 2$ structures were allowed. Part-way through the test, a partial structure report was discovered for one of the molecules, which could no longer be considered a blind test. Hence, a second molecule from the same category (small, rigid with common atom types) was offered to the participants as a replacement. Success rates within the three submitted predictions were lower than in the previous tests – there was only one successful prediction for any of the three 'blind' molecules. For the 'simplest' rigid molecule, this lack of success is partly due to the observed structure crystallizing with two molecules in the asymmetric unit. As in the 2001 blind test, there was no success in predicting the structure of the flexible molecule. The results highlight the necessity for better energy models, capable of simultaneously describing conformational and packing energies with high accuracy. There is also a need for improvements in search procedures for crystals with more than one independent molecule, as well as for molecules with conformational flexibility. These are necessary requirements for the prediction of possible thermodynamically favoured polymorphs. Which of these are actually realised is also influenced by as yet insufficiently understood processes of nucleation and crystal growth.

1. Introduction

This paper reports the results of a third blind test of crystal structure prediction (CSP), hosted by the Cambridge Crystallographic Data Centre and which we will refer to as CSP2004. Researchers have been struggling with the challenge of reliably predicting the observed crystal structure(s) of a given molecule, which shares many of its difficulties with the prediction of protein folding. Similarities and differences between the two types of problem have recently been discussed (Dunitz & Scheraga, 2004). Both problems depend on our ability to calculate the most stable arrangements of large molecular aggregates from general principles and both

are also subject to mostly unknown kinetic influences. Blind tests have been reported in both crystal structure prediction (Lommerse et al., 2000; Motherwell et al., 2002) and protein folding (Moult et al., 1999) communities, and both have generated considerable interest. Therefore, the present collaboration was organized as a follow-up to the 1999 (Lommerse et al., 2000) and 2001 (Motherwell et al., 2002) CSP blind tests and the results presented here should be digested in the context of the earlier studies.

Periodically repeating the blind test can assess progress in the field, but the necessarily small number of molecules studied in each blind test means that the choice of one slightly more difficult or easier molecule (whose difficulty often cannot be judged until the predictions have been made) can have as much or more influence on the results as the methodological advances made since the last such project. In total, 11 molecules have now been studied under blind test conditions [the four molecules in the current blind test are numbered (VIII)–(XI)] and the successes and failures in these overall results point to the areas where improved methodologies are most urgent.

2. Organization and approach

The organization of this latest collaboration, CSP2004, was essentially the same as the first two, which have already been published (CSP1999, Lommerse et al., 2000, and CSP2001, Motherwell et al., 2002). Invitations to participate were sent to ca 25 research groups known to be active in the field and 18 of these groups participated. We initially intended to include three molecules in the test, retaining the same three categories of molecules as in CSP1999 and CSP2001:

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

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

(3) molecules with several torsional degrees of flexibility. To be suitable, a crystal structure had to be of high quality with all atoms located and no disorder.

All methods of crystal structure prediction involve the generation of possible crystal structures, followed by some form of ranking of their likelihood of being observed. Results of the CSP2001 blind test suggested that the generation of all the likely crystal structures in the most common space groups, with one rigid molecule in the asymmetric unit, is no longer a major problem. Therefore, the challenges for search methods are flexible molecules, less common space groups and crystals with more than one independent molecule. We therefore relaxed the limitations on space group that were used in the first two blind tests – candidate crystals for CSP2004 were allowed to be in any space group and only restricted to $Z' < 3$. Therefore, unlike the first two blind tests, participants had to contemplate the possibility that the molecules crystallize in uncommon space groups, or with $Z' = 2$.

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. After considerable effort, we collected six candidates for category 1, two for category 2 and six for category 3. (Here, we make a plea to the crystallographic community; suitable structures of molecules that fit the three categories were difficult to find. We would be grateful for contacts from crystallographers with suitable unpublished structures for the next blind test.)[1]

To choose which of these molecules were used in the blind test, we took sketches of the molecular structures to a second independent colleague, this time a non-crystallographer with no access to the crystal structures – Professor Jeremy Sanders, University Chemical Laboratory, Cambridge. He was asked to choose one molecule from each of the three categories. The contributing crystallographers agreed to postpone the publication of these structures and they were held until April 2004 by Professor Spek. The molecular diagrams, as shown in the first three rows of Table 1, were sent by email to all of the participants on 1 October 2003. Participants were asked to send their predictions for each molecule to Professor Spek by 1 April 2004. We kept the format the same as in previous blind tests, allowing each participating group to submit three predictions for each molecule, with reasons for the prediction. Analysis of longer lists of generated structures provided useful insight following the CSP2001 workshop, so participants were encouraged to submit extended lists of their other predicted structures separately from their 'official' three, but preferably in order of likelihood (i.e. structures 4, 5, etc.).

On 31 December 2003 we became aware[2] that, while unpublished, the crystal structure of molecule VIII (hydantoin) had been presented in a poster session of a meeting of the British Crystallographic Association in March 2002 (Yu & Schwalbe, 2002). Many of the participants had already spent considerable effort on their calculations for this molecule, so we agreed that predictions should still be submitted for molecule (VIII), with participants indicating whether information contained in the conference abstract (space group, unit-cell dimensions and a diagram of the hydrogen bonding, Yu & Schwalbe, 2002) had influenced their predictions. While most of the participants agreed to submit their predictions without the use of information in the partial structure report, the predictions for molecule (VIII) could no longer be considered truly 'blind'. Hence, a second molecule from category 1 was chosen as a replacement – its molecular diagram and conditions of crystallization (last row of Table 1) were circulated on 7 January 2004. The timeline for the blind test and dates of the workshop could not be rearranged, so predictions for this molecule [henceforth referred to as molecule (XI)] had to be submitted along with (VIII) to (X). Thus, only half the time available for the other targets was available for predictions on molecule (XI).

On 5 April 2004 the experimental crystal structures of all four molecules were circulated to each participant for post-analysis of their predictions and, on 4 and 5 May 2004 a workshop was held at the Cambridge Crystallographic Data Centre to discuss the results.

3. Methodologies

Methodologies varied significantly amongst the 18 groups who submitted predictions. For brevity, we provide very short descriptions of the methodologies in Table 2. References are provided for most of the methods used and details specific to the work on the CSP2004 molecules are available in the supplementary information.[3]

Although the techniques vary, all methods of crystal structure prediction involve three general steps:

(i) building a molecular structure from the chemical bonding diagram;

(ii) searching for plausible packing arrangements of the molecule;

(iii) ranking the generated structures by some criteria – usually energy – although additional or alternate criteria are sometimes used.
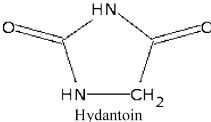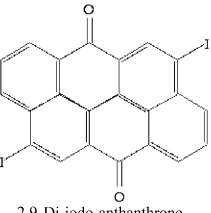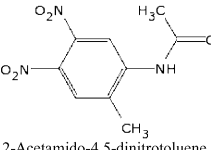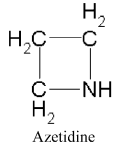
There have been no major advances in step (i) since CSP2001. Some new algorithms for structure generation (step ii) have been applied since the last blind test (particularly the development of search algorithms designed to take advantage of modern computing technology). The refinement of energy models for step (iii) continues partly through further para-meterization of empirically derived model potentials, but also the formulation of new models for intermolecular interactions, *e.g.* the semiclassical density sums, or PIXEL, approach (Gavezzotti, 2002, 2003a). No major progress has been made since CSP2001 in defining the non-energetic criteria of ranking putative structures.

3.1. Methods of generating the molecular structure

Two approaches can be used to treat the molecular structure during crystal structure predictions; the molecule can be treated as flexible and varied along with the crystal structure, usually through bond stretching, angle bending and torsional terms in a force-field. Alternatively, the molecule can be treated as rigid throughout the calculations. In the latter approach, the molecular structure is usually generated *via* gas-phase quantum chemical calculations and it is assumed that the environment of the crystal has no effect on the bond lengths, angles and torsions. Here, the choice of method used to generate the rigid molecular structure can be crucially important – the effect of assumed molecular structure on crystal energy calculations can be large (Beyer & Price, 2000).

---

[3] Supplementary data for this paper are available from the IUCr electronic archives (Reference: DE5014). Services for accessing these data are described at the back of the journal.

Table 1
Diagrams and crystallization conditions for the molecules of CSP2004.

Molecule		Crystallization conditions
(VIII)	Hydantoin	The compound was dissolved in methanol with heat, filtered and evaporated slowly at room temperature
(IX)	2,9-Di-iodo-anthanthrone	The compound was dissolved in boiling *o*-dichlorobenzene at 453 K. The solution was cooled to 433 K, put into a heating bath, and slowly cooled from 433 to 303 K with a cooling rate of 0.5 K h^{-1}
(X)	2-Acetamido-4,5-dinitrotoluene	Crystallization conditions not known
(XI)	Azetidine	The sample was crystallized from the pure liquid in a capillary using Boese's (Boese & Nussbaumer, 1994) laser-assisted zone-refinement method, at $T = 170$ K

3.2. Generating trial crystal structures

Methods for generating crystal packing arrangements are diverse and those used during this and earlier blind tests include the stepwise building-up of dimers and layers, grid-based systematic searches, genetic algorithms, Monte Carlo and random searches. Since random sampling often exhibits gaps and clusters of points, some participants in CSP2004 instead chose to follow a low-discrepancy Sobol' sequence (Press *et al.*, 1992; Sobol', 1967), which ensures a more uniform, and thus more efficient, sampling. Another difference between search methods is the use of symmetry; because of the large search space involved in finding all possible crystal structures, many methods make use of space-group symmetry to reduce the computing time required. For these methods, each space group (and Z') is searched separately, so the choice of which symmetries to consider is crucial. Ideally, all space groups would be searched, with varying numbers of molecules in the asymmetric unit for each. However, computing time is finite and the majority of known structures of organic molecules populate a handful of space groups, so it is usual to

Table 2
Summary of methodologies.

