

Global Optimization-Based Method for Deriving Intermolecular Potential Parameters for Crystals

Yelena A. Arnautova,[†] Jaroslaw Pillardy,[†] Cezary Czaplewski,^{†,‡} and Harold A. Scheraga^{*,†}

Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, and
Faculty of Chemistry, University of Gdansk, Sobieskiego 18, 80-952 Gdansk, Poland

Received: August 27, 2002

A new global optimization-based method for deriving parameters of atom–atom potentials for crystals has been developed. The method consists of two parts (force minimization and a novel Monte Carlo method, Vector Monte Carlo [VMC]) to minimize a vector target function consisting of three components. The first component depends on the rank of the minimized experimental structure among all structures found by a global search; the second one is a function that allows minimization of deviations between experimental and minimized experimental structures; and the third one is a penalty function that provides the best fit to the available heats of sublimation. The method enables the limitations of any chosen form of the potential to be evaluated and finds the best possible set of parameters for a given form. The method has been applied to sets of saturated hydrocarbon and ether molecules. For hydrocarbons, the resulting set of potential parameters describes observed crystal structures with high accuracy. Significant improvement was achieved in the case of ethers for which the experimental structures of the selected molecules (except dioxane) were stable upon local energy minimization and were found as global minima of the refined potential. Our results for dioxane are in agreement with those of Mooij et al. (*J. Phys. Chem. A* **1999**, *103*, 9883) and show that potential energy calculations with simple “6-exp-1” energy functions are not adequate for predicting the crystal structure of dioxane.

1. Introduction

Over the past decade, the problem of crystal structure prediction (CSP) for small organic molecules, based solely on molecular composition and potential energy, has been attracting considerable attention. The CSP blind tests organized by the Cambridge Crystallographic Data Centre^{1,2} to evaluate existing methods and programs demonstrated the significant recent progress achieved in this field. The results of the tests also emphasized the problems associated with CSP.

It would not be an exaggeration to say that practically any modern crystal computation is based on the method of atom–atom potential functions, which was proposed by Kitaigorodsky more than 35 years ago.³ Many different potentials ranging from simple empirical⁴ to fully ab initio⁵ can be found in the literature. Because experimental data (crystal structure parameters, sublimation enthalpies, and lattice frequencies) provide much information about nonbonded interatomic interactions in crystals, they have been used for many years for deriving potential parameters and evaluating their quality.^{6–16} Nonbonded potential parameters are usually fitted empirically to the crystal structures and sublimation data for a large range of organic compounds or for several groups of related molecules. Several potentials with all or some of the parameters obtained from ab initio calculations were proposed recently.^{5,17} Despite the abundance of available sets of parameters, development of accurate and transferable potentials remains a very difficult task.

The common feature of almost all methods mentioned above is the way in which the quality of the resulting potential

parameters is evaluated. For a long time, the main criterion has been the ability of a potential to reproduce observed crystal structures and sublimation enthalpies for the molecules under consideration. Only recently was it realized that this criterion is necessary but not sufficient^{5,18–20} to assess the quality of a potential. The main assumption used in CSP is that the observed structures correspond to lowest energy minima of a potential energy surface; therefore, an accurate and realistic potential should be able to reproduce this situation. Errors in potentials can make prediction virtually impossible since they often lead to a situation in which many crystal packings with energies much lower than the energy of the experimental structure will be found.

The main requirements for an accurate “predictive” potential can be formulated as follows:¹⁹ first of all, it should be able to reproduce the experimental crystal structure (and some other properties, such as sublimation enthalpies) with good accuracy, and second, the crystal structures corresponding to the low-energy minima found for the potential should represent possible crystal structures, and one of them, possibly the global minimum, should correspond to the most stable observed structure. These criteria¹⁹ were used for evaluation of some popular force fields. Mooij et al.¹⁸ also used them for developing ab initio potentials.

Unfortunately, for many classes of molecules, the traditional ways of deriving potential parameters do not always lead to accurate predictive potentials. By considering many experimental crystal structures for different molecules^{6–16} or large sets of dimers and trimers of different configurations,⁵ one may hope to obtain an accurate potential because, taken together, these approaches are supposed to provide sufficient information about interactions between different types of atoms at different

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

[†] Cornell University.

[‡] University of Gdansk.

distances and mutual orientations of molecules. However, calculations carried out by using some of the popular force fields^{18,19} lead to unsatisfactory results for many molecules. There are many possible reasons for this. In some cases, thermodynamic and kinetic effects can play a very important role. Another possible reason is the inadequate form and parameters of the potentials used. It was shown^{17,18,21} that use of more sophisticated potential functions (for example, those including a more accurate description of electrostatic interactions and those taking polarization into account) can greatly improve the results of predictions (i.e., relative stabilities of the predicted structures and accuracy of the structural parameters). Unfortunately, such potential functions are usually computationally very expensive and cannot be used for a global search. In many cases, it is not known if use of such sophisticated potentials is absolutely necessary for successful prediction or if reasonably good results can be achieved just by improving the parameters of simple potentials. Finally, it would be good to have a method for designing potentials, which would guarantee that the potentials will satisfy both main criteria discussed above.

Global search methods used for CSP provide much information about the potential energy surface of a crystal. This information can be used for refinement of potentials. For example, such an approach was used for development of force fields for proteins.²² In this approach,²² the potential parameters were optimized by an iterative algorithm that maximizes the free energy difference between the native structure and the nonnative states of the protein. A similar approach has been used simultaneously for developing new global optimization-based methods for parameter refinement for proteins²³ and crystals. The last of these methods is presented in this paper.

Two groups of molecules, saturated hydrocarbons and ethers, were selected for testing the new method for deriving potential parameters. Crystals of saturated hydrocarbons have been studied for many years.^{6,9,14} Because development of a force field usually starts by obtaining potential parameters for aliphatic carbon and hydrogen, a number of potentials describing carbon–hydrogen interactions can be found in the literature.^{6,9,11,24} In the approaches in which the properties of crystals were considered, the quality of the potentials was assessed only by minimization of lattice energies starting from experimentally observed structures.

The most detailed theoretical study of crystals of saturated hydrocarbons was carried out by Williams.^{9,14,25,26} Williams focused mostly on designing an accurate description of electrostatic interactions. It was shown^{25,26} that a poor fit is obtained for the *ab initio* electrostatic potential around *n*-alkane and cycloalkane molecules using only net atomic charges. The fit can be improved by using additional methylene bisector site charges. However, only a small improvement of the crystal structure parameters was achieved. Van de Streek et al.^{27,28} came to the conclusion that the crystal structures for long even-*n*-alkanes are not sensitive to electrostatic effects. Moreover, it was shown²⁷ that electrostatic interactions are not important for polymorphism of saturated hydrocarbons. At the same time, molecular dynamics calculations carried out by the same authors^{27,28} clearly indicated that thermal effects play an important role in crystals of saturated hydrocarbons.

A completely different approach to the description of intermolecular interactions in hydrocarbon crystals was proposed by Price.²⁹ A simple one center intermolecular potential for methane, with anisotropic repulsion and isotropic dispersion contributions, was designed and later adapted to model CH₃ and CH₂ groups in saturated hydrocarbons. The resulting

potential scheme was tested by local energy minimization starting from the experimental structures of *n*-pentane, *n*-hexane, and *n*-octane and appeared to perform as well as widely used isotropic atom–atom potentials.

There are not many CSP studies of saturated hydrocarbons. CSP has been carried out for *n*-hexane and *n*-octane by Chaka et al.³⁰ and for *n*-hexane, *n*-octane, adamantane, and iceane by Gao and Williams.³¹ Unfortunately, the authors^{30,31} evaluated the potentials used^{30,31} only by their ability to reproduce the experimental structures and did not provide any information about the energetic ranking of the minimized experimental structures. Crystal structures of 1,1'-biadamantane were considered by Gavezzotti³² who found the experimental structure as a result of the search, but no information was provided about the relative energies of other low-energy minima. Propane is probably the most extensively studied hydrocarbon molecule, and it was one of the target molecules in the first CSP blind test.¹ Surprisingly, its observed crystal structure was predicted correctly (i.e., it was found as one of three lowest energy minima) by only one program, using an *ab initio* potential. Additional results presented by Mooij et al.¹⁸ showed that empirical potentials cannot provide correct CSPs for propane.

Generally, the experimental structures of saturated hydrocarbons are stable upon local energy minimization with the existing simple force fields. However, it is not clear if these force fields are adequate from the point of view of CSP.

Much less information is available about the quality of existing potentials with regard to CSP for ether molecules. The crystal structure of the low-temperature form of dioxane was used by Williams¹⁵ for deriving potential parameters for oxohydrocarbons, but it was shown by Mooij et al.¹⁸ that empirical and *ab initio* potentials are inadequate for CSP of dioxane. The potentials were not able to reproduce observed relative stabilities of two dioxane polymorphs.^{33,34} Local energy minimization also led to quite large structural deviations for the high-temperature form.¹⁸ Free energy calculations for two dioxan polymorphs were carried out by van Eijck.³⁵ In this work,³⁵ the effects of lattice vibrations and thermal expansion were taken into account. The results showed that optimization of molecular geometry during energy minimization, together with inclusion of zero-point energy, led to some improvement of the calculated transition enthalpy. However, the enthalpy of transition between the two forms was still calculated with the wrong sign (−1.3 kJ/mol) compared to the experimental value (2.4 kJ/mol). It was also shown that the temperature-dependent vibrational energies are very similar for both forms and, therefore, do not change the relative stabilities of the two forms. It was suggested that the poor agreement between the calculated and experimental values of the transition enthalpy can be caused by use of the harmonic approximation in the free energy calculations. The experimental structures of tetrahydrofuran and 1,2-dimethoxyethane were found as global minima by Schmidt³⁶ and Arikawa et al.³⁷

In this paper, we present a new global optimization-based method for deriving nonbonded potential parameters. Our main goal was to design a method capable of finding a set of parameters for a given form of a potential, which would satisfy the following requirements: (i) the potential should be able to reproduce the experimental structure within a certain accuracy; (ii) 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; (iii) the energy value for the observed structure should be close to the

experimental enthalpy of sublimation. The new method was applied to derive accurate and transferable potentials providing that the experimental crystal structures of selected hydrocarbon and ether molecules correspond to lowest energy minima of the potential energy. No consideration is paid to entropic effects in these calculations.

