Crystal Structure Prediction by Global Optimization as a Tool for Evaluating Potentials: Role of the Dipole Moment Correction Term in Successful Predictions†

Jaroslaw Pillardy, Ryszard J. Wawak, Yelena A. Arnautova, Cezary Czaplewski, and Harold A. Scheraga*

Contribution from the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301

Received August 18, 1999. Revised Manuscript Received November 29, 1999

Abstract: A recently proposed method for surmounting the multiple-minima problem in protein folding is applied here to the prediction of crystal structures by global optimization of a potential energy function. The method, self-consistent basin-to-deformed-basin mapping, locates a group of large basins (regions of attraction of single minima) containing low-energy minima in the original energy surface, by coupling these groups of minima in the original surface to basins in a highly deformed energy surface, which contains a significantly reduced number of minima. The experimental crystal structures of formamide, imidazole, and maleic and succinic anhydrides were predicted as the global minima of the AMBER potential and were found among the lowest-energy minima for the DISCOVER potential. The results of the predictions serve as tests for evaluating the two potentials and may serve as a guide for potential refinements. Another important goal of this study was to clarify the role of the dipole moment contribution in calculations of the crystal electrostatic energy when the dipole moment of the unit cell is nonzero. Contrary to some practices, it is suggested that the use of the Ewald summation formula alone, without correcting for the dipole moment of the unit cell, is not the proper way to compute the electrostatic energy of a crystal and may lead to wrong predictions.

1. Introduction

The problem of crystal structure prediction has attracted considerable attention in recent years.[1−13] This is not surpising because the possibility of crystal structure prediction is of great importance for many branches of theoretical and applied chemistry. Crystal structure prediction plays an important role in many fields in which the design of new organic solids with desired physical properties is involved.[1,6,14,15] Another important consideration in crystal structure prediction theory is the problem of polymorphs.[2−4,7] In view of the fact that the properties of crystalline polymorphs may depend strongly on their crystal structure, knowledge of all possible crystal structures for a given molecule is critical for production, application, and storage of a wide range of organic compounds from drugs to energetic materials. A successful method for crystal structure prediction would also lead to a better understanding of intermolecular forces and of processes occurring during crystallization and crystal growth. In principle, the prediction should identify the structure that has the lowest free energy under appropriate conditions. Since this is computationally impractical, the conventional simplification is to search for the global minimum of the potential energy. With this simplification, the prediction of possible crystal structures requires an effective method for finding all significant low-energy minima of the potential energy, as well as a reliable potential that can reproduce the main features of the crystal energy and structure.

Despite much effort, no really efficient method for global optimization of crystal potential energy has been developed. The main problems are the existence of a very large number of local minima on the potential energy surface and its high dimensionality. Other obstacles include the large number of interatomic interactions that must be considered in the energy computations[16] and the need to use a mathematical device such as the Ewald summation[17] to calculate the electrostatic energy.

* To whom correspondence should be addressed. Phone: (607) 255-4034. Fax: (607) 254-4700. E-mail: has5@cornell.edu.

† Presented at the XVIII International Crystallography Congress, Glasgow, August, 1999.

(1) Dzyabchenko, A. V.; Pivina, T. S.; Arnautova, E. A. *J. Mol. Struct.* **1996**, *378*, 67.

(2) Chaka, A. M.; Zaniewski, R.; Youngs, W.; Tessier, C.; Klopman, G. *Acta Crystallogr.* **1996**, *B52*, 165.

(3) van Eijck, B. P.; Spek, A. L.; Mooij, W. T. M.; Kroon, J. *Acta Crystallogr.* **1998**, *B54*, 291.

(4) Williams, D. E. *Acta Crystallogr.* **1980**, *A36*, 715.

(5) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622.

(6) Holden, J. R.; Du, Z.; Ammon, H. L. *J. Comp. Chem.* **1993**, *14*, 422.

(7) Dzyabchenko, A. V. *Kristallografiya* **1989**, *34*, 226.

(8) Gibson, K. D.; Scheraga, H. A. *J. Phys. Chem.* **1995**, *99*, 3765.

(9) Karfunkel, H. R.; Gdanitz, R. J. *J. Comput. Chem.* **1992**, *13*, 1171.

(10) Williams, D. E. *Acta Crystallogr.* **1996**, *A52*, 326.

(11) Mooij, W. T. M.; van Eijck, B. P.; Price, S. L.; Verwer, P.; Kroon, J. *J. Comput. Chem.* **1998**, *19*, 459.

(12) Wawak, R. J.; Gibson, K. D.; Liwo, A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1743.

(13) Wawak, R. J.; Pillardy, J.; Liwo, A.; Gibson, K. D.; Scheraga, H. A. *J. Phys. Chem. A* **1998**, *102*, 2904.

(14) Dzyabchenko, A. V.; Agafonov, V. *Proc. Hawaii Int. Conf. Syst. Sci.* **1995**, *28*, 237.

(15) Molchanova, M. S.; Pivina, T. S.; Arnautova, E. A.; Zefirov, N. S. *J. Mol. Struct.* **1999**, *465*, 11.

(16) For an average organic crystal structure (10 atoms in the molecule and 4 molecules in the unit cell), the number of interatomic interactions which must be computed approaches 300 000 for a cutoff consisting of three layers of unit cells on each side of the reference unit cell. To obtain a local minimum structure, one needs to carry out several hundred energy and gradient evaluations.

(17) Ewald, P. *Ann. Phys.* **1921**, *64*, 253.

10.1021/ja9929990 CCC: $19.00
Published on Web 01/22/2000

Various approaches to crystal structure prediction have been proposed during the last several years; their main feature is to simplify the global optimization problem by restricting the search space. One approach is based on a statistical analysis of the experimental data from the Cambridge Structural Database (CSD)[18] to figure out the most popular arrangements of molecules in crystals. Symmetry elements and space groups favorable for formation of close packing of molecules with different symmetry were first derived theoretically by Kitaigorodsky.[19,20] Subsequent analyses[21−25] of increasing amounts of structural data contained in the CSD confirmed his results and provided extensive information about the distribution of space groups for molecular crystals. This information has been used widely by different researchers for crystal structure prediction. For example, Dzyabchenko used information about the most probable space groups to search for the lowest local minima of the lattice energy. In his work,[26] a systematic search for optimal packing of benzene was carried out within the framework of several most popular space groups, taking the molecular and energy hypersurface symmetries into account. The same approach was used later for predictions of crystal structures of organic nitramines.[1] Subsequent investigations included more space groups to increase the probability of finding the lowest-energy minima. Thus, Chaka et al.[2] considered 13 of the most frequent space groups in their prediction of the crystal structures of eight hydrocarbons. Van Eijck et al.[3] generated all possible crystal structures of benzene by a systematic grid search within 31 space groups.

Another approach to crystal structure prediction involves an extrapolation of the results obtained for molecular clusters to crystals. In this case, clusters that serve as starting models for crystal structure calculations are built in accordance with the most common types of coordination spheres. Williams[4] assumed a center of symmetry in clusters having an odd number of benzene molecules. Later Gavezzotti[5] proposed a more elaborate method that included the construction of small clusters using some typical symmetry elements; the full crystal structure was then built by translation of the clusters, yielding the most frequent space groups for organic compounds. The decision to accept a given cluster as a building block for subsequent crystal structure calculations was made on the basis of its energy. A similar cluster-based strategy was implemented in the MOLPAK program by Holden et al.,[6] which searches for possible crystal packings of minimal unit cell volume.

Several attempts were made to predict crystal structures by a systematic search within the whole space of structural parameters. In ref 7, four benzene molecules in a primitive cell of trivial symmetry ($P1$) were allowed to move freely and independently. A similar approach was used by Gibson and Scheraga,[8] who minimized the crystal energy of benzene with no constraints other than the existence of a lattice.

Taken together, these studies show that the use of symmetry constraints is quite effective and in most cases locates the experimental structure. However, there is always a possibility that a given molecule can form a crystal structure in one of the less frequently populated space groups. Also, a systematic search does not guarantee that the global minimum and all lowest-energy minima of the potential energy will be found.

A number of search methods that assume no crystal symmetry have been described. Karfunkel and Gdanitz[9] employed a Monte Carlo simulated annealing search strategy to locate the crystal energy global minimum, and all local minima inside an energy window, for several organic molecules containing heteroatoms and polar groups. A similar methodology was utilized in the MPA program by Williams.[10] In ref 11, both a systematic search and a Monte Carlo approach were used to generate possible crystal structures of acetic acid. In general, methods based on a Monte Carlo algorithm do not restrict the search space, although they fix the number of molecules in the unit cell. They are successful in locating a group of minima on the potential energy surface, but they do not solve the global optimization problem.

Some progress toward a solution of the global optimization problem for crystals has been achieved recently.[12,13] A promising approach to the multiple-minima problem involves methods based on deforming and smoothing the original energy surface, thereby greatly reducing the number of minima (occasionally to a single minimum) and simplifying the conformational search.[27−35]

Deformation (smoothing) of the original energy surface $f(\mathbf{x})$ is a procedure that deliberately alters the functional form of f in order to remove barriers between minima, making them merge together, and, therefore, significantly reducing their number. It is achieved by transforming the original function $f(\mathbf{x})$ into a new function $F(\mathbf{x},a)$ (a being a deformation parameter), with the additional requirement that $F(\mathbf{x},0) = f(\mathbf{x})$. The function F should be constructed in a way that ensures that the number of minima will decrease while the deformation parameter a increases. A good example of the function $F(\mathbf{x},a)$ appears in the diffusion equation method (DEM),[28] in which the deformed function is a solution of the diffusion equation with f being the initial boundary condition for the diffusion time (being a deformation parameter here) equal to zero.

The simplest approach to deformation-based global optimization is to track the lowest-energy minimum on the highly deformed potential energy surface back to the undeformed surface; however, this approach is successful only for relatively simple systems.[36] Usually, the lowest-energy minimum on the highly deformed energy surface does not correspond directly to the global minimum of the original surface, even when there is only one minimum left at the highest deformation, and trajectories connecting highly deformed and undeformed minima

(18) Allen, F. H.; Kennard, O.; Taylor, R. *Acc. Chem. Res.* **1983**, *16*, 146.
(19) Kitaigorodsky, A. I. *Molecular Crystals and Molecules*; Academic Press: New York, 1973.
(20) Pertsin, A. J.; Kitaigorodsky, A. I. *The Atom-Atom Potential Method: Applications to Organic Molecular Solids*; Springer: New York, 1987.
(21) Belsky, V. K.; Zorkii, P. M. *Acta Crystallogr.* **1977**, *A33*, 1004.
(22) Mighell, A. D.; Himes, V. L.; Rodgers, J. D. *Acta Crystallogr.* **1983**, *A39*, 737.
(23) Chernikova, N. Yu.; Bel'skii, V. K; Zorkii, P. M. *J. Struct. Chem.* **1991**, *31*, 661.
(24) Padmaja, N.; Ramakumar, S.; Viswamitra, M. A. *Acta Crystallogr.* **1990**, *A46*, 725.
(25) Brock, C. P.; Dunitz, J. D. *Chem. Mater.* **1994**, *6*, 1118.
(26) Dzyabchenko, A. V. *Zh. Strukt. Khim.* **1984**, *25*, 85.
(27) Pillardy, J.; Piela, L. *J. Phys. Chem.* **1995**, *99*, 11805.
(28) Piela, L.; Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem.* **1989**, *93*, 3339.
(29) Kostrowicki, J.; Piela, L.; Cherayil, B. J.; Scheraga, H. A. *J. Phys. Chem.* **1991**, *95*, 4113.
(30) Kostrowicki, J.; Piela, L. *J. Optimiz. Theory App.* **1991**, *69*, 269.
(31) Pillardy, J.; Olszewski, K. A.; Piela, L. *J. Phys. Chem.* **1992**, *96*, 4337.
(32) Pillardy, J.; Olszewski, K. A.; Piela, L. *J. Mol. Struct.* **1992**, *270*, 277.
(33) Amara, P.; Hsu, D.; Straub, J. E. *J. Phys. Chem.* **1993**, *97*, 6715.
(34) Oresič, M.; Shalloway, D. *J. Chem. Phys.* **1994**, *101*, 9844.
(35) Pappu, R. V.; Hart, R. K.; Ponder, J. W. *J. Phys. Chem. B* **1998**, *102*, 9725.
(36) Wawak, R. J.; Wimmer, M. M.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 5138.

