Derivation of a New Force Field for Crystal-Structure Prediction Using Global Optimization: Nonbonded Potential Parameters for Hydrocarbons and Alcohols

Yelena A. Arnautova,[†] Anna Jagielska,[†] Jaroslaw Pillardy,[‡] and Harold A. Scheraga*,[†]

Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, and Computational Biology Service Unit, Cornell Theory Center, Cornell University, Ithaca, New York 14853-1301

A new procedure has been developed for deriving accurate nonbonded potential parameters for crystal-structure prediction. The method consists of two steps. First, an initial set of potential parameters is derived by fitting to ab initio interaction energies of dimers. Second, this initial set is refined to satisfy the following criteria: the parameters should reproduce the observed crystal structures and sublimation enthalpies accurately; the observed crystal structure should correspond to the global or to one of the lowest-energy minima of the potential. These goals are achieved by applying our new global optimization-based method for deriving nonbonded potential parameters (*J. Phys. Chem. B* **2003**, *107*, 712), which consists of a force-minimization and a vector Monte Carlo [VMC] method, to information about the potential energy surface of the crystal. The procedure enables one to obtain the best possible set of parameters for crystal-structure prediction for a given form of the potential. The computed nonbonded parameters of the "6-exp-1" potential, derived by using the new procedure, are presented for aliphatic and aromatic hydrocarbons and alcohols.

1. Introduction

An accurate transferable force field plays a crucial role in many fields of theoretical and computational chemistry. Predictions of the conformations of biological molecules and the structures of crystals are probably two of the most challenging problems of which the solution is impossible without developing highly accurate potentials. The fact that accurate potentials still require further development and that success of almost any simulation depends strongly on the accuracy of the force field was once again demonstrated by the results of recent blind tests on the folding of proteins[1] and on a similar venture in crystal-structure prediction.[2,3]

In the present work, we focus only on one aspect of force-field development, namely, derivation of the nonbonded parameters of atom−atom potentials. Several different approaches have been proposed[4−28] to derive nonbonded parameters. They may be divided into two groups: methods based on ab initio calculations[22,23] and methods that rely on experimental data.[4−21,24−28] In the case of ab initio potentials, the parameters are fitted to reproduce the interaction energies of dimers or trimers of various molecules over a wide range of distances; therefore, it can be expected that the information about the global shape of the potential is sufficiently accurate. Another advantage of this method is its independence of experimental data, some of which are not always available. Because ab initio calculations essentially pertain to the gas phase (multibody effects can be taken into account, to a certain extent, by considering trimers), transferability of the resulting potential parameters to the condensed phase has to be checked carefully. It was shown by Mooij et al.[22] that, by including an accurate description of electrostatic interactions, polarization, and anisotropy of inter-

actions for some atom types in a potential model, it is possible to obtain nonbonded parameters that provide a highly accurate description of the crystal state. However, in the case of simpler potential models, which have to be used for global optimization of macromolecules and crystals, it may be expected that the accuracy of the ab initio parameters will not be extremely high.

The second group of methods for deriving nonbonded parameters relies on the use of experimental data available for the liquid and crystal states.[4−21,24−28] In some of the force fields, such as OPLS,[18] AMBER,[24] and GROMOS,[28] simulations of liquids were carried out to obtain parameters for the van der Waals part of the potentials. A large amount of experimental structural and thermodynamic data are available for crystals. They have long been recognized as an extremely important source of information about interatomic interactions and have been used for deriving nonbonded parameters.[5−17,19−27] Because it is reasonable to believe that some intermolecular interactions in crystals are essentially the same as the corresponding interactions in more flexible molecular aggregates (for example, nonbonded interactions within a protein molecule), experimental crystal data have been used not only for development of force fields for crystal calculations[5,6,9−13,15,25] but also for deriving van der Waals parameters of general all-atom potentials[7,8,16,17] for a wide range of molecules including proteins.

Practically all nonbonded potentials based on the use of experimental crystal data (e.g., crystal-structure parameters, sublimation enthalpies, lattice frequencies) were derived to reproduce these experimental values.[4−13,25−27] It was recognized only recently[22,29−31] that the ability of a potential to predict the observed crystal structure as the global or one of the lowest minima of the potential energy is as important as its ability to reproduce the observed crystal structure and sublimation enthalpy. Unfortunately, many of proposed force fields are still insufficiently accurate to predict the crystal structures of many molecules correctly.[2,3] There are at least several possible reasons why the traditional method for deriving potentials leads to

* To whom correspondence should be addressed. Tel: (607)255-4034. Fax: (607)254-4700. E-mail: has5@cornell.edu.
† Baker Laboratory of Chemistry and Chemical Biology.
‡ Computational Biology Service Unit.

unsatisfactory results for crystal-structure prediction. For some molecules, thermodynamic and kinetic effects can influence the crystal structure significantly. Another reason could be the inadequate mathematical form of the potential function to describe intermolecular interactions correctly. It was shown[23,29] that, in some cases, a more sophisticated description of electrostatics and inclusion of polarization can improve the results of prediction. However, a complicated potential form greatly extends the time of the calculations and limits the possibility to search the interaction-energy hypersurface exhaustively. Therefore, potentials for global search are usually an optimal compromise between accuracy and simplicity (i.e., efficiency) in global search-based methods. Finally, the reason for the failure of structure predictions can reside in the procedure used to derive the parameters of the potential function. In almost all methods, the main criterion for evaluating parameters has been the ability to reproduce experimental crystal structures and enthalpies of sublimation.[4-13,25-27] This criterion is a necessary but not a sufficient one.[29,30] By fitting potentials to reproduce experimental structures of many different molecules, one often does not obtain enough information about the interactions at all ranges of distances, from very short to very long. Therefore, such potentials are optimized only for the ranges of distances that occur in the experimental structures selected for the fitting. Such a procedure makes the method highly dependent on the availability of experimental data containing sufficient information about interactions over a wide range of distances. The fact that the interatomic interactions are not well reproduced for some distances can lead to the appearance of nonphysical minima with energies lower than the energy of the experimental structure. Such a potential makes prediction theoretically impossible, because the main assumption in global search-based methods is that the low-energy minima correspond to the observed crystal structures.

We define the necessary features of the potential function appropriate for crystal-structure prediction[30] as follows. First of all, the potential should reproduce experimental structures with a certain accuracy. Second, the lowest minima of the potential should represent the observed crystal structures with one of them (in an ideal case, the global minimum) corresponding to the most stable polymorph. An accurate potential should also reproduce the enthalpies of sublimation of experimental structures. These criteria have been used[30] to evaluate some popular, as well as newly developed, force fields.[22]

The idea of using the results of a *global* search, not only for potential evaluation but also for parameter refinement, was realized in the development of a new method for parameter refinement.[32] The method, called vector Monte Carlo (VMC),[32] makes use of information about the potential energy surface of the crystal available from global energy optimization and is designed so that the refined parameters satisfy all of the requirements listed above for an accurate potential. The method was tested[32] on two groups of molecules: aliphatic hydrocarbons and ethers. Significant improvement of the parameters of the potential was achieved for ethers for which, in almost all cases, the new potential predicted the experimental crystal structure as the global minimum. In the case of hydrocarbons, the new potential reproduced the experimental structures with very good accuracy; however, like already-existing potentials, it still could not reproduce the observed relative stabilities of different polymorphs of some hydrocarbon molecules. It was also shown[32] that the VMC method coupled with force minimization[32] enables one to evaluate the limitations of a given potential form and to find the best possible set of parameters. It was also

pointed out that the VMC optimization is rather time-consuming; thus, it is able to search only a limited space of parameters around a starting point efficiently (in this sense, our use of VMC alone is not strictly a *global* optimization procedure). Therefore, the performance of VMC depends on the quality of the starting set of parameters, which should already be quite accurate from the point of view of the criteria discussed earlier.

Recognizing the importance of developing an accurate general all-atom force field applicable to different systems (such as crystals and biological molecules), we decided to focus in this work on deriving the nonbonded potential parameters for some common types of atoms. We consider this work as a first step toward development of a general all-atom force field.