2. Methods

2.1. Form of the Potential Energy Function. As pointed out in the Introduction, the availability of a highly accurate potential is crucial for successful CSP. An improvement in the quality of a potential can be achieved by using more sophisticated potential models. Taking into account that the search for the global minimum of the potential energy of a crystal requires a huge number of energy evaluations, the use of such potentials can be prohibitively expensive. Therefore, some balance between the complexity of the potential and the computational cost has to be found. The simple “6-exp-1” potential functions with net atomic charges are considered in this paper. Additional non-atomic charge sites can be used when necessary to represent the molecular electrostatic potential better.

Nonbonded potential parameters from the W99 force field¹⁴ were used as a starting point for refinement. Net charges were obtained by fitting the ab initio electrostatic potential in the space around the molecules. The SCF 6-31G** basis set (GAMESS³⁸) was used for the molecular potential calculations.

The pairwise additive function

$$V_{ij} = -A_{ij}r_{ij}^{-6} + B_{ij}\exp(-C_{ij}r_{ij}) + q_iq_jr_{ij}^{-1} \quad (1)$$

was used for modeling dispersion–repulsion and electrostatic interactions, respectively, between the i th and the j th sites of different molecules.

To decrease the number of potential parameters, combination rules for heteroatomic nonbonded interactions such as

$$A_{ij} = \sqrt{A_{ii}A_{jj}}; B_{ij} = \sqrt{B_{ii}B_{jj}}; C_{ij} = (C_{ii} + C_{jj})/2$$

are widely used. The use of the rules is not completely justified and may lead to lower accuracy in crystal calculations.^{17,21} Calculations with and without the use of combination rules were carried out in this work, and they are discussed in the Results section. In our calculations, the adjustable parameters were A_{ij} , B_{ij} , and C_{ij} in the case of 6-exp-1 potential functions (The program could also be used for refinement of the parameters of the 6-12-1 potential. In this case, the adjustable parameters are the well depths, ϵ_{ij}^0 , and equilibrium radii, r_{ij}^0). Details of the method used for lattice energy calculations can be found in the work of Pillardy et al.³⁹

2.2. Procedure for Deriving Potential Parameters. *Vector Monte Carlo (VMC) Method.* To obtain potentials satisfying the requirements discussed in the Introduction, a special procedure for deriving potential parameters has to be designed. It has to be able to minimize the value of a target function including three main components. The first component depends on the rank of the minimized experimental structure among all structures found by a global search; the second one is a penalty function providing the best fit to the available heats of sublimation; and the third one is a function allowing minimization of deviations between experimental and minimized experimental structures. For n molecules, $3n$ components must be considered.

Because our goal is to develop potentials for which the minimized experimental structure would be lowest in energy,

the energy gap between the minimized experimental structure and the lowest energy structure found by a global search (and different from the minimized experimental structure) can be used as the first component, $G_i(a)$, of the target function:

$$G_i(a) = E_{\text{me}}^i(a) - \min E_k^i(a) \quad (2)$$

where a is a set of potential parameters; E_{me}^i is the energy of the minimized i th experimental structure; and E_k^i is the energy of minimum k in the set of minima found by global optimization.

Deviations of lattice energies from sublimation enthalpies ΔH_{subl}^i are usually minimized by using a harmonic penalty function:

$$P_i = (E_{\text{me}}^i - \Delta H_{\text{subl}}^i)^2 \quad (3)$$

Different functions can be used to minimize the structural deviation. In the present paper, the following function was used

$$R_i = w_c \cdot \text{diff}_c + w_a \cdot \text{diff}_a + w_t \cdot \text{diff}_t \quad (4)$$

where w_c , w_a , and w_t are empirical weights.

$$\text{diff}_c = (|a - a_e| + |b - b_e| + |c - c_e| + |\alpha - \alpha_e| + |\beta - \beta_e| + |\gamma - \gamma_e|) \quad (5)$$

with a_e , b_e , c_e , α_e , β_e , and γ_e being the experimental unit cell parameters;

$$\text{diff}_a = \sum_{i=1}^{\text{nmol}} \Omega_i / \text{nmol} \quad (6)$$

$$\text{diff}_t = \sum_{i=1}^{\text{nmol}} (fr_i - fr_i^e)^2 / \text{nmol} \quad (7)$$

The quantities diff_c , diff_a , and diff_t correspond to unit cell parameters, orientations of all molecules in the unit cell, and translational positions (in fractional coordinates fr_i), respectively, and describe deviations between experimental (e) and minimized experimental structures. Rotational angles⁴⁰ Ω_i characterize similarities of molecular orientations in the experimental and minimized experimental structures. Ω_i is found as

$$\Omega_i = \arccos(w_{11} + w_{22} + w_{33} - 1)/2 \quad (8)$$

where w_{11} , w_{22} , and w_{33} are diagonal elements of the matrix $\mathbf{W} = \mathbf{R}_1^{-1}\mathbf{R}_2$, relating rotation matrices, \mathbf{R}_1 and \mathbf{R}_2 , of corresponding molecules in two structures.

The simplest way to construct the target function is to take a weighted sum of all three components. The main drawback of such an approach is the difficulty in choosing the proper weights or (in the case of variable weights) the way to adjust the weights during optimization. Another problem that emerges is connected with the method of target function optimization with respect to the parameters a . Because crystal structures and lattice energies depend on the parameters (a) of the potential, local energy minimization of all structures with respect to the structural parameters must be carried out for each new set a . However, in this situation, standard gradient-based minimization procedures are not applicable to local minimizations with respect to the potential function parameters a , because of numerical instability of the target function, which depends on the parameters a only implicitly.

To avoid the problem with weights, the target function can be represented as a vector \mathbf{F} consisting of $3n$ components f_i each of which can be optimized independently:

$$\mathbf{F} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{3n} \end{pmatrix} \begin{pmatrix} G(a) \\ P(a) \\ R(a) \end{pmatrix} \quad (9)$$

where $\mathbf{G} = (G_1, \dots, G_n)$; $\mathbf{P} = (P_1, \dots, P_n)$; $\mathbf{R} = (R_1, \dots, R_n)$; n is the number of different molecules used for the parameter refinement. Optimization of a vector target function can be carried out by using a Monte Carlo-based method.

Our procedure for deriving potential parameters, called VMC,²³ is based on the Metropolis Monte Carlo method and consists of the following steps: (1) random perturbation of a parameter a_j ; (2) calculation of a new value of the target function \mathbf{F} ; (3) many-dimensional Metropolis test for the new value of \mathbf{F} ; and (4) test for the "minimum" value of \mathbf{F} . In step 1, the parameter a_j chosen at random is multiplied by a random number from the interval $[1 - \text{pert}, 1 + \text{pert}]$. Some restrictions are applied to the possible values of the potential parameters. In the case of the 6-12 form of the potential, the parameters ϵ_0 and r_0 should have physical significance; therefore, they should be allowed to vary within predefined intervals. Because it is not so easy to define the allowed ranges of the values for parameters A_{ij} , B_{ij} , and C_{ij} of the 6-exp-1 potential, the only restriction applied was the requirement that the minimum of the potential for a given A_{ij} , B_{ij} , and C_{ij} should exist and it should lie within the interval from 1 to 5 Å. Calculation of a new target function \mathbf{F} (step 2) included energy minimization for all structures with the modified set of parameters a .

The acceptance or rejection of a proposed step is determined by an iterative procedure, called a many-dimensional Metropolis test. The algorithm for this test used in the VMC method is presented in Figure 1. The first part (corresponding to step 3) is an iterative procedure, which starts with Metropolis tests carried out for all components f_i . The probability of acceptance is calculated as follows:

$$p = \exp \left[-\frac{(\beta_i^{\text{new}} - \beta_i^{\text{old}})}{\beta_i^{\text{avg}}} \right], \text{ where } \beta_i^{\text{avg}} = \frac{|\beta_i^{\text{new}}| + |\beta_i^{\text{old}}|}{2} \quad (10)$$

A vector $\mathbf{d}=(d_1, d_2, \dots, d_{3n})$ is introduced to mark the components accepted as a result of the Metropolis tests. The variable d_i equals 1 (0) if the i th criterion has (not) been accepted. If all the criteria f_i have been accepted, the proposed move is accepted. If all are rejected, the proposed move is rejected. If some (but not all) criteria are accepted (values of d_i equal to 1), those accepted criteria are removed from the list of those to be tested, and the iteration continues. The Metropolis tests are repeated only for the components rejected in the previous iteration. The iterations are repeated until all the remaining criteria (or just one criterion) f_i are accepted or rejected. A larger number of Metropolis tests is performed for an f_i , which has a lower probability to be accepted; therefore, the total probabilities of acceptance for all components are similar after finishing the many-dimensional Metropolis test. Temperatures (β_i) can change independently for each component (\mathbf{G} , \mathbf{P} , or \mathbf{R}) of \mathbf{F} . During the optimization, the temperatures are adjusted dynamically, which means that the acceptance ratio is calculated every N_{it} number of Monte Carlo steps and the values of β_i are adjusted to keep the ratio in the range from 40 to 60%.

