

An Assessment of Lattice Energy Minimization for the Prediction of Molecular Organic Crystal Structures

Graeme M. Day,[†] James Chisholm,[‡] Ning Shan,[†] W. D. Sam Motherwell,[‡] and William Jones^{*,†}

The Pfizer Institute for Pharmaceutical Materials Science, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK, CB2 1EW, and The Pfizer Institute for Pharmaceutical Materials Science, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, UK, CB2 1EZ

Received June 10, 2004; Revised Manuscript Received August 18, 2004

ABSTRACT: Lattice energy searches for theoretical low-energy crystal forms are presented for 50 small organic molecules, and we compare the experimentally observed crystal forms to these lists of hypothetical polymorphs. For each known crystal, the relative stability is calculated with respect to the global minimum energy structure, and we determine the number of unobserved structures lower in energy than the experimental form. The distributions of these relative energies and their rankings in the predicted lists are used to determine the efficacy of lattice energy minimization in crystal structure prediction. Although a simple form for the interaction energies has been used, the calculations produce almost a third of the known crystals as the global minimum in energy, and approximately a half of the known structures are within 1 kJ/mol of the global minimum. Molecules with no hydrogen-bonding capacity are most likely to be found close to the global minimum in lattice energy, while increasing the number of possible hydrogen-bond donor–acceptor combinations leads to less reliable predictions.

1. Introduction

Computational methods of predicting the most likely crystal structures of a molecule from its atomic connectivity alone could be of great value in several areas of materials chemistry. The predicted properties of such hypothetical polymorphs could be used as a screen for molecular crystals with novel properties, such as a high nonlinear optical response. Furthermore, the forecasting of energetically competitive polymorphs may help avoid the potentially disastrous problems associated with unanticipated crystal forms appearing late in the processing cycle of a pharmaceutical molecule.

The field of crystal structure prediction (CSP) has benefited from many detailed studies of well-chosen molecules, such as paracetamol¹ and *p*-dichlorobenzene,² as well as collaborations comparing alternate methodologies for small sets of molecules.^{3–5} Surveys of larger groups of molecules^{6–11} have also been valuable, but have usually been limited to specific classes of molecules, often in the development of model potentials. Such detailed studies of single molecules or small families have the advantage of being able to fine-tune the prediction methodology as well as the possibility of performing extensive experimental investigations in parallel with the theoretical study.

A more general survey does not allow for fine-tuning of the method for each molecular system and so may not give as reliable results for the individual molecules. However, it can provide a broader picture that can answer different questions:

(i) How often have experimentally observed crystals not found the global free energy minimum (i.e., are

kinetic products of crystallization) and, so, what success rate can we expect from CSP by global minimization methods?

(ii) Crystals of which types of molecules are more likely to be predictable by lattice energy minimization (see the recent survey asking this question¹²), and are there molecular characteristics that make them amenable to this most common of prediction methods?

(iii) Where the energy is not the sole factor determining the final crystal structure, what other calculated properties or nonenergetic descriptors can help determine the most likely observable structures generated by CSP searches?

Differences in methodology among the growing number of prediction studies make these questions difficult to answer through literature surveys. Predictions are especially dependent on the model used to calculate energies, and, indeed, different schemes for calculating intermolecular energies can give contrasting results for a given molecule. Pyridine is a good example—an atom–atom model potential with electrostatics described by atomic multipoles found that there are many lower energy structures than the known polymorphs,¹³ but by calculating interaction energies for the same hypothetical crystals by the semiclassical density sums (SCDS) pixel approach,^{14,15} the global minimum does correspond to one of the known forms.¹⁵ The approaches are different and each describes certain contributions to the interaction energy very well. The discrepancies are therefore noteworthy and highlight the sensitivity of crystal structure predictions to the approach used to describe the intermolecular energy. Such variations in results make it difficult to draw general conclusions from individual studies of particular molecules. However, we hope that, over a large set of molecules, using a consistent computational method, such variations will not obscure the overall picture and we can make

* To whom correspondence should be addressed. Tel: +44 (0) 1223 336468. Fax: +44 (0) 1223 336362. E-mail: wj10@cam.ac.uk.

[†] University of Cambridge.

[‡] Cambridge Crystallographic Data Centre.

conclusions about the general approach to crystal structure prediction. We have, therefore, performed a study of a large group of organic molecules, with a variety of shapes, sizes, and constituent functional groups. This diverse group of molecules is limited only by the criteria of being (a) small and reasonably rigid (less than 25 atoms) and (b) composed of common atom types, for which high-quality parameters are available for the model potential.

There are numerous software packages that generate lists of hypothetical crystal packings for a given molecule, many of which are described in Verwer and Leusen's review of crystal structure prediction methods.¹⁶ The problem of searching phase space is being solved by the use of clever algorithms and the rapid increase in available computing power. At least for small, rigid molecules, the generation of all of the plausible structures should now rarely be a stumbling point, at least when there is only one symmetrically independent molecule ($Z' = 1$). It is the post-search ranking of structures that is the main obstacle to reliable crystal structure prediction for small, rigid molecules. In the present study, we only consider the lattice energy of the hypothetical crystals. However, the collection of these predicted crystal structures can act as a future test set for other methods of ranking the probabilities of observing computed structures.

We chose to start with a form of model potential that requires no particularly specialized modeling software—the pairwise additive, isotropic repulsion–dispersion model with the charge distribution described by atomic point charges. Higher quality models, such as those explicitly including polarization,^{17,18} using atomic multipoles for more accurate electrostatics,^{19–21} anisotropic atom–atom repulsion,^{2,22} or more detailed partitioning of the molecular charge density,^{14,15} are certain to yield better results, but we leave the effects of such improvements in the model for future study. Furthermore, lattice energy calculations ignore the potentially important effects of lattice vibrational energy differences, especially the entropy – enantiotropic polymorphs are proof that the temperature dependence of relative stabilities is important. A computational study of known polymorphs²³ suggests that entropy differences of up to 15 J/mol·K between polymorphs are possible, although almost two-thirds of polymorphic pairs differ by less than 4 J/mol·K—just over 1 kJ/mol at room temperature. Inclusion of vibrational contributions to the energy will be important in fine-tuning the energy ranking of predicted structures, but we have not calculated these contributions in this work.

2. Choice of Molecules

As we have mentioned, we chose a set of small, rigid organic molecules for our test set. Many practical applications of CSP will need to deal simultaneously with conformational flexibility and the search for close packed structures of the molecule. This increases the difficulty of the problem because all relevant conformations must be searched in the generation of the trial crystal structures, and the models for the inter- and intramolecular energies must be accurate and well balanced. Some of the difficulties in combining inter-

and intramolecular energy calculations have been highlighted by case studies of conformational polymorphism.^{24,25} Flexible molecule force fields are continually improving, but, as we see from the results of recent blind tests,^{3–5} the added difficulties dramatically lower the success rates of CSP for flexible molecules. With this in mind, we initially chose to study rigid molecules and we searched the Cambridge Structural Database for suitable small molecules, using the following criteria.

(i) Molecules containing only the elements C, H, N, and O and a range of common functional groups, molecular sizes, masses, and shapes.

(ii) We defined “small” as less than 25 atoms and only took molecules that we judged as reasonably rigid. Some of the molecules we chose do have a potentially flexible torsion angle or rotation, but it was necessary to include these to cover a range of functional groups (for example, amide substituted aromatic rings). We are not addressing the problems related to molecular flexibility in this work, so when choosing molecules with potentially flexible functional groups, we only included molecules whose calculated gas phase conformation was close to the conformation observed in the crystal, e.g., torsion angles within about 5°. For example, the calculated out-of-plane torsion angle of the amide group in benzamide is about 21°, which is reasonably close the 25–28° observed in the various determinations of its crystal structure.

(iii) The molecule must have at least one known crystal structure with all atomic positions accurately located and no disorder.

As a final criterion, we chose to limit ourselves to the nine most common space groups: $P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $P2_1$, $C2/c$, $Pbca$, $Pnma$, $Pna2_1$, and $Pbcn$; about 95% of homomolecular organic crystals are observed in these space groups.²⁶ Furthermore, we have only considered crystals with $Z' \leq 1$. While these restrictions would not be necessary when studying a particular molecule, they restrict the search space and allowed us to study a greater number of molecules. As searching the less common space groups is only a technical detail, the exclusion of molecules that crystallize with less common symmetries is a small sacrifice for allowing a larger sample of molecular types and, so, more general results and conclusions.