Taking into account the advantages of deriving potential parameters from ab initio calculations, the following combined approach was adopted in this paper. First, we derived a starting parameter set by fitting potential parameters to ab initio interaction energies of dimers of selected molecules. As the next step, the quality of the starting parameter set was evaluated by using local- and global-energy minimizations. If the parameters satisfied all of the requirements listed above, they were accepted as final; otherwise, the VMC method was used for their refinement. It may happen that ab initio calculations fail to provide a reasonable set of potential parameters. In this case, force minimization[32] can be used to obtain a starting set for the VMC procedure.

The "6-exp-1" form of the potential seems to be accurate enough to describe the nonbonded interactions for a broad range of molecules and still simple enough to be used in global-search calculations. Therefore, in this paper, we focused on improvement of the parameters for this form of the potential and on checking its limitations before using a more complicated mathematical description of nonbonded interactions.

Three groups of molecules were considered: aliphatic hydrocarbons, aromatic hydrocarbons, and alcohols. We present new parameters of an all-atom nonbonded "6-exp-1" potential function for the atom types present in these molecules. The improved potentials, obtained as a result of ab initio calculations followed by refinement using the global optimization-based VMC method, are able to reproduce the experimental structures and sublimation enthalpies with high accuracy, as well as to predict the experimental crystal structures as global minima of the potentials.

2. Method

The procedure used in this work for deriving nonbonded potential parameters involved the following steps: first, an initial set of force-field parameters is obtained by fitting to ab initio interaction energies of dimers; second, the accuracy of the parameters is evaluated by local- and global-energy minimizations for a set of small organic molecules containing the same types of atoms as the molecules in the dimers; third, if the accuracy of the resulting parameters is not sufficiently high, that is, if the potential does not satisfy all three requirements listed above, refinement using the VMC procedure is carried out with the ab initio parameters used as a starting point. The following sections describe the methods used at each step of the procedure.

2.1. Form of the Potential Energy Function. The energies of the dimers, as well as the lattice energies of the crystal structures, were calculated as a sum of pairwise atom−atom interactions. The energy of interaction between atoms i and j was considered to be a sum of nonbonded, that is, exchange

Derivation of a New Force Field

J. Phys. Chem. B, Vol. 107, No. 29, 2003 **7145**

repulsion and dispersion attraction, and electrostatic energies

$$E_{ij}^{\text{inter}} = E_{ij}^{\text{nbond}} + E_{ij}^{\text{electr}} \tag{1}$$

The nonbonded energy, E_{ij}^{nbond}, was modeled by using the "6-exp" potential function

$$E_{ij}^{\text{nbond}} = -A_{ij}r_{ij}^{-6} + B_{ij}\exp(-C_{ij}r_{ij}) \tag{2}$$

where r_{ij} is the distance between atoms i and j of different molecules (monomers); A_{ij}, B_{ij}, and C_{ij} are parameters of the potential. The electrostatic interactions were modeled by using the Coulomb formula

$$E_{ij}^{\text{electr}} = \frac{q_i q_j}{r_{ij}} \tag{3}$$

where q_i and q_j are point charges and the dielectric constant was taken as unity.

To reduce the number of parameters in a fitting procedure, combination rules for the nonbonded parameters A_{ij}, B_{ij}, and C_{ij} are used for describing heteroatomic interactions. The commonly used[3,9,14,25-27] combination rules

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

were selected for our calculations, where A_{ii}, B_{ii}, and C_{ii} are parameters of pairwise interaction between atoms of the same type i. For some systems, having strong interactions between different atom types (for example, for systems with intermolecular hydrogen bonds), the combination rules may provide a poor approximation for heteroatomic interactions. In such cases, the rules can be released at the stage of the VMC refinement of parameters; however, in this work, use of the combination rules for hydrogen-bonded atoms led to satisfactory results. A hydrogen-bonded atom was treated as a separate atom type, with different parameters, compared to the hydrogen atoms bonded to carbon. No special hydrogen bond potential was used; we assume that the "6-exp-1" potential includes a correct description of the interactions involved in a hydrogen bond.

Point charges, q_i, were localized on atoms and fitted to reproduce the electrostatic potential of the monomers, calculated with the Hartree−Fock wave function in the 6-31G* basis set. The *resp* method[24a] implemented in the Amber 6.0 program[24b] was used for fitting charges. Point charges derived in this way were used to calculate the electrostatic interactions (eq 3) between monomers in the dimers and in crystals.

2.2. Derivation of Initial Nonbonded Force-Field Parameters: Ab Initio Calculations. Initial nonbonded parameters of the force field were derived by fitting the "6-exp" potential function (eq 2) to the ab initio van der Waals energy landscape. The van der Waals energies at each point were obtained by subtracting the electrostatic part of the interaction energy from the total ab initio interaction energy.

Because the goal of the ab initio calculations was to obtain a starting point for further VMC parameter optimization, we limited the ab initio calculations to interactions between the monomers within the dimers. The multibody interactions are included by subsequent VMC parameter optimization. All calculations were carried out with the Gaussian program.[33] The dimers consisted of two identical monomers. The geometry of the isolated monomers was fully optimized at the level of the density functional theory (DFT)[34] with the B3LYP functional and 6-311G** basis set and was kept frozen in the dimer

geometries. DFT calculations have been shown[35] to reproduce the experimental structures for many groups of molecules well at reasonably short calculation times compared to other ab initio methods that include electron correlation. The geometries of the dimers were generated on a grid of the Euler angles and spherical coordinates of the second monomer. For each homodimer, a set of about 1000 configurations, covering interatomic distances from 2 to 10 Å, was used to calculate interaction energies. For each configuration of a dimer, the total energy was calculated at the level of Møller−Plesset perturbation theory up to second order (MP2) with the 6-31G* basis set. This level of calculation should include the major part of the dispersion energy[36] and, at relatively low computational cost, provide a sufficient level of accuracy for the interaction energy to derive an initial nonbonded parameter set. If the mathematical form of our force field describes the shape of the interaction energy surface correctly, possible discrepancies from the experimental values of the interaction energy should be corrected at the further stage of refinement (VMC) of the parameters. The ab initio interaction energy was calculated by subtracting the energies of the monomers from the energy of the dimer. The energies of the monomers in the configuration of the dimer were corrected for the basis set superposition error (BSSE), which can constitute a significant contribution to the interaction energy, especially for small basis sets.[37] The BSSE was corrected by using the Boys−Bernardi counterpoise method[38] in which the energy of the dimer and the energies of the monomers have to be calculated with the same basis set, the basis set of the dimer.

2.3. Fitting Method. The parameters of the "6-exp" potential (eq 2) were fitted to reproduce the nonbonded part of the ab initio interaction energy by minimizing the least-squares function:

$$F_{\text{fit}}(\mathbf{a}) = \sum_k [E_{\text{abinitio}}^{(k)}(\mathbf{r}_k) - E_{\text{nbond}}^{(k)}(\mathbf{a};\mathbf{r}_k)]^2 \tag{4}$$

where $E_{\text{abinitio}}^{(k)}(\mathbf{r}_k)$ is the nonbonded ab initio energy of the dimer in configuration k, $E_{\text{nbond}}^{(k)}(\mathbf{a};\mathbf{r}_k)$ is the nonbonded energy of the dimer in configuration k calculated with the potential (eq 2), $\mathbf{r}_k$ is the vector of atomic coordinates, and $\mathbf{a}$ is the vector of the parameters of the force field.

To improve the accuracy of the fit in the region of the minimum of the nonbonded energy, the sum of squares in eq 4 was weighted, using exponential weights w_k:

$$w_k = \exp\lfloor-(E_{\text{abinitio}}^{(k)}(\mathbf{r}_k) - E_{\text{min}})/c\rfloor \tag{5}$$

where E_{min} is the minimal energy in the set of energy values for the given type of dimer, and c is a constant. This type of weighting leads to better fitting in the vicinity of the energy minimum ($w_k \to 1$) and lower accuracy for the regions of the potential distant from the minimum ($w_k \to 0$). The value of the constant c was adjusted within the range from 5 to 20 separately for each group of molecules (e.g., aliphatic and aromatic hydrocarbons and alcohols) to obtain the best fit to the ab initio interaction energy.