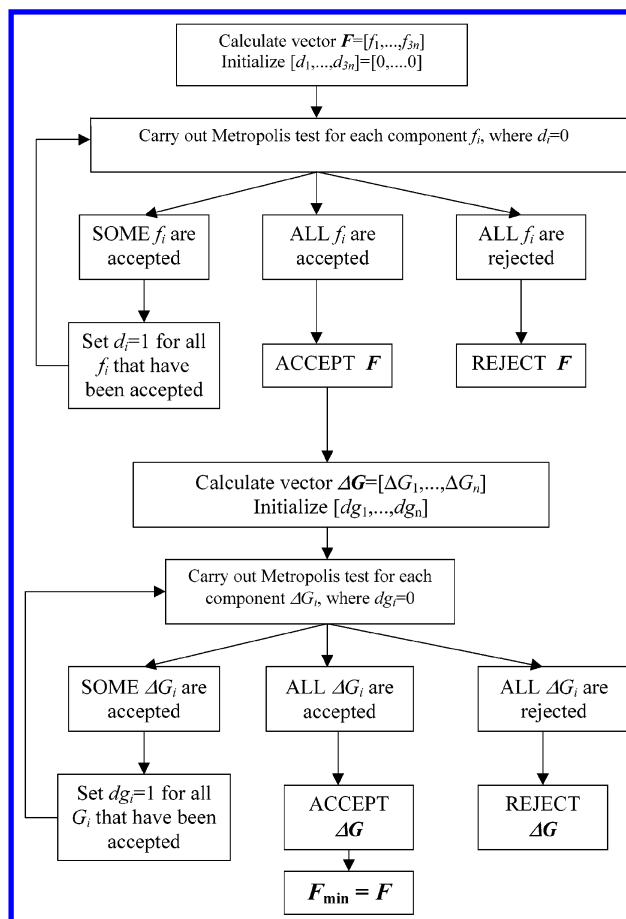


Figure 1. Flowchart of the many-dimensional Metropolis test.

If the new value of \mathbf{F} is accepted, it has to be compared with the minimum value of \mathbf{F} (step 4). In most cases, the starting set of parameters a (which is usually taken from existing force fields) provides acceptable accuracy for the minimized experimental structure (\mathbf{R}). The penalty function \mathbf{P} plays only a secondary role (see Section 3) in the parameter optimization; therefore, the minimum value of \mathbf{F} can be chosen based on the value of \mathbf{G} . In this part of the algorithm, the many-dimensional Metropolis test is applied only to the \mathbf{G} components of the target function \mathbf{F} . The temperatures used for this step are set according to the initial value of G_i for each molecule (for example, if G_i is small, a relatively high value of the temperature is used) at the beginning of each VMC run.

Steps 1–4 are repeated until convergence for the minimum value of \mathbf{F} is achieved. Optimization can be continued with lower temperatures for the \mathbf{R} component of \mathbf{F} to minimize structural deviations further. If the values of the \mathbf{G} and \mathbf{R} parts of the vector function \mathbf{F} are still very large and cannot be decreased, a different potential model should be considered. For example, more atom types or additional interaction sites can be introduced (such a situation occurred in the calculations for the hydrocarbon molecules and dioxane and is discussed in the Results section); in some cases, the form of the potential function has to be changed. This is a very important feature of the method, enabling one to pinpoint problems with a particular form of the nonbonded potential function.

Force Minimization. The algorithm described above includes local energy minimization, repeated for all minima at each step of refinement of the potential. This part of the method is very time-consuming and, if more than one molecule is considered, enables only a relatively small part of the parameter space of

the potential around a starting point to be explored. This means that the initial set of potential parameters should already be quite good. In some cases, one of the existing force fields can be used. However, for some molecules or even whole classes of molecules, the existing potentials give completely unsatisfactory results.

To find a good starting set (or sets) of parameters, the following procedure is used. Because a calculation of the two parts **P** and **R** of our vector function does not require minimization of energy for many structures, this can be carried out very rapidly. Thus, we can first search a very wide range of potential parameters *a* and try to find sets corresponding to low values of the function $FF(a) = \Sigma P_k(a) + \Sigma R_k(a)$. The function $FF(a)$ can be also rewritten in the form used by Williams and Houpt:⁴¹

$$FF(a) = \sum_k \sum_i w_i^k [F_i^k(x_i, a)]^2 + w' \sum_k (E_c^k(a) - \Delta H_{\text{subl}})^2 \quad (11)$$

where

$$F_i^k(x_i, a) = - \frac{\partial E_c^k}{\partial x_i}$$

is the equilibrium force due to a change of the energy E_c^k of the *k*th experimental crystal structure with respect to the *i*th structural parameter; x_i is a structural parameter; and w_i^k and w' are empirical weights. The requirement for $FF(a)$ to be as close to zero as possible means that all forces should vanish at the observed structures, i.e., these structures are local minima of the 6-exp-1 potential with parameters *a*.

From the point of view of a wide search range, form 11 of the function FF is more convenient than just a sum of **P** and **R** components since estimation of its value for a set of parameters *a* requires only one energy and one gradient evaluation for the experimental structure. In the current work, a random search was used to find sets of potential parameters corresponding to low values of FF . As the next step, we check if any of these sets corresponds to low values of our target function **F**. This is carried out by local energy minimizations for all structures found for the molecules under consideration. Again, a few (usually 2–3) best sets of parameters are selected and used as starting points for VMC parameter optimization.

Force minimization provides not only starting points *a* for parameter refinement but is also a tool for fast evaluation of the adequacy of a given potential form for a selected set of molecules. In some cases, a good parameter set can already be obtained as a result of force minimization. On the other hand, inability of the procedure to find parameter sets corresponding to low values of the function FF suggests that a different potential model should be considered. Both of these situations occurred for ethers (Section 4.4).

Summary of Procedure. The refinement of potential parameters using the VMC method requires the following information: (i) a starting set of potential parameters; (ii) parameters of experimental crystal structures for a given set of molecules; (iii) sublimation enthalpies (if available); and (iv) a set of minima found for the molecules by global search using the selected starting parameter set. The starting set of parameters can be taken from the literature or found by force minimization. In principle, refinement of potential parameters is an iterative process and includes two steps. The first step is refinement of parameters carried out by optimization of the target function **F** using the VMC method. It involves local energy minimizations

carried out at each step of the refinement for a set of minima found by the initial global search. The second step is evaluation of the resulting potential model. To evaluate new parameters, local and global energy minimizations have to be carried out for molecules used in the parameter refinement. This step is important because the minima used for the refinement are those obtained for the starting set of parameters. For any other parameter set, including the final one, the shape of the potential energy surface may be different, which means that low-energy minima other than those considered in the refinement procedure may exist. The minimized experimental structures for the molecules should correspond to the global minima or should be among the few lowest energy structures. If this is not the case, the refinement of the potential parameters has to be repeated using a new set of minima. If a good set of parameters (i.e., those satisfying all of the requirements listed in the Introduction) is obtained, global optimizations should also be carried out for similar molecules to check the transferability of the potential. In an ideal case, all important intermolecular interactions should be represented in the experimental crystal structures used in the refinement; therefore, transferability of the potential can be improved by extending the set of crystals used for refinement.

2.3. Global Optimization. According to eq 2, the function $G_i(a)$ defining the relative stability of the minimized experimental structure can be calculated if information about the spectrum of low-energy minima is available. Such information can be obtained by CSP calculations. A method best suited for this purpose would be one that focuses on the low-energy regions of the potential energy surface. In the present work, all global optimization runs were carried out by using the Conformation Family Monte Carlo (CFMC) method.³⁹ The method is designed for prediction of crystal structures of organic molecules without assuming any symmetry constraints except for the number of molecules in the unit cell. The central element of the CFMC method is a database of low-energy structures, which is an ensemble of structures clustered into families. The structures in this database are improved iteratively by a Metropolis type Monte Carlo procedure (including energy minimizations at each step) in which the search is biased toward the regions of the lowest energy families. The efficiency of the method has been demonstrated by predicting crystal structures of a set of rigid and flexible organic molecules.³⁹

Because our parameter refinement procedure requires as comprehensive a list of low-energy minima as possible, global energy optimization becomes the most important part of the method. To find all low-energy crystal packings for a given molecule, several runs with different number of molecules *z* in the unit cell have to be carried out. The global optimization runs for large *z* would be too time-consuming; besides, we are not guaranteed that our search method will be efficient for so high a dimensional space of structural parameters. Therefore, all global optimization runs for this paper were carried out for the same number of molecules *z* in the unit cell as in the experimental crystal structures. Two global optimization runs were carried out for each chiral molecule: one with the L (or R) enantiomer and the other one with the racemic mixture.