The set that we chose consists of 50 molecules (Table 1), ranging in molecular weights from 28 to 181 amu. We deliberately chose some molecules that are known to be polymorphic so that the total number of crystals is 62 (ignoring those in the less common space groups that we are not considering). The sampling of space groups within this set (58% $P2_1/c$, 11% $Pbca$, 10% $P\bar{1}$, 6% $Pna2_1$, 6% $P2_12_12_1$, 5% $P2_1$, and 3% $C2/c$) is reasonably representative of what is observed in the CSD.²⁶

3. Computational Methods

One of our criteria for the choice of molecules was that they were reasonably rigid, so that we could treat the molecular structure as fixed during crystal structure prediction. Crystal structure searches were performed using idealized molecular structures, generated from density functional theory calculations. We optimized the gas-phase structure using the program Dmol3,²⁷ with

Table 1. Test Set of Small Organic Molecules and Results of the Crystal Structure Prediction







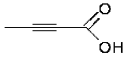

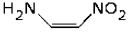












molecule	molecular formula	CSD Refcode	space group	diagram	N_{lower}^1	ΔE (kJ/mol) ¹
ethylene	C ₂ H ₄	ETHLEN01 (ETHLEN02)	$P2_1/c$ $(Im3m)^2$		2	+0.06
acetonitrile	C ₂ H ₃ N	QQQCIV01 (QQQCIV02)	$P2_1/c$ $(Cmc2_1)^2$		8 0	+0.55 -0.20
methylamine	CH ₃ N	METAMI	$Pcab$		3	+0.25
nitromethane	CH ₃ NO ₂	NTROMA	$P2_12_1$		4	+0.95
formamide	CH ₃ NO	FORMAM	$P2_1/c$		30	+1.55
acetic acid	C ₂ H ₄ O ₂	ACETAC	$Pna2_1$		3	+1.25
tetrolic acid	C ₄ H ₄ O ₂	TETROL TETROL01	$P\bar{1}$ $P2_1$		12 4	+1.40 +0.86
glycine	C ₂ H ₃ NO ₂	GLYCIN02 GLYCIN (GLYCIN01)	$P2_1/c$ $P2_1$ $(P3_1)^2$		2 98	+1.27 +5.19
(Z)-2-nitroethenamine	C ₂ H ₄ N ₂ O ₂	EDAWIP	$P2_1/c$		8	+1.83
2-azetidinone	C ₃ H ₅ NO	FEPNAP	$P\bar{1}$		0	-0.61
1,2,4-triazole	C ₂ H ₃ N ₃	TRAZOL	$Pbca$		3	+0.51
isoxazol-3-ol	C ₃ H ₃ NO ₂	NEZMUA	$P2_1/c$		7	+2.23
succinic anhydride	C ₄ H ₄ O ₃	SUCANH	$P2_12_1$		0	-3.18
succinimide	C ₄ H ₅ NO ₂	SUCCIN	$Pbca$		12	+2.54
oxazolidine-2,5-dione	C ₃ H ₃ NO ₃	OXAZDO	$C2/c$		11	+4.23
vinylene carbonate	C ₃ H ₂ O ₃	VINYLC	$P2_1/c$		0	-0.74
2-oxazolidinone	C ₃ H ₅ NO ₂	OXAZIL	$P2_1/c$		0	-1.64
5-methylene-oxazolidin-2-one	C ₄ H ₅ NO ₂	YOBQAH	$P2_1/c$		31	+1.84
3,5-pyrazolidinedione	C ₃ H ₄ N ₂ O ₂	DUNVEN	$P2_1/c$		66	+5.06
1,2,4-triazolidine-3,5-dione	C ₂ H ₃ N ₃ O ₂	KOXRIY	$P2_1/c$		45	+4.23
parabanic acid	C ₃ H ₂ N ₂ O ₃	PARBAC	$P2_1/c$		8	+2.44

Table 1 (Continued)





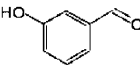
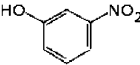
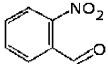
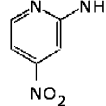
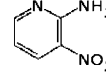
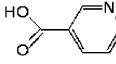
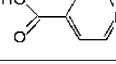
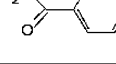
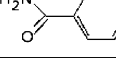
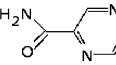
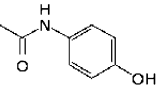
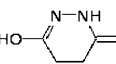
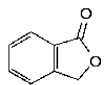
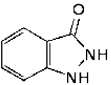
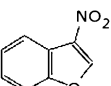
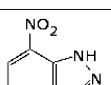
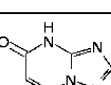
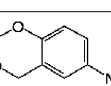
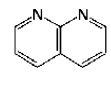
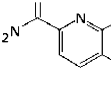
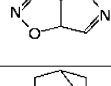
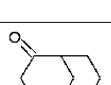
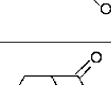
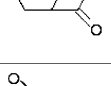
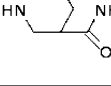
molecule	molecular formula	CSD Refcode	space group	diagram	N_{lower}^1	ΔE (kJ/mol) ¹
bicyclo(3.2.0)hept-1(5)ene	C ₇ H ₁₀	JIZREP	<i>P2₁/c</i>		10	+0.48
3,4-cyclobutylfuran	C ₆ H ₈ O	XULDUD	<i>Pbca</i>		0	-0.36
		XULDUD01	<i>P2₁/c</i>		24*	+2.48*
benzene	C ₆ H ₆	BENZEN	<i>Pbca</i>		2	+0.16
		BENZEN03	<i>P2₁/c</i>		0	-0.16 (-1.12 with PV ³)
pyrimidine	C ₄ H ₄ N ₂	PRMDIN	<i>Pna2₁</i>		5	+0.61
<i>m</i> -hydroxybenzaldehyde	C ₇ H ₆ O ₂	XAYCIJ	<i>Pna2₁</i>		Only including observed conformation	
					5	+0.61
					Including all 4 possible conformations	
					25	+2.24
<i>m</i> -nitrophenol	C ₆ H ₅ NO ₃	MNPHOL01	<i>P2₁,2₁,2₁</i>		3	+1.61
		MNPHOL02	<i>P2₁/c</i>		12	+2.70
<i>o</i> -nitrobenzaldehyde	C ₇ H ₅ NO ₃	NIBZAL	<i>P2₁</i>		10	+1.53
2-amino-4-nitropyridine	C ₅ H ₅ N ₃ O ₂	SEGRUR	<i>P2₁/c</i>		43	+7.30
2-amino-3-nitropyridine	C ₅ H ₅ N ₃ O ₂	AMNTPY	<i>P2₁/c</i>		0	-1.18
		AMNTPY01	<i>P2₁/c</i>		0	-0.67
		AMNTPY02	<i>P2₁/c</i>		- ⁴	- ⁴
nicotinic acid	C ₆ H ₅ NO ₂	NICOAC	<i>P2₁/c</i>		0	-0.14
isonicotinic acid	C ₆ H ₅ NO ₂	ISNICA	<i>P1̄</i>		6	+1.35
benzamide	C ₇ H ₇ NO	BZAMID	<i>P2₁/c</i>		3 *	+3.15 *
nicotinamide	C ₆ H ₆ N ₂ O	NICOAM	<i>P2₁/c</i>		0	-0.96
pyrazinamide	C ₅ H ₅ N ₃ O	PYRZIN	<i>P2₁/c</i>		24	+2.72
		PYRZIN01	<i>P2₁/c</i>		104 *	+5.88 *
		PYRZIN02 (PYRZIN05)	<i>P1̄</i> (<i>Pa</i>) ²		59	+4.19
paracetamol	C ₈ H ₉ NO ₂	HXACAN	<i>Pbca</i>		11	+3.33
		HXACAN01	<i>P2₁/c</i>		0	-1.32
1,2-dihydropyridazine-3,6-dione	C ₄ H ₄ N ₂ O ₂	MALEHY01	<i>P2₁/c</i>		2	+0.84
		MALEHY10	<i>P1̄</i>		2	+0.55
		MALEHY12	<i>P2₁/c</i>		19	+3.59

Table 1 (Continued)

molecule	molecular formula	CSD Refcode	space group	diagram	N_{lower}^1	ΔE (kJ/mol)
phthalide	$C_8H_6O_2$	HEZQUY	$P2_1/c$		2	+0.82
indazolinone	$C_7H_8N_2O$	FADMIG	$P2_1/c$		9	+2.22
3-nitrobenzofuran	$C_8H_5NO_3$	NIVBUP	$P\bar{1}$		0	-0.13
7-nitroindazole	$C_7H_5N_3O_2$	FULKUS	$P2_1/c$		0	-0.03
4,5-dihydro-5-oxo-(1,2,4)-triazolo(1,5-a)pyrimidine	$C_5H_4N_4O$	QAJYIJ	$Pna2_1$		0	-0.01
6-nitro-1,3-benzodioxin	$C_8H_7NO_4$	GEYWIQ	$Pbca$		1	+0.73
1,8-naphthyridine	$C_8H_6N_2$	NAPTYR	$P2_1/c$		0	-0.93
quinoline-2-carboxamide	$C_{10}H_8N_2O$	QUINCB10	$P2_1/c$		6	+2.15
3a,6a-dihydro-isoxazolo(5,4)isoxazole	$C_4H_4N_2O_2$	DIHIXL10	$P2_1/c$		2	+0.68
norbornene	C_7H_{10}	HOBBOB	$P2_1/c$		4	+0.58
bicyclo(3.3.1)nonane-2,6-dione	$C_9H_{12}O_2$	HEBBEV	$C2/c$		0	-4.39
3-aza-bicyclo(3.3.1)nonane-2,4-dione	$C_8H_{11}NO_2$	BOQQUT	$P2_1/c$		6	+3.37
3-diazabicyclo(3.3.1)nonane-2,6-dione	$C_7H_{10}N_2O_2$	DOGTIC	$P2_12_12_1$		0	-0.30