Because the dependence of the nonbonded energy in the function F_{fit} (eq 4) on the parameters is nonlinear, the function F_{fit} was minimized with the secant unconstrained minimization solver (SUMSL)[39] minimizing procedure. Potentially, the function F_{fit} can have more than one minimum in the multidimensional space of the force-field parameters. Local minimization leads to the minimum that is closest to the starting point, which does not have to be the global minimum. To explore a wide range of parameter space, many local minimizations were carried

out, starting from the parameter sets generated on the grid of potential parameters. The initial search was carried out within the range from −50% to +50% of the values of the parameters A and B given by Williams[25] and within the range from 3 to 5 Å^{-1} for the parameter C (for the parameter A for the hydrogen atom connected to oxygen, having the value zero in Williams' force field, the range from 0 to 20 Å^6·kcal/mol was initially searched). Three or five equally displaced values were considered for each dimension in the parameter space (depending on the coarseness of the search and the dimensionality of the parameter space). To verify whether the resulting parameter sets gave physically reasonable equilibrium distances between pairs of atoms, a one-dimensional nonbonded "6-exp" function with the resulting parameters was minimized with respect to the atom–atom distance for each pair of atom types used at each given stage of the fitting procedure. After minimization, the parameter sets giving any equilibrium atom–atom distances longer than 5 Å or shorter than 1.8 Å or giving no minimum of the "6-exp" function were excluded; they were considered as "nonphysical" (except for the atom pairs involving the hydrogen atom attached to oxygen of which, as in many other force fields,[22,24,25] we allowed description only by electrostatic interactions and nonbonded repulsion (with the parameter $A = 0$). However, in most cases, minimizations from different starting sets of parameters led to the same minimum of the function F_{fit} and the same final set of parameters, providing physically reasonable equilibrium distances of the "6-exp" function. If more than one minimum was found in the parameter space, usually only one corresponded to a physically acceptable parameter set.

2.4. Global Energy Optimization. The potential derived from the dimer calculations is a gas-phase potential. It cannot be expected to be extremely accurate when applied to the crystal state. To check the transferability of the ab initio parameters to the condensed phase, local- and global-energy minimizations have to be carried out for all selected molecules. Global optimization also provides information about the spectrum of low-energy minima necessary for the parameter refinement.

In the present work, all global optimization runs were carried out by using the Conformation-Family Monte Carlo (CFMC) method.[40] This method is designed for predicting 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 local-energy minimizations at each step carried out by using the SUMSL algorithm) 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.[40] It has also been used as a part of the VMC parameter refinement method.[32]

All global optimization calculations in 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.

A detailed description of the method used for potential energy calculations has been published.[40] It should be noted that, in the "6-exp" potential function, the nonbonded energy of an atom–atom interaction may go to minus infinity at short

distances because of the much slower changes in the "exp" part of the potential compared with the r^{-6} part. This problem is extremely important during a global search when many structures with very short interatomic contacts are generated. One of the possible solutions of this problem (implemented in our program) is to use an "$a \cdot r^n + b$" type approximation of the nonbonded energy between $r = 0$ and $r = r_{max} + (r_{min} - r_{max})/4$, where r_{max} and r_{min} are the distances at which the maximum and minimum, respectively, of the "6-exp" function appear (in this work, $n = 6$). The parameters a and b should be chosen to ensure that the energy and its first derivatives are continuous. However, use of this approximation may lead to slightly higher (than real) values of the lattice energy for minimized structures. Therefore, the approximation was used only for the local-energy minimizations carried out during a global search. As a final step of the search, all structures found by our method were optimized using a pure "6-exp" potential.

2.5. Refinement of Potential Parameters: The VMC Method. To obtain potentials satisfying the requirements discussed in the Introduction, a special VMC procedure for deriving potential parameters was designed. The VMC method minimizes a target function $\mathbf{F(a)}$ (eq 6) consisting of the three main components optimized independently at different temperatures (if a set of n molecules is used for parameter refinement, the vector function $\mathbf{F(a)}$ has $3n$ components).

$$\mathbf{F(a)} = \begin{pmatrix} f_1 \\ f_2 \\ ... \\ f_{3n} \end{pmatrix} = \begin{pmatrix} \mathbf{G(a)} \\ \mathbf{P(a)} \\ \mathbf{R(a)} \end{pmatrix} \tag{6}$$

where $\mathbf{G(a)} = (G_1, ..., G_n)$, $\mathbf{P(a)} = (P_1, ..., P_n)$, and $\mathbf{R(a)} = (R_1, ..., R_n)$.

The first component $\mathbf{G(a)}$ (eq 7) of the vector is the gap between the energy of the minimized experimental structure and the lowest-energy minimum, found by a global search for a given molecule. At the end of the refinement, the gap should be negative, that is, the experimental structure should have the lowest energy among all minima found.

$$G_i(\mathbf{a}) = E^i_{m.e.}(\mathbf{a}) - \min E^i_k(\mathbf{a}) \tag{7}$$

where $E^i_{m.e.}$ is the energy of the minimized ith experimental structure; E^i_k is the energy of minimum k in the set of minima found by global optimization (the set does not include the minimum corresponding to the minimized experimental structure).

The second element $\mathbf{P(a)}$ (eq 8) of the minimized vector function is a harmonic penalty function on the deviation of the lattice energy from the experimental enthalpy of sublimation, ΔH^i_{subl},

$$P_i(\mathbf{a}) = (E^i_{m.e.}(\mathbf{a}) - \Delta H^i_{subl})^2 \tag{8}$$

The third element $\mathbf{R(a)}$ (eq 9) is the function minimizing the structural deviations between the minimized experimental structure and the experimental structure:

$$R_i(\mathbf{a}) = w_c \text{diff}_c(\mathbf{a}) + w_a \text{diff}_a(\mathbf{a}) + w_t \text{diff}_t(\mathbf{a}) \tag{9}$$

where w_c, w_a, and w_t are empirical weights.

$$\text{diff}_c = (|a - a_e| + |b - b_e| + |c - c_e| + |\alpha - \alpha_e| + \\ |\beta - \beta_e| + |\gamma - \gamma_e|) \tag{10}$$

Derivation of a New Force Field

J. Phys. Chem. B, Vol. 107, No. 29, 2003 **7147**

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

$$\text{diff}_a = \sum_{i=1}^{n\text{mol}} \Omega_i / z \qquad (11)$$

$$\text{diff}_t = \sum_{i=1}^{n\text{mol}} (\text{fr}_i - \text{fr}_i^e)^2 / z \qquad (12)$$

where z is the number of molecules in the unit cell. 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 angles[41] Ω_i characterize similarities of molecular orientations in the experimental and minimized experimental structures. Ω_i is found as

$$\Omega_i = \arccos(w_{11} + w_{22} + w_{33} - 1)/2 \qquad (13)$$

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 target function $\mathbf{F(a)}$ of eq 6 is minimized using the VMC method.[32] The refinement of potential parameters using the VMC method requires the following information: (1) a starting set of potential parameters; (2) parameters of the experimental crystal structures for a given set of molecules; (3) sublimation enthalpies (if available); (4) a set of minima found for the molecules by a global search using the selected starting set of potential parameters (this set of minima was restricted to energies that were 3 kcal/mol above the global minimum; if the experimental structure was not the global minimum, then energies of 3 kcal/mol above the energy of the minimized experimental structure plus all minima below the energy of the minimized experimental structure). In principle, refinement of the potential parameters is an iterative process and includes two steps. The first step is refinement of the parameters carried out by optimization of the target function $\mathbf{F(a)}$ by 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 are carried out for molecules used in the parameter refinement. The whole procedure may be iterated, if necessary, by using the new set of minima and the new potential parameters as a starting point. After a good set of parameters (i.e., those satisfying all of the requirements listed in the Introduction) is found, global optimizations are also carried out for similar molecules to check the transferability of the potential.

