Beyond the Isotropic Atom Model in Crystal Structure Prediction of Rigid Molecules: Atomic Multipoles versus Point Charges

3 AUTHORS, INCLUDING:

Graeme Day
University of Southampton
111 PUBLICATIONS **3,478** CITATIONS

Beyond the Isotropic Atom Model in Crystal Structure Prediction of Rigid Molecules: Atomic Multipoles versus Point Charges

Graeme M. Day,[†] W. D. Sam Motherwell,[‡] and William Jones*,[†]

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

ABSTRACT: The lattice energies of predicted and known crystal structures for 50 small organic molecules with constrained (rigid) geometries have been calculated with a model potential whose electrostatic component is described by atom-centered multipoles. In comparison to previous predictions using atomic point charge electrostatics, there are important improvements in the reliability of lattice energy minimization for the prediction of crystal structures. Half of the experimentally observed crystal structures are found either to be the global minimum energy structure or to have calculated lattice energies within 0.5 kJ/mol (0.1 kcal/mol) of the global minimum. Furthermore, in 69% of cases, there are five or fewer unobserved structures with lattice energies calculated to be lower than that of the observed structure. The results are promising for the advancement of global lattice energy minimization for the ab initio prediction of crystal structures and confirm the utility of representing electrostatic contributions to the energy with atom-centered multipoles.

1. Introduction

The challenging task of predicting a molecule's most probable crystal structure(s) from knowledge of atom connectivity alone is usually approached through global minimization of the lattice energy.[1−4] However, it is the exception, rather than the norm, for the calculated energy of an observed structure to be well separated from those of other (calculated) hypothetical crystals. Thus, with normally very small energy differences among the most probable structures, it is advisable to use the best possible model for the packing energy in such crystal structure predictions. In a previous study,[5] the usefulness of lattice energy minimization, using a simple model for intermolecular interactions, was evaluated for a set of 50 rigid organic molecules, which, including polymorphs, provided a target set of 62 crystals. For 29% of these crystal structures, no lower energy unobserved structure was located in the searches, and in 56% of cases, five or fewer unobserved structures were found with lower lattice energies than the known form. During this study, we noticed that the crystal structures of molecules with hydrogen bond capability were less consistently predicted with energies close to the global minimum of the search, and there were generally more unobserved structures lower in energy than the known polymorphs when compared to molecules with no, or relatively unimportant, hydrogen bonding capacity. This observation may point to a greater tendency for the crystallization of hydrogen bonding molecules into kinetically trapped, metastable structures. However, it also raised concerns over the adequacy of the model potential to correctly predict the relative energies of competing hydrogen bond motifs.

The electrostatic component is the dominant contributor to the binding energy in hydrogen bonds, and these interactions generally show preferred orientations around the acceptor site.[6] Thus, the most obvious starting point for improving lattice energy calculations for the hydrogen bonding molecules is to use a better electrostatic model. The atomic point charges used in the earlier study[5] probably provide insufficient detail of the electron density distribution to adequately model these most important interactions in molecular crystals. This has been recognized for some time, and more elaborate methods for modeling the electrostatic interactions between molecules have been proposed. The computationally simplest approach is to define non-nuclear interaction sites at, for example, proposed positions of lone electron pairs and to place point charges at these extra positions.[7,8] Williams and Weller[7] noticed the inadequacy of atomic-centered charges for aromatic nitrogen and used such an electrostatic model in later parametrizations of nonbonded model potentials.[9−12] An even more elaborate representation of the molecular charge distribution can be obtained by adding higher order multipoles (dipole, quadrupole, etc.) to the atomic sites. Typical root-mean-square (rms) errors in the electrostatic potential around organic molecules are halved when dipoles and quadrupoles are included in addition to atomic charges,[13] and atomic multipole models give better predictions than atomic charges of hydrogen bond geometries in small molecular complexes.[14,15]

Multipoles are becoming more commonly used in modeling of molecular organic materials, as more software is being written that can handle the mechanics of multipole−multipole interactions in energy minimization,[16] lattice dynamics,[17,18] and molecular dynamics.[19−21] The superior performance of atomic multipole models to the simpler point charge has been demonstrated for

* To whom correspondence should be addressed. E-mail: wj10@cam.ac.uk.
† University of Cambridge.
‡ The Cambridge Crystallographic Data Centre.

molecular organic crystals by comparing the distortion of energy-minimized crystals away from their experimentally determined structures,[13,22] as well as in an ability to reproduce observed phonon frequencies (and hence free energies)[18] and mechanical properties.[17] The improvement for rigid molecules is clear, while the application to flexible molecules is currently hampered by problems with modeling the conformational dependence of atomic multipoles, which contribute to both the inter- and intramolecular energies.[13]

Crystal structure prediction studies using atomic multipole models have been reasonably successful,[23-28] but these studies have mostly examined individual systems in detail, not larger sets of diverse molecules that might lead to more general conclusions. Furthermore, there is a lack of direct comparisons between crystal structure prediction results using atomic multipole models and with the computationally cheaper atomic charges. Blind tests of crystal structure prediction[1,2] and a subsequent study of the 1999 blind test molecules[29] allow for some comparison, but the small set of molecules and other differences in the computational models make it difficult to draw general conclusions about the improvements that the more elaborate electrostatic model produces. We have, therefore, investigated how the lattice energy predictions for our set of 50 molecules are affected by replacing point charges by atomic multipoles.

2. Computational Methods

The molecular structures were taken from the gas phase (density functional theory, DFT) calculations of our previous study.[5] We could not easily partition the Dmol[30] calculated electron density into atomic multipoles, so we recalculated the DFT wave function using the program CADPAC.[31] We tried to keep the method as close as possible to that used for the electrostatic potential (ESP) fitted atomic charges from our previous work,[5] employing the B3P91 functional and 6-31G(d,p) basis set. Atomic multipoles were generated from the electron density using a distributed multipole analysis (DMA),[32,33] including multipoles up to $l = 4$ (i.e., charge, dipole, quadrupole, octupole, and hexadecapole) on each atom. The same *exp-6* repulsion−dispersion model as in our previous study was employed, using Williams' parameters fitted to the structures and lattice energies of a large set of small organic molecules;[11,12] these were summed to a 15 Å cutoff. Hence, the only change from our previous calculations is in the electrostatic model. Ewald summation was used for charge−charge, charge−dipole, and dipole−dipole interactions, while all higher order electrostatic terms (up to R^{-5}) were summed to a 15 Å cutoff between molecular centers of mass. The increase in computing time for lattice energy minimization using this multipole approach relative to the point charge calculations is about 10-fold.

The sets of crystal structures are those found in the simulated annealing searches described in our previous paper. For the present work, we kept all structures within 15 kJ/mol of the global minimum of our searches with the W99 + point charge model in the nine most common space groups for molecular organic crystals ($P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $P2_1$, $C2/c$, $Pbca$, $Pnma$, $Pna2_1$, and $Pbcn$).[34] Two of the molecules in our set have poly-

morphs in less common space groups (the $P3_2$ polymorph of glycine and Pa polymorph of pyrazinamide); we include these in our sets of structures for completeness. All structures were re-energy-minimized with the W99 + multipoles model potential, and as with the previous study, the validity of all of the final structures was tested by checking the stability of the elastic stiffness matrix and the ($\mathbf{k} = 0$) intermolecular phonon frequencies. These calculations were performed with the DMAREL[16-18] crystal structure modeling program.

