

# Crystal structure prediction of small organic molecules: a second blind test

W. D. Sam Motherwell,<sup>a\*</sup>  
Herman L. Ammon,<sup>b</sup> Jack D.  
Dunitz,<sup>c</sup> Alexander  
Dzyabchenko,<sup>d</sup> Peter Erk,<sup>e</sup>  
Angelo Gavezzotti,<sup>f</sup>  
Detlef W. M. Hofmann,<sup>g</sup>  
Frank J. J. Leusen,<sup>h</sup> Jos P. M.  
Lommerse,<sup>i</sup> Wijnand T. M.  
Mooij,<sup>h,p</sup> Sarah L. Price,<sup>j</sup> Harold  
Scheraga,<sup>k</sup> Bernd Schweizer,<sup>c</sup>  
Martin U. Schmidt,<sup>l</sup> Bouke P. van  
Eijck,<sup>m</sup> Paul Verwer<sup>n</sup> and  
Donald E. Williams<sup>o†</sup>

<sup>a</sup>Cambridge Crystallographic Data Centre, 12  
Union Road, Cambridge CB2 1EZ, UK,

<sup>b</sup>Department of Chemistry and Biochemistry,  
University of Maryland, College Park, MD  
20742-2021, USA, <sup>c</sup>Organic Chemical Labora-  
tory, ETH-Zurich, CH-8093 Zurich, Switzerland,

<sup>d</sup>Karpov Institute of Physical Chemistry, Voront-  
sovo pole 10, 103064 Moscow, Russia,

<sup>e</sup>Performance Chemicals Research, BASF AG,  
67056 Ludwigshafen, Germany, <sup>f</sup>Dipartimento

di Chimica Strutturale e Stereochimica Inorga-  
nica, via Venezian 21, 20133 Milano, Italy,

<sup>g</sup>GMD-SCAI, Schloss Berlinghoven, D-53754 St  
Augustin, Germany, <sup>h</sup>Accelrys Ltd, 230/250 The  
Quorum, Barnwell Road, Cambridge CB5 8RE,  
UK, <sup>i</sup>Doelenstraat 17, 5348 JR Oss, The

Netherlands, <sup>j</sup>Centre for Theoretical and  
Computational Chemistry, Department of  
Chemistry, University College, 20 Gordon  
Street, London WC1H 0AJ, UK, <sup>k</sup>Baker Labora-

tory of Chemistry, Cornell University, Ithaca, NY  
14853-1301, USA, <sup>l</sup>Clariant GmbH, Pigment  
Technology Research, G834, D-65926 Frankfurt  
am Main, Germany, <sup>m</sup>Bijvoet Centre for

Biomolecular Research, Utrecht University,  
Padualaan 8, 3584 CH Utrecht, The Nether-  
lands, <sup>n</sup>Solid State Chemistry Group and CMBI,  
University of Nijmegen, PO Box 9010, 6500 GL

Nijmegen, The Netherlands, <sup>o</sup>Department of  
Chemistry, University of Louisville, Louisville,  
KY 40292-2001, USA, and <sup>p</sup>Astex Technology  
Ltd, 250 Cambridge Science Park, Cambridge  
CB4 0WE, UK

† Deceased

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

© 2002 International Union of Crystallography  
Printed in Great Britain – all rights reserved

The first collaborative workshop on crystal structure predic-  
tion (CSP1999) has been followed by a second workshop  
(CSP2001) held at the Cambridge Crystallographic Data  
Centre. The 17 participants were given only the chemical  
diagram for three organic molecules and were invited to test  
their prediction programs within a range of named common  
space groups. Several different computer programs were used,  
using the methodology wherein a molecular model is used to  
construct theoretical crystal structures in given space groups,  
and prediction is usually based on the minimum calculated  
lattice energy. A maximum of three predictions were allowed  
per molecule. The results showed two correct predictions for  
the first molecule, four for the second molecule and none for  
the third molecule (which had torsional flexibility). The  
correct structure was often present in the sorted low-energy  
lists from the participants but at a ranking position greater  
than three. The use of non-indexed powder diffraction data  
was investigated in a secondary test, after completion of the *ab  
initio* submissions. Although no one method can be said to be  
completely reliable, this workshop gives an objective measure  
of the success and failure of current methodologies.

## 1. Introduction

Two major challenges appear to confront the predictive ability  
of theoretical and computational chemistry today: one is  
protein folding and the other is crystallization of organic  
compounds. There are obvious similarities. Both involve  
delicate balances between attractions and repulsions at the  
atomic level, between potential energy and entropic contri-  
butions to the free energy, and between thermodynamic and  
kinetic factors. Blind tests on the folding of proteins have been  
conducted in recent times (Orengo *et al.*, 1999). Here we  
report on a similar venture in crystal structure prediction  
(CSP) carried out in two stages in 1999 and 2001. Although  
early lack of progress in CSP was termed a 'continuing  
scandal' in *Nature* in 1988 (Maddox, 1988), and in spite of  
isolated claims of minor victories, the problem is now gener-  
ally recognized to be much more difficult than had been  
apparent. It is now seen to be not so much a matter of  
generating stable crystal structures but rather one of selecting  
one or more from many almost equi-energetic possibilities.  
Our successes and failures point the way to a better under-  
standing of the polymorphism phenomenon and also have  
practical implications for crystal engineering and design.

## 2. Approach and methodology

This paper reports on the results of a second blind test, known  
as CSP2001, which was part of a collaborative workshop held

Received 14 January 2002

Accepted 27 March 2002

*Dedicated in memoriam Jan  
Kroon*

at the Cambridge Crystallographic Data Centre (CCDC) in May 2001. The results of the first blind test, CSP1999, have already been published (Lommerse *et al.*, 2000). The arrangement of the blind test was as in CSP1999. Personal invitations were sent to about 25 researchers known to be active in the field and a total of 18 individuals agreed to participate. The list of unpublished structures was collected by personal contacts with about 30 laboratories known to be active in the small-molecule field. To give a reasonable chance of success within the practical computation limits of known computer programs, the maximum number of atoms including H atoms was set as 40; the space group was required to be in one of the ten most frequent as recorded in the Cambridge Structural Database (CSD) (Allen & Kennard, 1993), *i.e.*  $P2_1/c$ ,  $P\bar{1}$ ,  $P2_12_12_1$ ,  $C2/c$ ,  $P2_1$ ,  $Pbca$ ,  $Pna2_1$ ,  $Cc$ ,  $Pbcn$  and  $C2$  (in CSD frequency order); there should be one molecule per asymmetric unit and no solvent molecules or co-crystals. It was specified to the experimentalists that there should be no disorder, and the positions of all H atoms should be located experimentally. There were three categories of perceived difficulty for prediction:

- (i) rigid molecule with only C, H, N and O atoms, less than 25 atoms,
- (ii) rigid molecule with some less common elements (*e.g.* Br), less than 30 atoms,
- (iii) flexible molecule with two degrees of acyclic torsional freedom, less than 40 atoms.

An independent referee, Professor Tony Kirby, University Chemical Laboratory, Cambridge, was asked to select one molecule from each category and, if possible, to avoid molecules likely to be of near-planar conformation, as this turned out to be a bias in the CSP1999 selection. The referee had no access to the space group or crystal structure information, only to a list of chemical diagrams. The selected three chemical diagrams, IV, V and VI (Fig. 1), were sent by e-mail to the participants on 11 October 2000. The participants were asked to submit a maximum of three prediction structures for each molecule to the referee by midnight of 25 March 2001, with reasons for their selection and presentation in order of confidence. These are referred to in this paper as the 'ab initio predictions'.

An optional secondary test of prediction was also arranged, where the participants were supplied with simulated X-ray powder diffraction patterns for each molecule as extra information. They were given a second deadline date of 11 April 2001. The patterns were generated by CCDC after obtaining the experimental coordinates from the referee on 26 March 2001. These secondary submissions are known as the 'powder-assisted predictions' and are given in a separate section towards the end of this paper. On 12 April 2001, the experimental crystal structures were released to all participants, giving some time for post-analysis and preparation for the workshop meeting held in Cambridge on 10–11 May 2001.

To assist the reader in assessing the overall success and failure rate in these tests, the results of the CSP1999 workshop have been included in this paper. The full list of molecules for both workshops (Fig. 1), the full range of computer program

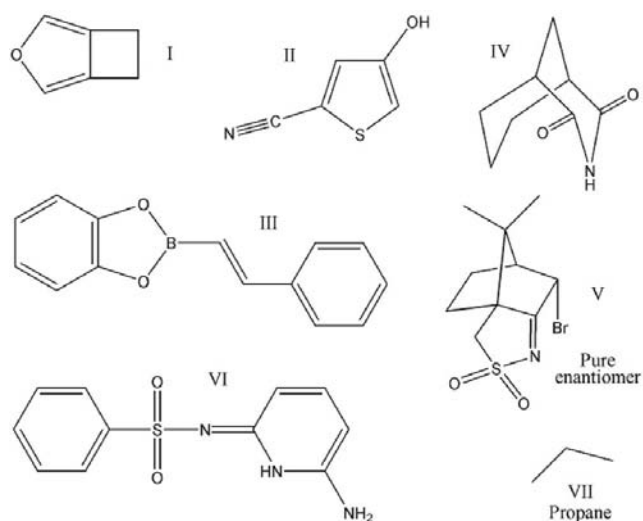
methodology (Table 1) and a summary of the results (Table 2) are given as combined tables for CSP1999 and CSP2001.

### 3. Methodology

Methods in the CSP tests are summarized in Table 1. Comprehensive reviews of computer methodology for crystal structure prediction have been published where many references are given to detailed publications (Gdanitz, 1997; Verwer & Leusen, 1998). All the methods involve three stages:

- (a) construct a three-dimensional molecular model either by molecular mechanics methods or by analogy with other CSD structures;
- (b) search through many thousands of hypothetical crystal structures built from the trial molecule in various space groups, including some searches that did not assume symmetry constraints;
- (c) select structures according to some criterion, usually the calculated lattice energy.