¹ ΔE is the calculated energy difference between the experimentally observed crystal structure and the lowest energy unobserved crystal structure. N_{lower} is the number of predicted, but unobserved, crystal structures that are lower in energy than the experimentally observed crystal structure. These measures are described in section 4.3. ² Polymorphs in space groups that were not considered in this work are listed in parentheses. ³ For the high-pressure polymorph of benzene, we added the PV energy term to the calculated energies of all the predicted structures to determine its position in the ranked list of hypothetical structures. ⁴ AMNTPY02 relaxed to the same potential energy minimum as AMNTPY01 (see text). *Indicates a crystal that was not located in the first four repeats of the simulated annealing search and the search was repeated until the structure was found.

the PW91 functional²⁸ and the double numerical polarized (DNP) basis set.²⁹ Several of the molecules in our list have multiple possible "rigid" conformations. For example, in *m*-nitrophenol, the proton of the alcohol

group could point away from (as is observed in the known polymorphs) or toward the nitro group; both planar geometries are stable in the gas phase and separated by a ~ 5 kJ/mol energy barrier. For the 11

molecules where there are such alternate conformations, we optimized geometries for each conformer (see section 4.4).

3.1 Model Potential. Besides the molecular structures, the other main ingredient in crystal structure prediction is a model for the interaction energies between molecules. Much of the variability of results in previous prediction studies results from differences in the choice of model potential.^{3–5} As this choice is crucial to the success of a study, we tested several models on a subset of 15 crystals from our test set.

There are many choices for the model potential, not only in the functional form used, but in the approach used to parametrize the function. As already mentioned, we restricted ourselves to simple functional forms of intermolecular potential in the present study, leaving improvements requiring more specialized software and increased computational expense to later investigations. We chose a set of model potentials that we consider suitable for the type of molecules that we are studying here.

Two model potentials derived specifically for rigid molecules were tested, both using the *exp-6* (Buckingham) functional form for nonbonded interactions.

(a) The FIT model uses the C, N, O, and nonpolar H parameters from Cox and Williams' work.^{30,31} The parameters were fitted to reproduce the structures and lattice energies of a large set of non-hydrogen-bonded hydrocarbons, oxohydrocarbons and azahydrocarbons. To complete the FIT model, parameters for polar hydrogen atoms, H(–N) and H(–O), were later fitted by Price and co-workers to sets of hydrogen-bonded organic molecular crystals, in conjunction with an accurate electrostatic model.^{9,20}

(b) The W99 model^{32–34} is a more recent set of parameters, derived in much the same way as the FIT model, but allowing the parameters for each atom type to depend on the bonding environment of the atom. For example, the parameters describing aromatic and aliphatic carbon were varied separately in the fitting procedure. Another detail of the W99 model potential is that the center of the interaction for all hydrogen atoms is shifted 0.1 Å along the X–H bond (X = C, N, O), toward the heavy atom. Thus, for all of our calculations with the W99 model, we foreshortened all DFT optimized X–H bond lengths by 0.1 Å.

We also tested the use of the nonbonded parameters from three flexible molecule force fields: Dreiding, CVFF95 and COMPASS.

(c) Dreiding³⁵ is the oldest of these three and one of the most commonly used force fields for molecular simulation. We tested the Dreiding-X variation of the potential, which uses the same *exp-6* functional form as FIT and W99. A specific 12–10 model for hydrogen bonding is included within the Dreiding model, although the use of an explicit expression for hydrogen bonding is probably redundant when the repulsion, dispersion, and electrostatic models are sufficiently accurate. Thus, we tested the Dreiding model with and without this hydrogen bond term.

(d) CVFF95^{36–39} force field was developed specifically for materials science applications and the nonbonded parameters were parametrized to the crystal structures

of small organic molecules, employing a 12–6 (Lennard-Jones) form.

(e) The final model that we tested is COMPASS,⁴⁰ a force field for condensed phase simulations. The non-bonded parameters of the 9–6 model were parametrized to structures and properties of the condensed phase. In contrast to the normal approach of parametrization by comparing lattice energy minimized models to experiment, molecular dynamics simulations were used to explicitly include temperature when fitting the model to experiment.

Both CVFF95 and COMPASS can supply their own atomic charges, based on bond increments. We tested the models with their own charges as well as atomic charges specifically fitted for each molecule. The charges were derived by fitting to the molecular electrostatic potential (ESP) of the DFT calculated charge density, calculated within the Dmol3 program.²⁷ The grid of points used in the fitting was taken from the internal and external atomic radii suggested in Dmol3 and smoothly varying weighted fitting points across border layers at the internal and external edges of the area around each atom.²⁷ These ESP atomic charges were used as the electrostatic model for the FIT, W99, and Dreiding calculations.

To test the models, we took the known crystal structures from the Cambridge Structural Database and replaced the molecular structure with the DFT optimized geometry. The crystal was then energy minimized using the Cerius2 software, keeping the DFT geometry fixed. We measured the root-mean-squared change in the lattice parameters *a*, *b*, and *c* and used the overall root-mean-squared change in the lattice parameters to judge the performance of the potential (Table 2).

The ESP atomic charges are a clear improvement over those calculated from transferable bond increments and the cost of calculating the molecular electrostatic potential for these small molecules is minimal. The best performance was with the W99 model, which was developed specifically for these types of molecules, while COMPASS+ESP followed closely behind. We proceeded with the W99+ESP model potential for all of the crystal structure searches presented here. We only used structural comparisons to guide our choice of model potential in this work, so are not making definite conclusions regarding the accuracy of these models. Sublimation enthalpies have been determined for a few of the crystals in our list and these are given, along with the W99+ESP calculated lattice energies, in the Supporting Information. There is good agreement for some of the crystals, while the lattice energy underestimates the sublimation enthalpy for several of the hydrogen-bonded crystals. Previous studies^{41,42} suggest that an improved electrostatic model would improve the agreement between lattice energies and sublimation enthalpies.

3.2 Crystal Structure Search. There are numerous methods of generating the possible crystal packings of a molecule,¹⁶ as compared in blind tests of crystal structure prediction.^{3–5} Some of these methods build up crystals from stable molecular clusters,^{43–46} while others search phase space systematically,^{6,7,47,48} by Monte Carlo,^{49–53} Molecular Dynamics,⁵⁴ or genetic algorithms^{55–57} or even completely randomly.^{58,59} Some of these methods may be more efficient or more effective

Table 2. Root Mean Squared % Changes in Lattice Parameters on Energy Minimization for the Model Potentials Tested

crystal Refcode	FIT	W99	Dreiding		CVFF95		COMPASS	
			with H-bond	without H-bond	bond increment charges	ESP charges	bond increment charges	ESP charges
ETHLEN	6.55	2.64	3.07	3.07	12.72	2.48	11.03	3.21
NTROMA	4.81	2.09	2.75	2.75	5.73	4.60	2.03	0.89
TRAZOL	6.88	4.68	6.87	6.22	17.09	6.78	21.17	5.67
NEZMUA	5.64	3.80	3.03	2.50	3.37	2.19	4.19	2.22
OXAZIL	2.23	3.12	3.90	1.15	1.98	1.94	2.90	3.25
SEGRUR	2.35	2.55	2.00	1.93	5.59	2.11	6.11	3.53
α PYRZIN	2.34	1.09	2.60	2.42	4.95	3.83	2.70	3.53
β PYRZIN	6.86	2.69	4.38	4.79	2.41	2.99	0.59	2.84
γ PYRZIN	5.23	2.80	2.61	2.89	0.40	0.92	1.72	1.73
δ PYRZIN	4.92	2.79	3.36	4.13	6.26	1.85	6.67	1.98
HEZQYU	2.75	1.42	1.39	1.39	1.85	1.28	1.79	1.77
QAJYIJ	6.57	3.50	2.83	2.83	17.72	4.23	12.52	4.31
GEYWIQ	2.25	1.38	1.41	1.41	2.43	1.80	7.78	0.92
QUINCB	2.45	0.97	1.56	2.27	2.31	2.72	1.96	2.22
HEBBEV	1.58	0.91	0.60	0.60	2.65	1.35	1.41	2.30
overall	4.65	2.66	3.18	3.04	7.91	3.12	7.83	2.96

for different types of molecules, but our aim is not to compare search methods. Instead, we are interested in the ranking of structures, given a list generated by one of these types of searches. We chose the simulated annealing algorithm of Karfunkel and Gdanitz,^{16,51–53} as implemented in the Accelrys Polymorph Predictor module of the Cerius2 software suite.⁶⁰ This search method is well tested and seems to be effective for a range of types of molecules.