3. Results and Discussion

As in our previous study,[5] we discuss the results in terms of a relative energy, ΔE, and the number of predicted structures that are lower in energy than the observed crystal structure, N_{lower}. The relative energy is defined with respect to the lowest energy, unobserved, structure; a negative value implies that no structures with lower energy than the observed crystal are found in the search, save possibly another known polymorph. As an example illustrating ΔE and N_{lower}, the lattice energies and densities of the predicted structures of paracetamol are presented in Figure 1. Here, each filled point represents a distinct predicted crystal structure and the open red symbols are the experimentally observed structures after energy minimization with the W99 + multipoles model potential. One of the known polymorphs (open red diamond) corresponds to the lowest energy predicted structure ($N_{\text{lower}} = 0$), and the next most stable predicted structure is 0.08 kJ/mol higher in energy ($\Delta E = -0.08$ kJ/mol). There are five unobserved predicted crystal structures lower in energy than the other polymorph (open red square, $N_{\text{lower}} = 5$), and the lowest energy unobserved structure is 1.65 kJ/mol more stable ($\Delta E = +1.65$ kJ/mol). If lattice energy could predict all crystal structures (assuming all polymorphs are known), then our target is $N_{\text{lower}} = 0$ and $\Delta E < 0$ for all crystal forms of a given molecule.

The relative energies and values of N_{lower} for the individual molecules are presented in Table 1, and the differences between the two models for each molecule are summarized in a plot of the ΔE with the point charge model against ΔE with atomic multipoles for each crystal (Figure 2). The 11 points in the lower left quadrant represent crystals that are predicted as more stable than any hypothetical crystal with both electrostatic models, while the majority of crystals are in the upper right quadrant; these are less stable than at least one unobserved hypothetical crystal in both sets of calculations.

Most of the crystals of hydrocarbons (black crosses) and polar, non-hydrogen-bonding molecules (blue triangles) lie close to the line $\Delta E_{\text{charges}} = \Delta E_{\text{multipoles}}$, demonstrating a small effect of changing the electrostatic model. For almost all of such molecules, the change in ΔE is less than 1 kJ/mol (bounded by the dashed gray lines) and uncertainties in the rest of the model potential are at least as important as the changes with the electrostatic model. The crystals of hydrogen bonding molecules are the most sensitive to the change in the electrostatic model, and in general, $\Delta E_{\text{multipoles}} < \Delta E_{\text{charges}}$. There are a few molecules for which the energy relative to the global minimum increases with the supposedly better electrostatics, and these are discussed
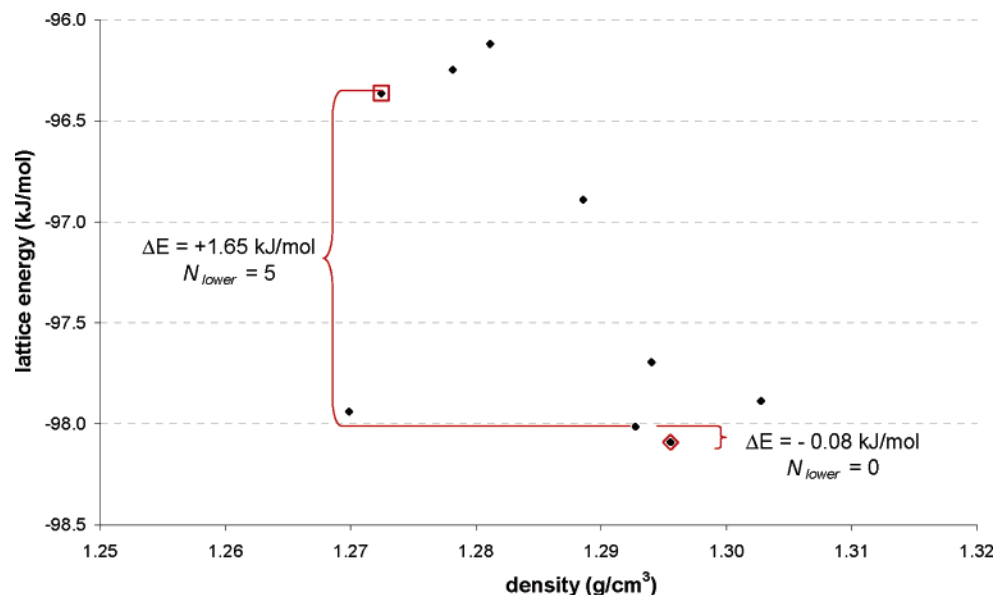
Figure 1. Predicted and observed crystal structures of paracetamol, using multipole electrostatics. Observed crystals are shown as red open symbols. For definitions of ΔE and N_{lower}, see the text.

below. However, before we address any of the individual cases, the overall distributions of ΔE and N_{lower} are compared from the point charge and atomic multipoles calculations (Figures 3–6).

3.1. Overall Differences between Atomic Charge and Multipole Predictions. There is an overall shift of the distribution to lower values of N_{lower} with the multipolar electrostatic model (Figure 3). While there is a small increase in the number of structures with $N_{lower} = 0$ (i.e., a "perfect" prediction) from 19 (point charges) to 21 (atomic multipoles), 44 of the 64 known crystal structures now have five or fewer predicted but unobserved structures with lower lattice energy; this is a noticeable improvement from the 36 crystals with $N_{lower} \leq 5$ with the simpler model. The cutoff for success in the recent blind tests of crystal structure prediction[1,2] has been for the observed structure to be found in the first three predictions. Our results suggest that the expected success rate from lattice energy minimization with a well chosen model potential and a point charge electrostatic description is approximately 42% (27 of 64 have $N_{lower} < 3$), increasing to 55% (35 of 64) with the multipole model of electrostatics.

As important as the low end of the N_{lower} distribution is the lowering of the upper limit. There were several poorly ranked crystals in the earlier study; for example, the four polymorphs of pyrazinamide (**35** in Table 1) had $N_{lower} = 0$, 24, 59, and 104, the β form of glycine (**8**) had $N_{lower} = 98$, and the known crystal of 3,5-pyrazolidinedione (**19**) also had a poor ranking ($N_{lower} = 66$). These are all lowered significantly ($N_{lower} = 0$, 7, 10, and 32 for pyrazinamide, $N_{lower} = 4$ for β-glycine, and $N_{lower} = 30$ for 3,5-pyrazolidinedione), and the highest N_{lower} in our test set is reduced from 104 to 47.

