

# Revisiting the Blind Tests in Crystal Structure Prediction: Accurate Energy Ranking of Molecular Crystals

Aldi Asmadi,<sup>†</sup> Marcus A. Neumann,<sup>‡</sup> John Kendrick,<sup>†</sup> Pascale Girard,<sup>‡</sup> Marc-Antoine Perrin,<sup>§</sup> and Frank J. J. Leusen<sup>\*,†</sup>

*Institute of Pharmaceutical Innovation, University of Bradford, Bradford, West Yorkshire, BD7 1DP, United Kingdom, Avant-garde Materials Simulation SARL, 30 bis, rue du vieil Abreuvoir, 78100 Saint Germain en Laye, France, and Laboratoire de Chimie Physique, UPRES-EA4066, Faculté de Pharmacie, Université Paris Descartes, 4 avenue de l'observatoire, 75006 Paris, France*

Received: July 22, 2009; Revised Manuscript Received: October 23, 2009

In the 2007 blind test of crystal structure prediction hosted by the Cambridge Crystallographic Data Centre (CCDC), a hybrid DFT/MM method correctly ranked each of the four experimental structures as having the lowest lattice energy of all the crystal structures predicted for each molecule. The work presented here further validates this hybrid method by optimizing the crystal structures (experimental and submitted) of the first three CCDC blind tests held in 1999, 2001, and 2004. Except for the crystal structures of compound **IX**, all structures were reminimized and ranked according to their lattice energies. The hybrid method computes the lattice energy of a crystal structure as the sum of the DFT total energy and a van der Waals (dispersion) energy correction. Considering all four blind tests, the crystal structure with the lowest lattice energy corresponds to the experimentally observed structure for 12 out of 14 molecules. Moreover, good geometrical agreement is observed between the structures determined by the hybrid method and those measured experimentally. In comparison with the correct submissions made by the blind test participants, all hybrid optimized crystal structures (apart from compound **II**) have the smallest calculated root mean squared deviations from the experimentally observed structures. It is predicted that a new polymorph of compound **V** exists under pressure.

## Introduction

Since 1999, a series of blind tests in crystal structure prediction (CSP) has been hosted by the Cambridge Crystallographic Data Centre (CCDC) to assess the progress and development of computational methods in the field of CSP. Up to the present, four blind tests have been organized which have taken place in 1999 (CSP1999),<sup>1</sup> 2001 (CSP2001),<sup>2</sup> 2004 (CSP2004),<sup>3</sup> and most recently in 2007 (CSP2007).<sup>4</sup> Given only the molecular structures and in some cases information about the number of independent molecules in the asymmetric unit, participants are invited to predict the crystal structures of three to four compounds, for which the experimental crystal structures are known but not available to the participants. Participants are allowed to submit up to three predictions for each compound. Until the latest blind test, the rate of success in predicting the experimentally observable crystal structures was variable, with no method successfully predicting the crystal structures of all compounds in the test.

All successful predictions have used the calculated lattice energy as their main selection criterion. This approach is based on the assumption that the experimental crystal structure corresponds to the global minimum of the lattice energy hypersurface, provided that the observed structure is thermodynamically stable at the experimental pressure and temperature.<sup>5,6</sup> The inability of lattice energy calculation methods to consistently predict the correct crystal structures indicated that either effects

such as entropy, zero-point energy, thermal energy and, possibly, crystallization kinetics could not be ignored or the underlying methods (usually a force field) for calculating the lattice energy were not sufficiently accurate.<sup>7</sup>

To address this issue, a hybrid DFT (density functional theory) method has recently been developed and parametrized to yield accurate lattice energies suitable for CSP.<sup>7</sup> In the hybrid method, the lattice energy of a molecular crystal is computed from the DFT total energy (using the Vienna Ab Initio Simulation Package, VASP<sup>8–11</sup>) and an empirical van der Waals correction obtained from an atom–atom pairwise sum of damped interactions, which are of the form  $C_6/R^6$ . In the most recent blind test (CSP2007),<sup>4,12</sup> for each of the four compounds, the hybrid method correctly predicted the experimental crystal structure as the structure with the lowest lattice energy, indicating that the method has great potential in the ranking of potential crystal structures. The work described here extends the validation of the hybrid method by using it to rerank all the submitted and experimental structures of the first three blind tests (CSP1999, CSP2001, and CSP2004).

## Methodology

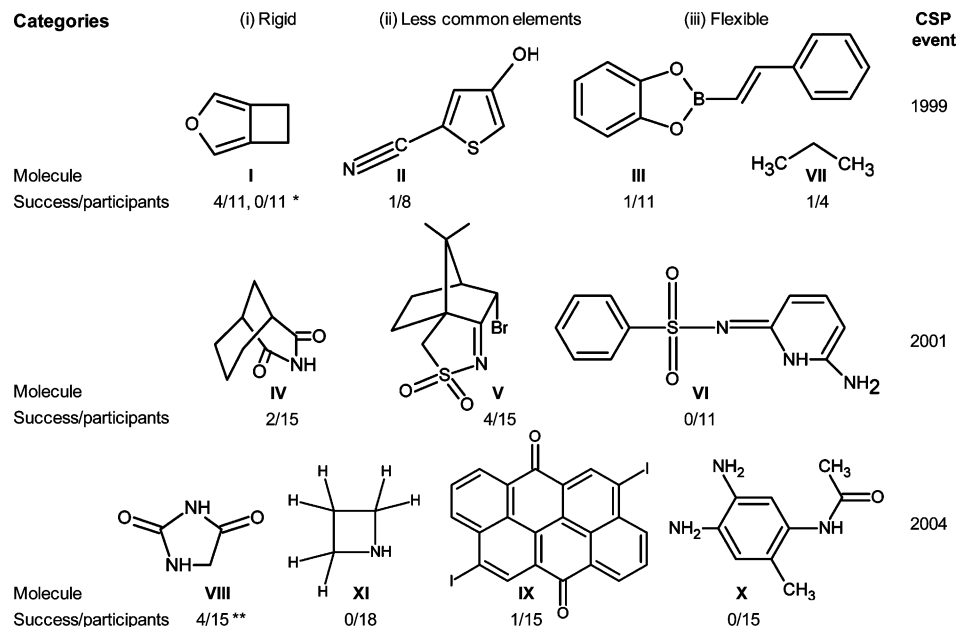
The lattice energies of all crystal structures submitted as part of CSP1999, CSP2001, and CSP2004 were minimized using the hybrid method implemented in GRACE version 1.0.<sup>13</sup> DFT calculations were performed using VASP 4.6 with projector-augmented wave (PAW) potentials to describe the interactions between ions and electrons and the PW91 exchange–correlation functional with the Vosko, Wilk, and Nusair interpolation formula. A plane-wave basis set cutoff energy of 520 eV was used, and Brillouin zone integrations were performed on a grid

\* Corresponding author. E-mail: f.j.j.leusen@bradford.ac.uk. Phone: +44(0)1274 236144.

<sup>†</sup> University of Bradford.

<sup>‡</sup> Avant-garde Materials Simulation SARL.

<sup>§</sup> Université Paris Descartes.



**Figure 1.** Summary of results from the first three CCDC blind tests. The number of successful predictions and the number of participants attempting predictions is given for each compound. \* = Two polymorphs were found experimentally, and the structures were determined. One structure, the *Pbca* form, was considered metastable, and predicted by four participants. The other structure, the *P2<sub>1</sub>/c* form, was considered stable, and was not predicted by any group. \*\* = The predictions were not blind (a partial crystal structure had been reported prior to the blind test, which was discovered halfway through CSP2004).<sup>3</sup>

with a *k*-point spacing of approximately  $0.07 \text{ \AA}^{-1}$ . The electronic wave function convergence on total energy was set to  $2.1 \times 10^{-6} \text{ kJ} \cdot \text{mol}^{-1}$  per atom for minimizations. The van der Waals correction was given by a pairwise sum over all atoms. The cutoff radii were chosen automatically by the program to obtain an average accuracy of  $1.7 \times 10^{-4} \text{ kJ} \cdot \text{mol}^{-1}$  per atom. Spline interpolants were used to ensure that the van der Waals correction smoothly decays to zero close to the cutoff radii. The  $C_6$  coefficients for heteroatomic interactions were calculated by a combination rule from atomic coefficients that have been fitted to molecular  $C_6$  coefficients obtained from experimental dipole oscillator strength distributions. The parameters of the damping function have been fitted to reproduce a representative set of 31 crystal structures determined at low temperatures.<sup>7</sup> Minimizations were considered complete when energies were converged to better than  $1.0 \times 10^{-3} \text{ kJ} \cdot \text{mol}^{-1}$  per atom, atomic displacements to  $3 \times 10^{-3} \text{ \AA}$ , and maximum atomic forces to  $2.9 \text{ kJ} \cdot \text{\AA}^{-1}$  per mole of atoms.

