Energy Ranking of Molecular Crystals Using Density Functional Theory Calculations and an Empirical van der Waals Correction

Marcus A. Neumann*,† and Marc-Antoine Perrin‡

Avant-garde Materials Simulation SARL, 30 bis rue du Vieil Abreuvoir, 78100 Saint Germain en Laye, France, and Aventis Pharma S.A., Centre de Recherche de Paris, 13 quai Jules Guesde, 94400 Vitry-sur-Seine, France

Received: January 7, 2005; In Final Form: May 23, 2005

By combination of high level density functional theory (DFT) calculations with an empirical van der Waals correction, a hybrid method has been designed and parametrized that provides unprecedented accuracy for the structure optimization and the energy ranking of molecular crystals. All DFT calculations are carried out using the VASP program. The van der Waals correction is expressed as the sum over atom−atom pair potentials with each pair potential for two atoms A and B being the product of an asymptotic $C_{6,A,B}/r^6$ term and a damping function $d_{A,B}(r)$. Empirical parameters are provided for the elements H, C, N, O, F, Cl, and S. Following Wu and Yang, the C_6 coefficients have been determined by least-squares fitting to molecular C_6 coefficients derived by Meath and co-workers from dipole oscillator strength distributions. The damping functions $d_{A,B}(r)$ guarantee the crossover from the asymptotic $C_{6,A,B}/r^6$ behavior at large interatomic distances to a constant interaction energy at short distances. The careful parametrization of the damping functions is of crucial importance to obtain the correct balance between the DFT part of the lattice energy and the contribution from the empirical van der Waals correction. The damping functions have been adjusted to yield the best possible agreement between the unit cells of a set of experimental low temperature crystal structures and their counterparts obtained by lattice energy optimization using the hybrid method. On average, the experimental and the calculated unit cell lengths deviate by 1%. To assess the performance of the hybrid method with respect to the lattice energy ranking of molecular crystals, various crystal packings of ethane, ethylene, acetylene, methanol, acetic acid, and urea have been generated with Accelrys' Polymorph Predictor in a first step and optimized with the hybrid method in a second step. In five out of six cases, the experimentally observed low-temperature crystal structure corresponds to the most stable calculated structure.

I. Introduction

The desire to predict the crystal structure of molecular compounds from the molecular geometry alone has triggered a substantial amount of research in the past and is motivated by the wide range of potential scientific and technological applications. Most solid-state properties such as density, hardness, color, morphology, or solubility crucially depend on the crystal packing. The ability to predict crystal structures would open up the door to the computation of solid-state properties with nothing more but the molecular geometry as a starting point, thus making it possible to select molecules for certain solid-state applications before they are even synthesized.

The state of the art in the field of polymorph prediction has been assessed by a series of blind tests organized by the Cambridge Crystallographic Data Center.[1,2] The blind tests reveal a fundamental shortcoming of all current approaches, which is the inability to calculate lattice energy differences with an accuracy that is higher than or at least comes close to typical energy differences of about 0.01−0.1 kcal/mol found between the low-energy crystal structures of a given molecule. Even though crystallization kinetics may complicate things further, it can be expected that in most cases only the stable polymorph and some of the metastable polymorphs with slightly higher lattice energies can actually be crystallized. As a consequence, polymorph prediction is bound to fail if the energy ranking is too inaccurate to allow for the reliable identification of the most stable crystal structures. In this paper, we present a new parametrization of a previously described hybrid approach for the calculation of lattice energies. To our knowledge, our parametrization is the first to offer the accuracy required for polymorph prediction.

Current methods used for lattice energy calculation can be roughly divided into empirical, semiempirical, and ab initio approaches. Empirical approaches use force fields which consist of a set of functional forms and parameters that have been fitted to experimental data and/or high-level ab initio calculations. If well parametrized, force fields can be appropriate for polymorph prediction,[1,2] but in general sufficiently accurate parameters are not readily available for most molecules. In contrast to force fields, ab initio calculations do not need to be parametrized, but involve the numerical solution of Schrödinger's equation for the electronic motion, thus requiring significantly more CPU time and computer memory. Since the exact solution of Schrödinger's equation is impossible for all but the simplest cases, various approximate techniques have been developed of which density functional theory (DFT) calculations are of particular interest, as they offer a good compromise between speed and accuracy for crystal structures with cell volumes of 100−1000 Å^3 and above. Unfortunately, the last statement is only true for salts and covalently bonded framework structures

* To whom correspondence may be addressed. E-mail: SARLAMS@aol.com. Phone: +33 6 25 05 33 29.
† Avant-garde Materials Simulation.
‡ Aventis Pharma.

10.1021/jp050121r CCC: $30.25
Published on Web 07/21/2005

such as zeolites. Because DFT calculations make use of the independent electron approximation, they fail to incorporate long-range dispersive interactions (van der Waals interactions) which result from electron correlation effects and play an important role in molecular crystals. Accordingly, pure DFT calculations are not appropriate for the structure optimization of molecular crystals.

Several attempts have been reported to overcome this problem. Some authors have tried to incorporate van der Waals interactions in the DFT formalism.[3−6] Although promising, these approaches either are currently too time-consuming for practical applications or suffer from other limitations. A more pragmatic approach, already explored by several authors[7−9] and also used in this work, is to combine DFT calculations with an empirical van der Waals correction. The empirical correction is defined as the sum over atom−atom pair potentials with each pair potential for two atoms A and B being the product of an asymptotic $C_{6,A,B}/r^6$ term and a damping function $d_{A,B}(r)$ (see eqs 3 and 5). The damping function is required to counteract the divergence of the $C_{6,A,B}/r^6$ term at short interatomic distances r. At large interatomic distances, the damping function is equal to 1.

It seems that the high potential of such a hybrid approach for the accurate energy ranking of molecular systems has not been realized in previous papers. Regarding the work of Elstner and co-workers,[7,8] this is not really surprising, since the authors use an approximate DFT method and put the emphasis on system size rather than accuracy. Wu and Yang,[9] on the other hand, use high-level DFT calculations combined with good-quality C_6 coefficients but fail to realize the importance of a careful parametrization of the damping functions. At first thought, one may judge that the damping functions are of secondary importance, because they intervene at short distances only, while the van der Waals correction involves a large number of long-range interactions. In fact, the opposite is true. Since the pair potentials decay as $1/r^6$, the van der Waals correction is dominated by interactions between neighboring atoms and strongly depends on the form of the damping functions. To look at this point from a different angle, let us consider an isolated pair of neutral atoms. At large interatomic distances, the DFT calculation yields zero and the empirical pair potential correctly reproduces the long-range behavior of the interaction energy. At short interatomic distances, the damping function guarantees that the contribution of the empirical pair potential to the total energy is negligible, while the DFT calculation correctly describes the short-rage repulsion between the two atoms. At intermediate distances close to the sum of the two van der Waals radii, however, both the DFT calculation and the empirical pair potential yield a significant energy contribution and the damping function needs to be chosen such that both contributions sum up to the correct total energy. In other words, the careful adjustment of the damping function is essential to obtain the right balance between the two energy contributions at intermediate interatomic distances.

In this work, we closely follow the strategy of Wu and Yang[9] for the derivation of the C_6 coefficients. In addition, we derive appropriate damping functions from a set of experimental low-temperature crystal structures.

The outline of this paper is as follows: Details of the hybrid method are presented in section II. In section III, we describe the parameter refinement of the C_6 coefficients and the damping functions. In section IV, we compare our structural results to those obtained from pure DFT calculations and from a hybrid approach using the empirical potentials proposed by Wu and Yang. In section V, we use our hybrid method to rank various crystal structures of ethane, ethylene, acetylene, methanol, acetic acid, and urea with respect to their lattice energies and we compare the predicted crystal structures to the corresponding experimental crystal structures. Finally, in section VI, we conclude.