Contributor	Molecules attempted	Programs	Ref.	Search generation	Space groups considered	Molecular model	Lattice energy/fitness function		Other criteria used to select submissions
							Electrostatic	Other	
Ammon	(VIII)–(XI)	*MOLPAK*, *ROTPAK*, *WMIN*, *PMC*	(a)	Grid-based systematic	$Z' = 1$ in $P1$, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, $P2_12_12$, $C2/c$, $Pbca$, $Pbcn$, $Pna2_1$, $Pca2_1$, $C2$, Cc	(VIII), (IX), (XI) rigid, (X) flexible	Atomic charges	Specifically fitted 6-exp	
Boerrigter	(VIII)–(XI)	*Cerius*2 Polymorph Predictor	(b)	Monte Carlo simulated annealing	No symmetry information used – $P1$ with $Z' = 1$, 2, 4	Rigid for search, flexible for energy minimization (VIII)–(X)	Atomic charges	Dreiding 6-12	Structural isotropy and density
Day	(VIII)–(XI)	*Cerius*2 Polymorph Predictor, *DMAREL*	(c)	Monte Carlo simulated annealing	$Z' = 1$ [(VIII)–(XI)] and 2 [(VIII), (XI)] in $P1$, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, $C2/c$, $Pbca$, $Pbcn$, $Pna2_1$, $Pnma$; $Z' = 1$ in $Pca2_1$, $P2_12_12$, Cc, $C2$, $C2/m$, $P2_1/m$, $P2/c$	Rigid throughout	Atomic multipoles [(VIII), (IX), (XI)] or charges (X)	Empirical 6-exp; specifically fitted anisotropic 6-exp (IX)	Free energy [(XI) only]
Della Valle & Venuti	(VIII), (IX), (XI)	*WMIN*, *IONIC*, *PLATON*	(d)	Structures generated using a low discrepancy Sobol' sequence	$Z' = 2$, $P1$	Rigid throughout	Atomic charges	Empirical 6-exp	Free energy
Dzyabchenko	(IX)–(XI)	*PMC*	(e)	Grid-based systematic	$Z' = 1$, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, $C2/c$, $Pbca$, $Pna2_1$, Cc, $C2$, Pc, $Pca2_1$, $Pbcn$	(VIII), (IX), (XI) rigid, (X) flexible	Atomic charges + bond dipoles	Empirical 6-12; specifically fitted anisotropic iodine (IX)	
van Eijck	(VIII)–(XI)	*UPACK*, *XTINKER*	(f)	Randomly generated starting structures	$Z' = 1$ in $P1$, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, $C2/c$, $Pbca$, $Pbcn$, $Pna2_1$, $Pca2_1$, Cc, $C2$, Pc, $Z' = 2$ in $P1$, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$	Flexible throughout	Atomic multipoles (VIII), (XI) or charges (IX), (X)	Empirical 6-exp	
Erk	(VIII)–(XI)	Systematic Search, *Cerius*2 Polymorph Predictor	(g)	Grid-based systematic and Monte Carlo simulated annealing	(VIII)–(XI) $Z' = 1$ in $P\bar{1}$, $P2_1/c$, $C2/c$, $P2_12_12_1$, (VIII) $Z' = 2$ and 4 in $P1$, and (XI) $Z' = 2$, 4, 8 in $P1$	Rigid for search, partly flexible for energy minimization	Atomic charges	Dreiding 6-12 (VIII), (IX), (X); Dreiding 6-12, CVFF300 6-12 and Momany 6-12 (XI)	

Table 2 (continued)

Contributor	Molecules attempted	Programs	Ref.	Search generation	Space groups considered	Molecular model	Lattice energy/fitness function		Other criteria used to select submissions
							Electrostatic	Other	
Facelli, Bazterra & Ferraro	(VIII), (X), (XI)	*MGAC*	(h)	Modified genetic algorithm	$Z' = 1$ and 2, in $P1$, $P\bar{1}$, $P2_1$, $C2$, Pc, Cc, $P2_1/c$, $C2/c$, $P2_12_12_1$, $Pca2_1$, $Pna2_1$, $Pbcn$, $Pbca$, $Pnma$	Flexible throughout	Atomic charges	GAFF 6-12	
Hofmann	(VIII)–(XI)	*FlexCryst*	(i)	Random search with calibrated cell	$Z' = 1$ in $P\bar{1}$, $P2_1$, $C2$, Cc, $P2_1/c$, $C2/c$, $Pca2_1$, $Pna2_1$, $Pbcn$, $Pbca$	Rigid	Trained potentials		
Leusen	(VIII)–(XI)	Materials Studio Polymorph	(j)	Monte Carlo simulated annealing	$Z' = 1$ in $P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $C2/c$, $P2_1$, $Pbca$, $Pna2_1$, Cc, $Pbcn$, $C2$, $P1$ and $Pca2_1$, $Z' = 2$ in $P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $C2/c$ and $P2_1$	Rigid for search, flexible for energy minimization	Atomic charges	CVFF 6-12	
Liang	(VIII)–(XI)	Materials Studio Polymorph	(j)	Monte Carlo simulated annealing	$Z' = 1$ in $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, $C2/c$, $Pbca$, $Pbcn$, $Pna2_1$, $Pca2_1$, Cc, $C2$, $P1$	Rigid for search, flexible for energy minimization	Atomic charges	Dreiding 6-12 (VIII, IX); Compass 6-9 (X); PCFF 6-9 (XI)	
Motherwell	(IX), (XI)	*RANCEL*	(k)	Genetic algorithm	Common space groups, $Z' = 1$	Rigid throughout	None	Empirical 6-exp	CSD based criteria
Pantelides & Karamertzanis	(VIII)–(XI)	Crystal Predictor	(l)	$O(10^5)$ structures generated using a low-discrepancy Sobol' sequence	$Z' = 1$ [(VIII)–(XI)] and 2 [(VIII), (IX), (XI)] in 59 most common space groups	(VIII), (IX), (XI) rigid, (XI) partly flexible	Optimal satellite charge model	Empirical 6-exp	
Price, Lewis, Nowell & Torrisi	(VIII)–(XI)	*MOLPAK*, *DMAREL*	(m)	Grid-based systematic	$Z' = 1$ in $P1$, $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$, $P2_12_12$, $C2/c$, $Pbca$, $Pbcn$, $Pna2_1$, $Pca2_1$	Rigid throughout	Atomic multipoles	Empirical 6-exp	Morphological and second derivative properties
Torrisi					$Pbcn$, $Pna2_1$, $Pca2_1$				Derivative properties
Scheraga & Arnautova	(VIII)–(XI)	*CRYSTALG*	(n)	Conformation family Monte Carlo (CFMC)	No symmetry information used – $P1$ with varying Z (= 2, 4, 8)	(VIII), (IX), (XI) rigid, (X) partly flexible	Atomic charges	Empirical 6-exp, specifically fitted anisotropic iodine (IX)	
Schmidt	(X), (XI)	*CRYSCA*	(o)	Randomly generated structures	$Z' = 1$, $P1$, $P\bar{1}$, $P2_1$, Cc, $P2_1/c$, $C2/c$, $P2_12_12_1$, $Pna2_1$, $Pbca$	(XI) flexible, (XI) rigid	Atomic charges	Empirical 6-exp	

Table 2 (continued)

Contributor	Molecules attempted	Programs	Ref.	Search generation	Space groups considered	Molecular model	Lattice energy/fitness function		Other criteria used to select submissions
							Electrostatic	Other	
Schweizer & Dunitz	(VIII), (X), (XI)	*ZIP-PROMET*, *PIXEL*	(*p*)	Stepwise construction of dimers and layers	$Z' = 1$ in $P\bar{1}$, $P2_1$, $P2_1/c$, $C2/c$, $P2_12_12_1$, $Pbca$	Rigid throughout	Energy minimization with exp-6 UNI potential and energy calculations with pixel-based method		
Verwer	(VIII), (IX), (XI)	Materials Studio Polymorph	(*j*)	Monte Carlo simulated annealing	$Z' = 1$ in $P1$, $P\bar{1}$, $P2_1$, $P2_1/c$, $C2/c$, $C2$, Cc, $P2_12_12_1$, $Pbca$, $Pna2_1$, $Pca2_1$, $Pbcn$	Rigid for search, flexible for energy minimization	Atomic charges	Dreiding 6-12 (VIII), (IX); Compass 6-9 (XI)	

References: (*a*) Holden *et al.* (1993), Busing (1981), Dzyabchenko *et al.* (1999); (*b*) Accelrys Inc. (1997); (*c*) Accelrys Inc. (1997), Price *et al.* (2001), Willock *et al.* (1995); (*d*) Della Valle *et al.* (2003), Busing (1981), Signorini *et al.* (1991), Spek (1980–2004); (*e*) Dzyabchenko (2004); (*f*) van Eijck (2002), van Eijck & Kroon (2000), Mooij *et al.* (1999), van Eijck *et al.* (2001*a*); (*g*) Erk (1999), Accelrys Inc. (1997); (*h*) Bazterra *et al.* (2002*a,b*, 2004*a,b*); (*i*) Hofmann & Lengauer (1997), Hofmann & Apostolakis (2003), Kuleshova *et al.* (2005); (*j*) Accelrys Inc. (2004); (*k*) Motherwell (2001); (*l*) Karamertzanis & Pantelides (2004, 2005); (*m*) Holden *et al.* (1993), Price *et al.* (2001), Willock *et al.* (1995); (*n*) Pillardy *et al.* (2001), Arnautova *et al.* (2003*a,b*); (*o*) Schmidt & Englert (1996), Schmidt & Kalkhof (1997); (*p*) Gavezzotti (1999–2000, 2004).

restrict searches to a set of the most commonly observed. Much more extensive searches (*e.g.* involving hundreds of thousands of candidate structures in large numbers of space groups) are now carried out by some researchers by exploiting parallel computation on distributed computer networks. Three participants did not use the information about the most common space groups to generate trial structures. Instead, they considered the space group $P1$ with different numbers (up to eight) of independent molecules and checked for symmetry in the final structures, using *CRYCOM* (Dzyabchenko, 1994), *PLATON* (Spek, 1980–2004) or the symmetry finder in *Cerius*2 (Accelrys Inc., 1997).

