Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

A test of crystal structure prediction of small organic molecules

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

[a]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England, [b]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA, [c]Organic Chemistry Laboratory, Swiss Federal Institute of Technology, ETH-Zentrum, CH-8092 Zurich, Switzerland, [d]Dipartimento di Chimica Strutturale e Stereochimica Inorganica, Universita' di Milano, via Venezian 21, 20133 Milano, Italy, [e]Institute for Algorithms and Scientific Computing, German National Research Centre for Information Technology (GMD-SCAI), Schloss Birlinghoven, D-53754 St Augustin, Germany, [f]Molecular Simulations Ltd, 230/250 The Quorum, Barnwell Road, Cambridge CB5 8RE, England, [g]Department of Crystal and Structural Chemistry, Bijvoet Centre for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands, [h]Centre of Theoretical and Computational Chemistry, Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, England, [i]Clariant GmbH, Pigment Technology Research, G834, D-65926 Frankfurt am Main, Germany, [j]CAOS/CAMM Centre (CMBI), University of Nijmegen, PO Box 9010, 6500 GL Nijmegen, The Netherlands, and [k]Department of Chemistry, University of Louisville, Louisville, KY 40292, USA

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

Received 26 October 1999
Accepted 15 February 2000

A collaborative workshop was held in May 1999 at the Cambridge Crystallographic Data Centre to test how well currently available methods of crystal structure prediction perform when given only the atomic connectivity for an organic compound. A blind test was conducted on a selection of four compounds and a wide range of methodologies representing the principal computer programs currently available were used. There were 11 participants who were allowed to propose at most three structures for each compound. No program gave consistently reliable results. However, seven proposed structures were close to an experimental one and were classified as 'correct'. One compound occurred in two polymorphs, but only one form was predicted correctly among the calculated structures. The basic problem with lattice energy based methods of crystal structure prediction is that many structures are found within a few kJ mol^{-1} of the global minimum. The fine detail of the force-field methodology and parametrization influences the energy ranking within each method. Nevertheless, present methods may be useful in providing a set of structures as possible polymorphs for a given molecular structure.

1. Introduction

This paper reports on the results of a blind test of the major methodologies for *ab initio* crystal structure prediction, which was part of a collaborative workshop held at the Cambridge Crystallographic Data Centre (CCDC) in May 1999. This workshop concentrated on the prediction of crystal structures of organic compounds from the molecular structure, as expressed by the atomic connectivity. Since the question raised in the review by Gavezzotti (1994), entitled 'Are crystal structures predictable?', many workers have continued to battle with what is recognized as a very difficult problem, in the hope that the answer may move forward from a blunt 'No' to 'Sometimes!'. The crystal-structure prediction problem has similarities to the protein-folding prediction problem. Both problems involve unsolved questions regarding the choice of force field, the existence of many almost equi-energetic minima in a multi-dimensional energy space, the relative importance of thermodynamic and kinetic factors, including possible nucleation steps, and others. There are also, of course, important and obvious differences. The crystal structure prediction problem is of considerable practical importance for the industrial processing of molecular materials, such as pharmaceuticals, because the different polymorphic forms of a given compound generally have significantly different physical properties, such as solubility, bio-availability, crystal habit,

crystal size, colour *etc*. As polymorphism affects many aspects of manufacture of substances, and the possibility of phase changes during the shelf-life of a product, there are also many practical reasons for pursuing the aim of predicting all possible crystal structures of a compound. Of course, the prediction problem is greatly simplified if powder diffraction data and information on unit-cell dimensions or space groups are known (see *e.g.* Shankland *et al.*, 1997; Andreev *et al.*, 1997; Engel *et al.*, 1999). However, the scope of this workshop was limited to *ab initio* prediction of the experimental crystal structure.

There are differences of opinion as to the frequency of polymorphism under normally accessible conditions of crystallization by the variation of solvent, temperature and pressure. Although some investigators believe that polymorphism is extremely common, the Cambridge Structural Database (CSD; Allen *et al.*, 1991) contains only *ca* 3% of structures with a molecule appearing in more than one crystal form. It should be remembered, however, that most crystal structures are determined in order to elucidate the molecular structure and the experimentalist will naturally choose what appears to be the best single crystal from a sample, rather than search for variants.

Most of the methods presented here are based on a calculation of the lattice energy of a set of motionless molecules. The structure with the lowest lattice energy is taken to be the thermodynamically stable polymorph at all temperatures, in other words, free-energy contributions from the vibrational enthalpy and entropy are ignored. In any case, there are reasons for believing that the crystallization process is under kinetic rather than thermodynamic control, so the polymorph obtained under given conditions is not necessarily the thermodynamically most stable one.

In addition to the energy methods, a start has been made in making use of statistical data derived from the experimental coordinates stored in the CSD. The prediction of the most likely polymorph is based on similarity of the structure patterns observed in the CSD.

2. Approach and methodology

It should be explained at the outset that this workshop was devised as a collaborative experiment arranged privately between a group of colleagues and was not intended to be a public invitation to the crystallographic community to participate. It was considered important to cover the main methods of crystal structure prediction, but no claim is made that every published computer program has been represented. Also, in order to conduct a practical 2-day workshop discussion, the number of participants was limited. CCDC offered to host the meeting and to organize a blind test on a small number of unpublished but accurate crystal structures.

Requests were circulated to about 20 structural crystallography laboratories in order to obtain a set of three unpublished organic crystal structures. To give a reasonable chance of success within the estimated capabilities of the participating programs, the maximum size of the molecule was restricted to 30 atoms, including hydrogen, and the space group was stipulated to be one of the more commonly observed ones with one molecule per asymmetric unit, *i.e.* $Z' = 1.0$, although no defined list of allowed space groups was provided. It was also specified that submissions should belong to three categories of perceived difficulty for prediction:

(i) a small, rigid molecule with only C,H,N,O atoms allowed and less than 20 atoms,

(ii) a small, rigid molecule with some less common elements, and

(iii) a molecule with some small amount of conformational freedom.

There was some difficulty in obtaining a number of suitable unpublished structures mainly because of the 30 atom restriction, but finally a list of approximately 20 molecules was presented to an independent referee, Dr P. Raithby, University of Cambridge, who selected three of them for the test. Chemical diagrams of the three selected compounds were sent to all participants on 18 December 1998 (Fig. 1). The participants were invited to work independently in their own laboratories using their own programs to predict the most likely crystal structures. They were allowed a maximum of three predictions per compound and asked to give reasons and degree of confidence for their selection.

In addition to the three selected compounds the unpublished structure of propane was offered by one of the crystallographic laboratories and added as an option for prediction. At this stage no crystal structural details had been made available to anyone outside the originating laboratory, not even to the referee. On 12 April 1999, all submissions of

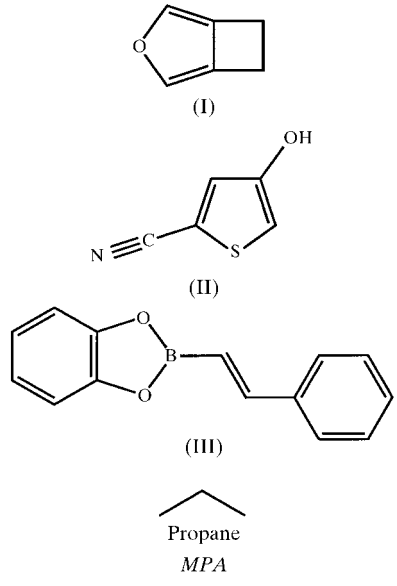

Figure 1
Molecular diagrams of the organic compounds presented to the authors in December 1998 for crystal structure prediction. (I): 3-oxabicyclo(3.2.0)hepta-1,4-diene (Boese & Garbarczyk, 1998); (II): 4-hydroxy-2-thiophenecarbonitrile (Blake *et al.*, 1999); (III): 2-(2-phenylethenyl)-1,3,2-benzodioxaborole (Clegg & Scott, 1998); propane (Boese *et al.*, 1999*a*,*b*).

Table 1
Overview of methodologies applied for crystal structure prediction for the blind test.

Name	Molecular model	Crystal structure generation and search	Fitness function	Minimization by	References
CRYSCA	Flexible	Random	Empirical energy	Derivatives	Schmidt & Englert (1996)
DMAREL	Rigid	Pregenerated *MOLPAK* starting set	Electrostatic multipole plus empirical exp-6	Derivatives	Willock *et al.* (1995)
MOLPAK	Rigid	Patterns of coordination	Empirical energy	Derivatives	Holden *et al.* (1993)
MPA	Flexible	Random or rotational grid	Empirical energy	Derivatives	Williams (1999)
MSI-PP	Flexible	Random (Monte Carlo)	Empirical energy	Derivatives	Karfunkel *et al.* (1993), Verwer & Leusen (1998)
UPACK	Flexible	Grid or random	Empirical energy	Derivatives	van Eijck & Kroon (1999)
UPACK + *ab initio*	Flexible	Pre-generated starting set	*Ab initio* derived energy	Derivatives	Mooij, van Duijneveldt *et al.* (1999)
Zip-Promet	Rigid	Building into one-dimensional, two-dimensional and three-dimensional	Empirical energy	Simplex	Gavezzotti (1999)
FlexCryst	Rigid	Building into one-dimensional, two-dimensional and three-dimensional	Statistical potentials	Grid refinement	Hofmann & Lengauer (1997)
PackStar	Rigid	Random (SA)	Statistical fit	Simulated annealing	Lommerse & Motherwell (2000)
Rancel	Rigid	Random (GA)	Statistical fit	Genetic algorithm	Motherwell (1999)

predicted structures were to be received by the referee, who then obtained the experimental coordinates which were revealed to all participants 7 days later. It had not been known up to this point that (I) had been found in two polymorphs, in *Pbca* and $P2_1/c$. Polymorph 1 (in *Pbca*) of this compound was obtained by low-temperature crystallization using a sealed capillary. At the end of the X-ray measurement, the cooling broke down and the crystal melted. By cooling the capillary again, polymorph 2 ($P2_1/c$) was formed and polymorph (I) was never seen again, neither in the original nor in any other capillary, irrespective of the applied temperatures (Boese, 1999). This is probably another case of a disappearing polymorph (Dunitz & Bernstein, 1995).

Comprehensive reviews of computerized methods for crystal structure prediction have been published (Gdanitz, 1998; Verwer & Leusen, 1998) and the reader is referred to references therein. There are four major components to any crystal structure prediction program:

(i) the input molecular model which may allow for conformational flexibility,

(ii) a method for generating and searching for crude initial crystal structures,

(iii) some sort of energy or fitness function to describe packing efficiency, and

(iv) an efficient means of reducing the packing hyperspace by evaluation of the fitness function so that a (global) minimum can be located.

The programs used in this test involved a wide range of all four components: rigid or flexible models, various approaches to create crude but feasible initial crystal structures, many kinds of force fields and several minimization methods. They are listed according to these four components in Table 1.

(i) The molecular model normally defines a rigid molecule, which can be generated by conventional *ab initio* calculations. Flexibility of the model is usually controlled by rotation of rigid fragments about selected bonds (*e.g. CRYSCA* and *MPA*). In some cases the full molecular structure is relaxed in the crystal structure at some point near the calculated minimum energy packing (*e.g. UPACK* and *MSI-PP*).

(ii) Initial crystal structures can be generated either (1) randomly (*e.g.* Schmidt & Englert, 1996), (2) using an (extensive) grid (*e.g.* van Eijck *et al.*, 1995), (3) by building up the structure systematically, for example, by using selected symmetry operators (Gavezzotti, 1991; Hofmann & Lengauer, 1997), (4) *via* Monte Carlo Simulated Annealing (Karfunkel *et al.*, 1993) or (5) utilizing a pattern for a coordination sphere (Holden *et al.*, 1993).

(iii) The energy or fitness function is based on different assumptions and subject to various approximations. Most energy calculations rely on an empirical component derived by fitting against a few selected experimental crystal stuctures and any thermodynamic data, such as sublimation enthalpies, that may be available. For simplicity, electrostatic interactions can be represented by point charges on atoms, but more elaborate models use multipole approximations. We do not report here all details of the force-field parameters, but provide literature references to the programs. Alternatively, methods which rely on direct information from the CSD can be employed, using observed contact geometries between chemical groups and/or atoms.

(iv) Minimization methods are basically of two types: (1) using the first or second derivatives of the energy potentials to find the pathways to the lattice energy minima or (2) by stochastic methods, such as simulated annealing and genetic algorithms.

3. Program details

Each participant worked independently and brief details now follow on the programs used with the author's name added (*italics*). Any opinions are those expressed by the participant in the workshop and are not necessarily shared by all. Space does not permit full details of calculations as would be the case

in a normal scientific paper, but sufficient references are provided and the reader may contact the participants directly if necessary.

3.1. *CRYSCA* (*Schmidt*)

The program *CRYSCA* is based on energy minimizations starting from random packings (Schmidt & Englert, 1996; Schmidt & Kalkhof, 1997). Intramolecular degrees of freedom can be handled also.