To give an indication of the computer resources needed to perform the calculations reported here, the elapsed time for a calculation of the energy and its gradients on molecule **VI** in space group *P2<sub>1</sub>/c* with 28 atoms in the asymmetric unit takes about 20 min using four Xeon E5462 processors (in parallel) running at 2.80 GHz. Each of these processors has four cores and 8 GB of memory; i.e., the simulation can access 32 GB in total distributed over the 16 cores involved in the calculation. A typical geometry optimization requires 15–20 evaluations of the energy and gradient, making the optimization take around 6 h using 16 cores.

The quality of the minimization results was evaluated using three measures: deviation ( $\Delta$ , in %) of each lattice parameter of the hybrid optimized unit cell relative to the experimental cell, unit cell deformation ( $D$ , in %) as described previously,<sup>7</sup> and root mean squared deviation (rmsd) measuring the atomic displacement between the experimental structures and the hybrid optimized structures. The latter was calculated using the crystal

packing similarity tool implemented in the Materials Module of Mercury CSD 2.0<sup>14,15</sup> which overlays a cluster of 16 molecules.

Optimized structures were grouped together if they belonged to the same minimum on the hybrid lattice energy hypersurface. The calculated lattice energies for these structures are essentially the same. In addition, the X-ray powder diffraction pattern was simulated for every structure using the Reflex module of Materials Studio 4.0,<sup>16</sup> and compared visually to the patterns simulated for the other structures in a group. Only crystal structures with nearly identical powder patterns were considered a match. Structures sharing a similar packing arrangement but with different energies were discarded as a match. Likewise, structures with the same lattice energy but a difference in their simulated powder patterns were also discarded as a match.

## Results and Discussion

All blind test molecules used in 1999, 2001, and 2004, the categories each of them belonged to, as well as the results of the participants, are summarized in Figure 1. The molecules were selected to test different aspects of the algorithms used in CSP. In total, participants submitted 406 predictions during the first three blind tests: 103 submissions made by 11 groups in CSP1999, 123 by 15 groups in CSP2001, and 180 by 18 groups in CSP2004. Results are presented for optimizations using the hybrid method on crystal structures for molecules **I–VIII**, **X**, and **XI**. Calculations on crystal structures for molecule **IX** were not carried out, since the empirical van der Waals correction has not yet been parametrized for iodine. A few structures were excluded as the submitted molecular structures differed from the structures specified in the blind tests. This left 369 crystal structures to be considered (including the 11 experimental structures, which encompasses two polymorphs of molecule **I**, that were the targets of the respective blind tests, and the two additional polymorphs of molecules **IV** and **VI** that were discovered afterward).

TABLE 1: Lattice Optimization Results of the Original Experimental Crystal Structures Using the Hybrid Method

structure <sup>b</sup>		CSD <sup>c</sup> reference	space group	$T_{\text{exp}}^d$ (K)	density (g·cm <sup>-3</sup> )	unit cell parameters <sup>a</sup>				$D^e$ (%)	rmsd <sup>f</sup> (Å)	HB <sup>g</sup>
						$a$ (Å)	$b$ (Å)	$c$ (Å)	$\beta$ (deg)			
<b>I_1</b>	exp	XULDUD	$Pbca$	293	1.280	5.31	12.65	14.54	90.00	0.78	0.062	
	hybrid				1.270	5.34	12.68	14.55	90.00			
	$\Delta^h$ (%)				-0.76	0.49	0.22	0.06	0.00			
<b>I_2</b>	exp	XULDUD01	$P2_1/c$	293	1.324	4.95	9.85	9.68	90.57	3.54	0.100	
	hybrid				1.278	4.97	10.04	9.81	90.98			
	$\Delta^h$ (%)				-3.46	0.37	1.90	1.32	0.45			
<b>II</b>	exp	GUFJOG	$P2_1/n$	283–303	1.489	7.52	8.33	9.06	100.19	5.39	0.489	2D
	hybrid				1.494	7.53	8.33	9.11	103.19			
	$\Delta^h$ (%)				0.30	0.13	0.00	0.60	3.00			
<b>III</b>	exp	QAMTAZ	$P2_1/c$	160	1.328	6.84	7.63	21.42	96.45	2.60	0.078	
	hybrid				1.309	6.80	7.71	21.64	96.49			
	$\Delta^h$ (%)				-1.43	-0.64	1.06	1.03	0.04			
<b>IV</b>	exp	BOQQUT	$P2_1/a$	297	1.338	7.70	10.61	9.34	95.03	3.25	0.083	1D
	hybrid				1.383	7.63	10.50	9.23	95.54			
	$\Delta^h$ (%)				3.34	-0.88	-1.07	-1.19	0.54			
<b>V</b>	exp	BOQWIN	$P2_12_12_1$	283–303	1.606	7.26	10.64	15.63	90.00	2.20	0.096	
	hybrid				1.633	7.15	10.62	15.66	90.00			
	$\Delta^h$ (%)				1.66	-1.53	-0.22	0.20	0.00			
<b>VI</b>	exp	UJIRIO	$P2_1/c$	183	1.464	8.25	8.96	15.09	91.21	7.38	0.114	1D
	hybrid				1.454	8.54	9.00	14.84	87.92			
	$\Delta^h$ (%)				-0.71	3.54	0.39	-1.69	-3.60			
<b>VII</b>	exp	JAYDUI	$P2_1/n$	30	0.803	4.15	12.61	6.98	91.28	3.13	0.074	
	hybrid				0.777	4.20	12.77	7.03	91.07			
	$\Delta^h$ (%)				-3.24	1.12	1.29	0.65	-0.23			
<b>VIII</b>	exp	PAHYON01	$C2/c$	190	1.699	9.35	12.18	7.23	104.59	7.05	0.183	1D
	hybrid				1.689	8.98	12.29	7.34	103.67			
	$\Delta^h$ (%)				-0.59	-3.99	0.94	1.55	-0.88			
<b>X</b>	exp	HAMTIZ01	$P2_1/n$	150	1.528	12.57	4.85	17.27	99.16	0.66	0.064	1D
	hybrid				1.527	12.60	4.84	17.25	98.98			
	$\Delta^h$ (%)				-0.07	0.28	-0.10	-0.11	-0.18			
<b>XI</b>	exp	XATMOV	$P2_1/c$	170	1.007	9.51	9.12	9.79	117.47	4.94	0.107	
	hybrid				1.008	9.67	8.95	9.94	119.03			
	$\Delta^h$ (%)				0.10	1.65	-1.81	1.51	1.33			

<sup>a</sup> Lattice parameters  $\alpha$  and  $\gamma$  are 90° in all of these structures due to space group constraints. <sup>b</sup> exp = experimental crystal structure; hybrid = hybrid optimized crystal structure. **I\_1** and **I\_2** are polymorphs of molecule **I**. <sup>c</sup> CSD stands for Cambridge Structural Database.<sup>17,18</sup> <sup>d</sup>  $T_{\text{exp}}$  = temperature of the experimental structure determination. <sup>e</sup>  $D$  is calculated as defined in ref 7. <sup>f</sup> rmsd is obtained using the crystal packing similarity tool in Mercury CSD 2.0,<sup>14,15</sup> with a 16-molecule comparison (ignoring all hydrogen atoms). <sup>g</sup> 1D or 2D in the HB column describes the dimensionality of the hydrogen bonding network in the crystal lattice. A blank indicates that no hydrogen bond is present in the crystal. <sup>h</sup> Deviation from experimental value; a positive value of  $\Delta$  indicates an expansion.

Table 1 presents the experimental data available for the original 11 structures in the Cambridge Structural Database<sup>17,18</sup> together with results of the hybrid method optimizations. Superpositions of the experimental crystal structures with their corresponding hybrid optimized crystal structures are provided in the Supporting Information. The crystal structure obtained by optimizing the experimental structure of molecule **VIII** shows the largest relative deviation in one of the unit cell parameters. The cell dimension along the  $a$  axis is compressed by 3.99% of its experimental value. In comparison to the other optimizations of the same crystal performed by the participants of CSP2004 using a variety of methods, a similar distortion was observed. As explained in the report on CSP2004,<sup>3</sup> the occurrence of these distortions is due to the choice of force field, the assumed molecular geometry, or the neglect of temperature in the calculations.

Optimization using the hybrid method usually results in cell deformations ( $D$ ) of 3% or less.<sup>7</sup> However, this is greatly influenced by the type of intermolecular interaction present in the lattice. Previous work has shown that crystal structures stabilized by networks of hydrogen bonds can show relatively large deviations using the hybrid method. This might be due to the density functional itself, which may be less accurate for hydrogen bonds than for other interactions.<sup>7</sup> This issue is highlighted by the results for molecules **II**, **VI**, and **VIII**, where

cell deformations greater than 5% are observed between the hybrid optimized structures and their corresponding experimental structures. Among the 11 experimental structures, optimization of the crystal structure of molecule **VI** yields the largest cell deformation, 7.38%. Its rmsd is 0.114 Å, which is the third highest after the optimized structures of molecule **II** at 0.489 Å and molecule **VIII** at 0.183 Å. The optimized unit cell of molecule **II** has a cell deformation of 5.39%, which is the third largest after molecules **VI** and **VIII**.