3. Experimental Data Used for Derivation and Validation of Parameters

The molecules used for deriving potential parameters are shown in Figure 2. The reference codes, space groups, and parameters of the experimental structures are given in Tables 1 and 2 and were obtained from the Cambridge Structural

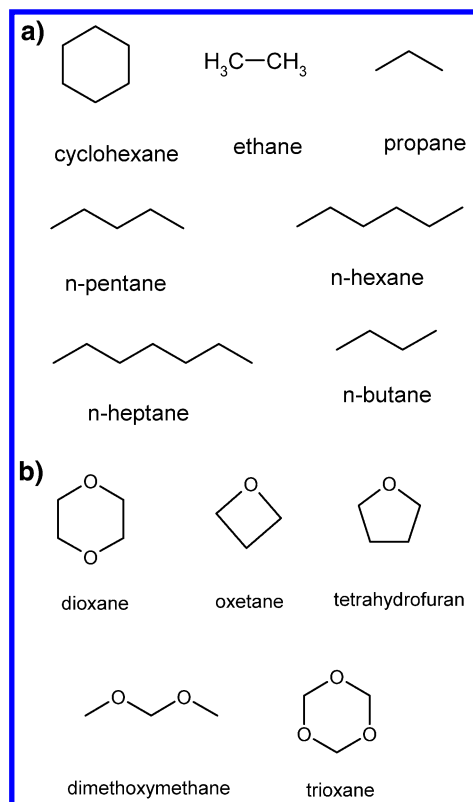


Figure 2. Molecules used for parameter refinement: (a) hydrocarbons and (b) ethers.

Database (CSD).⁴² If more than one experimental structure for a given compound was present in the CSD, the data corresponding to the lowest temperature (or lowest *R* factor) were selected. Molecules were considered to be rigid. Molecular structural parameters were taken as they are in the crystals except that C–H bond lengths were changed to give the average experimental value of 1.083 Å based on neutron diffraction data.⁴³

As seen from Tables 1 and 2, the types of packing vary considerably for the molecules under consideration. Practically all of the molecules considered here have symmetry higher than *C*₁ and retain it partially in the crystal state. In almost all crystal structures used, the molecules occupy special positions⁴⁴ on a symmetry element in the unit cell. Only in the cases of JAYDUI, HEPTANE02, and LIQVUC are the molecules in general positions.⁴⁴

Polymorphism is very common among the molecules considered in this work. Cyclohexane crystals can exist in three polymorphic forms:⁴⁵ two stable (forms I and II) and one metastable. The most stable low-energy form II exists below 186 K. The high-temperature form I is stable between 186 and 279.8 K. This polymorph is characterized by orientational disorder; therefore, we selected only form II for our calculations.

Ethane crystallizes in two polymorphic modifications.^{46,47} The high-temperature form is disordered and exists within a very narrow range of temperatures just below the melting point (*T* = 90.3 K). Again, the low-temperature modification was considered.

n-Butane can be obtained in three crystalline forms.⁴⁸ All three of them have two molecules in the unit cell and space group *P*2₁/*c*. The high-temperature phase I, which exists above 108 K, is disordered. There are two possible polymorphs, which may coexist below 108 K: the stable form III and metastable form II. The forms II and I have very similar unit cells and

TABLE 1: Results of Local and Global Energy Minimizations Carried out for Saturated Hydrocarbons with Various Potentials

	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	α (deg)	β (deg)	γ (deg)	<i>E</i> ^a	Δ <i>E</i> ^b	rank ^c
cyclohexane: CYCHEX ^d (<i>C</i> 2/ <i>c</i> , <i>z</i> = 4($\bar{1}$))									
exp	11.23	6.44	8.20	90.0	108.8	90.0	11.10		
W99	11.62	6.36	8.65	90.0	108.3	90.0	9.41	0.11	2
new	11.27	6.26	8.14	90.0	107.6	90.0	8.91	0.00	1
propane: JAYDUI (<i>P</i> 2 ₁ / <i>n</i> , <i>z</i> = 4(1))									
exp	4.15	12.61	6.98	90.0	91.3	90.0	3.41		
W99	4.41	12.85	6.97	90.0	91.5	90.0	5.89	0.05	3
new	4.18	12.51	6.71	90.0	90.9	90.0	5.36	0.00	1
Ethane: ETHANE01 (<i>P</i> 2 ₁ / <i>n</i> , <i>z</i> = 2($\bar{1}$))									
exp	4.23	5.62	5.85	90.0	90.4	90.0	4.90		
W99	4.50	5.68	5.71	90.0	91.3	90.0	4.50	0.00	1
new	4.25	5.54	5.57	90.0	90.7	90.0	3.88	0.05	3
<i>n</i> -pentane: PENTANE01 (<i>P</i> bcn, <i>z</i> = 4(2))									
exp	4.14	9.03	14.62	90.0	90.0	90.0	10.0		
W99	4.34	8.94	15.02	90.0	90.0	90.0	9.05	0.00	1
new	4.15	8.56	14.77	90.0	90.0	90.0	8.27	0.00	1
<i>n</i> -hexane: HEXANE01 (<i>P</i> $\bar{1}$, <i>z</i> = 1($\bar{1}$))									
exp	4.13	4.69	8.54	83.4	87.3	75.2	12.10		
W99	4.29	4.67	8.71	83.8	89.6	76.4	10.89	0.00	1
new	4.09	4.48	8.59	83.6	88.3	77.2	9.81	0.06	2
<i>n</i> -heptane: HEPTANE02 (<i>P</i> $\bar{1}$, <i>z</i> = 2(1))									
exp	4.12	4.69	20.35	78.1	81.8	74.3	13.80		
W99	4.27	4.66	20.63	78.1	84.7	76.7	12.24	0.00	1
new	4.09	4.46	20.35	78.4	83.9	77.2	11.11	0.18	2
<i>n</i> -butane, form III: DUCKOB (<i>P</i> 2 ₁ / <i>c</i> , <i>z</i> = 2($\bar{1}$))									
exp	4.11	7.62	8.09	90.0	118.6	90.0	17.16		
W99	4.40	7.46	8.36	90.0	118.9	90.0	17.16	0.20	3
new	4.19	7.43	7.97	90.0	119.6	90.0	15.74	0.15	2
<i>n</i> -butane, form II: DUCKOB02 (<i>P</i> 2 ₁ / <i>c</i> , <i>z</i> = 2($\bar{1}$))									
exp	5.71	5.17	7.87	90.0	105.9	90.0			
W99	5.73	5.24	8.27	90.0	107.9	90.0	17.56	0.0	1
new	5.56	5.07	7.86	90.0	106.7	90.0	16.05	0.0	1

^a *E* (kcal/mol) is the lattice energy of the minimized experimental structure for W99 and the new sets of potential parameters. In line exp, *E* is the experimental sublimation enthalpy.⁴⁹ ^b Difference between the energy of the minimized experimental structure and the energy of the lowest minimum found (kcal/mol). ^c Rank refers to the minimized experimental structure with respect to the global energy minimum.^d CSD reference code.

molecular packings. In these packings, the longest molecular axis is at almost 90° to that of the symmetry-related molecule. This is in contrast to phase III in which they are nearly parallel. Polymorphic forms III and II of the *n*-butane, as well as the *n*-heptane crystal structure, were used for checking the transferability of the potential parameters obtained for hydrocarbons.

Polymorphism is not so general among the crystal structures of ethers. Only in the case of dioxane was more than one crystalline form obtained experimentally.^{33,34} The low-temperature phase II of dioxane exists below 272.1 K. A second form (dioxane-I) exists in the temperature region from 272.9 to 284.1 K. Both phases have no disorder, and both were used for parameter refinement.

The crystal structure of trioxane⁴² was used for checking the transferability of the potentials obtained for ethers. The available enthalpies of sublimation (Δ*H*_{subl})⁴⁹ were used for energy scaling of the intermolecular potentials (Table 1). The traditional way of deriving experimental lattice energies (*E*) from enthalpies of sublimation using the formula

$$\Delta H_{\text{subl}} = -E - 2RT \quad (12)$$

is based on a number of approximations and is not very accurate.⁵⁰ On the other hand, the experimental error in measurements of Δ*H*_{subl} in many cases is larger than any

TABLE 2: Results of Local and Global Energy Minimizations Carried out for Ethers with Various Potentials^a

	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	α (deg)	β (deg)	γ (deg)	<i>E</i>	ΔE	rank
Dioxane, form II: CUKCIU02 ($P2_1/n$, $z = 2(\bar{1})$)									
exp	5.72	6.46	6.13	90.0	99.9	90.0	8.51		
W99	5.95	6.45	6.77	90.0	101.2	90.0	10.98	1.59	11
set1 ^b	5.84	6.19	6.66	90.0	105.9	90.0	11.85	1.37	9
set2	5.84	6.31	6.48	90.0	105.2	90.0	12.09	1.54	8
Dioxane, form I: CUCKIU10 ($P2_1/n$, $z = 2(\bar{1})$)									
exp	4.58	9.18	5.82	90.0	99.63	90.0	7.95		
W99	4.43	7.99	6.96	90.0	93.24	90.0	12.55	0.0	1
set1	4.27	9.21	5.83	90.0	95.46	90.0	13.14	0.0	1
set2	4.22	9.23	5.77	90.0	96.31	90.0	13.57	0.0	1
Oxetane: CIVXIO ($Pnam$, $z = 4(m)$)									
exp	8.62	6.38	6.04	90.0	90.0	90.0			
W99	9.09	6.81	6.26	90.0	90.0	90.0	7.68	0.49	12
set1	8.83	6.40	6.16	90.0	90.0	90.0	8.11	0.00	1
set2	8.77	6.41	6.11	90.0	90.0	90.0	8.24	0.00	1
Dimethoxymethane: LIQVUC ($P2_1/n$, $z = 4(1)$)									
exp	6.17	8.01	9.26	90.0	107.7	90.0			
W99	6.53	8.15	9.46	90.0	107.5	90.0	9.55	0.0	1
set1	6.44	7.92	9.07	90.0	109.6	90.0	10.20	0.0	1
set2	6.46	7.89	9.03	90.0	111.1	90.0	10.38	0.0	1
Tetrahydrofuran: BUNJAV01 ($C2/c$, $z = 4(2)$)									
exp	6.19	8.94	7.84	90.0	106.4	90.0			
W99	6.66	9.15	8.12	90.0	104.4	90.0	8.07	0.78	30
set1	6.44	8.87	7.76	90.0	107.7	90.0	8.77	0.00	1
set2	6.41	8.85	7.75	90.0	106.4	90.0	8.85	0.00	1
Trioxane: TROXAN ($R3c$, $z = 2(3)$)									
exp	6.03	6.03	6.03	101.1	101.1	101.1	13.8		
W99	6.31	6.31	6.31	102.0	102.0	102.0	10.65	0.0	1
set1	6.23	6.23	6.23	103.3	103.3	103.3	11.97	0.0	1
set2	6.23	6.23	6.23	103.1	103.1	103.1	11.57	0.0	1