The search algorithms are quite diverse, and force fields range from simple transferable atom–atom potentials to elaborate computer-intensive models for the electrostatic and other contributions to the intermolecular potential. One or two models included explicit allowance for polarization effects. The most common selection criterion is the global minimum in lattice energy, and the most important discovery for CSP within the past decade is the recognition that many discrete structural possibilities exist within an energy window of only a few  $\text{kJ mol}^{-1}$  above the global minimum. For example, for acetic acid there are about 100 calculated structures within  $5 \text{ kJ mol}^{-1}$  (Mooij *et al.*, 1998), although only one polymorph at ambient pressure has been found experimen-



**Figure 1**  
The molecular diagrams given to the participants in the CSP workshops (I–III, VII for CSP1999; IV–VI for CSP2001). Experimental structure references: I (Boese & Garbarczyk, 1998), II (Blake *et al.*, 1999), III (Clegg *et al.*, 2001), IV (Howie & Skakle, 2001), V (Fronczek & Garcia, 2001), VI (Hursthouse, 2001), VII (Boese *et al.*, 1999).

**Table 1**

Overview of methodologies applied for crystal structure prediction for the blind test.

Contributor	Molecules attempted	Program/approach	Reference	Molecular model	Search generation
Methods employing lattice-energy minimization for generation of structures					
Gavezzotti	III, V	<i>ZIP-PROMET</i>	<i>a</i>	Rigid	Stepwise construction of dimers and layers
Schweizer & Dunitz	I, IV	<i>ZIP-PROMET</i>	<i>a</i>	Rigid	Stepwise construction of dimers and layers
Williams	I–VII	<i>MPA</i>	<i>b</i>	Flexible	Lattman grid systematic
Erk	IV–VI	<i>SySe</i> and <i>PP</i>	<i>c</i>	Flexible	Grid-based systematic
van Eijck	I, III–VII	<i>UPACK</i>	<i>d</i>	Flexible	Grid-based and random
Dzyabchenko	IV–VI	<i>PMC</i>	<i>e</i>	Flexible	Symmetry-adapted grid systematic
Schmidt	I–VI	<i>CRYSKA</i>	<i>f</i>	Flexible	Random plus steepest descent
Ammon	I–VI	<i>MOLPAK</i>	<i>g</i>	Rigid	Grid-based systematic
Price	I–V	<i>DMAREL</i>	<i>h</i>	Rigid	Using <i>MOLPAK</i>
Scheraga	IV–VI	<i>CRYSTALG</i>	<i>i</i>	Flexible	Conformation family Monte Carlo
Verwer & Leusen	I–III, VII	<i>Polymorph Predictor (PP)</i>	<i>j</i>	Flexible	Monte Carlo simulated annealing
Leusen	IV–VI	<i>Polymorph Predictor (PP)</i>	<i>j</i>	Flexible	Monte Carlo simulated annealing
Verwer	IV–VI	<i>Polymorph Predictor (PP)</i>	<i>j</i>	Flexible	Monte Carlo simulated annealing
Mooij	I, III, VII	Multipole crystal optimizer	<i>k</i>	Flexible	By van Eijck ( <i>UPACK</i> )
Mooij	IV–VI	Multipole crystal optimizer	<i>k</i>	Flexible	By Leusen & Verwer ( <i>PP</i> )
Methods based on statistical data from CSD					
Hofmann	I–III	<i>FlexCryst</i>	<i>l</i>	Rigid	Grid-based systematic
	IV–VI	<i>FlexCryst</i>	<i>m</i>	Rigid	Grid-based systematic
Lommerse	I–V, VII	<i>Packstar</i>	<i>n</i>	Rigid	Monte Carlo simulated annealing
Motherwell	I–V, VII	<i>Rancel</i>	<i>o</i>	Rigid	Genetic algorithm

Contributor	Lattice energy/fitness function		Other features used to select three submissions
	Electrostatic	Other	
Methods employing lattice-energy minimization for generation of structures			
Gavezzotti	None	Empirical	
Schweizer & Dunitz	Atom charges	6-exp	
Williams	Atom charges + extra sites	6-exp	
Erk	Atom charges	6-exp	
van Eijck	Atom charges	6-exp or 6–12	Free Energy
Dzyabchenko	Atom charges	6-exp or 6–12	
Schmidt	Atom charges	6-exp	Volume, chemical intuition
Ammon	Atom charges	6-exp	Density
Price	Atom multipoles	Empirical /derived	Morphology and elastic constants
Scheraga	Atom charges	6-exp or 6–12	
Verwer & Leusen	Atom charges	Dreiding 6–12	
Leusen	Atom charges	CVFF 6–12	
Verwer	Atom charges	Dreiding 6–12	
Mooij	Atom multipoles	<i>Ab initio</i> 6-exp + polarization	
Mooij	Atom multipoles	Dreiding 6-exp	
Methods based on statistical data from CSD			
Hofmann	Statistical potentials		
	Trained potentials		
Lommerse	CSD group contacts		
Motherwell	None	6-exp	Energy plus fitting of CSD contacts

References: (a) Gavezzotti (1991); (b) Williams (1996); (c) Erk (1999); (d) van Eijck & Kroon (2000); (e) Dzyabchenko *et al.* (1999); (f) Schmidt & Englert (1996); (g) Holden *et al.* (1993); (h) Beyer *et al.* (2001); (i) Pillardy *et al.* (2001); (j) Verwer & Leusen (1998); (k) Mooij *et al.* (1999); (l) Hofmann & Lengauer (1997); (m) Apostolakis *et al.* (2001); (n) Lommerse *et al.* (2000); (o) Motherwell (2001).

tally. Most search methods included the ‘correct’ structure somewhere in the list, but it was frequently not the structure with the lowest lattice energy. Besides, small changes in the potentials can reshuffle the energy ordering. Most calculated structures are ‘temperature-less’ in the sense that no temperature is specified in the computational procedure, but some include estimates of the free energy. There are also attempts to use pattern recognition based on the Cambridge Structural Database of experimentally determined molecular crystals. Although the importance of the kinetic aspects of crystal nucleation and growth is widely recognized, they remain largely unexplored.

#### 4. Overview of results

The submitted results for the *ab initio* predictions are given for molecules IV (Table 3), V (Table 4) and VI (Table 5). For the combined tests CSP1999 and CSP2001, the correct predictions are summarized in Table 2. Since there were so many contributors who worked independently, it was thought best to provide first an overview of the results (§4) and some general conclusions (§6). In the supplementary material,<sup>1</sup> we provide

<sup>1</sup>Supplementary data for this paper are available from the IUCr electronic archives (Reference: BK0108). Services for accessing these data are described at the back of the journal.

**Table 2**

Summary of successful predictions.

The experimental structures are labelled Expt and printed in bold. For the experimental structures,  $P$  gives the number of successful predictions, and for the predicted structures,  $P$  is the order of confidence in the three submissions allowed. RMSD-Pack is the calculated r.m.s. deviation of the non-H atom positions from experimental positions. The decision as to a correct solution has been based on a visual assessment of the packing diagrams.

Molecule	$P$	Space group	$a$ (Å)	$b$ (Å)	$c$ (Å)	$\beta$ (°)	RMSD-Pack (Å)
<b>I Expt stable</b>	<b>0</b>	<b><math>P2_1/c</math></b>	<b>4.954</b>	<b>9.845</b>	<b>9.679</b>	<b>90.57</b>	
<b>I Expt Metastable</b>	<b>4</b>	<b><math>Pbca</math></b>	<b>5.309</b>	<b>12.648</b>	<b>14.544</b>	<b>90</b>	
Schweizer	1	$Pbca$	5.182	12.554	14.336	90	0.204
Williams	1	$Pbca$	5.125	12.503	14.104	90	0.277
Verwer & Leusen	1	$Pbca$	5.372	12.570	15.131	90	0.231
van Eijck	3	$Pbca$	5.276	12.468	14.390	90	0.525
<b>II Expt</b>	<b>1</b>	<b><math>P2_1/n</math></b>	<b>7.516</b>	<b>8.322</b>	<b>9.059</b>	<b>101.19</b>	
Verwer & Leusen	2	$P2_1/n$	7.234	8.299	9.210	104.53	0.427
<b>III Expt</b>	<b>1</b>	<b><math>P2_1/c</math></b>	<b>6.835</b>	<b>7.634</b>	<b>21.422</b>	<b>96.45</b>	
van Eijck	1	$P2_1/c$	6.763	7.758	20.940	98.32	0.214
<b>IV Expt</b>	<b>3</b>	<b><math>P2_1/c</math></b>	<b>9.388</b>	<b>10.606</b>	<b>7.704</b>	<b>95.03</b>	
Leusen	3	$P2_1/c$	9.182	10.509	8.024	83.02	0.261
Mooij	2	$P2_1/c$	9.229	10.406	7.963	96.13	0.200
<b>V Expt</b>	<b>3</b>	<b><math>P2_12_12_1</math></b>	<b>7.264</b>	<b>10.639</b>	<b>15.633</b>	<b>90</b>	
Price	1	$P2_12_12_1$	7.177	10.413	16.223	90	0.347
Williams†	3	$P2_12_12_1$	6.930	10.660	15.580	90	0.263
van Eijck‡	1	$P2_12_12_1$	7.119	9.984	15.891	90	0.777
Ammon§	1	$P2_12_12_1$	7.128	10.394	16.354	90	0.364
<b>VI Expt</b>	<b>0</b>	<b><math>P2_1/c</math></b>	<b>8.251</b>	<b>8.964</b>	<b>15.087</b>	<b>91.21</b>	
<b>VII Expt</b>	<b>1</b>	<b><math>P2_1/n</math></b>	<b>4.148</b>	<b>12.612</b>	<b>6.977</b>	<b>91.28</b>	
Mooij	1	$P2_1/n$	4.057	12.568	6.777	91.66	0.163

† Williams submitted a structure in space group  $Cc$ , which is an error. If ignored, this makes the rank  $P = 2$ . ‡ Correct packing but a large value 0.777 is due to molecular conformation differences because of an inadequate force field. § Although strictly speaking not allowed within the rules of the blind test, this result was the global minimum within chiral space groups. Structures in centrosymmetric space groups for the racemate were submitted in error.