To find a good starting set (or sets) of parameters for the VMC refinement, ab initio calculations can be used. However, it may be expected that, in some cases, they may not be able to provide a reasonable starting set of parameters. In this case, the following procedure, called force minimization and presented earlier,[32] can be used. First, a wide-range random search is carried out to find sets of potential parameters $\mathbf{a}$ corresponding to low values of the function:

$$\text{FF}(\mathbf{a}) = \sum_k \sum_i w_i^k [F_i^k(x_i,\mathbf{a})]^2 + w' \sum_k (E_e^k(\mathbf{a}) - \Delta H_{\text{subl}})^2 \qquad (14)$$
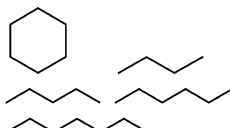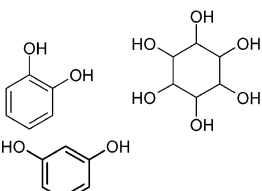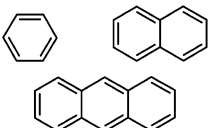
where

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

is the equilibrium force due to a change of the energy E_e^k of the kth experimental crystal structure with respect to the ith structural parameter; x_i is a structural parameter; and w_i^k and w' are empirical weights. The requirement for FF($\mathbf{a}$) to be as close to zero as possible means that all forces should vanish at the observed structures, that is, these structures are local minima of the "6-exp-1" potential with parameters $\mathbf{a}$. Second, we check whether any of the resulting sets corresponds to low values of the target function $\mathbf{F(a)}$ (eq 6). This is carried out by local-energy minimizations for all structures found for the molecules under consideration. Again, a few (usually two to three) best sets of parameters are selected and used as starting points for VMC parameter optimization.

3. Experimental Data Used for Derivation of Parameters

The molecules used for potential-parameter derivation and evaluation are shown in Table 1. The corresponding experimental crystal structures were taken from the Cambridge Structural Database (CSD).[42] The molecules were selected according to the following criteria: (1) the molecules should contain only C, H (or D), and O atoms; molecules with D were used only for geometry, but the potential parameters for D were taken as those for H (the atom types are listed in Table 1); (2) the observed structures should have no disorder; (3) in the case of X-ray diffraction data, the discrepancy factor R should be less than 5% (if several structures are available for a given molecule, the one with lowest R-factor was used); (4) if several structures for a given molecule obtained at different temperatures are available, the one corresponding to lowest temperature is used; (5) the structure obtained by neutron diffraction is preferred, if available; (6) crystal structures with the number of molecules in the unit cell (z) less or equal to 4.

TABLE 1: Atom Types, Molecules Used in ab Initio Calculations, and Molecules Used for Evaluation and Refinement of the Potential Parameters

Atom type	*Ab initio* energies of dimers[a]	Test molecules[b]
Aliphatic carbon (C1) and hydrogen (HC)	CH$_4$ H$_3$C—CH$_3$	
Oxygen bonded to two other atoms, one of which is hydrogen (O1); Hydrogen bonded to oxygen (HO)	HO-CH$_3$	
Aromatic carbon (CA); Hydrogen bonded to aromatic carbon (HA)		

[a] These molecules were used for deriving ab initio parameters.
[b] These molecules were used to test for transferability of the derived parameters and for parameter refinement.

Because the main assumption used in our parameter refinement procedure is that the observed crystal structure for a given molecule corresponds to the global minimum of the potential energy, the possible existence of another unknown crystal polymorph for this molecule can create difficulties during the parameter refinement. This may happen when the experimental structure used for minimization of the function $G_i(\mathbf{a})$ for a given molecule i is not the thermodynamically most stable crystal form. In this case, the VMC procedure will not be able to further decrease the value of $G_i(\mathbf{a})$, which will always be positive. Such behavior may be interpreted not only as an indication of the possible existence of another more stable polymorph but also as a sign of the inadequacy of the potential model used. To avoid this ambiguity, the crystal structures of only experimentally well-studied molecules were selected for our calculations. The literature available for these molecules was carefully checked to find any information about the existence of more than one polymorphic form and about their relative stabilities.

Because the goal of this work is to obtain van der Waals potential parameters, all molecules were considered to be rigid. Molecular structural parameters were taken as they appear in the crystals, except that C−H and O−H bond lengths were changed to the average experimental values of 1.083 and 0.983 Å, respectively, based on neutron diffraction data.[43]

The selected experimental structures demonstrate a variety of molecular packings (Tables 3−5). Molecules of almost all aliphatic and aromatic hydrocarbons have symmetry higher than C_1 and retain it partially in the crystal state, that is, flexible molecules such as aliphatic hydrocarbons may lose some of the gas-phase symmetry but retain some of it because of crystal packing. As a result, they often occupy special positions[44] on a symmetry element in the unit cell (Tables 3 and 4). In the case of alcohols, all molecules except YEPNOW have symmetry C_1 in the crystal (Table 5).

Many of the hydrocarbon molecules considered here can exist in several different polymorphic forms. When the experimental data for more than one polymorph of a particular molecule were available, the most stable form was used. The only exception was *n*-butane for which good experimental data are available for the most stable form III (DUCKOB) and the metastable form II (DUCKOB02), and both were selected (Table 3) for testing the resulting parameters.

Polymorphism is not as common among the crystal structures of alcohols. Only in the case of methanol was there more than one experimental crystalline form.[45−47] Methanol crystallizes in at least three, or perhaps four, different forms. The most stable α-form exists at 1 atm pressure and temperatures below 156 K. In this form, the chains of hydrogen-bonded molecules are puckered[46] (Figure 1A). As temperature increases, the methanol crystal undergoes a polymorphic phase transition to the high-temperature β-form. The transition involves a change from puckered chains to chains that are planar on average (due to the large thermal motion perpendicular to the plane of the chains),[46] and the planar chains are reoriented so that they become parallel to each other. The high pressure γ-form of methanol exists above 3.5 GPa.[47] Neutron diffraction data, available for the most stable α-form, were used in this work for deriving the van der Waals parameters for alcohols.

The available enthalpies of sublimation (ΔH_{subl})[48] were used for energy scaling of the intermolecular potentials (by P_i of eq 8, Tables 3−5). Because there are many uncertainties in comparing lattice energies and corresponding sublimation enthalpies, we decided to follow the approach proposed by Gavezzotti[49] and assumed that $\Delta H_{subl} \approx -E$. Taking into
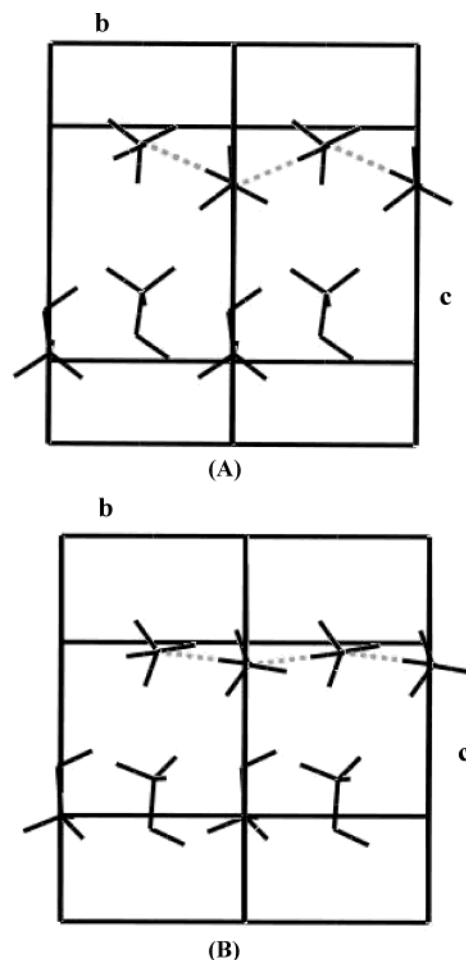

Figure 1. Crystal structures of α-form of methanol: (A) experimental; (B) minimized experimental structure with the *final*0 set of potential parameters.

account that an average experimental error is about 2 kcal/mol or more, a deviation of a few kilocalories per mole between the lattice energy and the corresponding sublimation enthalpy was considered as acceptable.

