Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

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

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

[a]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK, [b]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742-2021, USA, [c]Organic Chemical Laboratory, ETH-Zurich, CH-8093 Zurich, Switzerland, [d]Karpov Institute of Physical Chemistry, Vorontsovo pole 10, 103064 Moscow, Russia, [e]Performance Chemicals Research, BASF AG, 67056 Ludwigshafen, Germany, [f]Dipartmento di Chimica Strutturale e Stereochimica Inorganica, via Venezian 21, 20133 Milano, Italy, [g]GMD-SCAI, Schloss Berlinghoven, D-53754 St Augustin, Germany, [h]Accelrys Ltd, 230/250 The Quorum, Barnwell Road, Cambridge CB5 8RE, UK, [i]Doelenstraat 17, 5348 JR Oss, The Netherlands, [j]Centre for Theoretical and Computational Chemistry, Department of Chemistry, University College, 20 Gordon Street, London WC1H 0AJ, UK, [k]Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853-1301, USA, [l]Clariant GmbH, Pigment Technology Research, G834, D-65926 Frankfurt am Main, Germany, [m]Bijvoet Centre for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands, [n]Solid State Chemistry Group and CMBI, University of Nijmegen, PO Box 9010, 6500 GL Nijmegen, The Netherlands, [o]Department of Chemistry, University of Louisville, Louisville, KY 40292-2001, USA, and [p]Astex Technology Ltd, 250 Cambridge Science Park, Cambridge CB4 0WE, UK

† Deceased

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

Received 14 January 2002
Accepted 27 March 2002

Dedicated in memoriam Jan Kroon

The first collaborative workshop on crystal structure prediction (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 contributions 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 generally 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 equienergetic possibilities. Our successes and failures point the way to a better understanding 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

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 mole cules 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 infor mation. 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 experi mental 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 refer ences 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 kJ mol^{-1} above the global minimum. For example, for acetic acid there are about 100 calculated struc tures within 5 kJ mol^{-1} (Mooij *et al.*, 1998), although only one polymorph at ambient pressure has been found experimen

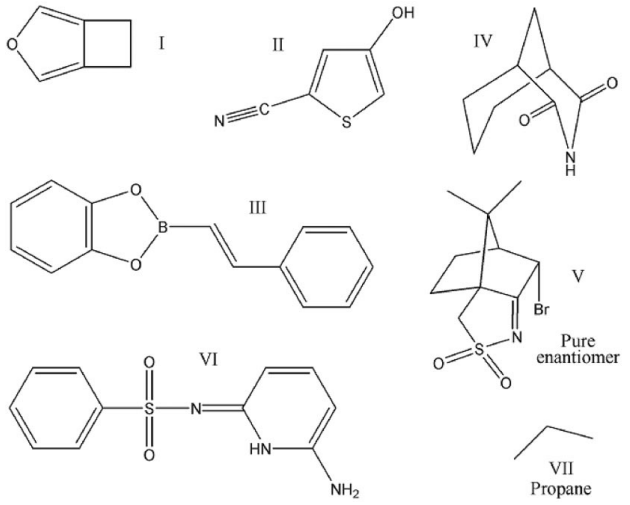


Figure 1
The molecular diagrams given to the participants in the CSP workshops (I III, VII for CSP1999; IV VI for CSP2001). Experimental structures references: I (Boese & Garbarczyk, 1998), II (Blake *et al.*, 1999), III (Clegg *et al.*, 2001), IV (Howie & Skakle, 2001), V (Fronczek & Garcia, 2001), 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	*ZIP-PROMET*	*a*	Rigid	Stepwise construction of dimers and layers
Schweizer & Dunitz	I, IV	*ZIP-PROMET*	*a*	Rigid	Stepwise construction of dimers and layers
Williams	I VII	*MPA*	*b*	Flexible	Lattman grid systematic
Erk	IV VI	*SySe* and *PP*	*c*	Flexible	Grid-based systematic
van Eijck	I, III VII	*UPACK*	*d*	Flexible	Grid-based and random
Dzyabchenko	IV VI	*PMC*	*e*	Flexible	Symmetry-adapted grid systematic
Schmidt	I VI	*CRYSCA*	*f*	Flexible	Random plus steepest descent
Ammon	I VI	*MOLPAK*	*g*	Rigid	Grid-based systematic
Price	I V	*DMAREL*	*h*	Rigid	Using *MOLPAK*
Scheraga	IV VI	*CRYSTALG*	*i*	Flexible	Conformation family Monte Carlo
Verwer & Leusen	I III, VII	*Polymorph Predictor* (*PP*)	*j*	Flexible	Monte Carlo simulated annealing
Leusen	IV VI	*Polymorph Predictor* (*PP*)	*j*	Flexible	Monte Carlo simulated annealing
Verwer	IV VI	*Polymorph Predictor* (*PP*)	*j*	Flexible	Monte Carlo simulated annealing
Mooij	I, III, VII	Multipole crystal optimizer	*k*	Flexible	By van Eijck (*UPACK*)
Mooij	IV VI	Multipole crystal optimizer	k	Flexible	By Leusen & Verwer (*PP*)
Methods based on statistical data from CSD					
Hofmann	I III	*FlexCryst*	*l*	Rigid	Grid-based systematic
	IV VI	*FlexCryst*	*m*	Rigid	Grid-based systematic
Lommerse	I V, VII	*Packstar*	*n*	Rigid	Monte Carlo simulated annealing
Motherwell	I V, VII	*Rancel*	*o*	Rigid	Genetic algorithm