II. The Hybrid Method

In our approach, we define the total lattice energy E_{tot} as the sum of a DFT component E_{DFT} and an empirical van der Waals correction E_{vdW}

$$E_{tot} = E_{DFT} + E_{vdW} \quad (1)$$

For the calculation of the DFT energy component and its derivatives with respect to atomic positions and lattice parameters, we use version 4.5 of the total energy and molecular dynamics program VASP (Vienna ab initio simulation program) developed at the "Institut für Materialphysik der Universität Wien".[10−13] Because VASP is used as a black box, we do not describe the VASP program in this paper. However, we briefly discuss the program settings chosen to run the calculations.

Standard projector-augmented wave (PAW) potentials[14] are used to describe the interaction between ions and electrons. Regarding the exchange correlation functional, we have tried several options, namely, the local density approximation (LDA) and three different versions of the generalized gradient approximation (GGA). LDA and GGA with the revised Perdew−Burke−Ernzerhof[15] functional were discarded right from the start, because—even without empirical potentials—they predict unrealistically strong binding forces for the crystals of simple molecular compounds such as N_2. The Perdew−Wang-91 functional[16,17] and the Perdew−Burke−Ernzerhof functional,[18] on the other hand, passed the initial test and reproduce the ordered phases of N_2 fairly well, if combined with an appropriate empirical correction. All results presented in this paper have been obtained with the Perdew−Wang-91 functional and the Vosko, Wilk, and Nusair interpolation formula.[20] The Perdew−Wang-91 functional was preferred over the Perdew−Burke−Ernzerhof functional, because the PAW potentials delivered with VASP have been parametrized using the Perdew−Wang-91 functional. For the Perdew−Burke−Ernzerhof functional, a detailed analysis of the performance of our hybrid approach has not yet been carried out. Concerning the size of the basis set, we use a plane wave cutoff energy of 520 eV and a k-point spacing[19] of roughly 0.7 $Å^{-1}$. For these values, the lattice energy differences between the various crystal packings of one and the same small molecules (ethane, methanol, urea) are typically converged to within 0.01−0.02 kcal/mol (throughout this paper, kcal/mol refers to 1 mol of molecules with 1 kcal = 4.19 kJ) or better and the Pulay pressure due to the finite cutoff energy is less than 0.1 GPa. To avoid confusion, it is important to stress that our numerical target accuracy of 0.01−0.02 kcal/mol does not refer to absolute lattice energies but to the lattice energy differences obtained for one and the same molecule. A problem arises when the k-point generation scheme attributes a single k-point to one or more reciprocal unit cell vectors of a given crystal structure. In that case, the curvature of the electronic band structure in the corresponding directions is not taken into account at all, resulting in a significant decrease of the accuracy of the calculated relative lattice energies. To avoid this effect without increasing the CPU time requirements too much, all lattice energy optimizations are followed by a single point energy calculation with at least 2 k-points for each direction of the reciprocal lattice.

We now turn our attention to the empirical correction term E_{vdW}. We define the van der Waals correction as the sum over individual atom−atom pair potentials, and for the van der Waals correction per unit cell we obtain

$$E_{VdW} = \sum_i \sum_j \sum_{\vec{k}}{}' \frac{1}{2} v_{t_i,t_j}(r_{i,j,\vec{k}}) \tag{2}$$

Here the first two sums run over all atoms in the unit cell and the third sum runs over all lattice vectors. The prime indicates that self-interactions ($i = j$ and $|\vec{k}| = 0$) are excluded. Each pair potential depends on the atom types t_i and t_j and the distance $r_{i,j}$, $\vec{k}$ between atom i, translated by $\vec{k}$, and atom j. The factor 1/2 avoids double counting. We use atom types that depend on the atomic number and the number of covalently bonded atoms. To truncate the infinite lattice sum in eq 2, we use a cutoff radius of 18 Å and a spline function that gradually switches off the pair potentials between 15 and 18 Å. A continuum correction ensures that the energy contribution from the neglected pair potentials is approximately taken into account.

Each pair potential is split into the product of a damping function $d_{A,B}(r)$ and a term that describes the asymptotic behavior at large interatomic distances

$$v_{A,B}(r) = d_{A,B}(r) \frac{C_{6,A,B}}{r^6} \tag{3}$$

The damping function must avoid the divergence of the pair potential at short interatomic distances. At large distances, the damping function should be equal to 1. Wu and Yang[9] have used a damping function originally proposed by Mooij at al.[21]

$$d_{A,B}(r) = \left(1 - \exp\left[-c_{damp}\left(\frac{r}{R_{A,B}{}'}\right)^3\right]\right)^2 \tag{4}$$

We slightly generalize this damping function by introducing a form factor n. In addition, we absorb the coefficient c_{damp} in the crossover distances $R_{A,B}$.

$$d_{A,B}(r) = \left(1 - \exp\left[-\left(\frac{r}{R_{A,B}}\right)^{3/n}\right]\right)^{2n} \tag{5}$$

$$R_{A,B} = \frac{R_{A,B}{}'}{c_{damp}{}^{n/3}} \tag{6}$$

For $n = 1$, both damping functions are identical. Equation 5 retains the main features of eq 4. Using either eq 4 or eq 5 in eq 3, the pair potential $v_{A,B}$ is a monotonic function of r and approaches a constant value at short interatomic distances (see Figure 1). The crossover from the asymptotic long-range r^{-6} behavior to the constant short-range behavior occurs in a region close to the crossover distance $R_{A,B}$. With decreasing value of the form factor n, the transition becomes more abrupt. We use the same form factor n for all damping functions.

To limit the number of independent empirical parameters, heteroatomic interaction parameters are derived from homoatomic interaction parameters by means of two combination rules. For the crossover distances $R_{A,B}$, we use an arithmetic combination rule

$$R_{A,B} = \frac{1}{2}(R_{A,A} + R_{B,B}) \tag{7}$$

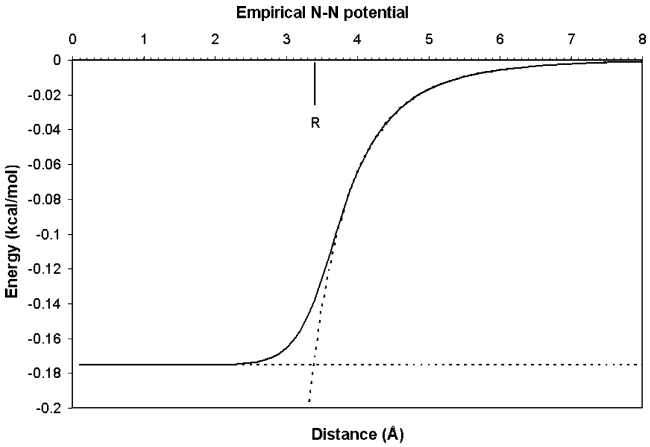

Figure 1. Empirical potential for a pair of nitrogen atoms after parameter refinement. The asymptotic behavior at short and large distances is also indicated. R is the crossover distance.

For the C_6 coefficients, we use the combination rule proposed by Wu and Yang

$$C_{6,A,B} = \frac{2(C_{6,A,A}{}^2 C_{6,B,B}{}^2 N_{eff,A} N_{eff,B})^{1/3}}{(C_{6,A,A} N_{eff,B}{}^2)^{1/3} + (C_{6,B,B} N_{eff,A}{}^2)^{1/3}} \tag{8}$$

where the parameters N_{eff} are the effective electron numbers according to Halgren[22] (see table deposited as Supporting Information).

The parameter refinement and the validation studies described in the following sections require the optimization of numerous crystal structures. Since the DFT component makes the calculation of energies and energy derivatives with the hybrid method very time-consuming, we have developed our own structure optimization program which is particularly suited for the crystal structures of small to medium size organic molecules.