The distributions of ΔE show similar improvements. While approximately half of the observed structures are within 1 kJ/mol of the most stable structure with the point charge model, half of the known crystals are within 0.5 kJ/mol ($\sim$0.1 kcal/mol) of the global minimum with the multipole model (Figure 4). While the differences in the distributions of the relative energies are

less remarkable above $\Delta E = 1$ kJ/mol, the upper limit on ΔE is decreased significantly, from 7.3 kJ/mol (charges) to 5.1 kJ/mol (multipoles).

For crystal structure prediction (CSP) to become a useful tool, improvements at the upper and lower ends of the distributions are important. Increasing the number of structures that are very close to the global minimum demonstrates that we are becoming more successful at predicting the most stable possible crystal structure; knowing that the crystal of, say, a potentially important pharmaceutical molecule is the most stable possible polymorph would give confidence in furthering its development into a dosage form. Lowering the upper limit in ΔE or N_{lower} that must be considered in order to be confident of finding all possible polymorphs is beneficial for developing CSP as a tool where knowledge of the absolute relative energies of the predicted crystals is less important, for example, to aid the structure solution from poor powder X-ray data[35] or to find all of the likely packing and hydrogen bonding motifs for a given molecule in crystal engineering.[36] In such applications, it is most important that we choose as small a set of predicted structures as possible in which we are confident that the real structure or structures will be found. Lowering the upper limits on ΔE by 2.2 kJ/mol and N_{lower} from 104 to 47 is significant progress for such applications.

The cumulative distribution of ΔE (Figure 5) should be a useful guide for future applications of crystal structure prediction. A selection of the lowest energy predicted structures are often considered as candidates for the most likely observable crystal structure or new polymorphic form and are then chosen for closer analysis, perhaps for the prediction of properties[25] or more accurate lattice[26] or free energy[28,37] calculations. This distribution provides an estimate of the confidence that the real crystal structures will be among the candidates chosen. For example, for a 95% confidence level of having included all real structures, Figure 5 shows that all structures within 5 kJ/mol of the global minimum should be considered from predictions with the W99 +

Table 1. The Test Set of Small Organic Molecules and Results of the Crystal Structure Prediction with Both Electrostatic Models
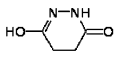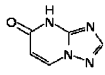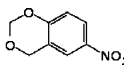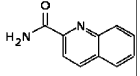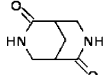
	molecule	molecular formula	CSD Refcode	space group	diagram	atomic charges		atomic multipoles	
						N_{lower}[1]	ΔE (kJ/mol)[1]	N_{lower}[1]	ΔE (kJ/mol)[1]
1	ethylene (+)	C_2H_4	ETHLEN	$P2_1/c$		2	+0.06	1	+0.07
2	acetonitrile (▲)	C_2H_3N	QQQCIV	$P2_1/c$		8	+0.75	0	-0.14
				$Cmcm$		0	-0.20	6	+1.52
3	methylamine (○)	CH_5N	METAMI	$Pbca$		3	+0.25	0	-0.13
4	nitromethane (▲)	CH_3NO_2	NTROMA	$P2_12_12_1$		4	+0.95	0	-0.11
5	formamide (○)	CH_3NO	FORMAM	$P2_1/c$		30	+1.55	0	-0.81
6	acetic acid (○)	$C_2H_4O_2$	ACETAC	$Pna2_1$		3	+1.25	9	+2.34
7	tetrolic acid (○)	$C_4H_4O_2$	TETROL	$P\bar{1}$		12	+1.40	18	+2.18
			TETROL01	$P2_1$		4	+0.86	11	+1.54
8	glycine (○)	$C_2H_5NO_2$	GLYCIN02	$P2_1/c$		2	+1.27	1	+0.65
			GLYCIN	$P2_1$		98	+5.19	4	+1.36
			GLYCIN01	$P3_2$		28[4]	+3.15[4]	0[4]	-2.28[4]
9	(Z)-2-nitroethenamine (○)	$C_2H_4N_2O_2$	EDAWIP	$P2_1/c$		8	+1.83	6	+1.94
10	2-azetidone (○)	C_3H_5NO	FEPNAP	$P\bar{1}$		0	-0.61	0	-0.19
11	1,2,4-triazole (○)	$C_2H_3N_3$	TRAZOL	$Pbca$		3	+0.51	7	+2.2
12	isoxazol-3-ol (○)	$C_3H_3NO_2$	NEZMUA	$P2_1/c$		7	+2.23	3	+1.23
13	succinic anhydride (▲)	$C_4H_4O_3$	SUCANH	$P2_12_12_1$		0	-3.18	0	-2.26
14	succinimide (○)	$C_4H_5NO_2$	SUCCIN	$Pbca$		12	+2.54	2	+0.64
15	oxazolidine-2,5-dione (○)	$C_3H_3NO_3$	OXAZDO	$C2/c$		11	+4.23	0	-1.15
16	vinylene carbonate (▲)	$C_3H_2O_3$	VINYLC	$P2_1/c$		0	-0.74	0	-0.77
17	2-oxazolidinone (○)	$C_3H_5NO_2$	OXAZIL	$P2_1/c$		0	-1.64	0	-1.07
18	5-methylene-oxazolidin-2-one (○)	$C_4H_5NO_2$	YOBQAH	$P2_1/c$		31	+1.84	4	+0.16
19	3,5-pyrazolidinedione (○)	$C_3H_4N_2O_2$	DUNVEN	$P2_1/c$		66	+5.06	30	+3.92
20	1,2,4-triazolidine-3,5-dione (urazole) (○)	$C_2H_3N_3O_2$	KOXRIY	$P2_1/c$		45	+4.23	18	+4.03
21	parabanic acid (○)	$C_3H_2N_2O_3$	PARBAC	$P2_1/c$		8	+2.44	1	+0.12

Table 1 (Continued)