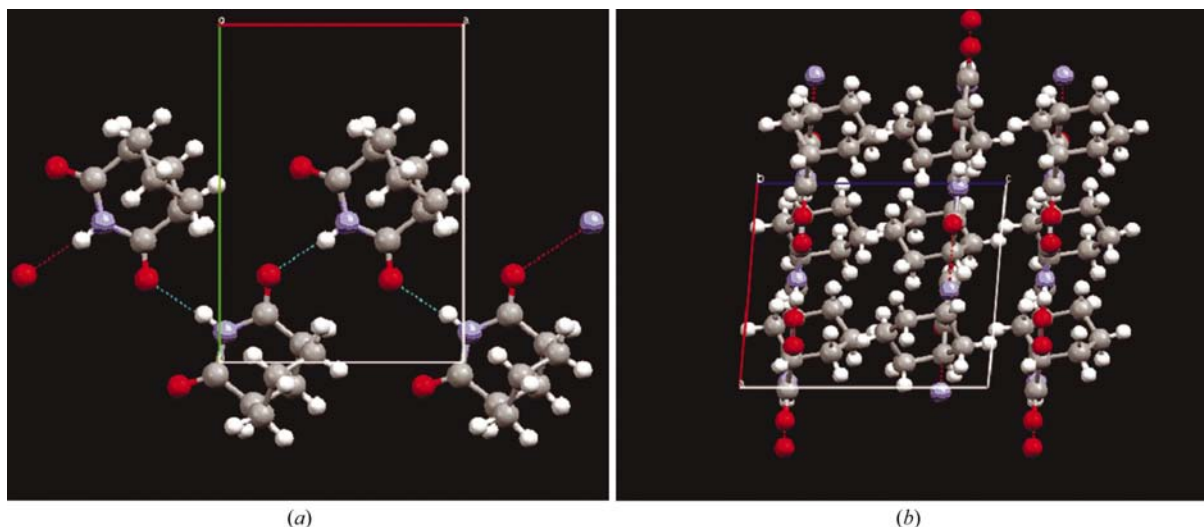
details of calculations and discussions prepared by each participant, under a named author subsection.

#### 4.1. Description of the experimental structures

A few comments on the experimentally determined structures are now given to demonstrate some of the challenges of prediction.

Compound IV (Howie & Skakle, 2001), in  $P2_1/a$ , shows hydrogen bonding in the packing diagram in Fig. 2. Inspection of related molecules in the CSD – those containing the

CH–CO–NH–CO–CH group in a ring system, with no other strong hydrogen-bond donors or acceptors – shows both dimer R2,2,(8) and catemer S1,1,(4) hydrogen-bond motifs (Allen *et al.*, 1999). The observed hydrogen-bond motif is a catemer, –NH···OC– mediated by the glide-plane operator in the  $a$  direction, and is almost exactly planar with N and O deviations of *ca.* 0.15 Å from the least-squares plane through the C, N, O and H atoms. The N–H···O distance of 1.973 Å is typical from CSD surveys, with almost optimal geometry: angles N–H···O = 171° and H···O=C = 129°, calculated using a normalized neutron N–H distance of 1.009 Å. The

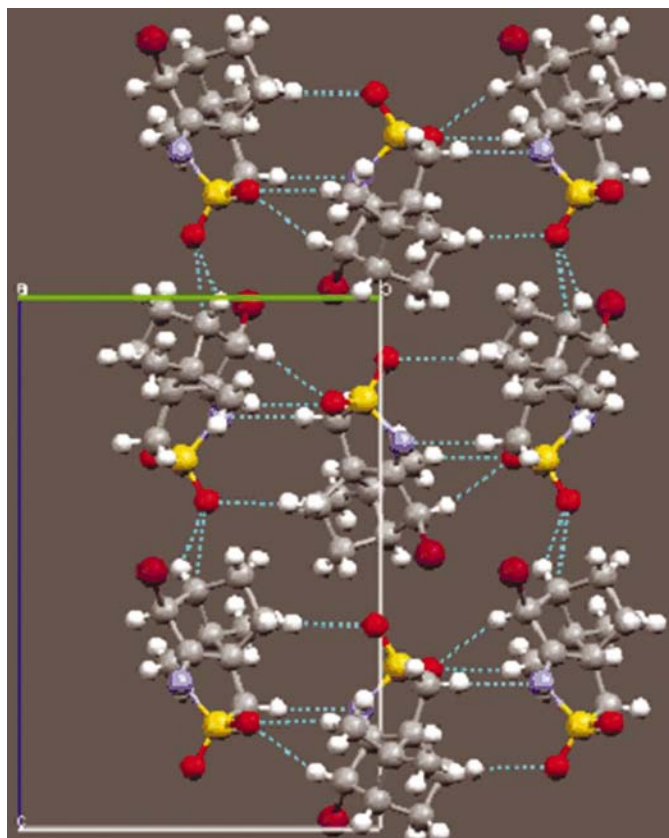


**Figure 2** Packing diagram for IV (a) showing hydrogen-bonded chains and (b) showing packing of chains.

other carbonyl O takes no part in hydrogen bonding. It was noted that there is a rather short intermolecular H···H contact of 2.118 Å between methylene groups related by a crystallographic centre of symmetry, but such contacts are found in some CSD structures of rather similarly sized molecules (e.g. AZTCDO10 2.199, BADNUP 2.157, 2.178).

Compound V (Fronczek & Garcia, 2001), in  $P2_12_12_1$  and known in advance to be a pure enantiomer, has no strong hydrogen-bonding groups, and the packing diagram (Fig. 3) does not show any particularly dominant group–group interactions. Intermolecular contacts are normal compared to similar molecules in the CSD; the O atoms have several C–H···O contacts (2.365, 2.381, 2.425, 2.593, 2.646 Å) substantially below the van der Waals radius sum. The Br atoms show no close contacts but do form a Br···Br chain distance of 4.427 Å using the screw axis along  $a$ . The five-membered ring containing S and N is infrequent in the CSD, but there is an entry for the de-brominated compound ROLBOJ, which has a similar ring conformation.

Compound VI (Hursthouse, 2001), in  $P2_1/c$ , is strongly hydrogen bonded (Fig. 4), forming a ribbon network running in the  $b$  direction mediated by the screw axis. It is notable that all donor H atoms are satisfied, and all acceptor O and N atoms are involved. It was observed that the bond lengths appear to be of low accuracy, despite the excellent hydrogen-



**Figure 3**

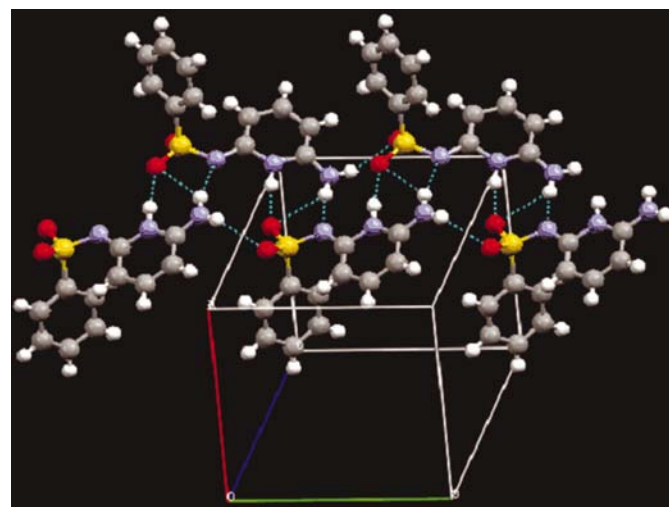
Packing diagram for V. There is no strong hydrogen bonding, but several C–H···O contacts are apparent. All contacts less than the sum of the van der Waals radii are shown.

bonding scheme, and subsequent communication with the laboratory revealed that there was a problem with very small crystals and a very low number of collected intensities. It was requested that a constrained refinement be made using the known phenyl geometry and isotropic temperature factors. The coordinate differences between the first and second refinements do not invalidate the accuracy of the packing arrangement for the purposes of this blind test. Apart from the two flexible torsional angles, an additional difficulty for CSP was that the S–N=C–N configuration might be either *cis* or *trans*.

#### 4.2. Comparison of calculated structures with experimental

A preliminary inspection of the submitted results using standard visualizer programs quickly revealed that many structures were completely different from the experimentally determined ones. The structures that visually seemed to show the same packing arrangement and similar cell dimensions were generally easy to accept as ‘correct’ as regards the overall packing arrangement. As in the CSP1999 test, we used the comparison method by Lommerse (Lommerse *et al.*, 2000) to compare the molecular coordination shell and derive an r.m.s. deviation for the non-H atoms for all atoms in the reference molecule and its 12 neighbours (RMSD-Pack; these calculations were performed by Lommerse before the workshop event). The lists of unit cells, space groups and RMSD-Pack are given for molecules IV (Table 3), V (Table 4) and VI (Table 5).

For correct structures in CSP1999, this figure was found to be in the range 0.163–0.525 Å. In practice, ‘incorrect’ structures show such a large RMSD that there is no problem in deciding; in this test, the range for correct structures was 0.200–0.364 Å. Only one case was found where there was a difficult decision, with a larger RMSD of 0.777 (van Eijck structure V, rank 1). This structure has the same symmetry-related 12 neighbours in the molecular coordination shell as



**Figure 4**

Packing diagram for VI. Selective view showing the hydrogen-bonding scheme, mediated by a screw axis along  $b$ . Note that all H donors are satisfied, and all acceptors have at least one H contact.

**Table 3**

Submitted results for molecule IV.

Results are presented in the space-group settings as submitted. Correct predictions are given in bold type. RMSD-Pack is calculated by the Lommerse method and is only given when a meaningful fit could be found within a certain tolerance.