The choice of the coordinate system is of particular importance as it affects the shape of the potential energy well around the energy minimum on the potential energy hypersurface. For the description of intramolecular degrees of freedom, we follow Baker, Kessi, and Delly[23] and use delocalized internal coordinates. Molecular displacements in the unit cell are described by whole molecule translations and rotations. Unit cell changes are defined in terms of an anisotropic compression or expansion of the crystal lattice, whereby the molecular centers follow the lattice deformation while the molecular geometries remain unchanged. Our choice of coordinates results in an increased harmonicity of the potential energy well such that crystal structures can be optimized very efficiently using a conjugate gradient or a quasi-Newton algorithm.[24] The quasi-Newton algorithm requires an initial estimate of the inverse Hessian matrix which is calculated using the Dreiding force field[25] with QEq charges.[26] Our optimization routine takes space group symmetry explicitly into account and can deal with atoms and molecules on special positions. All atomic positions are allowed to vary independently. By optimizing several crystal structures from different starting points, we have checked that our minimization routine in conjunction with the hybrid method is able to locate potential energy minima to within the desired accuracy.

With the hybrid method, we are aiming at an accuracy of about 0.01 kcal/mol for the calculation of relative lattice energies. Readers familiar with DFT calculations may object that this value is significantly smaller than the average error of several kcal/mol (ref 27) which is obtained when pure DFT

TABLE 1: Molecular C_6 Coefficients,[a] in Atomic Units (hartree*a_0^6), Obtained from Dipole Oscillator Strengths Distributions[31,32]

A	B	C_6	A	B	C_6
C_2H_2	CH_3OH	211.8	C_2H_2	HCl	163.0
C_2H_2	C_2H_5OH	328.7	C_2H_2	HBr	210.2
C_2H_2	C_3H_7OH	445.1	C_6H_6	HF	176.4
C_6H_6	CH_3OH	615.3	C_6H_6	HCl	473.5
C_6H_6	C_2H_5OH	954.9	C_6H_6	HBr	610.6
C_6H_6	C_3H_7OH	1293	C_2H_2	SH_2	210.1
HF	HF	19.0	C_2H_2	SF_6	332.3
HF	HCl	48.85	C_2H_2	CS_2	418.5
HF	HBr	62.45	C_2H_2	COS	286.4
HCl	HCl	130.4	C_6H_6	SH_2	610.1
HCl	HBr	167.8	C_6H_6	SF_6	966.3
HBr	HBr	216.6	C_6H_6	CS_2	1216
C_2H_2	HF	60.69	C_6H_6	COS	831.9

[a] We have used these data in conjunction with Table 1 from ref 9 for the refinement of atomic C_6 coefficients.

calculations are applied to the calculation of potential energy barriers for chemical reactions. Indeed, chemical reactions involve bond breaking as well as bond creation, and the approximate treatment of the exchange-correlation functional leads to relatively large error bars. It has to be kept in mind, though, that the chemical bonds in the different crystal packings of one and the same molecule are essentially the same. The energy differences between crystal polymorphs arise from changes of the molecular conformation and from intermolecular interactions, which can be seen as a small perturbation on top of the chemical bonding. For the purpose of lattice energy ranking, only these perturbations need to be taken into account accurately, since the chemical bond energies fall out of the equation.

III. Parameter Refinement

We have used two separate approaches to determine the C_6 coefficients on one hand and the form factor as well as the crossover distances on the other hand.

A. C_6 Coefficients. C_6 coefficients are, at least in principle, measurable quantities that can be determined independently of the precise nature of the DFT component, since the contribution of the later to the van der Waals energy is negligible at large interatomic distances. Following Wu and Yang,[9] we have fitted atomic C_6 coefficients to molecular C_6 coefficients derived from dipole oscillator strength distribution by Meath and co-workers.[28−34] The method employed by Meath and co-workers is a complex mixture of theoretical and experimental techniques.

The atomic C_6 coefficients depend on the hybridization state, which is closely related to the number of covalently bonded neighbors. Accordingly, we have defined the atom types H1 (hydrogen, one covalent bond), C2 (carbon, two bond partners, sp-hybridization), C3, C4, O1 (oxygen, one bond partner, sp_2-hybridization), and O2. For nitrogen, we have used a single atom type N_ because of the limited number of molecular C_6 coefficients available for nitrogen-containing molecules. Going beyond the work of Wu and Yang, we have also introduced the atom types F1, Cl1, and Br1 for fluorine, chlorine, and bromine as well as the atom types S1, S2, S3, S4, S5, and S6 for sulfur.

To obtain C_6 coefficients for the additional atom types, we have extended the data set used by Wu and Yang (see Table 1 of ref 9) by the values presented in Table 1. Molecular C_6 coefficients were calculated from atomic C_6 coefficients via

TABLE 2: Homoatomic C_6 Coefficients Derived from Molecular C_6 Coefficients[a]

atom type	C_6 Wu and Yang (hartree a_0^6)	C_6 this work (hartree a_0^6)	C_6 this work (kcal mol^{-1} Å^6)
H1	2.827	2.849	39.25
C2	29.76	29.91	412.2
C3	27.35	27.57	379.9
C4	22.06	21.91	301.9
N_	19.30	19.10	263.1
O1	12.95	12.79	176.3
O2	11.61	11.55	159.2
F		7.366	101.5
Cl		94.59	1303
Br		169.5	2335
S1		134.5	1853
S2		127.7	1760
S3			(1532)
S4			(1304)
S5			(1076)
S6		61.58	848.5

[a] Values in parentheses have been obtained by linear interpolation.

$$C_{6,\mathrm{mol1,mol2}} = \sum_{i\in \mathrm{mol1}} \sum_{j\in \mathrm{mol2}} C_{6,t_i,t_j} \tag{9}$$

where the two sums run over all atoms of the interacting molecules and t_i and t_j are the corresponding atom types. Heteroatomic C_6 coefficients were calculated from homoatomic C_6 coefficients according to eq 8. The homoatomic C_6 coefficients were adjusted as to minimize

$$F = \sum_i \left(\frac{C_{6,\mathrm{calc},i} - C_{6,\mathrm{exp},i}}{C_{6,\mathrm{exp},i}} \right)^2 \tag{10}$$

where the sum runs over all data points and the values of $C_{6,\mathrm{exp,i}}$ are taken from the publications of Meath and co-workers. Table 2 shows the optimized homoatomic C_6 coefficients together with the C_6 coefficients published by Wu and Yang. Our parameter refinement results in a root-mean-square deviation of 0.76%. The largest deviations between the calculated molecular C_6 coefficients and the target values, in the range of 2−3%, are found for pairs of molecules involving CS_2 or COS. For the elements H, C, N and O, our C_6 coefficients are essentially identical to the values obtained by Wu and Yang, all relative differences being smaller than 1.3%. This good agreement confirms the high stability of the refinement procedure already mentioned by Wu and Yang. It is important to note that only the C_6 coefficients of the sulfur atom types S1, S2, and S6 have been adjusted in our refinement. The values for all other sulfur atom types were obtained by linear interpolation.

B. Form Factor and Crossover Distances. Unlike the C_6 coefficients, the form factor and the crossover distances do not represent measurable quantities. Both types of parameters affect the empirical potentials predominantly in the region of short and intermediate interatomic distances, where it is important to obtain the right balance between the energy contributions of the DFT component and the empirical correction term. In general, the contribution of the DFT component to the van der Waals energy is difficult to quantify, since the DFT formalism results in a total energy that also comprises contributions from other sources such as covalent bonding, electrostatic interactions, and charge transfer. It must be stressed that the contribution from the DFT component depends on several factors such as

the choice of the basis set and the choice of the exchange-correlation functional. Therefore, an optimized set of empirical parameters is only valid in conjunction with the particular DFT approach that was used during the parameter refinement. In particular, the empirical parameters presented in this paper are only guaranteed to give good results in conjunction with the DFT calculations described at the beginning of section II.

It is worth mentioning that the empirical correction may implicitly cover more than just the missing part of the van der Waals interaction. Indeed, since the form factor and the cross over radii are adjusted to achieve best complementarity with the DFT component, any shortcomings of the later will be corrected for as long as they fit into the mathematical framework of the empirical correction. One may also hope that many body effects are implicitly taken into account to a certain extent, despite the fact that the empirical correction is based on pair potentials only. With regard to the above comments, one may suggest to use the neutral term "damped-C_6 correction" to describe the empirical part of the hybrid method. Nonetheless, we prefer the term "van der Waals correction", because it captures the physical meaning of the dominant contribution to the correction term.