	molecule	molecular formula	CSD Refcode	space group	diagram	atomic charges		atomic multipoles	
						N_{lower}[1]	ΔE (kJ/mol)[1]	N_{lower}[1]	ΔE (kJ/mol)[1]
22	bicyclo(3.2.0)hept-1(5)ene (+)	C_7H_{10}	JIZREP	$P2_1/c$		10	+0.48	28	+1.69
23	3,4-cyclobutylfuran (▲)	C_6H_6O	XULDUD	*Pbca*		0	-0.36	1	+0.20
			XULDUD01	$P2_1/c$		24	+2.48	47	+3.80
24	benzene (+)	C_6H_6	BENZEN	*Pbca*		2	+0.16	3	+0.19
			BENZEN03	$P2_1/c$		0	-0.16 (-1.12 with PV[2])	4	+0.82 (+1.05 with PV[2])
25	pyrimidine (▲)	$C_4H_4N_2$	PRMDIN	$Pna2_1$		5	+0.61	3	+0.32
26	*m*-hydroxybenzaldehyde (○)	$C_7H_6O_2$	XAYCIJ	$Pna2_1$		Only including observed conformation			
						5	+0.61	0	-0.27
						Including all 4 possible conformations			
						25	+2.24	1	+1.15
27	*m*-nitrophenol (○)	$C_6H_5NO_3$	MNPHOL01	$P2_12_12_1$		3	+1.61	1	+0.04
			MNPHOL02	$P2_1/c$		12	+2.70	1	+0.19
28	*o*-nitrobenzaldehyde (▲)	$C_7H_5NO_3$	NIBZAL	$P2_1$		10	+1.53	9	+1.93
29	2-amino-4-nitropyridine (○)	$C_5H_5N_3O_2$	SEGRUR	$P2_1/c$		43	+7.30	5	+3.13
30	2-amino-3-nitropyridine (○)	$C_5H_5N_3O_2$	AMNTPY	$P2_1/c$		0	-1.18	0	-1.43
			AMNTPY01	$P2_1/c$		0	-0.67	0	-0.81
			AMNTPY02	$P2_1/c$		-[3]	-[3]	-[3]	-[3]
31	nicotinic acid (○)	$C_6H_5NO_2$	NICOAC	$P2_1/c$		0	-0.14	2	+0.44
32	isonicotinic acid (○)	$C_6H_5NO_2$	ISNICA	$P\bar{1}$		6	+1.35	0	-0.75
33	benzamide (○)	C_7H_7NO	BZAMID	$P2_1/c$		3	+3.15	0	-1.38
34	nicotinamide (○)	$C_6H_6N_2O$	NICOAM	$P2_1/c$		0	-0.96	0	-4.93
35	pyrazinamide (○)	$C_5H_5N_3O$	PYRZIN	$P2_1/c$		24	+2.72	32	+2.86
			PYRZIN01	$P2_1/c$		59	+4.19	7	+0.88
			PYRZIN02	$P\bar{1}$		104	+5.88	10	+1.05
			PYRZIN05	*Pa*		0[4]	-0.89[4]	0[4]	-2.78[4]
36	paracetamol (○)	$C_8H_8NO_2$	HXACAN	*Pbca*		11	+3.33	5	+1.65
			HXACAN01	$P2_1/c$		0	-1.32	0	-0.07

Table 1 (Continued)

	molecule	molecular formula	CSD Refcode	space group	diagram	atomic charges		atomic multipoles	
						N_{lower}[1]	ΔE (kJ/mol)[1]	N_{lower}[1]	ΔE (kJ/mol)[1]
37	maleic hydrazide (○)	$C_4H_4N_2O_2$	MALEHY01	$P2_1/c$		2	+0.84	19	+3.94
			MALEHY10	$P\bar{1}$		2	+0.55	22	+4.33
			MALEHY12	$P2_1/c$		19	+3.59	27	+5.07
38	phthalide (▲)	$C_8H_6O_2$	HEZQUY	$P2_1/c$		2	+0.82	0	-1.14
39	indazolinone (○)	$C_7H_6N_2O$	FADMIG	$P2_1/c$		9	+2.22	6	+1.71
40	3-nitrobenzofuran (▲)	$C_8H_5NO_3$	NIVBUP	$P\bar{1}$		0	-0.13	2	+0.24
41	7-nitroindazole (○)	$C_7H_5N_3O_2$	FULKUS	$P2_1/c$		0	-0.03	1	+0.93
42	4,5-dihydro-5-oxo-(1,2,4)-triazolo(1,5-a)pyrimidine (○)	$C_5H_4N_4O$	QAJYIJ	$Pna2_1$		0	-0.01	6	+3.69
43	6-nitro-1,3-benzodioxin (▲)	$C_8H_7NO_4$	GEYWIQ	$Pbca$		1	+0.73	5	+1.12
44	1,8-naphthyridine (▲)	$C_8H_6N_2$	NAPTYR	$P2_1/c$		0	-0.93	0	-1.14
45	quinoline-2-carboxamide (○)	$C_{10}H_8N_2O$	QUINCB10	$P2_1/c$		6	+2.15	2	+1.05
46	3a,6a-dihydro-isoxazolo(5,4)isoxazole (▲)	$C_4H_4N_2O_2$	DIHIXL10	$P2_1/c$		2	+0.68	0	-0.55
47	norbornene (+)	C_7H_{10}	HOBBOP	$P2_1/c$		4	+0.58	1	+0.59
48	bicyclo(3.3.1)nonane-2,6-dione (▲)	$C_9H_{12}O_2$	HEBBEV	$C2/c$		0	-4.39	0	-4.09
49	3-aza-bicyclo(3.3.1)nonane-2,4-dione (○)	$C_8H_{11}NO_2$	BOQQUT	$P2_1/c$		6	+3.37	2	+0.15
50	3-diazabicyclo(3.3.1) nonane-2,6-dione (○)	$C_7H_{10}N_2O_2$	DOGTIC	$P2_12_12_1$		0	-0.30	16	+3.21

[1] The ratings of the success of the prediction, N_{lower} and ΔE, are explained in the text. [2] 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. [3] AMNTPY02 relaxed to the same potential energy minimum as AMNTPY01 (see ref 5). [4] These crystals are not in the space groups considered in the searches in our previous study (ref 5). The values in parentheses are obtained by lattice energy minimizing the experimental structure and comparing the energy to the predicted structures in the nine most common space groups ($P2_1/c$, P-1, $P2_1$, $P2_12_12_1$, $C2/c$, $Pbca$, $Pna2_1$, $Pnma$, and $Pbcn$).