^a For symbols and units, see footnotes to Table 1. ^b set1 and set2 refer to the two best sets of the potential parameters found by force minimization for the ethers.

correction applied to obtain the experimental value of the lattice energy. We followed the approach proposed by Gavezzotti⁵⁰ and used the experimental values of sublimation enthalpies in our calculations without any corrections, keeping in mind, however, that an average experimental error is about 2 kcal/mol. For this reason, the requirement that a potential should be able to reproduce experimental values of the sublimation enthalpies played only a secondary role in deriving parameters.

4. Results and Discussion

4.1. Local and Global Energy Minimizations. Before refining potential parameters for the selected molecules, we wanted to check how other force fields perform for these molecules. For this purpose, we chose the W99 force field of Williams.^{14,15} To assess the quality of the force field parameters, both local and global energy minimizations were carried out for all molecules shown in Figure 2. The results of these calculations are given in Tables 1 and 2.

The results show that the W99 force field performs quite well in the case of hydrocarbons. Structural deviations between the experimental and the minimized experimental structures are relatively small (maximum average deviation is 3.2% for DUCKOB). The largest deviation of 7.1% was obtained for the unit cell parameter *a* of DUCKOB. For all molecules, the original symmetry of the experimental structure was preserved after local energy minimization. The results of global optimization runs carried out for the seven hydrocarbon molecules are shown in the last two columns of Table 1. In all of runs, the minimized experimental structures were found among the three lowest energy minima. When the minimized experimental

structures did not correspond to the global minima (CYCHEX, JAYDUI, and DUCKOB), their energies were only slightly higher than the energies of the corresponding global minima. For *n*-butane, the W99 force field was not able to reproduce the observed relative stabilities of the two low-temperature polymorphs (forms II and III). The metastable form II (DUCKOB02) was found to be 0.2 kcal/mol more stable than lowest temperature form III (DUCKOB). From the point of view of the requirements formulated for a potential in the Introduction, the W99 force field performs well for the hydrocarbons. Even though the energy differences between the minimized experimental structures and the global minima for cyclohexane, propane, and *n*-butane are very small, we decided to try to improve the parameters of the force field using the VMC method.

The results of local and global energy minimizations carried out for five ether molecules, using the W99 force field, are given in Table 2. In the case of ethers, the W99 potential parameters gave rather unsatisfactory results. The deviations of the unit cell parameters of the minimized experimental structures from their experimental values are generally larger than for the hydrocarbons. The largest average deviation (10.6%) was obtained for CUKCIU10. The unit cell parameter *c* of this structure was almost 20% larger after energy minimization with the W99 parameters. As a result of the local energy minimization, a significant expansion of the unit cell took place for all structures.

The minimized experimental structure of dioxane form II (CUKCIU02) is 1.59 kcal/mol higher than the global minimum and was found as minimum number 11. The observed relative stabilities of the two polymorphs of dioxane were not reproduced, i.e., the high-temperature form I (CUCKIU10) was found to be the global minimum. It is much more stable than the low-temperature form II (according to the experimental data,⁵¹ form II is 0.56 kcal/mol more stable). The minimized experimental structures of oxetane and tetrahydrofuran corresponded to minima number 12 and 30, respectively. In general, the results of the global searches support the suggestion that the W99 parameters are not adequate for the selected ether molecules.

The global search runs carried out for the hydrocarbon and ether molecules using the W99 parameters provided the information about the potential energy minima necessary for the parameter refinement. Because hundreds of minima were found for each molecule, it would be computationally too expensive to consider all of them in the VMC procedure. Moreover, it will be shown in the next section that the VMC method does not require a complete set of high-energy minima in order to refine the potential parameters. Therefore, we selected, from the list of structures found by global search for each molecule, all packings with energies lower than the energy of the corresponding minimized experimental structure (E_{exp}) plus all structures with energies within 2 kcal/mol above E_{exp} .

4.2. Test of the VMC Method for One Molecule (Oxetane).

The VMC method for refining potentials was first tested with one molecule. We chose oxetane as a test molecule because the parameters of the W99 force field did not perform well for this molecule. In the first run of the refinement, the W99 parameters were used as a starting point. All together, 66 low-energy crystal packings found by the global search for oxetane were considered. The parameters A_{HH} , B_{HH} , C_{HH} , A_{CC} , B_{CC} , and C_{CC} of the potentials, as well as A_{OO} , B_{OO} , and C_{OO} , were allowed to vary, and the combination rules were used to calculate the parameters for heteroatomic interactions. As a result of the refinement, the experimental structure became the

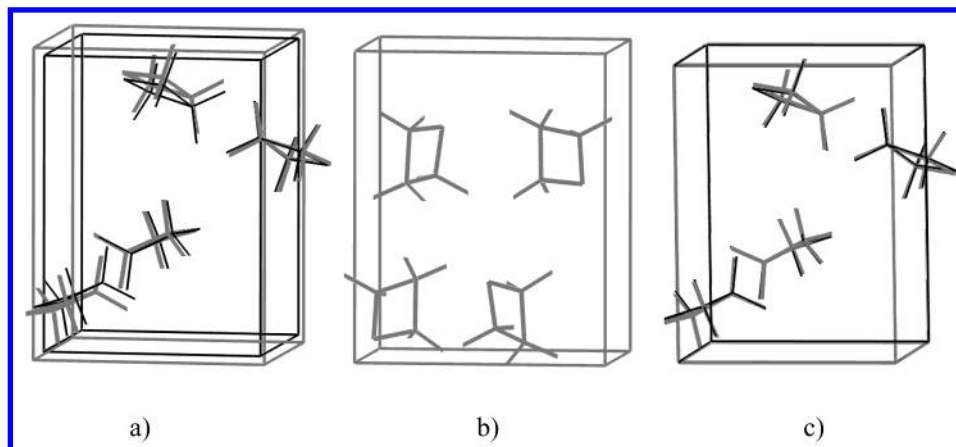


Figure 3. Crystal structures of oxetane: (a) experimental (grey) and minimized experimental structure (black) for the W99 potential (found as minimum 12); (b) global minimum structure for the W99 potential; (c) experimental (grey) and minimized experimental structure (black) for the refined set of potential parameters (found as the global minimum).

most stable one. Although the global search carried out with the resulting set of parameters showed that no new structures with energies lower than the energy of the minimized experimental structure appeared, the structural deviations between the experimental and the minimized experimental structures were very large.

An additional run at a lower temperature, which was set for the component R in order to reduce the structural deviation between the experimental and the minimized experimental structures, did not lead to a change in the order of minima. Several sets of parameters found by force minimization were also used as starting points for the VMC procedure. In all of these runs, the correct order of minima for oxetan was obtained only when a high temperature was set for the part R of the target function F (to allow large structural deviations). These results clearly show that the 6-exp-1 potential functions with the combination rules for heteroatomic parameters A_{ij} , B_{ij} , and C_{ij} are not able to describe the crystal structure of oxetane.

In the next run, the parameters of $H\cdots C$, $H\cdots O$, and $C\cdots O$ interactions were allowed to vary independently (no combination rules). As a result of this run, the set of parameters satisfying the three criteria listed in the Introduction was obtained. A subsequent global search carried out using these parameters showed that the minimized experimental structure was indeed the most stable one. The structural deviations were also small (Figure 3). Figure 4a,b shows how the R and G components of the target function F changed during the parameter refinement. Figure 5 shows changes in the energies of the different crystal structures of oxetane. The minimized experimental structure (CIVXIO) became the most stable one, i.e., the G component became negative, relatively early in the run (Figure 4b). It took some additional time to improve the structural parameters (component R decreased from 2.0 for iteration 50 to 0.6 at the end of the run, Figure 4a). At the end of the refinement, the changes in the R and G components were very small, which indicated that convergence was achieved. The energies of the high-energy crystal packings changed in a very similar way during the calculations (Figure 5). This suggests that even if some higher energy structures were missed during the global search, this should not affect the results of the potential refinement.