3.3. Ranking of structures

As in CSP1999 and CSP2001, the final ranking of crystal structures is still almost exclusively based on the minimized lattice energy, so varies from participant to participant depending on the choice of model for the intra- and intermolecular energies. The lattice energy is 'temperature-less', in that entropic contributions to the free energy are by and large neglected, although some groups have improved the thermodynamic model to include estimates of the vibrational contributions to the free energy. Others have introduced additional criteria (such as calculated mechanical properties and morphologies or structural isotropy) as an initial attempt to consider kinetic differences between predicted structures. Previous blind tests and many other studies show that dozens of possibilities often exist within a few kJ mol^{-1}, sometimes hundreds within, say, 5 kJ mol^{-1} of the calculated global minimum (Day *et al.*, 2004). Therefore, the required accuracy of calculated relative energies is very high and any additional or alternate criteria that can reliably discriminate between likely and unlikely structures will be valuable. Still other participants have chosen scoring criteria based on statistical comparisons to the known structures in the Cambridge Structural Database as an alternative to lattice energy minimization.

3.4. Treatments of the molecular flexibility in (X)

We summarize the various methods of treating the flexible molecule (X) in Table 3, adopting the nomenclature shown in Fig. 1 for the six exocyclic torsion angles.

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 many of the participants.

4.1. Experimental structures

Small rigid molecules fitting our criteria for category 1 with no published crystal structure are now uncommon. Therefore, a colleague was asked to grow crystals of hydantoin [molecule (VIII)], a readily available molecule with no reported crystal structure, as a candidate for CSP2004. The compound was obtained from Aldrich and recrystallized from methanol by slow evaporation.[4] The molecules are linked into chains *via* pairs of N—H$\cdots$O$=$C ($R_{NO} = 2.913$ Å) interactions forming centrosymmetric motifs (Fig. 2, left), which pack into a layered crystal structure in the space group $C2/c$ (Fig. 2, right). The

---

[4] We thank Mr Andrew Trask and Dr David J. Watkin for hydantoin crystallization and X-ray structure determination, respectively.

Table 3
Summary of conformational treatment of molecule (X).

| Contributor	Conformational treatment	Treatment of intramolecular energy
Ammon	Limited flexibility of φ allowed in the search	Semi-empirical PM3 calculations
Boerrigter	Searches were started from two conformers: $\varphi = -50.4°$; $\omega = -4.5°$; $\psi = -34.7°$; $\tau_1 = -170.7°$; $\tau_2 = 39.8°$; $\tau_3 = 38.0°$ and $\varphi = -151.2°$; $\omega = 175.7°$; $\psi = 32.9°$; $\tau_1 = -170.7°$; $\tau_2 = 39.8°$; $\tau_3 = 38.0°$	Dreiding 2.21 force field
Day	Searches were performed with a limited series of low-energy rigid conformations, varying φ in the range $-45°$ to $+45°$ with the remaining torsion angles optimized for the isolated molecule.	DFT calculations
Dzyabchenko	The molecule was treated as a set of six linked rigid fragments, linked by quadratic potentials and allowed to vary independently. Searches were started with a planar conformation, except for the two nitro groups, which were rotated 90° from the molecular plane.	Specifically derived torsion potentials for aromatic π-energy.
van Eijck	ω was fixed at 180°, torsional parameters for φ, τ_2 and τ_3 fitted to analogues of X. φ, τ_2 and τ_3 were searched randomly, along with cell parameters and molecular positions. τ_1 and ψ set to zero in the search. All torsion angles were allowed to vary during energy minimization.	Specifically derived bond-stretch, angle and torsion potentials.
Erk	Searches were started from two conformers: $\tau_1 = 180°$, $\tau_2 = \tau_3 = 0°$, $\omega = 180°$ and $\varphi = \pm 30°$.	Dreiding 2.21 force field
Facelli, Bazterra & Ferraro	The torsion angles were not allowed to vary due to a typo in the input file; ω was incorrectly fixed at 0°.	GAFF (Generic Amber Force Field)
Hofmann	Searches were started from four conformations after conformational analysis with *MOMO* (Gemmel *et al.*, 1999).	
Leusen	All searches were started from a 'saddle-point' conformation ($\varphi = 0°$, $\omega = 180°$, $\psi = 180°$, $\tau_1 = 180°$, $\tau_2 = 90°$, $\tau_3 = 90°$), as any relevant conformation can be reached from this starting point during lattice energy minimization with flexible molecules.	CVFF force field
Liang	All searches were started from a 'saddle point' conformation ($\varphi = 0°$, $\omega = 180°$, $\psi = 180°$, $\tau_1 = 180°$, $\tau_2 = 90°$, $\tau_3 = 90°$), as any relevant conformation can be reached from this starting point during lattice energy minimization with flexible molecules.	PCFF force field
Pantelides & Karamertzanis	The molecule was treated as a set of seven linked rigid fragments. These fragments were allowed to vary independently during both the global search and the local minimization. All other internal degrees of freedom were fixed at their gas-phase optimal values.	A continuous and differentiable function of the torsion angles using interpolants of SCF calculated energies
Price & Nowell	Searches were performed with a limited series of low-energy rigid conformations.	ΔE_{intra} and the DMA electrostatic model from MP2 calculations
Scheraga & Arnautova	All 6 (*i.e.* τ_1, τ_2, τ_3, φ, ω, ψ) torsional angles were allowed to vary during the global search and local energy minimization.	Third-order Fourier expansion fitted to MP2 calculations
Schmidt	$\psi = \omega = 0°$ and $\tau_1 = 180°$ were fixed, while φ, τ_2, τ_3 were allowed to vary simultaneously with cell parameters and molecular positions, starting from random values.	6-term cosine series fitted to DFT calculations for $E(\varphi)$, empirically fitted potential for $E(\tau_2, \tau_3)$.
Schweizer & Dunitz	Searches were performed with one rigid conformation, chosen after a CSD search showed three structures with $\varphi = 0°$ out of five hits. The geometry of this conformation was taken from an optimization of the isolated molecule: $\varphi = 0°$, $\omega = 180°$, $\psi = 180°$, $\tau_1 = 180°$, $\tau_2 = 30°$, $\tau_3 = 47°$.	Only one conformation considered.
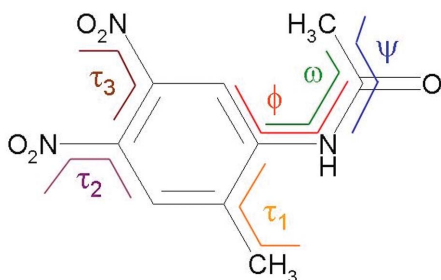
Figure 1
Definition of torsion angles in molecule (X).

partial structure reported at the BCA conference in 2002 is now published (Yu *et al.*, 2004).

Compound (IX) is a derivative of a commercial pigment (2,9-dibromoanthanthrone is commercially produced as a scarlet pigment for use in coatings). The diiodo derivative was synthesized by the pigments research group of Hoechst in the eighties. Due to the low solubility, harsh crystallization conditions had to be applied to grow a single crystal. The molecule crystallizes in layers (Paulus *et al.*, 2005), with molecules on an inversion centre in the space group $P2_1/c$, and with neighbouring molecules forming close $I \cdots O$ contacts

(Fig. 3, left). The C—I···O angle is almost 180° and the I···O contact distance of 3.144 Å is consistent with the polar flattening observed in contact distances with halogen atoms (Nyburg & Faerman, 1985; the sum of van der Waals radii is approx 3.4 Å). These slightly buckled sheets stack parallel to the (10$\bar{2}$) plane (Fig. 3, right).

The crystals of several nitrotoluene derivatives, synthesized almost a century ago, were discovered in the School of Agricultural Sciences at the University of Milano and submitted as candidates for CSP2004 (Demartin *et al.*, 2004). One of these, 2-acetamido-4,5-dinitrotoluene, was selected as molecule (X); it crystallizes in the space group $P2_1/c$ with one molecule per asymmetric unit (Watkin *et al.*, 2004). The amide substituent is twisted out of the plane of the benzene ring, with $\varphi = 41°$. This twist orients the polar H atom at a sufficient angle away from

the ring methyl substituent to form N—H···O═C hydrogen bonds ($R_{NO} = 2.865$ Å) forming molecular chains along the *b* axis of the unit cell (Fig. 4, left). Within these chains, the aromatic parts of the molecules stack in an offset face-to-face manner, while the main interaction between antiparallel chains is the interlocking of acetamide methyl groups between the two nitro substituents (Fig. 4, right).

Azetidine [molecule (XI)] is a low melting compound, so a sample of the pure liquid (obtained from Aldrich) was crystallized in a capillary at 170 K using the laser-assisted zone-refinement technique of Boese (Boese & Nussbaumer, 1994). The molecule is puckered in the observed crystal structure (Parsons, 2005) and, predictably, packs with N—H···N hydrogen bonding. There are several possible arrangements for hydrogen bonding and the observed crystal structure is built up from the chains shown in Fig. 5 (left), with N···N distances of 3.121 and 3.102 Å. These chains contain two azetidine molecules in the asymmetric unit and pack in an antiparallel manner along the crystallographic *b* axis in the space group $P2_1/c$ (Fig. 5, right).

4.2. Comparison of calculated and experimental structures

The predicted and experimental crystal structures were compared with two similarity measures. We performed the first screen for correct predictions using Chisholm's *COMPACK* program (Chisholm & Motherwell, 2005), which searched the lists of submitted crystal structures for predictions that reproduced the experimental structure to within a specified tolerance. The *COMPACK* program avoids the use of space-group symmetry by defining a crystal structure in terms of a molecular packing environment; the experimental structure is represented by a molecule and its coordination shell, which we took as the closest 12–16 molecules. The relationships between these molecules are then represented by a set of interatomic distances and angles, which are searched for in the target structures. If the distances and angles match to within specified tolerances, then the coordination spheres are overlaid and a root-mean-squared deviation (RMSD) in atomic positions is calculated for all 12–16 molecules. In these calcula-

Figure 2
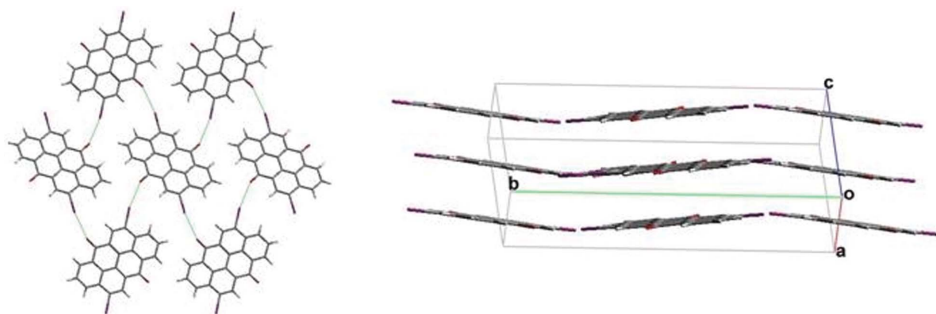Crystal structure of molecule (VIII). Hydrogen bonds are indicated by thin blue lines.