In our refinement procedure, we adjust the form factor and the crossover distances to obtain the best possible agreement between the unit cells of a set of experimental low temperature crystal structures and the corresponding unit cells obtained by lattice energy minimization with the hybrid method. The choice of low-temperature crystal structures, ideally in the range from 0 to 50 K, is important to reduce the influence of lattice vibrations on the unit cell parameters. At temperatures close to the zero point, an experimental crystal structure essentially corresponds to the minimum of the lattice energy hypersurface, if we neglect small deviations that arise from the presence of zero-point vibrations in conjunction with anharmonicities of the potential energy well. Hence, our refinement procedure minimizes the deviation between the positions of the minima on the true lattice energy hypersurfaces and their theoretical counterparts. The molecular packing is very sensitive to changes of the empirical potentials, while the molecular geometries are hardly affected. We therefore limit our structural comparisons to unit cell changes as an indicator for molecular rearrangements.

To quantify the deviation between two unit cells by a single value, one may simply calculate the weighted sum over the squared deviations found for individual unit cell parameters. However, this approach has two disadvantages. First, there are two different types of parameters, cell lengths and cell angles, and the relative weighting of the two contributions is somewhat arbitrary. Second, unit cells can frequently be defined in several equivalent ways which do not result in the same value for the deviation between two similar unit cells. We therefore use a different approach that is independent of the unit cell choice and does not require the distinction between cell lengths and cell angles. The deviation between two unit cells is described in terms of an anisotropic expansion/compression along three mutually perpendicular axes and the overall difference is characterized by a single number that corresponds to the sum over the individual expansions/compressions expressed in percent. We use two slightly different measures Δ and $\tilde{\Delta}$ of the unit cell deformation, which are defined in more detail in appendix A.

In our refinement procedure, we minimize

$$F = \sum_i \tilde{\Delta}_i \tag{11}$$

where the sum runs over all pairs of experimental and calculated unit cells and $\tilde{\Delta}_i$ is defined in eq A8.

The parameter optimization based on eq 11 is not as straightforward as it may seem. We use the Powell algorithm[24] which does not require derivatives of eq 11 with respect to the adjustable parameters. At first sight, one may be tempted to carry out a full lattice energy minimization of all crystal structures for each new set of adjustable parameters. Unfortunately, a single crystal structure optimization requires typically 1 day of CPU time on a state-of-the-art PC. Consequently, the CPU time requirements get out of hand, even if the refinement is run on several PCs in parallel. To reduce the CPU time requirements, we use an iterative approach which exploits the fact that the time-consuming DFT component of the lattice energy hypersurface remains unchanged throughout the entire refinement, because only the parameters of the empirical van der Waals correction are affected by the refinement procedure.

Each iteration starts with a complete optimization of all crystal structures including lattice parameters and atomic positions. After the initial optimization, all molecules are treated as rigid bodies until the start of the next iteration. For each crystal structure, a set of points is chosen around the minimum of the lattice energy hypersurface such that the Hessian matrix can be constructed from the energy gradients at these points. The energy gradients of the DFT component are calculated and stored for the minimum and the surrounding points. Finally, the adjustable parameters are optimized by means of the Powell algorithm. Whenever the evaluation of eq 11 requires the optimization of a crystal structure for a new set of empirical parameters, we calculate the energy gradients of the empirical component at the initial energy minimum and the surrounding points and combine it with the stored DFT gradients to obtain the total energy gradients. From these, we construct the Hessian matrix H and we use the harmonic approximation to calculate the new lattice energy minimum according to

$$\vec{x}_{\text{new}} = \vec{x}_{\text{old}} + H^{-1}\vec{g} \tag{12}$$

where the coordinate vectors $\vec{x}_{new}$ and $\vec{x}_{\text{old}}$ comprise lattice changes as well as whole molecule translations and rotations. The vector $\vec{g}$ is the total energy gradient at $\vec{x}_{\text{old}}$. In general, two to four iterations are required to obtain convergence of the empirical parameters.

The set of crystal structures that we have used in our parameter refinement is presented in Table 3. For each crystal structure, Table 3 gives the compound name, the CSD reference code or a literature reference, the sum formula, the experimental temperature, the number of independent unit cell parameters, and the unit cell volume. The crystal structure of cubic acetylene has been measure at 143 K, but the unit cell parameters were extrapolated down to 0 K from the known temperature dependence between 134 and 176 K assuming a parabolic behavior ($a_{143K} = 6.102$ Å, $a_{0K,\text{extrapolated}} = 5.998$ Å). The crystal structure of γ-nitrogen has been determined at a pressure of 407 MPa. Accordingly, we have carried out lattice enthalpy optimizations including a PV term for this crystal structure. The last column of Table 3 indicates the unit cell deformation (see eq A7) for the experimental and the calculated unit cell after parameter refinement. The chosen structures cover a variety of chemical bonds and hydrogen bonding motifs. Sulfur atoms are present with coordination numbers ranging from 1 to 5.

We have refined a single form factor for all empirical potentials and one crossover distance per element. For bromine, the crossover distance has not yet been determined. As a starting point, we have taken the parameters of Wu and Yang, choosing

TABLE 3: Crystal Structures Used for Parameter Refinement

compound	reference	composition	T_{exp} (K)	N_{par}	V_{exp} (Å^3)	D (%)
benzene	BENZEN06	C_6H_6	15	3	462.5	2.5
butane	DUCKOB	C_4H_{10}	5	4	222.7	1.6
propane	JAYDUI	C_3H_8	30	4	364.9	2.6
durene	ref 35	$C_{10}H_{14}$	1.5	4	403.3	3.7
p-xylene	ref 36	C_8H_{10}	4	4	310.9	1.5
orthorhombic acetylene	ACETYL11	C_2H_2	15	3	208.2	5.0
cubic acetylene	ref 37	C_2H_2	143 → 0	1	215.8	0.4
acetic acid	ACETAC05	$C_2H_4O_2$	4	3	289.2	2.1
formic acid	FORMAC02	CH_2O_2	5	3	193.0	1.2
furan	BUNJAV02	C_4H_8O	5	4	401.4	2.6
methanol	METHOL02	CH_4O	15	3	200.5	6.8
terephthalic acid	TEPHTH03	$C_8H_6O_4$	2	6	169.0	4.1
hexamethylenetetramine	HXMTAM10	$C_6H_{12}N_4$	15	1	332.4	1.3
α-nitrogen	ref 38−39	N_2	20	1	181.4	1.3
γ-nitrogen	ref 40	N_2	20	2	80.0	1.9
nitromethane	NTROMA13	CH_3NO_2	4	3	275.3	3.4
N,*N*′-diformohydrazide	FOMHAZ16	$C_2H_4N_2O_2$	15	4	177.9	2.2
N-hydroxy-methaneimidamide	FORAMO01	CH_4N_2O	16	3	277.2	10.0
glyoxime	GLOXIM11	$C_2H_4N_2O_2$	9	4	178.6	5.9
urea	UREAXX12	CH_4N_2O	12	2	145.1	3.1
N-scetyl-L-cysteine	NALCYS02	$C_5H_9NO_3S$	16	6	174.7	1.6
dimethyl sulfide	ref 41	C_2H_6S	5	6	179.3	4.1
bis(trifuoromethyl)disulfane	RIKXEO	$C_2F_6S_2$	113	6	321.0	2.4
2,3-dichloropyrrole	MILMOJ	$C_4H_3Cl_2N$	88	3	257.4	0.9
1,4,4,4-tetrafluorobuta-1,3-diene	IDIXIC	$C_4H_2F_4$	120	4	238.3	2.7
chlorofluoromethane	IHERES	CH_2ClF	30K	4	133.9	2.4
O-methyl-thiocarbamate	CONFAM	C_2H_5NOS	123	4	419.4	3.3
(((fluoro)carbonyl)methylidene)sulfur tetrafluoride	FOPNUT	C_2HF_5OS	105	4	245.7	1.6
1,2-ethanedithiol	QEKXOT	$C_2H_6S_2$	130	3	455.3	4
thiourea *S*,*S*-dioxide	SURDOX02	CH_4N_2OS	110	3	412.0	3.9
fluorosulfonylmethylidene sulfur difluoride oxide	TUNWII	$CHF_3O_3S_2$	130	4	553.6	2.1

TABLE 4: vdW Radii[42] and Homoatomic Crossover Distances before and after Refinement[a]

element	vdW radius (Å)	crossover distance from Wu and Yang (Å)	crossover distance after refinement (Å)
H	1.2	1.575	3.200
C	1.7	2.231	3.884
N	1.55	2.034	3.384
O	1.52	1.995	2.837
F	1.47	1.929	2.988
Cl	1.75	2.297	3.419
S	1.80	2.362	4.205
		$n = 1.0$	$n = 0.25$

[a] The distances before refinement correspond to the damping functions of Wu and Yang. The form factor is also indicated.