4.3. Optimization of Potential Parameters for Hydrocarbons. The VMC method was first applied to improve the potential parameters for aliphatic carbon and hydrogen. The first five molecules in Table 1 were used for the refinement, and the last two, *n*-heptane and *n*-butane, were used to check the

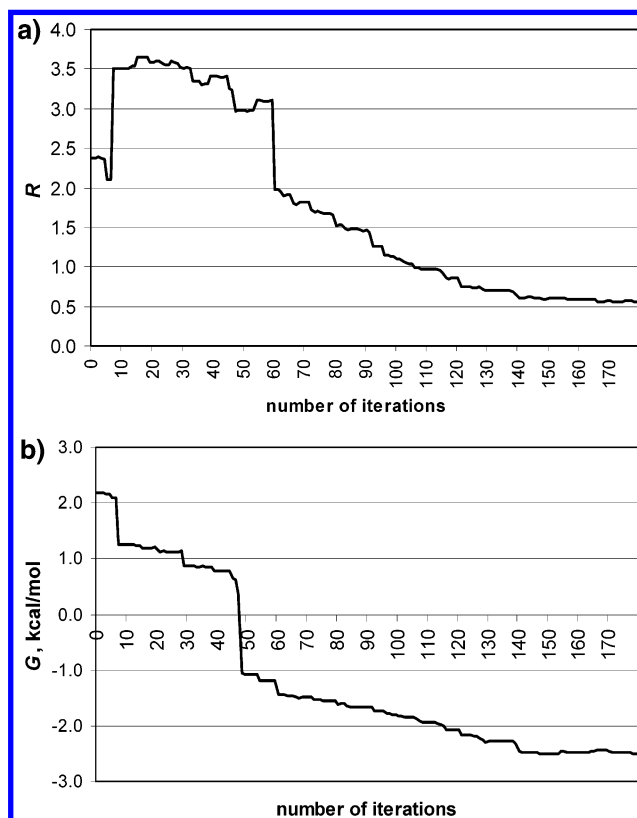


Figure 4. Changes in the R (a) and G (b) components of the target function for oxetane during parameter refinement.

transferability of the resulting parameters. The W99 parameters with the combination rules were used as a starting point in the VMC procedure. Unfortunately, no visible improvement of the potential parameters was achieved. In the next run, the parameters for $C\cdots H$ interactions were allowed to vary (no combination rules) as well. The final set of parameters was of similar quality as the initial W99 parameters.

To check if there is a set of parameters describing the hydrocarbon crystals better than W99, force minimization was carried out. In these calculations, all parameters of the potentials were varied independently. No set of parameters was found for which the minimized experimental structures of all hydrocarbon molecules under consideration would be global minima. The new set of parameters new (Table 3) was obtained as a result of the force minimization. A subsequent VMC run started from

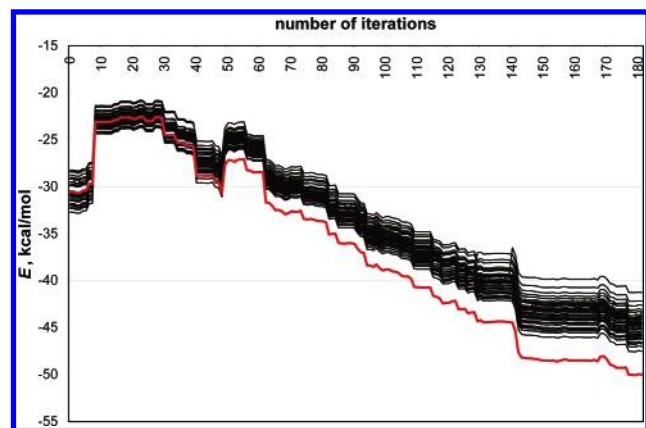


Figure 5. Changes in the energies of the oxetane crystal structures during potential refinement. The red line corresponds to the minimized experimental structure; the black lines correspond to the different crystal structures of oxetane found by the global search. Number of iterations refers to the iterations in which a new value of the function F was accepted.

TABLE 3: Nonbonded Potential Parameters Obtained as a Result of the Refinement for Saturated Hydrocarbons

interaction	new		
	A ($\text{\AA}^6 \cdot \text{kcal/mol}$)	B (kcal/mol)	C (\AA^{-1})
H...H	32.069	2996.798	3.906
H...C	78.749	26945.541	4.282
C...C	339.355	34318.479	3.331

the new set of parameters did not lead to any noticeable improvement of the results. The new set gave slightly smaller structural deviations (maximum average deviation 2.2%) between the experimental and the minimized experimental structures than W99 (Table 1). However, the discrepancies between the lattice energies and the corresponding sublimation enthalpies were larger than in the case of W99. By contrast with the results obtained for W99, the minimized experimental structures of cyclohexane and propane were found as global minima of the new potentials and the minimized experimental structures of ETHANE01 and *n*-hexane as minima number 3 and 2, respectively. The new potential parameters are transferable to other saturated hydrocarbon molecules (Table 1), although they were not able to reproduce the observed relative stabilities of the two *n*-butane polymorphs.

Figure 6 presents a comparison of the parameter sets W99 and new. At large distances, the curves nearly superimpose, but at short distances, the potentials look very different. The equilibrium distances are shorter for H...H and H...C interactions in the case of new (3.1 and 3.2 \AA , respectively) than in the case of W99 (3.2 and 3.6 \AA). The equilibrium C...C distance is unusually long (4.3 \AA) for the new set. The energy distributions between different types of interactions also differ for the two sets of parameters.

The force minimization revealed the existence of many almost equally good sets of potential parameters. This result suggests that strong correlations exist between the parameters A , B , and C^{41} of the potentials and also between different types of interactions (H...H, H...C, and C...C). Another explanation is that the interatomic interactions in crystals of saturated hydrocarbons are rather weak; this leads to a "flat" potential energy surface with very broad and shallow minima. It also follows from the fact that for many hydrocarbons, more than one polymorph was found experimentally. This tendency to form several polymorphic forms is visible from the results of the global search. For all seven molecules, many structures within

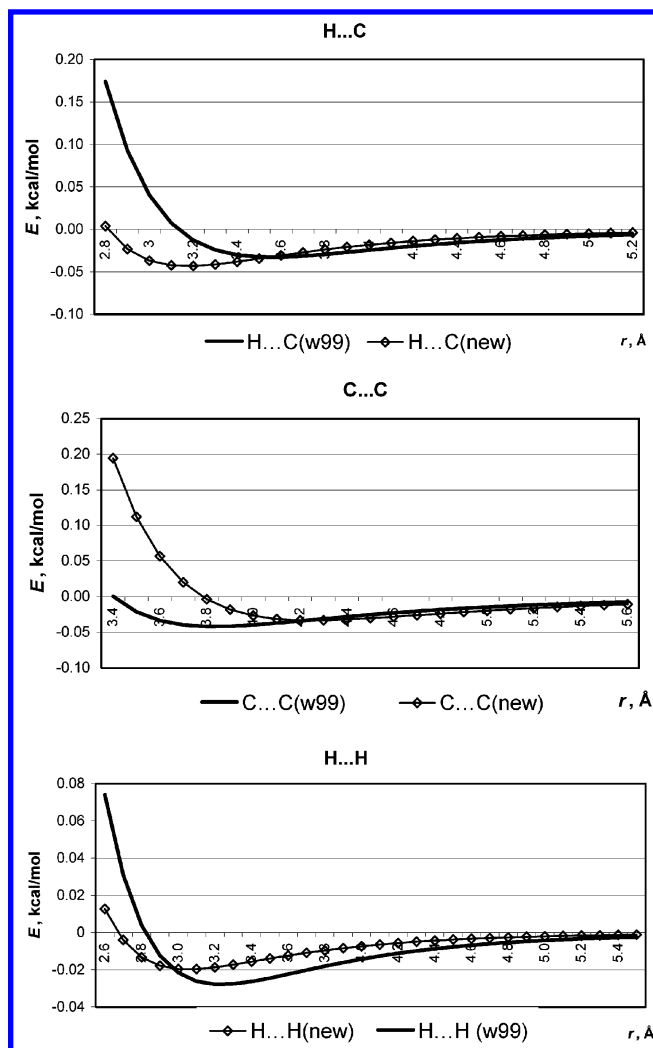


Figure 6. Dispersion-repulsion potential curves for H...C, C...C, and H...H interactions.

a narrow range of energies (energy differences between structures ≈ 0.05 kcal/mol) were found for all considered sets of parameters.

The results presented in Table 1 show that the simple 6-exp-1 potentials with the point charges located at atomic sites correctly describe the crystal structures of saturated hydrocarbons. However, they fail to reproduce the observed stabilities of different polymorphs of these molecules. There may be several possible explanations for these results. First of all, thermal effects play an important role in the crystals of hydrocarbons. Another explanation is associated with an inadequate description of electrostatic interactions. As a part of this work, we carried out local and global energy minimizations for ethane, cyclohexane, and *n*-pentane using the charge model proposed by Williams²⁵ and Williams and Abrahams²⁶ and the W99 and new potential parameters. The presence of the additional point charges^{25,26} did improve the fit to the electrostatic potential, but their effect on the structures and relative energies of different crystal structures was very small. These results are in agreement with the conclusion^{27,28} that electrostatic interactions are not very important for the polymorphism of saturated hydrocarbons.

4.4. Optimization of Potential Parameters for Ethers. The molecules used for the parameter refinement are shown in Figure 2b. The crystal packings found by global search for dioxane, oxetane, dimethoxymethane, and tetrahydrofuran were used in

TABLE 4: Nonbonded Potential Parameters Obtained as a Result of the Refinement for Ethers

Set1			
interaction	A ($\text{\AA}^6 \cdot \text{kcal/mol}$)	B (kcal/mol)	C (\AA^{-1})
H...O	148.373	24400.657	4.665
C...O	340.734	45432.793	3.162
O...O	125.299	48432.390	3.868
Set2			
interaction	A ($\text{\AA}^6 \cdot \text{kcal/mol}$)	B (kcal/mol)	C (\AA^{-1})
H...O	144.929	10681.594	4.493
C...O	322.245	57193.279	3.233
O...O	190.810	44642.327	3.681

the force minimization and the VMC procedures. Trioxane was selected for testing the transferability of the potential parameters.