The CSD was used to obtain information on the preferred molecular conformations in the solid state. Compound (I) has no major degrees of freedom. Correspondingly, the energy minimizations were performed using a rigid molecule. Compound (II) was calculated twice using two search models representing the two possible in-plane orientations of the OH group (out-of-plane orientations are scarce for hydroxyl groups connected to an aromatic system). Compound (III) contains two major intramolecular degrees of freedom: rotations around the single bond B—C between the benzodioxaborole and the ethene fragment, and around the C—C bond between the ethene and the phenyl fragment. A planar conformation is possibly preferred by an isolated molecule, but a search in the CSD revealed several crystal structures of similar compounds showing remarkable deviations from planarity. Therefore, these two intramolecular degrees of freedom were included in the energy calculations from the start. The corresponding torsional potentials were calculated by AM1 and *ab initio* methods, and fitted by analytical functions. For all compounds the lattice energy was calculated by a 6-exp potential with an additional Coulombic term. The van der Waals parameters for C, H, B, N and O atoms were taken from Schmidt (1995, 1999). For S the following parameters were used: A(S,S) = 6000 Å^6 kJ mol^{-1}, B(S,S) = 1 000 000 kJ mol^{-1}, C(S,S) = 3.56 Å^{-1}. The force field works with charges calculated by the charge-iteration method (Howell *et al.*, 1977). Summation was carried out in direct space. In polar space groups, the results were confirmed by Ewald summation. The energy minimizations started from a set of several hundred random packings, *i.e.* random values within sensible ranges were assigned for lattice parameters, molecular positions, orientations and the intramolecular degrees of freedom. The energy was minimized by a special steepest-descent algorithm. The resulting crystal structures were checked for high densities, sensible packing arrangements (*e.g.* no missed hydrogen bonds) and possible higher symmetries.

With this method the crystal structures of pentamethylferrocene and the azo pigment $C_{18}H_{19}N_3O_4$ had been predicted successfully (Schmidt, 1995; Schmidt & Englert, 1995).

3.2. *DMAREL* (*Price*)

This group sought to use the best intermolecular potential, based on a distributed multipole electrostatic model, that they could estimate within the time constraint. This was challenging because all three molecules differed significantly from the range of C/H/N/O rigid polar molecules for which the approach has been developed and shown to be successful (Price & Wibley, 1997; Potter *et al.*, 1999).

The molecular model structures were obtained by *ab initio* optimization of an MP2 6-31G** wavefunction using *Gaussian*98 (Frisch *et al.*, 1998). The electrostatic energy was calculated from the atomic charges, dipoles, quadrupoles, octapoles and hexadecapoles derived by a distributed multipole analysis (Stone & Alderton, 1985) of this wavefunction using *CADPAC* (Amos & Andrews, 1995), except in the case of (III), where only the SCF charge density could be analysed. Empirically fitted potential parameters for the 6-exp repulsion–dispersion potential (with usual combining rules) were used for C and H (Williams & Cox, 1984) throughout, O (Cox *et al.*, 1981) in (I) and (II), and N (Williams & Cox, 1984) in (II), with the Hp in (II) being assumed to be the same as the fitted (N—)Hp (Coombes *et al.*, 1996). The S···S parameters were estimated by adapting the S(thiazole) parameters from the MMFF force field (Halgren, 1992). For (I) and (II), these choices were assessed by considering the reproduction of the crystal structures of FURANE10, WEYXON, TCTHPH and GAKTAN from the CSD. The repulsion parameters for all interactions involving either B or O were derived from the MP2 6-31G** charge density of 2-(2-phenylethenyl)-1,3,2-benzodioxaborole (Tsui & Price, 1999), using a novel method (Nobeli & Price, 1999) based on the assumption that the repulsion potential at any relative orientation of two molecules is proportional to the overlap of their charge densities. The corresponding R^{-6} dispersion parameters were derived using the Slater–Kirkwood formulae, using experimental atomic polarizabilities (Ketelaar, 1953) for all but B, for which a recent *ab initio* value was available (Doerksen & Thakkar, 1999), in conjunction with the effective number of electrons relationship derived by Halgren (1992). The initial close-packed starting structures were generated using *MOLPAK* (Holden *et al.*, 1993), with new estimated parameters for S and B interactions, generating 25 structures within each of the 20 packing types provided. All the starting structures in $P\bar{1}$, $P2_1/c$, $P2_12_12_1$ and $P2_1$ ($25 \times 9 = 225$) were minimized using *DMAREL* (Willock *et al.*, 1995), plus a variable proportion of the structures in $P1$, $C2/c$, $Pca2_1$ and $Pbca$, as judged by their initial lattice energy and time available.

3.3. *MOLPAK* (*Ammon*)

A crystal structure can be envisioned as a molecule in close contact with a number of surrounding molecules related to the central molecule by unit-cell symmetry. Detailed analyses of numerous triclinic, monoclinic and orthorhombic crystal structures revealed a small number of common coordination geometries or patterns; the most common coordination number was 14. *MOLPAK* was designed to construct the highest density, hypothetical packing arrangements for a molecule (search probe) based on the coordination geometries (23 at present). Initial models of (I)–(III) were built with either *Chem*3D (CambridgeSoft Corporation, 1998) or *PC Spartan Pro* (Wavefunction Inc., 1999); *ab initio* optimizations

were performed with a non-local DFT method and 6-31G* basis set with either the *Gaussian*94 (Frisch *et al.*, 1995) or *Gaussian*98W (Frisch *et al.*, 1998) programs (B3LYP/6-31G* option). The calculations of (II) utilized search probes representing the two possible OH conformations. All unique orientations of a search probe were examined (180° rotation in 10° steps about three Eulerian axes) to create 6859 (19^3) packing arrangements for each coordination geometry. The packing calculations were accelerated by the use of only the repulsive term of a standard 6–12 potential and pre-determined repulsion energy thresholds. For each geometry, the 50 highest density arrangements were refined by lattice energy minimization with either the *WMIN* (atom-centred monopole model; Busing, 1981) or *DMAREL* (Willock *et al.*, 1995). *WMIN* is not parameterized for sulfur and the force-field coefficients for an ether oxygen were utilized.

3.4. *MPA* (*Williams*)

Starting coordinates for molecules (I), (II) and (III) were obtained from molecular mechanics energy minimizations. These coordinates were then optimized using *ab initio* quantum mechanics with *Gaussian*94 (Frisch *et al.*, 1995) with a HF 6-31G** basis set. The C—H bond distances obtained were reset to 1.09 Å for H bonded to saturated C and 1.08 Å for H bonded to a doubly bonded C for uniformity. The structure of propane, including hydrogen positions, was taken from Williams (1994), which was similarly optimized with a 6–31++G** basis set. The hydrogen positions in this molecule were not reset.

The MEPs of optimized molecules (I), (II) and (III) were calculated with a HF/6-31G** basis set on a fourfold tessellation icosahedral geodesic grid (Spackman, 1996) around each molecule. The successive layers were placed at 1.6, 1.8, 2.0, 2.2 and 2.4 times the van der Waals radii of the atoms. The MEPs were fitted with the program *PDM*97 (Williams, 1997). For propane, additional methylene bisector sites (Williams, 1994) were used to obtain a better representation of the MEP. For the Biosym force field (Biosym Technologies, 1969), methylene bisector sites were added at a distance of 0.81 Å for methylene and 0.41 Å for methyl groups. For the W99 force field both bisector distances were set at 0.6 Å. C—H distances were foreshortened by 0.1 Å.

Non-bonded intermolecular energy in the crystals was modelled by pairwise sums of atom–atom terms of the (12-6-1) or (exp-6-1) type. Molecules were assumed to be rigid. The cut-off distance for lattice summation was 8 Å and accelerated convergence (Williams, 1971) was used to improve the accuracy of lattice sums. Lattice energy minimizations were made in the 18 most popular space groups, which comprise about 93% of all organic crystals (Bauer & Kassner, 1992). Initially, the molecule was placed in a large 30 Å orthogonal cell in a Lattman grid rotational orientation (Williams, 1973) and the molecular centre at 1/4,1/4,1/4 fractional cell coordinates. During a first stage only cell edge lengths were varied, followed by a second stage in which all symmetry-allowed parameters were varied.

3.5. *MSI-PP* (*Verwer & Leusen*)

For molecules (I), (II) and (III), the initial structures were taken from geometry optimizations at the HF 6-31G* level using *Gaussian*94 (Frisch *et al.*, 1995). The results of these calculations were also used to obtain atomic charges, by fitting to the electrostatic potential. For molecules (I) and (III), the CHelpG (Breneman & Wiberg, 1990) method was used; point charges optimally reproducing the ESPs of both conformations of molecule (II) were obtained with the RESP method (Bayly *et al.*, 1993).

The model used for propane consisted of the molecular structure, optimized within the Dreiding 2.21 force field (Mayo *et al.*, 1990), employing no atomic charges.

Crystal structure predictions were carried out using the Polymorph Predictor technology (Leusen, 1996; Leusen *et al.*, 1999) as implemented in the *Cerius*^{2} molecular modelling package (Molecular Simulations Inc., 1999). The ten most common space groups ($P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $P2_1$, $C2/c$, $Pbca$, $Pnma$, $Pna2_1$, $Pbcn$, $P1$) were considered for molecules (I), (II) and (III), starting from each low-energy conformer of molecule (II) in a separate prediction run. In the case of propane, the 17 most common space groups were investigated with $Z' = 1$, as well as $P1$ with all integer values of Z' up to $Z' = 8$.

All predictions were performed using the Dreiding 2.21 force field using Ewald summation to ensure convergence of long-range intermolecular interactions. Molecules were kept rigid during the Monte Carlo Simulated Annealing search procedure, but were fully flexible in the final lattice energy minimization step.

3.6. *Promet* (*Gavezzotti, Schweizer & Dunitz*)

Promet (Gavezzotti, 1991, 1999) starts with a systematic generation of dimers or strings of rigid molecules over or along the most common symmetry operators in organic crystal chemistry (the inversion centre, the screw axis or the glide plane). The most cohesive among these oligomers are then translated in space as necessary to generate crystal structures in the most common space groups for organic crystals. A sometimes unfavourable consequence of this *aufbau* style is that it cannot reach crystal structures in which no particularly stable substructures are present. Approximate crystal structures thus generated are then refined by lattice energy minimization (*Minopec*, a new lattice energy minimizer: Gavezzotti, 1999) with respect to cell parameters and rigid-body displacements, under the space-group symmetry constraints.

Promet library potentials have the often criticized 6-exp chargeless form (Filippini & Gavezzotti, 1993; Gavezzotti & Filippini, 1994), but user-supplied potentials supplemented with atomic charge parameters were also used. The calculation of electrostatic energies was performed only in the point-charge approximation without forced convergence.

For (I) the molecular dimensions were taken from results of an *ab initio* calculation (MACSPARTAN, 6-31G* basis set assuming C_{2v} symmetry). Library potentials were supple-

mented in most of the calculations by R^{-1} terms over atomic point charges resulting from a fit of the electrostatic potential. For (III) the library potentials without charges were used. For the crystal structure generation the search was restricted to space groups $P\bar{1}$, $P2_1$, $P2_1/c$, $P2_12_12_1$ and $Pbca$, all with $Z' = 1$. The enthalpy of sublimation of (I) was taken to be that of isomeric phenol (65–69 kJ mol^{-1}). This can provide a useful check on the calibration of the electrostatic terms.

Compound (II) was not considered, since it is an exercise in the prediction of molecular conformation also (the position of the hydroxyl hydrogen).

3.7. Standard *UPACK* (*van Eijck*)

Propane was the only molecule that could be treated routinely with the standard *UPACK* force fields (van Eijck & Kroon, 1999). Compound (II) was not attempted owing to the lack of adequate parameterization of the CN group.

For (I) and (III) the geometry of the free molecule was optimized at the 6-31G* SCF level, and charges were obtained from ESP fittings using *MOLDEN* (Schaftenaar & Noordik, 1999).

Force-field parameters for bond distances and angles were adjusted to reproduce the *ab initio* values and reasonable guesses were taken for the corresponding force constants. All torsional angles involving Csp^2 atoms were restrained to planarity with the aid of a harmonic potential. Intermolecular Lennard–Jones parameters were taken from the all-atom OPLS force field (Jorgensen *et al.*, 1996); for propane this force field was used without adaptations. In (III) the non-bonded parameters for the B atom were transferred from the Csp^2 atom.

Crystal structures were generated within fixed space groups, using either a grid search or a random search technique. An estimated density was applied as a constraint. For each compound the space groups $P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, Cc, $Pca2_1$, $C2$, $P1$, $Pbcn$ and Pc were investigated. The final set consisted of all structures within an energy window of 6 kJ mol^{-1}.

3.8. *UPACK* plus *ab initio* based energy minimization (*Mooij*)

The crystal structures generated by van Eijck (see §3.7) were investigated by energy minimizations with an *ab initio* intermolecular potential. This potential was derived from high-level *ab initio* calculations at the SCF + MP2 level. The functional form involves atomic multipole moments, atomic dipole polarizabilities, a damped dispersion term and an exponential repulsion term with some anisotropic features. Propane was the only compound that could be treated with our recently developed parameterization for alkanes, ethers and alcohols (Mooij, van Duijneveldt *et al.*, 1999). For the other molecules we determined parameters for the Csp^2—H fragment from *ab initio* calculations on ethene dimers and methanol–ethene dimers, following the methods outlined by Mooij, van Duijneveldt *et al.* (1999). Details of the parameterization will be made available elsewhere (Mooij, 2000).