Figure 3
Crystal structure of molecule (IX). I···O contacts are indicated by thin green lines.
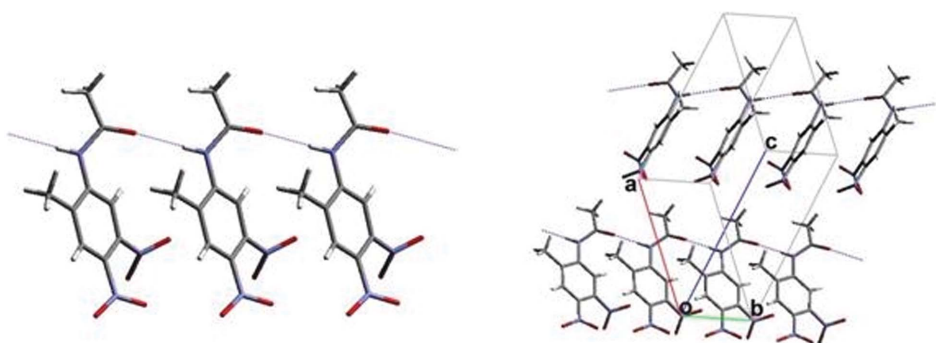

Figure 4
Crystal structure of molecule (X). Hydrogen bonds are indicated by thin blue lines.

tions, we ignored the H atoms because of the usual uncertainties in the X-ray determination of their positions. Typically, a matched crystal structure gave an RMSD of less than 1 Å for a 16-molecule coordination sphere. A second screen for correct predictions was performed by Dzyabchenko on the three predictions by each group using the program *CRYCOM* (Dzyabchenko, 1994). This method also generates an RMSD value for the similarity between structures based on a comparison of the fractional atomic coordinates for the reference molecule, suitably transformed by space-group operators for maximum overlap. These RMSD values are directly comparable to those reported in CSP2001. The two methods found the same set of correct predictions amongst the submitted structures for all four molecules and examples of

the best simulations of the true crystal structures are shown in Fig. 6. In cases where the structure was not found in the submitted list, many of the participants energy minimized the experimental crystal structure using the same energy model as used in the predictions to test how well their methods describe the energy minimum. These results are presented in an 'energy minimized in post-analysis' section of each table.

We present the results and discuss the difficulties encountered for molecules (IX), (X) and (XI), whose predictions were purely blind, in §4.4. Although many of the predictions for molecule (VIII) were performed without recourse to the information in the BCA conference abstract (Yu & Schwalbe, 2002), we distinguish the molecule (VIII) results by presenting them in a separate section.

4.3. Molecule (VIII) predictions

From the 18 participants, 15 attempted the prediction of molecule (VIII) – the other three (Dzyabchenko, Motherwell and Schmidt) abandoned their calculations when the conference report was discovered. Most of those who did submit predictions had either finished their calculations before the conference abstract was discovered or continued their calculations with no bias towards the observed space group or hydrogen-bonding pattern. The results for this molecule are in line with previous blind test results for molecules of this category (Fig. 7). In CSP1999, the small rigid molecule was known to have two polymorphs, one of which was successfully predicted by 4 of 11 participants, while the other was not predicted amongst anyone's 'top 3'. Only one polymorph is known for the small rigid molecule studied in CSP2001 and two out of 15 groups had successful predictions in their top three. For molecule (VIII), four groups found the observed structure in their three official predictions, as ranks 1, 1, 1 and 2. Six of the others had found the correct crystal structure, but higher in their list of generated structures (Table 4). This molecule crystallized in one of the most common space groups for organic molecules ($C2/c$, $Z = 8$), so was easily located by most search methods that rely on the use of these groups. On the other hand, searching in $P1$ with eight inde-
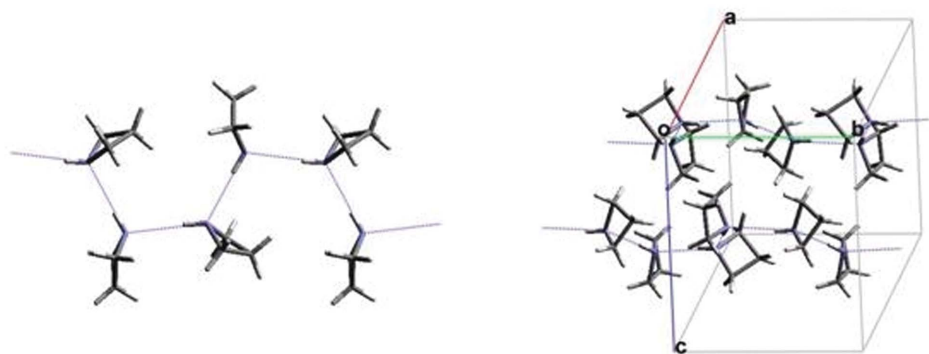

Figure 5
Crystal structure of molecule (XI). Hydrogen bonds are indicated by thin blue lines.

crystal of VIII (green) and
Pantelides.VIII.1 (red)

crystal of IX (green) and
Day.IX.1 (red)

crystal of X (green) and
vanEijck.X.135 (red)

crystal of XI (green), Day.XI.4 (red)
and Day.XI.4 with symmetry relaxed (blue)
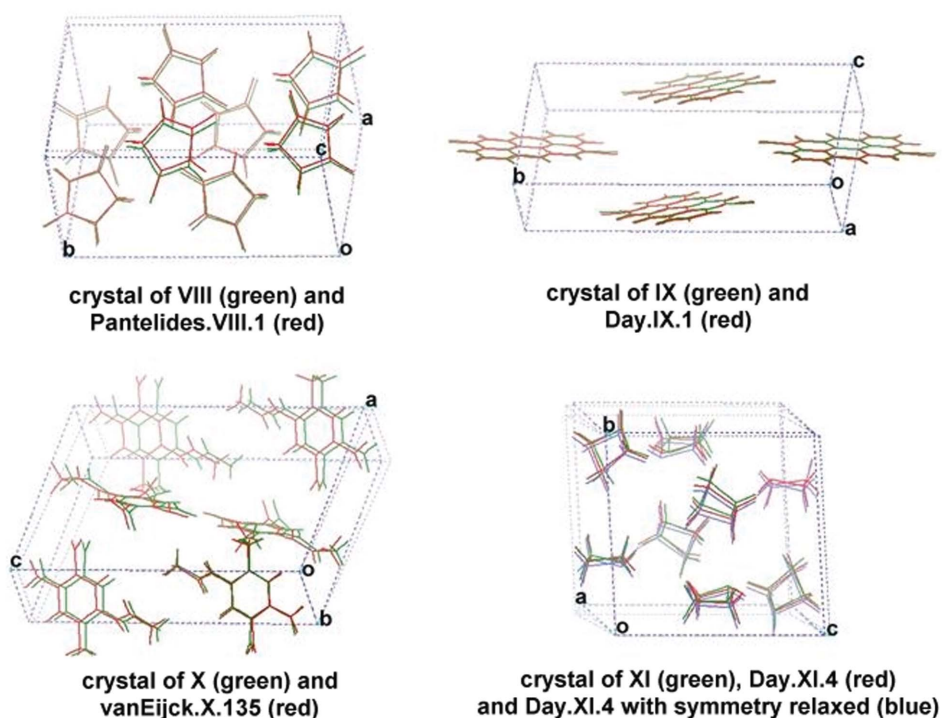
Figure 6
Overlays of unit cells of observed and sample predicted structures of molecules (VIII)–(XI).

research papers

Table 4
Lattice parameters and RMSD for the experimental and predicted structures of molecule (VIII).

NS = not submitted amongst first 3 or in extended lists. 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. The value in parentheses is calculated using CRYCOM.

	Rank	ΔE (kJ mol^{-1})	Density (g cm^{-3})	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	RMSD (Å)
Experimental (T = 190 K)	–	–	1.669	9.3538 (7)	12.1757 (11)	7.2286 (6)	90	104.593 (4)	90	–
Calculations continued as 'blind' after the abstract was discovered										
Predicted amongst first three										
Day	1	−0.25†	1.577	8.962	12.287	7.857	90	102.96	90	0.62 (0.110)
van Eijck	1	−1.12†	1.656	8.745	12.216	7.722	90	103.29	90	0.60 (0.106)
Facelli	2	+0.08	1.683	8.769	12.087	7.598	90	101.27	90	0.61 (0.128)
Pantelides & Karamertzanis	1	−0.34†	1.620	8.974	12.091	7.751	90	102.55	90	0.53 (0.105)
Present in list, outside of first three predictions										
Erk	5	+3.76	1.629	8.711	12.279	7.851	90	103.62	90	0.72
Hofmann	41	+4.44	1.629	9.066	12.230	7.569	90	103.47	90	1.06
Liang	8	+11.72	1.596	8.725	12.410	7.910	90	103.37	90	0.79
Price & Lewis	−‡	−0.20	1.635	8.766	12.185	7.779	90	101.93	90	0.58
Price & Lewis	−‡	+0.08	1.637	8.908	12.229	7.620	90	101.96	90	0.95
Scheraga & Arnautova	5	+1.79	1.677	8.801	12.133	7.551	90	100.51	90	0.73
Verwer	4	+3.05	1.632	8.764	12.176	7.845	90	103.38	90	0.67
Present in list, but with large deviations from the observed structure										
Leusen	3	+0.17	1.610	8.212	12.074	8.559	90	103.28	90	1.70 (0.175)
Calculations using some of the structural information reported in (Yu & Schwalbe, 2002)										
Ammon	−§	–	1.587	9.008	12.283	7.758	90	102.63	90	0.60 (0.148)
Schweizer & Dunitz	−¶	+1.7	1.770	8.665	11.836	7.481	90	101.81	90	0.73 (0.150)
Energy minimized in post-analysis										
Della Valle	NS	+0.95	1.494	9.381	12.194	7.864	90	98.52	90	0.60

† ΔE for the global minimum is calculated with respect to the second-lowest energy prediction by that research group. ‡ Predictions were not ranked on lattice energy alone – a good reproduction of the experimental structure was found as the global minimum in lattice energy in the search with a planar molecule (first line), and second lowest in energy when searched with an 'envelope' molecular structure (second line). However, these were not amongst the submitted predictions, as the three for submission were chosen fairly arbitrarily on energy, growth rate and mechanical properties as described in the supplementary material. § Only one structure submitted as the lowest energy in the search with dimer geometry based on the BCA conference abstract (Yu & Schwalbe, 2002). ¶ Submitted as the lowest energy structure found in *C2/c*, the space group of the observed crystal structure. This was the highest density structure in the entire search, but some lower energy structures were found in other space groups.

pendent molecules is a very difficult task, even for rigid molecules. Nevertheless, the experimental structure was located by one of the participants (Scheraga & Arnautova) who carried out a global search in $P1$, Z = 8.