$n = 1.0$ for the form factor and calculating the initial crossover distances according to eq 6 with $c_{damp} = 3.54$ and $R_{A,A}' = 2R_{A,vdW}$, where $R_{A,vdW}$ are the atomic van der Waals radii according to Bondi.[42] Table 4 lists the van der Waals radii, the initial crossover distances corresponding to the damping functions of Wu and Yang, and the crossover distances after parameter refinement. For the form factor n, our parameter refinement yields a value of 0.25.

Since the program development and the parameter refinement have proceeded in parallel in several phases, it is difficult to quote the exact CPU time requirements of the refinement process. We currently run our calculations on a cluster of eight PCs, each machine having an Intel Pentium 4 processor running at 2.8 GHz, a 800 MHz front side bus, and 1Gb of double channel PC3200 memory. With the latest version of our program, a complete parameter refinement from scratch would probably require 1 to 2 months.

In total, our data set contains 31 crystal structures with 109 independent lattice parameters compared to only 8 adjustable empirical parameters (1 form factor and 7 homoatomic crossover distances), and the refinement problem clearly is well determined. After parameter refinement, we obtain an average unit cell deformation of 3.0%. Despite the low number of adjustable parameters, the agreement between the optimized and the experimental crystal structures is excellent.

Initially, it has not been obvious at all that only eight adjustable parameters would be sufficient to obtain an acceptable result. After the positive outcome of our first parameter refinement, we have tried to further improve the agreement between the calculated and the experimental unit cells by increasing the number of adjustable parameters. Using either the complete set of structures or the hydrocarbon subset, we have explored several possibilities, including the joint refinement of crossover distances and C_6 coefficients, the refinement of one crossover distance per atom type, the refinement of independent parameters for hydrogen bonds, and the use of a more complex combination rule with one adjustable parameter instead of eq 7. All these attempts have in common that there was no significant improvement of the overall agreement, while many attempts resulted in unrealistic crossover distances or C_6 coefficients. Our conclusion is that the potential of our empirical correction based on atom−atom pair potentials is fully exploited using C_6 coefficients derived from dipole oscillator strength distributions, a single form factor, and one crossover distance per element. The remaining discrepancies are not necessarily related to an incorrect description of the van der Waals interactions. According to Table 3, the biggest unit cell deformations of 10.0%, 6.8%, and 5.9% are observed for *N*-hydroxy-methaneimidamide, methanol, and glyoxime, three strongly hydrogen bonded crystals, while crystal structures dominated by van der Waals interactions such as benzene (2.5%), nitromethane (3.4%), or hexamethylenetetramine (1.3%) have significantly smaller unit cell deviations. It seems that the remaining structural differences are in part related to a certain

TABLE 5: Unit Cell Deformations and Relative Volume Errors Obtained with Three Different Methods: Pure DFT, Hybrid Method with Refined Parameters (this work), Hybrid Method with Parameters from Wu and Yang[a]

	Δ (%)			ΔV_{rel} (%)		
compound	pure DFT	this work	Wu & Yang	pure DFT	this work	Wu & Yang
benzene	47	2.5	22.2	32.2	0.7	−19.9
butane	24.1	1.6	32.8	27.2	−0.6	−27.9
propane	22.6	2.6	37.7	25.3	−2.6	−31.4
durene	21.4	3.7	23.5	23.8	−0.9	−20.9
p-xylene	24.5	1.5	22.8	27.8	−0.1	−20.4
orthorhombic acetylene	20.5	5	20.8	22.2	−1.8	−18.7
cubic acetylene	15.3	0.4	16.3	16.5	−0.4	−15.0
acetic acid	22	2.1	13.4	19.0	−0.03	−12.5
formic acid	23.9	1.2	12.4	22.2	−1.2	−11.7
furan	20.7	2.6	25.7	23.0	−0.4	−22.6
methanol	21.4	6.8	24.8	14.3	−3.2	−21.9
terephthalic acid	33.2	4.1	12.7	25.3	−0.1	−12.0
hexamethylenetetramine	15.6	1.3	19.7	16.8	−1.3	−17.9
α-nitrogen	19.2	1.3	7.7	21.1	−1.2	−7.4
γ-nitrogen	13.9	1.9	7.9	14.9	0.7	−7.6
nitromethane	15.3	3.4	11.2	16.5	0.1	−10.6
N,*N*′-diformohydrazide	25.1	2.2	12.2	22.8	−1.2	−11.5
N-hydroxy-methaneimidamide	14.2	10	17.5	13.1	−2.3	−16.0
glyoxime	23.9	5.9	16.3	18.6	−1.9	−15.0
urea	5.5	3.1	14.2	5.6	−3.1	−13.2
average	21.465	3.16	18.59	20.4	−1.0	−16.7

[a] The relative volume changes are calculated according to $\Delta V_{rel} = (V_{calc} - V_{exp})/V_{exp}$.

lack of accuracy of our DFT calculations with respect to hydrogen bonding.

The form factor on one hand and the crossover distances on the other hand are strongly correlated. Our final value of 0.25 for the form factor corresponds to a rapid transition at the crossover distance from the long-range r^{-6} behavior to the constant short-range part of the empirical potential. The crossover distances are significantly larger than typical covalent bond lengths, which are therefore hardly affected by the additional empirical potentials. As an example, Figure 1 shows the empirical potential for a pair of nitrogen atoms.

IV. Structural Results

In the previous section, we have seen that crystal structure optimizations with the hybrid method typically result in unit cell deformations of about 3% compared to experimental low-temperature crystal structures. Accordingly, for a randomly chosen direction, the optimized unit cell will typically deviate from the corresponding experimental cell by 1%. Similarly, calculated and experimental unit cell lengths disagree by about 1% on average. To fully appreciate these results, they need to be compared to the performance of pure DFT calculations on one hand and to hybrid energy calculations with parameters from Wu and Yang on the other hand.

Since Wu and Yang only have published empirical parameters for H, C, N, and O, we have limited the comparison to the corresponding subset of crystal structures in Table 3. Table 5 lists the unit cell deformations and relative volume errors obtained from pure DFT calculations, the hybrid method with our parameters, and the hybrid method with parameters from Wu and Yang. In all three cases, the DFT calculations have been carried out as described at the beginning of section II. For the empirical correction of Wu and Yang, we have used eq 5 with $n = 1$ and the parameters from the second column of Table 2 and the third column of Table 4.

With our own parameters, the hybrid method yields an average unit cell deformation of 3.2%, compared to 21.5% for pure DFT calculations and 18.6% for the hybrid approach with the parameters of Wu and Yang. Regarding the unit cell volume, pure DFT calculations result in unit cells that are too large by 20.4% as expected, because the attractive part of the van der Waals interaction is missing. With empirical parameters from Wu and Yang, on the other hand, the unit cell volume is typically underestimated by 16.7%, since the crossover distances are too short and the resulting attractive forces are too strong. In contrast, our own hybrid approach with fully refined parameters underestimates the unit cell volumes by only 1% on average. The above comparisons illustrate that our hybrid approach makes it possible to reproduce crystal structures of molecular compounds with unprecedented accuracy. Compared to pure DFT calculations, unit cell errors are typically reduced by 1 order of magnitude.