The search consists of several steps and is repeated separately for each space group.

(i) The first stage is the simulated annealing, during which crystal structures are built and randomly modified. During a heating phase, high energy barriers can be crossed and so the whole of the potential energy surface is sampled. This is followed by a cooling phase, during which the probability of crossing energy barriers decreases.

(ii) After the simulated annealing, the resulting structures are clustered to eliminate duplicates.

(iii) The hypothetical structures are energy minimized.

(iv) The structures are clustered again to remove duplicates that converged to the same minimum during energy minimization.

There are several parameters that can be adjusted during this routine. The simulated annealing is governed by the heating rate (T_h), cooling rate (T_c), and two parameters to signal the end of the heating phase: a maximum temperature (T_{max}) and a limit on the number of consecutive accepted structures, N_{accept} . The heating is ended when either the temperature reaches T_{max} or N_{accept} consecutive random moves are accepted. The clustering is controlled by a tolerance on the measure of similarity. The lower this number, the closer structures must be to be considered identical, so high tolerances risk throwing away distinct crystals. During the first clustering, a maximum number of structures, N_{max} , are chosen for energy minimization. Although it is possible that a high energy trial structure can relax to be one of the most stable crystals, the number of energy minimizations can be limited, by discarding any structures above N_{max} in the energy ranked list following the simulated annealing. The settings for the parameters are a balance of computing time and con-

Table 3. Simulated Annealing and Clustering Parameters

simulation parameter	value used
heating rate, T_h	0.025
cooling rate, T_c	0.001
maximum temperature, T_{max}	100000 K
maximum number of consecutive accepted structures, N_{accept}	12
clustering tolerance	0.10
structures kept after clustering	10000

fidence that the search generates a complete set of possible low-energy hypothetical crystals.

We took most of our settings from the default values in the highest performance (and most computationally intensive) settings provided by the Cerius2 software. The heating and cooling rates as well as maximum temperature and number of consecutive accepted moves in the heating were left unchanged (Table 3). We have developed our own algorithm for the clustering of crystal structures, so used the Cerius2 clustering with very tight settings. The tolerance for comparing structures was decreased to 0.10 from the default value of 0.13, which resulted in very few structures being discarded. We tested the performance of the search on a subset of our molecules and found that, with this clustering tolerance, the default number of structures to energy minimize (1000 per space group), was too low, missing about 5% of structures within 5 kJ/mol of the global minimum in the final list. While doubling N_{max} to 2000 lowered the number of missed structures to about 1% of those in the lowest 5 kJ/mol, we decided to include all distinct simulated annealing structures in the energy minimization by increasing N_{max} to 10000.

The simulation was repeated several times for a subset of our molecules, combining final lists after each repeat of the simulated annealing. We kept track of the number of new structures in the low energy list after each repeat and found that the search was converged after three repeats for this subset of molecules. To be more confident of adequately searching phase space, we added an extra repeat and performed four searches for each molecule in our set. On average, one repeat of the search (simulated annealing, clustering, and energy minimization) for the nine space groups took about 6–8 hours on our workstation (one 600 MHz processor of an SGI Octane), so we used about 1 day of computing time per molecule with the final settings.

For those molecules with multiple rigid conformations, the search was repeated for any conformations with molecular energy within 10 kJ/mol of the most stable. The low energy lists of crystals structures for each conformation were then combined, using the sum of molecular (gas-phase DFT energy) and lattice (from the model potential) energies as a measure of their relative stabilities.

3.3 Symmetry Constraints and Unstable Structures. Energy minimizers generally do not guarantee that a minimum in the energy is found, but stop at any stationary point (i.e., the forces are all zero to within some tolerance). Such points on the potential energy surface could be minima or saddle points, which are a minimum in energy in at least one direction, but a maximum along at least one other path. Such structures are especially likely to be found when using space group symmetry in the energy minimizations, as is done in the energy minimizations within the Polymorph Predictor package. Therefore, there may be many such cases in the final lists we generate with Polymorph Predictor. These saddle point structures are not observable crystals, so must be removed from the lists of hypothetical structures or pushed downhill toward the nearest minimum on the energy surface. This task is automated in another crystal structure modeling program, DMAREL,^{19,61} which can calculate elastic constants and phonon frequencies for rigid molecules. Either an unstable elastic constant matrix or a negative phonon frequency indicates that the structure is at a saddle point. In either case, DMAREL performs a line search in the direction of decreasing energy and resumes the energy minimization at the lowest point along this path.

The final structures within 15 kJ/mol of the global minimum from the Polymorph Predictor calculations were reminimized using DMAREL, initially using space group symmetry. Their stabilities (phonon frequencies and elastic constants) were then checked and structures were pushed to a minimum where required. In some cases, the original space group symmetry is conserved, but some structures required at least one symmetry constraint to be removed to reach a stable minimum. In these cases, all symmetry was removed from the structure and we identified the remaining symmetry after energy minimization. In some cases, this leads to space groups that were not included in the initial searches and even to structures with more than one molecule in the asymmetric unit. As yet, only $\mathbf{k} = 0$ phonon calculations are implemented in DMAREL, so it is still possible that some structures at the end of this procedure are unstable to unit cell doubling.

For comparison to the lists of predicted crystals, the known crystal structures for each molecule were taken from the CSD and lattice energy minimized, with the molecular structure exchanged for the idealized gas-phase geometry. A table of unit cell parameters for the lattice energy minimized structures is available as supplementary information.

3.4 Clustering. The list of optimized predicted structures often contains several duplicate structures, and it is helpful if these duplicates are removed to produce a list of unique structures. To cluster our list, we have developed an algorithm, COMPACT,⁶² to identify whether two crystal structures are the same

to within specified tolerances. The algorithm avoids the use of space group symmetry but rather defines a crystal structure in terms of its molecular packing environment. A molecule together with its coordination shell, say the closest 14 closest molecules, is chosen to represent the reference crystal structure. Relationships between these molecules are described by distance and angle constraints which can then be searched for in a target structure. Tolerances are placed on distances and angles to control the level to which relative positions and orientations of the molecules must match. This method was also used to locate the position of the observed structure within the predicted list.

4. Results and Discussion

4.1 Are All of the Experimental Structures Generated? Of the 62 crystals in the space groups that we searched, all but four were located in the final lists: benzamide (BZAMID), the monoclinic polymorph of 3,4-cyclobutylfuran (3-oxabicyclo(3.2.0)hepta-1,4-diene) (XULDUD01), the β polymorph of pyrazinamide (PYRZIN01), and form III of 2-amino-3-nitropyridine (AMNTPY02). The result is pleasing; if we were studying a particular molecule on its own, more than four repeats of the simulated annealing would be run to be more confident in having a complete search of phase space. Furthermore, an experienced user of the software could perhaps adjust other settings (e.g., the heating and cooling rates) to further optimize the efficiency of the search. All of the missed structures are in the space group $P2_1/c$. As this is the most common space group for molecular organic crystals, it is probably a good idea for future prediction studies to search this symmetry more intensely than others. For three of these four missed crystals, we added additional repeats of the simulated annealing in $P2_1/c$ until the known structure was found. For XULDUD01, the known structure was located in the fifth search, BZAMID in the 12th, and PYRZIN01 in the 13th. The simulated annealing algorithm clearly has difficulties with the latter two crystals; they may inhabit narrow potential energy wells in otherwise inaccessible areas of phase space. We have not compared search methods in this work, but if the cause of the difficulty in finding these structures is the narrowness of the potential energy well, then other search methods would probably have similar problems in finding these crystal structures.

The four missed structures remind us that we have not produced complete sets of crystal structures for our molecules; there may be other low-energy structures missing. We know this in advance, because of our limiting of the space groups considered. Furthermore, few of the molecules have been subjected to polymorph screens, so it is likely that undiscovered structures exist. More extensive calculations could show up more unrealized low-energy structures, while more extensive experiments could discover more real polymorphs. We expect, however, that the results of either would have a small influence on the overall results of our study.