Name	Space group	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)	RMSD-Pack (Å)
Experimental Ammon	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>a</i>	7.704	10.606	9.338	90.0	95.0	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	10.159	7.927	9.899	90.0	77.0	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	7.623	12.255	8.341	90.0	90.0	90.0	
Dzyabchenko	<i>P</i> $\bar{1}$	7.307	5.835	10.233	76.8	95.1	111.5	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	8.900	7.840	13.047	90.0	126.1	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	9.232	8.550	12.156	90.0	127.7	90.0	
Erk	<i>P</i> $\bar{1}$	5.667	6.450	10.918	86.8	81.7	79.2	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	9.096	8.146	10.650	90.0	97.1	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	10.065	8.021	10.146	90.0	104.8	90.0	
Hofmann	<i>Pbca</i>	12.031	11.527	11.719	90.0	90.0	90.0	
	<i>P</i> $\bar{1}$	6.949	6.801	8.124	87.4	89.5	85.1	
	<i>P</i> $\bar{1}$	6.819	5.937	10.416	90.5	92.4	62.8	
Leusen	<i>P</i> $\bar{1}$	6.892	6.423	10.368	77.2	82.8	61.3	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>a</i>	9.958	7.596	10.474	90.0	105.2	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	11.538	5.955	11.346	90.0	90.0	90.0	
Lommerse	<b><i>P</i><sub>2</sub><sub>1</sub>/<i>a</i></b>	<b>8.024</b>	<b>10.509</b>	<b>9.182</b>	<b>90.0</b>	<b>83.0</b>	<b>90.0</b>	<b>0.261</b>
	<i>P</i> <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	8.091	9.500	9.998	90.0	90.0	90.0	
	<i>Pbca</i>	11.579	11.785	11.145	90.0	90.0	90.0	
Mooij	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	6.567	10.529	12.407	90.0	77.4	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	10.247	7.706	9.962	90.0	76.3	90.0	
	<b><i>P</i><sub>2</sub><sub>1</sub>/<i>c</i></b>	<b>9.229</b>	<b>10.406</b>	<b>7.963</b>	<b>90.0</b>	<b>83.9</b>	<b>90.0</b>	<b>0.200</b>
Motherwell	<i>Pbca</i>	11.974	11.366	11.560	90.0	90.0	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	8.037	6.527	14.097	90.0	90.0	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub>	6.288	7.926	7.668	90.0	100.7	90.0	
Price	<i>Pbca</i>	11.748	11.638	11.152	90.0	90.0	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	11.129	6.142	15.531	90.0	134.3	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	6.144	7.094	18.148	90.0	87.4	90.0	
Scheraga	<i>Pbca</i>	11.526	11.859	11.482	90.0	90.0	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	10.112	7.918	9.697	90.0	77.0	90.0	
	<i>Pbca</i>	12.003	11.196	11.379	90.0	90.0	90.0	
Schweizer	<i>P</i> $\bar{1}$	9.976	7.173	5.707	109.9	104.1	83.9	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	8.434	6.543	15.774	90.0	88.4	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	6.199	15.101	10.352	90.0	116.9	90.0	
Schmidt	<i>C</i> 2/ <i>c</i>	11.295	12.271	12.965	90.0	81.7	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	10.087	7.415	9.793	90.0	103.6	90.0	
	<i>Pbca</i>	11.394	11.696	10.948	90.0	90.0	90.0	
van Eijck	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	9.284	8.541	11.721	90.0	128.1	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	10.262	7.537	9.826	90.0	104.5	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	11.232	11.292	5.916	90.0	90.0	90.0	
Verwer	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	9.071	7.843	12.596	90.0	56.0	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>n</i>	9.132	8.108	10.662	90.0	97.0	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	10.171	7.990	10.034	90.0	75.9	90.0	
Williams	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	6.226	10.901	12.482	90.0	76.8	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	10.420	7.480	9.910	90.0	77.1	90.0	
	<i>P</i> <sub>2</sub> <sub>1</sub> / <i>c</i>	6.370	12.160	10.180	90.0	102.0	90.0	
	<i>C</i> 2/ <i>c</i>	22.280	10.290	6.890	90.0	96.2	90.0	

the experimental structure, and the higher RMSD is explained by the fact that the intramolecular force field was unable to reproduce the correct puckering of the five-membered ring.

At the workshop discussion, it was decided that a more detailed comparison of submitted structures to the experimental structures would be of interest. Also those participants who had energy-ranked lists of structures (lowest energy is rank 1, next rank 2 *etc.*) were invited to contribute these to identify whether a match with the experimental structure could be found at a rank higher than 3. The comparison results for molecules IV, V and VI are given in Tables 6, 7 and 8, respectively, and these tables allow a comparison of how accurately the different force fields reproduced these structures. The *ab initio* submissions are given first in each table, followed by higher-ranked structures from the lists with

energy differences from the lowest value. In some cases where the structure was not found in the energy list, authors have presented a 'minimized experimental' (ME) structure, using their relevant force field to test how well the force field does describe this energy minimum. In these cases, the structure has no rank in the list, but the symbol ME is given.

This detailed comparison of structures with the experimental reference was performed after the workshop by Dzyabchenko, using the program *CRYCOM* (Dzyabchenko, 1994). The first step was to bring each pair of structures (target and reference) to the same space-group setting whenever they were not the same. Atom connectivity matching was automatically carried out by the CSD program *GEOM78*. The rigid-body parameters (*i.e.* the centre of mass coordinates and the three Euler angles of both molecules) were calculated with

**Table 4**

Submitted results for molecule V.

Results are presented in the space-group settings as submitted. Correct predictions are given in bold type. RMSD-Pack is calculated by the Lommerse method and is only given when a meaningful fit could be found within a certain tolerance.

Name	Space group	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)	RMSD-Pack (Å)
Experimental <b>Ammon</b>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	7.2643	10.6393	15.6331	90	90	90	<b>0.364</b>
	<b><i>P</i>2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></b>	<b>10.394</b>	<b>16.354</b>	<b>7.128</b>	<b>90</b>	<b>90</b>	<b>90</b>	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	10.799	12.802	8.608	90	90	90	
Dzyabchenko	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	10.595	11.524	9.884	90	90	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	12.959	10.44	8.36	90	90	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	7.906	8.931	15.959	90	90	90	
Erk	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	13.351	8.524	10.083	90	90	90	
	<i>P</i> 2 <sub>1</sub>	8.04	10.508	7.446	90	104.4	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	14.319	11.008	7.571	90	90	90	
Gavezzotti	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	7.463	14.716	10.96	90	90	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	11.858	7.015	13.178	90	90	90	
	<i>P</i> 2 <sub>1</sub>	6.9771	12.04	7.422	90	116.1	90	
Hofmann	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	11.720	9.638	10.058	90	90	90	
	<i>P</i> 1	6.874	9.962	8.441	95.5	80.6	100.3	
	<i>P</i> 2 <sub>1</sub> / <i>c</i>	10.876	9.285	15.602	90	49.9	90	
Leusen	<i>P</i> 2 <sub>1</sub> / <i>c</i>	10.718	9.285	16.000	90	49.2	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	7.336	12.11	13.343	90	90	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	12.391	6.924	13.628	90	90	90	
Lommerse	<i>P</i> 2 <sub>1</sub>	7.158	10.485	8.247	90	76.1	90	
	<i>P</i> 2 <sub>1</sub>	7.711	10.744	8.16	90	97.7	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	9.486	11.243	11.584	90	90	90	
Mooij	<i>P</i> 2 <sub>1</sub>	7.481	9.233	9.095	90	97.14	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	13.144	7.228	11.939	90	90	90	
	<i>P</i> 2 <sub>1</sub>	7.096	10.549	8.545	90	112.8	90	
Motherwell	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	10.746	9.982	10.848	90	90	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	7.955	8.485	16.424	90	90	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	7.602	14.106	10.353	90	90	90	
<b>Price</b>	<i>P</i> 2 <sub>1</sub>	8.804	10.919	8.224	90	46.67	90	<b>0.347</b>
	<b><i>P</i>2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></b>	<b>16.223</b>	<b>10.413</b>	<b>7.177</b>	<b>90</b>	<b>90</b>	<b>90</b>	
	<i>P</i> 2 <sub>1</sub>	7.218	10.703	8.269	90	67.55	90	
Scheraga	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	10.859	12.907	8.562	90	90	90	
	<i>P</i> 2 <sub>1</sub>	7.215	11.266	8.811	90	60.31	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	9.967	11.528	10.76	90	90	90	
Schmidt	<i>P</i> 2 <sub>1</sub>	7.309	10.236	8.454	90	78	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	8.920	9.214	13.332	90	90	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	6.742	12.018	13.687	90	90	90	
van Eijck†	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	7.277	8.708	17.461	90	90	90	<b>0.777</b>
	<b><i>P</i>2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></b>	<b>9.985</b>	<b>15.891</b>	<b>7.119</b>	<b>90</b>	<b>90</b>	<b>90</b>	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	7.949	11.386	12.397	90	90	90	
Verwer	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	14.651	8.524	8.716	90	90	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	7.178	13.323	12.216	90	90	90	
	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	12.853	7.381	12.375	90	90	90	
Williams‡	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	11.171	10.679	10.013	90	90	90	
	<i>C</i> <i>c</i>	6.91	15.97	10.53	90	81.26	90	
	<i>P</i> 2 <sub>1</sub>	8.12	10.81	6.95	90	70.28	90	
	<b><i>P</i>2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></b>	<b>10.66</b>	<b>6.93</b>	<b>15.58</b>	<b>90</b>	<b>90</b>	<b>90</b>	<b>0.263</b>

† van Eijck's result has the correct crystal packing, but a large RMSD owing to differences in molecular model conformation. The RMSD given was calculated by Dzyabchenko's program. ‡ Williams's submission in *Cc* is an error for this chiral molecule.

reference to a defined set of molecular axes. These six parameters together with the unit-cell parameters were used as the basis for comparison.

The target structure (*i.e.* the prediction structure) was matched against the reference experimental structure with all equivalent descriptions (ED) of the former taken into account. These ED were generated from the original one by changing the direction and the origin of the unit-cell axes in all possible ways compatible with the given space group; these are given by the so-called affine normalizer group derivative for the space group. Whenever assignment of local axes allowed ambiguity because of molecular symmetry (as in compound

IV), the list of ED was further expanded by virtue of the point-group operations.

For the flexible molecule VI, the comparison involved consideration of sets of rigid-body parameters of three constituent fragments: SO<sub>2</sub>, the phenyl and the remaining hetero-N-aromatic group. Each fragment was treated independently as if it were a single rigid molecule, with its particular point-group symmetry taken into account and with a common condition that the cell transformation and origin shift are the same.

As a result of this rigid-body treatment, the deviations in the cell dimensions, the net centre of mass translation and the

**Table 5**

Submitted results for molecule VI.

Results are presented in the space-group settings as submitted. There were no correct predictions. RMSD-Pack is calculated by the Lommerse method and is only given when a meaningful fit could be found within a certain tolerance.