To examine the effect of the empirical correction on the atomic coordinates, we have optimized the atomic positions for all compounds in Table 3, starting from the experimental crystal structures and keeping the unit cell parameters constant. For each compound, the root-mean-square deviation between the calculated and the experimental positions of all non-hydrogen atoms was evaluated. With the hybrid method, we obtained root-mean-square deviations in the range from 0.005 to 0.12 Å around an average value of 0.041 Å. Pure DFT calculation performed marginally worse, yielding an average root-mean-square deviation of 0.045 Å. It can be concluded that the empirical correction hardly affects the accuracy of the calculated atomic coordinates.

V. Energy Ranking

Since we have used only structural data to refine the empirical parameters, it cannot be taken for granted that the hybrid method is appropriate for accurate lattice energy calculations, even though one may argue that crystal structures result from a fine balance between all interactions involved, so that lattice energies obtained with the hybrid method cannot be completely wrong if the experimental crystal structures are correctly reproduced.

The task of assessing the energy accuracy of the hybrid method is far from trivial. The available experimental data are mostly free energy differences between different solid or solid and liquid phases at fairly high temperatures. The computation of these free energy differences requires the accurate determination of all vibrational modes, a job which is not feasible with

our current software and hardware because of the important CPU time requirements of the hybrid method.

To avoid the calculation of free lattice energies, we again focus on low-temperature crystal structures. Our working assumption is that small and fairly rigid molecules will always find a way to adopt the most stable crystal structure at any temperature, if the cooling rate is not too fast. In other words, the lattice energy of the experimental low-temperature crystal structure should be lower than the lattice energy of any other molecular arrangement in most cases. Since lattice vibrations are not completely negligible, even at very low temperature, because of the presence of zero-point vibrations, it may sometimes happen that the correct energy ranking can only be reproduced if free lattice energies are calculated rather than lattice energies. However, in the low temperature limit, lattice energy differences and free lattice energy differences are almost the same, and the energy difference between the lattice energy of the most favorable molecular packing and the lattice energy of the experimental low temperature crystal structure after lattice energy minimization can be expected to be either zero or rather small in all cases, provided that the energy calculations are sufficiently accurate.

To test the energy accuracy of the hybrid method, we have generated a large number of crystal packings for acetylene, ethylene, ethane, methanol, acetic acid, and urea using Accelrys' Polymorph Predictor in Cerius[2] (ref 43). The first three molecules have been chosen because they are the simplest model systems for single, double, and triple covalent bonds. Their crystal structures are dominated by van der Waals interactions. For the last three molecules, hydrogen bonding yields an important contribution to the lattice energy. An important factor for the selection of the six tests molecules was their small size, which favorably affects the computational cost. The use of acetylene for both the parameter refinement and the energy ranking study is somewhat problematic, because it leaves the transferability of the empirical correction unchecked for the C2 force field type. The above concern also applies to methanol, acetic acid, and urea, even though to a much lesser extend. In fact, the last three compounds share their force field types with many other structures used for parameter refinement. Discarding these three compounds would only have a marginal effect on the resulting empirical parameters.

For each molecule, the Polymorph Predictor was run twice with fine settings for the 17 most frequent space groups, which cover roughly 95% of all experimentally observed crystal structures. The CVFF force field[44] was used for ethylene, ethane, methanol, acetic acid, and urea. The Polymorph Predictor in conjunction with the CVFF force field performs extremely well in these five cases and the experimental crystal structure is always among the two most stable predicted crystal structures. For acetylene, the CVFF force field fails to optimize the experimental crystal structure correctly and we have used the Universal Force Field[45] with QEq[26] charges instead. The case of acetylene illustrates one of the major shortcomings of force field based energy calculations, namely, that each of the currently existing force fields has a limited range of applicability and may perform well for some molecules while it completely fails for others.

After structure generation with the Polymorph Predictor, we have reranked the most stable crystal structures found for each molecule with the hybrid method. When reranking crystal structures, one typically would like to be sure to have found either the N most stable crystal structures or all crystal structures in a certain energy window above the most stable one. Since accurate energy calculations are time-consuming, the question is for how many of the force field predicted crystal structures the lattice energies should be minimized again using the hybrid method.

Let us suppose that we have a list of m ranked crystal structures for which lattice energies E_i have been calculated using a fast but inaccurate method. The lattice energies E_i increase with the index i. Proceeding from low to high lattice energies E_i, we reoptimize the crystal structures with a more accurate method and obtain new lattice energies E_i'. If we already have calculated new lattice energies for all $i \leq m'$, we can determine the average energy shift and its standard deviation

$$\overline{\Delta E} = \sum_{i=1}^{m'} \frac{1}{m'}(E_i' - E_i) \tag{13}$$

$$\sigma_{\Delta E} = \sqrt{\frac{1}{m'-1}\sum_{i=1}^{m'}(E_i' - E_i - \overline{\Delta E})^2} \tag{14}$$

If we assume that the individual energy shifts follow a Gauss distribution, the probability that all remaining crystal structures will have new lattice energies that are greater than a certain energy limit E_{lim} is given by

$$p(E_i' > E_{\mathrm{lim}}\ \forall\ i > m') = \prod_{i=m'+1}^{m}\left(\frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{E_i + \overline{\Delta E} - E_{\mathrm{lim}}}{\sqrt{2}\sigma_{\Delta E}}\right)\right) \tag{15}$$

If we want to find all structures in a certain energy window, E_{lim} is the sum of the lowest new lattice energy and the size of the energy window. If we are interested in finding the N most stable crystal structures instead, E_{lim} is the Nth lowest value of the new lattice energies. We can stop optimizing further crystal structures when $p(E_i' > E_{\mathrm{lim}}\ \forall\ i > m')$ is greater than a certain target probability.

In the case of acetylene, we wanted to be 99% sure to have optimized the 20 most stable crystal structures. In total, 122 crystal structures had to be considered. The lattice energies found after reranking with the hybrid method in a 1 kcal/mol energy window above the most stable crystal structure are shown in Figure 2. The most stable predicted crystal structure corresponds to the experimental low-temperature crystal structure measured at 15 K. The second most stable structure (rank 2) is very similar to the most stable one, and the lattice energies of the two structures differ by only 0.025 kcal/mol. Both crystal structures consist of identical layers, but the stacking of the layers is different in the two cases (ABAB versus ABCDABCD). Rank 3 is a crystal structure between the two experimentally observed phases of acetylene. It is interesting to note that the crystal structure predicted with the Polymorph Predictor which yields the rank 3 structure is very close to the cubic high-temperature phase of acetylene. However, this agreement is lost after crystal structure optimization with the hybrid method, probably because the high-temperature phase corresponds to a saddle point rather than a minimum of the potential energy hypersurface. Rank 3 is 0.20 kcal/mol less stable than rank 1, compared to 0.43 kcal/mol for the cubic high-temperature phase after lattice energy minimization in the space group $Pa\bar{3}$. The two most stable crystal structures of acetylene are shown in Figure 3.

In the case of ethylene, ethane, methanol, acetic acid, and urea, we have optimized 15, 15, 15, 20, and 11 crystal structures,

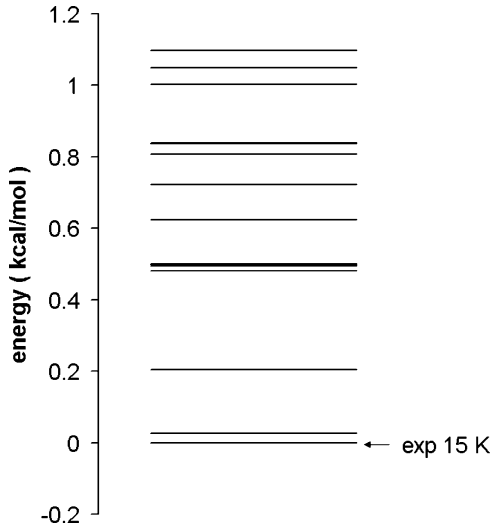

Figure 2. Lattice energies obtained for acetylene after crystal structure optimization with the hybrid method. The arrow indicates the experimental low-temperature crystal structure.