For each compound, atomic multipole moments were determined by fitting to the electrostatic potential, as implemented in the *MOLDEN* program (Schaftenaar & Noordik, 1999); for propane a correlated wavefunction was used [MP2/EZ(2df)/DZ(2p)], whereas for (I) and (III) the Hartree–Fock level was used [SCF/DZ(2d^0)]. In (III) the non-bonded parameters for B were taken over from Csp^2. The molecules were all treated as fully flexible during energy minimization. For intramolecular interactions the MM3 force field as implemented in the *TINKER* program (Ponder, 1998) was used, including its van der Waals and electrostatic parameters. All crystal energy minimizations were performed with a recently developed program (Mooij, van Eijck & Kroon, 1999).

3.9. *FlexCryst* (*Hofmann*)

The initial structures of (I), (II) and (III) were minimized with *MINDO* assuming a planar geometry. For (II) two minima were found differing in the hydroxyl conformation. For each of these molecules potential crystal structures were calculated in the space groups $P1$, $P\bar{1}$, $P2_1$, $P2_1/c$ and $P2_12_12_1$. In all cases it was assumed that the crystal structure contained only one molecule in the asymmetric unit. The crystal structures of each molecule were ranked according to the FlexCryst scoring function and the three crystal structures lowest in energy were selected for each compound.

The program builds up crystal structures systematically (Hofmann & Lengauer, 1997, 1998), using methods analogous to those described for *Promet*. The algorithm works in a discrete space (mesh constant 1 Å) and during the generation the four vectors (origin, **a**, **b** and **c**) defining a crystal structure are forced to be grid points. The choice of the mesh constant balances accuracy *versus* calculation time, a finer grid allows a larger number of crystal structures. All generated structures are compared with the CSD. It has been proved that there exists a connection between observed frequencies of interatom distances or distances between functional groups and the energy *via* the Boltzmann equation. These frequencies of interatom distances are used to create a scoring function. Finally, the structures are ranked according to their score without further minimization.

3.10. *PackStar* (*Lommerse*)

In *PackStar*, CSD information is represented in scatterplots in the IsoStar library (Bruno *et al.*, 1997), which show preferred orientations of pairs of interacting functional groups. From these scatterplots, a crystal field environment can be built around a given molecule, representing the interaction of an atom or small chemical group with the molecule. This field can be expressed in scalar propensity maps for each interaction (Verdonk *et al.*, 1999) or as a propensity vector map which preserves information about the directionality of an interaction (Lommerse & Motherwell, 2000).

For a given crystal structure a total cost is calculated representing a fit with the high propensity spaces for each individual interaction. It is assumed that the unknown

Table 2
Experimental crystal structure data and crystal structure predictions for (I).

P: experimental polymorph number or structure prediction number. The RMSD values were calculated and all comparisons made against the *Pbca* structure (polymorph 1) only. For polymorph 2 none of the solutions were correct. The rows in bold in the table can be judged as correct predictions. For an example, see Fig. 2(*b*). Full coordinate sets for all results given in this table have been deposited.

Method	*P*	Space group	*a* (Å)	*b* (Å)	*c* (Å)	α (°)	β (°)	χ (°)	Density (g cm^{-3})	RMSD Conf.	RMSD Pack.
Experimental	1	*Pbca*	5.309	12.648	14.544	90.00	90.00	90.00	1.288		
Experimental	2	$P2_1/c$	4.954	9.845	9.679	90.00	90.57	90.00	1.333		
CRYSCA	1	$P2_12_12_1$	7.202	7.909	7.979	90.00	90.00	90.00	1.384	0.0972	2.0
CRYSCA	2	$P2_12_12_1$	7.346	7.562	8.378	90.00	90.00	90.00	1.352	0.0972	2.0
CRYSCA	3	$P2_1$	5.100	7.926	5.748	90.00	94.98	90.00	1.359	0.0892	2.0
DMAREL	1	$P2_1$	6.085	7.811	5.439	90.00	97.39	90.00	1.227	0.0123	2.0
DMAREL	2	$P2_1/c$	8.707	5.420	12.467	90.00	120.83	90.00	1.245	0.0120	2.0
DMAREL	3	$P\bar{1}$	6.324	5.432	8.061	74.40	96.99	105.12	1.224	0.0123	2.0
MOLPAK	1	$P2_1/c$	8.458	5.368	12.857	90.00	59.33	90.00	1.253	0.0120	2.0
MOLPAK	2	$P2_1$	6.682	5.395	7.100	90.00	90.00	90.00	1.229	0.0120	2.0
MOLPAK	3	$P2_1/n$	5.456	7.126	13.059	90.00	90.00	90.00	1.239	0.0120	2.0
MPA	**1**	***Pbca***	**5.125**	**12.503**	**14.104**	**90.00**	**90.00**	**90.00**	**1.392**	**0.0129**	**0.277**
MPA	2	$P2_1$	5.114	6.864	6.638	90.00	81.67	90.00	1.364	0.0130	2.0
MSI-PP	**1**	***Pbca***	**5.372**	**12.570**	**15.131**	**90.00**	**90.00**	**90.00**	**1.231**	**0.0365**	**0.231**
MSI-PP	2	$P2_12_12_1$	11.556	8.112	5.412	90.00	90.00	90.00	1.240	0.0367	2.0
MSI-PP	3	$Pna2_1$	7.817	5.489	11.986	90.00	90.00	90.00	1.223	0.0375	2.0
UPACK	1	$C2/c$	25.029	7.341	5.304	90.00	76.23	90.00	1.329	0.0131	2.0
UPACK	2	$P2_1/c$	19.478	5.321	7.430	90.00	37.82	90.00	1.332	0.0168	2.0
UPACK	**3**	***Pbca***	**5.276**	**12.468**	**14.390**	**90.00**	**90.00**	**90.00**	**1.329**	**0.0134**	**0.525**
UPACK/ ab initio	1	$P2_1$	5.320	8.268	5.252	90.00	91.29	90.00	1.362	0.0677	2.0
UPACK/ ab initio	2	$P2_1/c$	5.868	5.250	17.043	90.00	64.03	90.00	1.333	0.0667	2.0
UPACK/ ab initio	3	$Pca2_1$	10.416	5.226	8.536	90.00	90.00	90.00	1.354	0.0686	2.0
Zip-Promet	**1**	***Pbca***	**5.182**	**12.554**	**14.336**	**90.00**	**90.00**	**90.00**	**1.349**	**0.0127**	**0.204**
Zip-Promet	2	$P\bar{1}$	5.344	6.846	7.399	84.86	71.96	66.96	1.329	0.0131	2.0
Zip-Promet	3	$P2_1/c$	8.764	5.268	12.138	90.00	123.69	90.00	1.349	0.0124	2.0
FlexCryst	1	$P2_12_12_1$	4.211	7.460	18.571	90.00	90.00	90.00	1.078	0.0378	2.0
FlexCryst	2	$P2_1/c$	12.728	4.211	15.231	90.00	45.00	90.00	1.090	0.0377	2.0
FlexCryst	3	$P2_1/c$	10.817	4.150	14.414	90.00	123.69	90.00	1.168	0.0377	2.0
PackStar	1	$P2_12_12_1$	7.100	7.130	10.180	90.00	90.00	90.00	1.221	0.0129	2.0
PackStar	2	$Pna2_1$	7.870	12.130	5.630	90.00	90.00	90.00	1.170	0.0130	2.0
PackStar	3	Cc	6.471	12.327	7.432	90.00	66.36	90.00	1.158	0.0129	2.0
Rancel	1	$P\bar{1}$	5.404	7.562	6.786	83.70	81.24	106.55	1.213	0.0130	2.0
Rancel	2	*Pbca*	10.305	13.335	7.585	90.00	90.00	90.00	1.207	0.0129	2.0
Rancel	3	$P2_12_12_1$	12.487	9.705	4.123	90.00	90.00	90.00	1.259	0.0130	2.0

experimental structure has an optimal fit with the available CSD data, which should lead to a low cost. Using simulated annealing the lowest cost is searched for in a given set of space groups ($P1$, $P\bar{1}$, $P2_1$, $P2_1/c$, $P2_12_12_1$, $C2$, Cc, $C2/c$, $Pbca$, $Pnma$, $Pna2_1$). A more detailed description of the method will become available in due course (Lommerse & Motherwell, 2000).

The (fixed) geometry of the molecule was calculated by an *ab initio* calculation at the 6-31G** level using *CADPAC* (Amos & Andrews, 1995). A grid was used to store values of the IsoStar-derived density maps around the molecule. The grid size was 0.5–0.8 Å in order to obtain statistically meaningful densities. For (I), the molecule was described by a furan and two methylene groups. Scatterplots of these groups with methylene, aromatic C—H, aromatic C and Osp^3 were retrieved from the IsoStar library and converted into four propensity vector maps representing all interactions of furan with itself. For (II), the molecule was segmented into thiophene, aromatic cyano and aromatic hydroxyl, interacting with cyano, hydroxyl, aromatic C—H, aromatic C and Ssp^3. Compound (III) was segmented into phenyl, 1,2-dioxophenyl, (*E*)-ethene and boron interacting with aromatic C—H, aromatic C, Osp^3 and B. Finally, propane was treated using methyl and methylene interacting with themselves.

3.11. *Rancel* (*Motherwell*)

Crystal structures were calculated based on the same rigid molecular models as in *PackStar* in a selected range of space groups: $P\bar{1}$, $P2_1$, $P2_1/c$, $C2/c$, $P2_12_12_1$, $Pca2_1$, $Pna2_1$ and $Pbca$. A genetic algorithm is used on a fixed population to search for a minimum in a fitness function calculated from the intermolecular atomic distances (Motherwell, 1999). The fitness function is expressed in terms of the observed frequency distributions of the intermolecular distances in the CSD. The function Q is the sum of squares of differences between smoothed frequency curves sampled at 0.2 Å intervals. These curves were derived from CSD moleclues that were similar in chemical constitution and molecular size to the query molecule. Visual inspection of search results was used to select the CSD molecules. Penalty functions are used to add a positive quantity to the function when distances fall below the closest contacts observed in the CSD sample or when certain expected distances are not found.

For each query structure the GA search procedure was performed ten times for each of the eight space groups, using a set of frequency curves derived from about ten CSD molecules. Structures with low Q values were inspected visually and empirical packing energies calculated as a check for close packing, using the table from Gavezzotti (1994). Checks for voids of radius 2 Å in the cell space also confirmed that all low Q structures were indeed closely packed.

4. Overview of the results

Space group, cell dimensions, crystal structure density and RMSD (root-mean-square deviation) values for the molecular conformation, as well as the crystal structure packing, are

Table 3
Experimental crystal structure data and crystal structure predictions for (II).

See headnote to Table 2.

Method	*P*	Space group	*a* (Å)	*b* (Å)	*c* (Å)	α (°)	β (°)	χ (°)	Density (g cm^{-3})	RMSD Conf.	RMSD Pack.
Experimental		$P2_1/n$	7.516	8.332	9.059	90.00	100.19	90.00	1.499		
CRYSCA	1	$P2_12_12_1$	6.957	8.457	9.078	90.00	90.00	90.00	1.567	0.0976	1.425
CRYSCA	2	$P2_1/c$	9.894	5.416	9.728	90.00	91.47	90.00	1.607	0.0975	2.0
CRYSCA	3	$P2_12_12_1$	5.493	6.197	15.326	90.00	90.00	90.00	1.604	0.4535	2.0
DMAREL	1	$P2_12_12_1$	9.455	7.009	8.535	90.00	90.00	90.00	1.480	0.0662	1.333
DMAREL	2	$Pna2_1$	17.224	3.726	8.552	90.00	90.00	90.00	1.525	0.0663	2.0
DMAREL	3	$P2_1/c$	5.105	8.075	12.945	90.00	90.72	90.00	1.569	0.0662	2.0
MOLPAK	1	$P2_1/n$	3.835	11.518	12.082	90.00	73.41	90.00	1.637	0.4369	2.0
MOLPAK	2	$P2_1/c$	12.568	3.788	11.272	90.00	107.11	90.00	1.632	0.4368	2.0
MOLPAK	3	$P\bar{1}$	5.650	3.741	12.583	96.91	76.77	92.87	1.629	0.4369	2.0
MPA	1	$P2_1/c$	7.780	9.382	8.186	90.00	83.22	90.00	1.411	0.0382	2.0
MPA	2	$P2_12_12_1$	9.187	8.356	7.487	90.00	90.00	90.00	1.457	0.0382	1.412
MSI-PP	1	$P2_1/a$	10.235	8.228	6.951	90.00	114.13	90.00	1.567	0.0968	2.0
MSI-PP	**2**	***P*2$_1$/*n***	**7.243**	**8.299**	**9.210**	**90.00**	**104.53**	**90.00**	**1.562**	**0.0991**	**0.427**
MSI-PP	3	$P2_1$	10.529	7.314	3.902	90.00	65.91	90.00	1.526	0.0978	2.0
FlexCryst	1	$P\bar{1}$	9.539	8.367	7.071	63.93	72.75	43.39	1.203	0.0826	2.0
FlexCryst	2	$P\bar{1}$	6.164	7.141	10.630	128.25	121.26	82.17	1.386	0.4422	2.0
FlexCryst	3	$P2_1/c$	6.000	3.879	24.515	90.00	90.00	90.00	1.467	0.4423	2.0
PackStar	1	$P2_1/n$	3.690	15.790	13.510	90.00	131.30	90.00	1.416	0.0584	2.0
PackStar	2	$P2_12_12_1$	4.400	8.650	15.750	90.00	90.00	90.00	1.397	0.0584	2.0
PackStar	3	$Pna2_1$	8.670	5.490	12.670	90.00	90.00	90.00	1.388	0.4330	2.0
Rancel	1	$C2/c$	12.747	8.253	10.958	90.00	97.08	90.00	1.464	0.0646	2.0
Rancel	2	$P2_1$	7.555	10.057	3.819	90.00	96.09	90.00	1.451	0.0646	2.0
Rancel	3	$P2_1/c$	3.880	14.587	10.331	90.00	91.89	90.00	1.433	0.0645	2.0

Table 4
Experimental crystal structure data and crystal structure predictions for (III).