Name	Space group	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)	RMSD-Pack (Å)
Experimental	$P2_1/c$	8.2506	8.9643	15.087	90	91.21	90	
	$P\bar{1}$	11.508	6.676	7.614	85.9	95.3	81.2	
Ammon	$P2_1/c$	7.551	23.099	6.794	90	82.5	90	
	$P2_1/c$	7.739	6.683	22.817	90	82.9	90	
	$Pbca$	10.862	8.379	23.845	90	90	90	
	$Pbca$	9.317	9.85	24.697	90	90	90	
Dzyabchenko	$Pbca$	9.351	10.345	22.923	90	90	90	
	$C2c$	12.634	7.6702	24.832	90	81.0	90	
	$C2c$	16.505	10.896	14.139	90	62.5	90	
Erk	$P2_1/c$	9.369	16.983	7.932	90	70.1	90	
	$P\bar{1}$	10.886	7.632	8.062	120.8	93.9	97.6	
	$P\bar{1}$	5.385	11.543	10.84	69.8	65.2	3.6	
Hofmann	$P2_1/c$	10.743	15.792	7.107	90	111.9	90	
	$P2_1/a$	15.941	8.976	7.801	90	86.0	90	
	$P2_1/a$	11.893	13.649	7.569	90	114.0	90	
Leusen	$P2_1/c$	8.086	8.674	16.118	90	98.0	90	
	$P\bar{1}$	10.663	8.738	9.473	92.3	55.7	60.2	
	$P2_1/c$	14.106	5.895	16.626	90	126.1	90	
Mooij	$Pbca$	23.316	8.798	10.753	90	90	90	
	$P2_1/c$	9.008	12.857	15.817	90	133.51	90	
	$P2_1/c$	7.656	11.14	17.797	90	118.81	90	
Scheraga	$P2_1/c$	7.921	14.937	11.197	90	101.27	90	
	$C2c$	22.866	5.533	16.734	90	91.3	90	
	$P\bar{1}$	6.852	7.775	11.157	83.7	73.6	69.9	
Schmidt	$P2_1/c$	4.946	9.306	23.119	90	96.5	90	
	$P\bar{1}$	9.847	30.491	21.458	4.87	90.49	90.51	
	$P\bar{1}$	9.338	16.234	8.379	34.45	94.10	71.97	
van Eijck	$P2_1/c$	13.021	7.681	11.940	90	62.98	90	
	$P2_1/a$	7.010	24.520	6.657	90	85.2	90	
	$An$	7.054	24.541	6.602	90	93.8	90	
Verwer	$Pbca$	24.384	7.134	13.281	90	90	90	
	$P2_1/c$	13.31	12.03	7.15	90	101.26	90	
	$P2_1/c$	14.06	11.73	6.98	90	76.54	90	
Williams	$Pbca$	7.83	11.99	23.96	90	90	90	

net rotation angle of the molecule (or each of the rigid fragments in molecule VI) were determined.

In addition, an alternative method of comparison was performed in an atom–atom matching mode, where each atom of the molecule was formally treated as an independent fragment, with rotation ignored. This automatically resulted in fully standardized lists of coordinates where respective atoms of both reference and target occur in the same order. These lists were then used to calculate the RMSD (packing and molecular) and the torsion angles in molecule VI.

The full list of all experimental and submitted structure coordinates, authors' calculated energies, comments on selection criteria, and comparisons of standardized sets of coordinates with the experimental structures is provided as supplementary material. This supplement also contains a personal account by each author of their procedures and discussion of their criteria for selection.

### 4.3. Overall success rate

The crystal structure of molecule IV was predicted correctly in two cases out of 15 submissions, with energy ranking of 2 and 3, respectively. This might be an unexpectedly poor success rate, since we have a rigid molecule with limited element types (C, H, N, O) and the hydrogen-bonding

empirical parameters for  $\text{NH}\cdots\text{O}=\text{C}$  have been studied and developed over many years. Nevertheless, it is noteworthy that hydrogen bonding was not treated explicitly in the two correct predictions for molecule IV. In CSP1999, the comparable structure (I in *Pbca*) was correctly predicted four times out of 11 submissions, with energy ranks 1, 1, 1 and 3. If the rule had been relaxed in 2001 to allow five predictions using energy ranking, there would have been three more correct predictions allowed (see Table 6). If we had allowed only one prediction then in fact no-one would have succeeded! Examination of energy differences shows that within the first three ranks we are dealing with a range of only  $0.5 \text{ kJ mol}^{-1}$  (Leusen) and  $1.0 \text{ kJ mol}^{-1}$  (Mooij), which is certainly within the error of the energy calculation. Other predictions showed a range of only about  $2\text{--}8 \text{ kJ mol}^{-1}$ , which is of the same order as the uncertainty caused by neglect of entropy.

The crystal structure of molecule V was predicted correctly four times out of 15 submissions, with energy ranks 1, 1, 1 and 3. As mentioned already, the structure submitted by van Eijck is correctly packed, but the RMSD is higher because of differences in the molecular conformation, so it can be counted as a correct prediction of the packing structure. This compound containing Br required significant work in order to model the contribution of Br to the intermolecular forces



**Table 6**

Molecule IV: comparison of predicted structures with the experimental structure.

(a) Cell data  $sa$ ,  $sb$ ,  $sc$  and  $s\beta$  are relative deviations  $(\text{Cal} - \text{Exp})/\text{Exp}$  in percents.  $E$ -rank refers to the order in the sorted list of ascending energy values by this author. ME means minimized experimental coordinates.  $\Delta E$  is the energy difference from the lowest-energy structure in the list found by this author.

Structure	Choice or $E$ -rank	$\Delta E$ (kJ mol <sup>-1</sup> )	$a$ (Å)	$b$ (Å)	$c$ (Å)	$\beta$ (°)	$sa$ (%)	$sb$ (%)	$sc$ (%)	$s\beta$ (%)
Experimental			9.34	10.61	7.71	95.0	0	0	0	0
<i>Ab initio</i>										
Leusen	3	0.5	9.18	10.51	8.02	97.0	-1.7	-0.9	4.1	2.0
Mooij	2	0.2	9.23	10.41	7.96	96.1	-1.1	-1.9	3.3	1.1
$E$ -rank > 3 or ME										
Ammon	ME	5.6	9.15	10.66	7.73	95.1	-2.0	0.5	0.3	0.1
Dzyabchenko	31	6.1	9.16	10.60	7.73	95.9	-1.9	-0.3	0.2	1.0
Erk $PP^\dagger$	116	23.	9.47	10.77	7.85	96.2	1.5	1.5	1.9	1.2
Hofmann	358	-	9.22	10.20	7.63	94.6	-1.3	-3.8	-0.9	-0.4
Price	ME	7.8	9.31	10.59	7.91	94.5	-0.2	-0.1	2.7	0.5
Scheraga	5	4.3	9.09	10.59	7.80	95.3	-2.6	-0.2	1.2	0.2
Schmidt‡	9	2.9	8.94	10.52	7.69	95.4	-4.3	-0.8	-0.2	-0.4
Schweizer & Dunitz§	(1)	0	9.73	11.41	7.86	95.6	4.2	7.6	2.0	0.6
van Eijck‡	5	2.4¶	9.10	10.51	7.79	97.0	-2.6	-0.9	1.0	2.0
Verwer	209	12.8	9.53	10.71	7.84	96.1	2.0	1.0	1.8	1.2
Williams	4	-	9.22	10.68	7.79	95.1	-1.2	0.7	1.2	0.1
Powder-assisted based on indexation or Rietveld refinement ††										
Dzyabchenko‡‡	(31)		9.34	10.59	7.71	95.0	0.0	-0.1	0.1	-0.0
Schmidt§§	(9)		9.33	10.60	7.67	94.6	-0.1	-0.1	-0.4	-0.4
Verwer	(209)		9.35	10.64	7.70	94.9	0.1	0.3	0.0	-0.1

(b) Deviations of the predicted structures from the experimental as defined by the rigid-body net translation ( $t$ ) and net rotation ( $\omega$ ) of the molecule and the atomic r.m.s. deviations, conformational (Conf) and packing. Pack-D is defined as the r.m.s. of the  $N$  ( $N$  = number of atoms in the asymmetric unit, H atoms included) quantities  $d = |A_{\text{mean}}(\mathbf{x}_{\text{cal}} - \mathbf{x}_{\text{exp}})|$  calculated for the respective atoms of the predicted and observed structures, where  $\mathbf{x} = (x, y, z)$  fractional coordinates of the atom and  $A_{\text{mean}}$  is the matrix to convert the fractional coordinates into orthogonal ones based on the mean cell parameters of the two unit cells. Pack-L is defined in Lommerse *et al.* (2000).

Structure	$t$ (Å)	$\omega$ (°)	RMSD (Å)		
			Conf	Pack-D	Pack-L
<i>Ab initio</i>					
Leusen	0.049	2.5	0.149	0.171	0.261
Mooij	0.031	1.6	0.104	0.127	0.200
$E$ -rank > 3 or ME					
Ammon	0.030	1.0	0.084	0.102	
Dzyabchenko	0.090	4.6	0.126	0.206	
Erk $PP^\dagger$	0.071	3.4	0.130	0.161	
Hofmann	0.091	3.8	0.145	0.167	
Price	0.049	0.8	0.095	0.099	
Scheraga	0.201	5.5	0.088	0.292	
Schmidt‡	0.049	2.3	0.080	0.141	0.188
Schweizer/Dunitz	0.152	4.8	0.208	0.205	
van Eijck‡	0.055	0.8	0.149	0.189	0.221
Verwer	0.047	2.8	0.119	0.129	
Williams	0.038	1.0	0.050	0.072	
Powder-assisted based on indexation or Rietveld refinement					
Dzyabchenko‡‡	0.051	2.5	0.126	0.168	
Schmidt§§	0.043	2.1	0.080	0.117	0.114
Verwer	0.047	2.8	0.100	0.128	0.128

† PP – the *Polymorph Predictor* result. ‡ Submitted as a powder-assisted structure; selected by comparison with experimental powder diagram. § Rank 1, but not submitted in the blind test. ¶ 2.4 is the free energy value, pure energy is 2.0. †† Ranking refers to the closest *ab initio* structure. ‡‡ Structure found by energy minimization based on the experimental cell and then refined by fitting of calculated X-ray intensities *versus* observed with constrained geometry. §§ Structure refined by energy minimization with experimental cell but no intensity fit.

adequately, and it shows a higher success rate than the comparable category of molecule V in CSP1999, where only one correct prediction occurred out of eight submissions. We can note from Table 7 that if the test had allowed five predictions per molecule, there would have been one further success. The energy differences between the global and the

third-lowest minima are rather larger for molecule V [approximately 3.6 (Ammon), 4.1 (Price), 0.5 (Williams) and 1.6 kJ mol<sup>-1</sup> (van Eijck)], but this may reflect the relatively smaller number of chiral space groups allowed.

Molecule VI, with its flexibility, *cis-trans* possibility and many hydrogen-bonding possibilities, showed no successful

**Table 7**

Molecule V: comparison of predicted structures with experimental structure.

(a) Headings as Table 6. An *E*-rank ME means minimized experimental coordinates. Cell parameters *a*, *b*, *c* and their percentage deviations *sa*, *sb*, *sc* of calculated versus observed are given.  $\Delta E$  is the energy difference to the lowest-energy structure.