	Lattice energy/fitness function		
Contributor	Electrostatic	Other	Other features used to select three submissions
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	*Ab initio* 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,[1] we provide

[1] 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)	*β* (°)	RMSD-Pack (Å)
I Expt stable	**0**	***P2$_1$/c***	**4.954**	**9.845**	**9.679**	**90.57**	
I Expt Metastable	**4**	***Pbca***	**5.309**	**12.648**	**14.544**	**90**	
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
II Expt	**1**	***P2$_1$/n***	**7.516**	**8.322**	**9.059**	**101.19**	
Verwer & Leusen	2	*P*2$_1$/*n*	7.234	8.299	9.210	104.53	0.427
III Expt	**1**	***P2$_1$/c***	**6.835**	**7.634**	**21.422**	**96.45**	
van Eijck	1	*P*2$_1$/*c*	6.763	7.758	20.940	98.32	0.214
IV Expt	**3**	***P2$_1$/c***	**9.388**	**10.606**	**7.704**	**95.03**	
Leusen	3	*P*2$_1$/*c*	9.182	10.509	8.024	83.02	0.261
Mooij	2	*P*2$_1$/*c*	9.229	10.406	7.963	96.13	0.200
V Expt	**3**	***P2$_1$2$_1$2$_1$***	**7.264**	**10.639**	**15.633**	**90**	
Price	1	*P*2$_1$2$_1$2$_1$	7.177	10.413	16.223	90	0.347
Williams†	3	*P*2$_1$2$_1$2$_1$	6.930	10.660	15.580	90	0.263
van Eijck‡	1	*P*2$_1$2$_1$2$_1$	7.119	9.984	15.891	90	0.777
Ammon§	1	*P*2$_1$2$_1$2$_1$	7.128	10.394	16.354	90	0.364
VI Expt	**0**	***P2$_1$/c***	**8.251**	**8.964**	**15.087**	**91.21**	
VII Expt	**1**	***P2$_1$/n***	**4.148**	**12.612**	**6.977**	**91.28**	
Mooij	1	*P*2$_1$/*n*	4.057	12.568	6.777	91.66	0.163

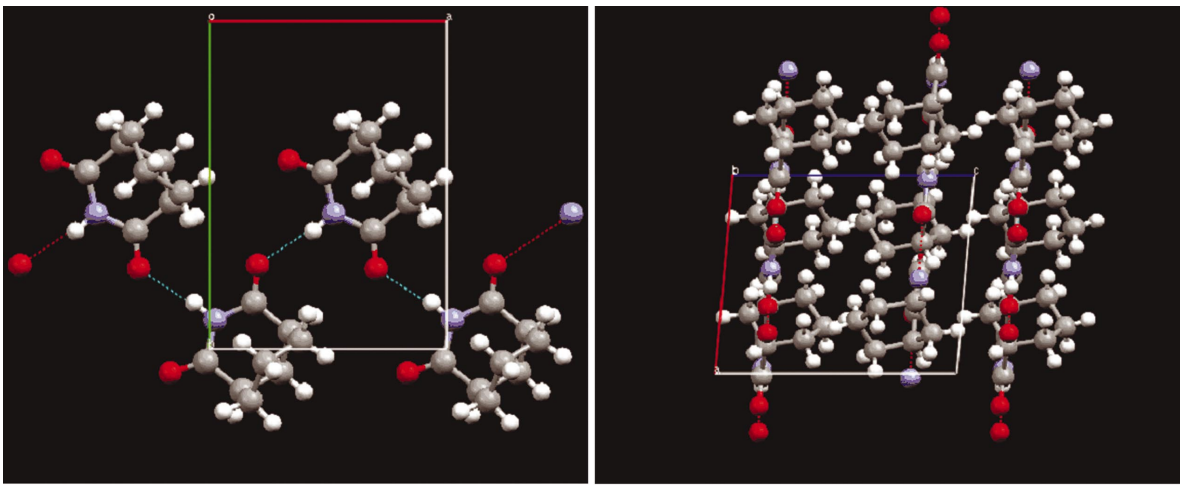
† Williams submitted a structure in space group *Cc*, which is an error. If ignored, this makes the rank *P* 2. ‡ Correct packing but a large value 0.777 is due to molecular conformation differences because of an inadequate force field. § Although strictly speaking not allowed within the rules of the 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 *P*2$_1$/*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