One of the best agreements between optimized and experimental structures is observed for molecule **X** (for which the molecular packing in the lattice is also stabilized by hydrogen bonding). Lattice optimization of this structure produces the smallest deformation with a  $D$  value of 0.66%, as well as a small rmsd, 0.064 Å. Its unit cell parameters deviate in the range 0.10–0.28% from the experimental values. Comparatively small structural differences are observed in crystal structures whose crystal packings are dominated by van der Waals interactions. This is confirmed by the optimized crystal structures of molecules **I** (both polymorphs), **III**, **V**, **VII**, and **XI**, which produce small cell deformations, with  $D$  values of 0.78% (**I\_1**), 3.54% (**I\_2**), 2.60% (**III**), 2.20% (**V**), 3.13% (**VII**), and 4.94% (**XI**).

Tables 2, 3, and 4 tabulate the energy rankings predicted by the hybrid method for all structures (except for compound **IX**)

TABLE 2: Energy Rankings Obtained by Lattice Optimizations of the CSP1999 Structures Using the Hybrid Method

rank	entry <sup>a</sup>	density (g·cm <sup>-3</sup> )	ΔHyb.E <sup>b</sup> (kJ·mol <sup>-1</sup> )	rank	entry <sup>a</sup>	density (g·cm <sup>-3</sup> )	ΔHyb.E <sup>b</sup> (kJ·mol <sup>-1</sup> )
<b>Molecule I</b>							
<b>1</b>	<b>exp_1, gsd_1, v-L_1, wil_1</b>	<b>1.270</b>	<b>0.00</b>	13	moo_3	1.259	2.41
<b>2</b>	<b>exp_2</b>	<b>1.278</b>	<b>0.46</b>	14	amm_1, pri_2	1.251	2.58
3	van_3	1.277	0.49	15	gsd_2, mot_1	1.234	2.64
4	wil_2	1.258	1.29	16	lom_3	1.236	2.86
5	gsd_3	1.264	1.38	17	v-L_3	1.229	4.21
6	moo_1, pri_1, smd_3	1.233	1.74	18	hof_2, hof_3	1.221	5.22
7	moo_2	1.256	1.78	19	van_2	1.245	5.50
8	v-L_2	1.248	1.82	20	lom_1, smd_2	1.161	6.02
9	smd_1	1.268	1.87	21	amm_3, lom_2	1.245	6.09
10	amm_2	1.235	1.88	22	mot_2	1.213	6.20
11	hof_1	1.260	2.14	23	van_1	1.244	6.31
12	pri_3	1.246	2.19	24	mot_3	1.211	7.00
<b>Molecule II</b>							
1	pri_1, smd_1, wil_2	1.494	0.00	11	lom_3	1.501	8.04
<b>2</b>	<b>exp, v-L_2</b>	<b>1.494</b>	<b>0.05</b>	12	lom_2	1.449	11.42
3	v-L_3	1.503	0.06	13	lom_1, mot_3	1.422	12.03
4	v-L_1	1.491	0.18	14	smd_3	1.475	12.09
5	pri_2	1.500	1.09	15	smd_2	1.529	12.62
6	wil_1	1.457	2.01	16	amm_1	1.532	20.48
7	hof_1	1.461	2.92	17	hof_2	1.432	21.24
8	mot_1	1.428	3.11	18	amm_2	1.529	25.25
9	mot_2	1.442	3.39	19	amm_3	1.535	28.81
10	pri_3	1.432	5.90	20	hof_3	1.511	30.23
<b>Molecule III</b>							
<b>1</b>	<b>exp, moo_1, van_1</b>	<b>1.311</b>	<b>0.00</b>	13	v-L_1	1.305	6.34
2	wil_2	1.285	0.58	14	gsd_1	1.287	6.76
3	pri_1, pri_3, wil_1	1.294	0.81	15	mot_3	1.226	7.05
4	moo_2	1.272	2.31	16	amm_3	1.282	7.41
5	v-L_2	1.300	3.98	17	v-L_3	1.291	7.91
6	van_3	1.260	4.13	18	smd_1	1.284	7.95
7	lom_1	1.260	4.27	19	amm_1	1.288	8.47
8	amm_2, pri_2, van_2	1.293	4.50	20	gsd_2, hof_1	1.248	8.55
9	mot_2, smd_2	1.285	5.09	21	lom_2	1.270	13.00
10	smd_3	1.291	5.30	22	mot_1	1.238	14.45
11	moo_3	1.271	5.58	23	hof_2	1.213	19.06
12	lom_3	1.264	5.91	24	hof_3	1.230	20.41
<b>Molecule VII</b>							
<b>1</b>	<b>exp, moo_1</b>	<b>0.784</b>	<b>0.00</b>	8	van_3	0.780	0.56
2	wil_1	0.785	0.10	9	lom_1	0.775	0.91
3	wil_2	0.778	0.19	10	lom_2	0.769	0.97
4	van_2	0.777	0.22	11	v-L_2	0.770	1.16
5	moo_2	0.775	0.28	12	mot_1	0.761	1.24
6	mot_3, v-L_1, van_1	0.781	0.38	13	mot_2	0.762	1.32
7	moo_3	0.776	0.45	14	lom_3	0.723	3.29

<sup>a</sup> Each subtable shows the optimized crystal structures starting from the experimental structure as well as from the predictions made by the CSP1999 participants. Bold font represents entries that correspond to the experimental structures. Every three-character abbreviation in the entries stands for one of the following: exp = experimental, amm = Ammon, gsd = Gavezzotti, Schweizer, and Dunitz, hof = Hofmann, lom = Lommerse, moo = Mooij, mot = Motherwell, pri = Price, smd = Schmidt, van = van Eijck, v-L = Verwer and Leusen, and wil = Williams. <sup>b</sup> ΔHyb.E is the relative hybrid energy between the optimized crystal structure and the global minimum.

submitted to CSP1999, CSP2001, and CSP2004, respectively. The rankings are based on the relative hybrid energy for each compound. Each table consists of several subtables, each corresponding to a particular blind test molecule. Each entry is designated by the following nomenclature: “xyz\_number”. The three characters “xyz” indicate the name(s) of the blind test participant(s) who submitted the original structure. The “number” refers to the rank of the prediction when it was submitted. All hybrid optimized crystal structures listed in Tables 2, 3, and 4 are available as a CIF file in the Supporting Information.

For 8 out of the 10 molecules, the crystal structure with the lowest hybrid lattice energy corresponds to the experimental structure. For molecule **II**, the calculated energy difference between the crystal structure with the lowest energy and the experimental structure is just 0.05 kJ·mol<sup>-1</sup>. Such energy differences are smaller than the numerical errors in the calculation due to the plane-wave energy cutoff, the *k*-point spacing, or the accuracy at which the geometry optimization is stopped. These two optimized structures have the same density but different molecular packings in the lattice, as illustrated by the

comparison of the simulated X-ray powder diffraction data for these two structures (shown in the Supporting Information).

Molecule **V** is the only molecule that is predicted by the hybrid method to have crystal structures significantly lower in energy than the lattice energy calculated for its experimental structure. The experimental structure is ranked fourth with an energy difference of 1.52 kJ·mol<sup>-1</sup> relative to the lowest energy structure. Lower rank structures were obtained by some of the CSP2001 participants using different methods. The rank 1 predictions made by Ammon, Price, and van Eijck were correct, while Williams' third structure corresponded to the experimental data. Ranking the experimental structure as the fourth most stable structure may be due to a shortcoming of the computational approach (neglect of zero-point energies and entropy, or inherent inaccuracy of the hybrid method), but it could also be an indication that there are further polymorphs of molecule **V** that remain to be discovered. Compared to the most stable predicted structure, the relative energy of the optimized experimental structure is within the energy range of likely observable polymorphs.<sup>19</sup> This ambiguity may be lifted by