4. Results and Discussion

4.1. Derivation of Initial Values of Potential Parameters by ab Initio Calculations. Initial parameters, based on ab initio calculations described in section 2.2, were derived sequentially for small sets of molecules containing a few atom types at each step of the derivation. The molecules chosen for ab initio calculations of the interaction energy in dimers are shown in Table 1 (second column). First, the parameters for aliphatic carbon (C1) and hydrogen connected to aliphatic carbon (HC) were derived by fitting to the ab initio interaction energy for dimers of methane and dimers of ethane. Then, these parameters were fixed and used for the carbon and hydrogen atoms in methyl groups at the next stages of derivation of the parameters for the other atom types: oxygen connected to two other atoms one of which is hydrogen (O1) and hydrogen connected to oxygen (HO). The parameters for the O1 and HO atom types were derived from the ab initio interaction energy of methanol dimers. To check for the transferability of the C1 and HC parameters from aliphatic hydrocarbons to alcohols, the values obtained at the first stage for hydrocarbons were used as the initial parameters for the methyl group to model the interactions in the methanol dimers, and *all* parameters were then optimized. Because this relaxation of the C1 and HC parameters did not

Derivation of a New Force Field

J. Phys. Chem. B, Vol. 107, No. 29, 2003 **7149**

TABLE 2: Potential Parameters Obtained by ab Initio Calculations (ab initio), and the Final Parameters after VMC Refinement or Force Minimization (*final* and *final*0)

atom type	method	A, Å^6·kcal/mol	B, kcal/mol	C, Å^{-1}
		Hydrogen		
HC	ab initio (same as final)	13.89	1567.7	3.81
HA	ab initio	10.75	3335.9	4.36
	final	41.75	3392.1	3.88
HO	ab initio	14.10	63.2	3.06
	*final*0^a	0.00	2238.8	2.59
	final^a	3.25	137.8	3.69
		Carbon		
C1	ab initio (same as final)	650.65	30869.1	3.16
CA	ab initio	457.30	40605.9	3.32
	final	445.65	50360.5	3.39
		Oxygen		
O1	ab initio	372.00	11403.8	3.15
	*final*0^a	716.81	87027.8	3.64
	final^a	332.72	20231.9	3.46

^{a} Two sets of parameters corresponding to the models with zero (*final*0) and nonzero (*final*) attraction for alcohol hydrogen are given for HO and O1 atom types.

change their values significantly (the differences being smaller than 10% of the initial values), the best parameter set for C1 and HC atom types obtained for hydrocarbons was used in further calculations. To include a description of hydrogen bonds with the "6-exp-1" potential in the parameters for O1 and HO atom types, special attention was paid in the ab initio calculations for methanol dimers to sample the configuration range for which a hydrogen bond is formed. To derive parameters for aromatic carbon (CA) and hydrogen connected to aromatic carbon (HA), ab initio calculations for various arrangements of benzene dimers were used. Good accuracy of fitting was obtained, that is, in all cases, the root-mean-square error of fitting of the nonbonded potential to the ab initio energy sets was within the range 0.01−0.05 kcal/mol. All initial parameters, derived by fitting to ab initio interaction energies, are shown in Table 2.

4.2. Evaluation and Refinement of Potential Parameters. Because we are interested in modeling the crystal state, it is now necessary to assess whether the model with ab initio potential parameters is appropriate for simulation of the condensed phase. This is accomplished by carrying out local and global minimizations of the lattice energy for the test molecules in column 3 of Table 1. First, starting from the observed crystal structures, the energies were minimized locally with respect to all of the degrees of freedom without using any constraints of space-group symmetry. This test provided an evaluation of the ability of the potential to reproduce the observed crystal structures (parameters and space groups) and sublimation enthalpies. To test the derived nonbonded parameters further, global optimization calculations were carried out for all test molecules.

Because our calculations correspond to 0 K, whereas experimental structures are derived at a nonzero temperature, we cannot expect to obtain extremely high accuracy for the structural parameters because of the neglect of thermal motion.[50] On the other hand, use of experimental crystal data obtained at different temperatures for parameter refinement means that thermal expansion effects will be included implicitly to some degree in the model. The following criteria were used to evaluate the accuracy of a potential. A parameter set providing structural deviations below 5% and deviations between lattice energy and the corresponding sublimation enthalpy of 2−3 kcal/mol was considered as sufficiently accurate. The accurate parameter set

also had to predict the observed crystal structure as one of the lowest-energy minima of the lattice energy (preferably the global one).

To evaluate how the parameters obtained in this work perform compared to those from known force fields, global- and local-energy minimizations were also carried out with the W99 force field.[25,26] By contrast to the original model[9] used by Williams, who moved the repulsion center of the H atom closer to the X atom in the X−H bond, no such foreshortening of the repulsion center of hydrogen was introduced. Potentially the omission of this foreshortening could lead to larger structural deviations when comparing our results of local-energy minimizations with those reported by Williams. However, comparison of the results presented by Williams[25,26] for pentane, hexane, heptane, naphthalene, anthracene, and methanol (with such foreshortening) with the results obtained here without foreshortening for the same molecules provided no indication that such foreshortening is necessary for more accurate crystal calculations.

Tables 3−5 contain experimental unit cell parameters (rows "exp"), deviations of the lattice parameters (in percent) obtained as a result of local-energy minimizations with different potentials, and lattice energies of the minimized experimental structures and their rank among all structures found by the global search.

Aliphatic Hydrocarbons. Seven aliphatic hydrocarbons (column 3 of Table 1) were selected for testing the van der Waals parameters for aliphatic carbon C1 and hydrogen HC. The results presented in Table 3 show that, in the majority of cases, the experimental structures are stable upon local-energy minimization. For all of these molecules, the original symmetry and space group of the experimental structure were preserved after local-energy minimization. The average deviation of the unit cell parameters from their experimental values was found to be less than 4%. The largest deviations were obtained for the unit cell parameter *a* of the ethane and propane crystal structures (8.0% and 5.8%, respectively). It was shown[51,52] that thermal motion plays an important role in hydrocarbon crystals. Because ethane and propane are the smallest in the homologous series of aliphatic hydrocarbons tested, it can be expected that thermal vibrations can have a strong effect on their crystal structures. Molecular orientations were found to be very similar for the experimental and minimized experimental structures of all molecules (Table 3, column "mol rot"); however, an exceptionally large molecular rotation (∼15° from the orientation in the experimental structure) was obtained for cyclohexane with both the W99 and ab initio potentials.

Comparison of the lattice energies obtained with ab initio potentials with the corresponding sublimation enthalpies demonstrated that the sublimation enthalpies were underestimated for all crystal structures considered. This result may be caused by neglecting multibody interactions in the ab initio calculations because they were carried out only for dimers. Underestimation of the dispersion energy by the MP2 calculations may be another source of the discrepancies between the sublimation enthalpies and the calculated lattice energies.

The results of the global optimization calculations, starting with the ab initio parameters and carried out for the aliphatic hydrocarbons, show that the "ab initio" parameters perform quite well for this class of molecules. In all cases, the minimized experimental structure was found as one of few lowest-energy minima. For *n*-pentane, *n*-hexane, *n*-heptane, and form II of *n*-butane, it was the global minimum, and for the other molecules, the energy of the minimized experimental structure was less than 0.25 kcal/mol above the global minimum.

TABLE 3: Results of Local- and Global-Energy Minimizations Carried Out for Aliphatic Hydrocarbons with Various Potentials