point charge model, while those within 4 kJ/mol should be sufficient with the W99 + multipoles model. The number of distinct structures per unit of energy tends to increase rapidly a few kJ/mol above the global
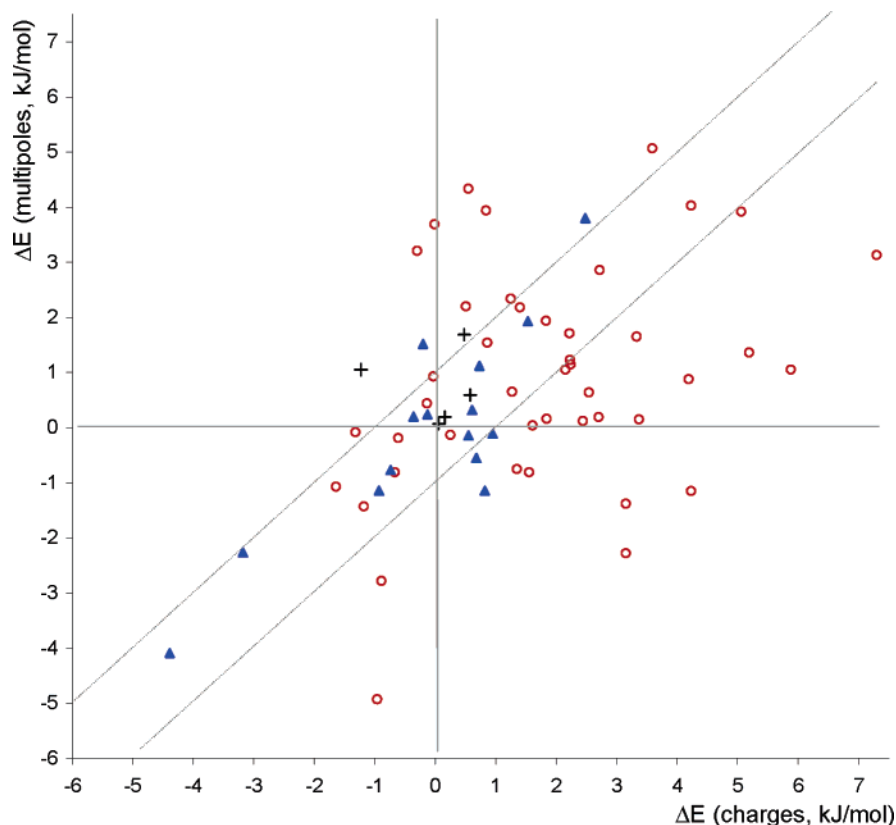
Figure 2. Plot of ΔE from multipole calculations against ΔE from atomic charge calculations. Crosses are hydrocarbons; blue (filled) triangles are polar, non-hydrogen-bonding molecules; and red (open) circles are hydrogen bonding molecules. (The classification is indicated in Table 1.) The gray lines delimit the area where the change in ΔE is less than 1 kJ/mol.
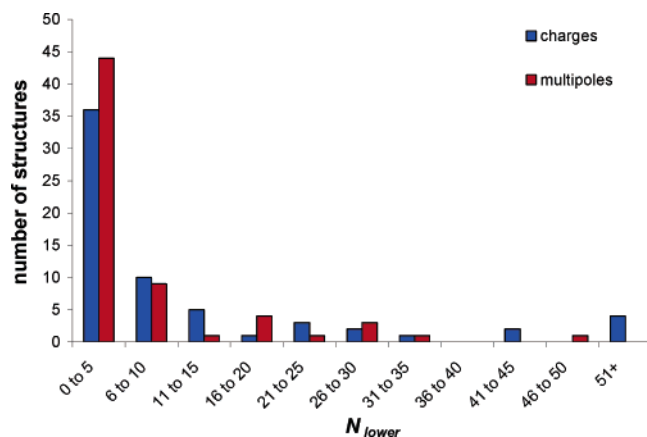

Figure 3. Distributions of N_{lower} with the atomic point charge and atomic multipole models.

Figure 4. Distributions of ΔE with the atomic point charge and atomic multipole models.

minimum, so small improvements of the ΔE cutoff can seriously reduce the number of structures that must be considered. We observe (Figure 5) that, for a given ΔE cutoff, the W99 + multipole predictions give a confidence approximately 10% higher than the W99 + point charge calculations.

It is only through systematically improving our energy calculations that we will be able to define the "real" distribution of crystal structures on their energy landscapes and, hence, determine the propensity for metastable structures to be observed. Improvements of the repulsion and dispersion models and inclusion of polarization and lattice vibration contributions to the free energy would all refine the calculated relative

energies of the predicted crystals. Ignoring the thermal contributions to the relative energies is probably the biggest limitation of current calculations. A computational study[38] of 204 pairs of known polymorphs suggests that entropy differences up to 15 J/(mol K) (=4.5 kJ/mol at room temperature) between polymorphs are possible (although almost two-thirds of polymorphic pairs differ by less than 4 J/(mol K) = 1.2 kJ/mol at room temperature). The lattice energy differences between almost all of the predicted and real crystal structures are smaller than possible entropy differences, so consideration of lattice vibrational contributions to the energy is likely to be important for any further improvements to the success of crystal structure prediction.

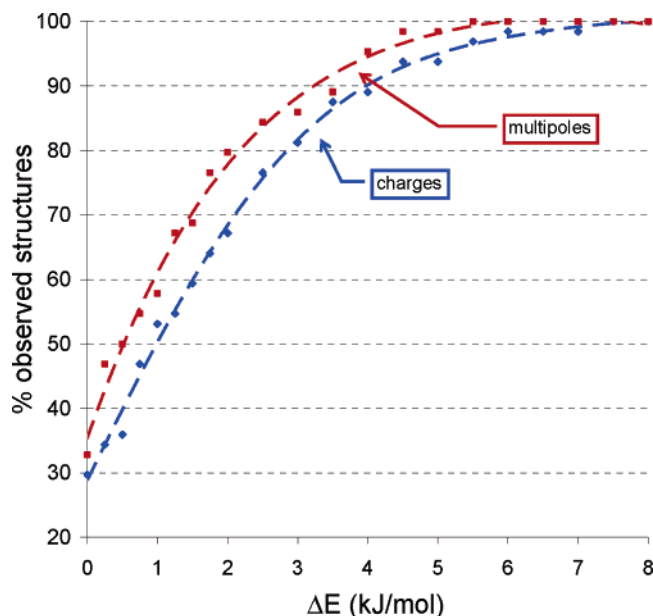**Figure 5.** Cumulative distributions of ΔE with the atomic point charge and atomic multipole models.

3.2. Hydrogen Bonding. In our earlier study, the presence of intermolecular hydrogen bonding was judged to be the greatest factor affecting the reliability of the lattice energy predictions.[5] With the multipole electrostatic model mainly improving the results for hydrogen bonding molecules, the difference in reliability for the two classes of molecule is now less dramatic (Figure 6). However, most of the structures with high ΔE (say >2 kJ/mol) are hydrogen bonded crystals. Even with the highly accurate electrostatic model, errors in lattice energy calculations are still greater when hydrogen bonds are involved, and some specific cases are discussed below. However, the distributions are sufficiently different that we are now confident in proposing that this is a real difference; metastable crystal structures are more commonly observed for molecules that can form hydrogen bonds in the crystal. High energy barriers will separate local minima on the crystal's potential energy surface where rearrangement of hydrogen bonds is required to transform between crystal structures. Therefore, the intermolecular contacts that are favorable in the early stages of crystal nucleation and growth may persist in the fully developed crystal, even where crystals of lower total energy may exist. It seems plausible that non-hydrogen-bonding molecules, on the other hand, may explore possible arrangements more freely, increasing the likelihood of finding the global minimum in energy.

3.3. Individual Cases. As well as changes in the overall results, it is worth commenting on a few of the individual molecules. First, we address those for which the more elaborate electrostatic model gives improved predictions.

(a) Oxazolidine-2,5-dione. One of the biggest changes in ΔE is for oxazolidine-2,5-dione (**15**), whose known crystal (OXAZDO) was 4.23 kJ/mol above the global minimum in our search with the point charge electrostatic model. With one donor and three acceptors, this molecule has many possibilities for hydrogen bonding in the crystal. The observed structure has hydrogen

bond dimers (Figure 7, left), while the global minimum of the W99 + point charge search and many of the other low energy structures show hydrogen bond chain motifs (Figure 7, right). With the atomic multipoles model, the relative energies change dramatically and the observed crystal structure is now lower in energy than any of the other predicted crystals ($\Delta E = -1.15$ kJ/mol). This re-ranking is understandable in terms of the geometries of the hydrogen bonding. The hydrogen bond geometry around the oxygen in the best packed chain structures is nearly linear, while real crystals usually favor C=O⋯H angles near 120°,[6] with the hydrogen approaching the oxygen atom's lone pair density. Indeed, the C=O⋯H angle in the dimers of the known crystal is 121°. The atomic point charge on the oxygen atom cannot provide this orientational bias and favors the wrong motif. The atomic multipoles, on the other hand, do describe the anisotropy of the electron density around the oxygen. Hence, the dimer motif is favored and the lattice energy of the real structure is lowered relative to that of the structures with hydrogen bond chains.