often branch during a reversal of the deformation.[12,13] A possible solution to this problem is to track back more than one minimum in the reversing procedure, and to try to detect branching of trajectories by using a local search in the vicinity of each trajectory. This approach was far more successful than the single- or multiple-trajectory approach, and was applied in the theoretical prediction of crystal structures of hexasulfur and benzene.[12,13] However, it does not work sufficiently well for highly demanding applications, such as large Lennard-Jones clusters[37] or polypeptide chains.[38]

In the present paper, we apply a recently proposed method for global minimization, the self-consistent basin-to-deformed-basin mapping (SCBDBM)[38] method. The underlying principle is to locate a group of large basins containing low-energy minima (superbasins) in the original energy surface. This is achieved by coupling the superbasins in the original energy surface to basins in a highly deformed energy surface by iterative cycles, each of which reverses the deformation and then deforms the energy surface again from the newly found low-energy structures, until a degree of self-consistency is attained. A more detailed description of this algorithm is given in the Global Optimization Algorithm section. The method has been applied successfully to predict low-energy structures of polyalanine chains of length up to 100 amino acid residues[38], and to locate the global minima of Lennard-Jones argon clusters containing up to 100 atoms.[37]

Any attempt to predict a crystal structure theoretically, on the basis of the minimization of the potential energy, requires a physically reasonable energy function. The model potential should satisfy two criteria: (a) It should reproduce the experimental structure within a certain accuracy; (b) The crystal structures corresponding to the lowest-energy minima found for the potential should represent possible crystal structures, and one of them, possibly the global minimum, should correspond to the observed structure. If both criteria are satisfied, two situations are possible as a result of crystal structure prediction. First, the minimized experimental structure corresponds to the global minimum, which means that the potential is correct. Second, the minimized experimental structure is found among many low-energy structures that are close in energy to the global minimum. It has been noted[2,6,9] that the latter situation is quite typical, especially for molecules of regular shape without pronounced bumps and hollows. In these cases, crystal structure prediction provides a list of possible structures instead of a single structure. To eliminate hypothetical structures, it may be necessary to take into account such factors as entropy effects, kinetics of crystallization, or the effects of the environment. On the other hand, the force field used in the computations may not be accurate enough, and it may be necessary to refine the potential parameters or change the functional form of the potential (for example, by using a more accurate charge distribution model or by including polarization effects) in order to render all the hypothetical structures as energetically unfeasible.

Since any method for predicting crystal structures deals with a huge number of minima in a very large search space, and involves a large number of energy and gradient computations, the potential energy functions must be simple and computationally inexpensive. The most common simplification separates the potential energy into a sum of pairwise interatomic interactions, which are usually taken to be sums of repulsive and attractive terms (van der Waals potential) and the electrostatic energy. The parameters for the van der Waals potential are usually optimized by fitting to experimental structures.[39−45] Ab initio calculations of the energies of dimers and small clusters can be used to obtain the parameters for some molecules but still are not applicable for very large organic molecules. Various models have been proposed to describe electrostatic interactions in crystals.[39,40] Most popular models use isolated charges positioned on the atomic nuclei, derived either from experimental structures[39], or by matching the charges to ab initio molecular electrostatic potentials.[40,46] More sophisticated potentials, including representation of the molecular charge distribution by sets of multipoles on each atomic site, have been proposed recently.[47] In general, polarization effects play an important role in crystals, especially for conjugated systems and systems with hydrogen bonds, and must be taken into account to provide potential directionality. Unfortunately, existing polarizability models are not accurate enough[48] or are too elaborate[49] to be used for massive computations. The common way to evaluate the quality of a potential is to check its ability to reproduce the experimental structure as a local minimum of the potential, with the hope that the potential will be able to describe the entire energy surface correctly. This satisfies the first criterion in the previous paragraph, but not necessarily the second one.

The goal of the present study was to apply the SCBDBM method[38] to the calculation of crystal structures of polar organic molecules, and to evaluate two popular potential energy functions, DISCOVER[42,43,50] and AMBER,[46] according to the two criteria a and b. Both potentials have simple functional forms (9−6−1 and 12−6−1 for DISCOVER and AMBER, respectively), and contain electrostatic parameters obtained in different ways. Our results suggest that the AMBER potential is better suited for crystal structure prediction than DISCOVER.

A further goal of the present work was to clarify the role of the dipole moment of the unit cell in energy computations. Even if the experimental crystal structure corresponds to a nonpolar space group, the unit cells in the structures considered during global search computations may have large dipole moments, because no symmetry constraints other than the periodic condition are assumed during the search. A large dipole moment of the unit cell raises questions about the very definition of the electrostatic energy of a crystal, since this energy depends not only on the arrangement of the molecules in the unit cell, but also on the shape of the macrocrystal, and on the choice of the reference unit cell in the computations. A rigorous mathematical approach requires the addition of a correction term, in the form of a surface integral, to the usual Ewald summation.[51−54] Our

(37) Pillardy, J.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem.* **1999**, *103 A*, 9370.
(38) Pillardy, J.; Liwo, A.; Groth, M.; Scheraga, H. A. *J. Phys. Chem. B* **1999**, *103*, 7353.
(39) Williams, D. E.; Starr, T. L. *Comput. Chem.* **1977**, *1*, 173.
(40) Williams, D. E.; Weller, R. R. *J. Am. Chem. Soc.* **1983**, *105*, 4143.
(41) Williams, D. E.; Houpt, D. J. *Acta Crystallogr.* **1986**, *B42*, 286.
(42) Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319.
(43) Lifson, S.; Hagler, A. T.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5111.
(44) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361.
(45) Gavezzotti, A.; Filippini, G. *J. Phys. Chem.* **1994**, *98*, 4831.
(46) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(47) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047.
(48) Le Sueur, C. R.; Stone, A. J. *Mol. Phys.* **1994**, *83*, 292.
(49) Stone, A. J. *Mol. Phys.* **1985**, *56*, 1065.
(50) Ewig, C. S.; Thacher, T. S.; Hagler, A. T. *J. Phys. Chem. B* **1999**, *103*, 6998.
(51) de Leeuw, S. W.; Perram, J. W.; Smith, E. R. *Proc. R. Soc. London* **1980**, *A373*, 27.
(52) Smith, E. R. *Proc. R. Soc. London* **1981**, *A375*, 475.

results obtained with the AMBER and the DISCOVER force fields suggest that computations of crystal energies using the Ewald summation with no dipole correction, and without the restriction to nonpolar crystals, lead to artifact structures among the lowest energy structures; sometimes, the prediction may be wrong because of this factor alone. In contrast to our previous calculations,[12,13] which treated only rigid molecules with no dipole moment, the rigid molecules considered here each have a dipole moment. Finally, the global optimization approach is carried out without resort to information about the space group, and normally the number of molecules in the unit cell is also a predicted value;[12,13] however, in this paper, we used experimental information about that number to save computational time.

2. Calculation of the Potential Energy Function in Crystals: Role of the Dipole Moment of the Unit Cell

Assuming a given pairwise potential, the energy per molecule of a real crystal in vacuo should theoretically be calculated as the total energy of the entire cluster of periodically arranged molecules, divided by the number of molecules comprising the cluster; i.e., the sum of all pairwise interactions would have to be computed. Since only large crystals are of interest here, this way of calculating the crystal energy is computationally impossible. A natural way to circumvent the problem seems to be to restrict the computations to a chosen unit cell inside the crystal and to its interactions with other unit cells within a certain cutoff. However, as discussed later in this section, this way of computing the energy of the crystal is not always correct, even when imposing huge cutoffs.

The sum of all pairwise interactions can be represented in the form of a sum over all unit cells of the crystal, where each term of the sum is the internal energy of the unit cell, and half the sum of all interactions between the atoms of that unit cell and all other atoms. Because of the computational cost, the latter summation must be restricted to only atoms satisfying certain cutoff criteria relative to the unit cell; moreover, a single unit cell is chosen as a reference cell, representative of all cells in the crystal. The energy defined in this way may be called a "simplified" energy. Three basic questions must be asked: Is the final value of the "simplified energy" independent of the choice of the order of summation, i.e., of the way the infinite lattice summation is approximated by a finite summation? Is the final value of the "simplified energy" independent of the choice the reference unit cell? How is the true energy of a real crystal in vacuo related to the "simplified" energy?

The problem simplifies if the pairwise interactions are only of Lennard-Jones type (e.g., for the crystal of S_6). In this case, all pairwise interactions decrease in the order of $1/r^6$, and the sum of all interactions between the reference cell atoms and all other atoms converges unconditionally. Consequently, the limiting value of the energy is independent of the order of summation, and high accuracy can be achieved with relatively small cutoffs. The limiting value is the same for any choice of the reference unit cell unless it is close to the surface of the crystal; since the number of such unit cells is proportional to the $^2/_3$ power of the overall size of the crystal, the simplified energy is an accurate measure of the true Lennard-Jones energy of the crystal in vacuo.

The presence of an electrostatic term in pairwise interatomic interactions greatly complicates the answer to the three questions above, because the $1/r$ convergence rate of the electrostatic interactions does not guarantee the unconditional convergence of the sum of interactions between the atoms in the reference cell and all other atoms.

If the dipole moment of the unit cell is zero, unconditional convergence can still be achieved, if, additionally, it is assumed that only complete, uncharged molecules (or uncharged units such as Na^+ Cl^-) are considered in the computations. The sum of all interactions between the *atoms* of the reference unit cell and all other *atoms* of a finite crystal can be represented as a sum of interactions between the *reference unit cell* and all *molecules* not belonging to that unit cell. This is a triple sum, in which the outer sum is taken over all molecules not belonging to the reference cell, the middle sum is taken over all atoms of those molecules, and the inner sum is over all atoms in the reference cell. Since the dipole moment of the unit cell is zero, and the molecules are uncharged, the outer sum exhibits at worst a $1/r^4$ behavior as a sum of quadrupole−dipole interactions and is unconditionally convergent. As in the case of pure Lennard-Jones interactions, the true electrostatic energy per molecule of a crystal in vacuo is the same as the "simplified" energy computed by using the cutoff approach, for any cutoff criterion and any choice of the reference unit cell. However, the need to sum over pairs, each comprising a unit cell and a molecule, to achieve the $1/r^4$ behavior, rather than pairs of atoms, greatly increases the computational cost.

Fortunately, if the dipole moment of the unit cell is zero, the cutoff approach can be replaced by the well-known Ewald summation,[13,17,54] which enables a limiting value of the electrostatic energy to be computed very efficiently, and with practically unlimited accuracy. The true electrostatic energy of a crystal in vacuo can then be calculated with good accuracy and reasonable computational cost, if the cutoffs for the Ewald summation both in the real and reciprocal space are not too small, and the constant α in the Ewald summation is properly chosen.[13] The dipole moment of the unit cell is guaranteed to be zero if all computations are carried out with the assumption that the crystal symmetry is defined by one of the nonpolar space groups, since the dipole moments of the molecules in the unit cell cancel out; however, that assumption is stronger than the requirement that the dipole moment of the unit cell be zero. In papers dealing with crystal energy computations, where nonpolar space groups are assumed, the dipole moment issue is usually not mentioned (e.g., refs 1 and 2), and in ref 9 it is not addressed at all, although no symmetry elements are assumed.