compound	method	a^a	b^a	c^a	α^b	β^b	γ^b	mol rot^c	E^d	ΔE^e	rank^f
cyclohexane (CYCHEX),^{g} $C2/c$, $Z = 4(\bar{1})$	exp	11.23	6.44	8.20	90.0	108.8	90.0		11.1		
	W99	3.5	−1.2	5.5	0.0	−0.5	0.0	15.7	−9.4	0.11	2
	ab initio	2.3	−0.3	2.8	0.0	−0.9	0.0	15.1	−8.5	0.24	3
propane (JAYDUI),^{g} $P2_1/n$, $Z = 4(1)$	exp	4.15	12.61	6.98	90.0	91.3	90.0		6.8		
	W99	6.3	1.9	−0.1	0.0	0.2	0.0	1.98	−5.9	0.05	3
	ab initio	5.8	1.9	1.0	0.0	−0.5	0.0	1.6	−4.4	0.12	4
ethane (ETHANE01),^{g} $P2_1/n$, $Z = 4(\bar{1})$	exp	4.23	5.62	5.85	90.0	90.4	90.0		4.9		
	W99	6.4	1.1	−2.4	0.0	0.9	0.0	2.5	−4.5	0.00	1
	ab initio	8.0	2.8	−2.9	0.0	−1.6	0.0	3.0	−3.0	0.01	2
n-pentane (PENTANE01),^{g} $Pbcn$, $Z = 4(2)$	exp	4.14	9.03	14.62	90.0	90.0	90.0		10.0		
	W99	4.8	−0.9	2.7	0.0	0.0	0.0	0.0	−9.1	0.00	1
	ab initio	3.9	−1.6	3.2	0.0	0.0	0.0	0.1	−7.3	0.00	1
n-hexane (hexane01),^{g} $P1$, $Z = 1(1)$	exp	4.13	4.69	8.54	83.4	87.3	75.2		12.1		
	W99	3.9	−0.4	1.9	0.5	2.6	1.6	5.6	−10.9	0.00	1
	ab initio	3.1	−0.6	1.6	1.1	1.6	0.8	3.8	−8.9	0.00	1
n-heptane (HEPTANE02),^{g} $P\bar{1}$, $Z = 2(1)$	exp	4.12	4.69	20.35	78.1	81.8	74.3		13.8		
	W99	3.6	−0.6	1.4	0.0	3.5	3.2	2.9	−12.2	0.00	1
	ab initio	2.7	−0.9	1.5	0.6	3.3	1.2	2.2	−10.2	0.00	1
n-butane, form III (DUCKOB),^{g} $P2_1/c$, $Z = 2(\bar{1})$	exp	4.38	7.58	8.35	90.0	119.9	90.0		8.6		
	W99	0.5	−1.6	0.1	0.0	−0.8	0.0	3.2	−8.6	0.20	3
	ab initio	0.0	0.0	0.0	0.0	0.0	0.0	4.1	−6.8	0.21	3
n-butane, form II (DUCKOB02),^{g} $P2_1/c$, $Z = 2(\bar{1})$	exp	5.71	5.17	7.87	90.0	105.9	90.0		-		
	W99	0.3	1.4	5.1	0.0	1.9	0.0	7.2	−8.78	0.00	1
	ab initio	0.7	1.2	4.2	0.0	0.9	0.0	9.2	−6.96	0.00	1

^{a} In line exp, lattice constants are given in Å; in lines W99 and ab initio, percent deviations from the experimental (exp) values are given. ^{b} In line exp, unit cell angles are given in deg; in lines W99 and ab initio, percent deviations from the experimental (exp) values are given. ^{c} Molecular rotations from the experimental positions in deg. ^{d} Lattice energy of the minimized experimental structure in kcal/mol. In line exp, the experimental sublimation enthalpy is given. ^{e} Difference ($E_{m.e.}$ − min$E$) between the energy of the minimized experimental structure and the energy of the lowest minimum found (kcal/mol). ^{f} Rank refers to the minimized experimental structure with respect to the global energy minimum. ^{g} Cambridge Structural Database reference code.[42]

TABLE 4: Results of Local- and Global-Energy Minimizations Carried Out for Aromatic Hydrocarbons with Various Potentials

compound	method	a^a	b^a	c^a	α^b	β^b	γ^b	mol rot^c	E^d	ΔE^e	rank^f
benzene (BENZEN),^{g} $Pbca$, $Z = 4(\bar{1})$	exp	7.46	9.66	7.03	90.0	90.0	90.0		10.8		
	W99	1.1	−1.2	0.9	0.0	0.0	0.0	2.1	−10.1	0.20	2
	ab initio	−1.2	−2.7	0.4	0.0	0.0	0.0	4.4	−7.6	0.00	1
	final^h	−0.7	−1.8	1.9	0.0	0.0	0.0	4.9	−9.5	0.00	1
anthracene (ANTCEN14),^{g} $P2_1/a$, $Z = 2(\bar{1})$	exp	8.55	6.02	11.17	90.0	124.6	90.0		23.3		
	W99	0.1	1.2	0.4	0.0	0.4	0.0	2.3	−21.9	0.00	1
	ab initio	1.6	−2.8	−0.7	0.0	−1.6	0.0	4.4	−17.5	0.00	1
	final^h	1.6	−0.5	−0.4	0.0	1.4	0.0	2.9	−20.7	0.00	1
naphthalene (NAPHTA14),^{g} $P2_1/a$, $Z = 2(\bar{1})$	exp	8.09	5.94	8.65	90.0	124.6	90.0		17.4		
	W99	2.3	2.4	2.2	0.0	0.9	0.0	2.5	−16.1	0.00	1
	ab initio	3.6	−0.8	−0.9	0.0	2.7	0.0	6.8	−12.7	0.00	1
	final^h	3.6	1.2	0.3	0.0	2.1	0.0	4.8	−15.2	0.00	1

^{a} In line exp, lattice constants are given in Å; in lines W99 and ab initio, percent deviations from the experimental (exp) values are given. ^{b} In line exp, unit cell angles are given in deg; in lines W99 and ab initio, percent deviations from the experimental (exp) values are given. ^{c} Molecular rotations from the experimental positions in deg. ^{d} Lattice energy of the minimized experimental structure in kcal/mol. In line exp, the experimental sublimation enthalpy is given. ^{e} Difference ($E_{m.e.}$ − min$E$) between the energy of the minimized experimental structure and the energy of the lowest minimum found (kcal/mol). ^{f} Rank refers to the minimized experimental structure with respect to the global energy minimum. ^{g} Cambridge Structural Database reference code.[42] ^{h} Results of local- and global-energy minimizations carried out with the set of nonbonded parameters obtained as a result of the refinement of ab initio parameters by using the VMC procedure.

In general, the results obtained from local- (columns 3−10 of Table 3) and global-energy minimizations (columns 11 and 12) with the "ab initio" and W99 potentials are very similar; larger structural deviations in unit cell parameters and molecular rotations occur for the same molecules and for the same unit cell parameters with both potentials. However, W99 reproduces the observed sublimation enthalpies noticeably better. Curves for HC···HC and C1···C1 interactions, presented in Figure 2, show that the main difference between the potentials lies in the strength of the HC···HC interactions, which are significantly weaker for the "ab initio" parameters. For the C1···C1 inter-

actions, both the potential well depth and the equilibrium distance are larger for the "ab initio" potential (−0.065 kcal/mol and 4.2 Å compared to −0.042 kcal/mol and 3.8 Å for W99).

The VMC method was applied to improve the potential parameters for aliphatic carbon and hydrogen. The ab initio parameters were used as the starting point. The first five molecules in Table 3 were used for the refinement, and the last two, n-heptane and n-butane, were used to check for the transferability of the resulting parameters. No visible improvement of the potential parameters was achieved.

Derivation of a New Force Field

J. Phys. Chem. B, Vol. 107, No. 29, 2003 **7151**

TABLE 5: Results of Local- and Global-Energy Minimizations Carried Out for Alcohols with Various Potentials