For the ranking of the predicted structures, many model potentials are available that have been parameterized to similar molecules, so most participants found the experimentally observed crystal structure either at or within a few kJ mol^{-1} of the global minimum in lattice energy (Table 4). Almost all of the predicted structures are compressed in the a direction and expanded along **c**, indicating a consistent problem with either most of the force fields used, the assumed molecular geometry or the temperature-less nature of lattice-energy minimization. Besides these distortions of the lattice, the predicted structures reproduce the experimentally observed crystal very well. The lowest RMSD in atomic positions are around 0.5–0.7 Å for a sphere of 16 molecules using Chisholm's comparison algorithm (Chisholm & Motherwell, 2005) or about 0.1 Å from Dzyabchenko's comparison method (Dzyabchenko, 1994). These are similar to the deviations in the predictions of the corresponding molecule in CSP2001. The overlay in Fig. 6 is typical of most of

the correctly predicted structures and demonstrates the quality of these predictions.

One exception is Leusen's predicted structure: his third predicted structure does correspond to the same lattice-energy minimum obtained by energy minimizing the observed structure, so is the closest to this structure that could be predicted with his choice of force field. However, the deviations are too great to be considered a successful prediction (high RMSD
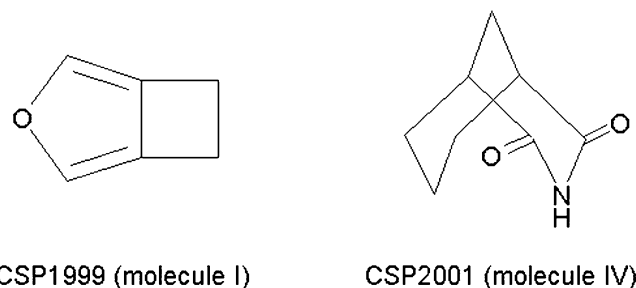

Figure 7
Previous blind test molecules in category 1 (simple rigid molecules).

Table 5
Lattice parameters, ΔE and RMSD for the experimental and predicted structures of molecule (XI).

NS = not submitted amongst first 3 or in extended lists. 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.

	Rank	ΔE (kJ mol^{-1})	Density (g cm^{-3})	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	RMSD (Å)
Experimental (T = 170 K)	–	–	1.007	9.507 (3)	9.122 (3)	9.790 (3)	90	117.469 (4)	90	–
None predicted amongst first three										
Present in list, outside of first three predictions										
Day	4†	+0.23	0.979	9.832	8.983	10.054	90	119.22	90	0.46
Lowered symmetry	–‡	(−0.15)	0.980	9.830	8.990	10.059	92.41	119.24	87.52	0.65
van Eijck	30†	+1.01	0.988	10.025	8.795	10.275	90	122.11	90	0.32
Lowered symmetry	–‡	(+0.93)	0.989	10.029	8.788	10.282	89.99	122.20	89.95	0.42
Pantelides & Karamertzanis	24†	+0.76	0.986	9.796	8.954	10.025	90	119.02	90	0.44
Lowered symmetry	–‡	(+0.76)	0.986	9.798	8.956	10.03	90.16	119.06	91.44	0.59
Present in list, but with large deviations from the observed structure										
Leusen	> 100	+9.71	0.973	8.644	9.955	9.264	90	102.07	90	–§
Energy minimized in post-analysis										
Ammon	NS	+5.33	0.942	9.714	9.082	10.091	90	115.25	90	0.76
Della Valle & Venuti	NS	+4.61	0.962	9.210	9.143	9.869	90	108.36	90	1.26
Dzyabchenko	NS	+13.72	0.975	9.563	9.199	9.675	90	113.91	90	1.31
Facelli	NS	+5.03	1.087	10.303	8.161	9.790	90	122.00	90	1.74
Hofmann¶	NS	+2.70	1.051	9.459	8.942	9.748	89.62	118.94	89.90	0.15
Liang	NS	+6.28	0.987	9.723	9.202	9.950	90	120.33	90	0.78
Price & Torrisi	NS	+0.67	0.977	9.955	8.939	10.201	89.71	121.05	86.25	0.78
Scheraga & Arnautova	NS	+1.50	0.977	9.692	9.302	9.933	90	119.98	90	0.63
Schmidt	NS	+2.01	1.047	9.821	8.633	9.966	90	120.99	90	–

† The structure found in $P2_1/c$ minimized to a saddle point on the potential energy surface, with a very small negative eigenvalue of the Hessian. ‡ The unstable $P2_1/c$ structures, after minimization with symmetry constraints were removed, resulting in a final structure in the space roup $P\bar{1}$ with Z' = 4. The lowering of the energy is very small for van Eijck (0.09 kJ mol^{-1}) and Pantelides & Karamertzanis (0.001 kJ mol^{-1}), while the structure lowers in energy by 0.38 kJ mol^{-1} with Day's model, making this structure lower in energy than any other hypothetical structure. § Deviations from the experimental structure are so pronounced that no RMSD could be calculated. ¶ Minimization in space group $P1$.

and errors in lattice vectors of > 10%) – here, the search algorithm has located the correct structure, but the predictions are a failure due to inadequacies of the chosen force field.

A newly discovered polymorph of a similar molecule (barbituric acid) indicates that slight molecular deformations can be an issue even for what seems to be a typical rigid molecule (Lewis *et al.*, 2004). However, for (VIII) the choice of different molecular structures had only a small effect on search results – Price & Lewis found the experimental crystal structure as a global minimum with a planar model, and the second lowest with an envelope molecular structure.

Besides molecule (VIII), participants in this third blind test felt that the chosen molecules were generally more difficult prediction targets than the molecules of previous blind tests.

4.4. Purely blind predictions – molecules (IX), (X) and (XI)

4.4.1. Molecule (XI). Molecule (XI) was attempted by all 18 groups, but no correct predictions were made amongst the first three predictions of any of the participants. Here, a large factor in the poor predictive success is that the molecule crystallizes with two molecules in the asymmetric unit. In fact, because of limitations of some methods or constraints on computing time,[5] only eight of the 18 groups could have produced Z' = 2, $P2_1/c$ structures in their structure generation. Of these, three participants did generate the observed structure, but ranked it outside of the top 3 (Table 5). Leusen encountered similar problems as molecule (VIII) – the closest

lattice-energy minimum to the true structure was present in his extended list, but severely distorted from the experimentally observed structure. Again, the structure generation found the appropriate minimum, but the force field was unsatisfactory.

Crystallization with two molecules in the asymmetric unit was not the only difficulty with molecule (XI). Some participants had problems modelling this molecular crystal, even after the experimentally observed crystal was made available; the strained four-membered ring is not typical of molecules used in parameterizing many of the common force fields. Some force fields predicted a planar ring structure, so the experimental crystal structure could not be adequately modelled. Others found that even quantum mechanical calculations did not reproduce the observed molecular structure as accurately as expected for such a small molecule. The observed molecular structure is atypical of such azetidine rings – the bond lengths are shorter than in the crystals of other molecules with azetidine rings – the C—C bond lengths of 1.517 (4), 1.514 (4), 1.510 (4) and 1.514 (3) Å are all much shorter than the average observed bond lengths of 1.548 Å (σ = 0.018 Å; Allen *et al.*, 1987). Similarly, the C—N bond lengths in the crystal of (XI) – 1.474 (3), 1.465 (4), 1.470 (3) and 1.465 (4) Å – are all at least 0.1 Å shorter than the average bond lengths in other crystals of azetidines (1.484 Å, σ = 0.018 Å; Allen *et al.*, 1987). The sensitivity of lattice-energy calculations to the assumed molecular structure is such that such abnormalities led to large variations in the calculated relative energies (Table 5). Compounding the problem was the observation that many distinct crystal structures exist in a small energy window around the global minimum (*e.g.* Day found 20 within

[5] Recall that molecule (XI) was not added to the list of targets until midway through the test.

Table 6
Lattice parameters and RMSD for the experimental and predicted structures of molecule (IX).

NS = not submitted amongst first three or in extended lists. 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 the H atoms. The value in parentheses is calculated using *CRYCOM*.