It cannot be emphasized too strongly that the Ewald summation is merely a mathematical trick for speeding convergence. All assumptions associated with the way it was derived should be carefully verified before it is applied, one of the most important being that the dipole moment of the unit cell is zero. The Ewald summation cannot be treated as a universal "physical" way to compute the electrostatic energy. If the unit cell has a nonzero dipole moment, the computation of the true energy of the crystal in vacuo becomes more complicated. With a given choice of a reference unit cell, the "simplified" electrostatic energy can be computed as the sum of the Ewald energy and a correction term accounting for the dipole−dipole interaction between the reference cell and all other unit cells. A precise mathematical derivation of appropriate formulas can be found in ref 51, and in a more compact form in ref 54. Since the size of the macrocrystal is assumed to be large, the correction can be represented as a two-dimensional surface integral.[51,54]

(53) Deem, M. W.; Newsam, J. M.; Sinha, S. K. *J. Phys. Chem.* **1990**, *94*, 8356.
(54) van Eijck, B. P.; Kroon, J. *J. Phys. Chem. B* **1997**, *101*, 1096.

Table 1. Parameters of Experimental and Locally Minimized Experimental Structures for Crystals with Zero Dipole Moment of the Experimental Unit Cell[a]

method	space group	energy,[b] kcal/mol	V,[c] Å^3	a, Å	b, Å	c, Å	α, deg	β deg	γ, deg
			Formamide						
exptl	$P2_1/n$, $Z=4$		217.1	3.54	8.95	6.97	90.0	101.1	90.0
DISCOVER	$P\bar{1}$, $Z=4$	−16.36	220.5	3.54	9.79	6.50	94.9	96.6	98.3
AMBER	$P2_1/n$, $Z=4$	−22.19	202.6	3.48	9.56	6.10	90.0	95.5	90.0
			Imidazole						
exptl	$P2_1/c$, $Z=4$		347.3	7.57	5.37	9.78	90.0	119.1	90.0
DISCOVER	$P2_1/c$, $Z=4$	−19.51	340.1	8.10	4.88	9.89	90.0	119.5	90.0
AMBER	$P2_1/c$, $Z=4$	−19.56	324.5	8.07	4.80	9.85	90.0	121.8	90.0
			Maleic anhydride						
exptl	$P2_12_12_1$, $Z=4$		434.6	7.18	11.23	5.39	90.0	90.0	90.0
DISCOVER	$P2_12_12_1$, $Z=4$	−19.59	402.8	6.52	12.18	5.07	90.0	90.0	90.0
AMBER	$P2_12_12_1$, $Z=4$	−23.03	394.3	6.76	11.49	5.07	90.0	90.0	90.0
			Succinic anhydride						
exptl	$P2_12_12_1$, $Z=4$		440.5	6.96	11.71	5.40	90.0	90.0	90.0
DISCOVER	$P2_12_12_1$, $Z=4$	−22.31	428.3	6.68	12.04	5.33	90.0	90.0	90.0
AMBER	$P2_12_12_1$, $Z=4$	−27.01	418.0	6.78	11.62	5.31	90.0	90.0	90.0

[a] The runs with and without the spherical dipole moment correction term produced identical results and maintained zero dipole moment of the unit cell. [b] Energy per molecule. [c] Volume of the unit cell.

The value of the correction term depends on the shape of the macrocrystal; this is where the conditional convergence of the electrostatic interactions shows up. Analytical formulas are available for several simple shapes such as a sphere, a cube, and a rectangular parallelepiped;[51,54] for example, the correction term for a spherical crystal or a cubic-shaped crystal takes the form[51] $2\pi p^2/3V$, where p and V are the values of the dipole moment and the volume, respectively, of the unit cell. If the macrocrystal takes the shape of an infinitely long needle with the dipole moment of the unit cell in the needle direction, or a shape of an infinitely large platelet with the dipole moment in its plane, the correction term vanishes. To compute the true energy numerically, all "simplified" energies must be averaged over all reference unit cells because the "simplified" energies also depend on the choice of the reference unit cell.[54] Since every unit cell in the crystal may be considered as the reference cell, numerical computation of the true electrostatic energy involves a five-dimensional grid evaluation of the integrand in the surface integral. Accurate computation of the true energy of a crystal whose dipole moment is not zero is not only computationally expensive but also varies with the shape of the macrocrystal.

The above considerations seem to imply that the computation of the electrostatic energy of crystals makes sense only if all energy computations, *including those for intermediate structures appearing in global and local minimizations*, are carried out within the class of crystals whose unit cells have reasonably small dipole moment. There is an ongoing discussion of this matter in the literature,[51−54] but some papers[53,54] suggest that there should be no difference in the treatment of crystals with zero and nonzero dipole moment; they propose using only the Ewald summation and neglecting all other effects, including the surface integral that describes the contribution of the dipole moment. They argue[54] that a polar crystal in vacuo will assume the shape of a needle or a platelet, with the dipole moment along the needle or in the plane of the platelet, respectively, to remove the otherwise positive energy resulting from a nonzero dipole moment. If the crystal were surrounded by a medium of very high dielectric constant, the surface effect would be canceled *exactly*[53,54] by external charges that accumulate on the surface; this is sometimes referred to as a thin foil boundary condition.

We challenge that point of view. Since the existence of the dipole moment correction term is a mathematical fact and that term may account for more than 50% of the total energy, it cannot simply be dismissed without strong theoretical and numerical evidence. In our opinion the statement, that crystals in vacuo can always find a shape for which the surface effect would vanish (a needle or platelet) and the electrostatic energy would be at its lowest, cannot be applied to the physical situation, because it implies that the ratio between the surface area of the ends of the needle and its total surface area, or between the surface area of the edge of the platelet and its total surface area, is infinitesimally small. If the ratio between the length of the needle and its thickness is more realistic, the surface effect is significant. To determine this, we calculated the average dipole moment correction term for a square cross-section needle with the dipole moment along the needle, by using the surface integral from ref 54 which involves five-dimensional numerical evaluations of the integrand; two ratios between the length of the needle L and the side a of the square were considered, namely 10 and 5. The averaging was done over all reference cells lying no closer than 10% of a to the side walls of the needle and no closer than 10% of L to the ends of the needle to make sure that all necessary assumptions[54] for applying the formula were satisfied. With the ratios 10:1 and 5:1, the surface effect accounted for about 5% and 15% of the spherical correction term $2\pi p^2/3V$, respectively. If the averaging was done by including more molecules lying closer to the boundaries of the needle, those numbers would be higher. For the structure of formamide with energy −21.9 kcal/mol calculated with AMBER in Table 3, for example, 5% of the spherical correction term would amount to more than +0.6 kcal/mol, which would separate this high dipole moment structure from the lowest energy structures.

If the crystal is surrounded by a medium of a very high dielectric constant, and the external charges are "borrowed" from that medium and compensate the surface effect, the crystal can therefore no longer be treated computationally in isolation from that medium.

To provide further support for our opinion, we carried out theoretical crystal structure predictions, based on global minimization of the potential energy function for six polar molecules, with both AMBER and DISCOVER. These are presented in the Results and Discussion section.

3. Global Optimization Algorithm

3.1. SCBDBM Method. Since details of the method are given elsewhere,[38] we present only a brief description here. We wish

Table 2. Parameters of Experimental and Locally Minimized Experimental Stuctures for Crystals with Nonzero Dipole Moment of the Experimental Unit Cell

method	space group	energy,[a] kcal/mol	V,[b] Å^3	D,[c] D	E_D,[d] kcal/mol	a, Å	b, Å	c, Å	α, deg	β, deg	γ, deg
				Pyrimidine							
exptl	$Pna2_1$, $Z = 4$		403.7			11.55	9.46	3.69	90.0	90.0	90.0
DISCOVER											
with[e]	$Pna2_1$, $Z = 4$	−14.19	416.2	3.5	0.23	11.32	9.98	3.68	90.0	90.0	90.0
without[f]	$Pna2_1$, $Z = 4$	−14.46	416.9	4.1	0.33	11.05	9.99	3.77	90.0	90.0	90.0
AMBER											
with	$Pna2_1$, $Z = 4$	−15.28	394.7	3.7	0.35	11.01	9.83	3.64	90.0	90.0	90.0
without	$Pna2_1$, $Z = 4$	−15.64	394.9	4.0	0.34	10.88	9.84	3.69	90.0	90.0	90.0
				Formic acid							
exptl	$Pna2_1$, $Z = 4$		193.0			10.24	3.52	5.36	90.0	90.0	90.0
DISCOVER											
with	$Pna2_1$, $Z = 4$	−14.69	198.6	4.2	0.66	10.66	3.38	5.51	90.0	90.0	90.0
without	$Pna2_1$, $Z = 4$	−15.43	196.4	4.6	0.80	10.69	3.29	5.58	90.0	90.0	90.0
AMBER											
with	$Pna2_1$, $Z = 4$	−18.75	190.7	0.2	0.00	10.53	3.50	5.17	90.0	90.0	90.0
without	$Pna2_1$, $Z = 4$	−18.75	190.8	0.3	0.00	10.54	3.50	5.17	90.0	90.0	90.0

[a] Energy per molecule. [b] Volume of the unit cell. [c] Dipole moment of the unit cell. [d] Dipole moment spherical correction term. [e] Results of minimization with the dipole moment spherical correction term included. [f] Results of minimization with the dipole moment spherical correction term not included.

Table 3. Parameters of Calculated Structures Obtained by Global Optimization Compared with Locally Minimized Experimental Structures for Formamide with Discover and Amber[a]

method	space group[b]	energy,[c] kcal/mol	V,[d] Å^3	D,[e] D	E_D,[f] kcal/mol	a, Å	b, Å	c, Å	α, deg	β, deg	γ, deg
				DISCOVER[g]							
minimized exptl	$P\bar{1}$, $Z = 4$	−16.36	220.5	0.0	0.00	3.54	9.79	6.50	94.9	96.6	98.3
I[i] (min exptl)	$P\bar{1}$, $Z = 4$	−16.36	220.5	0.0	0.00	3.54	9.79	6.50	94.9	96.6	98.3
II	$P1$, $Z = 2$	−16.25	213.6	0.0	0.00	5.23	6.62	7.02	108.9	69.0	102.1
III	$Pcmn$, $Z = 4$	−16.15	214.9	0.0	0.00	4.88	5.67	7.76	90.0	90.0	90.0
				DISCOVER[h]							
minimized exptl	$P\bar{1}$, $Z = 4$	−16.36	220.5	0.0	0.00	3.54	9.79	6.50	94.9	96.6	98.3
I	Pn, $Z = 2$	−16.37	216.5	12.9	9.37	5.08	6.44	3.49	90.0	108.7	90.0
II	Pb, $Z = 4$	−16.29	217.7	12.9	9.27	6.49	5.09	7.34	90.0	116.2	90.0
III	Pa, $Z = 4$	−16.25	220.7	13.0	9.27	5.09	6.67	6.91	90.0	70.0	90.0
				AMBER[g]							
minimized exptl	$P2_1/n$, Z = 4	−22.19	202.6	0.0	0.00	3.48	9.56	6.10	90.0	95.5	90.0
I (min. exp.)	$P2_1/n$, $Z = 4$	−22.19	202.6	0.0	0.00	3.48	9.56	6.10	90.0	95.5	90.0
II	$P2_12_12_1$, $Z = 4$	−22.19	207.9	0.0	0.00	3.58	6.16	9.42	90.0	90.0	90.0
III	$P2_1/n$, $Z = 4$	−22.18	203.1	0.0	0.00	5.73	6.24	6.05	90.0	110.0	90.0
				AMBER[h]							
minimized exptl	$P2_1/n$, $Z = 4$	−22.19	202.6	0.0	0.00	3.48	9.56	6.10	90.0	95.5	90.0
I (min. exp.)	$P2_1/n$, $Z = 4$	−22.19	202.6	0.0	0.00	3.48	9.56	6.10	90.0	95.5	90.0
II	$P\bar{1}$, Z = 4	−21.98	205.3	1.4	0.07	3.48	9.52	6.21	92.9	91.2	92.2
III	$P1$, $Z = 4$	−21.94	197.2	18.9	13.72	4.92	6.80	7.00	60.9	82.4	74.3

[a] All structural parameters, volumes, and dipole moments of the unit cells are reported for representation with $Z = 4$. [b] Space group symmetry and number of molecules in the unit cell (Z) for the final structure (when it is possible to represent the final structure with a smaller number of molecules in the unit cell, the space group corresponding to this representation is presented, but all other parameters are given for $Z = 4$). [c] Energy per molecule. [d] Volume of the unit cell. [e] Dipole moment of the unit cell. [f] Dipole moment spherical correction term. [g] Results of minimization with the dipole moment spherical correction term included. [h] Results of minimization with the dipole moment spherical correction term not included. [i] Roman numbers denote the order of the minima obtained in the run in ascending order of the energy; if the minimized experimental structure was found in the run, it is reported with the note min exptl.

to find the global minimum of an energy function $f(\boldsymbol{x})$, where $\boldsymbol{x}$ is the collection of the lattice vectors, and parameters defining the position and orientation of each rigid molecule in the unit cell.