The lack of success with the remaining crystal (2-amino-3-nitropyridine, AMNTPY02) is due to shortcomings of lattice energy minimization with a simple model potential. When relaxing the experimental crystal on the W99+point charge potential energy surface, the

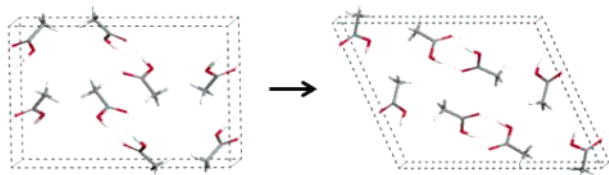


Figure 1. Relaxing one of the acetic acid saddle point (unstable) crystals.

stepped sheet structure relaxes to the energy minimum of one of the other polymorphs (AMNTPY01), with planar hydrogen-bonded molecular sheets. The model fails to reproduce the energy barrier between the two forms, so we can only expect to be able to predict the polymorph with the stable minimum. A similar observation was made with a similar trimorphic molecule, 2-amino-5-nitropyrimidine,⁶³ where a polymorph with buckled hydrogen-bonded sheets collapses to another known polymorph on energy minimization, even with an elaborate model of the electrostatic interactions. Such crystals, with closely related polymorphs, are a challenge for computer modeling. AMNTPY02 is excluded from the results reported in the rest of this study.

Although we only searched nine space groups, the structures do have the opportunity to adopt higher symmetry space groups during energy minimization, or lower symmetry space groups during the stability checking and reminimizations (section 3.3). As a result, we did find one known crystal structure that is not in our list of space groups, the $Cmc2_1$ low-temperature phase of acetonitrile was located in searches of both $P2_1$ and $Pna2_1$ and is the global minimum of the lattice energy search for this molecule (Table 1). Therefore, in total (excluding AMNTPY02 and including the $Cmc2_1$ polymorph of acetonitrile), we consider the prediction of 62 crystal structures in the remainder of this study.

4.2 Reranking of Structures after Reminimization. The use of DMAREL and the calculation of elastic constants and phonon frequencies to test the stability of the predicted structures proved to be very important. Of the 50 molecules, the stability testing and reminimizations introduced new low-energy crystal structures in most cases. On average, 2 of the 10 lowest energy structures for each molecule resulted from this process and would not have been located had we not performed these extra calculations beyond the initial energy minimizations. The “top 10” list of structures of only 6 of the 50 molecules was unaffected by these calculations.

As an example, four of the 10 lowest energy crystal structures of acetic acid were found during this process and would have been missed if the search had been stopped after the first energy minimizations immediately following the simulated annealing search. One of these structures is shown in Figure 1. On the left is structure 80 (counting upward from the global minimum) of the list from Polymorph Predictor, an orthorhombic ($Pbcn$) crystal 4.48 kJ/mol above the global minimum. The stability calculations showed that this structure is unstable to breaking the symmetry constraints of the mirror plane that keeps the methyl groups in close contact. The symmetry was relieved and the lattice parameters shifted during energy minimization to allow the methyl groups to reorient to the symmetry relaxed structure on the right of Figure 1, a

triclinic ($P\bar{1}, Z' = 4$) crystal 2.56 kJ/mol lower in energy than its orthorhombic “parent” and seventh in the final list.

4.3 Location of the Observed Structures in the Ranked Lists. The locations of the observed crystal structures in the lists of predicted crystals (Table 1) are described by their energy relative to the most stable unobserved crystal in the list (ΔE) and the absolute position of the observed crystals in the list, counting upward from the global minimum. Instead of presenting the position of the observed crystal by its absolute rank, we report the number of lower energy hypothetical, but unobserved, crystal structures, N_{lower} . In this way, a “perfect” prediction yields a negative ΔE and $N_{\text{lower}} = 0$. Furthermore, all the known crystal forms for polymorphic systems can reach such a ranking simultaneously. The three structures that were not found in the original search, but were located after an increased number of simulated annealing repeats, are included in Table 1 and are indicated with an asterisk.

A positive relative energy tells how much the experimentally observed crystal is apparently less stable than the theoretical global minimum structure, while a negative value indicates that there are no lower energy unobserved structures in the list. We illustrate this with two contrasting results, using plots of the lattice energy against density for the lowest energy structures to summarize the results of the CSP for these two molecules (Figure 2). The upper plot shows the results for bicyclo(3.3.1)nonane-2,6-dione. For this rather bulky molecule, the lowest energy structure corresponds to the known crystal (CSD refcode HEBBEV) and there are no competing structures within 4.4 kJ/mol. For this crystal, ΔE is -4.4 kJ/mol and $N_{\text{lower}} = 0$; this crystal is easily predicted and polymorphs seem unlikely.

The lower plot in Figure 2 summarizes results for the much smaller 3,5-pyrazolidinedione. Here, there is no large gap in energy separating any of the lowest energy structures and, with the model potential used here, the known crystal (refcode DUNVEN) is 67th in the list ($N_{\text{lower}} = 66$), 5.1 kJ/mol above the global minimum. With such small energy differences between structures, this crystal is a much greater challenge for prediction by lattice energy minimization. Furthermore, if the energy calculations are correct, and entropy differences between structures are small, the known crystal is metastable to many more densely packed potential polymorphs.

Of the 62 known crystals, 18 are predicted with lower energy than any of the unobserved structures; 17 are global minima and the two polymorphs of 2-amino-3-nitropyridine rank 1 and 2 in energy. This represents a success rate of 29% for such a “perfect” prediction. Of these, most (12) are within 1 kJ/mol of the nearest unobserved crystal and a further three are within 2 kJ/mol; results as in Figure 2a, where the known crystal is far more stable than any others, are quite rare. The worst ranked crystal is the β form of pyrazinamide (refcode PYRZIN01); there are 105 lower energy hypothetical crystals in the list. On relative energy, all of the structures are within 7.3 kJ/mol of the global minimum. The known crystal of 2-amino-4-nitropyridine (CSD refcode SEGRUR) has this highest ΔE . Considering all of the predicted structures as potential poly-

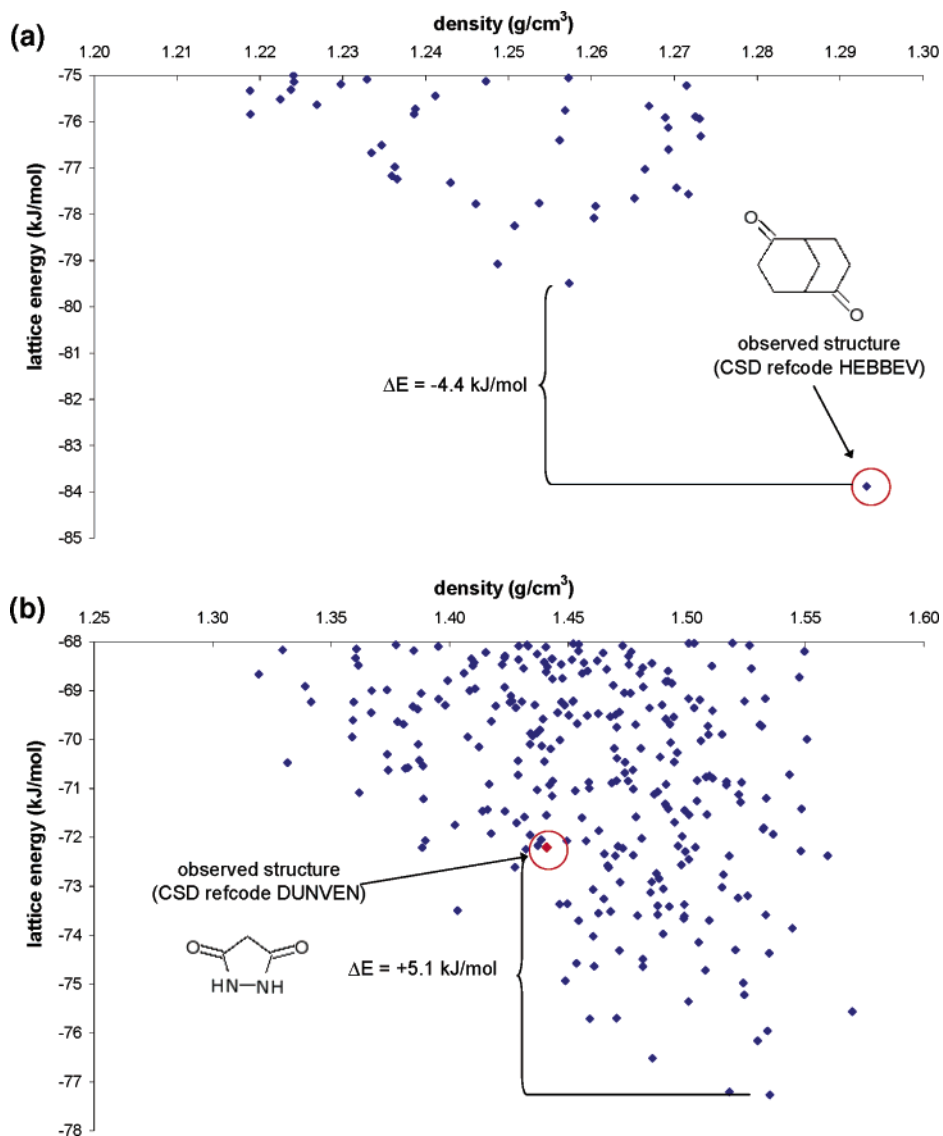


Figure 2. Predicted structures of (a) bicyclo(3.3.1)nonane-2,6-dione and (b) 3,5-pyrazolidinedione.