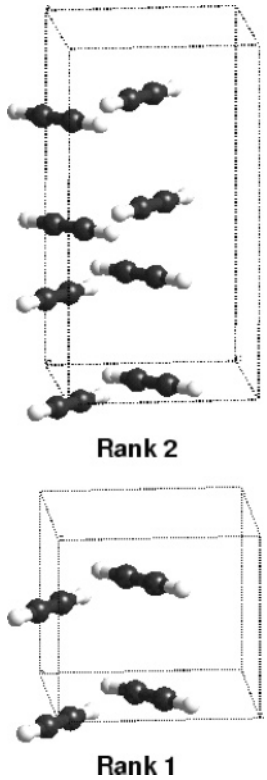

Figure 3. The two most stable calculated crystal structures obtained for acetylene.

respectively, without applying the probabilistic approach described in this section. Since the energy ranking of the Polymorph Predictor seems to be fairly reliable in these five cases, it is unlikely that we have missed any relevant crystal structures. For all six molecules, Table 6 lists the temperature at which the experimental low-temperature crystal structure has been determined, the force field used for polymorph prediction, the rank of the experimental crystal structure according to the force field, the rank obtained with the hybrid method, and the energy difference between the two most stable crystal structures, again calculated with the hybrid method.

The most striking result is the fact that the hybrid method finds the experimental low-temperature crystal structure as the most stable crystal structure in five out of six cases. Because

TABLE 6: Rank of the Experimental Low-Temperature Crystal Structure[a]

compound	T_{exp} (K)	force field	rank with force field	rank with hybrid method	ΔE_{12} (kcal/mol)
acetylene	15	UFF	12	1	0.025
ethylene	85	CVFF	1	2	0.005
ethane	85	CVFF	1	1	0.002
methanol	15	CVFF	1	1	0.012
acetic acid	4	CVFF	2	1	0.064
urea	12	CVFF	2	1	0.123

[a] ΔE_{12} is the energy difference between rank 1 and rank 2 calculated with the hybrid method.

of the general nature of the hybrid method, it can be expected to perform equally well for other molecules, thus offering a widely applicable solution to the energy ranking problem in the field of polymorph prediction. In addition, the high success rate confirms our assumption that the experimentally observed low-temperature crystal structures correspond to the molecular packing with the lowest lattice energy in most cases.

Another important result is the small size of the energy differences between the two most stable crystal structures. A more detailed analysis reveals that the two most stable crystal structures are very similar in all cases except for urea. For acetylene, ethylene, ethane, and acetic acid, the two most stable crystal structures consist of identical layers that are stacked differently. In the case of methanol, the two most stable structures are built up of one-dimensional hydrogen bonded chains with identical arrangements of the methyl groups and the oxygen atoms. The only difference is the direction of the hydrogen bonds in neighboring chains. In the case of urea, the two most stable crystal structures are truly different and their lattice energy difference is relatively large.

It is interesting to compare the observed lattice energy differences to the expected numerical accuracy of about 0.01−0.02 kcal/mol of the hybrid method (see section II). For ethylene and ethane, the calculated energy differences are significantly lower than the expected accuracy and the energy ranking with the hybrid method is wrong in one out of two cases. For all other molecules, the energy differences are slightly higher than or at least in the same range as the expected accuracy and the energy ranking with the hybrid method is correct.

Even though the effect of zero-point vibrations could be expected to be small, the excellent success rate of our approach is slightly surprising. For instance, it is well-known that changes of the hydrogen bond stretch frequencies by ± 100 cm^{-1} depending on the chemical environment are not uncommon. Accordingly, one may expect that the neglect of zero-point vibrations introduces an error of ± 0.14 kcal/mol ($=50$ cm^{-1}) per hydrogen bond. The high success rate of our approach despite these uncertainties may be explained by a compensation of errors.

With respect to practical applications, the small size of some of the energy differences is somewhat problematic, because it limits our ability to predict the most stable crystal structure. However, the situation is much more favorable than it may seem at first sight. One important application of polymorph prediction lies in the field of pharmaceutical research, where it is crucial to make sure that there is no other polymorph which is significantly less soluble and thus less bioavailable than the crystal forms that have already been observed experimentally. The stable polymorph always has the lowest solubility. Let us suppose that there are several low-energy crystal structures with similar crystal packings and virtually identical lattice energies. If none of these structures has been observed experimentally,

the most stable and least soluble crystal form clearly is still missing and it is advisable to carry out further crystallization experiments. If, on the other hand, one of the low-energy crystal structures has already been crystallized, it is unclear if the experimental crystal structure corresponds to the stable polymorph, but this ambiguity is tolerable, because the other low-energy crystal structures essentially have the same solubility. In this case, the calculation also tells us that there are probably crystallization conditions that favor the selective crystallization of other low-energy crystal structures or the simultaneous crystallization of several low-energy forms.

We close the present section with two unrelated remarks. Unlike the experimentally observed low-temperature crystal structure of methanol, the second most stable crystal structure of methanol has a permanent dipole moment. Therefore, one may try to obtain this crystal structure by crystallization in a strong electrostatic field. Our second remark is related to the question why nobody has ever observed a crystal structure of acetic acid with a dimmer motif. Our results indicate that the most stable dimmer packing is less stable by 0.71 kcal/mol than the experimental low-temperature crystal structure.

VI. Conclusion

In this paper, we have shown that the crystal structures and the relative lattice energies of molecular crystals can be computed fairly accurately by combining high level DFT calculations with a properly adjusted empirical van der Waals correction. All present results indicate that our hybrid approach provides the accuracy required for polymorph prediction. However, six energy ranking studies are surely not enough to prove its general applicability. Further validation work is currently under way, including energy ranking studies for the six molecules of the two CCDC blind tests[1,2] on polymorph prediction. In parallel, we are also working on improvements of our hardware and software to be able to deal with more complex crystal structures. In particular, it still needs to be demonstrated that the hybrid approach can provide satisfactory results for flexible molecules with various stable conformations.

The excellent energy ranking results obtained with our hybrid method are also interesting from a more fundamental point of view. After years of virtually no progress in the field of polymorph prediction, certain voices claim that the current inability to reliably predict crystal structures is not related to the inaccuracy of the energy calculations employed but to the complete neglect of crystallization dynamics. Contrary to this point of view, our results support the idea that the relative lattice energies are the dominant factor for the experimental appearance of polymorphs, at least as far as small and rigid molecules are concerned.

It is important to realize that the high accuracy of the hybrid method with respect to lattice energy differences does not necessarily imply that it is suited equally well for the calculation of absolute lattice energies. To elucidate this point, we are planning to carry out a comparison of experimental and calculated sublimation enthalpies. Because the determination of sublimation enthalpies requires calculations for both molecular crystals and isolated molecules, further software development is necessary before we can address this interesting issue.

The hybrid method described in this paper provides more accurate lattice energy differences than any other method applicable to relatively large molecular systems. Despite this fact, it must not be forgotten that the experimental order of stability depends on free lattice energy differences rather than lattice energy differences. The accurate determination of the effect of zero-point vibrations and temperature on the energy ranking remains an important challenge for the future.

At present, it is very difficult to reliably assess the accuracy of the hybrid method with respect to the calculation of lattice energy differences. While we know that our numerical error is of the order of 0.01−0.02 kcal/mol, the additional error due to the inherent inaccuracy of the hybrid method is difficult to quantify. In principle, the overall accuracy can be derived from the comparison of the experimental and the calculated relative stabilities for a sufficiently large set of polymorphic crystal structures. However, such a comparison only makes sense if free lattice energy differences are calculated instead of lattice energy differences, because otherwise the observed discrepancies are likely to be dominated by the neglect of entropy effects and zero-point vibrations rather than the shortcomings of the hybrid method itself. As a result of our current inability to calculate free lattice energies with the hybrid method, a thorough analysis of the attainable accuracy needs to be postponed until later. Nevertheless, a rough estimate can be derived from the results presented in Table 6. For methanol, acetylene, acetic acid, and urea, our energy ranking is correct, while the energy differences between the stable and the first metastable polymorph range from 0.012 to 0.12 kcal/mol. It thus seems that for small molecules the relative stability in the zero temperature limit can be predicted with an accuracy of less than 0.1 kcal/mol, despite the neglect of zero-point vibrations.