The results presented in Table 2 show that the W99 force field is not quite adequate for the ether crystals. Therefore, as a first step of the potential refinement, we tried to improve the parameters of the W99 force field. We also want to design potentials that would be transferable between hydrocarbon and ether molecules; thus, in the first VMC run, the parameters of C...C and H...H interactions were kept fixed at the values given by the W99 force field. The combination rules were applied to calculate heteroatomic parameters; hence, the parameters for O...O interactions were the only variables in this run. Unfortunately, no noticeable improvement of the parameters was achieved. The force minimization carried out for these molecules also did not find any sets of parameters that were better than the starting point (the W99 parameters).

As the next step, the parameter set new obtained by the force minimization for the hydrocarbon molecules was considered. The parameters for H...H, C...H, and C...C interactions were kept fixed, and the force minimization was carried out with parameters for H...O, C...O, and O...O interactions as variables (without combination rules). Several different sets of parameters corresponding to low values of the target function F were found (the only exception was the value of component G for dioxane, which was relatively high for all sets). We selected two best sets (set1 and set2, Table 4) for global optimization. The results of the local and global energy minimizations for the ether molecules are shown in Table 2 (set1 and set2). The results of local minimizations for dimethoxymethane and trioxane are similar for the three sets of parameters (W99, set1, and set2). Significant improvement was achieved in the case of oxetane and tetrahydrofuran. For these molecules, the experimental structures minimized with the parameters from set1 and set2 were found as global minima of the potential energy. Average deviations of the unit cell parameters from their experimental values for CIVXIO and LIQVUC were also smaller for both new sets as compared with the results obtained for the W99 potentials. The only molecule for which almost no improvement took place was dioxane. To find out why this happened, we decided to consider this molecule separately. The results of these calculations are discussed in Section 4.5.

Figure 7 shows the differences between the W99 parameters and the new sets obtained by force minimization. The well depths for all types of interactions are smaller for set1 and set2 than for W99. The interaction between carbon and oxygen atoms, as described by both set1 and set2, is very weak with a long equilibrium distance (4.8 \AA). In the case of O...O interactions, the equilibrium distances are similar for all three sets of parameters (W99, set1, and set2), although the well depth is more than two times bigger for the W99 potential. The difference between the W99 potential parameters and the

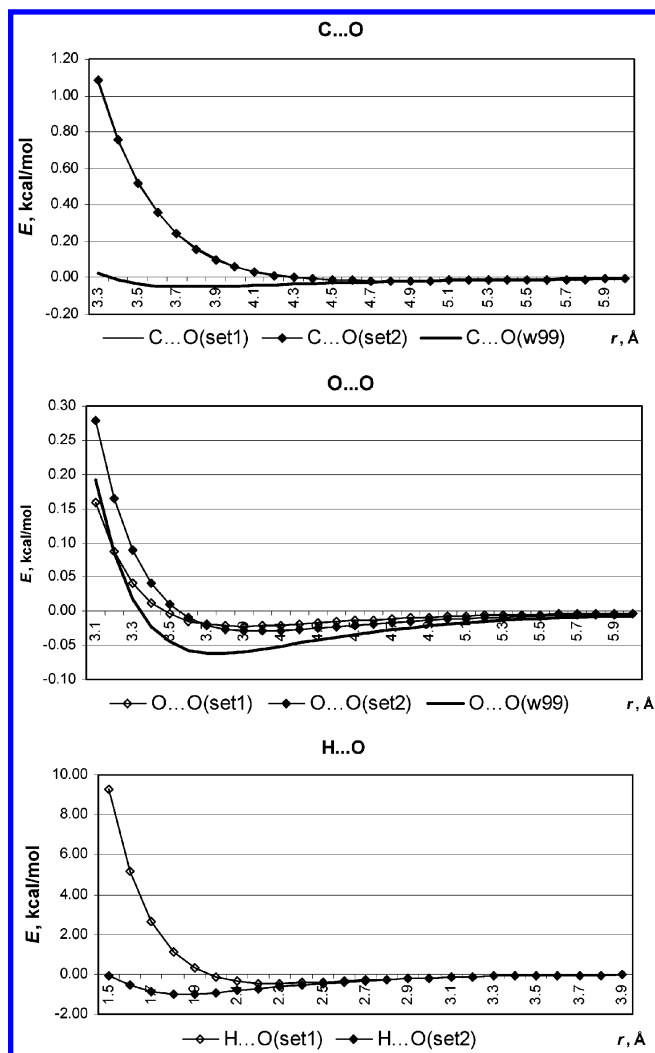


Figure 7. Dispersion–repulsion potential curves for C...O, O...O, and H...O interactions. The H...O curve for the W99 potential is not shown. The well depth and the equilibrium distance for the W99 potential are equal to -0.04 kcal/mol and 3.5 \AA , respectively.

parameters obtained by the force minimization is especially visible in the case of hydrogen–oxygen interactions. The new sets of parameters give significantly shorter equilibrium distances (2.3 and 1.9 \AA for set1 and set2, respectively) between oxygen and hydrogen atoms and much deeper potential wells than W99.

4.5. Refinement of Potential Parameters for Dioxane.

Dioxane is known to crystallize in two different forms. Form I is stable in the 11 K temperature range just below the melting point ($T = 284$ K). As temperature decreases, form II becomes more stable. However, the results of energy minimization carried out for both polymorphs with several sets of potential parameters showed that form I is always the most stable one with an energy difference of 1 – 2 kcal/mol. Moreover, form I was found as the global minimum. The main difference between the two forms is in the value of the electrostatic contribution to the total lattice energy. It is 1 – 2 kcal/mol larger for form I than for form II. The electrostatic energy of polymorph I is the largest among all of the structures found by global search.

Exactly the same behavior was noticed by Mooij et al.¹⁸ In their work, six different force fields (including DREIDING⁵² and OPLS⁵³) were used for CSP of dioxane. Two of the force fields were obtained by ab initio procedures and included, besides the usual dispersion and repulsion terms, atomic

TABLE 5: Results of Local and Global Energy Minimizations Carried out for Dioxane^a

	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	α (deg)	β (deg)	γ (deg)	E	ΔE	rank
Dioxane, form II: CUKCIU02 ($P2_1/n$, $z = 2(\bar{1})$)									
exp	5.72	6.46	6.13	90.0	99.9	90.0	8.51		
calcd	5.54	5.59	6.88	90.0	105.2	90.0	12.9	0.08	3
Dioxane, form I: CUCKIU10 ($P2_1/n$, $z = 2(\bar{1})$)									
exp	4.58	9.18	5.82	90.0	99.6	90.0	7.95		
calcd	3.94	8.89	5.95	90.0	99.5	90.0	12.7	0.20	5

^a For symbols and units, see footnotes to Table 1.

multipole moments for modeling electrostatic interactions and a polarization term. For all of the force fields, the results were very similar to those described in the previous paragraph. One of the explanations proposed by Mooij et al.¹⁸ was that their results “are not due to model errors, but that form I really has the best packing energy at 0 K. The only way to explain this ... is to assume that above a certain temperature dioxan-II is favored over such an optimized dioxan-I structure by vibrational entropy, and that above 273 K the true dioxan-I structure is again favored over dioxan-II”. A second explanation was that force field errors (i.e., errors in the repulsion term of the potential) cause the wrong energy ordering of the dioxane polymorphs. The first possibility was checked by van Eijck³⁵ who carried out free energy calculations for two dioxan polymorphs.

Our method of potential refinement allows us to check the second explanation and find out if potential energy calculations with 6-exp-1 potential functions designed specifically for dioxan can reproduce the observed crystal structures and stabilities for this molecule. We decided to carry out potential refinement separately for dioxane and allow all potential parameters to vary independently during the optimization. As a result of the VMC optimization, we obtained a set of parameters, which properly describes the relative stabilities of the two crystal forms of dioxane (Table 5). Form II was 0.2 kcal/mol more stable than form I. However, form II was found as minimum three with an energy of 0.08 kcal/mol higher than the energy of the global minimum. The correct stability of the two forms was also achieved at the expense of very large structural deviations between the unit cell parameters of the experimental, and the minimized experimental structures for both forms (Table 5, lines calcd).

We also tried different ways of modeling electrostatic interactions in dioxane crystals. For example, a global optimization run was carried out using the W99 force field and the charge model with additional sites near oxygen atoms to model electron pairs. Unfortunately, use of this charge model did not influence the results. Differences in the relative contributions of electrostatic and dispersion–repulsion energies to the total energy of different dioxane polymorphs suggest that general reduction of charges ($q_{\text{new}} = q \cdot k$, where $0 < k < 1$) would be favorable for the higher stability of form II over form I. Such reduction is equivalent to the use of a dielectric constant $\epsilon > 1$. We checked several values of the reduction coefficient k ranging from 0.5 to 1.0. The value 0.5 was selected since it had the strongest effect on the electrostatic contribution to the total energy. When the reduced charges were used, the energy difference between the two forms of dioxane became smaller (because of reduced electrostatic contribution for form I) although the observed stability of the dioxane polymorphs was again achieved only at the expense of large structural deviations of the structural parameters from their experimental values.