Consider a mapping $F(\boldsymbol{x},a)$, where a defines the extent of a deformation, such that $F(\boldsymbol{x},a)$ becomes smoother with a gradual decrease in the number of minima when a increases; assume that $F(\boldsymbol{x},0) = f(\boldsymbol{x})$. With increasing a, the number of minima gradually decreases, because some of the minima merge into one. As deformation proceeds, groups of individual minima are first merged, defining *superbasins* of these groups of minima for certain values of the deformation parameter a. As the deformation parameter increases, the superbasins from the smaller deformation (*lower-order superbasins*) also merge, constituting higher-order superbasins. Finally, for a very high deformation, only a few minima remain.

A logical procedure for locating the global minimum of $f(\boldsymbol{x})$ would be first to locate the highest-order (most deformed) superbasin related to this minimum and then to locate within it the superbasins of gradually lower order (lower deformation) that still contain this minimum, until the deformation is fully reversed. The major difficulty in proceeding in this manner is that there is no straightforward relation between the values of F at its minima and the corresponding minimum values of f. Therefore, one can never tell which superbasin corresponds to the global minimum of the original energy function, based only on the "energy" relations between superbasins. Consequently, it is not sufficient to reverse the deformation once only, to find the global minimum of f, even if a multitrajectory search is carried out during the reversing procedure. To surmount this problem, we propose a self-consistent procedure that finds the coupling relations between superbasins of different order, by

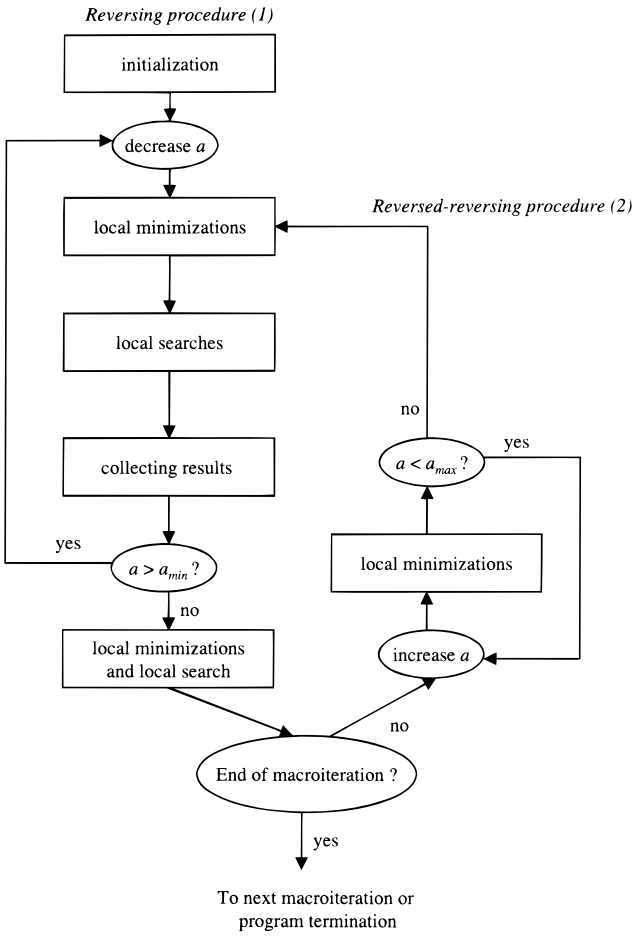

Figure 1. Block diagram of the reversing procedure, coupled with the reversed-reversing procedure within a single macroiteration.

iterating steps consisting of reversing the deformation and then reintroducing the deformation. The maximum number of trajectories to be followed is fixed at an arbitrary number p.

The procedure is outlined in Figure 1. It consists of a series of *macroiterations*. Each macroiteration establishes the coupling between superbasins of consecutive order and contains a self-consistent procedure within it. In macroiteration i, the parameter $a^{(i)}$, which controls the deformation, changes between two extreme values $a^{(i)}_{\max}$ and $a^{(i)}_{\min}$. For macroiteration $i+1$, $a^{(i+1)}_{\max} = a^{(i)}_{\min}$ and $a^{(i+1)}_{\min} = a^{(i)}_{\min}/\Delta$ (or 0 in the last macroiteration), where Δ is a logarithmic step length. The first macroiteration is initialized with randomly generated and minimized conformations in the most deformed space, while each subsequent macroiteration starts where the previous one left off.

Within each macroiteration, minima found on the most deformed surface with $a = a^{(i)}_{\max}$ are tracked back to the least deformed surface with $a = a^{(i)}_{\min}$, by decreasing the deformation parameter a and searching locally for new minima at each reversal step, in the vicinity of the minima already found; this is the reversing procedure. Subsequently, new minima found on the least deformed surface are tracked to the most deformed surface, by gradually increasing the parameter a and carrying out local minimizations without any local search at each step. This is the reversed-reversing procedure. This cycle is continued until no new minima are found on the least deformed surface or until the maximum allowed number of cycles is reached.

3.2. Application to Crystal Structure Prediction. The smoothing deformation chosen in the present work is the distance scaling deformation which was used in our earlier work,[13] where the distance scaling method (DSM) was applied for predicting the structure of benzene. It is very simple to implement, is designed to work with pairwise interactions, was shown to perform reasonably well in finding the global minima of Lennard-Jones and water clusters,[27,31,32] and has been applied successfully to predict the crystal structures of benzene and hexasulfur molecules.[13] In the DSM,[27] the site−site distance r_{ij} in the pairwise interaction is transformed into $\tilde{r}_{ij}$ as follows:

$$\tilde{r}_{ij} = \frac{r_{ij} + a r_{0,ij}}{1 + ba}$$

The parameter $r_{0,ij}$ in eq 1 is the position of the minimum of the undeformed pairwise interaction term. When the deformation parameter a is increased, the original function of the site−site distance (e.g., the Lennard-Jones potential) is flattened, but the position of its minimum and the function value at the minimum remain the same if the value of the parameter b is taken to be 1 (as in the original formulation[27] of the DSM). The parameter b controls the position of the minimum and remains constant during the calculations. For values of b greater or equal to 1, the position of the minimum of the deformed site−site function shifts to larger values of the site−site distance while, for $b < 1$, it shifts toward zero; if the deformation parameter a is greater than $1/(1-b)$, the two-body potential becomes totally attractive. Application of this deformation to a pairwise potential makes it relatively long-ranged by diminishing energy barriers between minima, by lowering repulsion for all values of b, and lowering attraction if $b > 0$. For a pairwise interaction which has a minimum, like a Lennard-Jones interaction, the value of $r_{0,ij}$ should be chosen as the position of this minimum; for electrostatic interactions, $r_{0,ij}$ should be large enough to ensure that the Coulomb interaction is weak at this distance, and this energy contribution will effectively be eliminated at large deformations. We choose the same value of $r_{0,ij} = r_{0,\mathrm{elec}}$ for all electrostatic terms, so that no interactions are smoothed faster than others. The Ewald summation was used to speed the calculations of the electrostatic energy; a detailed description of the way the Ewald summation was applied for the deformed electrostatic interactions can be found in ref 13.

When the deformation parameter a increases, the crystal shrinks because the pairwise interactions become more long-ranged and more molecules attract each other; in addition, the repulsive part of the potential is flattened by the deformation. If the parameter b of the deformation equals 1, the unit cell tends to collapse, and molecules overlap in space for large deformations. Many local minima are not removed, and an efficient reversing procedure becomes impossible. If the parameter b is set to a value larger than 1, molecules are forced to stay apart by shifting the local minima of the pairwise interactions to larger intermolecular distances. Smaller values of $r_{0,\mathrm{elec}}$ in the deformation of the electrostatic parts of the potential make the electrostatic part of the potential vanish more slowly. To test different values of b and $r_{0,\mathrm{elec}}$, several reversed-reversing procedures were carried out, starting with a set of 9 randomly generated local minima of the possible crystal structures of formamide (one of the molecules considered in this work). The values of $b = 1.75$ and $r_{0,\mathrm{elec}} = 1$ and the logarithmic scale of deformation were chosen to achieve the maximum degree of merging between minima, so as to distribute the merging of minima uniformly throughout the deformation. The relative energies of the 9 minima of the formamide crystal along the reversed-reversing trajectories are shown in Figure 2; abrupt vertical drops in relative energy represent merging of minima. The plot shows that there is no need to divide the

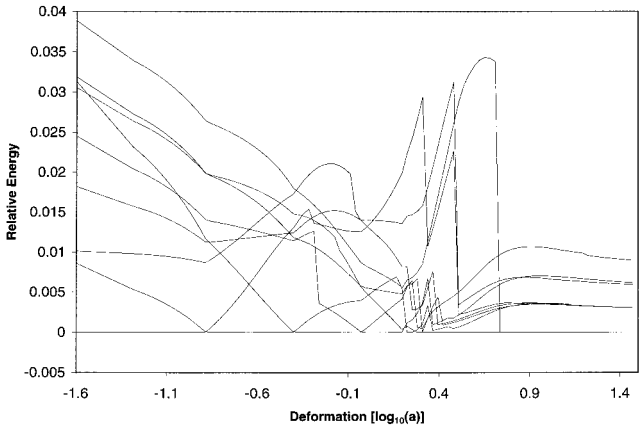

Figure 2. Changes in the order of the energies during the reversed-reversing procedure for the 9 low-energy, randomly generated minima in the undeformed energy surface of the formamide crystal. The vertical axis describes the relative energy difference between the current minimum and the lowest-energy minimum for a given value of the deformation parameter (the difference between the energy of the current minimum and the energy of the lowest-energy minimum in the sample, divided by the absolute value of the latter). The procedure starts from the 9 minima in the undeformed energy surface ($a \sim 0$). With increasing deformation, the trajectories ultimately merge to four minima near $a = 30$ ($\log a = 1.5$). Merging of minima is represented by abrupt vertical drops in relative energies.

deformation interval into more than two parts (i.e., only two macroiterations are sufficient), because there is only one region of frequent merging of minima, between $a = 1.5$ and 2.5 ($\log a = 0.2-0.4$). Also, the maximum value of the deformation parameter could be chosen as 30 ($\log a = 1.48$), because all merging occurred for lower values of a. We chose to use two macroiterations, the first with $a^1_{\max} = 30$ and $a^1_{\min} = 2.18$, and the second with $a^2_{\max} = 2.18$ and $a^2_{\min} = 0$.

A side effect of avoiding the collapse problem is the relatively high flattening of the potential in the rotational degrees of freedom due to an increase in the distance between molecules coupled with increasingly flattened pairwise interactions. This behavior, however, may be corrected by using a properly designed local search as discussed below.

A local search plays a very important role in the algorithm because it detects branching during the reversing procedure. However, this search should be carried out in the vicinity of a starting minimum; otherwise the relationship between minima may be lost (i.e. the newly found minimum may not be related to the previous one but belong to a completely different "tree" of trajectories). The simplest way to carry out the local search is by a random perturbation of the structures, followed by local energy minimization, as implemented in the multiple-trajectory perturbation approach;[12,13] in such local searches, we perturb all variables of the system. There are two drawbacks to this kind of local search: (i) Quite often, it fails to find another basin due to the relatively small perturbations of the lattice vectors and molecular positions (the perturbations have to be small in order carry out the search in the vicinity of a starting minimum). (ii) The search of the rotational degrees of freedom is inadequate. The latter results from the fact that the deformation significantly flattens the rotational part of the energy function, making the system very insensitive to rotations of molecules.