Structure	Choice or <i>E</i> -rank	$\Delta E$ (kJ mol <sup>-1</sup> )	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	<i>sa</i> (%)	<i>sb</i> (%)	<i>sc</i> (%)
Experimental			7.26	10.64	15.63	0.0	0.0	0.0
<i>Ab initio</i>								
Ammon†	1	0	7.13	10.39	16.35	-1.9	-2.3	4.6
Price	1	0	7.18	10.41	16.22	-1.2	-2.1	3.8
van Eijck	1‡	0	7.12	9.98	15.89	-2.0	-6.1	1.6
Williams	2	0.12	6.93	10.66	15.58	-4.6	0.2	-0.3
<i>E</i> -rank > 3 or ME								
Dzyabchenko	5	2.4	7.57	10.01	15.06	4.2	-5.9	-3.7
Erk <i>PP</i> §	High	22	7.20	10.60	15.99	-0.9	-0.3	2.3
Gavezzotti	14 or 15?	4.8	7.05	10.35	15.53	-2.9	-2.7	-0.6
Hofmann	746		7.20	10.50	15.50	-0.9	-1.3	-0.8
Leusen¶	70	10.4	7.36	10.94	15.59	1.3	2.8	-0.3
Mooij¶	9	6.7	7.14	10.80	15.38	-1.8	1.5	-1.6
Scheraga	ME		7.07	10.57	16.11	-2.7	-0.6	3.0
Schmidt	46	5.9	6.81	10.06	16.73	-6.2	-5.5	7.0
Verwer	6	1.2	7.30	10.96	15.06	0.5	3.0	-3.7
Powder-assisted based on indexation††								
Dzyabchenko‡‡	(5)		7.26	10.63	15.63	-0.0	-0.0	0.0
Verwer	(6)		7.26	10.63	15.62	-0.1	-0.0	-0.1

(b) Deviations in the rigid-body parameters and RMSDs for atoms.

Structure	<i>t</i> (Å)	$\omega$ (°)	RMSD (Å)		
			Conf	Pack-D	Pack-L
<i>Ab initio</i>					
Ammon†	0.192	2.6	0.137	0.279	0.364
Price	0.210	3.1	0.130	0.283	0.347
van Eijck	0.561	11.7	0.367	0.777	
Williams	0.160	1.6	0.098	0.187	0.263
<i>E</i> -rank > 3 or ME					
Dzyabchenko	0.245	14.7	0.204	0.581	
Erk <i>PP</i> §	0.890	15.4	0.223	1.041	
Gavezzotti	0.182	4.9	0.164	0.309	
Hofmann	0.192	0.7	0.000	0.190	
Leusen¶	0.300	5.9	0.186	0.387	0.400
Mooij¶	0.229	1.0	0.146	0.284	0.294
Scheraga	0.173	4.0	0.000	0.239	
Schmidt	0.530	8.0	0.157	0.649	
Verwer	0.313	3.0	0.168	0.373	
Powder-assisted based on indexation					
Dzyabchenko‡‡	0.084	6.5	0.176	0.289	0.286
Verwer	0.291	4.9	0.175	0.375	0.367

† Resubmitted after racemic structures given in error. ‡ Free-energy ranking (rank 4 if just potential energy). § *PP* – the *Polymorph Predictor* result. ¶ Submitted on powder step. †† Ranking refers to closest *ab initio* structures. ‡‡ Structure found by energy minimization based on the experimental cell and then refined by fitting of calculated X-ray intensities versus observed with constrained geometry.

prediction in 11 submissions. The equivalent category of molecule VI in CSP1999 was in fact correctly predicted once in 11 submissions. The difficulty with VI is presumably in the sensitivity of empirical hydrogen-bonding potentials to small movements in the H...N distance and, in the case of those models using multipole electrostatics, the fact that the electrostatics vary with molecular conformation. Table 8 shows the post-analysis energy minimum comparison results, where in some cases the correct structure was found but at a very high energy rank (*viz.* 54, 79 and 340), and in other cases the

correct structure did not appear at all in the list. It is interesting to see that the ME results show that the correct structure has an energy minimum with the various force fields with only quite small distortions.

#### 4.4. Comparison of energy minima

It was noticeable that several groups submitted incorrect structures that were very similar, and a summary of these coincident energy minima for the *ab initio* submissions is given in Table 9 for molecules IV and V, again prepared by

**Table 8**

Molecule VI: comparison of predicted structures with experimental structure.

(a) Cell data dimensions and their differences from experimental structure, *sa*, *sb*, *sc* and *sβ*, expressed as percentages of *a*, *b*, *c* and *β*. An *E*-rank ME means minimized experimental coordinates.

	<i>E</i> -rank	$\Delta E$ (kJ mol <sup>-1</sup> )	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\beta$ (°)	<i>sa</i> (%)	<i>sb</i> (%)	<i>sc</i> (%)	<i>sβ</i> (%)
Experimental			8.25	8.96	15.09	91.2	0.0	0.0	0.0	0.0
<i>E</i> -rank > 3 or ME										
Dzyabchenko	79‡	16.7	8.33	9.72	14.82	100.8	1.0	8.4	-1.7	10.5
Erk <i>PP</i> ‡	54	8.8	8.55	9.21	15.06	88.1	3.7	2.8	-0.2	-3.4
Hofmann	ME		8.30	8.80	15.00	90.5	0.6	-1.8	-0.5	-0.8
Mooij	ME	16.6	8.65	9.20	14.45	84.3	4.9	2.6	-4.2	-7.6
Scheraga	ME		9.17	10.43	13.00	92.2	11.1	16.3	-13.8	1.1
Schmidt	ME	6.6	8.26	8.90	14.90	95.0	0.1	-0.7	-1.3	4.2
van Eijck	340‡	13.0	8.41	9.18	14.24	91.5	1.9	2.4	-5.6	0.3
Verwer	733	32.7	8.60	8.85	15.35	86.9	4.2	-1.2	1.8	-4.7
Williams	ME		8.35	9.04	14.67	89.9	1.2	0.9	-2.8	-1.5
Powder-assisted based on indexation§										
Dzyabchenko¶			8.24	8.95	15.06	91.2	-0.1	-0.2	-0.2	0.0
Leusen <i>PS</i> ††			8.24	8.95	15.07	91.2	-0.1	-0.1	-0.1	0.0

(b) Deviation parameters *t* and  $\omega$  for three rigid molecular fragments constituting the flexible molecule V: the central SO<sub>2</sub> group, the heteroaromatic planar system C<sub>4</sub>H<sub>6</sub>N<sub>2</sub> and the phenyl residue C<sub>6</sub>H<sub>5</sub>. Torsion angles  $\tau_1$ (OSNC) and  $\tau_2$ (OSCC) for the predicted structures and their deviations ( $\Delta\tau$ ) from those in the experimental structure ( $\tau_1 = 508^\circ$ ,  $\tau_2 = 316^\circ$ ) are also reported.

	<i>t</i> (Å)			$\omega$ (°)			$\tau_1$ (°)	$\tau_2$ (°)	$\Delta\tau_1$ (°)	$\Delta\tau_2$ (°)	RMSD (Å)	
	SO <sub>2</sub>	C <sub>4</sub> H <sub>6</sub> N <sub>2</sub>	C <sub>6</sub> H <sub>5</sub>	SO <sub>2</sub>	C <sub>4</sub> H <sub>6</sub> N <sub>2</sub>	C <sub>6</sub> H <sub>5</sub>					Conf	Pack
<i>E</i> -rank > 3 or ME												
Dzyabchenko	0.877	0.584	0.491	29.5	14.6	21.4	11.9	55.4	-38.9	23.8	0.619	0.742
Erk <i>PP</i> ‡	0.219	0.143	0.165	0.8	6.1	7.1	72.7	24.0	22.0	-7.6	0.265	0.262
Hofmann	0.156	0.111	0.153	0.4	0.4	0.4	50.8	31.6	0.0	0.0	0.000	0.128
Mooij	0.269	0.113	0.302	10.9	4.6	18.2	66.0	39.3	15.3	7.7	0.322	0.379
Scheraga	1.296	1.412	0.925	10.8	10.7	10.7	50.8	31.6	0.0	0.0	0.000	1.337
Schmidt	0.448	0.489	0.395	16.8	10.0	26.8	49.1	9.3	-1.7	22.3	0.219	0.831
van Eijck	0.338	0.359	0.296	9.4	7.8	9.1	59.1	18.6	8.3	-13.0	0.254	0.469
Verwer	0.166	0.220	0.052	8.5	7.8	9.7	56.9	34.4	6.1	2.8	0.191	0.299
Williams	0.075	0.099	0.070	3.1	3.1	3.1	50.8	31.6	0.0	0.0	0.028	0.148
Powder-assisted based on indexation												
Dzyabchenko¶	0.160	0.036	0.112	4.2	6.6	5.2	54.1	33.0	3.3	1.4	0.255	0.303
Leusen <i>PS</i> ††	0.035	0.071	0.147	5.0	3.4	6.9	54.3	31.8	3.5	0.2	0.214	0.231

‡ Ranking refers to the *trans*-isomer with regard to the S–N=C–N fragment. ‡ Erk *Polymorph Predictor*, search in five space groups. § Ranking refers to closest *ab initio* structures. ¶ Structure found by energy minimization based on the experimental cell and then refined by fitting of calculated X-ray intensities versus observed with constrained geometry. †† Solution obtained with *Powder Solve*.

Dzyabchenko using *CRYCOM*. The comparisons show an interesting convergence of different programs and force fields on the same minima, and the table shows the positions of such minima when they occurred at least twice in the total set of submissions (minima that occurred just once are omitted). Since no minimum can be considered as definitive, and these minima are not necessarily for the correct experimental structure, an arbitrary choice was made as a reference, so that the differences could be tabulated within each group. It can be seen that some minima have a quite small range of *t*, especially minimum *m1*, with maximum *t* = 0.097 Å and  $\omega$  = 3.0°. Cell dimensions differences are generally less than 6%.

It might be concluded from this comparison that there are some well defined low-energy crystal packings that can be found for a wide variety of force fields. The frequency of finding a particular minimum (the number of coincidences) does not predict the most likely polymorph. It might be

thought that these sets of coincident minima, such as *m1* and *o1* for molecule IV, represent likely alternative experimental polymorphs that could be obtained by different crystallization conditions. This might indeed be the case, and we invite experimentalists to search for other polymorphs. However, the frequencies of coincidences could reflect similarities in the search paths taken by the various programs exploring the energy surface; this is a mathematical construct of fitting a molecule into a given cell 'box' with predefined space-group symmetry, which is acknowledged as having no physical reality in the crystallization process.

## 5. Secondary test using simulated powder diffraction data

A secondary test was arranged to challenge the performance of CSP programs in the frequently encountered real-labora-

**Table 9**

Assignment of various energy minima showing coincidence of predictions by more than one participant.