TABLE 3: Energy Rankings Obtained by Lattice Optimizations of the CSP2001 Structures Using the Hybrid Method

rank	entry <sup>a</sup>	density (g·cm <sup>-3</sup> )	ΔHyb.E <sup>b</sup> (kJ·mol <sup>-1</sup> )	rank	entry <sup>a</sup>	density (g·cm <sup>-3</sup> )	ΔHyb.E <sup>b</sup> (kJ·mol <sup>-1</sup> )
Molecule IV							
1	<b>exp, leu_3, moo_2</b>	<b>1.383</b>	<b>0.00</b>	12	mot_1	1.306	6.70
2	amm_1, erk_2, leu_1, moo_1, smd_1, srg_1, van_1, ver_2, wil_1	1.372	2.30	13	s-d_2	1.325	7.63
3	amm_2	1.347	3.59	14	pri_1	1.339	7.76
4	amm_3	1.347	4.86	15	dzy_2, smd_3	1.313	8.38
5	mot_2	1.317	5.23	16	hof_1, hof_3	1.302	8.45
6	lom_3, ver_3	1.278	5.26	17	leu_2, van_2	1.329	9.25
7	erk_3, moo_3, mot_3, pri_3, smd_2, srg_2, lom_2	1.319	5.31	18	s-d_1	1.328	9.65
8	wil_2	1.333	6.07	19	s-d_3	1.297	9.92
9	lom_1	1.349	6.17	20	dzy_3	1.258	10.02
10	pri_2	1.308	6.53	21	wil_3	1.248	10.26
11	dzy_1, erk_1, ver_1, van_3	1.321	6.65	22	hof_2	1.292	10.64
Molecule V							
1	leu_1, moo_1, ver_2	1.693	0.00	13	hof_1	1.684	7.35
2	wil_1	1.660	0.09	14	erk_1, leu_3, srg_3	1.643	8.12
3	moo_2, mot_3, pri_2, srg_1, wil_2	1.636	0.41	15	smd_1	1.679	9.49
4	<b>exp, amm_1, pri_1, van_1, wil_3</b>	<b>1.631</b>	<b>1.52</b>	16	smd_3	1.678	9.64
5	dzy_2, mot_1	1.654	2.13	17	gav_1, leu_2, ver_1, erk_2	1.668	10.17
6	amm_2, dzy_1, pri_3	1.662	2.91	18	hof_2	1.592	10.78
7	van_3	1.653	3.05	19	hof_3	1.560	11.92
8	amm_3, gav_3, moo_3, srg_2, ver_3	1.657	3.15	20	gav_2	1.637	12.37
9	lom_3	1.673	5.46	21	dzy_3	1.579	12.83
10	erk_3	1.652	5.66	22	smd_2	1.654	14.41
11	lom_2	1.653	6.19	23	lom_1	1.567	20.57
12	mot_2	1.640	6.32	24	van_2	1.627	22.38
Molecule VI							
1	<b>exp</b>	<b>1.454</b>	<b>0.00</b>	16	ver_1	1.438	22.81
2	moo_3	1.532	6.43	17	ver_3	1.405	23.38
3	erk_1	1.417	6.56	18	ver_2	1.409	23.89
4	van_3	1.467	11.29	19	amm_2	1.441	24.14
5	van_2, moo_1	1.446	11.67	20	erk_3, hof_3	1.412	24.84
6	van_1	1.426	12.57	21	moo_2	1.478	25.13
7	dzy_2	1.444	13.88	22	erk_2	1.425	25.63
8	smd_1	1.474	15.02	23	amm_1	1.497	26.72
9	smd_3	1.389	17.27	24	amm_3	1.491	29.40
10	leu_2	1.459	17.82	25	srg_1	1.392	29.44
11	dzy_3	1.427	18.09	26	hof_1	1.427	37.78
12	wil_1	1.468	18.63	27	smd_2	1.420	40.74
13	srg_2	1.392	18.91	28	srg_3	1.394	44.61
14	wil_2	1.455	20.68	29	leu_3	1.476	46.86
15	dzy_1	1.453	22.15	30	leu_1	1.472	50.90
				31	hof_2	1.421	53.74

<sup>a</sup> Each subtable shows the optimized crystal structures starting from the experimental structure as well as from the predictions made by the CSP2001 participants. Bold font represents entries that correspond to the experimental structures. Every three-character abbreviation in the entries stands for one of the following: exp = experimental, amm = Ammon, day = Day, dzy = Dzyabchenko, erk = Erk, gav = Gavezzotti, hof = Hofmann, leu = Leusen, lom = Lommerse, moo = Mooij, mot = Motherwell, pri = Price, smd = Schmidt, srg = Scheraga, s-d = Schweizer and Dunitz, van = van Eijck, ver = Verwer, and wil = Williams. <sup>b</sup> ΔHyb.E is the relative hybrid energy between the optimized crystal structure and the global minimum.

further experimental work. As can be seen from Figure 2, which shows relative hybrid lattice enthalpies obtained by enthalpy minimization as a function of pressure for the predicted crystal structures ranked 1–8 (rank numbers at 0 GPa), the relative stability of the rank 1 structure increases substantially with respect to all other forms and in particular with respect to the rank 4 structure. At 2.5 GPa, the rank 1 structure is 8 kJ·mol<sup>-1</sup> more stable than the second most stable crystal structure and 17 kJ·mol<sup>-1</sup> more stable than the rank 4 structure. This behavior is consistent with the high density of the rank 1 structure (see Table 3). If the rank 1 structure is truly more stable at 0 GPa than the optimized experimental structure at rank 4, but does not crystallize easily because of unfavorable nucleation dynamics, crystallization under pressure will help to produce this predicted form. If, at 0 GPa, the rank 1 structure is wrongly predicted to be more stable than the optimized experimental structure found at rank 4 because of errors in the energy calculations, then the rank 1 structure can be expected to be the more stable form at high pressure. At high pressure, the lattice energy gap between the two structures is likely to be bigger than the inaccuracy in the energy calculations. Hence,

in both scenarios, crystallization under pressure will help to produce the rank 1 structure. Assuming that the rank 1 structure can be obtained in this manner, the relative stability of the structures at ranks 1 and 4 (at 0 GPa) can be investigated experimentally. The calculations presented in this paper suggest that a new crystal form will appear under pressure. However, this does not necessarily mean that it will be the most stable, as only a limited number of crystal structures have been considered in this study and other crystal packing alternatives, for instance those with two or more molecules in the asymmetric unit, may be more stable. Note that this applies to all the compounds studied here; i.e., other, as yet undiscovered, crystal structures may be more stable than the structures considered in this study.

In CSP1999, molecule **I** was found to crystallize in two different space groups: polymorph **I\_1** in *Pbca* and polymorph **I\_2** in *P2<sub>1</sub>/c*. The experimental conditions suggested that the *Pbca* form is metastable at 293 K. Experimentally, this structure was only obtained once, and attempts to crystallize it since have always led to the other polymorph.<sup>1</sup> Four participants submitted the metastable *Pbca* polymorph as one of their three predictions

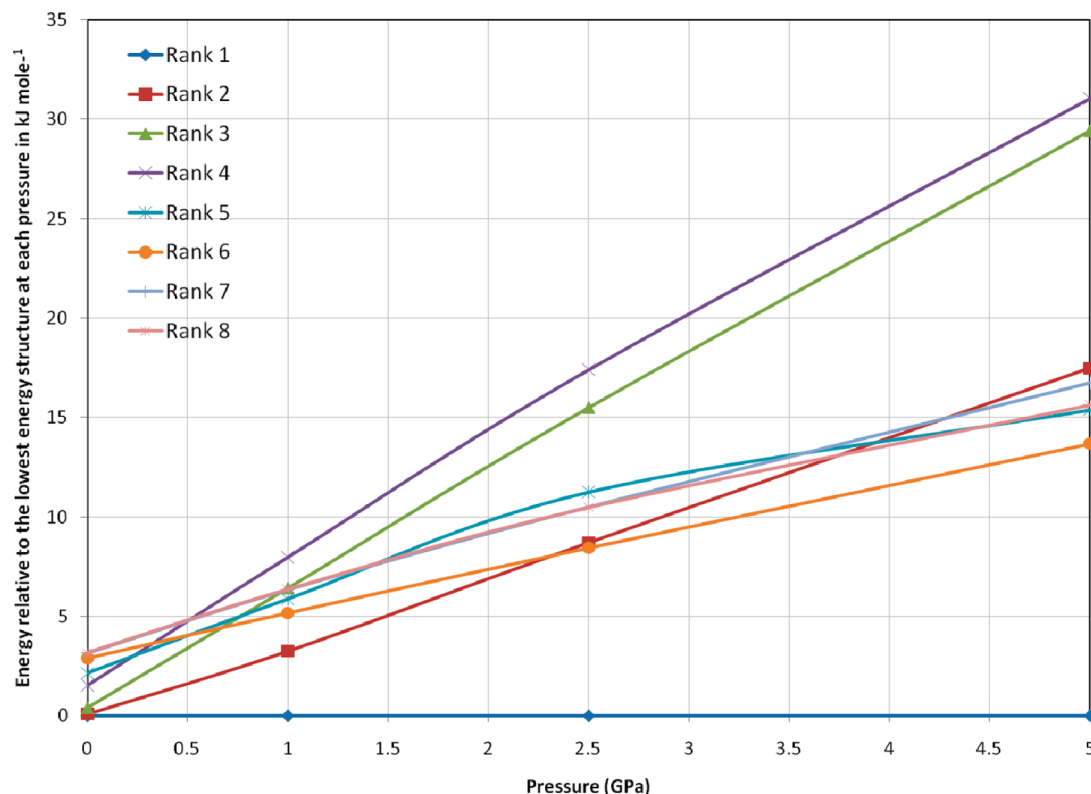
TABLE 4: Energy Rankings Obtained by Lattice Optimizations of the CSP2004 Structures Using the Hybrid Method