morphs, these results are consistent with estimates that lattice energy differences between polymorphs are usually smaller than 10 kJ/mol.^{64,65}

Our overall results are summarized in Figure 3a,b, where the predictions are binned into intervals of 5 for N_{lower} and 0.5 kJ/mol for ΔE . The distribution of relative energies is better than we expected, with 35% of structures either lower than any others or within 0.5 kJ/mol of the global minimum; 53% are within 1 kJ/mol. 56% of the observed crystals are separated by fewer than five hypothetical structures from the global minimum and 74% have $N_{\text{lower}} \leq 10$. These distributions could be used to guide future prediction studies. For example, if we wanted to construct a list of possibilities for the crystal structure of a new molecule, with 90% confidence that the real structure will occur in this list, all structures within a cutoff of 4 kJ/mol above the global minimum should be included as possibilities.

The distributions tail off fairly quickly at higher N_{lower} or relative energy and the results are encouraging for the crystal structure prediction of small, rigid molecules. The rate at which these distributions fall away is a combination of two factors. The modeling of energies is fairly crude; lattice energies ignore temperature and

only approximate the energies at $T = 0$ K. Furthermore, the simple form of model potential that we have used has known limitations, so the errors in lattice energies are probably as great as the relative energies that we are calculating for many of the molecules. The second factor is that metastable polymorphs do exist and it is possible that more stable polymorphs are experimentally realizable for some of the molecules in our set. Hence, the relative energies that we calculate for some of the higher ranked crystals could be a realistic measure of their metastability. To separate the effects of modeling deficiencies and our incomplete knowledge of the true number and stability of crystal forms, we could perform extensive polymorph screens on all 50 of the molecules. This is a mammoth task for this number of molecules. Alternatively, by systematically improving the theoretical basis of our energy ranking, we should get to a point where we can confidently estimate the proportion of crystals in our sample that are metastable.

4.4 Assessing the Importance of Alternate Conformations. For all but one of the molecules where alternate conformations exist, the conformation in the known crystal is the most stable in the gas phase (Table 4). Furthermore, for several of the molecules, the

Table 4. Relative Energies of Alternate Molecular Conformations

molecule	conformational difference from conformation in known crystal	molecular energy relative to conformation found in known crystal (kJ/mol)
<i>m</i> -hydroxybenzaldehyde	180° rotation of hydroxyl	+0.33
	180° rotation of aldehyde	+0.37
	rotation of hydroxyl and aldehyde	-3.27
nicotinic acid	180° rotation of COOH	+1.01
nicotinamide	180° rotation of amide	+3.40
paracetamol	180° rotation of hydroxyl	+1.42
<i>m</i> -nitrophenol	180° rotation of hydroxyl	+2.02
<i>o</i> -nitrobenzaldehyde	180° rotation of aldehyde	+9.52
isoxazol-3-ol	180° rotation of hydroxyl	+17
1,2-dihydropyridazine-3,6-dione	180° rotation of hydroxyl	+21.0
pyrazinamide	180° rotation of amide	+39.5
quinoline-2-carboxamide	180° rotation of amide	+45

conformational energy difference is too great to be balanced by improved crystal packing. Where the mo-

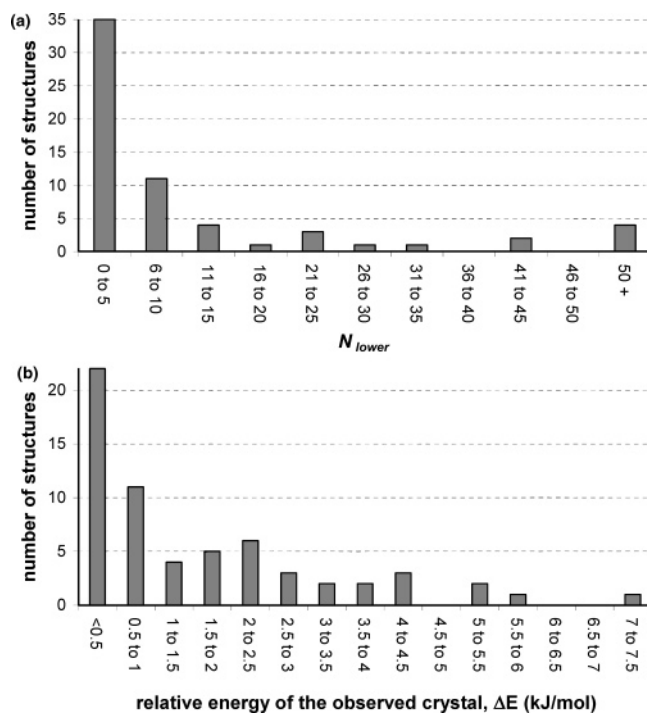


Figure 3. Distributions of the known crystals in our predicted lists.

lecular energy difference is only a few kJ/mol, these alternate conformations can give competitive low-energy crystals, so we performed the prediction for any conformations more stable than, or within 10 kJ/mol of the conformation in the known crystal. This meant extra searches for six of our molecules, each in one extra conformation, except *m*-hydroxybenzaldehyde, where three extra conformations had to be searched. For all of these molecules, the atomic charges were fitted separately to the DFT calculated charge distribution for each conformation.

For five of these molecules (nicotinic acid, nicotinamide, *m*-nitrophenol, *o*-nitrobenzaldehyde, and paracetamol), the global minimum lattice energy in the search with the alternate conformation is less stable than with the observed conformation. Hence, the alternate conformation is disfavored on both molecular and packing energies. On the sum of lattice energy and gas-phase monomer energy, the observed crystal is more stable than any of those found for the alternate conformation,

so the N_{lower} and ΔE based on the known conformation are unaffected.

The ranking of the known crystal was only affected for *m*-hydroxybenzaldehyde. For this molecule, one of the molecular conformations, with the hydroxyl and aldehyde substituents rotated 180 degrees from their orientation in the observed conformation, is predicted to be significantly more stable than the observed conformer. On lattice energy, the lowest energy crystal predicted for this conformation is 1.6 kJ/mol less stable than global minimum with the observed conformation, but is more stable on total (lattice + conformational) energy. In fact, 18 predicted crystals with this molecular conformation are more stable on total energy than the known form (refcode XAYCIJ). Of the two conformations that are less stable in the gas phase, one packs poorly, with worse lattice energies than the known crystal, while the other (with just the aldehyde rotated with respect to the known conformation) packs with better lattice energies, which approximately balance the higher molecular energy. On total energy, two crystals of this conformation are lower in energy than the known form. In all, by taking account of these additional conformations, there are an additional 20 calculated structures lower in energy than the observed crystal and the relative energy, ΔE , is increased by 1.63 kJ/mol. This ranking and relative energy were used in Figure 3a,b.

5. Discussion

5.1 Which Crystals Are Predictable? In addition to the overall statistics on the performance of lattice energy minimization for crystal structure prediction, we hoped that our results would distinguish between types of molecules that are more or less easily predicted than others.

The search procedure used here has been very successful at locating the experimentally observed crystal structures, in all but a few cases—XULDUD01, BZAMID, and PYRZIN01. It is difficult to say what the reasons are for this, although we can speculate that the potential energy well is narrow and difficult to access via the simulated annealing search. That two of these are aromatic amides suggests that such molecules may be problematic for the search routine, although the other two polymorphs of pyrazinamide and the crystal structure of nicotinamide were located without problems.