Application of a linear search provides a remedy for the first problem: the search was carried out along a randomly generated direction in multidimensional space and stopped when a new basin was found or a predefined maximum number of steps was exhausted. To circumvent the second problem, we designed a short systematic search over the rotational degrees of freedom of the molecules. The potential energy was computed for four consecutive values of each Eulerian angle of each molecule, while the rest of the molecules were kept fixed. The lowest-energy configuration was then chosen for subsequent local minimization.

The number of trajectories was chosen to be 10. The maximum number of cycles within the first macroiteration was set to 4 and within the second one to 10. During the local search, all variables were perturbed; the Eulerian angles of the molecules were perturbed randomly by no more than 20°, and the positions of the atoms and the values of the lattice vectors by no more than 1 Å.

With this choice of parameters, the method focuses on searching for a few lowest-energy structures and not on finding the complete set of the lowest-energy structures, which would require significantly larger computer resources. Because of this, only the first few computed structures are truly the lowest-energy ones; all others correspond to a sample of very low energy structures, and some structures may be missing in the set.

4. Molecular Models

For the calculations, we selected a number of small organic molecules which are interesting from the theoretical as well as the practical point of view. Several criteria were used for the selection. We were interested in molecules of biological importance, for which reliable potentials should be available. The molecules should be small, rigid, and contain C, H, N, and O atoms. To test the ability of a given potential to describe different types of interactions in crystals, such as hydrogen bonds and $\pi-\pi$ interactions, planar aromatic and hydrogen-bonded molecules were chosen. The crystal structures were chosen so that some of them had unit cells with zero dipole moments and others had nonzero dipole moments. The following molecules were selected: formamide, imidazole, maleic and succinic anhydrides, pyrimidine, and formic acid, all of them being assumed to be rigid.

We deduced molecular models from the experimental data. Since neutron diffraction experiments provide more accurate information about positions of hydrogen atoms, we used neutron diffraction data where available. The X-ray diffraction data were used for succinic and maleic anhydrides and for pyrimidine. For formamide and formic acid, molecular geometries obtained by neutron diffraction for corresponding deuterio compounds were used. All molecules under consideration except succinic anhydride are essentially planar; the largest deviation from the plane through the ring atoms occurrs for hydrogen atoms of maleic anhydride and is equal to 0.13 Å. Succinic and maleic anhydrides and pyrimidine have approximate point symmetry group *mm*2, and formamide, imidazole, and formic acid approximate point symmetry group *m* (Figure 3). The point symmetry groups were used to deduce symmetric molecular models. The positions of hydrogen atoms were shifted along the experimental bond directions to give the average experimental bond lengths of 1.09 Å for C−H and 1.04 Å for N−H, obtained by neutron diffraction.

The molecular models used in this work have geometrical parameters slightly different from those of the experimental ones; the issue of the influence of the geometrical model on the predicted structures is addressed in the next section.

5. Results and Discussion

5.1. Local Minimizations of the Experimental Structures. An important criterion that must be satisfied by any potential

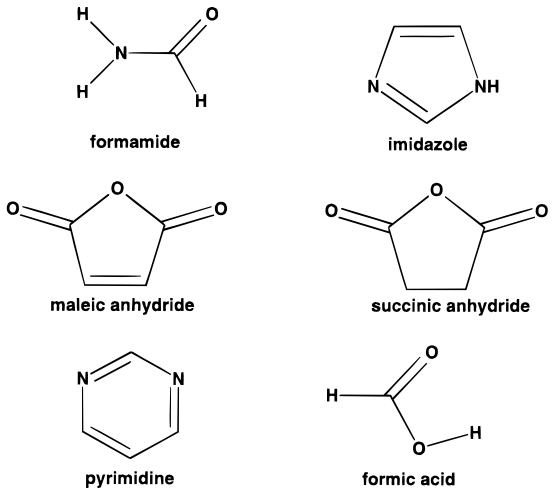

Figure 3. Molecular structures considered in this study.

is its ability to reproduce an experimental structure. For this purpose, we first conducted local minimizations for all molecules, starting from the experimental structures, with and without the dipole moment spherical correction term $2\pi p^2/3V$. The parameters of the experimental and the energy-minimized structures are given in Tables 1 and 2.

Formamide crystallizes in space group $P2_1/n$ with the number of molecules in the unit cell $Z = 4$.[55] The crystal structure consists of puckered sheets of molecules. Within the sheets, pairs of molecules associate about the centers of symmetry to form almost coplanar bimolecular units. Puckering of the sheets results from the tilt of these units relative to one another. Within each sheet, hydrogen bonds of two types exist: one type links molecules together to form bimolecular units (H···O bond length 1.92 Å; N−H···O angle 174.9°); the other type links bimolecular units together to form sheets (H···O bond length 1.85 Å; N−H···O angle 167.3°). After a local minimization with the DISCOVER potential, the crystal structure contained N−H···O bonds linking molecules within bimolecular units (bond length and bond angle 1.98 Å and 173.6°) and two types of hydrogen bonds linking bimolecular units within the sheets (bond lengths and bond angles 1.94 Å and 170.0°, and 1.91 Å and 172.9°, respectively). The orientation of the bimolecular units changed in such a way that the sheets of molecules became almost planar; as a result, the space group symmetry became $P\bar{1}$, with two independent molecules in the asymmetric unit. The average deviations of the unit cell vector lengths and angles from the experimental values were 5.4% and 6.3°, respectively.

When the crystal structure of formamide was optimized with the AMBER potential, the initial symmetry was preserved and the topology of the sheets was reproduced. However, the orientation of molecules changed significantly (Figure 4A). The length of the hydrogen bonds connecting the molecules in the bimolecular units was much shorter in the minimized structure (1.82 Å), which in turn led to significant shortening (12.5%) of the lattice parameter *c*. Other hydrogen bond parameters changed little. The bond lengths and N−H···O angles in the minimized structure were equal to 1.82 Å and 175.3°, and 1.81 Å and 171.5° for the first and the second types of hydrogen bonds, respectively. The average deviations of the unit cell parameters from the experimental values were 7% for lattice lengths and 5.6° for lattice angles.

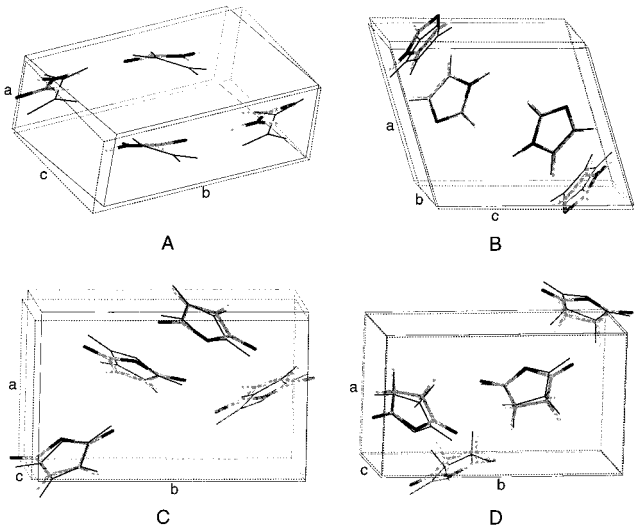

Figure 4. Experimental structures (thin line) and the locally minimized experimental structures with AMBER (thick line) for (A) formamide, (B) imidazole, (C) maleic anhydride, and (D) succinic anhydride. The minimized experimental structures are also the global minima of global minimization runs with the dipole moment correction term included.

The crystal structure of imidazole[56] (space group $P2_1/c$, $Z = 4$) is based on endless chains of planar molecules, connected by strong N−H···N hydrogen bonds (H···N distance 1.81 Å). Both the DISCOVER and the AMBER potential satisfactorily reproduce the experimental structure (Figure 4B), except that the hydrogen bond length is much longer in the case of the structure minimized with AMBER (1.98 Å). The average deviations of the unit cell parameters were 2.9% and 4.4% for the DISCOVER and the AMBER potentials, correspondingly. For both potentials, the largest deviation took place for the cell length *b* and did not exceed 10.5% (Table 1).

Maleic[57] and succinic[58] anhydrides are quite similar in their molecular and crystal structures and represent an example of isostructural compounds (Table 1). The final structures obtained with both the DISCOVER and the AMBER potential had the same space group symmetry as the experimental structures. In general, the DISCOVER potential reproduces the crystal structures of these compounds less well than AMBER. The average deviations of the unit cell parameters were 3.9% and 2.3% for maleic anhydride and 1.4% and 0.8% for succinic anhydride with DISCOVER and AMBER, respectively (Figure 4C,D). The deviations in unit cell parameters were smaller for succinic anhydride with both potentials.

Pyrimidine[59] forms a polar structure with symmetry $Pna2_1$, $Z = 4$. The crystal structure consists of stacks of nearly parallel molecules which overlap slightly. The axes of the stacks are parallel to the *c* axis. In all cases, the final structures obtained with DISCOVER and AMBER had the same space group symmetry as the experimental structure (Figure 5A). The deviations in the lattice parameters of pyrimidine were similar in magnitude (less than 4.0%), but the volumes of the unit cell increased with DISCOVER and decreased with AMBER (Table 2). The dipole moments of the unit cells in the pyrimidine structures minimized with and without the dipole moment correction term were nearly the same for DISCOVER and

(55) Torrie, B. H.; O'Donovan, C.; Powell, B. M. *Mol. Phys.* **1994**, *82*, 643.
(56) Martinez-Carrera, S. *Acta Crystallogr.* **1966**, *20*, 783.
(57) Marsh, R. E.; Ubell, E.; Wilcox, H. E. *Acta Crystallogr.* **1962**, *15*, 35.
(58) Ehrenberg, M. *Acta Crystallogr.* **1965**, *19*, 698.
(59) Furberg, S.; Grøgaard, J.; Smedsrud, B. *Acta Chem. Scand.* **1979**, *B33*, 715.

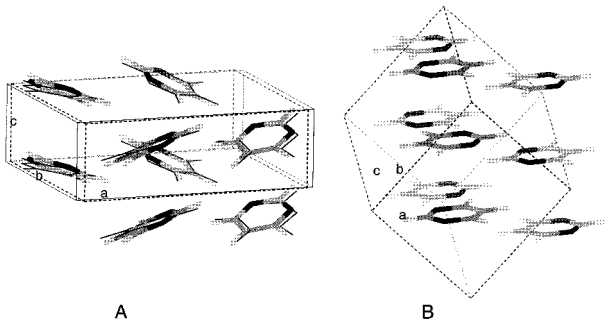

Figure 5. (A) Experimental structure (thin line) and the locally minimized experimental structure with AMBER without the dipole moment correction term included (thick line) for pyrimidine. (B) Global minimum structure of pyrimidine for the AMBER potential.

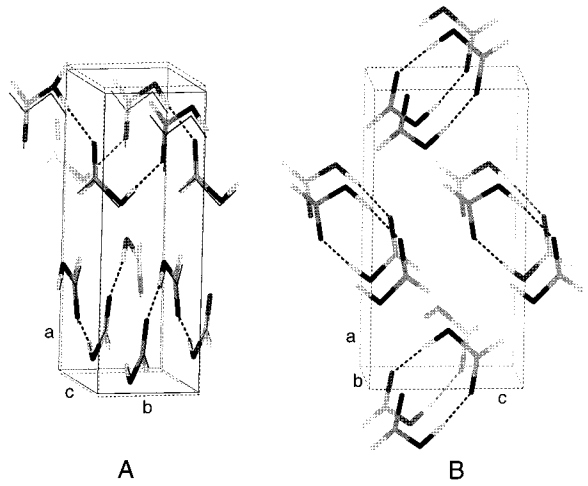

Figure 6. (A) Experimental structure (thin line) and the locally minimized experimental structure with AMBER without the dipole moment correction term included (thick line) for formic acid. (B) Global minimum structure of formic acid for the AMBER potential.

AMBER. However, with the dipole moment correction term included, the dipole moments of the unit cells are smaller. Values of the dipole moment correction did not exceed 2% of the total lattice energy.