rank	entry <sup>a</sup>	density (g·cm <sup>-3</sup> )	ΔHyb.E <sup>b</sup> (kJ·mol <sup>-1</sup> )	rank	entry <sup>a</sup>	density (g·cm <sup>-3</sup> )	ΔHyb.E <sup>b</sup> (kJ·mol <sup>-1</sup> )
Molecule VIII							
<b>1</b>	<b>exp, amm_1, day_1, fac_2, leu_3, pan_1, swz_1, van_1</b>	<b>1.689</b>	<b>0.00</b>	9	boe_1, pan_2, ver_2, srg_2	1.599	7.48
2	day_2, pan_3	1.685	0.78	10	cli_3	1.610	7.98
3	day_3, fac_3, p-L_1, van_2	1.680	0.99	11	cli_2	1.563	10.27
4	p-L_2	1.660	1.73	12	boe_3, del_1	1.498	10.29
5	leu_2	1.644	1.83	13	leu_1	1.594	10.31
6	p-L_3	1.622	2.96	14	del_2	1.487	10.33
7	boe_2	1.613	4.04	15	cli_1, ver_3	1.552	10.47
8	fac_1, srg_3, swz_2, ver_1	1.598	7.12	16	del_3	1.512	13.85
				17	van_3	1.561	14.54
Molecule X							
<b>1</b>	<b>exp</b>	<b>1.527</b>	<b>0.00</b>	21	day_3	1.491	9.88
2	srg_3, van_3	1.584	0.78	22	fac_3	1.557	11.07
3	dzy_3, van_2	1.505	3.65	23	swz_1	1.498	11.51
4	day_2	1.505	3.69	24	dzy_2	1.528	11.73
5	amm_3, van_1	1.518	3.88	25	swz_3	1.503	12.07
6	day_1	1.515	4.12	26	smd_2	1.483	12.67
7	leu_1	1.556	4.51	27	leu_2	1.528	13.16
8	p-n_3	1.547	4.76	28	cli_1	1.483	13.94
9	leu_3	1.532	4.91	29	cli_2	1.486	14.29
10	swz_2	1.521	5.30	30	fac_1	1.564	15.57
11	amm_2	1.549	5.34	31	smd_1	1.502	15.75
12	srg_2	1.531	5.43	32	p-n_1	1.478	15.88
13	srg_1, pan_1	1.550	5.68	33	p-n_2	1.509	16.71
14	erk_3	1.509	6.28	34	cli_3	1.498	17.18
15	erk_1	1.536	6.39	35	boe_1	1.558	18.71
16	dzy_1, pan_2	1.522	7.13	36	boe_2	1.401	19.63
17	pan_3	1.527	7.17	37	fac_2	1.568	27.83
18	smd_3	1.463	9.29	38	boe_3	1.332	28.20
19	amm_1	1.531	9.62	39	hof_3	1.440	35.66
20	erk_2	1.560	9.76	40	hof_1	1.415	46.91
Molecule XI							
<b>1</b>	<b>exp*</b>	<b>1.008</b>	<b>0.00</b>	20	mot_3	1.034	1.84
2	pan_2	1.028	0.42	21	dzy_2, p-t_3	1.007	2.00
3	pan_1, smd_2, srg_2, ver_2	1.045	0.43	22	ver_1	0.980	2.07
4	swz_1	1.016	0.52	23	ver_3	1.016	3.08
5	day_3*	1.036	0.55	24	dzy_3	1.029	3.19
6	van_2	1.033	0.62	25	boe_3*	0.984	3.51
7	cli_1	1.043	0.64	26	del_1	0.946	3.64
8	cli_2	1.042	0.67	27	fac_2	1.037	3.84
9	leu_3	0.995	0.68	28	fac_3*	1.040	4.00
10	pan_3*	1.033	0.75	29	fac_1*	1.031	4.01
11	cli_3, mot_2, srg_3*	1.030	0.79	30	erk_3	0.994	4.12
12	mot_1	1.036	0.92	31	srg_1	0.936	6.10
13	amm_1, day_1, erk_1, p-t_1, smd_1, van_1	0.996	0.97	32	hof_3	1.110	8.78
14	boe_1, day_2, del_2, erk_2*, van_3	0.987	0.99	33	boe_2*	1.037	10.58
15	amm_2	0.996	1.02	34	swz_3	1.057	10.60
16	smd_3	0.975	1.04	35	hof_2	1.045	10.67
17	amm_3	1.024	1.45	36	hof_1	1.056	10.72
18	swz_2	1.013	1.78	37	leu_2	1.071	12.30
19	del_3, dzy_1, p-t_2	1.020	1.80	38	leu_1	1.066	12.54

<sup>a</sup> Each subtable shows the optimized crystal structures starting from the experimental structure as well as from the predictions made by the CSP2004 participants. Bold font represents entries that correspond to the experimental structures. For molecule **XI**, entries marked with an asterisk are structures with two independent molecules in the asymmetric unit ( $Z' = 2$ ) while unmarked entries are structures with  $Z' = 1$ . Every three-character abbreviation in the entries stands for one of the following: exp = experimental, amm = Ammon, boe = Boerrigter, day = Day, del = Della Valle, dzy = Dzyabchenko, erk = Erk, fac = Facelli, hof = Hofmann, leu = Leusen, cli = Liang, mot = Motherwell, pan = Pantelides, p-L = Price and Lewis, p-n = Price and Nowell, p-t = Price and Torrisi, smd = Schmidt, srg = Scheraga, swz = Schweizer, ver = Verwer, and van = van Eijck. <sup>b</sup> ΔHyb.E is the relative hybrid energy between the optimized crystal structure and the global minimum.

for this molecule. However, the  $P2_1/c$  polymorph was not submitted by any participant. In postanalysis of the extended lists of structures generated by the participants in CSP1999, Gavezzotti-Schweizer-Dunitz and Verwer-Leusen, two of the three participants who had submitted the  $Pbca$  form as their rank 1 prediction, located the other experimental polymorph at 1.9 and 6.5 kJ·mol<sup>-1</sup> above the global minimum, respectively. In Verwer-Leusen's list of generated structures, this polymorph was found at rank 266. Further work was carried out by Mooij and Leusen in a subsequent computational study of CSP1999,<sup>20</sup> using the same force field as Verwer and Leusen in CSP1999 (Dreiding 2.21<sup>21</sup>) but combined with atomic multipoles, instead of atomic charges, to obtain a more accurate electrostatic model

for the molecule. The  $Pbca$  polymorph was still calculated as having the most stable lattice energy, while the  $P2_1/c$  form was found at 4.7 kJ·mol<sup>-1</sup> above the global minimum. Later, another computational study produced a similar result.<sup>22</sup> A list of packing alternatives was generated using a force field and atomic charges derived from fitting the electrostatic potential. The low energy structures were then reminimized using a software package called DMAREL<sup>23,24</sup> to remove any crystal structures possibly located on the saddle points of the lattice energy hypersurface, by computing the elastic constants and phonon frequencies of the crystals. Again, the  $Pbca$  polymorph was found to be the most stable. The  $P2_1/c$  structure was 2.8 kJ·mol<sup>-1</sup> higher in energy at rank 25. As shown in Table 2, the hybrid method



**Figure 2.** Relative enthalpy changes of the eight most stable crystal structures of molecule **V** as a function of pressure, obtained by lattice enthalpy minimization with the hybrid method. The rank refers to that found at 0 GPa. The energy differences at each pressure are taken from the energy of the most stable structure at that pressure.

**TABLE 5: Lattice Optimization Results for the Additional Polymorphs of Molecules IV and VI**

structure <sup>b</sup>		CSD reference	space group	$T_{\text{exp}}^c$ (K)	density (g·cm <sup>-3</sup> )	unit cell parameters <sup>a</sup>				$D^d$ (%)	rmsd <sup>e</sup> (Å)	HB <sup>f</sup>
						$a$ (Å)	$b$ (Å)	$c$ (Å)	$\beta$ (deg)			
<b>IV_2</b>	exp	BOQQU01	$P2_1/c$	250	1.338	7.671	10.55	18.89	95.58	2.50	0.090	1D
	hybrid				1.372	7.593	10.48	18.74	96.07			
	$\Delta^g$ (%)				2.54	-1.02	-0.63	-0.75	0.51			
<b>VI_2</b>	exp	UJIRIO02	$P2_1/c$	203	1.463	12.11	10.79	17.46	97.32	1.48	0.092	2D
	hybrid				1.477	12.09	10.77	17.39	98.00			
	$\Delta^g$ (%)				0.96	-0.16	-0.24	-0.43	0.70			

<sup>a</sup> Lattice parameters  $\alpha$  and  $\gamma$  are 90° in both of these structures due to space group constraints. <sup>b</sup> exp = experimental crystal structure; hybrid = hybrid optimized crystal structure. <sup>c</sup>  $T_{\text{exp}}$  is the temperature of experimental structure determination. <sup>d</sup>  $D$  is calculated as defined in ref 7. <sup>e</sup> rmsd is obtained using the crystal packing similarity tool in Mercury CSD 2.0,<sup>14,15</sup> with a 16-molecule comparison (ignoring all hydrogen atoms). <sup>f</sup> 1D or 2D in the HB column describes the dimensionality of the hydrogen bonding network in the crystal lattice. <sup>g</sup> Deviation from experimental value; a positive value of  $\Delta$  indicates an expansion.

predicts the  $Pbca$  polymorph (exp\_1) as the most stable form, closely followed by the  $P2_1/c$  polymorph (exp\_2) as the second most stable packing alternative in the list. The lattice energy difference between the two structures is 0.46 kJ·mol<sup>-1</sup>. Considering the neglect of zero-point vibrational energies and entropic contributions in these calculations, such a small value is a clear indication that either structure can be observed experimentally.