Besides these infrequent difficulties with the searching of phase space, the success of crystal structure prediction by lattice energy minimization depends on

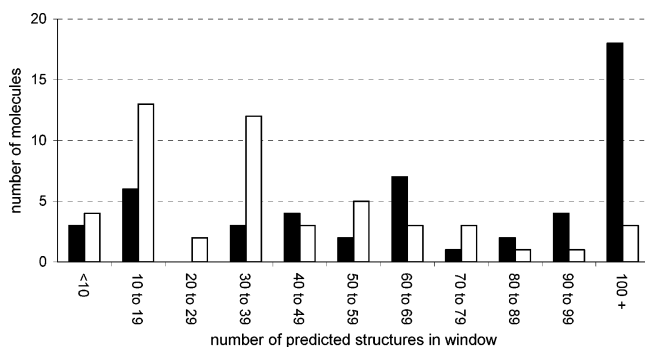


Figure 4. Density of structures in the lowest 5 kJ/mol (filled bars) and lowest 5% in lattice energy (unfilled bars).

the number of predicted low-energy structures and the computation of the relative energies of the predicted crystal structures. We tried correlating several molecular descriptors with the position of crystals in the predicted lists. No easily quantifiable descriptor, such as molecular mass or dipole moment, shows any correlation with how well that molecule's known crystal is predicted. It seems reasonable to suggest that particularly bulky molecules with very asymmetric shapes should have less possible ways to closely pack, so should be more easily predictable. This is not an easily quantified molecular property, but can be judged by eye. We deliberately included four similarly shaped molecules in our set, chosen for their obvious bulkiness. The four crystals (refcodes HOBBOB, HEBBEV, BOQQUT, and DOGTIC) all rank quite well ($N_{\text{lower}} = 4, 0, 6$ and 0 , respectively). Although less so, 3a,6a-dihydro-isoxazolo-(5,4)isoxazole (refcode DIHIXL) also has a distinctive bent shape and its $N_{\text{lower}} = 2$.

The density of structures in the lowest energy part of the list is a measure of the difficulty in predicting the correct structure. A large number of structures in a small energy range means small energy differences and, so, requires much more accurate energy calculations to predict the correct order of stability. We counted the number of structures within 5 kJ/mol of the global minimum for all of our molecules and the distribution (filled bars in Figure 4) shows that about a third of molecules have more than 100 distinct structures in this small energy window. The median of this distribution is about 67. Considering these densely packed lists, the low overall values of N_{lower} are a testament to the usefulness of the empirical atom–atom model potential.

We noticed that the smaller molecules with lower lattice energies generally have more molecules in the 5 kJ/mol energy window, so we also present a distribution that is less dependent on the size of the molecule and magnitude of the lattice energy—the number of structures within 5% of the global minimum (white bars in Figure 4). More than half of the molecules have between 10 and 40 structures in this relative energy range and the median of the distribution is 33. Of the five bulky molecules mentioned earlier, all but norbornene (refcode HOBBOB) have fewer than 20 structures within 5% of the global minimum. The distinctive shape of these molecules may be a reason for their having few low-energy close packed possibilities, helping their overall predictability.

5.2 Consequences of Hydrogen Bonding on Predicting Crystal Structures. Apart from global mo-

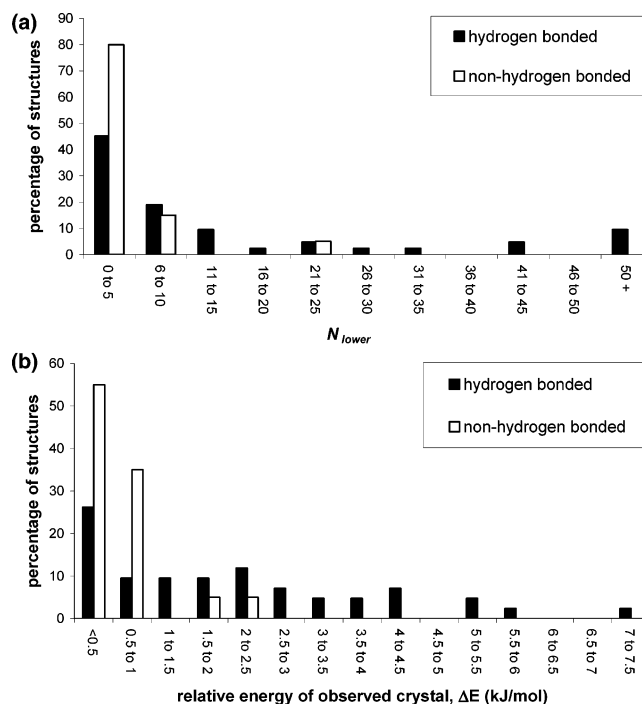


Figure 5. Distributions of (a) N_{lower} and (b) ΔE for hydrogen bonding (filled bars) and non-hydrogen bonding (unfilled bars) molecules.

lecular characteristics, local attributes may be important in determining the predictability of a molecule's crystal structure. Probably the most important such attribute is the presence or absence of hydrogen-bond donors and acceptors. Molecules with the ability to hydrogen bond behave very differently from those whose crystals have no such strong, directional interactions. To examine differences between hydrogen-bonding and non-hydrogen-bonding molecules, the distributions of N_{lower} and ΔE were split into these two categories. For this classification, the presence or absence of weak C–H...O interactions was ignored. Because of the different numbers of each type of molecule in our test set, the distributions are presented as percentages instead of the absolute number of structures in each bin (Figure 5).

Differences between the two classes of molecules are striking. For 80% of non-hydrogen-bonded crystals, there are five or fewer theoretical structures lower in energy than the observed crystal structure and there are never more than 25 lower in energy. In contrast, less than half of hydrogen-bonded crystals have N_{lower} between 0 and 5 and all of the poorly ranked crystals are heavily hydrogen bonded. Over half of the non-hydrogen-bonded crystals are found as either global minima or within 0.5 kJ/mol of the global minimum and 90% are within 1 kJ/mol (Figure 5b). Again, the distribution is much flatter for the hydrogen-bonding molecules; we must consider predicted structures up to 4.5 kJ/mol above the global minimum to include 90% of the observed structures.

The differences in results for the two classes of molecules are likely to be due partly to the increased difficulty in modeling the energies of hydrogen-bonded crystals. Such strong interactions involve very close contacts with hydrogen atoms, so the calculations are especially sensitive to slight changes in molecular

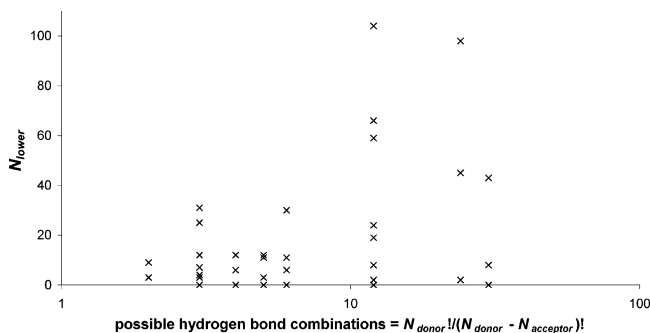


Figure 6. Ranking of the observed hydrogen bonding crystals versus the number of hydrogen bond acceptor–donor combinations.

geometry. The dominant contribution of electrostatics to these interactions also means that the simple atomic point charge model used here limits the quality of calculated energies far more for crystals with hydrogen bonding. Such a simple representation of the charge distribution cannot describe important features such as lone pairs, which dictate the directionality of hydrogen bonds and relative energies of different hydrogen-bond motifs. Thus, the errors in relative lattice energies are greater for molecules with hydrogen-bonding capability. As a result, molecules with a plurality of competing hydrogen-bond motifs are a greater challenge for CSP. For each molecule, a measure of the number of competing hydrogen bond options can be roughly quantified as the number of possible hydrogen-bond acceptors/donors pairwise combinations ($N_{\text{acceptor}}!/(N_{\text{acceptor}} - N_{\text{donor}}!)$). While this measure ignores differences in extended motifs, our results strongly suggest that molecules with more hydrogen-bonding possibilities have a higher likelihood of a poor lattice energy ranking (Figure 6).

While much of the difficulty with hydrogen-bonding molecules results from limitations of the model for the interaction energies, the distributions in Figures 5 and 6 could reveal a reality of crystallization. It seems sensible that strong, directional interactions lead to higher barriers between valleys on the potential energy surface. Such higher energy barriers facilitate the trapping of a crystal in a local energy minimum, possibly leading to the growth of metastable polymorphs. This is a likely contributor to our results, but we must reduce the errors in our modeling methods to confidently describe real differences between classes of molecules—higher quality energy calculations are desirable and are the subject of current investigations.

6. Conclusions

Lattice energy minimization, the most widely used method for the prediction of crystal structures, gives promising results for rigid organic molecules. This study of a diverse set of 50 molecules shows the success rate that can be achieved when using a carefully parametrized, but simple form of model potential—about half of all molecules can be found among the five lowest energy predicted structures, or within 1 kJ/mol of the global minimum in lattice energy. Due to differences in the details of modeling methods, these statistics would have been difficult to determine from a survey of the literature.¹²

We might expect that certain molecular characteristics affect the predictability of a crystal and there is some evidence that more bulky molecules have less possibilities for low-energy packing, so are easier to predict than those with simple shapes. However, the greatest difference in types of molecules is the presence or absence of hydrogen-bonding sites. The known crystals of almost all non-hydrogen-bonding molecules are at, or within 1 kJ/mol of, the global minimum. These relative energies are certainly well within the expected errors due to the model potential and omission of thermal effects in lattice energy calculations. There is a much greater variability of ranking for molecules with hydrogen bonding, with up to approximately 100 computed structures lower in energy than the known crystal. While part of this variability may be due to deficiencies in the modeling, the results do point toward a greater propensity for the formation of metastable polymorphs of hydrogen-bonded molecules. Investigations of the effect of improving the model potential and thermodynamic basis of the calculations are required to determine the true proportion of metastable crystals in our test set, which we hope is reasonably representative of all small organic molecules in the Cambridge Structural Database.