Compound				a^a	b^a	c^a	α^b	β^b	γ^b	Mol.rot.^{c}	E^d	ΔE^e	Rank^f
H₃COH	Methanol: METHOL02^g $P2_12_12_1$, $Z=4(1)$		exp.	4.87	4.64	8.87	90.0	90.0	90.0		-		
			W99	4.0	-5.2	1.3	0.0	0.0	0.0	19.9	-12.9	0.00	1
			final0^h	2.1	5.4	2.2	0.0	0.0	0.0	18.9	-9.9	0.00	1
			final^h	2.5	6.7	0.8	0.0	0.0	0.0	17.9	-8.7	0.00	1
HO⌒OH	Glycol: NOZKES^g $P2_12_12_1$, $Z=4(1)$		exp.	5.01	6.92	9.27	90.0	90.0	90.0		-		
			W99	0.3	0.6	-2.3	0.0	0.0	0.0	4.2	-18.9	0.78	2
			final0^h	0.4	2.6	1.8	0.0	0.0	0.0	4.5	-17.8	0.00	1
			final^h	-0.9	2.5	1.9	0.0	0.0	0.0	4.7	-16.0	0.00	1
(1,2-dihydroxybenzene structure)	1,2-dihydroxybenzene CATCOL12^g $P2_1/c$, $Z=4(1)$		exp.	10.08	5.52	10.94	90.0	118.5	90.0		19.3		
			W99	1.0	-2.4	-2.3	0.0	-3.2	0.0	6.4	-20.3	0.00	1
			final0^h	1.9	-2.2	0.8	0.0	-1.8	0.0	3.2	-20.3	0.00	1
			final^h	1.1	-1.8	0.3	0.0	-1.6	0.0	2.4	-18.9	0.00	1
(L-chiro-Inositol structure)	L-chiro-Inositol: FOPKOK^g $P2$, $Z=2(1)$		exp.	6.87	9.13	6.22	90.0	106.4	90.0		-		
			W99	2.0	-1.9	-0.5	0.0	0.5	0.0	1.4	-45.8	0.00	1
			final0^h	-0.1	0.9	0.6	0.0	0.1	0.0	1.3	-48.9	0.00	1
			final^h	-1.7	1.2	-1.4	0.0	-0.9	0.0	1.5	-44.7	0.00	1
	1,2,3/4,5,6-cyclohexanehexol: YEPNOW^g $P\bar{1}$, $Z=1(\bar{1})$		exp.	9.60	6.50	6.51	70.6	69.6	73.7		-		
			W99	1.3	-4.0	-1.8	0.0	-0.6	5.4	4.1	-54.9	3.45	2
			final0^h	1.9	-1.1	-0.5	0.6	0.7	4.1	3.5	-56.6	0.00	1
			final^h	1.7	-1.2	-0.8	0.1	-0.9	5.4	1.9	-51.2	0.00	1
	Muco-inositol: MUINOS^g $P2_1/c$, $Z=4(1)$		exp.	6.73	9.49	13.09	90.0	119.5	90.0		-		
			final0^h	2.4	0.8	-0.5	0.0	-0.1	0.0	1.8	-51.8	0.00	1
			final^h	0.9	0.3	-0.5	0.0	0.1	0.0	1.3	-46.6	0.00	1
HO⌒OH	m-Dihydroxybenzene: RESORA13^g $Pna2_1$, $Z=4(1)$		exp.	10.53	9.53	5.60	90.0	90.0	90.0		22.3		
			final0^h	-0.8	1.4	1.9	0.0	0.0	0.0	1.5	-21.6	0.00	1
			final^h	-0.5	1.0	2.1	0.0	0.0	0.0	1.9	-19.6	0.00	1

^{a} In line exp, lattice constants are given in Å; in lines W99 and ab initio, percent deviations from the experimental (exp) values are given. ^{b} In line exp, unit cell angles are given in deg; in lines W99 and ab initio, percent deviations from the experimental (exp) values are given. ^{c} Molecular rotations from the experimental positions in deg. ^{d} Lattice energy of the minimized experimental structure in kcal/mol. In line exp, the experimental sublimation enthalpy is given. ^{e} Difference ($E_{m.e.} -$ min$E$) between the energy of the minimized experimental structure and the energy of the lowest minimum found (kcal/mol). ^{f} Rank refers to the minimized experimental structure with respect to the global energy minimum. ^{g} Cambridge Structural Database reference code.[42] ^{h} Results of local- and global-energy minimizations carried out with the set of nonbonded parameters obtained by force minimization[32] for the models with zero (*final*0) and nonzero (*final*) attraction for alcohol hydrogen.
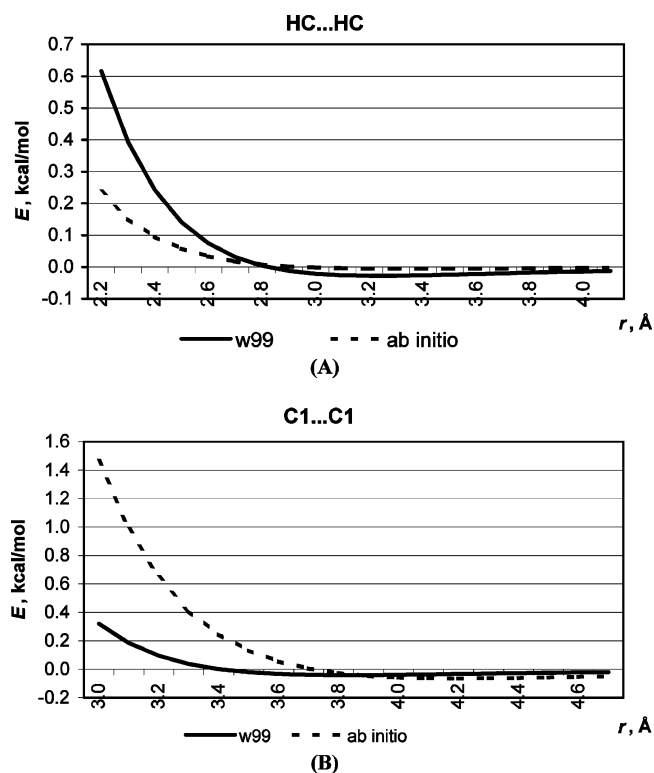

Figure 2. Dispersion−repulsion potential curves for (A) HC···HC and (B) C1···C1 interactions in aliphatic hydrocarbons with the W99 and ab initio (same as final) potentials.

The results presented in this paper are very similar to those obtained in our previous work[32] in which the same group of saturated hydrocarbons was used for testing the VMC procedure. The set of nonbonded parameters obtained from the ab initio calculations (Table 2) provides comparable accuracy for the crystal calculations as the set found by force minimization.[32] The results of global-energy minimizations presented here support the conclusion[32] that the interatomic interactions in crystals of aliphatic hydrocarbons are rather weak, which leads to a "flat" potential energy surface with very broad and shallow minima. The results obtained in this work are also in agreement with those presented by other authors.[51,52] Taking into account the role of thermal effects in crystal structures of aliphatic hydrocarbons,[51,52] it appears that probably no further improvement of the results can be achieved by using a simple static model. The "ab initio" potential parameters were accepted as final for aliphatic carbon and hydrogen and used later for other molecules containing the same atom types.

Aromatic Hydrocarbons. The crystal structures of benzene, naphthalene, and anthracene were selected for evaluation and refinement of the ab initio van der Waals parameters describing the interactions between aromatic carbon (CA) and aromatic hydrogen (HA). The results of global- (columns 11 and 12) and local-energy minimizations (columns 3−10) for these molecules are given in Table 4. The structural deviations obtained as the result of local-energy minimizations of the experimental structures were less than 4% for all molecules. The largest deviation was obtained for the unit cell parameter *a* of naphthalene.

Generally, the results obtained with the "ab initio" and W99 potentials were very similar, except for the lattice energies, which were significantly lower for the "ab initio" potential than the observed sublimation enthalpies. For example, the lattice energy of the anthracene crystal structure obtained with the "ab initio" parameters was 5.8 kcal/mol smaller than the experimental value of the sublimation enthalpy. On the other hand, the result of the global search starting with the "ab initio" potential showed some improvement in the potential parameters compared to the W99 force field. The experimental structure
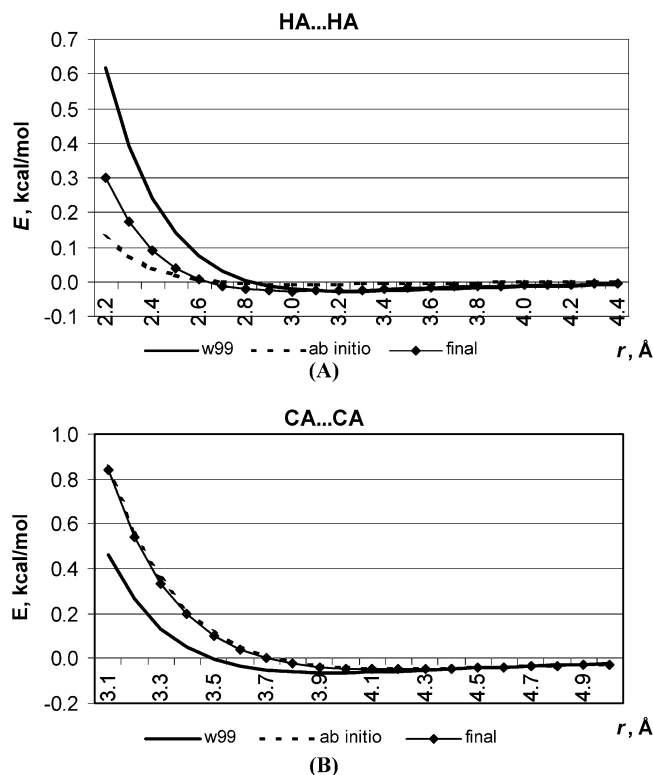
Figure 3. Dispersion−repulsion potential curves for (A) HA···HA and (B) CA···CA interactions in aromatic hydrocarbons with the W99, ab initio, and final potentials.

of benzene was located as a global minimum for the "ab initio" and final potentials, whereas it was only minimum number 2 for W99.