Method	*P*	Space group	*a* (Å)	*b* (Å)	*c* (Å)	α (°)	β (°)	χ (°)	Density (g cm^{-3})	RMSD Conf.	RMSD Pack.
Experimental	1	$P2_1/c$	6.835	7.634	21.422	90.00	96.45	90.00	1.336		
CRYSCA	1	$P\bar{1}$	6.288	8.802	9.999	84.21	74.82	87.78	1.397	0.0953	2.0
CRYSCA	2	$P\bar{1}$	6.294	8.969	9.945	100.77	103.94	94.91	1.400	0.0947	2.0
CRYSCA	3	$P2_1/c$	6.491	25.517	6.964	90.00	111.93	90.00	1.387	0.0963	2.0
DMAREL	1	$P2_1/c$	11.434	5.461	19.476	90.00	105.60	90.00	1.267	0.1065	2.0
DMAREL	2	$P2_1/c$	9.570	5.653	28.352	90.00	49.34	90.00	1.276	0.1068	2.0
DMAREL	3	$P2_1/c$	12.438	5.050	19.280	90.00	98.07	90.00	1.238	0.2227	2.0
MOLPAK	1	$P2_1/c$	17.851	3.847	29.759	90.00	146.49	90.00	1.315	0.1057	2.0
MOLPAK	2	$P2_1/c$	9.667	5.717	28.345	90.00	133.45	90.00	1.305	0.1056	2.0
MOLPAK	3	$P\bar{1}$	15.235	3.853	10.559	81.39	90.03	111.34	1.303	0.1056	2.0
MPA	1	$P2_1/c$	11.769	4.407	21.354	90.00	85.49	90.00	1.344	0.1155	2.0
MPA	2	$Pbca$	7.193	27.411	11.563	90.00	90.00	90.00	1.302	0.1155	2.0
MSI-PP	1	$C2/c$	12.205	7.540	27.606	90.00	62.68	90.00	1.315	0.0563	2.0
MSI-PP	2	$P\bar{1}$	6.942	12.681	7.191	87.37	64.02	86.18	1.307	0.1283	2.0
MSI-PP	3	$P2_1/n$	28.708	10.402	3.948	90.00	72.93	90.00	1.317	0.1923	2.0
UPACK	**1**	***P*2$_1$/*c***	**6.763**	**7.758**	**20.940**	**90.00**	**98.32**	**90.00**	**1.365**	**0.0476**	**0.214**
UPACK	2	$P2_1/c$	9.915	5.466	27.812	90.00	46.11	90.00	1.367	0.1326	2.0
UPACK	3	$C2/c$	23.603	6.571	16.355	90.00	61.07	90.00	1.337	0.1830	2.0
UPACK/ab initio	1	$P2_1/c$	5.824	8.238	23.449	90.00	91.91	90.00	1.320	0.1233	0.955
UPACK/ab initio	2	$Pca2_1$	8.354	5.426	25.394	90.00	90.00	90.00	1.289	0.1393	2.0
UPACK/ab initio	3	$P2_1/c$	6.809	17.327	9.667	90.00	90.14	90.00	1.301	0.1474	2.0
Zip-Promet	1	$P2_12_12_1$	10.139	3.509	30.605	90.00	90.00	90.00	1.363	0.0845	2.0
Zip-Promet	2	$P2_1$	10.154	3.504	15.315	90.00	87.33	90.00	1.364	0.0990	2.0
FlexCryst	1	$P2_1$	15.556	4.000	10.000	90.00	85.13	90.00	1.197	0.0644	2.0
FlexCryst	2	$P2_1/c$	16.000	4.000	20.000	90.00	90.00	90.00	1.160	0.0644	2.0
FlexCryst	3	$P2_1/c$	13.038	4.000	24.413	90.00	94.40	90.00	1.169	0.0644	2.0
PackStar	1	Cc	11.570	6.000	19.990	90.00	68.80	90.00	1.147	0.1157	2.0
PackStar	2	$P2_1/n$	3.860	14.500	25.940	90.00	64.50	90.00	1.133	0.1157	2.0
PackStar	3	Cc	6.990	10.030	25.170	90.00	46.70	90.00	1.156	0.1157	2.0
Rancel	1	$P2_1/c$	7.913	12.970	11.750	90.00	87.73	90.00	1.232	0.1156	2.0
Rancel	2	$P\bar{1}$	9.440	5.149	13.349	108.08	76.48	82.96	1.266	0.1156	2.0
Rancel	3	$P2_12_12_1$	13.876	12.030	7.196	90.00	90.00	90.00	1.236	0.1156	2.0

listed for all 36 proposed crystal structures in Tables 2–5, one table for each molecular structure.

4.1. RMSD values for the molecular conformations

RMSD values for the molecular conformations were calculated with the coordinates of the heavy atoms only, except for (II), where it was necessary to include the hydroxyl hydrogen.

4.2. Comparison of crystal packings

Although all calculated structures were compared visually with the experimental structure, a method was developed that automatically quantifies the similarity of crystal packings. The method is based on the exploration of the coordination sphere around a central molecule in a crystal structure (Gavezzotti & Filippini, 1995). For a crystal with $Z' = 1$, every molecule has an identical environment, *i.e.* the same coordination sphere. For our comparison, the six closest molecules around a central molecule in the experimental structure are selected, according to the distances between the geometrical centres (arithmetical centroids) of the molecules. Also, the relative orientation of each of these molecules is described by two reference vectors.

For each predicted crystal structure, the 12 closest molecules around a central molecule are first selected and all 665 280 permutations of six molecules out of the 12 molecules are then tested. Structures are rejected if any of the molecular orientations differ with respect to the experimental structure by more than 30°. For the remaining structures, atomic coordinates were superimposed using a

least-squares fit. The vectors describing the molecular orientations in each structure are shown in Fig. 2(*a*). The lowest RMSD values (RMSD pack) are given in Tables 2–5 and an example of the fit for (I) is shown in Fig. 2(*b*).

Since seven correct predictions have been made, it is interesting to calculate more precise deviations of these structures from the experimental data in terms of cell lengths, cell angles *etc.* from the experimental structure. An overview of these details is given in Table 6. All successful predictions are visualized in their unit cells in Figs. 3(*a*)–(*d*).

5. Discussion of results from each program by the participants

The author of each section is given in *italics*. The predicted structures submitted by each participant have been deposited as supplementary material.[1]

5.1. *CRYSCA* (*Schmidt*)

CRYSCA predicted all crystal structures as possible polymorphs, but the experimental structures were not predicted as the most probable ones.

5.1.1. Compound (I). More than 100 crystal structures were found within 5 kJ mol^{-1} above the minimum energy. In several cases, the resulting packings adopted a higher symmetry during the minimization, even tetragonal symmetry ($P4_12_12$) occurred. Entropic effects were neglected as usual. Since entropy differences between organic polymorphs are typically of the order of 0–5 kJ mol^{-1}, all calculations have an intrinsic uncertainty of $\sim$5 kJ mol^{-1} in the free energy. Therefore, it was virtually impossible to guess which of the calculated crystal structures would be the experimental one. The guessing failed! Although the assumed input molecular geometry was incorrect (four-membered ring twisted by 14° rather than being planar), both polymorphs were found in the calculations. The difference between the calculated and experimental crystal structures of polymorph 1 (polymorph 2) were Δa = −4.6% (0.3%), Δb = −0.6% (3.2%), Δc = 1.4% (−2.9%), $\Delta\beta$ = 0 (−1.0%) for the lattice parameters, ΔX = 0.39 Å (0.26 Å) for the molecular position and $\Delta\tau$ = 4.8° (5.1°) for the molecular orientation. Polymorph 1 corresponded to the third-best structure in *Pbca*, having an energy difference of 4.54 kJ mol^{-1} from the global minimum found in $P2_1/c$. Polymorph 2 was found ranked 87 in $P2_1/c$ with an energy difference of 5.61 kJ mol^{-1}. The results would be improved by using the correct molecular geometry. The remaining energy difference can be explained either by entropic effects or by inadequacies in the description of the intermolecular interactions, especially in the charges.

5.1.2. Compound (II). Crystal structure prediction was attempted for (II), although the applied force field and charges were not reliable for this molecule. The structure assumed to be the most probable has very similar structural features to the experimental structure, but with an incorrect space group ($P2_12_12_1$ rather than $P2_1/n$). The correct structure was 2.9 kJ mol^{-1} lower in energy (rank 24, Δa = −3.8%, Δb = 1.0%, Δc = −1.2%, $\Delta\beta$ = 3.8%, ΔX = 0.55 Å and $\Delta\tau$ = 4.9°).

5.1.3. Compound (III). The calculations revealed five energy minima within an energy window of only 1 kJ mol^{-1} and many more minima within 5 kJ mol^{-1}. Most low-energy minima showed a deviation from planarity of $\sim$1–10°. In the

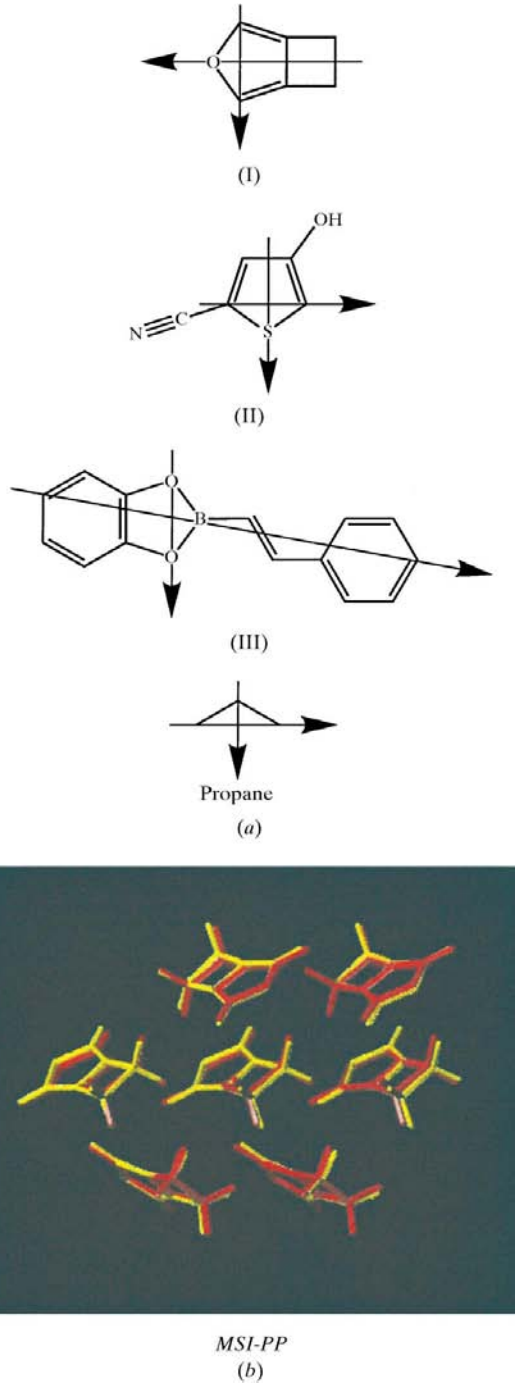

Figure 2
(*a*) Vectors describing molecular orientations; (*b*) Superimposition of (I) for the experimental structure (yellow) with the predicted structure by *Zip-Promet* (red).

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: BK0070). Services for accessing these data are described at the back of the journal.

experimental structure the ethene fragment is almost coplanar with the benzodioxoborole fragment ($\varphi = 1^\circ$), whereas the phenyl ring is rotated by $\varphi' = 7^\circ$. The experimental structure corresponded to the packing ranked 32nd with an energy of 4.36 kJ mol^{-1} above the global minimum energy. The geometry of this packing was quite close to the experimental structure ($\Delta a = -2.0\%$, $\Delta b = 1.5\%$, $\Delta c = -2.4\%$, $\Delta\beta = 1.5\%$, $\Delta X = 0.12$ Å and $\Delta\tau = 4.3^\circ$). Even the intramolecular torsion angles were similar ($\varphi = 3^\circ$, $\varphi' = 7^\circ$).

If additional data had been available, *e.g.* a powder diagram (even a non-indexable one), *CRYSCA*, and probably all other methods as well, would have found all crystal structures correctly. The accuracy of almost all calculated packings is high enough for a subsequent Rietveld refinement.

advantage in discriminating between different favourable interaction motifs and predicting the cell dimensions. The two possible orientations of the hydroxyl group gave gas-phase molecular structures differing by 2 kJ mol^{-1}, so both were searched. Since the lower energy conformer had a lower global minimum in the lattice energy by 1.8 kJ mol^{-1}, the observed conformer was favoured by both the inter- and intramolecular energy terms, although not strongly.