(b) Benzamide, Nicotinamide, and Pyrazinamide. Three of the molecules for which ΔE is lowered most dramatically are benzamide (**33**), nicotinamide (**34**), and pyrazinamide (**35**). These molecules are structurally very similar, being amide substituted benzene, pyridine, and pyrazine. These typify the kind of molecule for which atomic point charges give an insufficient description of the electrostatics. The π-electron density above and below the aromatic ring and around the amide C=O probably requires atomic dipoles and quadrupoles for an accurate representation, and the lone pair density at the ring nitrogen atoms and carbonyl oxygen is vitally important in the description of the hydrogen bonding interactions. For nicotinamide, the known crystal structure (NICOAM) was already the global minimum ($\Delta E = -0.96$ kJ/mol) with the simple model and, in the new calculations, the energy gap to the second most stable structure is almost 5 kJ/mol; this crystal is almost certainly the thermodynamically most stable structure, and polymorphism seems unlikely, at least within the space groups considered in this study. Benzamide (BZAMID), which forms chains of dimers in the crystal, was less well predicted in our earlier study: there were three predicted structures lower in energy than the known crystal, up to 3.15 kJ/mol more stable than the known form. In our newer calculations, BZAMID is now 1.38 kJ/mol lower in energy than all of the other predicted crystals.

There are four known polymorphs of pyrazinamide, and the improvement in the predictions with the atomic multipoles model has already been mentioned. The γ polymorph, in the space group Pa, has a lattice energy of -79.07 kJ/mol with the W99 + atomic charges model, 0.89 kJ/mol lower in energy than any of the other calculated structures. With the W99 + multipoles model, the γ form is 2.78 kJ/mol more stable than any of the predicted structures; as for nicotinamide, the energy gap between this observed structure and predicted competing forms is increased with the multipolar electrostatic model. ΔE values for the other polymorphs were $+2.72$ kJ/mol (α form), $+4.19$ kJ/mol (δ form), and $+5.88$ kJ/mol (β form) with atomic charges and $+2.86$ kJ/mol (α form), $+0.88$ kJ/mol (δ form), and $+1.05$ kJ/
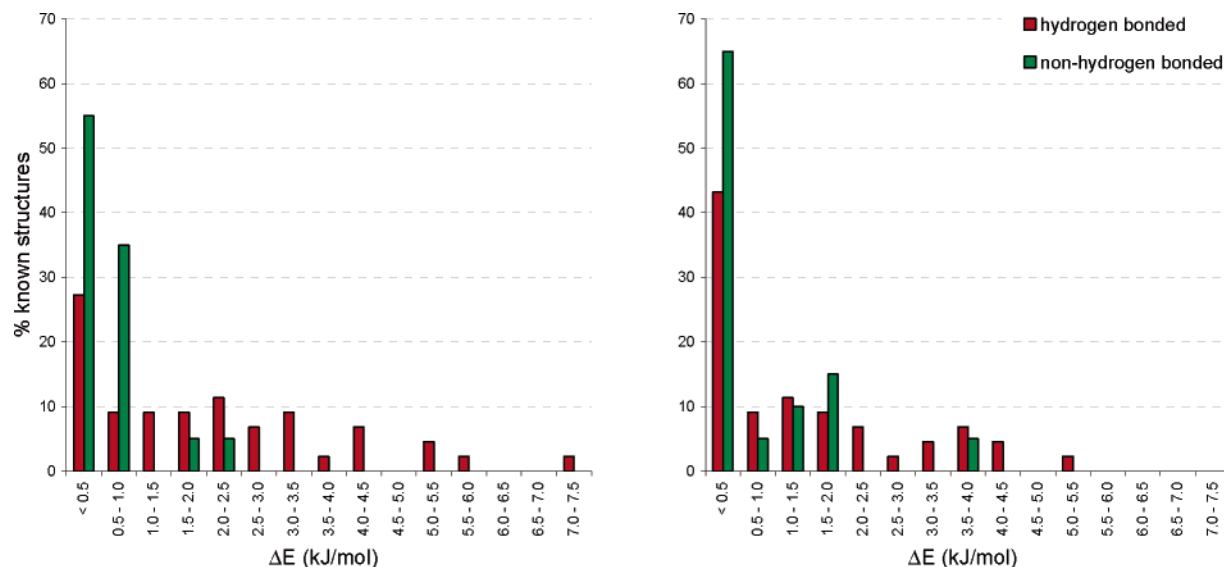
Figure 6. Relative energy distributions for hydrogen bonding and non-hydrogen-bonding molecules from the W99 + atomic charge (left) and W99 + multipoles (right) calculations.
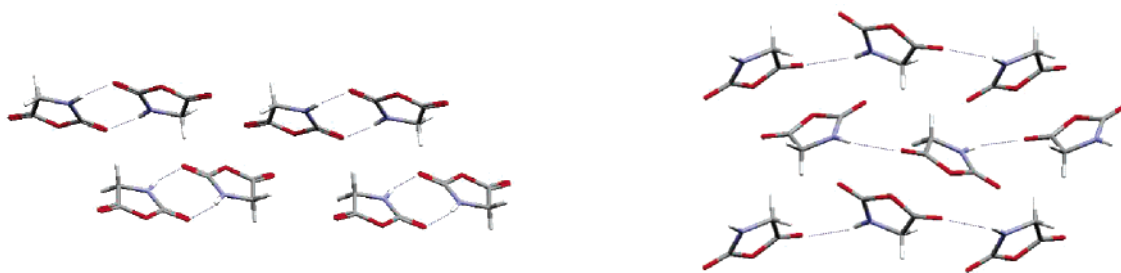

Figure 7. Hydrogen bonding in the known (left) and W99 + point charge lowest energy (right) crystal structures of oxazolidine-2,5-dione.

mol (β form) with atomic multipoles; the relative energies of the β and δ forms are noticeably lowered.