(*a*) (*b*)

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 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*.

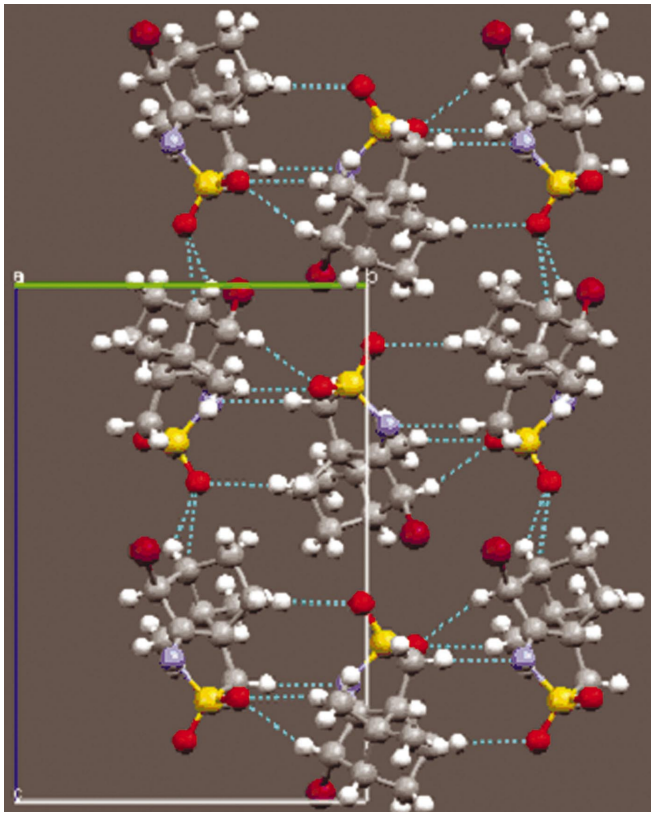


Figure 3
Packing diagram for V. There is no strong hydrogen bonding, but several C H···O contacts are apparent. All contacts less than the sum of the van der Waals radii are shown.

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

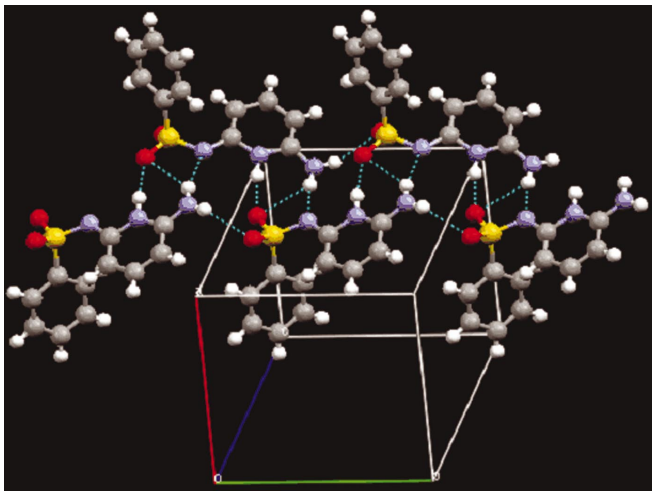


Figure 4
Packing diagram for VI. Selective view showing the hydrogen bonding scheme, mediated by a screw axis along *b*. Note that all H donors are satisified, and all acceptors have at least one H contact.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.