After the results of CSP2001 were published, new polymorphs of molecules **IV**<sup>25</sup> and **VI**<sup>26</sup> were found experimentally. Both new polymorphs, denoted **IV\_2** and **VI\_2** in this work, have two independent molecules in the asymmetric unit, which explains why they were not found as part of CSP2001, since the participants were told that there would be one independent molecule per asymmetric unit. Table 5 shows the changes in unit cell parameters resulting from optimization of these two experimental structures with the hybrid method. CIF files and superpositions of the minimized structures with their experimental counterparts are available in the Supporting Information.

Despite the presence of hydrogen bonding networks in these polymorphs, both optimized structures are in good agreement with the experimental observations. Relatively small deformations (and rmsd's) of 2.50% (and 0.090 Å) and 1.48% (and 0.092 Å) are observed for **IV\_2** and **VI\_2**, respectively. In terms of the calculated lattice energies, the hybrid method suggests that the original polymorphs of molecules **IV** and **VI** are more stable by 0.56 and 0.92 kJ·mol<sup>-1</sup>, respectively. These results rank the new polymorphs in second place for both molecules.

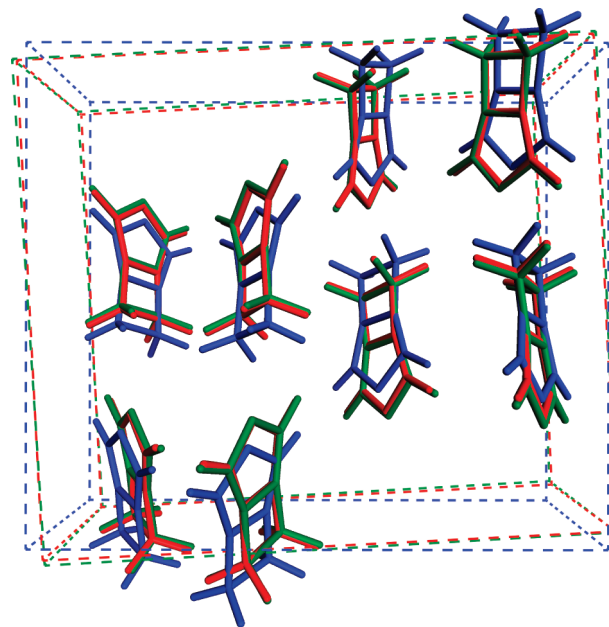
Application of the hybrid method caused significant reordering of the relative stabilities of the crystal structures reported in Tables 2, 3, and 4. There appears to be little correlation between the ranking obtained with the hybrid method and the ranking reported by the participants in CSP1999, CSP2001, and CSP2004. Furthermore, some equivalent structures are found as a result of the reoptimization with the hybrid method. These structures, which basically converged to the same minimum on the lattice energy hypersurface, are clustered together into groups in Tables 2, 3, and 4. Various algorithms can be proposed to

determine the similarity of one crystal structure to another.<sup>27–31</sup> The outcome of the clustering is somewhat arbitrary, since it depends on which method is used.<sup>32</sup>

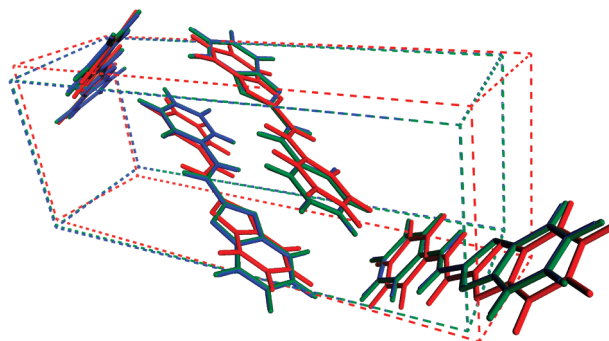
CSP is typically divided into two steps. The first is the generation of hypothetical crystal structures using a structure generation method, and the second is the ranking of these structures according to lattice energy. In practice, it is difficult to know if the search for hypothetical structures is complete. An indication of completeness within a certain energy window is given when previously predicted structures are found again, without the prediction of any new structures. Identical structures can then be removed in the clustering procedure. In this study, the structure generation is provided by the work of other participants in the blind tests, and is thus restricted to the (at most) three structures submitted by each group for each molecule. The number of clusters formed by optimizing the crystal structures for each compound can be used as an indicator to assess the completeness of the collection of minima considered in this study. The greater the number of clusters containing more than one structure, the greater is the confidence that the collection of structures is complete. The results for molecules **IV**, **V** (Table 3), and **VIII** (Table 4) indicate that the sampling of crystal packings is adequate. Results for these molecules are at variance to the distribution of clusters obtained for molecules **II** (Table 2), **VI** (Table 3), **VII** (Table 2), and **XI** (Table 4), which indicates that the collection of low energy minima for **II**, **VI**, **VII**, and **XI** is probably incomplete. The likely reason for the lack of completeness lies with the limited selection of up to only three structures for each molecule from each participant, and the differing force fields that have led to their selection. Where the force fields are very different, it is to be expected that the sampling of the potential energy surface will be different too. One way to resolve this would be to perform hybrid DFT calculations on structures predicted within a wider energy window. Molecule **XI** is different as most groups only considered crystal lattices with  $Z' = 1$ , whereas a full CSP considering  $Z' = 2$  needs to be conducted to find the experimental structure.

Most of the correct predictions in CSP1999, CSP2001, and CSP2004 were clustered together with their corresponding experimental structures after optimization with the hybrid method. However, one correct prediction for molecule **I** by van Eijck (van\_3) was optimized by the hybrid method to a structure at rank 3,  $0.49 \text{ kJ} \cdot \text{mol}^{-1}$  above the lowest energy structure. The rmsd (as reported in CSP1999<sup>1</sup>) of van Eijck's submission was the highest among the correct predictions: 0.525, in comparison to 0.204 (Gavezzotti-Schweizer-Dunitz), 0.231 (Verwer-Leusen), and 0.277 (Williams), but still acceptable as a correct prediction. These results are a further indication of just how difficult it is to determine the similarity between two crystal structures. In this case, the deviation of van Eijck's predicted structure from the experimental *Pbca* structure is sufficiently large for it to find itself in a different minimum on the potential energy surface. A comparison of the experimental *Pbca* polymorph of molecule **I**, van Eijck's prediction, and van Eijck's structure optimized with the hybrid method is shown in Figure 3.

In CSP1999, only one submitted structure for molecule **III** was considered a successful prediction (the rank 1 structure submitted by van Eijck). Mooij's rank 1 structure, moo\_1, did not meet the criteria to be a correct solution due to its high rmsd (0.955) from the experimental structure (compared with 0.214 for van Eijck's prediction).<sup>1</sup> However, the hybrid method optimized Mooij's structure to the global minimum. Figure 4 gives a visual comparison of the experimental structure of



**Figure 3.** Superposition of the experimental *Pbca* polymorph of molecule **I** (in blue) with van Eijck's third-ranked prediction (in red) and the corresponding hybrid optimized structure (in green).



**Figure 4.** Superposition of the experimental structure of molecule **III** (in blue) with Mooij's first-ranked prediction, prior to optimization (in red) and after optimization (in green) with the hybrid method.

molecule **III** with Mooij's structure prior to and after optimization with the hybrid method.

Most experimental crystal structures of the blind test molecules are reproduced well by the hybrid method. To compare its performance with the methods used in CSP1999, CSP2001, and CSP2004, Table 6 summarizes the rmsd's of the official correct predictions and the hybrid optimized structures in comparison to the experimental structures. Apart from molecule **II**, the hybrid optimized structures give the smallest rmsd values. The deviations between the experimental structures and the hybrid optimized structures are in the range 0.062–0.183 Å. For molecule **II**, a relatively large deviation of 0.489 Å is observed. As discussed earlier, the presence of hydrogen bonding in the crystal lattice may contribute to a relatively large geometrical deviation. For molecule **V**, even though the experimental structure corresponds to the energetically unfavorable rank 4 structure, the predicted geometry gives the smallest value of all the rmsd's obtained for this molecule.