The nine lowest energy structures found in the search, with lattice energies between −83.5 and −82.3 kJ mol^{-1}, were examined. They all contain chains of molecules held together by N···H—O hydrogen bonds. In most cases this chain is planar (or nearly so) with the S atoms on the same side of the chain. The chains pack to give close (∼3.5 Å)

5.2. *DMAREL* (*Price*)

5.2.1. Compound (I). It was clear at the outset that the electrostatic contribution would not be important in (I) (we estimate <17% of the lattice energy) and our potential did not give a satisfactory reproduction of the structure of furan (RMSD errors of 5.9% in the cell lengths). Post-analysis showed that the reproduction of the $P2_1/c$ polymorph was sensitive to the position of the H-atom interaction sites (RMSD errors of 5.9% using the experimental structure, 6.6% for the *ab initio* optimized structure), although the *Pbca* form was satisfactorily reproduced (0.4 and 1.9%).

The $P2_1/c$ form was found within 3 kJ mol^{-1} (approximately 16th lowest) of the global minimum. The *Pbca* form was within 1 kJ mol^{-1} of the global minimum and would have been our third guess, if it had been generated by the search procedure. This clearly shows the plurality of structures formed by simple-shaped non-polar molecules, which makes the energy rankings very sensitive to small details of the intermolecular potential, such as the location of the hydrogen interaction sites.

5.2.2. Compound (II). Compound (II) was clearly the one where an accurate electrostatic model should provide an

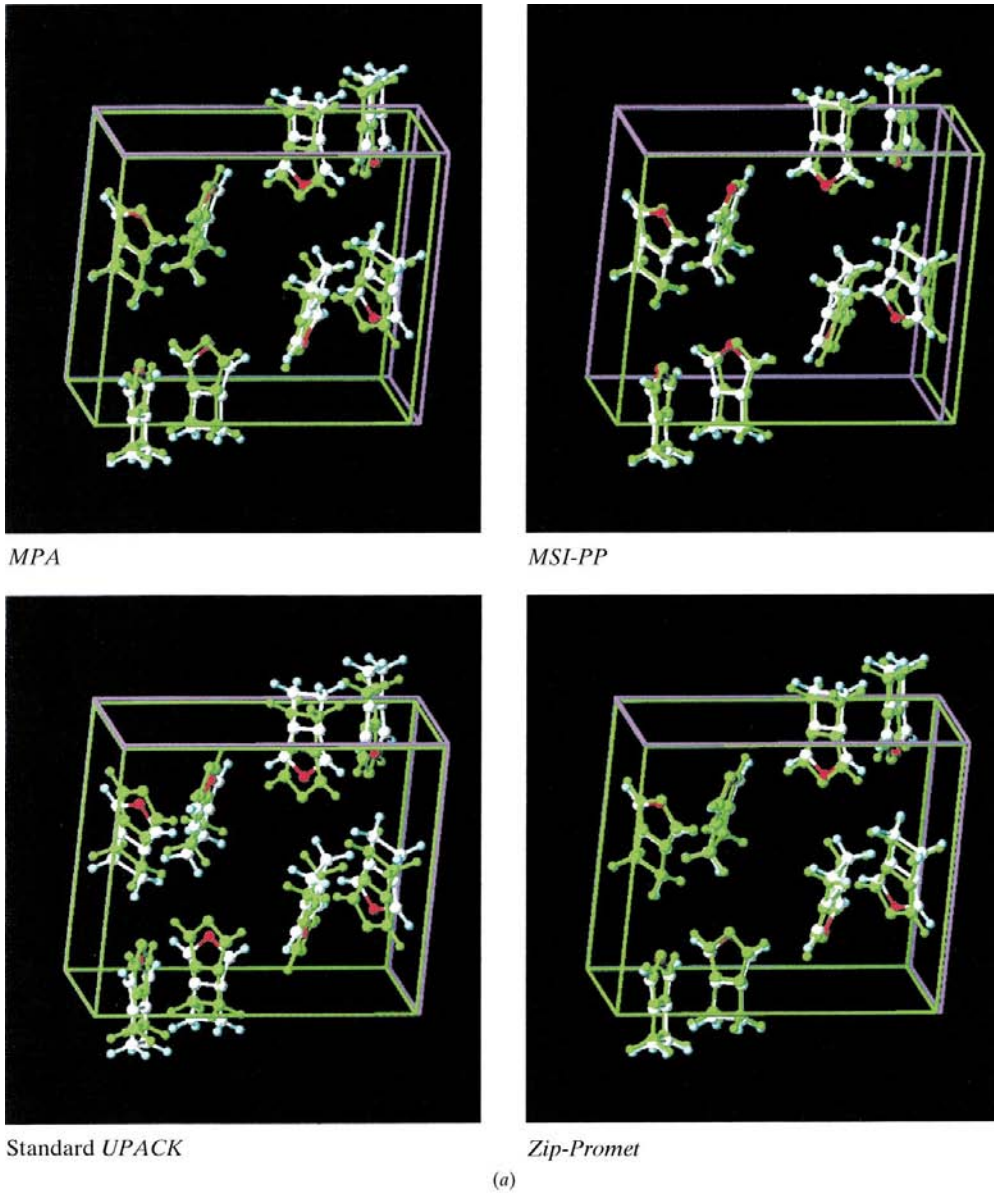

Figure 3
Comparison of the experimental unit cell of (*a*) (I), (*b*) (II), (*c*) (III) and (*d*) propane (coloured by atom type) with the successfully predicted unit cells (in green), including symmetry-related molecules (also in green). The origins of the cells have been superimposed.

S···N contacts between the chains. However, there are a variety of packings of this chain motif, differing little in energy, with the chains having the closest S···N contacts antiparallel in $P2_12_12_1$ and $P2_1/c$, and parallel in $Pna2_1$, $P2_1$, $P\bar{1}$, $C2/c$ and $P1$. Thus, the global minimum ($P2_12_12_1$) with antiparallel chains, the third lowest energy $Pna2_1$ (-83.1 kJ mol^{-1}) with parallel chains and the non-planar chain ($P2_1/c$, -82.9 kJ mol^{-1}) were submitted as guesses. The second lowest energy structure (-83.3 kJ mol^{-1}) with very similar antiparallel chains was the correct space group ($P2_1/c$), but contained some qualitative errors in the chain packings. Energy minimization of the experimental crystal structure (using the *ab initio* molecular structure) gave an energy of -1.5 kJ mol^{-1} lower than the global minimum found and an RMSD error in the cell lengths of 1.8%. Thus, the main problem was that the search procedure did not find the experimental packing of the chains, amongst the many possibilities that differed little in energy.

5.2.3. Compound (III). The *ab initio* calculations revealed that there was an extremely flat torsional potential for the rotation of the phenyl ring and so it was clear that this angle would be determined by the packing forces, and that an accurate structure prediction would be impossible with our rigid molecule methodology. This torsion angle was 19° for the lowest energy conformer found, but the energy loss on rotation to a planar molecule was only 0.9 kJ mol^{-1}. The crystal structure search using a rigid molecule with a 19° torsion angle showed several low-energy structures in $P2_1/c$, with a global minimum of -109 kJ mol^{-1}. Searching this space group with the planar molecule produced even lower energy minima (by 3.7 kJ mol^{-1}). The experimental crystal structure was estimated to have a lattice energy of -114.1 kJ mol^{-1} and thus our calculated lattice energies are consistent with the observed structure being the energetically optimum structure. Using our new model potential, we were able to reproduce the observed experimental structure with errors of $\Delta a = -0.018$, $\Delta b = -0.316$, $\Delta c = 0.040$ Å, $\Delta\beta = 2°$, which was an encouraging sign for novel potential development methodology.

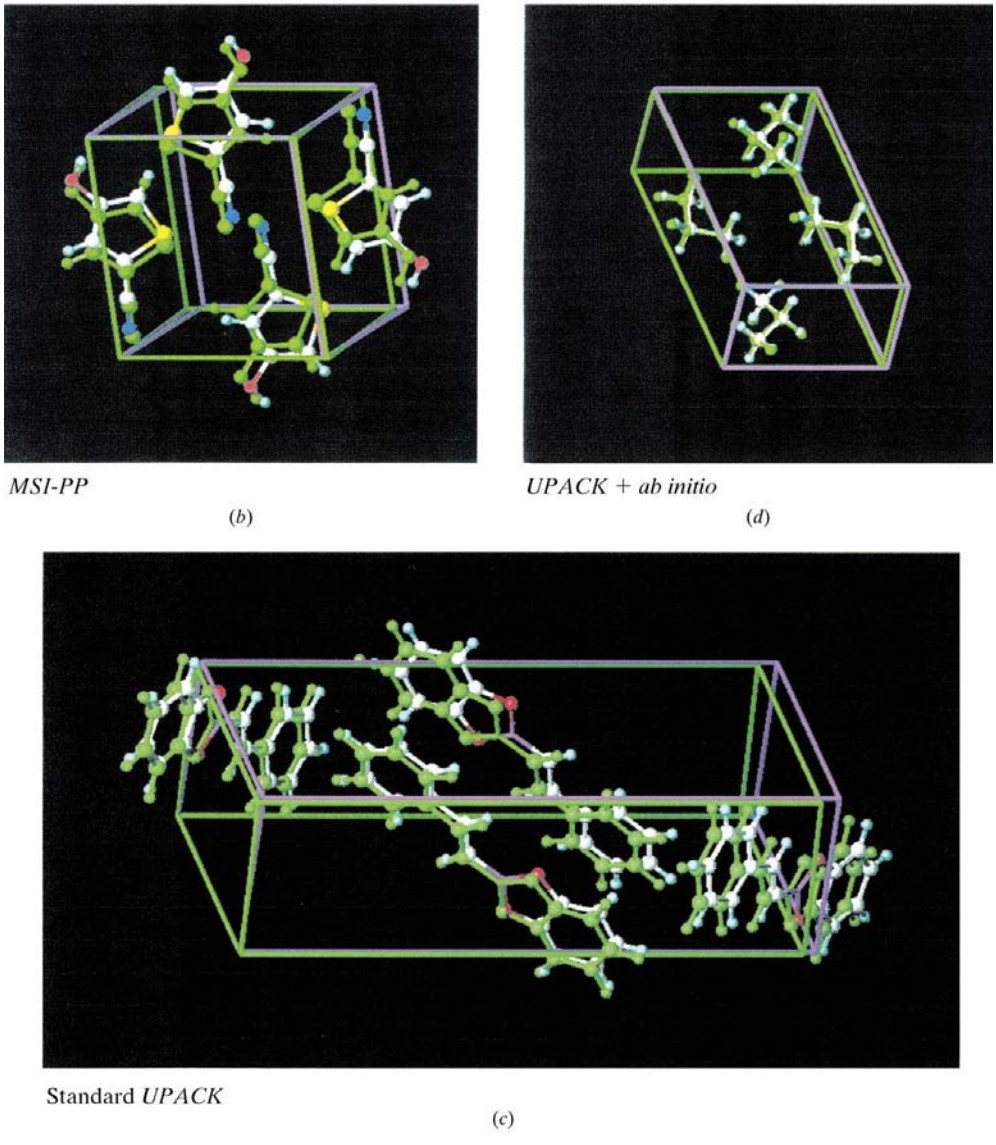

Figure 3 (continued)

5.3. *MOLPAK* (*Ammon*)

5.3.1. Compound (I). Owing to our interest in energetic materials for which high solid-state density is a desirable property, the highest predicted crystal density has generally been used, rather than the lowest lattice energy as an indication of the 'best' (correct) structure. The three solutions submitted for (I) can be characterized by (density, *WMIN* lattice energy, space group):

(1) 1.245 g cm^{-3}, -57.70 kJ mol^{-1}, $P2_1/c$;

(2) 1.221 g cm^{-3}, -57.70 kJ mol^{-1}, $P2_1$;

(3) 1.231 g cm^{-3}, -57.99 kJ mol^{-1}, $P2_1/c$.

Crystal energies, calculated with *Crystal*95 (Dovesi *et al.*, 1996; 6-21G/DFT) for a few of the solutions, indicated (2) as the lowest. The smallest *WMIN* lattice energy corresponded to (3), but the second lowest energy (–57.78 kJ mol^{-1}) corresponded to the correct solution for the *Pbca* polymorph. With *DMAREL* optimizations, this polymorph had the third smallest lattice energy

(−55.73 kJ mol^{-1}). *DMAREL* also found the $P2_1/c$ structure, but the lattice E (−53.43 kJ mol^{-1}) corresponded to the 20th smallest. The *WMIN* force field was woefully inadequate for the $P2_1/c$ polymorph; refinement of the experimental X-ray structure gave deviations in a, c and β of +27, −16 and −10%, respectively.

5.3.2. Compound (II). The densities and lattice energies ranged from 1.510–1.290 g cm^{-3} and −75.3 to −57.3 kJ mol^{-1}. Although neither quantity was successful in identifying the correct structure, it was contained within the top structures at 1.364 g cm^{-3} and −71.80 kJ mol^{-1}.