	Rank	ΔE (kJ mol^{-1})	Density (g cm^{-3})	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	RMSD (Å)
Experimental (T = 293 K)	–	–	2.299	4.202 (1)	20.956 (4)	9.276 (2)	90	100.62 (3)	90	–
Predicted amongst first three										
Day	1	−0.34†	2.296	4.224	20.925	9.306	90	101.05	90	0.19 (0.068)
Present in list, outside of first three predictions										
Boerrigter‡	100	+10.40	2.337	3.999	21.125	9.516	86.72	81.97	86.53	1.19
Della Valle & Venuti	35	+13.60	2.314	4.084	21.644	9.125	89.71	83.26	89.71	0.80
Dzyabchenko‡	62	+42	2.310	4.406	21.226	8.625	90	95.76	90	1.15
Hofmann‡	69	+21	2.187	4.474	21.950	8.901	90	104.12	90	1.64
Liang	10	+22	2.275	4.115	21.388	9.343	90	97.81	90	0.41
Pantelides & Karamertzanis	10	+3.36	2.341	4.126	21.044	9.224	90	98.59	90	0.39
Price	∼ 40	+17	2.287	4.189	20.788	9.488	90	101.15	90	0.82
Scheraga & Arnautova	10	+4.48	2.270	4.081	21.348	9.499	90	99.44	90	0.40
Energy minimized in post-analysis										
van Eijck	NS	+10.0	2.337	4.107	21.121	9.243	90	98.40	90	0.48

† ΔE for the global minimum is calculated with respect to the second lowest energy prediction. ‡ Reported as matches to the experimental structure, but with extreme deviations.

1 kJ mol^{-1}; Pantelides & Karamertzanis found 77 unique stationary points within 1 kJ mol^{-1}). The accuracy required in the energy ranking is evident in Pantelides & Karamertzanis' and van Eijck's predictions; the observed structure is within *ca* 1 kJ mol^{-1} of the global minimum in both cases, but ranked 24th and 30th in their lists, respectively. Given this observation, it was somewhat surprising that the calculation of the thermal expansion and vibrational contributions to the free energy (see Day; Della Valle & Venuti) led to minimal rearrangements of the stability ordering.

A final comment on the experimentally observed structure of (XI) – at least four participants reported that this structure corresponds to a saddle point on the potential energy surface – small negative eigenvalues of the Hessian were observed, indicating a saddle point on a flat potential energy surface. This flat potential energy surface and instability further complicated the calculations on this molecule. Relaxing the crystal structure along the path to lower energy breaks the $P2_1/c$ symmetry of the structure, shifting it to a $P\bar{1}$, $Z' = 4$ crystal (blue overlay in Fig. 6). This shift lowers the energy by a tiny amount – 0.008 (Pantelides & Karamertzanis), 0.08 (van

Eijck), 0.04 (Price & Torrisi) and 0.4 kJ mol^{-1} (Day). For the latter, this energy lowering is enough to make this structure more stable than the three submitted structures. At least on this (exp-6 + multipoles) potential energy surface, the experimental structure appears to correspond to the global minimum in lattice energy. With such a low energy barrier, the observed $P2_1/c$ symmetry may be a thermal average over two symmetrically equivalent lower symmetry structures. Such a flat potential energy surface would also lead to a large amplitude degree of freedom with an important entropy increase and possible disorder, although none is reported in the crystal structure determination (Parsons, 2005).

In short, while molecule (XI) may at first seem to be a typical 'simple' target, there were a multitude of complications making this a very difficult system to predict.

4.4.2. Molecule (IX). Molecule (IX) was attempted by 15 groups and the experimental structure was present in only one participant's 'top 3' – as the global minimum in lattice energy from Day's search. Eight others predicted the correct structure, but outside of their best three (Table 6) and several of these show large distortions away from the observed structure – RMSD values greater than 1 Å are fairly high for rigid molecules. Predictions for molecule (IX) were less successful than for the corresponding molecules in previous blind tests (Fig. 8), which were predicted in the best three guesses by 1 of 8 (CSP1999) and 4 of 15 (CSP2001) groups.

The poor success rate may be partly due to the size of the molecule (35 atoms), but is largely a consequence of problems in modelling iodine, whose repulsive wall shows an extreme anisotropy; Nyburg and Faerman found the axial ratio of end-on to side-on contact distances is 0.83 (Nyburg & Faerman, 1985). The lattice parameters in the one correct prediction are all within 1% of the observed values and the small RMSD shows very good structural reproduction (see overlay in Fig. 6). For this prediction, a sophisticated and computationally expensive model potential had to be developed specifically, using anisotropic atom–atom repulsion and electrostatic
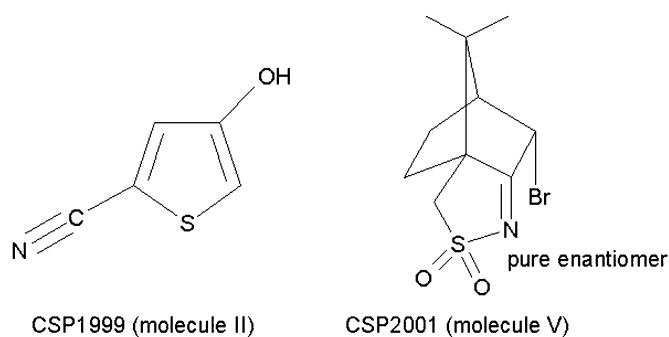

CSP1999 (molecule II)   CSP2001 (molecule V)

Figure 8
Previous blind test molecules in category 2 (rigid molecules with challenging functional groups).

Table 7
Lattice parameters and RMSD for the experimental and predicted structures of molecule (X).

NS = not submitted amongst first three or in extended lists. 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.

|	Rank	ΔE (kJ mol^{-1})	Density (g cm^{-3})	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	RMSD (Å)
Experimental (T = 150 K)	–	–	1.528	12.5693 (4)	4.8531 (1)	17.2663 (5)	90	99.1624 (15)	90	–
None predicted amongst first three										
Present in list, outside of first three predictions										
Day	5–9†	(+2.36)†	1.456	12.423	5.005	17.687	90	97.19	90	0.75
Dzyabchenko	41	+117	1.507	13.468	4.745	16.582	90	95.91	90	1.17
van Eijck	135	+7.08	1.484	12.574	4.840	17.711	90	96.52	90	0.50
Erk	18	+6.86	1.430	13.407	4.905	17.109	90	99.00	90	0.63
Leusen	> 100	+35	1.476	12.448	4.728	18.827	90	103.7	90	1.74
Pantelides & Karamertzanis	> 100	+20	1.440	13.019	4.980	17.117	90	96.20	90	0.71
Schmidt	> 100	+28	1.578	13.016	4.847	16.042	90	95.67	90	0.81
Energy minimized in post-analysis										
Ammon	NS	−4.10	1.432	13.427	4.922	16.949	90	97.96	90	0.68
Facelli	NS	−3.49	1.521	12.521	4.879	17.411	90	100.92	90	0.26
Hofmann‡	NS	+14.66	1.512	12.478	4.959	17.222	90.14	99.65	89.84	0.14
Price & Nowell	NS	Approx. +5§	1.430	13.483	4.919	16.914	90	97.82	90	0.67
Scheraga & Arnautova	NS	+10.78	1.411	12.245	5.139	18.204	90	100.62	90	1.11

† Searches in three conformations were ranked separately; this was found as the third in one of the three conformations considered, so would have had an overall rank between 5 and 9. ΔE given with respect to the lowest-energy crystal structure from searches with that conformation. ‡ Minimization in space group $P1$. § Estimating the intramolecular energy from MP2/6-31G** partially optimized experimental molecular geometries.

models to correctly represent the I atoms' contribution to the lattice energy. The parameterization and testing of such models was hindered by the poor quality of the reported crystal structures of similar compounds, *e.g.* CSD refcodes IOANTQ, IONTRQ, IANTRQ, whose R factors are 0.106, 0.130 and 0.0175, with no hydrogen positions reported.

4.4.3. Molecule (X). Molecule (X) (the conformationally flexible molecule) was attempted by 15 groups, but there were no correct 'top three' predictions. For comparison, the flexible molecule in CSP1999 (Fig. 9, left) was only predicted once from 11 submissions, while there were no successes from 11 submissions in CSP2001 (Fig. 9, right). Problems with simultaneously treating the conformation and crystal packing meant that when a close match to the experimental structure was found in seven of the participants' extended lists, the ranking on lattice energy was generally poor and deviations in lattice parameters were often higher than for the rigid molecules (Table 7). The overlay in Fig. 6 shows the best reproduction (van Eijck's 135th ranked structure) of the experimentally observed crystal structure.

Improving the predictability of such molecules requires advances in search procedures for the packing of conformationally flexible molecules, as well as greater accuracy in models of intra- and intermolecular energies. It was found that all of the torsion angles defined in Fig. 1 are important in determining the packing of molecule X – τ_1, φ, ψ and ω all influence the intermolecular hydrogen bonding, while τ_2 and τ_3 have an important effect on close packing of the molecules. More details are available in the individual accounts in the supplementary information. The balance of the calculated intra- and intermolecular energies was crucial – the gas-phase minimum energy structures from quantum chemical calculations had the acetamide substituent in the plane of the benzene ring ($\theta = 0°$), with the polar H atom shielded by the *ortho*-methyl substituent, preventing efficient intermolecular hydrogen bonding. The energy required for the observed 41.5° twist of the acetamide substituent was calculated to be between 5 and 14 kJ mol^{-1}, and the improvement in calculated lattice energies with this conformation did not compensate for this conformational energy. Crystals in which conformations of the constituent molecules are far above the gas-phase energy minimum must be critically stabilized by non-bonded interactions in the crystal. Such cases may be the greatest challenge for energy models in crystal structure prediction. One approach is the coupling of quantum mechanics calculations with very accurate force fields during energy minimization and this has led to successful predictions for some molecules where crystal and vacuum conformations are entirely different (van Eijck *et al.*, 2001*a,b*).
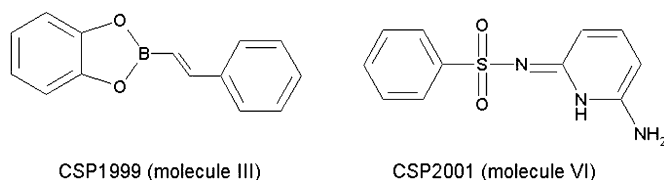
Figure 9
Previous blind test molecules in category 3 (flexible molecules).