(c) Glycine. Glycine (**8**) is the only molecule in our test set that crystallizes with a zwitterionic molecular structure. We expect that the electrostatic interactions are the dominant contribution to the lattice energy and the form of the electrostatic model should be important. Our earlier calculations (W99 + charges) found the α, β, and γ forms with ΔE = +1.27, +5.19, and +3.15 kJ/mol, respectively, and N_{lower} = 2, 98, and 28. With the W99 + multipoles model, all three forms have much better rankings (N_{lower} = 1, 4, and 0) and relative energies (ΔE = +0.65, +1.36, and −2.28 kJ/mol); as expected, the changes with the different electrostatic models are important. Our calculated order of stability with the W99 + multipoles model agrees with calorimetric studies, which found that at ambient and low temperatures, the order of stability is $\gamma > \alpha > \beta$.[39]

(d) *m*-Hydroxybenzaldehyde. The predictions for *m*-hydroxybenzaldehyde (**26** in Table 1, XAYCIJ) are complicated by the existence of four possible planar conformations.[5] With the point charge model, the known crystal was found as the sixth lowest in energy in a search with the observed conformation only. Density functional theory calculations found that the molecular conformation in the known crystal is higher in energy than one of the other possible conformations that can be generated by rotating the hydroxyl and aldehyde moieties by 180°. While this alternate conformation was

found to pack less favorably than the observed conformation, the total energies (conformational + W99 + charges lattice energy) of 18 of the crystals predicted for this conformation were lower than the total energy of the observed crystal. Furthermore, two of the predicted crystals of a third molecular conformation also had lower total energies than XAYCIJ. With the atomic multipoles model for the lattice energies, the known structure is the most stable crystal in the search with the known molecular conformation and the lattice energies of the alternate conformations are relatively disfavored. So, on total energy, only one unobserved crystal structure now has a lower calculated total energy than the known crystal. Thus, assuming that XAYCIJ is the thermodynamically stable form, the multipolar electrostatic model is more successful both at correctly ordering the possible crystal structures within the known conformation and at predicting the generally more favorable lattice energies for the observed molecular conformation.

Despite the overall improved results of the multipole model over the atomic point charge calculations, a few of the molecules have higher relative energies after the newer calculations. Are these cases where the multipoles model is less accurate than the atomic charges, or are the calculations pointing to real metastable crystals? It is easier to rationalize the cases where the more elaborate calculations improve the predictions. These molecules are good candidates for further calcula-
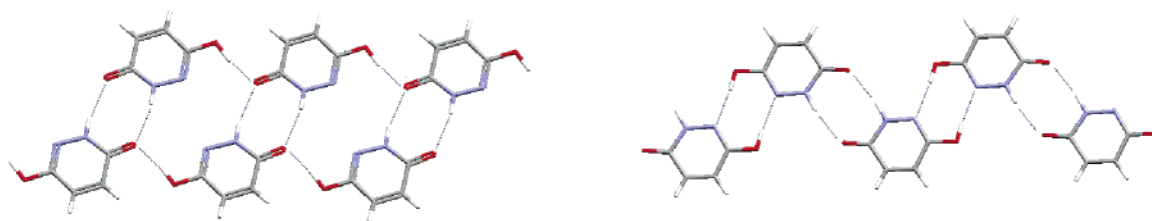
Figure 8. Hydrogen bonding in the known polymorphs (left) of maleic hydrazide and the lowest energy structures with the W99 + multipoles model potential (right).

tions to investigate errors in the models and possibly experimental investigation; even for molecules where the polymorphism has been investigated extensively, it is possible that lower energy polymorphs are realizable after remarkable time-lags.[40]

(e) Maleic Hydrazide. The most extreme case in our test set is maleic hydrazide (**37**); three polymorphs are known, and its polymorphism has been well studied.[41] All three have the same hydrogen bonding motif: O–H···O=C chains, linked together by N–H···O=C dimer interactions (Figure 8, left). Two of these had low ΔE (+0.55 and +0.84 kJ/mol) and N_{lower} (both had N_{lower} = 2) in our original search with the atomic point charge model, while the third was significantly higher in the predicted list (ΔE = +3.59 kJ/mol, N_{lower} = 19). With the point charges replaced by multipoles, the energies of all three are increased relative to the lowest energy structure, to ΔE = +3.94, +4.33, and +5.07 kJ/mol, with N_{lower} = 19, 22, and 27. Almost all of the crystals that are now more stable than the known forms are made up of chains of molecules linked by pairs of O–H···N and N–H···O=C interactions (Figure 8, right). In fact, the three known polymorphs are the first, second, and sixth lowest energy structures that lack the O–H···N hydrogen bond. A possible explanation is that the W99 *exp-6* parameters are not well suited for this molecule; no molecules in the W99 parametrization set contained both oxygen and nitrogen,[11,12] so the relative energies of O–H···N and O–H···O interactions might be unreliable. The differences in the two sets of calculations are intriguing, and maleic hydrazide is currently a target for further calculations to resolve the energy differences between observed and predicted structures.

(f) 3-Diazabicyclo(3.3.1)nonane-2,6-dione and 4,5-Dihydro-5-oxo-(1,2,4)-triazolo(1,5-a)pyrimidine. The relative lattice energy of the observed crystal of 3-diazabicyclo(3.3.1)nonane-2,6-dione (**50**, DOGTIC) increased significantly with the atomic multipoles electrostatic model. It is surprising that the prediction for this molecule should worsen (from ΔE = −0.3 kJ/mol with the point charge model to +3.21 kJ/mol using atomic multipoles) while the relative energy of the structurally very similar 3-aza-bicyclo(3.3.1)nonane-2,4-dione (**49**, BOQQUT) was improved by more than 3 kJ/mol. Similarly, the known crystal of 4,5-dihydro-5-oxo-(1,2,4)-triazolo(1,5-a)pyrimidine (**42**, QAJYIJ) is assigned a high relative energy (ΔE = +3.69 kJ/mol) with the W99 + multipoles model. Unlike maleic hydrazide, most of the lower energy structures are characterized by the same hydrogen bonding (N–H···O=C) as in the observed form, so the energy differences between structures result from variations in the weaker interactions.

As far as we know, both molecules (**50** and **42**) were only crystallized under one set of conditions, with the primary aim of determining the molecular structure.[42,43] Further experimental investigation is therefore required to be certain that other polymorphs do not exist.

(g) 3,5-Pyrazolidinedione and 1,2,4-Triazolidine-3,5-dione (Urazole). 3,5-Pyrazolidinedione (**19**, DUNVEN) and 1,2,4-triazolidine-3,5-dione (urazole, **20**, KOXRIY) are the final poorly predicted crystals. Both are calculated to be approximately 4 kJ/mol less stable than the global minimum, with many lower energy predicted structures. The crystals of both molecules are heavily hydrogen bonded and so may be susceptible to being kinetically trapped in metastable crystal structures, and neither has been extensively studied for polymorphism. For these molecules, the two sets of calculations agree in assigning them high relative energies, so we are confident in the calculations. For this reason, these are the most promising targets for experimental polymorph screens of the 50 molecules studied.

4. Conclusions

The lattice energy crystal structure predictions of 50 small organic molecules have been compared using results based on two electrostatic models. Overall, the more elaborate model, with multipoles up to hexadecapole on each atom, is more successful at predicting the known crystal structures near the global minimum in lattice energy and with fewer lower energy structures than the simpler (atomic point charge) electrostatic model. We found that 32 of the 64 known crystal structures are either global minima or have calculated lattice energies within 0.5 kJ/mol of the lowest energy structure. This is compared to 23 of the known crystals within 0.5 kJ/mol of the global minimum with the atomic charge model. Furthermore, the lattice energy of a known crystal is now never higher than 5.1 kJ/mol above the lowest energy predicted structure (compared to 7.3 kJ/mol with atomic charges).