5.3.3. Compound (III). *Ab initio* geometry optimizations of planar and slightly non-planar models were performed to gauge the overall molecular planarity. All of the optimized models were completely planar and our calculations failed to predict the structure with a planar search probe. However, use of the experimental X-ray model and *WMIN* force field (with new B parameters; however, sp^2-hybridized carbon was a good substitute) easily yielded the correct structure. It had the lowest lattice energy (−118.16 kJ mol^{-1}); the next lowest lattice energy was −116.40 kJ mol^{-1}. There were two major differences between the X-ray and *ab initio* (B3LYP/6-31G*) models: in the X-ray model, the ethene–phenyl bond angle was larger by 4.7° and the ethene–phenyl torsion angle was 6°. Superposition of the benzodioxaborole portions of the two models (an almost perfect fit) showed an approximate 0.8 Å positional difference between the C atoms at the opposite ends (phenyls) of the models. In retrospect, it is surprising that such seemingly small differences between the *ab initio* and X-ray models separated success and failure, but the results indicate the importance of allowing some molecular flexibility during the packing process. This, however, will require a proper balance between the inter- and intramolecular components of the total energy.

5.4. *MPA* (*Williams*)

5.4.1. Compound (I). The force field of Cornell *et al.* (1995) was used.

Polymorph 1: Space group *Pbca* showed the lowest energy and the rank 1 energy in this space group corresponded to the observed polymorph 1. The predicted lattice constants were very close to the observed ones.

Polymorph 2: The observed space group for polymorph 2, $P2_1/c$, was fourth lowest in energy of the 18 space groups examined. The rank 1 energy structure in $P2_1/c$ did not correspond to the observed structure. This structure had lower energy than the relaxed observed structure, using this force field. Therefore, the predicted structure in $P2_1/c$ is possibly an unobserved third polymorph of (I).

5.4.2. Compound (II). The force field of *Biosym* was used, supplemented by the force field of Cornell *et al.* (1995) for sulfur. The observed space group $P2_1/c$ showed the lowest energy, but lattice constants were not correct because the predicted hydrogen-bonding scheme did not correspond to the observed scheme.

5.4.3. Compound (III). The force field of Williams & Cox (1984) was used for C, O and H, supplemented by the force field of Beringhelli *et al.* (1983) for the boron. Space group $P2_1/c$ showed the lowest energy, but none of the low-ranked energy structures in this space group corresponded to the observed structure.

5.4.4. Propane. The force field of Biosym was used. The lowest energy was obtained in the incorrect space group $P2_12_12_1$. The correct space group $P2_1/n$ gave the second lowest energy and the observed structure had energy rank 1 in this space group.

Only compound (I), polymorph 1, was correctly predicted. The failure to predict polymorph 2 was traced to the inadequacy of the force field. Relaxation of the observed structure of polymorph 2 yielded a maximum shift of 21.1% in the cell edges. For (II) the force field was better, yielding a maximum relaxation shift of 4.5% in the cell edges. However, the predicted hydrogen-bonding scheme in this structure was incorrect. For compound (III) the force field was excellent, yielding a maximum shift of only 1.3% in the cell edges. The crystal structure prediction failed because the molecular structure (for a gas molecule) was wrong. QM calculations produced a planar structure, but the observed structure is non-planar, so intermolecular forces distort this molecule. When the observed non-planar molecular structure was used, correct results were obtained for the crystal structure prediction. For propane the force field was inadequate, yielding a maximum shift of 11.2% in the cell edges. When an improved force field (Williams, 1999) was used the crystal structure was correctly predicted.

5.5. *MSI-PP* (*Verwer & Leusen*)

In total, 37 959, 58 715, 80 337 and 86 552 initial structures were optimized for (I), (II), (III) and propane, respectively. For each molecule, the results of all predictions were merged into a single set of predicted structures, ranked by energy, and three low-energy packings sufficiently differing from each other were selected. In all cases, the set of optimized structures contained the experimental polymorph(s) as obtained after minimization, showing that the sampling algorithm works sufficiently well for this type of molecule. The bottleneck is at the ranking stage, because many different structures are found within an energy range that is small compared with the errors which can be expected using a non-specific force field such as Dreiding.

For (I), the simulations found the experimentally observed *Pbca* polymorph as the global minimum. The $P2_1/c$ polymorph was also found by the search algorithm, but the force field used was not able to estimate its relative energy correctly (this polymorph ranked as number 266 in the list of predicted structures with an energy difference of 6.5 kJ mol^{-1} relative to the global minimum). The observed polymorph of (II) was predicted as the structure with the second lowest, only 0.24 kJ mol^{-1} above the predicted global minimum structure. For (III), the experimental crystal structure was found as the fifth lowest energy packing alternative – the relative energy is

Table 5
Experimental crystal structure data and crystal structure predictions for propane.

Method	P	Space group	a (Å)	b (Å)	c (Å)	α (°)	β (°)	χ (°)	Density (g cm^{-3})	RMSD Conf.	RMSD Pack.
Experimental	1	$P2_1/n$	4.148	12.612	6.977	90.00	91.28	90.00	0.807		
MPA	1	$P2_12_12_1$	8.064	4.074	10.538	90.00	90.00	90.00	0.850	0.0012	2.0
MPA	2	$P2_1/c$	4.072	8.082	10.510	90.00	90.18	90.00	0.851	0.0012	1.300
MSI-PP	1	$P\bar{1}$	6.700	8.578	4.295	67.76	60.56	88.05	0.753	0.0135	2.0
MSI-PP	2	$P2_1/n$	9.679	10.450	4.285	90.00	64.56	90.00	0.752	0.0143	2.0
MSI-PP	3	$P2_1/c$	9.930	6.388	6.198	90.00	87.73	90.00	0.750	0.0141	2.0
UPACK	1	$P\bar{1}$	4.051	6.487	7.636	76.28	82.31	60.30	0.870	0.0037	2.0
UPACK	2	$P2_1/c$	4.153	6.539	12.445	90.00	86.43	90.00	0.873	0.0047	0.726
UPACK	3	$P\bar{1}$	4.134	6.468	8.032	52.32	85.83	86.21	0.869	0.0048	2.0
UPACK*/*ab initio	**1**	$\mathbf{P2_1/c}$	**6.776**	**12.568**	**7.797**	**90.00**	**31.35**	**90.00**	**0.852**	**0.0027**	**0.163**
UPACK/ab initio	2	$P2_1/c$	25.401	4.067	13.066	90.00	14.95	90.00	0.846	0.0031	2.0
UPACK/ab initio	3	$P2_1/c$	6.554	4.058	13.530	90.00	76.14	90.00	0.843	0.0028	2.0
PackStar	1	$P2_12_12_1$	3.990	8.720	10.320	90.00	90.00	90.00	0.820	0.0088	2.0
PackStar	2	$P2_12_12_1$	4.190	8.190	11.020	90.00	90.00	90.00	0.779	0.0087	2.0
PackStar	3	$P2_12_12_1$	6.830	7.320	8.100	90.00	90.00	90.00	0.727	0.0087	2.0
Rancel	1	$Pna2_1$	9.150	10.136	4.140	90.00	90.00	90.00	0.767	0.0064	2.0
Rancel	2	$P\bar{1}$	15.014	4.336	4.142	102.53	69.19	60.35	0.768	0.0063	2.0
Rancel	3	$P\bar{1}$	4.181	8.143	6.793	76.01	59.76	77.70	0.764	0.0063	2.0

0.95 kJ mol^{-1}. The experimental structure of propane was ranked 15th, with a relative energy of 0.64 kJ mol^{-1} in our calculation.

With the exception of the $P2_1/c$ polymorph of (I), the unit-cell dimensions of the predicted crystal structures are within 5% of their true values, which is an expected result when using the Dreiding force field, and the molecular conformations and orientations in the cell are also close to those observed experimentally. Large deviations are found for the $P2_1/c$ polymorph of (I), both in cell parameters, where errors in individual cell constants accumulate to cause an error of more than 10% in the volume of the predicted cell, as well as in the packing itself: the planar dimer which is found experimentally is severely distorted in the predicted structure. The close C—H···O interaction which apparently favours the formation of the planar dimer did not lead to a lower energy in the Dreiding force field, as calculations on the isolated planar and distorted dimers show.

5.6. Standard *UPACK* (*van Eijck*)

To assess the success of the crystal structure predictions, the experimental structures were subjected to energy minimizations in the OPLS force field. The resulting structures were always present among the hypothetical ones, so in this respect the structure generation was successful. Further classification of success or failure depends on the degree of distortion of that structure from the experimental geometry, and on the ranking of its energy with respect to the hypothetical structure with the lowest energy.

For (I) and (III) some structures with low energy were recalculated with the Buckingham-type interatomic potential of Williams & Houpt (1986). Subjectively, some weight was given to these alternative energies and thus there are slight discrepancies between the rankings of the OPLS force field and the rankings as submitted in the blind test. The differences between the two force fields are not large, but sufficient for the experimental structure to be within the first three or not, which is what counted in this exercise.

5.6.1. Compound (I). There were 181 structures within 6 kJ mol^{-1}. The structure with the lowest energy corresponded to the experimental *Pbca* polymorph. On the other hand, the $P2_1/c$ polymorph is calculated with very poor geometry – it is almost unrecognisable – and the ranking is 57. A minor point of concern is that low-energy structures tend to lose their symmetry when the restrictions of space-group symmetry are removed. Despite the successful structure prediction for the *Pbca* polymorph, the force field obviously needs improvement.

5.6.2. Compound (III). Here only 99 structures were found within 6 kJ mol^{-1}. The experimental structure is ranked second, both with OPLS and Williams–Houpt potentials. For that reason this prediction was the only one on which we were offering a small bet (albeit with odds of 1 against 10) that the experimental structure would be either the first or second choice. That bet would have paid out. However, a very sobering observation is that the $P2_1/c$ structure proposed as the first choice by Price as well as by Williams was missed in the search, owing to an inadequate force field used in the first stage of the structure generation. In the OPLS force field this structure would have ranked first and the experimental structure would have been proposed as third.

5.6.3. Propane. The number of structures within 6 kJ mol^{-1} was as large as 558, so the confidence level of finding the correct structure was estimated to be almost zero. In fact, the experimental structure was ranked fifth with an energy difference of 0.23 kJ mol^{-1}. This is probably all that can be expected from a general-purpose force field such as OPLS. It is very gratifying that the *ab initio* force field developed by Mooij (see below) was able to select the correct structure out of this multitude of possibilities.

5.7. *UPACK* plus '*ab initio*' energy minimization (*Mooij*)

The *ab initio* potentials were developed in order to obtain more accurate estimates of the lattice energies of different crystal packings. This should result in more reliable predictions, both in terms of geometry of the predicted structures, as well as of energy and ranking. The *ab initio* potential for propane is thought to be fairly accurate. This is supported by the good reproduction of both high-level *ab initio* data (Mooij, van Duijneveldt, 1999) and the experimental crystal structures of some alkanes (Mooij, van Eijck & Kroon, 1999). However,

Table 6
Deviations for the successful prediction from the experimental structures.

ΔL: largest relative error in a cell axis; $\Delta\varphi$: largest deviation in a cell angle; ΔX: molecular translation (calculated *via* fractional coordinates); $\Delta\tau$: molecular rotation; ΔE: energy difference with the lowest-energy structure that was found.

Compound	Method	ΔL (%)	$\Delta\varphi$ (°)	ΔX (Å)	$\Delta\tau$ (°)	Rank	ΔE (kJ mol^{-1})
(I) (polymorph 1)	*MPA*	3.5	–	0.27	6.3	1	0
	MSI-PP	4.0	–	0.14	5.1	1	0
	UPACK	1.4	–	0.73	7.1	1	0
	Zip-Promet	2.4	–	0.19	4.4	1	0
(II)	*MSI-PP*	3.6	4.3	0.43	4.8	2	0.24
(III)	*UPACK*	2.3	1.9	0.11	5.6	2	0.35
Propane	*UPACK* + *ab initio*	2.9	0.4	0.04	1.5	1	0

the Csp^2—H potential developed for this test is much less accurate. Firstly, the fit to the ethene and ethene–methanol dimer energies was not very satisfactory (RMSD = 1.1 kJ mol^{-1} for the 45 bound dimers). It is interesting to note that the subset of 14 C—H···O-bonded methanol–ethene dimers are fitted much more accurately (RMSD = 0.17 kJ mol^{-1}). So, C—H···O hydrogen bonds appear to be accurately modelled by this potential, but there are larger flaws in the ethene–ethene part. Secondly, the transferability of parameters derived from ethene to aromatic hydrocarbons is questionable. Indeed, the potential is not capable of accurately reproducing the crystal structures of benzene and furan (molecular reorientations up to 19°). Obviously, there is room for improvement of this potential for aromatic compounds and it would have been better to parameterize on an aromatic system such as benzene. This would be computationally very expensive and was in any case impossible in the time frame of this blind test.

5.7.1. Compound (I). The three submitted structures did not contain either of the two observed polymorphs. The *Pbca* form was predicted as number 12 in our list of hypothetical structures, with an energy difference of 2.1 kJ mol^{-1}. The $P2_1/c$ form was number 5 at 1.2 kJ mol^{-1}. Indeed, the 'disappearing' of the *Pbca* polymorph suggests that the $P2_1/c$ form is the more stable one. Both predicted structures show considerable deviations from the experimental packings, with maximum changes in cell axes around 6% and molecular reorientations of 7°. These deviations are not unexpected, considering the results for furan. In the $P2_1/c$ polymorph the major 'building block' is a planar, doubly C—H···O bonded dimer. This motif is excellently reproduced by our *ab initio* potential. The failure of the standard force fields in predicting the $P2_1/c$ polymorph can be traced back to errors in the modelling of this dimer motif (see *MSI-PP*; Verwer & Leusen, 1998). Apart from our *ab initio* potential, only *DMAREL* predicted the $P2_1/c$ polymorph as a feasible structure, suggesting that an accurate electrostatic model is essential.