The 'correct' prediction minima are highlighted in bold. The column headed Choice reflects the ranking 1, 2 or 3 as given in the submission and does not always correspond to the absolute energy ranking. Minimum gives a label to each minimum for discussion elsewhere.  $t$  is the difference in centre of mass coordinates, and  $\omega$  is the rotation angle. The cell dimension differences,  $sa$ ,  $sb$ ,  $sc$  and  $s\beta$ , are expressed as percentages of  $a$ ,  $b$ ,  $c$  and  $\beta$ .

Author	Choice or <i>E</i> -rank	Minimum	$a$ (Å)	$b$ (Å)	$c$ (Å)	$\beta$ (°)	$t$ (Å)	$\omega$ (Å)	$sa$ (%)	$sb$ (%)	$sc$ (%)	$s\beta$ (%)
Molecule IV												
Space group $P2_1/c$												
Ammon	1	IV- <i>m</i> 1	10.16	7.93	9.90	77.0	(Reference)					
Erk	2	IV- <i>m</i> 1	10.15	8.02	10.06	75.2	0.029	3.0	-0.1	1.2	1.7	-2.3
Leusen	1	IV- <i>m</i> 1	10.47	7.60	9.96	74.8	0.097	2.3	3.1	-4.2	0.6	-2.8
Mooij	1	IV- <i>m</i> 1	10.25	7.71	9.96	76.3	0.058	2.2	0.9	-2.8	0.6	-0.9
Scheraga	1	IV- <i>m</i> 1	10.11	7.92	9.70	77.0	0.082	3.3	-0.5	-0.1	-2.1	0.1
Schmidt	1	IV- <i>m</i> 1	10.09	7.41	9.79	76.4	0.080	2.8	-0.7	-6.5	-1.1	-0.8
van Eijck	1	IV- <i>m</i> 1	10.26	7.54	9.83	75.4	0.070	1.0	1.0	-4.9	-0.8	-2.0
Verwer	2	IV- <i>m</i> 1	10.17	7.99	10.03	75.5	0.037	2.7	0.1	0.8	1.4	-1.4
Williams	1	IV- <i>m</i> 1	10.42	7.48	9.91	77.0	0.086	1.0	2.6	-5.6	0.1	0.1
Dzyabchenko	1	IV- <i>m</i> 2	8.98	7.84	13.05	53.9	(Reference)					
Erk	1	IV- <i>m</i> 2	9.10	8.15	13.12	53.7	0.035	0.8	1.3	3.9	0.6	0.2
van Eijck	3	IV- <i>m</i> 2	9.07	7.84	12.60	56.0	0.119	3.3	1.0	0.0	-3.5	-1.7
Verwer	1	IV- <i>m</i> 2	9.13	8.11	13.17	53.5	0.033	0.8	1.7	3.4	0.9	0.3
<b>Leusen</b>	<b>3</b>	<b>IV-<i>m</i>3</b>	<b>9.18</b>	<b>10.51</b>	<b>8.02</b>	<b>83.0</b>	<b>(Reference)</b>					
<b>Mooij</b>	<b>2</b>	<b>IV-<i>m</i>3</b>	<b>9.23</b>	<b>10.41</b>	<b>7.96</b>	<b>83.8</b>	<b>0.033</b>	<b>1.1</b>	<b>0.5</b>	<b>-1.0</b>	<b>-0.8</b>	<b>1.0</b>
Dzyabchenko	2	IV- <i>m</i> 4	9.23	8.55	12.16	52.3	(Reference)					
Schmidt	3	IV- <i>m</i> 4	9.28	8.54	11.72	51.9	0.083	5.4	0.6	-0.1	-3.6	0.3
Lommerse	3	IV- <i>m</i> 5	6.57	10.53	12.41	77.4	(Reference)					
Verwer	3	IV- <i>m</i> 5	6.23	10.90	12.48	76.8	0.382	11.1	-5.2	3.5	0.6	-0.8
Space group $Pbca$												
Erk	3	IV- <i>o</i> 1	11.53	11.72	12.03	90.0	(Reference)					
Lommerse	2	IV- <i>o</i> 1	11.14	11.58	11.78	90.0	0.214	3.9	-2.0	-3.3	-1.2	0.0
Mooij	3	IV- <i>o</i> 1	11.37	11.56	11.97	90.0	0.067	0.7	-0.5	-1.4	-1.4	0.0
Motherwell	3	IV- <i>o</i> 1	11.64	11.15	11.75	90.0	0.295	3.4	-2.4	1.0	-4.8	0.0
Price	3	IV- <i>o</i> 1	11.86	11.48	11.53	90.0	0.350	2.2	-4.2	2.9	-2.0	0.0
Scheraga	2	IV- <i>o</i> 1	11.20	11.38	12.00	90.0	0.193	1.2	-0.2	-2.9	-2.9	0.0
Schmidt	2	IV- <i>o</i> 1	10.95	11.39	11.70	90.0	0.156	1.6	-2.8	-5.0	-2.8	0.0
Space group $P2_12_12_1$												
Leusen	2	IV- <i>o</i> 2	5.95	11.35	11.54	90.0	(Reference)					
van Eijck	2	IV- <i>o</i> 2	5.92	11.29	11.23	90.0	0.067	1.8	-2.7	-0.7	-0.5	0.0
Molecule V												
Space group $P2_12_12_1$												
Leusen	1	V- <i>o</i> 1	7.34	12.11	13.34	90.0	(Reference)					
Mooij	1	V- <i>o</i> 1	7.23	11.94	13.14	90.0	0.343	2.5	-1.5	-1.4	1.5	0.0
Verwer	2	V- <i>o</i> 1	7.38	12.37	12.85	90.0	0.285	3.3	0.6	2.2	-3.7	0.0
Leusen	2	V- <i>o</i> 2	6.92	12.39	13.63	90.0	(Reference)					
Erk	2	V- <i>o</i> 2	7.57	11.01	14.32	90.0	0.374	12.6	-11.2	9.3	5.1	0.0
Gavezzotti	1	V- <i>o</i> 2	7.01	11.86	13.18	90.0	0.134	3.9	-4.3	1.3	-3.3	0.0
Verwer	1	V- <i>o</i> 2	7.18	12.22	13.32	90.0	0.127	4.4	-1.4	3.7	-2.2	0.0
Dzyabchenko	2	V- <i>o</i> 3	7.91	8.93	15.96	90.0	(Reference)					
Motherwell	1	V- <i>o</i> 3	7.95	8.48	16.42	90.0	0.432	11.0	0.6	-5.0	2.9	0.0
Dzyabchenko	1	V- <i>o</i> 4	8.36	10.44	12.96	90.0	(Reference)					
Price	3	V- <i>o</i> 4	8.56	10.86	12.91	90.0	0.288	10.7	-0.4	4.0	2.4	0.0
Ammon	2	V- <i>o</i> 4	8.61	10.80	12.80	90.0	0.311	9.9	3.0	3.4	-1.2	0.0
Gavezzotti	3	V- <i>o</i> 5	9.37	10.06	11.72	90.0	(Reference)					
Mooij	3	V- <i>o</i> 5	9.98	10.75	10.85	90.0	0.232	4.9	-7.4	6.5	6.8	0.0
Scheraga	2	V- <i>o</i> 5	9.97	10.76	11.53	90.0	0.139	6.0	-1.6	6.4	7.0	0.0
Verwer	3	V- <i>o</i> 5	10.01	10.68	11.17	90.0	0.180	3.4	-4.7	6.9	6.2	0.0
Ammon†	3	V- <i>o</i> 5	9.88	10.59	11.52	90.0	0.190	5.7	5.5	5.3	-1.7	0.0
<b>Price</b>	<b>1</b>	<b>V-<i>o</i>6</b>	<b>7.18</b>	<b>10.41</b>	<b>16.23</b>	<b>90.0</b>	<b>(Reference)</b>					
<b>van Eijck</b>	<b>1</b>	<b>V-<i>o</i>6</b>	<b>7.12</b>	<b>9.98</b>	<b>15.89</b>	<b>90.0</b>	<b>0.388</b>	<b>10.7</b>	<b>-2.0</b>	<b>-4.1</b>	<b>-0.8</b>	<b>0.0</b>

Table 9 (continued)

Author	Choice or <i>E</i> -rank	Minimum	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\beta$ (°)	<i>t</i> (Å)	$\omega$ (Å)	<i>sa</i> (%)	<i>sb</i> (%)	<i>sc</i> (%)	<i>s</i> $\beta$ (%)
<b>Williams</b>	<b>3</b>	<b>V-06</b>	<b>6.93</b>	<b>10.66</b>	<b>15.58</b>	<b>90.0</b>	<b>0.251</b>	<b>1.7</b>	<b>-4.0</b>	<b>2.4</b>	<b>-3.4</b>	<b>0.0</b>
<b>Ammon†</b>	<b>1</b>	<b>V-06</b>	<b>7.13</b>	<b>10.39</b>	<b>16.35</b>	<b>90.0</b>	<b>0.052</b>	<b>0.5</b>	<b>-0.7</b>	<b>-0.2</b>	<b>0.8</b>	<b>0.0</b>
Space group <i>P2</i> <sub>1</sub>												
Leusen	3	V- <i>m</i> 1	7.16	10.48	8.25	76.1	(Reference)					
Lommerse	3	V- <i>m</i> 1	7.48	9.23	9.09	82.9	0.264	18.1	4.5	-11.9	10.3	8.9
Scheraga	3	V- <i>m</i> 1	7.31	10.24	8.45	78.0	0.593	13.3	2.1	-2.4	2.5	2.5
Erk	1	V- <i>m</i> 1	7.45	10.51	8.04	75.6	0.015	4.1	4.0	0.2	-2.5	-0.7
Mooij												
Motherwell	3	V- <i>m</i> 2	6.77	10.92	8.80	62.1	0.459	13.1	-4.6	3.5	3.0	4.5
Price	2	V- <i>m</i> 2	7.22	10.70	8.63	67.5	0.158	2.4	1.7	1.5	1.0	-0.3
Scheraga	1	V- <i>m</i> 2	7.21	11.27	8.17	69.6	0.616	7.7	1.7	6.8	-4.4	-2.1
Williams	2	V- <i>m</i> 2	6.95	10.81	8.12	70.3	0.489	6.6	-2.1	2.5	-5.0	-2.8

† Resubmitted after racemic structures given in error.

tory situation, where some limited low-quality powder diffraction data have been collected for the substance. The key point about any such powder data is that it should not be indexable, *i.e.* the unit-cell dimensions cannot be determined. It is now well established in the literature that if a unit cell can be indexed, and a reasonable molecular model can be defined, then a real-space fitting of the observed diffraction profile can quickly and reliably lead to the correct crystal structure (David *et al.*, 1998; Engel *et al.*, 1999). Most of the programs currently available for this real-space structure solution method will have no difficulty with rigid molecules or two flexible torsions. Therefore an expectation of any such test was that the participants would not index any supplied powder pattern but would rely only on profile comparison to select the correct structure from their low-energy lists.