The results of calculations carried out for dioxane using simple 6-exp-1 potentials are in agreement with the results obtained by Mooij et al.¹⁸ for a set of known empirical and

newly developed ab initio potentials. They suggest that factors other than just errors in the potentials are responsible for the incorrect calculated relative stabilities of the dioxane polymorphs.

5. Conclusion

A new global optimization-based method for deriving parameters of atom–atom potentials for crystals has been developed. The method makes use not only of experimental information about structure and energy of observed crystal forms but also of information about the potential energy surface obtained as a result of global energy minimization. In principle, the method can be used in combination with any CSP procedure, which is able to provide a complete set of low-energy minima. However, it does not require exhaustive sampling of higher energy structures in order to optimize the parameters of the potential energy adequately.

The method includes two parts, force minimization and the VMC procedure, and enables one to find the best possible set of parameters for a given form of the potential. It is able to evaluate limitations of the chosen form of the potential and to provide indications of how to improve it (for example, by using a more accurate description of electrostatic interactions).

Two groups of molecules representing saturated hydrocarbons and ethers were considered. In the case of hydrocarbons, simple 6-exp-1 potentials with combination rules for heteroatomic interactions are able to describe the observed crystal structures with high accuracy; however, they cannot reproduce the observed relative stabilities of different polymorphs. The results of the calculations suggest that the correct description of hydrocarbon crystals can probably be achieved by taking thermal effects into account.

When the selected set of hydrocarbon molecules was supplemented by the set of small ethers in order to derive parameters for oxygen, the 6-exp-1 potential functions used with the combination rules were not sufficient to describe the crystal structures and the polymorphism of ethers. Remarkably good results were obtained when the combination rules were abandoned. For all of the ether molecules but one, the minimized experimental structures were found as the global minima of the optimized potential. The deviations of the unit cell parameters from their experimental values were also small. In the case of dioxane, the method was not able to find a set of parameters correctly describing both the crystal structures and the energy ordering of the dioxane polymorphs.

A separate parameter optimization was carried out for dioxane. All parameters of the potential were allowed to vary independently. As a result of the refinement, a set of parameters, which correctly describes the relative stabilities of the two crystal forms of dioxane, was found at the expense of encountering very large structural deviations. These results support the hypothesis that the unusual behavior of dioxane cannot be explained by errors in the parameters of 6-exp-1 potentials alone. The results of our and Mooij et al.¹⁸ calculations showed that static energy calculations with the simple 6-exp-1 energy functions as well as with more advanced ab initio potentials are not adequate for the CSP of dioxane. They also suggest that entropy effects can be responsible for the observed stability of the two dioxane polymorphs.

Acknowledgment. This research was supported by grants from the National Institutes of Health (GM-14312), the National Science Foundation (MCB00-03722), and the Fogarty Foundation (TW-1064). This work was carried out using computational

resources provided in part by (i) the Cornell Theory Center, which receives funding from Cornell University, New York State, and members of the Theory Center's Corporate Partnership Program, (ii) the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center, and (iii) with our own array of 193 dual-processor PC computers. We are grateful to Professor A. Liwo and Dr. A. V. Dzyabchenko for helpful discussions.

References and Notes

- (1) Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooij, W. T. M.; Price, S. L.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. *Acta Crystallogr. B* **2000**, *56*, 697.
- (2) Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Lommerse, J. P. M.; Mooij, W. T. M.; Price, S. L.; Scheraga, H.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. *Acta Crystallogr. B* **2002**, *58*, 647.
- (3) Pertsin, A. J.; Kitaigorodsky, A. I. *The Atom-Atom Potential Method. Applications to Organic Molecular Solids*; Springer-Verlag: Berlin, 1987; p 43.
- (4) Filippini, G.; Gavezzotti, A. *Acta Crystallogr. B* **1993**, *49*, 868.
- (5) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872.
- (6) Warshel, A.; Lifson, S. *J. Chem. Phys.* **1970**, *53*, 582.
- (7) Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319.
- (8) Momany, F. A.; Carruthers, L. M.; McGuire, R. F.; Scheraga, H. A. *J. Phys. Chem.* **1974**, *78*, 1595.
- (9) Williams, D. E.; Starr, T. L. *Comput. Chem.* **1977**, *1*, 173.
- (10) Lifson, S.; Hagler, A. T.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5111.
- (11) Cox, S. R.; Hsu, L.-Y.; Williams, D. E. *Acta Crystallogr. A* **1981**, *37*, 293.
- (12) Williams, D. E.; Cox, S. R. *Acta Crystallogr. B* **1984**, *40*, 404.
- (13) Ewig, C. S.; Thacher, T. S.; Hagler, A. T. *J. Phys. Chem. B* **1999**, *103*, 6998.
- (14) Williams, D. E. *J. Mol. Struct.* **1999**, *485*, 321.
- (15) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1.
- (16) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1154.
- (17) Tsui, H. H. Y.; Price, S. L. *Cryst. Eng. Commun.* **1999**, *1*, 24.
- (18) Mooij, W. T. M.; van Eijck, B. P.; Kroon, J. *J. Phys. Chem. A* **1999**, *103*, 9883.
- (19) Pillardy, J.; Wawak, R. J.; Arnautova, Y. A.; Czaplewski, C.; Scheraga, H. A. *J. Am. Chem. Soc.* **2000**, *122*, 907.
- (20) van Eijck, B. P. *J. Comput. Chem.* **2002**, *23*, 456.
- (21) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.
- (22) Hao, M.-H.; Scheraga, H. A. *J. Phys. Chem.* **1996**, *100*, 14540.
- (23) Pillardy, J.; Czaplewski, C.; Liwo, A.; Wedemeyer, W. J.; Lee, J.; Ripoll, D. R.; Arlukowicz, P.; Oldziej, S.; Arnautova, Y. A.; Scheraga, H. A. *J. Phys. Chem. B* **2001**, *105*, 7299.
- (24) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205.
- (25) Williams, D. E. *J. Comput. Chem.* **1994**, *15*, 719.
- (26) Williams, D. E.; Abraha, A. *J. Comput. Chem.* **1999**, *20*, 579.
- (27) van de Streek, J.; Verwer, P.; Bennema, P.; Vlieg, E. *J. Comput. Chem.* **2002**, *23*, 365.
- (28) van de Streek, J.; Verwer, P.; Bennema, P.; Vlieg, E. *Acta Crystallogr. B* **2002**, *58*, 677.
- (29) Price, S. L. *Acta Crystallogr. B* **1986**, *42*, 388.
- (30) Chaka, A. M.; Zaniewski, R.; Youngs, W.; Tessier, C.; Klopman, G. *Acta Crystallogr. B* **1996**, *52*, 165.
- (31) Gao, D.; Williams, D. E. *Acta Crystallogr. A* **1999**, *55*, 621.
- (32) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622.
- (33) Buschmann, J.; Müller, E.; Luger, P. *Acta Crystallogr. C* **1986**, *42*, 873.
- (34) Koritsanszky, T.; Strumpel, M. K.; Buschmann, J.; Luger, P.; Hansen, N. K.; Pichon-Pesme, V. *J. Am. Chem. Soc.* **1991**, *113*, 9148.
- (35) van Eijck, B. P. *J. Comp. Chem.* **2001**, *22*, 816.
- (36) Schmidt, M. U. *Kristallberechnungen Metallorganischer Molekülverbindungen*; Verlag Shaker, Technische Universität Aachen, Aachen, Germany, 1995.
- (37) Arikawa, T.; Tajima, N.; Tsuzuki, S.; Tanabe, K.; Hirano, T. *J. Mol. Struct. (THEOCHEM)* **1995**, *339*, 115.
- (38) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347.
- (39) Pillardy, J.; Arnautova, Y. A.; Czaplewski, C.; Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12351.
- (40) Dzyabchenko, A. V. *Acta Crystallogr. B* **1994**, *50*, 414.
- (41) Williams, D. E.; Houpt, D. J. *Acta Crystallogr. B* **1986**, *42*, 286.
- (42) Allen, F. H.; Kennard, O.; Taylor, R. *Acc. Chem. Res.* **1983**, *16*, 146.
- (43) Desiraju, G. R.; Steiner, T., Eds. *The Weak Hydrogen Bond in Structural Chemistry and Biology*; Oxford University Press: Oxford, 1999; p 507.
- (44) *International Tables for Crystallography*; Hahn, T., Ed.; Kluwer Academic Publishers: Dordrecht/Boston/Lancaster/Tokyo, 1987; Vol. A, p 25.
- (45) Kahn, R.; Fourme, R.; André, D.; Renaud, M. *Acta Crystallogr. B* **1973**, *29*, 131.
- (46) van Nes, G. J. H.; Vos, A. *Acta Crystallogr. B* **1978**, *34*, 1947.
- (47) Amoureux, J. P.; Foulon, M.; Muller, M.; Bee, M. *Acta Crystallogr. B* **1986**, *42*, 78.
- (48) Boese, R.; Weiss, H.-C.; Bläser, D. *Angew. Chem., Int. Ed.* **1999**, *38*, 988.
- (49) Bondi, A. *J. Chem. Eng. Data* **1963**, *8*, 371.
- (50) Gavezzotti, A., Ed. *Theoretical Aspects and Computer Modeling of the Molecular Solid State*; John Wiley & Sons: New York, 1997; p 97.
- (51) Jacobs, C. J.; Parks, G. S. *J. Am. Chem. Soc.* **1934**, *56*, 1513.
- (52) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- (53) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.