5. Discussion

5.1. Overall success rates

The overall success rates in this blind test are a combination of the successes and failures of different aspects of the prediction methodology – namely, the generation and the evaluation of possible structures.

research papers

5.1.1. The search problem. After CSP2001, it was felt by many of the participants that methodologies for generating the structures of rigid molecules corresponding to the common space groups with $Z' = 1$ were then reliable. For the category 1 molecule in CSP2001, 11 out of 15 participants generated the experimentally observed crystal structure somewhere in their lists. For molecule (VIII), 11 out of 14 groups (not counting the search using the reported dimer as a starting point) generated the correct structure. These success rates are fairly good and some of the methodologies used in more than one blind test appear to be consistently successful. Where it seems that simple rigid molecule searches need improvement is when more than one independent molecule is present in the asymmetric unit. Only eight out of 18 participants searched for structures with $Z' > 1$, some due to limitations of their search methodology and others because of limited computing resources and time. However, even those search algorithms that have the capacity to search with multiple independent molecules performed poorly; several participants reported poor sampling of phase space with $Z' > 1$ and, indeed, only three groups generated the correct structure for molecule (XI). The predictability of such structures is important – over 10% of homonuclear organic crystals have $Z' > 1$ (Steiner, 2000). Furthermore, the application of crystal structure prediction to salts and co-crystals requires the improvement of search methods with multiple molecular units. There has been success in searches for salt structures of flexible molecules, demonstrating that the search problem can be solved (Leusen, 2003). However, the time and computational expense required prohibit such predictions from becoming routine without significant increases in computing power. Some advances will come with increasing use of parallel computing and distributed computing networks, but methodological advances will also be needed.

In CSP2001 the search methods were as successful with the category 2 molecule as with category 1: 12 out of 15 submitted sets of structures contained the experimentally observed crystal structure. This year's molecule (IX) has internal symmetry which predisposes it to crystallize on a crystallographic inversion centre. This should have simplified the search, but the success rate dropped to only nine out of 15 participants generating the correct structure. Why the poor search results? This is probably due partly to the size of the molecule, but the main reason must be the critical difficulties in modelling the I atoms – both in quantum chemical calculations used to generate electrostatic models and in limitations of the spherical atom–atom method for interactions with I atoms, which are known to behave anisotropically. Most of the search algorithms rely on a reasonable model for interaction energies; minima on the model potential energy surface must be reasonably close to minima on the 'real' energy surface and the relative energies of the minima must also be reasonably accurate. Search methods must discard high-energy structures to avoid infinitely many solutions, so a model that wrongly assigns a high energy to the correct crystal structure may result in a failed search.

The search algorithms struggled with the flexible molecule. Only seven out of 15 groups generated the correct structure somewhere in their extended list – approximately the same success rate as in CSP2001 (four of 11). Apparently, little progress has been made in solving the problems involved in simultaneously searching both molecular conformations and crystal-packing alternatives. The problems are partly the same as for $Z' > 1$ crystals – the extra degrees of freedom quickly expand the phase space that must be sampled. Participants took different approaches to this problem. Some performed rigid-body searches on a range of conformations and took the conformational energy from quantum mechanical calculations. Others searched phase space with rigid molecular structures, but allowed these to adapt to the crystalline environment during energy minimization with some or all of the intramolecular coordinates free to vary. Still others included conformational flexibility from the start – defining intramolecular degrees of freedom during the search and treating them as parameters in the same way as changes to the cell dimensions and positions of the molecules in the unit cell. Each of these approaches had some success. Perhaps more important is the inadequacy of many current force fields in satisfactorily modelling both intra- and intermolecular energies, which are on such different scales. Energy models that cannot treat these contributions accurately and in a balanced manner will inevitably lead to inconsistent or failed searches. Attempts to couple atomistic intermolecular potentials with *ab initio* calculations of intramolecular energies during energy minimization, *e.g.* van Eijck *et al.* (2001*a,b*) and Pantelides & Karamertzanis (this work), are a possible way forward.

While there are still challenges in this structure generation aspect of crystal structure prediction, we are encouraged by the consistent success of some methods – two participants (Day; Pantelides & Karamertzanis) generated the experimental crystal structure somewhere in their list for all three molecules and others (van Eijck found three out of four, Dzyabchenko two out of three and Schmidt one out of two) missed only one of the molecules that they attempted. The comparison of methods and understanding of any failures should guide the continued advancement in search methodologies.

5.1.2. Ranking of the generated structures. Most methods still rely on calculated lattice energies as the sole scoring function for ranking predicted structures and the discrepancies in relative energies amongst the successful predictions and post-analysis energy minimizations highlight the variety of energy models applied to this test. The part of the energy models that has seen the most attention is the description of electrostatic interactions. There were several approaches used in CSP2004 to improve on the atomic point-charge model: atomic multipole expansions (Day; van Eijck; Price *et al.*), off-nuclear charges (Pantelides & Karamertzanis) and bond dipoles (Dzyabchenko), and the representation of the molecular charge density in terms of a large number (*ca* 10⁴) of charged pixels (Schweizer & Dunitz). A recent study (Day *et al.*, 2005) has directly compared the crystal structure prediction of a large set of rigid molecules using atomic point charges

and atomic multipole models. The results demonstrated an improved success rate from approximately 40% of the experimentally observed structures found amongst the lowest three in lattice energy with a point charge model to 55% using atomic multipole expansions. Optimized off-nuclear charge sites appear to be able to reproduce molecular electrostatic potentials to about the same accuracy as atomic multipole models (Karamertzanis & Pantelides, 2004) and bond dipoles give a comparable quality of molecular electrostatic potential, so such models should lead to similar improvements. The pixel approach, recently developed by Gavezzotti (Gavezzotti, 2002, 2003a) and used by Schweizer & Dunitz in this year's blind test, is an exciting new methodology for quantifying intermolecular interactions and shows promise for rationalizing and predicting organic crystal structures (Gavezzotti, 2003b,c). While all of these methodologies can offer improved reliability over the spherical atom–atom approach for small, rigid molecules, there is much work to be done in transferring such methods to larger molecules and improving on the consistently disappointing crystal structure prediction results for flexible molecules. For example, the conformational dependence of atomic multipoles means that a static set of multipoles offers little improvement over atomic charges in the energy minimization of conformationally flexible molecules (Brodersen et al., 2003). Better results are obtained by recalculating the multipoles as the conformation changes during energy minimization (van Eijck et al., 2001a,b); this procedure is promising, but currently too expensive for handling large molecules.

Model potentials can be fine-tuned even further by modifying the form and parameters of the model of van der Waals interactions specifically for the molecule being studied. Such an approach led to the only successful prediction for the truly blind molecules in this test [Day, molecule (IX)]. Currently, the time required to build such a molecule-specific model means that the approach is not generally applicable. However, systematic approaches to building such model potentials are developing (Day & Price, 2003; Mitchell & Price, 2000).

Besides lattice energy, several additional selection criteria have been tried. As an improvement within the thermodynamic model, a few participants calculated the vibrational contribution to the crystals' free energies (Price et al.) and free energy minimization was also attempted (Day; Della Valle & Venuti). For the molecules in this study, such calculations did not lead to significant reordering of the thermodynamic stabilities. It is also clear that thermodynamics do not control the crystallization process alone – kinetic factors can be crucial. Two groups (Boerrigter; Price et al.) re-ranked their lattice energy predictions by relating other calculated properties (e.g. structural isotropy, mechanical properties and morphology calculations) to possible kinetic preferences for certain structures. Such ideas have shown promise in other studies (Anghel et al., 2002; Beyer et al., 2001) and may play an important role in future crystal-structure prediction methodologies. However, in the current blind test, these extra scoring functions failed to improve on lattice-energy predictions; in fact, Price & Lewis would have correctly predicted the

structure of molecule (VIII) had they not submitted a slightly less stable structure with a predicted favourable growth rate instead of their global minimum with the planar molecular conformation.

6. Conclusions

Blind predictions have been attempted for a set of molecules in a test organized by the Cambridge Crystallographic Data Centre. Besides molecule (VIII), whose structure had unfortunately already been partially reported, the overall success rates remain poor and do not reflect any major advances since the previous blind test, CSP2001 (Motherwell et al., 2002). In the previous test, there were six successes within the three predictions allowed for each molecule. For the truly blind predictions [i.e. excluding molecule (VIII)] in the present test, there was only one success – the prediction of molecule (IX)'s crystal structure as a global minimum in lattice energy by one of the participants. Analysis of the lowest energy predictions, extended lists of generated structures and post-analysis of the experimentally observed crystal structures has highlighted possible reasons for the poorer success rates compared with the previous blind tests and so the blind test has been very useful for participants in pointing to the current obstacles to reliable crystal structure prediction. In fact, despite the lack of observable progress, there is a feeling among the participants that the programs and methodologies are improving.

In summary, search methods must be improved for crystals with multiple independent molecules and conformationally flexible molecules. To find the most stable arrangement of molecules, the basis of energy calculations still needs to be improved. Electrostatic models that are more detailed than the traditional atomic point charge model are becoming more common and seem to improve the reliability of lattice-energy calculations. Descriptions of the van der Waals interactions are still largely empirically derived and the growing amount of experimental data has been vital in the development of model potentials. While such models are often satisfactory, the treatment of less common atom types (such as iodine) and atoms in bonding environments that are not adequately represented in empirical parameterizations [such as the strained ring in molecule (XI)] may require more specifically developed models. The successful advances in energy models for rigid molecules (such as atomic multipoles and off-centre charge sites) should be extended to flexible molecules, for which energy models still need improvement. Most current approaches to the intermolecular energy are still based on atom–atom interactions, with or without atomic point charges or multipoles, and, while this may be computationally convenient, its physical basis leaves something to be desired. At the close intermolecular distances that occur in crystals, the accurate calculation of the electrostatic energy requires the interaction between the molecular charge density distributions rather than between a few localized charge monopoles, dipoles, quadrupoles, etc. As computational possibilities improve, the role of individual atomic nuclei as the loci of intermolecular attractions and repulsions will possibly need to

be replaced by better models based on more delocalized quantum-mechanical charge-density distributions (Gavezzotti, 2003*b,c*).

However, even with advances in our models for the energy, most methodologies continue to rely solely on global minimization of the lattice energy. A few participants included estimated vibrational contributions to the free energy of predicted structures through lattice dynamical calculations, thus accounting for entropy differences between polymorphs. In the present blind test, such calculations did not significantly change the stability order of the hypothetical crystals, so had little effect on the reliability of the predictions. Such calculations may become more common as such a capability is implemented in the programs and the required computing time becomes available for such calculations on the thousands of structures that are produced for a given molecule.