5.7.2. Compound (III). For (III) the energy-minimized experimental structure is predicted as being the most favourable. However, it does not qualify as a correct prediction, since deviations are very large; the molecule is reoriented by 17°, resulting in changes in cell axes up to 9%. As for (I), this supports the notion that the *ab initio* ethene-derived potential for such molecules is not as accurate as the potential for alkanes, ethers and alcohols. For (III) standard force fields with ESP-derived charges seem to be more accurate (see the results of van Eijck and Verwer & Leusen).

5.7.3. Propane. It is encouraging to see that for the compound for which we had the most confidence in the potential, the prediction is (the) most successful. Not only is the experimental structure predicted as the energetically most favourable one, but also the deviation from the experimental structure is low, with a maximum change in cell axes of 3%, and a molecular reorientation of 1.5°. Based on this result one could speculate that the use of an equally accurate intermolecular potential for the other compounds would produce similar results for their prediction.

5.8. *Zip-Promet* (*Gavezzotti, Schweizer & Dunitz*)

5.8.1. Compound (I). In each selected space group, 1000–2000 starting structures were generated and subjected to several cycles of energy minimization. Approximately half of the structures with the highest energy were eliminated after each cycle. The refinements were terminated when no significant change in energy occurred. Separate runs with *Minopec* both with and without charges were carried out – in the former case, without any forced convergence procedure but using the standard cut-off calculated by the *Promet* protocol.

According to our calculations, the *Pbca* structure submitted has the best lattice energy (−68 kJ mol^{-1}). However, as we found more than 20 structures in this and in other space groups within an energy window of 2–3 kJ mol^{-1}, and as the relative ranking of these structures vary with assumed atomic charges (which are 'reasonable' but not rigorously defined), and as the calculations refer to molecules at rest, there could be no real basis for confidence in our prediction. None of the participating groups guessed correctly the structure of the stable $P2_1/c$ polymorph, which was not even similar to any of the low-energy structures we found in this space group. We refined the experimental structure lattice energy to −65 kJ mol^{-1}, slightly higher than our best $P2_1/c$ crystal structure (−66.1 kJ mol^{-1}).

5.8.2. Compound (III). The molecule was taken to be flat, on the assumption that the intermolecular advantage of a flat molecule will overrule any (small) preference for twisted conformations. Boron was treated as carbon. The ten top-ranking (energywize) crystal structures all have a short cell edge between 3.5 and 5.6 Å. They belong to four different space groups and their lattice energy is between 126.7 and 125.0 kJ mol^{-1}: the first two were then submitted, by the rules of the game, as 'predictions', but with zero confidence level. The cell volume of the top-ranking calculated structure is

almost identical to that of the X-ray structure, whose calculated lattice energy is 125.0 kJ mol^{-1}. These results:

(i) show that the crystal density may be predictable even with the wrong structure;

(ii) leave no ground for blaming the prediction failure on bad chargeless potentials (indeed, the general results of this test demonstrate, in our opinion, that success is not potential-dependent);

(iii) confirm that reproducible crystal structure prediction is still a far-away goal.

A large number of crystal structures with overlapping molecules arranged in flat layers were found, because parallel pairs at a distance of 3.5 Å are very favourable precursors. Flat aromatics of small size will rather form herringbone patterns (Desiraju & Gavezzotti, 1989), but it was thought unfair to override the automatic *Zip-Promet* screenout procedures to accept low-stability precursors, such as herringbone ribbons. Besides, the override introduces some subjective judgement that was also thought inappropriate in a real blindfold test. Thus, we blame our present failure mainly on the *Promet aufbau* architecture. A switch has now been introduced which may force the search to proceed on screw axes or glide planes with a longer translation period, thus avoiding 3.5 Å cell edges.

5.9. *FlexCryst* (*Hofmann*)

The program, initially developed for fast structure generation (the calculation time for all three compounds amounts to 73 min) rather than for structure prediction, did not succeed for any of the compounds. The algorithm subdivides into two steps, the crystal structure generation and the crystal structure scoring. Both steps were analysed for possible improvements.

The program has no option for structure generation in space group *Pbca* so the polymorph of (I) in *Pbca* could not be generated. In the other cases, for which the experimental structure is space group $P2_1/c$, the inversion centres were located almost correctly, but the screw-rotation axis was not found in the correct position. Indeed, the program finds very similar structures assuming the correct direction of the screw axis in the three compounds. For comparison, the molecule of the experimental and the generated structure were superimposed and the difference between the unit cell vectors was calculated. For the three molecules the maximum distance between two unit cell corners was 1.04, 1.67 and 1.89 Å. Such distances are close enough to reproduce the correct structure after refinement.

The rank of the experimental structure has been calculated for the structures in $P2_1/c$, to check the quality of the scoring function. Among the generated structures the experimental structure ranks in positions 3687, 594 and 132. Owing to the algorithm, which generates up to 50 structures for each minimum, this indicates a slightly incorrect potential for (II) and (III). For the $P2_1/c$ polymorph of (I), which was not highly ranked by any method, this program gives a score outside any confidence limits. The predicted structures, based on very recently developed statistical potentials, are all more loosely packed compared with the experimental one, indicating that the scoring function needs further improvement. The quality of the scoring function depends on the possibility to separate overlapping effects of several atom pair interactions in functional groups and also on the available data in the CSD. Both effects are under investigation.

5.10. *PackStar* (*Lommerse*)

5.10.1. Compound (I). The experimentally based crystal field environments of cyclobutafuran did not reveal any specific favourable interaction: there were no high propensity areas or highly directional parts present in the propensity maps. The lowest cost structure (in $P2_12_12_1$, cost = −37) had a very different packing by comparison with both the experimental polymorphs. The key aromatic C—H···O interactions in the experimental structures were not reproduced, instead the weaker aliphatic C—H$_2$···O interactions were predominant. The costs for the experimental structures, −24 (*Pbca* polymorph) and −25 ($P2_1/c$ polymorph), were less favourable in the applied crystal fields. Two important reasons for these deviations were identified:

(i) an insufficiently large calculated interaction area around the molecules (sum of the van der Waals radii plus 0.5 Å), causing the crystal structure densities to be generally too low, and

(ii) the unusual short C—H···O distance (2.61 Å, *Pbca* polymorph), probably caused by the rather electron-deficient aromatic C—H group (owing to the adjacent furan oxygen), which was not well reproduced by the general aromatic C—H contact group used.

5.10.2. Compound (II). For (II), the important possible interactions were reproduced in the propensity maps: the cyano···H—O and O—H···O—H hydrogen bonds. All low-cost structures favoured the cyano···H—O hydrogen bonds, forming chains throughout the structures. Besides the chain, the calculated lowest cost structure (−40.5) demonstrated very similar features to those in the experimental structure, such as the formation of layers. However, in the experimental structure an unusually short S···O distance is present which was unfavourable according to the propensity maps. This forced an alternating 180° flip of the molecule within the chain. It is likely that this short S···O distance is caused by the strong electron-withdrawing cyano group adjacent to the sulfur atom, making the latter rather electron deficient by comparison with a general thiophene fragment. It is encouraging to observe that allowing for this short S···O distance, the cost of the experimental structure turned out to be the lowest (−40.8, not optimized).

5.10.3. Compound (III). Again, few specific favourable interactions were present in the propensity maps for the rigid planar molecule. Like the experimental structure (cost −57.8), the lowest cost structure (−66.3) as well as most other predicted structures showed the molecular herringbone structure or zigzag pattern. However, a number of close contacts found in the experimental structure were not allowed in the propensity maps, such as short C—H···O and C—

H···C contacts, so the experimental structure could not have been predicted with the applied crystal field environments. In all cases, the densities of the calculated structures were too low, revealing similar problems to the calculations for (I).

5.10.4. Propane. The solutions suggested by the program showed packings with densities close to the experimental structure, demonstrating that the small interaction space used in the calculations becomes less of a problem for (very) small molecules. The calculated packings were, however, different.

Summarizing, experimental data can be used directly for the construction of possible crystal structures. The method is generally applicable for a large variety of molecules, as long as sufficient experimental data are available. Although the packings are realistic, a number of problems using this approach have been revealed. Most importantly, care has to be taken building a molecule from its functional fragments. The properties of a group can be strongly influenced by adjacent groups. This was most strongly revealed by the rather acidic aromatic hydrogen in (I) and the electron-deficient sulfur in (II). The interaction space of a molecule, being the sum of the van der Waals radii plus 0.5 Å only, is clearly not sufficient, especially when weak interactions determine the final crystal structure. We are working on an improved approach to tackle both major and minor deficiencies of the current program.

5.11. *Rancel* (*Motherwell*)

5.11.1. Compound (I). Compound (I) gave a range of the cost function Q from 9.6 to 65.0. The first choice was made in space group $P\bar{1}$, Q = 23.3, rank 13, because of a strong similarity to one of the CSD reference sets (GOHWAB) and the lowest Q value in $P\bar{1}$. The second choice was the global minimum in the set of 80 runs (10 trials in 8 space groups) in *Pbca*, but the packing is wrong. When the experimental structures in *Pbca* and $P2_1/c$ became available, the Q values were very high (*ca* 64.0) and could not have led to the correct solution. With hindsight, the choice of reference molecules from the CSD was poorly made, because the molecule has a significantly lower melting point than the reference molecules, leading to generally higher frequency values at any given distance. Thus, molecular size should have been used in the selection and, for example, the frequency curves for furan (FURAN01) are quite close to the experimental data for (I). The only penalty condition was that all atoms of the furan ring should each have two H-atom contacts within 3.1 Å. Further examination of CSD molecules suggests strongly that conditions could have been set on C—H···O distances, involving the furan H. In particular, the centrosymmetric dimer motif for (I) in $P2_1/c$ was found exactly in several CSD structures of low molecular weight, planar conformation and no other hydrogen donors or acceptors.

5.11.2. Compound (II). The results for (II) showed the correct hydrogen-bonded translational chain, aligned with the monoclinic b axis 8.253 Å (*cf.* 8.332 Å experimentally), but wrongly packed in the space group $C2/c$, not $P2_1/c$. The second and third choice molecules showed different hydrogen-bonded chains. Subsequent inspection of CSD-similar molecules suggest that penalty conditions could have been set to expect C—H···O interactions and also allow a close S···O contact of 3.20 Å.

5.11.3. Compound (III). Compound (III) was calculated using a rigid planar model; the first choice was in $P2_1/c$, but is wrongly packed. Inspection of the CSD shows that the planar ribbon arrangement of molecules linked by weak C—H···O interactions is often observed in the packing of planar molecules. No CSD-derived penalty conditions had been set, although a small number of chemically similar molecules show these ribbons.

These calculations show that the use of such a simple fitness function based only on isotropic interatomic distances is insufficient for prediction of these small (nearly) planar molecules. It is possible that the Q function might have some discriminatory value when combined with lattice energy calculations and further work of this nature is planned.

6. Conclusions

The results of our workshop have provided an objective evaluation of the possibilities and limitations of current methods and claims of crystal structure prediction. The limitation to three predictions per method per compound may have been somewhat too severe, but the fact that seven correct predictions were made (five even as the first choice!) is encouraging. It demonstrates, at least for the classes of molecules considered here, that the prediction problem, although beset with difficulties, is not hopeless.

The occurrence of (I) in two polymorphs raises interesting questions. There is always the possibility and indeed the likelihood that still undetected, additional polymorphs of all four compounds may exist. Some of our 'unsuccessful' predictions might well find their targets in this undiscovered territory and thus increase our overall score. However, even if additional polymorphs were to be found, it is most unlikely that our present methods could predict the conditions under which one polymorph or the other would be formed. Compound (I) provides an interesting example. Although the *Pbca* form was first obtained, subsequent crystallizations consistently produced the $P2_1/c$ form, thus leading to the view that this is probably the more stable form. If the less stable *Pbca* form had not been found, our overall score of seven correct predictions would have fallen to 3. The more stable $P2_1/c$ form was not predicted by any participant, a failure that underlines the lack of reliability of the present methods.

Compound (II) has considerable electrostatic character compared with (I) and was not attempted by several participants. It may seem puzzling that *DMAREL* with its more elaborate multipole description of the electrostatic terms did not lead to a correct prediction, while *MSI-PP* was successful with only a point charge model. However, post-analysis with *DMAREL* gave an energy for the experimental structure below the global minimum of earlier runs, thus pointing to a failure in the search procedure rather than in the force field.

The experience with compound (III) may be taken to indicate the importance of molecular flexibility in the crystal

packing calculations. Even small deviations from molecular planarity may affect the ordering of different crystal packing energies.

Propane, the smallest molecule, was correctly predicted by only one program, using an *ab initio* potential. This was almost the only case where a successful participant expressed confidence in the reliability of his prediction, based on the goodness-of-fit to high-level *ab initio* data and results for related compounds.