Acknowledgment. We thank Dr. Neil Feeder and Dr. Pete Marshall for stimulating discussions and valuable input to the project. The referees for this paper are also thanked for their helpful comments. We also thank the Pfizer Institute for Pharmaceutical Materials for funding.

Supporting Information Available: Experimental and lattice energy minimized unit cells (Table S1); experimental and lattice energy minimized unit cells (Table S2); and structure references. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Beyer, T.; Day, G. M.; Price, S. L. *J. Am. Chem. Soc.* **2001**, *123*, 5086–5094.
- Day, G. M.; Price, S. L. *J. Am. Chem. Soc.* **2003**, *125*, 16434–16443.
- Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooij, W. T. M.; Price, S. L.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. *Acta Crystallogr.* **2000**, *B56*, 697–714.
- Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Lommerse, J. P. M.; Mooij, W. T. M.; Price, S. L.; Scheraga, H.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. *Acta Crystallogr.* **2002**, *B58*, 647–661.
- Day, G. M.; Motherwell, W. D. S.; Ammon, H.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A. V.; Dunitz, J. D.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Torrisi, A.; Nowell, H.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Schweizer, B.; Verwer, P. *Acta Crystallogr. B* **2004**, in preparation.
- van Eijck, B. P.; Mooij, W. T. M.; Kroon, J. *Acta Crystallogr.* **1995**, *B51*, 99.
- van Eijck, B. P.; Kroon, J. *J. Comput. Chem.* **1999**, *20*, 799–812.
- Boese, R.; Kirchner, M. T.; Dunitz, J. D.; Filippini, G.; Gavezzotti, A. *Helv. Chim. Acta* **2001**, *84*, 1561–1577.
- Beyer, T.; Price, S. L. *J. Phys. Chem. B* **2000**, *104*, 2647–2655.

- (10) Payne, R. S.; Roberts, R. J.; Rowe, R. C.; Docherty, R. J. *Comput. Chem.* **1998**, *19*, 1–20.
- (11) Pillardy, J.; Wawak, R. J.; Arnautova, Y. A.; Czaplewski, C.; Scheraga, H. A. *J. Am. Chem. Soc.* **2000**, *122*, 907.
- (12) Beyer, T.; Lewis, T.; Price, S. L. *CrystEngComm* **2001**, *3*, 178–212.
- (13) Anghel, A. T.; Day, G. M.; Price, S. L. *CrystEngComm* **2002**, *4*, 348–355.
- (14) Gavezzotti, A. *CrystEngComm* **2003**, *5*, 429–438.
- (15) Gavezzotti, A. *CrystEngComm* **2003**, *5*, 439–446.
- (16) Verwer, P.; Leusen, F. J. J. *Rev. Comput. Chem.* **1998**, *12*, 327–365.
- (17) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872–9882.
- (18) Mooij, W. T. M.; van Eijck, B. P.; Kroon, J. *J. Phys. Chem. A* **1999**, *103*, 9883–9890.
- (19) Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J. Comput. Chem.* **1995**, *16*, 628–647.
- (20) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352–7360.
- (21) Price, S. L. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: John Wiley and Sons: New York, 2000; Vol. 14, pp 225–289.
- (22) Mitchell, J. B. O.; Price, S. L.; Leslie, M.; Buttar, D.; Roberts, R. J. *J. Phys. Chem. A* **2001**, *105*, 9961–9971.
- (23) Gavezzotti, A.; Filippini, G. *J. Am. Chem. Soc.* **1995**, *117*, 12299–12305.
- (24) Buttar, D.; Charlton, M. H.; Docherty, R.; Starbuck, J. *J. Chem. Soc., Perkin Trans. 2* **1998**, 763–772.
- (25) Starbuck, J.; Docherty, R.; Charlton, M.; Buttar, D. *J. Chem. Soc., Perkin Trans. 2* **1999**, 677–691.
- (26) Brock, C. P.; Dunitz, J. D. *Chem. Mater.* **1994**, *6*, 1118–1127.
- (27) MS Modelling, Release 3.0.1, Accelrys Inc., A.: San Diego, 2004.
- (28) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13288.
- (29) Delley, B. *J. Chem. Phys.* **1990**, *92*, 508–517.
- (30) Cox, S. R.; Hsu, L.-Y.; Williams, D. E. *Acta Crystallogr.* **1981**, *A37*, 293–301.
- (31) Williams, D. E.; Cox, S. R. *Acta Crystallogr.* **1984**, *B40*, 404–417.
- (32) Williams, D. E. *J. Mol. Struct.* **1999**, *486*, 321–347.
- (33) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1–20.
- (34) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1154–1166.
- (35) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897–8909.
- (36) Hwang, M. J.; Stockfisch, T. P.; Hagler, A. T. *J. Am. Chem. Soc.* **1994**, *116*, 2515.
- (37) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. *J. Comput. Chem.* **1994**, *15*, 162.
- (38) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Hagler, A. T. *Isr. J. Chem.* **1994**, *34*, 195.
- (39) Peng, Z. W.; Ewig, C. S.; Hwang, M. J.; Waldman, M.; Hagler, A. T. *J. Phys. Chem. A* **1997**, *10*, 7243.
- (40) Sun, H. *J. Phys. Chem.* **1998**, *102*, 7338.
- (41) Day, G. M.; Price, S. L.; Leslie, M. *Cryst. Growth Des.* **2000**, *1*, 13–27.
- (42) Day, G. M.; Price, S. L.; Leslie, M. L. *J. Phys. Chem. B* **2003**, *107*, 10919–10933.
- (43) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622–4629.
- (44) Gavezzotti, A., Zip-Promet computer program; University of Milano, 1999–2000.
- (45) Hofmann, D. W. M.; Lengauer, T. *Acta Crystallogr.* **1997**, *A53*, 225–235.
- (46) Hofmann, D. W. M.; Lengauer, T. *J. Mol. Model.* **1998**, *4*, 132–144.
- (47) Holden, J. R.; Du, Z. Y.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422–437.
- (48) Chaka, A. M.; Zaniewski, R.; Youngs, W.; Tessier, C.; Klopman, G. *Acta Crystallogr.* **1996**, *B52*, 165.
- (49) Williams, D. E. *Acta Crystallogr.* **1996**, *A52*, 326.
- (50) Williams, D. E. MPA and MPG; University of Louisville: Louisville, 1996.
- (51) Karfunkel, H. R.; Gdanitz, R. J. *J. Comput. Chem.* **1992**, *13*, 1171.
- (52) Gdanitz, R. J. *Chem. Phys. Lett.* **1992**, *190*, 391.
- (53) Karfunkel, H. R.; Leusen, F. J. J.; Gdanitz, R. J. *J. Comput.-Aided Mater. Design* **1994**, *1*, 177.
- (54) Tajima, N.; Tanaka, T.; Arikawa, T.; Sukarai, T.; Teramae, S.; Hirano, T. *Bull. Chem. Soc. Jpn.* **1995**, *68*, 519.
- (55) Motherwell, W. D. S. *Nova Acta Leopold.* **1999**, *NF79*, 89–98.
- (56) Motherwell, W. D. S. *Mol. Cryst. Liq. Cryst.* **2001**, *356*, 559–567.
- (57) Bazterra, V. E.; Ferraro, M. B.; Facelli, J. C. *J. Chem. Phys.* **2002**, *116*, 5984–5991.
- (58) Schmidt, M. U.; Englert, U. *J. Chem. Soc., Dalton Trans.* **1996**, 2077–2082.
- (59) Schmidt, M. U.; Kalkhof, H.; Clariant GmbH: Frankfurt, 1997.
- (60) Cerius2, version 4.6, Accelrys Inc.: San Diego, 1997.
- (61) Price, S. L.; Willock, D. J.; Leslie, M.; Day, G. M. DMAREL, version 3.1, 2001.
- (62) Chisholm, J.; Motherwell, W. D. S. *J. Appl. Crystallogr.* **2004**, submitted.
- (63) Aakeroy, C. B.; Nieuwenhuyzen, M.; Price, S. L. *J. Am. Chem. Soc.* **1998**, *120*, 8986–8993.
- (64) Kitaigorodskii, A. I. *Molecular Crystals and Molecules*; Academic Press: New York and London, 1973; Vol. 29.
- (65) Chickos, J. S.; Braton, C. M.; Hesse, D. G.; Liebman, J. F. *J. Org. Chem.* **1991**, *56*, 927–938.

CG0498148