Of course, we recognize that crystallization is not an equilibrium process, so not under thermodynamic control. It seems that reasonable success can be achieved through lattice-energy minimization, which suggests that molecules often tend to find their lowest-energy arrangement during crystallization. However, in some cases, although we do not actually know with what frequency, metastable crystals will be favoured at the nucleation event and preserved during crystal growth. To thoroughly account for such effects in crystal structure prediction, a full treatment of nucleation at the molecular level is required and a detailed description of crystallization conditions would be necessary. While there has been some progress in the molecular-level modelling of nucleation and crystal growth, general application to the large number of hypothetical structures in CSP is currently far beyond computational feasibility. In the absence of the ability to model nucleation and growth, some participants have tried to relate simpler calculated properties (such as structural isotropy and crystal morphologies) to preferences for some of the low-energy predicted structures during nucleation and growth. Alternatively, statistical comparisons to the known structures in the Cambridge Structural Database may uncover kinetic preferences for some hypothetical structures (Desiraju, 2002). These extra criteria did not improve predictions for the molecules in this blind test, but the testing of such ideas is a necessary step towards including kinetic criteria in CSP and towards an understanding of crystallization and polymorphism.

In conclusion, this third blind test has provided an objective evaluation of the state of the art in crystal structure prediction. There are technical and fundamental obstacles to reliable predictions and the comparison of a variety of approaches and problems encountered while working on a difficult set of molecules has served to highlight the immediate areas requiring attention and methodological innovation.

References

Accelrys Inc. (1997). *Cerius*2. San Diego, USA.

Accelrys Inc. (2004). *MS Modelling*, Release 3.0.1. San Diego, USA.

Allen, F. H., Kennard, O., Watson, D. G., Brammer, L. & Orpen, A. G. (1987). *J. Chem. Soc. Perkin Trans. II*, pp. S1–S19.

Anghel, A. T., Day, G. M. & Price, S. L. (2002). *CrystEngComm*, **4**, 348–355.

Arnautova, Y. A., Jagielska, A., Pillardy, J. & Scheraga, H. A. (2003*a*). *J. Phys. Chem. B*, **107**, 7143–7154.

Arnautova, Y. A., Pillardy, J., Czaplewski, C. & Scheraga, H. A. (2003*b*). *J. Phys. Chem. B*, **107**, 712–723.

Bazterra, V. E., Ferraro, M. B. & Facelli, J. C. (2002*a*). *J. Chem. Phys.* **116**, 5984–5991.

Bazterra, V. E., Ferraro, M. B. & Facelli, J. C. (2002*b*). *J. Chem. Phys.* **116**, 5992–5995.

Bazterra, V. E., Ferraro, M. B. & Facelli, J. C. (2004*a*). *Int. J. Quant. Chem.* **96**, 312–320.

Bazterra, V. E., Ona, O., Caputo, M. C., Ferraro, M. B., Fuentealba, P. & Facelli, J. C. (2004*b*). *Phys. Rev. A*, **69**, 053202.

Beyer, T., Day, G. M. & Price, S. L. (2001). *J. Am. Chem. Soc.* **123**, 5086–5094.

Beyer, T. & Price, S. L. (2000). *CrystEngComm*, **34**, 183–190.

Boese, R. & Nussbaumer, M. (1994). *Tranformations and Interactions in Organic Crystal Chemistry*. International Union of Crystallography, Crystallographic Symposia, 7.

Brodersen, S. W., Leuson, F. J. J. & Engel, G. (2003). *Phys. Chem. Chem. Phys.* **5**, 4923–4931.

Busing, W. R. (1981). *WMIN*, Report ORNL-5747. Oak Ridge National Laboratory, Tennessee, USA.

Chisholm, J. & Motherwell, W. D. S. (2005). *J. Appl. Cryst.* **38**, 228–231.

Day, G. M., Chisholm, J., Shan, N., Motherwell, W. D. S. & Jones, W. (2004). *Cryst. Growth Des.* **4**, 1327–1340.

Day, G. M. Motherwell, W. D. S. & Jones, W. (2005). *Cryst. Growth Des.* **5**, 1023–1033.

Day, G. M. & Price, S. L. (2003). *J. Am. Chem. Soc.* **125**, 16434–16443.

Della Valle, R. G., Venuti, E. & Brillante, A. (2003). *J. Chem. Phys.* **118**, 807–815.

Demartin, F., Filippini, G., Gavezzotti, A. & Rizzato, S. (2004). *Acta Cryst.* B**60**, 609–620.

Desiraju, G. R. (2002). *Nat. Mater.* **1**, 77–79.

Dunitz, J. D. & Scheraga, H. A. (2004). *Proc. Nat. Acad. USA*, **101**, 14309–14311.

Dzyabchenko, A. V. (1994). *Acta Cryst.* B**50**, 414–425.

Dzyabchenko, A. V. (2004). *PMC*, Version 2004. Karpov Institute of Physical Chemistry, Moscow.

Dzyabchenko, A. V., Agafonov, V. & Davydov, V. A. (1999). *J. Phys. Chem. A*, **103**, 2812–2820.

Eijck, B. P. van (2002). *J. Comput. Chem.* **23**, 456–462.

Eijck, B. P. van & Kroon, J. (2000). *Acta Cryst.* B**56**, 535–542.

Eijck, B. P. van, Mooij, W. T. M. & Kroon, J. (2001a). *J. Comput. Chem.* **22**, 805–815.

Eijck, B. P. van, Mooij, W. T. M. & Kroon, J. (2001b). *J. Phys. Chem. B*, **105**, 10573–10578.

Erk, P. (1999). *Crystal Design: From Molecular to Application Properties in Crystal Engineering: From Molecules and Crystals to Materials*, edited by D. Braga, F. Grepioni & G. A. Orpen, NATO Science Series C, Vol. 538. Dordrecht: Kluwer Academic Publishers.

Gavezzotti, A. (1999–2000). *ZIP-PROMET*. University of Milano, Italy.

Gavezzotti, A. (2002). *J. Phys. Chem. B*, **106**, 4145–4154.

Gavezzotti, A. (2003a). *J. Phys. Chem. B*, **107**, 2344–2353.

Gavezzotti, A. (2003b). *CrystEngComm*, **5**, 429–438.

Gavezzotti, A. (2003c). *CrystEngComm*, **5**, 439–446.

Gavezzotti, A. (2004). *OPIX*. University of Milano, Italy.

Gemmel, E., Beck, H. Bolte, M. & Egert, E. (1999). *MOMO*. University of Frankfurt, Germany.

Hofmann, D. W. M. & Apostolakis, J. (2003). *J. Mol. Struct.* **647**, 17–39.

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. (2004). *Mol. Simulation*, **30**, 413–436.

Karamertzanis, P. G. & Pantelides, C. C. (2005). *J. Comput. Chem.* **26**, 304–324.

Kuleshova, L. N., Hofmann, D. W. M. & Antipin, M. Y. (2005). *Cryst. Rep.* **50**, 167–176.

Leuson, F. J. J. (2003). *Cryst. Growth Des.* **3**, 189–192.

Lewis, T. C., Tocher, D. A. & Price, S. L. (2004). *Cryst. Growth Des.* **4**, 979–987.

Lommerse, J. P. M., Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Gavezzotti, A., Hofmann, D. W. M., Leuson, 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.

Mitchell, J. B. O. & Price, S. L. (2000). *J. Phys. Chem. A*, **104**, 10958–10971.

Mooij, W. T. M., van Eijck, B. P. & Kroon, J. (1999). *J. Phys. Chem. A*, **103**, 9883–9890.

Motherwell, W. D. S. (2001). *Mol. Cryst. Liq. Cryst.* **356**, 559–567.

Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Dzyabchenko, A., Erk, P., Gavezzotti, A., Hofmann, D. W. M., Leuson, F. J. J., Lommerse, J. P. M., Mooij, W. T. M., Price, S. L., Scheraga, H., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P. & Williams, D. E. (2002). *Acta Cryst.* B**58**, 647–661.

Moult, J., Hubbard, T., Fidelis, K. & Perdersen, J. T. (1999). *Proteins: Struct. Funct. Genet.* **37**, 2–6.

Nyburg, S. C. & Faerman, C. H. (1985). *Acta Cryst.* B**41**, 274–279.

Parsons, S. (2005). To be submitted.

Paulus, E. F., Acs, A. & Schmidt, M. U. (2005). In preparation.

Pillardy, J., Arnautova, Y. A., Czaplewski, C., Gibson, K. D. & Scheraga, H. A. (2001). *Proc. Nat. Acad. Sci. USA*, **98**, 12351–12356.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in FORTRAN*. Cambridge Unviersity Press.

Price, S. L., Willock, D. J., Leslie, M. & Day, G. M. (2001). *DMAREL*. http://www.ucl.ac.uk/~ucca17p/dmarelmanual/dmarel.html.

Schmidt, M. U. & Englert, U. (1996). *J. Chem. Soc. Dalton Trans.* pp. 2077–2082.

Schmidt, M. U. & Kalkhof, H. (1997). *CRYSCA*. Frankfurt, Clariant GmbH.

Signorini, G. F., Righini, R. & Schettino, V. (1991). *Chem. Phys.* **154**, 245–261.

Sobol', I. M. (1967). *Comput. Math. Math. Phys.* **7**, 86–112.

Spek, A. L. (1980–2004). *PLATON*. http://www.cryst.chem.uu.nl/platon. Utrecht University, The Netherlands.

Steiner, T. (2000). *Acta Cryst.* B**56**, 673–676.

Watkin, D. J., Motherwell, W. D. S., Cooper, R. I. & Pantos, S. (2004). *Acta Cryst.* E**60**, o2295–o2297.

Willock, D. J., Price, S. L., Leslie, M. & Catlow, C. R. A. (1995). *J. Comput. Chem.* **16**, 628–647.

Yu, F. & Schwalbe, C. H. (2002). *Structure of hydantoin from a twinned crystal.* British Crystallographic Association Spring Meeting. Nottingham, UK.

Yu, F.-L., Schwalbe, C. H. & Watkin, D. J. (2004). *Acta Cryst.* C**60**, o714–o717.