Acknowledgment. The authors thank Aventis Pharma France for the generous funding provided for the project.

Appendix A

Let $\vec{a}_1$, $\vec{b}_1$, and $\vec{c}_1$ as well as $\vec{a}_2$, $\vec{b}_2$, and $\vec{c}_2$ be the lattice vectors of two similar unit cells. For each unit cell, the transformation from fractional coordinates to Cartesian coordinates is described by a matrix

$$L_i = (\vec{a}_i, \vec{b}_i, \vec{c}_i) \tag{A1}$$

where the lattice vectors are the column vectors of the matrix. The first set of lattice vectors is mapped onto the second set of lattice vectors by the transformation matrix

$$T_{1\rightarrow 2} = L_2 L_1^{-1} \tag{A2}$$

According to the single value decomposition theorem, $T_{1\rightarrow 2}$ can be decomposed into the product of an orthonormal matrix Q, a diagonal matrix D, and an orthonormal matrix U

$$T_{1\rightarrow 2} = QDU \tag{A3}$$

This equation can be rewritten as follows

$$T_{1\rightarrow 2} = RS \tag{A4}$$

$$R = QU \tag{A5}$$

$$S = U^T DU \tag{A6}$$

S is a symmetric matrix with three orthogonal eigenvectors and eigenvalues given by the diagonal elements d_1, d_2, and d_3 of the matrix D. The multiplication of the lattice vectors $\vec{a}_1$, $\vec{b}_1$, and $\vec{c}_1$ by the matrix S corresponds to an anisotropic deformation of the unit cell along three mutually perpendicular axes. R is a rotation matrix and does not change the geometry of the unit

cell. As a measure for the deviation between two unit cells, we define the unit cell deformation

$$\Delta = \frac{1}{2}\left(\sum_{i=1}^{3}\left|d_i - 1\right| + \sum_{i=1}^{3}\left|\frac{1}{d_i} - 1\right|\right) \qquad \text{(A7)}$$

If the two unit cells are identical, we obtain $d_1 = d_2 = d_3 = 1$ and $\Delta = 0$. In all other cases, Δ is greater than zero. A relative compression or expansion along a single direction by 1% yields a deformation Δ of roughly 1%, while a relative change of 1% along three perpendicular directions results in a deformation of approximately 3%. By the use of two separate sums in eq A7, we achieve that the two transformations $T_{1\rightarrow 2}$ and $T_{2\rightarrow 1}$ result in the same unit cell deformation. Thus eq A7 provides a well-defined, quantitative measure for the overall deviation between two unit cells. All unit cell deviations quoted in this paper have been calculated according to eq A7.

While the quantitative meaning of eq A7 is easy to grasp, it is inconvenient for the computational refinement of parameters because of the discontinuities of its first derivatives at $d_i = 1$. To avoid this disadvantage, we define a second measure

$$\tilde{\Delta} = \frac{1}{2}\left(\sum_{i=1}^{3}(d_i - 1)^2 + \sum_{i=1}^{3}\left(\frac{1}{d_i} - 1\right)^2\right) \qquad \text{(A8)}$$

Supporting Information Available: Table of effective number of electrons according to Halgren. 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.; Schweitzer, B.; Schmidt, M. U.; van Eijck, B. P.; Verver, P.; Williams, D. E. *Acta Crystallogr., Sect. B*: *Struct. Sci.* **2000**, *56*, 697.

(2) Motherwell, D. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W. M.; Leussen 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*: *Struct. Sci.* **2002**, *58*, 647.

(3) Rapcewiez, K.; Ashcroft, N. W. *Phys. Rev. B* **1991**, *44*, 4032.

(4) Andersson, Y.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **1996**, *76*, 102.

(5) Dobson, J. F.; Dinte, B. P. *Chem. Phys. Lett.* **1996**, *76*, 1780.

(6) Kohn, W.; Meir, Y.; Makarov, D. E. *Phys. Rev. Lett.* **1998**, *80*, 4153.

(7) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras E. *J. Chem. Phys.* **2001**, *114*, 5149.

(8) Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 29.

(9) Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *116*, 515.

(10) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47*, 558.

(11) Kresse, G.; Hafner, J. *Phys. Rev. B* **1994**, *49*, 14251.

(12) Kresse, G.; Furthmüller, *J. Comput. Mater. Sci.* **1996**, *6*, 15.

(13) Kresse, G.; Furthmüller, *J. Phys. Rev. B* **1996**, *54*, 11169.

(14) Kresse, G.; Joubert, D. *Phys. Rev. B* **1999**, *59*, 1758.

(15) Hammer, B.; Hansen, L. B.; Nørskov, J. K. *Phys. Rev. B* **1999**, *59*, 7413

(16) Wang, Y.; Perdew, J. P. *Phys. Rev. B* **1991**, *44*, 13298.

(17) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671.

(18) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865

(19) We use a Γ centred Monkhorst−Pack grid with the number N_i of k-points along the reciprocal lattice vector $\bar{b}_i$ given by $N_i = \max(1, l^*|\bar{b}_i| + 0.5)$ with $l = 15$.

(20) Vosko, S. H.; Wilk, L.; Nursair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(21) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveld-van de Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872.

(22) Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827.

(23) Baker, J.; Kessi, A.; Delley, B. *J. Chem. Phys.* **1996**, *105*, 192.

(24) Press: W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C++*; Cambridge University Press: Cambridge, 2002; Chapter 10.

(25) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III *J. Phys. Chem.* **1990**, *94*, 8897.

(26) Rappé, A. K.; Goddard, W. A., III *J. Phys. Chem.* **1991**, *95*, 3358.

(27) Zhao, Y.; Pu, J.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2004**, *6*, 673.

(28) Margoliash, D. J.; Meath, J. W. *J. Chem. Phys.* **1978**, *68*, 1426.

(29) Jhanwar, B. I.; Meath, W. J. *Mol. Phys.* **1980**, *41*, 1061.

(30) Jhanwar, B. I.; Meath, W. J. *Chem. Phys.* **1982**, *67*, 185.

(31) Kumar, A.; Meath, W. J. *Mol. Phys.* **1985**, *54*, 823.

(32) Kumar, A.; Meath, W. J. *Mol. Phys.* **1992**, *75*, 311.

(33) Burton, G. R.; Chan, W. F.; Cooper, G.; Brion, C. E.; Kumar, A.; Meath, W. J. *Can. J. Chem.* **1994**, *72*, 529.

(34) Kumar, A.; Meath, W. J. *Mol. Phys.* **1997**, *90*, 389.

(35) Neumann, M. A.; Johnson, M. R.; Radaelli, P. G.; Trommsdorff, J. P.; Parker, S. F. *J. Chem. Phys.* **1999**, *110*, 516.

(36) Prager, M.; David, W. I. F.; Ibberson, R. M. *J. Chem. Phys.* **1991**, *95*, 2473.

(37) McMullan, R. K.; Kvick, Å.; Popelier, P. *Acta Crystallogr., Sect. B*: *Struct. Sci.* **1992**, *48*, 726.

(38) Hörl, E. M.; Marton, L. *Acta Crystallogr.* **1961**, *14*, 11.

(39) Donohue, J. *Acta Crystallogr.* **1961**, *14*, 1001.

(40) Schuch, A. F.; Mills, R. L.; *J. Chem. Phys.* **1970**, *52*, 6000.

(41) Ibberson, R. M.; McDonald, P. J.; Pinter-Krainer, M. *J. Mol. Struct.* **1997**, *415*, 259.

(42) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.

(43) The Polymorph Predictor and Cerius2 are products of Accelrys Inc., 9685 Scranton Rd, San Diego, CA 92121-3752, USA.

(44) Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319.

(45) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.