The results of the workshop provide only partial answers to the question of the effectiveness of the four main components of the prediction programs, as mentioned in the *Introduction*.

(i) Molecular model: methods that use flexible rather than rigid molecular models are clearly more adaptable than rigid models, at the price of additional computer time. In the post-analysis of (III), several participants claim that the correct solution would have been found if the experimental molecular geometry had been available. For the $P2_1/c$ polymorph of (I), the energy calculations are very sensitive to the positioning of hydrogen in the crucial C—H$\cdots$O packing motif.

(ii) Generation of trial structures: most of the algorithms were able to produce structures that bore some resemblance to the experimental structure. However, the problems of handling very flexible molecules, uncommon space groups, or $Z' > 1$, are undoubtedly more difficult and were not addressed by this exercise.

(iii) Fitness functions: successful predictions were obtained only where energy was used as a fitness function. For reliable prediction, however, it will be necessary to include effects of molecular motion in the crystals by means of molecular or lattice dynamical methods. Statistical methods, based on the CSD data, have yet to establish their usefulness in predicting unknown structures. Ultimately, our lack of knowledge about thermodynamic and kinetic aspects of crystal nucleation and growth may be a hindrance that will be hard to surmount.

(iv) Minimization methods: most methods applied minimization to the generated crude structures by using first- or second-order derivatives of the energy function, or by the Simplex method. Success or failure of the various methods is not dependent on the minimization method as such, but on the search and generation algorithm and the fitness function. The minima found for the successful predictions were satisfactorily close to the experimental structures.

In summary, at the present stage of development, perhaps the best that can be expected from crystal structure prediction programs is to provide a list of possible candidates for experimentally observable polymorphs.

The organisers of the CCDC workshop thank Dr P. R. Raithby for acting as the independent referee for the blind test and R. Boese, A. J. Blake and W. Clegg for kindly providing the compounds with unpublished experimental crystal structures. BPvE thanks S. Polling for preliminary calculations on furan and on compound (I). DEW thanks A. Abraha for carrying out most calculations for this blind test. HLA thanks J. R. Holden and Z. Du for their important contributions and hard work. Financial support was provided by the Office of Naval Research (N000149710488); thanks to program officers R. Miller and J. Goldwasser. PV and FJJL thank J. van de Streek for carrying out the calculations on propane. PV and WTMM thank the NWO-CW/PPM-CMS-crystallization project for financial support. SLP thanks T. Beyer and H. H. Y. Tsui for carrying out the predictions on compounds (I) and (III), respectively, and the EPSRC for computational and financial support.

References

Allen, F. H., Davies, J. E., Johnson, O. J., Kennard, O., Macrae, C. F., Mitchell, E. M., Mitchell, G. F., Smith, J. M. & Watson, D. G. (1991). *J. Chem. Inf. Comput. Sci.* **31**, 187–204.

Amos, R. D. & Andrews, J. S. (1995). *CADPAC*6, ed. 6.0. University Chemical Laboratory, Cambridge, England. (Contributions from Alberts, I. L., Colwell, S. M., Handy, N. C., Jayatilaka, D., Knowles, P. J., Kobayashi, R., Koga, N., Laidig, K. E., Laming, G., Lee, A., Maslen, P. E., Murray, C. W., Rice, J. E., Simandiras, E. D., Stone, A. J., Su, M. D. & Tozer, D. J.)

Andreev, Y. G., Lightfoot, P. & Bruce, P. G. (1997). *J. Appl. Cryst.* **30**, 294–305.

Bauer, H. B. & Kassner, D. (1992). *Acta Cryst.* B**48**, 356–369.

Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. (1993). *J. Phys. Chem.* **97**, 10269–10280.

Beringhelli, T., Filippini, G., Gavezzotti, A. & Simonetta, M. J. (1983). *Mol. Struct.* **94**, 51–61.

Biosym Technologies (1969). *Discover*. Biosym Technologies, San Diego, CA, USA.

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.* B**55**, 963–974.

Boese, R. (1999). Personal communication.

Boese, R. & Garbarczyk, J. (1998). Unpublished results.

Boese, R., Weiss, H.-C. & Bläser, D. (1999*a*). *Angew. Chem.* **111**, 1042–1045.

Boese, R., Weiss, H.-C. & Bläser, D. (1999*b*). *Angew. Chem. Int. Ed. Engl.* **38**, 988–992.

Breneman, C. M. & Wiberg, K. B. (1990). *J. Comput. Chem.* **11**, 361–373.

Bruno, I. J., Cole, J. C., Lommerse, J. P. M., Rowland, R. S., Taylor, R. & Verdonk, M. L. (1997). *J. Comput.-Aided Mol. Des.* **11**, 525–537.

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

CambridgeSoft Corporation (1998). *Chem*3D. CambridgeSoft Corporation, Cambridge, MA, USA.

Clegg, W. & Scott, A. J. (1998). Unpublished results.

Coombes, D. S., Price, S. L., Willock, D. J. & Leslie, M. (1996). *J. Phys. Chem.* **100**, 7352–7360.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr, Fergusson D. M., Spellmeyer, D. C., Thomas, F., Caldwell, J. W. & Kollman, P. A. (1995). *J. Am. Chem. Soc.* **117**, 5179–5197.

Cox, S. R., Hsu, L. Y. & Williams, D. E. (1981). *Acta Cryst.* A**37**, 293–301.

Desiraju, G. R. & Gavezzotti, A. (1989). *Acta Cryst.* B**45**, 473–482.

Doerksen, R. J. & Thakkar, A. J. (1999). *J. Phys. Chem. A*, **103**, 2141–2151.

Dovesi, R., Saunders, V. R., Roetti, C., Causa, M., Harrison, N. M., Orlando, R. & Apra, E. (1996). *Crystal*95 *User's Manual.* University of Torino, Italy.

Dunitz, J. D. & Bernstein, J. (1995). *Acc. Chem. Res.* **28**, 193–200.

Eijck, B. P. van & Kroon, J. (1999). *J. Comput. Chem.* **20**, 799–812.

Eijck, B. P. van, Mooij, W. T. M. & Kroon, J. (1995). *Acta Cryst.* B**51**, 99–103.

Engel, G. E., Wilke, S., Konig, O., Harris, K. D. M. & Leusen, F. J. J. (1999). *J. Appl. Cryst.* **32**, 1169–1179.

Filippini, G. & Gavezzotti, A. (1993). *Acta Cryst.* B**49**, 868–880.
Frisch, M. J., Trucks, G. W., Schlegel, H. B., Gill, P. M. W., Johnson, B. G., Robb, M. A., Cheeseman, J. R., Keith, T., Petersson, G. A., Montgomery, J. A., Raghavachari, K., Al-Laham, M. A., Zakrzewski, V. G., Ortiz, J. V., Foresman, J. B., Cioslowski, J., Stefanov, B. B., Nanayakkara, A., Challacombe, M., Peng, C. Y., Ayala, P. Y., Chen, W., Wong, M. W., Andres, J. L., Replogle, E. S., Gomperts, R., Martin, R. L., Fox, D. J., Binkley, J. S., Defrees, D. J., Baker, J., Stewart, J. P., Head-Gordon, M., Gonzalez, C. & Pople, J. A. (1995). *Gaussian*94, Revision D.4. Gaussian, Inc., Pittsburgh PA, USA.
Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A. Jr, Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S. & Pople, J. A. (1998). *Gaussian*98W, Revision A.6. Gaussian, Inc., Pittsburgh PA, USA.
Gavezzotti, A. (1991). *J. Am. Chem. Soc.* **113**, 4622–4629.
Gavezzotti, A. (1994). *Acc. Chem. Res.* **27**, 309–314.
Gavezzotti, A. (1999). *Zip-Promet with Minopec.* University of Milan, Italy.
Gavezzotti, A. & Filippini, G. (1994). *J. Phys. Chem.* **98**, 4831–4837.
Gavezzotti, A. & Filippini, G. (1995). *J. Am. Chem. Soc.* **117**, 12299–12305.
Gdanitz, R. J. (1998). *Curr. Opin. Solid State Mater. Sci.* **3**, 414–418.
Halgren, T. A. (1992). *J. Am. Chem. Soc.* **114**, 7827–7843.
Hofmann, D. W. M. & Lengauer, T. (1997). *Acta Cryst.* A**53**, 225–235.
Hofmann, D. W. M. & Lengauer, T. (1998). *J. Mol. Mod.* **4**, 132–144.
Holden, J. R., Du, Z. Y. & Ammon, H. L. (1993). *J. Comput. Chem.* **14**, 422–437.
Howell, J., Rossi, A., Wallace, D., Haraki, K. & Hoffmann, R. (1977). *ICON.* Quantum Chemical Program Exchange No. 517.
Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives. J. (1996). *J. Am. Chem. Soc.* **118**, 11225–11236.
Karfunkel, H. R., Leusen, F. J. J. & Gdanitz, R. J. (1993). *J. Comput.-Aided Mater. Des.* **1**, 177–185.
Ketelaar, J. A. A. (1953). *Chemical Constitution.* Amsterdam: Elsevier.
Leusen, F J. J. (1996). *J. Cryst. Growth*, **166**, 900–903.
Leusen, F. J. J., Wilke, S., Verwer, P. & Engel, G. E. (1999). *Implications of Molecular and Materials Structure for New Technologies*, NATO Science Series E, Vol. 360, edited by J. A. K. Howard, F. H. Allen & G. P. Shields, pp. 303–314. Dordrecht: Kluwer Academic Publishers.
Lommerse, J. P. M. & Motherwell, W. D. S. (2000). In preparation.
Mayo, S. L., Olafson, B. D. & Goddard III, W. A. (1990). *J. Phys. Chem.* **94**, 8897–8909.
Molecular Simulations Inc. (1999). *Cerius*^{2}. Molecular Simulations Inc., 9685 Scranton Road, San Diego, CA, USA.
Mooij, W. T. M. (2000). Thesis. Utrecht, The Netherlands.
Mooij, W. T. M., van Duijneveldt, F. B., van Duijneveldt-van de Rijdt, J. G. C. M. & van Eijck, B. P. (1999). *J. Phys. Chem.* **103**, 9872–9882.
Mooij, W. T. M., van Eijck, B. P. & Kroon, J. (1999). *J. Phys. Chem.* **103**, 9883–9890.
Motherwell, W. D. S. (1999). *Nova Acta Leopoldina*, **NF79**, 89–98.
Nobeli, I. & Price, S. L. (1999). *J. Phys. Chem. A*, **103**, 6448–6457.
Ponder, J. W. (1998). *Tinker*, Version 3.6. Washington University of Medicine, USA.
Potter, B. S., Palmer, R. A., Withnall, R., Chowdhry, B. Z. & Price, S. L. (1999). *J. Mol. Struct.* In the press.
Price, S. L. & Wibley, K. S. (1997). *J. Phys. Chem. A*, **101**, 2198–2206.
Schaftenaar, G. & Noordik, J. H. (1999). *J. Comput.-Aided Mol. Des.* In the press.
Schmidt, M. U. (1995). *Kristallstrukturberechnungen metallorganischer Molekülverbindungen.* Aachen, Germany: Verlag Shaker.
Schmidt, M. U. (1999). *Crystal Engineering: From Molecules and Crystals to Materials*, edited by D. Braga and G. Orpen. Dordrecht: Kluwer Academic Publishers.
Schmidt, M. U. & Englert, U. (1995). European Crystallographic Meeting 16, Lund, Sweden, 6-11.8.1995.
Schmidt, M. U. & Englert, U. (1996). *J. Chem. Soc. Dalton Trans.* pp. 2077–2082.
Schmidt, M. U. & Kalkhof, H. (1997). *CRYSCA*. Clariant GmbH, Frankfurt Am Main, Germany.
Shankland, K., David, W. I. F. & Csoka, T. (1997). *Z. Kristallogr.* **212**, 550–552.
Spackman, M. A. (1996). *J. Comput. Chem.* **17**, 1–18.
Stone, A. J. & Alderton, M. (1985). *Mol. Phys.* **56**, 1047–1064.
Tsui, H. H. Y. & Price, S. L. (1999). *Cryst. Eng. Commun.* No. 007.
Verdonk, M. L., Cole, J. C. & Taylor, R. (1999). *J. Mol. Biol.* **289**, 1093–1108.
Verwer, P. & Leusen, F. J. J. (1998). *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd, pp. 327–365. New York: Wiley-VCH.
Wavefunction Inc. (1999). *PC Spartan Pro.* Wavefunction Inc., Irvine, CA, USA.
Williams, D. E. (1971). *Acta Cryst.* A**27**, 452–455.
Williams, D. E. (1973). *Acta Cryst.* A**29**, 408–414.
Williams, D. E. (1994). *J. Comput. Chem.* **15**, 719–732.
Williams, D. E. (1997). *Pdm*97. University of Louisville, Kentucky, USA.
Williams, D. E. (1999*a*). *J. Mol. Struct.* pp. 321–347.
Williams, D. E. (1999*b*). *J. Mol. Struct.* pp. 485–486.
Williams, D. E. & Cox, S. R. (1984). *Acta Cryst.* B**40**, 404–417.
Williams, D. E. & Houpt, D. J. (1986). *Acta Cryst.* B**42**, 286–295.
Willock, D. J., Price, S. L., Leslie, M. & Catlow, C. R. A. (1995). *J. Comput. Chem.* **16**, 628–647.