Formic acid crystallizes in space group $Pna2_1$ with four molecules in the unit cell.[60] The crystal structure consists of infinite planar chains of molecules connected by hydrogen bonds (2.62 Å). The chains are tightly packed in layers. The interactions between the chains are of the van der Waals type. The results in Table 2 show that the DISCOVER potential reproduced the experimental structure of formic acid with larger discrepancies than AMBER (average deviations 1.8−2.5% and 1.2%, respectively). The positions and attitudes of molecules in the structures minimized with either DISCOVER or AMBER were very similar to those observed experimentally. The minimized structures contain the same hydrogen bonds as the experimental structure, but the bond lengths are noticeably longer in the case of DISCOVER [O(−H)···O distance 2.86 Å]; as a result, the volume of the unit cell is larger than with AMBER. The structural parameters and volume of the unit cell obtained with the AMBER potential are fairly close to those in the experimental structure (Figure 6A), despite the zero van der Waals parameters for the hydroxyl hydrogen in the AMBER potential. As seen in Table 2, the dipole moment of the unit cell and the corresponding contribution to the lattice energy were much larger with the DISCOVER potential (4.2−4.6 D and 0.6−0.8 kcal/mol, respectively). With AMBER, the dipole moment of the unit cell was 0.2−0.3 D and the contribution to the energy was negligibly small. This is a result of completely different charge distributions with DISCOVER and with AMBER; the angle between the dipole moment vectors for the *molecular* model of formic acid with DISCOVER and with AMBER is 43°, and the molecular dipole moment is much larger with DISCOVER.

In general, the results presented in Tables 1 and 2 show that the deviations of the lattice parameters from their experimental values are nearly the same for the DISCOVER and the AMBER potentials. In all cases (except formamide with DISCOVER), the final structures had the same space group symmetry as the starting structures which were derived from the experimental structures. However, structures containing hydrogen bonds are quite poorly reproduced by the two potentials. The lengths of hydrogen bonds in the energy-minimized structures are longer than they are in the experimental structures, especially with the AMBER potential.

With both potentials, the volumes of the unit cells decreased during energy minimization (except for formamide, pyrimidine, and formic acid minimized with the DISCOVER potential). There are several possible reasons for this effect. Since the DISCOVER potential parameters were obtained from experimental data for nonzero temperature, thermal effects are partially included in these parameters. At the same time, thermal motions are usually anisotropic and, therefore, cannot be described correctly by using an isotropic potential [this effect may also be responsible for much larger than an average deviation in one of the cell edges (for example, *b* for imidazole)]. The optimization of the DISCOVER force field included calculations of lattice sums over the energy components, in which two layers of unit cells on each side of the reference cell were taken into account. For large unit cells, a cutoff of 50 Å was employed. In our work, the lattice summation was extended to 5 or 6 layers of unit cells surrounding the reference unit cell. This is much larger than the typical cutoffs used in deriving potentials and can easily exceed 50 Å. The potentials were not parametrized with such large cutoffs, and the extra long-range attractions in our calculations might cause the unit cell to contract. The AMBER nonbonded parameters were obtained from simulations of liquids and may therefore not be strictly applicable to crystals.

5.2. Global Minimizations. After local minimizations of the experimental structures were carried out, two independent global minimization runs were then carried out for each molecule and for each potential, with four molecules in the unit cell (however, for some of the resulting structures a representation with $Z = 2$ molecules in the unit cell was possible). In the first run, the spherical dipole moment correction term $2\pi p^2/3V$ was added to the standard Ewald summation, to compensate for any dipole moment arising in the crystal structures in the computations. In practice, this ensures that the unit cells in the final structures will not have large dipole moments, because the dipole moment correction acts as a penalty function. In the second run, the dipole moment correction was not added; i.e., the approach suggested by others[54] was explored. Since the first run focuses the search on structures with zero or low dipole moment of the unit cell (open circles in Figures 7 and 8), the second run (without the penalty function) usually adds structures with large dipole moment (filled diamonds in Figures 7 and 8) to the lowest-energy structures found in the first run.

For all molecules whose crystal structures have zero dipole moment, namely formamide, imidazole, and maleic and succinic anhydrides, the minimized experimental structures were found

(60) Nahringbauer, I. *Acta Crystallogr.* **1978**, *B34*, 315.

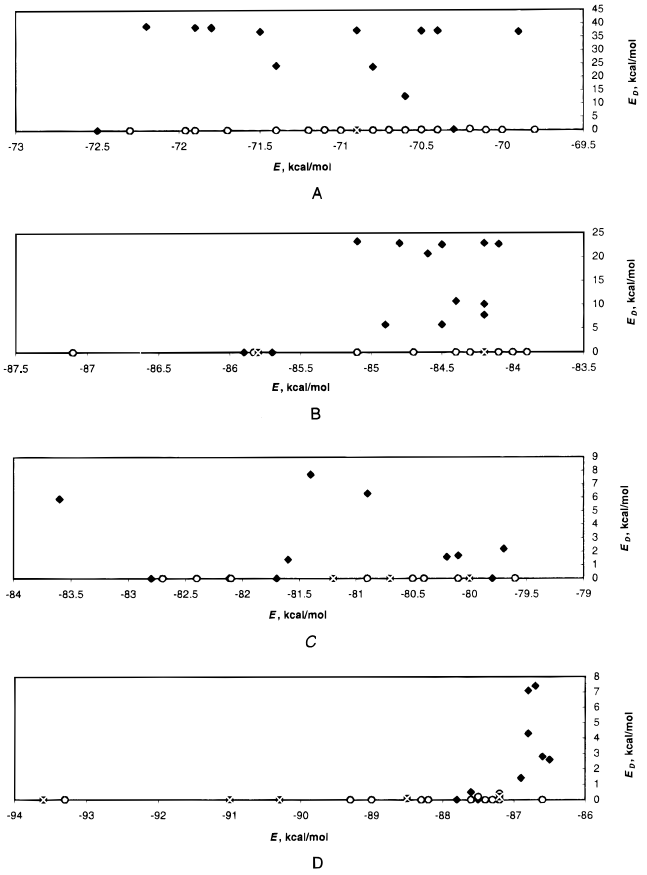

Figure 7. Dipole moment correction term (vertical axis) versus energy (horizontal axis) for the lowest-energy structures found with DISCOVER in the global minimization runs with the correction (open circles) and without the correction (filled diamonds). Structures found in both runs are marked by a cross in a black square. If the correction were included, the energy would be the sum of the values shown on both axes. Key: (A) formamide; (B) imidazole; (C) maleic anhydride; (D) succinic anhydride.

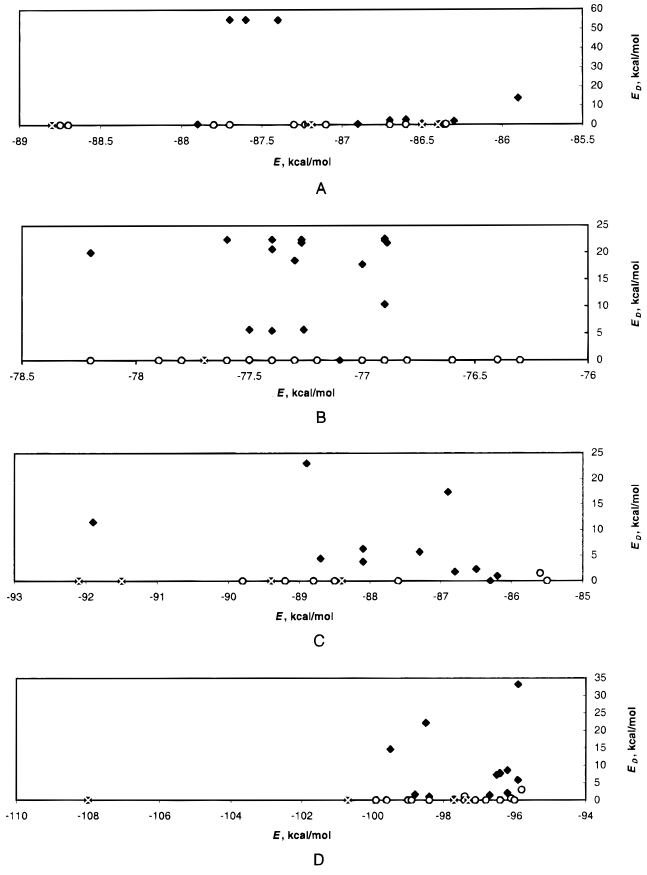

Figure 8. Same as Figure 7 but for the AMBER potential.

as the global minima of the AMBER potential in the runs with the correction (in the case of formamide another minimum with the same energy as the minimized experimental structure has been found); for the DISCOVER potential, they were found among the lowest-energy minima (see Tables 3−6). All the lowest-energy minima for both potentials (open circles in Figures 7 and 8) represented crystals whose unit cells had zero or negligibly small dipole moment. The prediction with the dipole moment correction included was successful for the AMBER potential (Figure 4) but only moderately successful for DISCOVER.

Although the unit cells of the experimental structures of those molecules have zero dipole moment, many artifact structures, whose unit cells had large dipole moments, were found among the lowest-energy minima in the runs without the correction term, with either potential (see Figures 7 and 8). With DISCOVER, a crystal structure with a large dipole moment was the global minimum for formamide, imidazole, and maleic anhydride (see Tables 3−5). With AMBER, a nonphysical structure whose unit cell had a large dipole moment had the same energy as the minimized experimental structure for imidazole (the structure with energy −19.5 kcal/mol in Table 4). This structure (Figure 9) contains chains of molecules connected by N−H···N bonds packed in layers; the chains in every second layer are parallel, whereas the neighboring layers are rotated with respect to each other. Molecular chains are also present in the locally minimized experimental structure (the global minimum of the run with the correction term included), but all chains are parallel to each other (Figure 10). The parameters of the hydrogen bonds presented in Figures 9B and 10B (the angle Φ between the projection of the hydrogen bond N···H on the plane of the acceptor imidazole molecule and the lone-pair direction, and the angle Θ between the hydrogen bond N···H and a line orthogonal to that plane) show that the hydrogen bonds in the high dipole moment structure (Figure 9B) are distorted, with the position of the hydrogen deviating from the average position observed experimentally for this kind of hydrogen bond,[61,62] whereas there is almost no distortion in the minimized experimental structure.

If the energies of the structures with large dipole moment of the unit cell, obtained as low-energy structures in the runs without the dipole moment correction term, are recalculated assuming a spherical shape for the macrocrystal and adding the corresponding correction term, they become dramatically higher than the energies computed by using only the Ewald summation plus the Lennard-Jones contribution (Figures 7 and 8; the recomputed energy can be obtained by adding the values on the horizontal and vertical axes). The lowest-energy structures of formamide may serve as an example. With the DISCOVER potential, the energy of the minimized experimental structure and the energies of the lowest-energy structures, computed without any correction term, are about −16.4 kcal/mol. Most of these structures have unit cells with large dipole moments, and inclusion of the spherical correction term brings their energies up to about −3.5 kcal/mol; such structures would never be considered in any global minimization method as candidates for the correct structure. The dipole moment contribution plays a major role here.

(61) Vedani, A.; Dunitz, J. D. *J. Am. Chem. Soc.* **1985**, *107*, 7653.
(62) Klebe, G. *Mol. Biol.* **1994**, *237*, 212.