It was not practical within the time schedule of CSP2001 to obtain real-laboratory X-ray powder diffraction data for the compounds IV, V and VI, although this would have been the ideal test. Instead, the CCDC arranged for a simulated powder pattern to be calculated from the experimental single-crystal coordinates after they were released by the referee on 26 March. The simulated patterns were given a reasonable amount of peak broadening, random noise was added and a laboratory background was added from other data collections. These patterns were provided to participants who had then approximately two weeks to work on their selections before the true coordinates were released. This optional test resulted in a total of 18 submissions, known here as the 'powder-assisted' results. The successful predictions are included in Tables 6, 7, 8 and 9. The simulated powder patterns and all submitted coordinates with authors' comments are included in the supplementary material.

For compound IV (Table 6), there were submissions from eight participants, of which two were judged to be correct using non-indexed powder data. These correct structures occurred in the full energy lists at ranks 5 and 9. There was the complication that some participants used a unit cell obtained by indexing the powder pattern, which was not the original

idea of the test; unfortunately it was not possible in the limited time available to produce simulated patterns with such peak broadening as to prevent indexing. Some other participants did not index but refined the cell obtained from their predictions using Rietveld refinement. Using indexed or refined cell data there were three more correct predictions, at energy ranks 9, 31 and 209.

For compound V (Table 7), there were submissions from seven participants, of which two were judged correct using non-indexed data, at energy ranks 9 and 70. There were a further two correct predictions using indexed cells, both at energy rank 5.

For compound VI (Table 8), there were submissions from three participants, none of which were successful using non-indexed data. One submission was successful using the indexed cell, at energy rank 79. (Another participant, after indexing, used a real-space structure-solution program, which, not surprisingly, produced a correct solution.)

## 6. Conclusions

Considering the combined results of CSP1999 and CSP2001 (Table 2), we can say that, for the rigid molecules I, II, IV, V and VII, the occurrence of predictions with an accuracy of a few percent in the cell dimensions represents significant progress. There was also a success with one flexible molecule, III, having two degrees of torsional freedom. If the arbitrary rule of submitting the best three structures from each method had been extended to six structures, say, the success quota would have been notably higher. However, it cannot yet be claimed that any of the methods used is consistently reliable. The total absence of success with molecule VI, which involves flexibility in only two torsion angles, is notable. Note also that if the rule had been to allow only one submission then six structures would be removed from Table 2.

In so far as an observed crystal structure represents a minimum in lattice energy, we need better search algorithms for flexible molecules and more than one molecule per

asymmetric unit, and more accurate models for intra- and intermolecular potentials. In some cases with two molecules per asymmetric unit, a solution is currently possible, provided one is prepared to spend the greatly increased time searching the larger parameter space. Such time will become available as computing power increases. Methods that have already enjoyed some success will become more successful. We hope that within a few years reasonably accurate and reliable crystal structure prediction will be possible for rigid molecules containing C, H, N, O, S, P and halogen atoms. Difficulties will remain for crystals with uncommon space groups or with several molecules in the asymmetric unit.

To include the role of temperature on crystal structure and properties, we need to compare free energies rather than lattice energies. Reasonable estimates of the vibrational enthalpy and entropy contributions are already available through lattice dynamic calculations, but contributions from other sources, such as impurities, vacancies, grain boundaries and the like, will long remain intractable.

Crystallization of a compound from solution or from the melt is a non-equilibrium process, the outcome of which is determined at some degree of supersaturation or supercooling by formation of viable nuclei. These need not be those of the most stable structure, and hence there is no guarantee that an observed crystal structure is the thermodynamically stable form under any given conditions (Dunitz & Bernstein, 1995). A full dynamic treatment of the nucleation and growth stages is therefore called for, but that level of evolutionary modelling is beyond our wildest dreams at present.

The Cambridge Structural Database consists mainly of crystal structures for those polymorphs that form suitable single crystals and were obtained under normal laboratory temperature and pressure. As this database grows from its current 250 000 entries at a rate of about 10% per year, the pattern information latent in it may become useful for selection of the most likely polymorphs from low-energy sets (Desiraju, 1995). However, it must be said that experience so far suggests that the database is still too small in that there are often insufficient examples of specific types of compounds and substituents when faced with a particular prediction challenge.

In conclusion, the results of these CSP blind tests have provided an objective evaluation of the possibilities and limitations of current methods of crystal structure prediction. Crystal structure prediction, although beset by fundamental and technical difficulties, is no longer scandalously hopeless.

The authors wish to acknowledge the pioneering contribution of Donald Williams to the science of force-field methods in crystal structure prediction. Professor Williams was fortunately able to make his contribution to this paper before his death in August 2001. We thank Professor Tony Kirby for acting as referee and Dr Howie, Dr Fronczek and Professor

Hursthouse for provision of their unpublished crystal structures IV, V and VI, respectively. WDSM thanks E. Pidcock for assistance in organization of the workshop results and K. Shankland for preparation of the simulated powder patterns. We acknowledge the assistance of many colleagues involved in developing our programs and methods. In particular, HLA thanks Z. Du and Dr J. R. Holden, and the Office of Naval Research for financial support; AD thanks Dr F. Affouard for assistance in arranging computations on the workstation of LDSM in Lille-1 University; SLP thanks Graeme Day for performing the calculations on IV and Helen Tsui for deriving the new potential parameters for V; HS thanks Y. Arnautova, J. Pillardy and C. Czaplowski. AD is grateful for financial support from Conseil Regional de Nord-Pas de Calais and Russian Foundation for Basic Research (grant No. 99-03-32962) and WTMM for financial support from the Marie Curie Industry Host Fellowship programme of the European Commission.

## References

- Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 1; 31–37.
- Allen, F. H., Motherwell, W. D. S., Raithby, P. R., Shields, G. P. & Taylor, R. (1999). *New J. Chem.* pp. 25–34.
- Apostolakis, J., Hofmann, D. W. M. & Lengauer, T. (2001). *Acta Cryst.* **A57**, 442–450.
- Beyer, T., Day, G. M. & Price, S. L. (2001). *J. Am. Chem. Soc.* **123**, 5086–5094.
- Blake, A. J., Clark, B. A. J., Gierens, H., Gould, R. O., Hunter, G. A., McNab, H., Morrow, M. & Sommerville, C. C. (1999). *Acta Cryst.* **B55**, 963–974.
- Boese, R. & Garbarczyk, J. (1998). Private communication.
- Boese, R., Weiss, H.-C. & Blaser, D. (1999). *Angew. Chem. Int. Ed. Engl.* **38**, 988–992.
- Clegg, W., Marder, T. B., Scott, A. J., Wiesauer, C. & Weissensteiner, W. (2001). *Acta Cryst.* **E57**, o63–o65.
- David, W. I. F., Shankland, K. & Shankland, N. (1998). *Chem. Commun.* **8**, 931–932.
- Desiraju, G. R. (1995). *Angew. Chem. Int. Ed. Engl.* **34**, 2311–2327.
- Dunitz, J. & Bernstein, J. (1995). *Acc. Chem. Res.* **28**, 193–200.
- Dzyabchenko, A. V. (1994). *Acta Cryst.* **B50**, 414–425.
- Dzyabchenko, A. V., Agafonov, V. & Davydov, V. A. (1999). *J. Phys. Chem. A*, **103**, 2812–2820.
- Eijck, B. P. van & Kroon, J. (2000). *Acta Cryst.* **B56**, 535–542.
- Engel, G. E., Wilke, S., König, O., Harris, K. D. M. & Leusen, F. J. J. (1999). *J. Appl. Cryst.* **32**, 1169–1179.
- Erk, P. (1999). *Crystal Engineering: From Molecules and Crystals to Materials*. NATO Science Series C, Vol. 538, pp. 143–161. Dordrecht: Kluwer.
- Fronczek, F. R. & Garcia, J. G. (2001). *Acta Cryst.* **E57**, o886–o887.
- Gavezzotti, A. (1991). *J. Am. Chem. Soc.* **113**, 4622–4629.
- Gdanitz, R. J. (1997). *Theoretical Aspects and Computer Modeling of the Molecular Solid State*, pp. 185–202. New York: John Wiley & Sons.
- Hofmann, D. & Lengauer, T. (1997). *Acta Cryst.* **A53**, 225–235.
- Holden, J. R., Du, Z. Y. & Ammon, H. L. (1993). *J. Comput. Chem.* **14**, 422–437.
- Howie, R. A. & Skakle, J. M. S. (2001). *Acta Cryst.* **E57**, o822–o824.
- Hursthouse, M. (2001). Private communication.

- Lommerse, J. P. M., Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Mooij, W. T. M., Price, S. L., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P. & Williams, D. E. (2000). *Acta Cryst.* **B56**, 697–714.
- Maddox, J. (1988). *Nature (London)*, **335**, 201–201.
- Mooij, W. T. M., van Eijck, B. P. & Kroon, J. (1999). *J. Phys. Chem.* **A103**, 9883–9890.
- Mooij, W. T. M., van Eijck, B. P., Price, S. L., Verwer, P. & Kroon, J. (1998). *J. Comput. Chem.* **19**, 459–474.
- Motherwell, W. D. S. (2001). *Mol. Cryst. Liq. Cryst.* **356**, 559–567.
- Orengo, C. A., Bray, J. E., Hubbard, T., LoConte, L. & Sillitoe, I. (1999). *Proteins: Structure, Function and Genetics*, Suppl. **3**, 149–170.
- Pillard, J., Arnautova, Y. A., Czaplowski, C., Gibson, K. D. & Scheraga, H. A. (2001). *Proc. Nat. Acad. Sci. USA*, **98**, 12351–12356.
- Schmidt, M. U. & Englert, U. (1996). *J. Chem. Soc. Dalton Trans.* pp. 2077–2082.
- Verwer, P. & Leusen, F. J. J. (1998). *Rev. Comput. Chem.* **12**, 327–365.
- Williams, D. E. (1996). *Acta Cryst.* **A52**, 326–328.