The improvements in these results over the calculations with the more commonly used atomic point charge models are important for all applications of crystal structure prediction, and with the steady increase of computer power that is available at low cost, the added computing expense (approximately a 10-fold increase for energy minimizations) is easily manageable for small molecules. Considering the factors that are ignored in the lattice energy approach, systematic improvements to the energy model and the related computing cost are bound to give diminishing returns. We can imagine that there is a "true" distribution corresponding to Figure 5 and improvements to the model will take us closer to this reality. A reliable description of the relative lattice energies of possible crystal structures is a necessary

basis for evaluating the importance of other factors (e.g., relative nucleation and growth rates) in determining a crystal's structure. For example, the atomic charge calculations might have motivated an exploration of kinetic reasons why hydrogen bonded dimers are found in the crystal structure of oxazolidine-2,5-dione (**15**) in preference to chains; energy calculations with the atomic multipoles model suggest that the preference is simply energetic and any exploration of other determining factors would have been ill-founded.

It is for such hydrogen bonding molecules that atomic multipoles provide the greatest improvement in reliability of lattice energy predictions. However, crystals of non-hydrogen-bonding molecules are still more reliably predicted than those of hydrogen bonding molecules; we now have confidence that this reflects the ability of such strong, directional interactions to kinetically trap thermodynamically metastable crystal structures.

References

(1) 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.
(2) 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.
(3) Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; 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*, submitted.
(4) The observed crystal does not necessarily correspond to the lowest in energy, and there have been some attempts to introduce kinetic scoring criteria to structures generated from lattice energy minimization.
(5) Day, G. M.; Chisholm, J.; Shan, N.; Motherwell, W. D. S.; Jones, W. *Cryst. Growth Des.* **2004**, *4*, 1327.
(6) Steiner, T. *Angew. Chem., Int. Ed.* **2002**, *41*, 48.
(7) Williams, D. E.; Weller, R. R. *J. Am. Chem. Soc.* **1983**, *105*, 4143.
(8) Karamertzanis, P. G.; Pantelides, C. C. *Mol. Simul.* **2004**, *30*, 413.
(9) Williams, D. E.; Cox, S. R. *Acta Crystallogr.* **1984**, *B40*, 404.
(10) Williams, D. E. *J. Mol. Struct.* **1999**, *486*, 321.
(11) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1154.
(12) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1.
(13) Brodersen, S. W.; Leusen, F. J. J.; Engel, G. *Phys. Chem. Chem. Phys.* **2003**, *5*, 4923.
(14) Buckingham, A. D.; Fowler, P. W. *J. Chem. Phys.* **1983**, *79*, 6426.
(15) Buckingham, A. D.; Fowler, P. W. *Can. J. Chem.* **1985**, *63*, 1985.
(16) Price, S. L.; Willock, D. J.; Leslie, M.; Day, G. M. *DMAREL*, version 3.1; 2001.
(17) Day, G. M.; Price, S. L.; Leslie, M. *Cryst. Growth Des.* **2000**, *1*, 13.
(18) Day, G. M.; Price, S. L.; Leslie, M. L. *J. Phys. Chem. B* **2003**, *107*, 10919.
(19) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.
(20) Leslie, M. L. *Mol. Phys.*, in press.
(21) Gray, A. E.; Day, G. M.; Leslie, M.; Price, S. L. *Mol. Phys.* **2004**, *102*, 1067.
(22) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.
(23) Mooij, W. T. M.; van Eijck, B. P.; Price, S. L.; Verwer, P.; Kroon, J. *J. Comput. Chem.* **1998**, *19*, 459.
(24) Mooij, W. T. M.; van Eijck, B. P.; Kroon, J. *J. Phys. Chem. A* **1999**, *103*, 9883.
(25) Beyer, T.; Day, G. M.; Price, S. L. *J. Am. Chem. Soc.* **2001**, *123*, 5086.
(26) van Eijck, B. P.; Mooij, W. T. M.; Kroon, J. *J. Phys. Chem. B* **2001**, *105*, 10573.
(27) Lewis, T. C.; Tocher, D. A.; Day, G. M.; Price, S. L. *CrystEngComm* **2003**, *5*, 3.
(28) Day, G. M.; Price, S. L. *J. Am. Chem. Soc.* **2003**, *125*, 16434.
(29) Mooij, W. T. M.; Leusen, F. J. *J. Phys. Chem. Chem. Phys.* **2001**, *3*, 5063.
(30) Delley, B. *J. Chem. Phys.* **1990**, *92*, 508.
(31) Amos, R. D.; with contributions from Alberts, I. L.; Andrews, J. S.; Colwell, S. M.; Handy, N. C.; Jayatilaka, D.; Knowles, P. J.; Kobayashi, R.; Koga, N.; Laidig, K. E.; Maslen, P. E.; Murray, C. W.; Rice, J. E.; Sanz, J.; Simandiras, E. D.; Stone, A. J.; Su, M.-D. *CADPAC*, version 6.0; University of Cambridge: Cambridge, 1995.
(32) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.
(33) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047.
(34) The sets of structures may have been slightly different if the original searches had been performed with the better electrostatic model, but this is unlikely to significantly affect the results.
(35) Tremayne, M.; Grice, L.; Pyatt, J. C.; Seaton, C. C.; Kariuki, B. M.; Tsui, H. H. Y.; Price, S. L.; Cherryman, J. C. *J. Am. Chem. Soc.* **2004**, *126*, 7071.
(36) Blagden, N.; Cross, W. I.; Davey, R. J.; Broderick, M.; Pritchard, R. G.; Roberts, R. J.; Rowe, R. C. *Phys. Chem. Chem. Phys.* **2001**, *3*, 3819.
(37) van Eijck, B. P. *J. Comput. Chem.* **2001**, *22*, 816.
(38) Gavezzotti, A.; Filippini, G. *J. Am. Chem. Soc.* **1995**, *117*, 12299.
(39) Boldyreva, E. V.; Drebushchak, V. A.; Drebushchak, T. N.; Paukov, I. E.; Kovalevskaya, Y. A.; Shutova, E. S. *J. Therm. Anal. Calorim.* **2003**, *73*, 409.
(40) Thallapally, P. K.; Jetti, R. K. R.; Katz, A. K.; Carrell, H. L.; Singh, K.; Lahiri, K.; Kotha, S.; Boese, R.; Desiraju, G. R. *Angew. Chem., Int. Ed.* **2004**, *43*, 1149.
(41) Katrusiak, A. *Acta Crystallogr.* **2001**, *B57*, 697.
(42) Haj, M. A.; Salas, J. M.; Quiros, M.; Molina, J.; Faure, R. *J. Mol. Struct.* **2000**, *519*, 165.
(43) Kostyanovsky, R. G.; Lyssenko, K. A.; El'natanov, Y. I.; Krutius, O. N.; Bronzova, I. A.; Strelenko, Y. A.; Kostyanosky, V. R. *Mendeleev Commun.* **1999**, 106.