Table 4. Parameters of Calculated Structures Obtained by Global Optimization Compared with Locally Minimized Experimental Structures for Imidazole with DISCOVER and AMBER[a]

method	space group	energy, kcal/mol	V, Å^3	D, D	E_D, kcal/mol	a, Å	b, Å	c, Å	α, deg	β, deg	γ, deg
				DISCOVER							
minimized exptl	$P2_1/c$, $Z = 4$	−19.51	340.1	0.0	0.00	8.10	4.88	9.89	90.0	119.5	90.0
I	$P\underline{2}_12_12_1$, $Z = 4$	−20.12	337.6	0.0	0.00	4.84	8.20	8.50	90.0	90.0	90.0
II	$P\underline{1}$, $Z = 4$	−19.88	333.9	0.0	0.00	5.09	8.19	8.20	86.9	99.0	91.9
III	$P1$, $Z = 2$	−19.88	332.6	0.0	0.00	5.09	8.19	8.87	66.0	85.6	97.2
XIV (min exptl)	$P2_1/c$, $Z = 4$	−19.51	340.1	0.0	0.00	8.10	4.88	9.89	90.0	119.5	90.0
minimized exptl	$P2_1/c$, $Z = 4$	−19.51	340.1	0.0	0.00	8.10	4.88	9.89	90.0	119.5	90.0
I	$Pna2_1$, $Z = 4$	−20.13	337.5	11.8	5.04	8.23	4.75	8.61	90.0	90.0	90.0
II	$P1$, $Z = 4$	−19.76	339.1	0.0	0.00	4.87	7.94	9.65	113.7	92.9	82.9
III	$P1$, $Z = 4$	−19.69	346.5	9.4	3.12	5.06	7.27	9.75	95.3	76.4	95.4
				AMBER							
minimized exptl	$P2_1/c$, $Z = 4$	−19.56	324.5	0.0	0.00	8.07	4.80	9.85	90.0	121.8	90.0
I (min exptl)	$P2_1/c$, $Z = 4$	−19.56	324.5	0.0	0.00	8.07	4.80	9.85	90.0	121.8	90.0
II	$P2_12_12_1$, $Z = 4$	−19.49	325.1	0.0	0.00	4.91	7.58	8.72	90.0	90.0	90.0
III	$P1$, $Z = 4$	−19.47	325.0	0.0	0.00	5.08	8.07	8.51	80.5	104.4	80.1
minimized exptl	$P2_1/c$, $Z = 4$	−19.56	324.5	0.0	0.00	8.07	4.80	9.85	90.0	121.8	90.0
I	$Pbn2_1$, $Z = 4$	−19.56	332.6	14.8	5.00	4.73	7.95	8.85	90.0	90.0	90.0
II	$P2_1$, $Z = 4$	−19.42	333.9	0.0	0.00	5.17	7.68	8.68	90.0	75.6	90.0
III	Pc, $Z = 4$	−19.39	325.6	15.5	5.60	7.24	4.55	9.91	90.0	84.8	90.0

[a] For symbols and units, see footnotes to Table 3.

Table 5. Parameters of Calculated Structures Obtained by Global Optimization Compared with Locally Minimized Experimental Structures for Maleic Anhydride with DISCOVER and AMBER[a]

method	space group	energy, kcal/mol	V, Å^3	D, D	E_D, kcal/mol	a, Å	b, Å	c, Å	α, deg	β, deg	γ, deg
				DISCOVER							
minimized exptl	$P2_12_12_1$, $Z = 4$	−19.59	402.8	0.0	0.00	6.52	12.18	5.07	90.0	90.0	90.0
I	$P\underline{2}_1/c$, $Z = 4$	−19.82	393.6	0.0	0.00	5.34	12.54	6.16	90.0	72.5	90.0
II	$P\underline{1}$, $Z = 2$	−19.79	393.6	0.0	0.00	6.01	7.47	9.71	109.2	103.2	75.3
III	$P1$ $Z = 2$	−19.72	395.0	0.0	0.00	3.84	8.89	11.69	97.1	94.2	86.8
V (min exptl)	$P2_12_12_1$, $Z = 4$	−19.59	402.8	0.0	0.00	6.52	12.18	5.07	90.0	90.0	90.0
minimized exptl	$P2_12_12_1$, $Z = 4$	−19.59	402.8	0.0	0.00	6.52	12.18	5.07	90.0	90.0	90.0
I	$Pna2_1$, $Z = 4$	−20.04	405.2	7.0	1.48	7.67	5.85	9.02	90.0	90.0	90.0
II	$P2_1/c$, $Z = 4$	−19.82	393.6	0.0	0.00	5.34	12.50	6.16	90.0	72.5	90.0
III	$P1$, $Z = 4$	−19.79	393.7	0.0	0.00	5.35	8.32	10.06	73.2	75.3	102.7
				AMBER							
minimized exptl	$P2_12_12_1$, $Z = 4$	−23.03	394.3	0.0	0.00	6.76	11.49	5.07	90.0	90.0	90.0
I (min exptl)	$P2_12_12_1$, $Z = 4$	−23.03	394.3	0.0	0.00	6.76	11.49	5.07	90.0	90.0	90.0
II	$P2_1/c$, $Z = 4$	−22.88	390.1	0.0	0.00	6.65	8.04	7.55	90.0	74.8	90.0
III	$P2_12_12_1$, $Z = 4$	−22.46	403.4	0.0	0.00	4.97	7.19	11.29	90.0	90.0	90.0
minimized exptl	$P2_12_12_1$, $Z = 4$	−23.03	394.3	0.0	0.00	6.76	11.49	5.07	90.0	90.0	90.0
I (min exptl)	$P2_12_12_1$, $Z = 4$	−23.03	394.3	0.0	0.00	6.76	11.49	5.07	90.0	90.0	90.0
II	$P\underline{c}$, $Z = 2$	−22.98	389.8	12.2	2.88	6.60	8.51	7.10	90.0	77.6	90.0
III	$P1$, $Z = 2$	−22.98	389.8	12.2	2.88	5.39	7.10	10.77	82.4	75.6	97.6

[a] For symbols and units, see footnotes to Table 3.

For the molecules whose experimental unit cells have nonzero dipole moment, pyrimidine and formic acid, the energy-minimized experimental structures were not found by the SCBDBM method; instead, some lower energy structures with zero dipole moment were found with or without the spherical dipole correction (see Tables7 and 8).

With pyrimidine, the minimized experimental structure was never found in any global optimization, because of its relatively high energy with both potentials. The global-minimum structure (Figure 5B) for the AMBER potential consists of stacks of parallel molecules which do not overlap with each other; this is compared to the minimized experimental structure in Figure 5A. The main difference lies in the orientation of the molecules within a stack; in the experimental structure all molecules in each stack have the same orientation, and in the global-minimum configuration every second molecule within each stack is rotated by 180°. Such an arrangement leads to zero dipole moment for the unit cell in Figure 5B.

Formic acid is another polar crystal for which the minimized experimental structure was never found during global optimizations. The energy of the global minimum with the AMBER potential was more than 0.3 kcal/mol lower than that of the minimized experimental structure. The two structures are shown in Figure 6. The experimental structure consists of infinite planar chains of molecules connected by hydrogen bonds, whereas the global minimum has a helixlike hydrogen bond arrangement. The hydrogen bonds in the global-minimum structure have distorted geometry,[62] with the hydrogen atom lying out of the plane of the acceptor carbonyl group. The AMBER potential[46] has special Lennard-Jones parameters for polar hydrogens (the repulsion for hydrogen that is bonded to nitrogen or oxygen atoms is lower than for hydrogen that is bonded to carbon, which represents the reduced charge density around polar hydrogens), but there is no special directional term for hydrogen bonds, as in the MM3 force field,[63] which can improve the accuracy. Similar problems with the hydrogen-bond geometry were observed in a structure of formamide with the AMBER potential.

To address the issue of the influence of the molecular model on the predictions, the lowest-energy structures obtained in the global optimization were reminimized locally with the experimental geometries (instead of the symmetrized geometries) in

(63) Lii, J.-H.; Allinger, N. L. *J. Comput. Chem.* **1998**, *19*, 1001.

Table 6. Parameters of Calculated Structures Obtained by Global Optimization Compared with Locally Minimized Experimental Structures for Succinic Anhydride with DISCOVER and AMBER[a]

method	space group	energy, kcal/mol	V, Å^3	D, D	E_D, kcal/mol	a, Å	b, Å	c, Å	α, deg	β, deg	γ, deg
				DISCOVER							
minimized exptl	$P2_12_12_1$, $Z=4$	−22.31	428.3	0.0	0.00	6.68	12.04	5.33	90.0	90.0	90.0
I (min exptl)	$P2_12_12_1$, $Z=4$	−22.31	428.3	0.0	0.00	6.68	12.04	5.33	90.0	90.0	90.0
II	$P2_1/c$, $Z=4$	−22.21	422.5	0.0	0.00	9.43	5.21	9.47	90.0	71.6	90.0
III	$P2_1/c$, $Z=4$	−21.62	423.5	0.0	0.00	8.46	5.73	9.39	90.0	68.3	90.0
minimized exptl	$P2_12_12_1$, $Z=4$	−22.31	428.3	0.0	0.00	6.68	12.04	5.33	90.0	90.0	90.0
I (min exptl)	$P2_12_12_1$, $Z=4$	−22.31	428.3	0.0	0.00	6.68	12.04	5.33	90.0	90.0	90.0
II	$P1$, $Z=2$	−21.37	425.5	7.9	1.79	5.75	8.86	9.06	69.2	99.4	94.5
III	$P1$, $Z=2$	−21.28	426.2	0.0	0.00	6.01	8.44	8.93	83.2	71.3	87.8
				AMBER							
minimized exptl	$P2_12_12_1$, $Z=4$	−27.01	418.0	0.0	0.00	6.78	11.62	5.31	90.0	90.0	90.0
I (min exptl)	$P2_12_12_1$, $Z=4$	−27.01	418.0	0.0	0.00	6.78	11.62	5.31	90.0	90.0	90.0
II	$P2_1/c$, $Z=4$	−25.18	431.5	0.0	0.00	5.64	7.55	10.45	90.0	104.3	90.0
III	$P1$, $Z=4$	−24.99	417.1	0.0	0.00	5.15	9.37	9.63	110.2	90.0	105.9
minimized exptl	$P2_12_12_1$, $Z=4$	−27.01	418.0	0.0	0.00	6.78	11.62	5.31	90.0	90.0	90.0
I (min exptl)	$P2_12_12_1$, $Z=4$	−27.01	418.0	0.0	0.00	6.78	11.62	5.31	90.0	90.0	90.0
II	$P2_1/c$, $Z=4$	−25.18	431.5	0.0	0.00	5.64	7.55	10.45	90.0	104.3	90.0
III	Pc, $Z=2$	−24.87	437.1	14.6	3.68	6.84	8.62	7.45	90.0	95.3	90.0

[a] For symbols and units, see footnotes to Table 3.

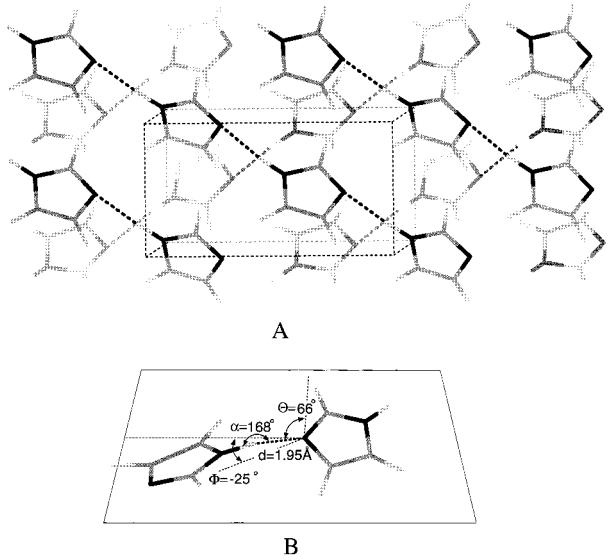

Figure 9. (A) Packing of imidazole molecules in the lowest energy structure found in the run without the correction term. Thick dotted lines indicate hydrogen bonds. (B) Geometrical parameters of intermolecular hydrogen bonds: the angle Φ between the projection of the hydrogen bond N···H on the plane of the acceptor imidazole molecule and the lone-pair direction and the angle Θ between the hydrogen bond N···H and a line orthogonal to that plane.