Among the blind tests conducted so far, CSP2004 showed the lowest success rate. In this blind test, Day predicted the crystal structure of molecule **IX** correctly, but unfortunately that compound could not be considered in this study because the empirical van der Waals correction used in the hybrid method has not yet been parametrized for iodine. Although not blind,



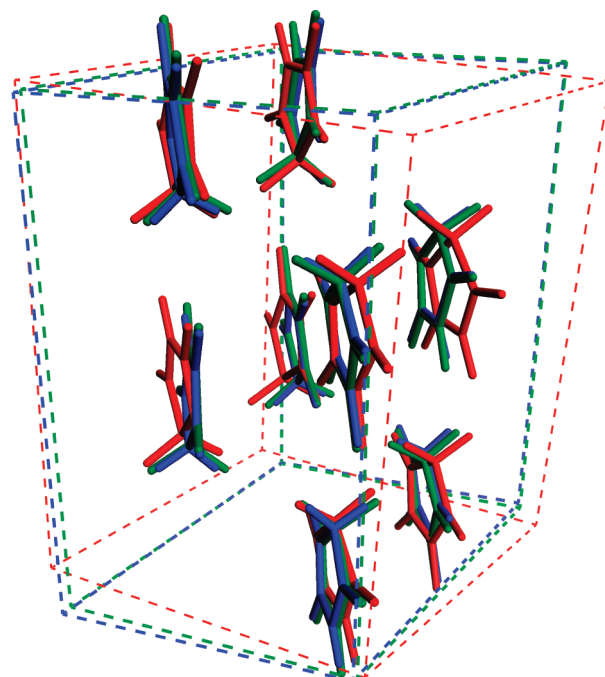
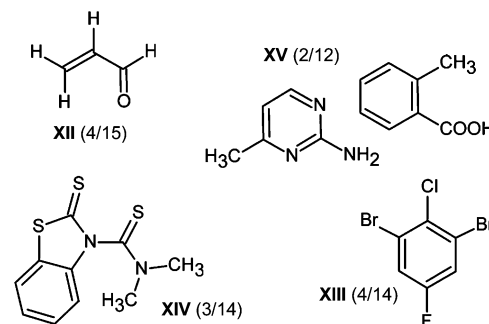
**TABLE 6: Comparison of the Geometries of Experimental Crystal Structures to Crystal Structures Optimized with the Hybrid Method and Structures Submitted by Participants in the Blind Tests**

structure	reference <sup>a</sup>	target <sup>a</sup>	rmsd <sup>b</sup>	structure	reference <sup>a</sup>	target <sup>a</sup>	rmsd <sup>b</sup>
<b>I_1</b>	<b>exp</b>	<b>hyb_opt</b>	<b>0.062</b>	<b>V</b>	<b>exp</b>	<b>hyb_opt</b>	<b>0.096</b>
	exp	gsd_1	0.211		exp	amm_1	0.329
	exp	van_3	0.645		exp	pri_1	0.320
	exp	v-L_1	0.203		exp	van_1	0.830
	exp	wil_1	0.307		exp	wil_3	0.259
<b>II</b>	<b>exp</b>	<b>hyb_opt</b>	<b>0.489</b>	<b>VII</b>	<b>exp</b>	<b>hyb_opt</b>	<b>0.074</b>
	exp	v-L_2	0.487		exp	moo_1	0.120
<b>III</b>	<b>exp</b>	<b>hyb_opt</b>	<b>0.078</b>	<b>VIII</b>	<b>exp</b>	<b>hyb_opt</b>	<b>0.183</b>
	exp	van_1	0.213		exp	day_1	0.378
<b>IV</b>	<b>exp</b>	<b>hyb_opt</b>	<b>0.083</b>		exp	fac_2	0.331
	exp	leu_3	0.225		exp	pan_1	0.332
	exp	moo_2	0.172		exp	van_1	0.352

<sup>a</sup>Please see Tables 2–4 for the key to the entries. exp = experimental crystal structure; hyb\_opt = hybrid optimized, experimental crystal structure. Results obtained with the hybrid method are highlighted in bold. <sup>b</sup>rmsd is calculated with the crystal packing similarity tool in Mercury CSD 2.0,<sup>14,15</sup> overlaying a cluster of 16 molecules between the “Reference” structure and the “Target” structure. All hydrogen atoms are ignored.

there were also some correct predictions for molecule **VIII**. Out of 36 submissions made by 15 groups, 12 groups located the experimental structure in their lists of generated structures. Five participants (Pantelides, Day, van Eijck, Leusen, and Facelli) had the experimental structure among their top three official predictions. Lattice parameters in these predictions (except for Leusen) were acceptably close to the experimental structure. In the case of Leusen, one of the submitted structures (leu\_3) does resemble the experimental data; however, it was regarded as unsuccessful due to its severely distorted lattice. Two groups (Ammon and Schweizer) submitted a structure calculated using structural information obtained from a conference abstract in which the unit cell dimensions and the space group of the crystal structure of molecule **VIII** were reported.<sup>33</sup> Their submissions did correspond to the experimental structure, but their predictions were considered to be invalid.<sup>3</sup> The hybrid method optimized all these predicted structures to the same structure as the optimized experimental structure. A comparison of the leu\_3 structure, before and after the hybrid optimization, with the experimental crystal structure of molecule **VIII** is provided in Figure 5.

CSP for flexible molecules (molecules **III**, **VI**, and **X**) and for crystal structures with more than one independent molecule in the asymmetric unit (molecule **XI**) has always proven to be difficult.<sup>1–3</sup> These systems are difficult because of the high number of geometrical variables; they test the structure generation algorithms used in CSP, as well as the accuracy of the lattice energy calculations. Flexible molecules may crystallize in a relatively unfavorable conformation which can be offset by favorable intermolecular interactions (a phenomenon called conformational polymorphism); hence, it is important that (i) all relevant conformations of a flexible molecule are considered in the search for crystal packing alternatives, (ii) crystal structure optimizations are performed with fully flexible molecules, and (iii) the lattice energy function used strikes an accurate balance between inter- and intramolecular interactions. Note that, from a CSP point of view, solvates, salts, and cocrystals can be regarded as structures with more than one independent molecule in the asymmetric unit. In CSP2001 and CSP2004, no participant successfully predicted the crystal structures of flexible molecules or crystal structures with more than one independent molecule in the asymmetric unit. In the extended lists of structures

**Figure 5.** Superposition of the known crystal structure of molecule **VIII** (in blue) with Leusen's third-ranked prediction, before optimization (in red) and after optimization (in green) with the hybrid method.**Figure 6.** Molecular diagrams of compounds used in the fourth blind test (CSP2007). The figures in parentheses (*s/p*) give the number of successful predictions (*s*) and the number of participants attempting predictions (*p*) for each compound. Compound **XV** is a cocrystal.

submitted by the participants, only 7 out of 15 (molecule **X**), 3 out of 18 (molecule **XI**), and 4 out of 11 (molecule **VI**) groups had generated the experimental crystal structures (extended lists were not submitted for molecule **III** in CSP1999). Poor geometrical agreements and relatively high lattice energies were observed for these structures. For molecule **X**, upon optimization with the hybrid method, none of the submitted structures corresponds to the optimized experimental structure. The presence of the experimental structure in the extended lists submitted by some participants indicates that for this molecule the search for candidate structures was adequate, but the accuracy of the method used for calculating the lattice energy was insufficient. For molecule **XI**, which crystallizes with two independent molecules in the asymmetric unit, six participants submitted the same structure as their lowest energy structure. This structure has only one independent molecule in the asymmetric unit and ranks 13th upon optimization with the hybrid method (see Table 4). Note that there are several other structures with one molecule in the asymmetric unit which have lower lattice energies.

For completeness, results of the most recent blind test (CSP2007) are presented here as well.<sup>4</sup> Figure 6 depicts the

**TABLE 7: Root Mean Squared Deviations of the Correct Predictions from the Experimental Crystal Structures in CSP2007<sup>a</sup>**

structure	reference <sup>b</sup>	target <sup>b</sup>	rmsd <sup>c</sup>	structure	reference <sup>b</sup>	target <sup>b</sup>	rmsd <sup>c</sup>
<b>XII</b>	<b>exp</b>	<b>nlk_1</b>	<b>0.127</b>	<b>XIV</b>	<b>exp</b>	<b>nlk_1</b>	<b>0.130</b>
	exp	amm_2	0.174		exp	pkm_1	0.222
	exp	b-t_1	0.156		exp	van_1	0.147
	exp	swz_2	0.183	<b>XV</b>	<b>exp</b>	<b>nlk_1</b>	<b>0.075</b>
<b>XIII</b>	<b>exp</b>	<b>nlk_1</b>	<b>0.082</b>		exp	van_3	0.294
	exp	amm_1	0.385				
	exp	day_1	0.159				
	exp	pkm_1	0.152				

<sup>a</sup> The figures presented in this table were taken from the full report of CSP2007.<sup>4</sup> Results obtained with the hybrid method are highlighted in bold. <sup>b</sup> Every three-character abbreviation in the entries stands for one of the following: exp = experimental, nlk = Neumann, Leusen and Kendrick, amm = Ammon, b-t = Boerrigter and Tan, day = Day, pkm = Price, Karamertzanis, Misquitta and Welch, swz = Schweizer, and van = van Eijck. <sup>c</sup> rmsd is calculated using COMPACT<sup>29</sup> by comparing 16 molecules in each crystal structure (excluding hydrogen atoms).

molecular structures of the fourth blind test compounds and summarizes the results achieved by the blind test participants for each compound. The rmsd's from the experimental data for the correctly predicted structures for all participants are shown in Table 7. The hybrid method predicted each of the four crystal structures correctly as rank 1 (nlk\_1 entries in Table 7).<sup>12</sup> The hybrid method also achieved the lowest rmsd's from the experimental structures for all of the correct predictions in CSP2007.