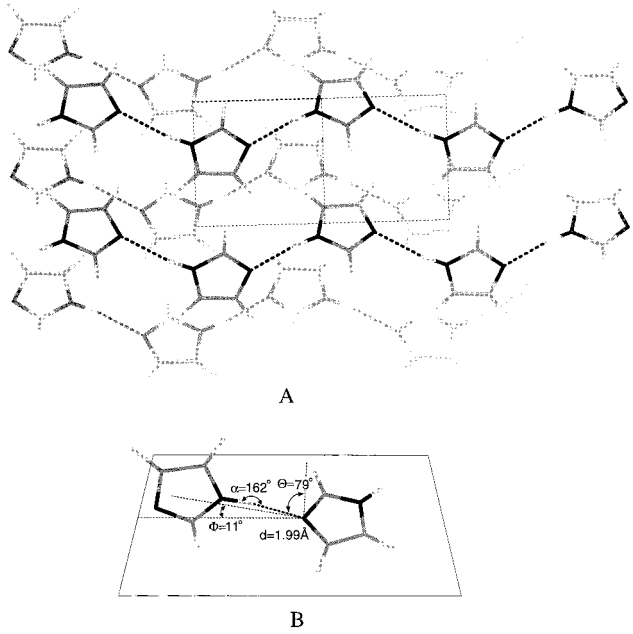

Figure 10. Same as Figure 9 but for the minimized experimental structure, found as the lowest energy structure in the run with the correction term included.

the case of formamide and pyrimidine (with the C−H and N−H bond lengths adjusted as described in section 4). These two molecules were chosen because the global optimization resulted in two structures (Table 3, structure III, DISCOVER with dipole moment contribution; Table 7, structure II, AMBER with dipole moment contribution) in which the symmetry of the crystal depended directly on the molecular geometry (i.e., the molecule occupied a special position in the mirror plane for formamide and on a 2-fold axis for pyrimidine). After local minimization, the structural parameters of the unit cells remained the same within 0.01 Å and 0.1°; the corresponding energies were lowered by less than 0.1 kcal/mol each. Because of the change in the molecular geometry, the symmetry of the unit cells for the two structures mentioned above was lowered to $P2_12_12_1$ for formamide and *Pc* for pyrimidine; all the other structures retained their previous symmetries. The experimental structures of formamide and pyrimidine were also reminimized with the experimental geometries, and the results were the same as those obtained with our symmetrized molecular model witin 0.01 Å and 0.1°.

5.3. Conclusions. The global optimization method presented here, SCBDBM, may be used as a reliable procedure not only for theoretical crystal structure prediction, but also as a tool for testing potentials, and, ultimately, for their refinements. It is a de novo type method, and in the case of crystal computations, it does not make use of any information other than the force field parameters and molecular geometry. All the structural parameters, space groups, and the number of molecules in the unit cell are results of the prediction.

As with many other potentials, the AMBER and DISCOVER force fields were parametrized to reproduce experimental structures without taking the features of the entire potential surface into account. This procedure may be misleading because the potentials derived in that manner may produce minima that are lower in energy than the minimized experimental structure;

Table 7. Parameters of Calculated Structures Obtained by Global Optimization Compared with Locally Minimized Experimental Structures for Pyrimidine with DISCOVER and AMBER[a]

method	space group	energy, kcal/mol	V, Å^3	D, D	E_D, kcal/mol	a, Å	b, Å	c, Å	α, deg	β, deg	γ, deg
				DISCOVER							
minimized exptl	$Pna2_1$, $Z = 4$	−14.19	406.7	3.5	0.23	11.21	9.94	3.65	90.0	90.0	90.0
I	$P2_1/n$, $Z = 4$	−15.01	409.6	0.0	0.00	3.97	10.35	9.97	90.0	86.6	90.0
II	$P2_12_12_1$, $Z = 4$	−14.89	412.1	0.0	0.00	5.76	7.07	10.11	90.0	90.0	90.0
III	$P1$, $Z = 4$	−14.74	415.0	0.0	0.00	6.97	6.97	9.86	69.3	110.7	90.0
minimized exptl	$Pna2_1$, $Z = 4$	−14.46	406.9	4.1	0.33	11.02	9.90	3.73	90.0	90.0	90.0
I	$P2_12_12_1$, $Z = 4$	−14.89	412.1	0.0	0.00	5.76	7.07	10.11	90.0	90.0	90.0
II	$P1$, $Z = 4$	−14.73	409.3	2.9	0.24	7.39	7.68	7.68	88.7	76.1	103.9
III	$P2_1$, $Z = 2$	−14.73	411.7	6.0	1.06	7.31	5.58	10.61	90.0	72.2	90.0
				AMBER							
minimized exptl	$Pn\underline{a}2_1$, $Z = 4$	−15.28	394.7	3.7	0.35	11.01	9.83	3.64	90.0	90.0	90.0
I	$P1$, $Z = 2$	−16.28	382.6	0.0	0.00	7.21	7.66	11.76	71.6	52.2	59.0
II	$I2/a$, $Z = 4$[b]	−16.27	382.6	0.0	0.00	6.66	9.30	6.84	90.0	64.5	90.0
III	$P\bar{1}$, $Z = 2$	−16.27	382.6	0.0	0.00	5.88	6.84	10.18	85.0	93.6	70.2
minimized exptl	$Pn\underline{a}2_1$, $Z = 4$	−15.64	394.9	4.0	0.34	10.88	9.84	3.69	90.0	90.0	90.0
I	$P\bar{1}$, $Z = 2$	−16.28	382.6	0.0	0.00	7.21	7.66	11.76	71.6	52.2	59.0
II	$P\bar{1}$, $Z = 2$	−16.27	382.6	0.0	0.00	5.88	6.66	10.42	83.9	85.4	108.4
III	$P1$, $Z = 4$	−16.15	328.5	0.0	0.00	6.77	6.91	9.96	80.4	70.2	60.7

[a] For symbols and units, see footnotes to Table 3. [b] Molecules occupy special positions on the 2-fold axis.

Table 8. Parameters of Calculated Structures Obtained by Global Optimization Compared with Locally Minimized Experimental Structures for Formic Acid with DISCOVER and AMBER[a]

method	space group	energy, kcal/mol	V, Å^3	D, D	E_D, kcal/mol	a, Å	b, Å	c, Å	α, deg	β, deg	γ, deg
				DISCOVER							
minimized exptl	$Pna2_1$, $Z = 4$	−14.69	198.6	4.2	0.66	10.66	3.38	5.45	90.0	90.0	90.0
I	$P2_1/a$, $Z = 4$	−15.81	188.7	0.0	0.00	6.31	5.04	6.59	90.0	115.8	90.0
II	$P1$, $Z = 4$	−15.75	191.8	0.0	0.00	5.39	6.15	6.28	84.8	72.1	75.4
III	$P2_1/c$, $Z = 4$	−15.74	192.7	0.0	0.00	3.39	5.64	10.51	90.0	85.1	90.0
minimized exptl	$Pn\underline{a}2_1$, $Z = 4$	−15.43	196.4	4.6	0.80	10.69	3.29	5.58	90.0	90.0	90.0
I	$P\bar{1}$, $Z = 2$	−15.75	191.8	0.0	0.00	5.39	6.15	6.28	84.8	72.1	75.4
II	$P1$, $Z = 2$	−15.72	190.2	0.0	0.00	5.45	5.60	6.45	88.1	90.9	74.9
III	$P2_1$, $Z = 2$	−15.60	195.9	4.5	0.86	6.02	5.64	6.41	90.0	63.9	90.0
				AMBER							
minimized exptl	$Pna2_1$, $Z = 4$	−18.75	190.7	0.2	0.00	10.53	3.50	5.17	90.0	90.0	90.0
I	$P2_12_12_1$, $Z = 4$	−19.17	190.4	0.0	0.00	10.32	3.53	5.22	90.0	90.0	90.0
II	$P2_1/c$, $Z = 4$	−19.06	190.9	0.0	0.00	3.61	9.33	5.67	90.0	87.2	90.0
III	$P2_1$, $Z = 2$	−18.99	189.0	0.6	0.02	6.25	5.24	6.26	90.0	67.2	90.0
minimized exptl	$Pna2_1$, $Z = 4$	−18.75	190.8	0.3	0.00	10.54	3.50	5.17	90.0	90.0	90.0
I	$P2_1/c$, $Z = 4$	−19.06	190.9	0.0	0.00	3.61	9.33	5.67	90.0	87.2	90.0
II	$P2_1$, $Z = 2$	−19.00	189.3	0.7	0.02	6.25	5.24	6.26	90.0	67.2	90.0
III	$P1$, $Z = 4$	−18.87	195.0	0.6	0.02	3.47	7.50	7.64	88.4	87.3	79.0

[a] For symbols and units, see footnotes to Table 3.

such lower-energy minima were not taken into account during the parametrization, and may not correspond to any real structures. To assess the quality of the potential (as described in the Introduction), it is necessary to confirm that the minimized experimental structure is either the global minimum of the potential energy surface or at least one of the lowest energy minima. This verification requires considerable computational effort, partly because of the high numerical expense to evaluate the values of the energy and gradient in the numerous local minimizations that have to be carried out. The SCBDBM method provides that verification because of the relatively small number of local minimizations required in the global search.

Another criterium to evaluate the potential is to compare the minimized experimental structure with the observed structure. The average deviations of the structural parameters obtained with both the DISCOVER and AMBER potentials were very similar and did not exceed 4% for all molecules except formamide. For this molecule, the deviations were larger and, in the case of the DISCOVER potential, the symmetry of the experimental structure was lost during minimization. In general, the observed structures were reproduced with reasonable accuracy; at the same time, both potentials were not able to capture the fine details of structures containing hydrogen bonds.

Since the global minimum corresponds to the minimized experimental structure, i.e., the predictions were successful, for imidazole, and maleic and succinic anhydrides, with the AMBER potential, in the global search with the correction term included, this suggests that the potential is reasonable for these molecules. For formamide two structures with almost equal energies (the energy difference being less than 0.01 kcal/mol) were found, one of them being the minimized experimental structure. In this case, it is hard to evaluate the prediction (and quality of the potential as well). The predictions for the same molecules under the same conditions with the DISCOVER potential, which were less successful, indicate flaws in the DISCOVER force field. For those same molecules, the existence of very low-energy structures with a high dipole moment with both potentials, in the runs in which the dipole correction term was omitted, suggests that calculating the electrostatic energy using the Ewald summation alone is incorrect.

This explanation does not pertain to the failure to predict the crystal structures of pyrimidine and formic acid in searches with no dipole moment correction. With both potentials, the dipole moment spherical correction terms of the locally minimized experimental structures cannot compensate for the differences between the energies of the minimized experimental structures

and those of the predicted nonpolar global minima structures. Clearly, the potentials are not adequate for these two molecules.

The present work does not provide an answer as to how to predict highly polar crystal structures, but it suggests that the dipole moment correction term may play a key role in all predictions, for polar and nonpolar crystals.

6. Computational Details

The program was parallelized on the fine-grain level; i.e., the energy and the gradient calculations were carried out in parallel.

All calculations were carried out on an IBM SP2 supercomputer at the Cornell Theory Center; the numerical expense depended on the size of the molecule, and the number of molecules in the unit cell. For maleic anhydride, with 4 molecules in the unit cell, full global optimization required 19 h, using 20 processors of the SP2 supercomputer. Typically, about 7000 local minimizations were carried out. Most of these minimizations were carried out while finding trajectories in the reversing and reversed-reversing procedures. In these cases, minimization is started from a previously minimized structure corresponding to a slightly different deformation parameter (slightly larger for the reversing procedure, slightly smaller for the reversed-reversing procedure), and, therefore, it is extremely fast. The real numerical expense is incurred during the local minimizations in the local search; the typical number of local searches was about 700.

All local minimizations were carried out using the SUMSL algorithm.[64] The CRYCOM program[65] was used for crystal structure comparison and space group determination.

Acknowledgment. This research was supported by grants from the National Science Foundation (MCB95-13167) and from the National Institutes of Health (GM-14312). The computations in this work were carried out at the Cornell Theory Center which is funded in part by NSF, New York State, the NIH National Center for Research Resources (P41RR-04293), the IBM Corp., and the CTC Corporate Partnership Program and the San Diego Supercomputer Center provided by the National Partnership for Advanced Computational Infrastructure (NSF cooperative agreement ACI-9619020). We thank K. D. Gibson for help in preparation of this paper and A. V. Dzyabchenko for making his CRYCOM program available for our computations.

JA9929990

(64) Gay, D. M. *ACM Trans. Math. Software* **1983**, *9*, 503.
(65) Dzyabchenko, A. V. *Acta Crystallogr.* **1994**, *B50*, 414.