## Conclusions

Lattice energy minimization with the hybrid method has been shown to be remarkably successful in predicting both the structures and the relative energies of the crystal structures of small organic molecules.

Using the hybrid DFT method, the experimental structures for 8 out of 10 molecules in the blind tests (12 out of 14 if CSP2007 is included) are found as the lowest energy crystal structures. For one molecule (**II**), the experimental structure is the second most stable structure with a very small energy difference relative to the global minimum. Only the experimental structure of compound **V** is found to optimize to a structure which is significantly less stable than the lowest energy crystal structure. It ranks fourth among the optimized crystal structures for that molecule, and under pressure, the appearance of a new polymorph is expected.

In comparison to the experimental structures, the structural agreement obtained with the hybrid method is very good, especially for crystal packings dominated by van der Waals interactions. For crystal packings stabilized by a network of hydrogen bonding, the hybrid method still gives excellent geometric results; however, the structural deviations can be slightly larger than those found for crystal structures without hydrogen bonding. The structural discrepancies between the hybrid optimized structures and the experimental structures are smaller than those achieved by any of the participants in CSP1999, CSP2001, CSP2004, and CSP2007, with the exception of molecule **II**.

The results presented in this study are based on the limited range of structures predicted in the previous blind tests. To obtain a definitive answer as to the accuracy of the hybrid DFT method for predicting the relative stabilities of crystal structures, consideration of a more complete set of structures would be required. It is probable that additional important, low-energy crystal structures have not been considered in this study, since they were not present among the set of starting structures. This is especially true for molecule **XI**, since it crystallizes with two independent molecules in the asymmetric unit which was considered by only a few of the participants in CSP2004. Thus,

the collection of starting structures for molecule **XI** in this study contained insufficient crystal structures with two independent molecules in the asymmetric unit. The results indicate that the collection of starting minima for molecules **II**, **VI**, and **VII** may also be incomplete. As a consequence, the lattice energy rankings presented here may be adversely affected if a full search for minima on the hybrid lattice energy hypersurface were conducted for each of these compounds.

For molecule **III** (CSP1999), it has been discovered that one of the structures submitted by a participant, which did not qualify as a successful prediction, converges to the same minimum as the experimental structure. A similar result is observed for one of the submitted structures for molecule **VIII** (CSP2004). The opposite is found for molecule **I**, where a correct prediction by one of the participants, which is close to the experimental structure, is found as a different minimum upon optimization with the hybrid method. These cases demonstrate the difficulty in determining similarity among crystal structures.

Three of the molecules studied here have known polymorphs. The two experimentally observed polymorphs of molecule **I** are found in this study as rank 1 and 2, separated by a small energy gap of only 0.46 kJ·mol<sup>-1</sup>. This result is an improvement over previous studies<sup>1,20,22</sup> in which the lattice energy of the stable *P2<sub>1</sub>/c* polymorph was calculated to be considerably higher than the lattice energy of the metastable *Pbca* polymorph. For molecules **IV** and **VI**, new polymorphs have been discovered after the blind test (CSP2001).<sup>25,26</sup> Upon minimization with the hybrid method, both structures are ranked in second place with small lattice energy differences separating them from the global minima (0.56 kJ·mol<sup>-1</sup> for **IV** and 0.92 kJ·mol<sup>-1</sup> for **VI**). The small lattice energy differences between the rank 1 and 2 structures in each of these three cases are further evidence of the accuracy of the hybrid method.

The results presented in this study suggest that accurate lattice energies are an appropriate selection criterion in CSP. However, it is necessary to include zero-point energy and entropic contributions in these calculations to accomplish a reliable approach for the prediction of polymorphic stability.

**Acknowledgment.** A.A., J.K., and F.J.J.L. thank the School of Life Sciences at the University of Bradford for funding this project.

**Supporting Information Available:** Superpositions of experimental crystal structures and their corresponding hybrid optimized structures, comparison of the simulated X-ray powder diffraction data between the global minimum (rank 1) and the optimized experimental crystal structure (rank 2) of molecule **II**, and CIF files containing the hybrid optimized crystal

structures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (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., Sect. B* **2000**, *56*, 697–714.
- (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., Sect. B* **2002**, *58*, 647–661.
- (3) Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; 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.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P. *Acta Crystallogr., Sect. B* **2005**, *61*, 511–527.
- (4) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J. S.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M. A.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A. K.; Schweizer, B. *Acta Crystallogr., Sect. B* **2009**, *65*, 107–125.
- (5) Verwer, P.; Leusen, F. J. J. *Rev. Comput. Chem.* **1998**, *12*, 327–365.
- (6) Price, S. L. *Phys. Chem. Chem. Phys.* **2008**, *10*, 1996–2009.
- (7) Neumann, M. A.; Perrin, M. A. *J. Phys. Chem. B* **2005**, *109*, 15531–15541.
- (8) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47*, 558–561.
- (9) Kresse, G.; Hafner, J. *Phys. Rev. B* **1994**, *49*, 14251–14269.
- (10) Kresse, G.; Furthmüller, J. *Comput. Mater. Sci.* **1996**, *6*, 15–50.
- (11) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, *54*, 11169–11186.
- (12) Neumann, M. A.; Leusen, F. J. J.; Kendrick, J. *Angew. Chem., Int. Ed.* **2008**, *47*, 2427–2430.
- (13) GRACE (the Generation, Ranking, and Characterization Engine) software package is a product of Avant-garde Materials Simulation SARL, 30 bis, rue du vieil Abreuvoir, F-78100 St-Germain-en-Laye, France, [info@avmatsim.eu](mailto:info@avmatsim.eu).
- (14) Macrae, C. F.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Shields, G. P.; Taylor, R.; Towler, M.; van de Streek, J. *J. Appl. Crystallogr.* **2006**, *39*, 453–457.
- (15) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. *Acta Crystallogr., Sect. B* **2002**, *58*, 389–397.
- (16) Materials Studio 4.0 software package is a product of Accelrys, Inc, 10188 Telesis Court, Suite 100, San Diego, CA 92121 (<http://accelrys.com>).
- (17) Allen, F. H.; Kennard, O.; Taylor, R. *Acc. Chem. Res.* **1983**, *16*, 146–153.
- (18) Allen, F. H. *Acta Crystallogr., Sect. B* **2002**, *58*, 380–388.
- (19) Bernstein, J. *Polymorphism in Molecular Crystals*; Oxford University Press: Oxford, U.K., 2002.
- (20) Mooij, W. T. M.; Leusen, F. J. J. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5063–5066.
- (21) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897–8909.
- (22) Day, G. M.; Chisholm, J.; Shan, N.; Motherwell, W. D. S.; Jones, W. *Cryst. Growth Des.* **2004**, *4*, 1327–1340.
- (23) Price, S. L.; Willock, D. J.; Leslie, M.; Day, G. M. *DMAREL*, version 3.1, University College London, London, United Kingdom, 2001.
- (24) Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J. Comput. Chem.* **1995**, *16*, 628–647.
- (25) Hulme, A. T.; Johnston, A.; Florence, A. J.; Fernandes, P.; Shankland, K.; Bedford, C. T.; Welch, G. W. A.; Sadiq, G.; Haynes, D. A.; Motherwell, W. D. S.; Tocher, D. A.; Price, S. L. *J. Am. Chem. Soc.* **2007**, *129*, 3649–3657.
- (26) Jetti, R. K. R.; Boese, R.; Sarma, J.; Reddy, L. S.; Vishweshwar, P.; Desiraju, G. R. *Angew. Chem., Int. Ed.* **2003**, *42*, 1963–1967.
- (27) Karfunkel, H. R.; Rohde, B.; Leusen, F. J. J.; Gdanitz, R. J.; Rihs, G. *J. Comput. Chem.* **1993**, *14*, 1125–1135.
- (28) Dzyabchenko, A. V. *Acta Crystallogr., Sect. B* **1994**, *50*, 414–425.
- (29) Chisholm, J. A.; Motherwell, S. *J. Appl. Crystallogr.* **2005**, *38*, 228–231.
- (30) Willighagen, E. L.; Wehrens, R.; Verwer, P.; de Gelder, R.; Buydens, L. M. C. *Acta Crystallogr., Sect. B* **2005**, *61*, 29–36.
- (31) Mooij, W. T. M.; van Eijck, B. P.; Price, S. L.; Verwer, P.; Kroon, J. *J. Comput. Chem.* **1998**, *19*, 459–474.
- (32) van Eijck, B. P. *Acta Crystallogr., Sect. B* **2005**, *61*, 528–535.
- (33) Yu, F.; Schwalbe, C. H. In *Structure of hydantoin from a twinned crystal*; British Crystallographic Association Spring Meeting, Nottingham, U.K., 2002.