Although, on average, the accuracy of the "ab initio" parameters is relatively high, the deviations of the lattice energies from the corresponding sublimation enthalpies were unacceptably large. To improve the accuracy of the "ab initio" potential, the VMC refinement was carried out using the "ab initio" values as a starting point. As a result, the set of parameters called "final" was obtained (Table 2). The major change in the parameters was for the HA···HA interactions, for which the well depth increased from −0.008 to −0.027 kcal/ mol, the value close to that of W99 potential (−0.028 kcal/ mol) (Figure 3). The equilibrium distances for the HA···HA interaction (3.0, 3.2, and 3.2 Å) were similar for all three sets of parameters (W99, "ab initio", and "final"). For CA···CA interactions, they were slightly longer (4.2 Å for "ab initio" and final, and 3.9 Å for W99).

Local- and global-energy minimizations, carried out for the final set, demonstrated that the improved parameters reproduce both the observed crystal structures and lattice energies of benzene, naphthalene, and anthracene accurately. For all three molecules, the experimental structures were found as the global minima of the potential.

Alcohols. Two other atom types considered here were oxygen (O1) and hydrogen (HO) in the alcohol group. Seven molecules with different numbers of alcohol groups were selected for evaluation and refinement of parameters (Table 5).

Analysis of the nonbonded parameters obtained from the ab initio calculations showed that they provide an unphysical description of the interactions between alcohol hydrogens, that is, only an attraction between two (HO) hydrogen atoms. To find nonbonded parameters for alcohol oxygen and hydrogen that would be adequate for crystal calculations, force minimiza-

tion[32] was carried out for the set of molecules in Table 5, consisting of glycol and catechol. Two different models were considered. It was shown that the best fit to the ab initio interaction energies of methanol dimers[22] or to the experimental crystal structures[26] can be obtained by a model involving only repulsion on the (HO) hydrogen-bonded hydrogens. Therefore, in the first model, the parameter A_{HO} of eq 2 was set to zero, and the only nonbonded variables were the parameters A_O, B_O, and C_O for the alcohol oxygen and the parameters B_{HO} and C_{HO} for the alcohol hydrogen. The second model included all six parameters as variables of the nonbonded potential for the alcohol group (A_{OH}, B_{OH}, C_{OH}, A_O, B_O, C_O). The best set of parameters found by force minimization for each model is presented in Table 2 (rows *final*0 and *final*, respectively). In the second model, in which the A_{HO} parameter was allowed to vary, its value was close to zero.

To assess the accuracy of the models, local-energy minimizations were carried out for the molecules in Table 5 using the *final*0 and *final* sets of potential parameters. The final geometries and lattice energies were compared with the experimental values (Table 5). Both the *final*0 and *final* models appeared to perform fairly well for all molecules except methanol (Table 5). The initial symmetry of the experimental structures was always preserved after energy minimization. The deviations of the unit cell parameters from their experimental values for all molecules, except the unit cell parameter b for methanol and the unit cell angle γ for YEPNOW, did not exceed 3%. For the *final*0 and *final* sets, the largest deviation in the unit cell parameters occurred for parameter b of methanol. The molecular rotations resulting from local optimization of the methanol experimental structure were also found to be larger than those for the other alcohol molecules. This may indicate that the parameters need some further improvement. However, very similar results were obtained for the α-form of methanol with other potentials, such as W99, Dreiding,[14] and OPLS.[18] On the other hand, the ab initio potential model used by Mooij et al.[22] reproduced the observed crystal structure of α-methanol with significantly higher accuracy. Comparison of the experimental and minimized experimental structures (panels A and B in Figure 1) shows that the main difference between them lies in the configuration of the hydrogen-bonded chains. The models used in this work, as well as in the force fields mentioned above,[14,18,25,26] could not reproduce the observed puckering of the hydrogen-bonded chains. It is seen from Figure 1 that the chain configuration obtained as a result of local-energy minimization is almost flat. Taking into account the fact that the ab initio model used by Mooij et al.[22] included an accurate description of electrostatic interactions, polarization, and anisotropy in the oxygen repulsion, the results obtained for the simple potentials are likely to be caused by absence of a description of the directionality of hydrogen bonds in all these force fields. It should be noted that the methanol crystal structure is probably the most sensitive to the fine details of a potential model because the methanol molecule does not contain any bulky functional groups, which could influence the crystal packing. This suggestion is also supported by the small structural deviations found for the other alcohol molecules. Comparison of hydrogen-bond configurations in the crystals of alcohols other than methanol showed that they are reproduced very well by the potentials developed in this work.

The resulting sets of nonbonded parameters, together with the parameters of the W99 force field, were validated by carrying out global-energy optimization (columns 12 and 13 of Table 5). The crystal structures of MUINOS and RESORA13
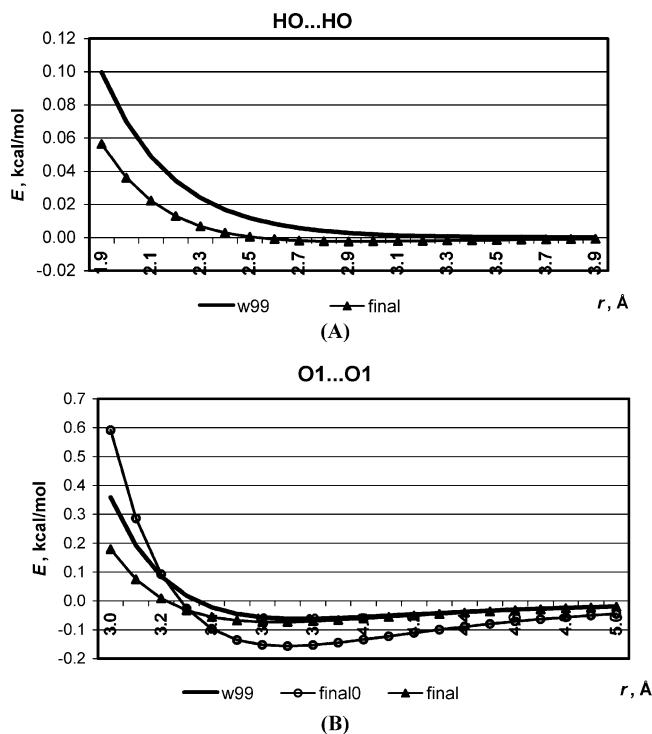
Derivation of a New Force Field

J. Phys. Chem. B, Vol. 107, No. 29, 2003 **7153**

Figure 4. Dispersion−repulsion potential curves for (A) HO⋯HO and (B) O1⋯O1 interactions obtained with the W99, *final*0, and *final* potentials.

were used for additional tests of the transferability of the parameters. For these molecules, global searches were carried out only with the *final*0 and *final* potential parameters. For the *final*0 and *final* sets, all of the minimized experimental structures were found as the global minima of the potential in contrast with the results for W99, which predicted the experimental structures of glycol and YEPNOW as minima number 2. One possible reason for this result is that our potential was designed especially for alcohols, whereas the parameters of the W99 force field were obtained by fitting to experimental data for a large set of H,C,O-containing molecules including quite different molecules [for example, the oxygen was considered to be the same in alcohol groups (−OH) and in dioxan (−OC)].

The potential curves describing the HO⋯HO and O1⋯O1 interactions for several sets of van der Waals parameters are shown in Figure 4. The repulsion for the hydrogen−hydrogen interactions is several orders of magnitude stronger for *final*0 compared to the W99 and the *final* potentials (Table 2); therefore, the *final*0 potential curve is not shown in Figure 4A. The O1⋯O1 potential for *final* is similar to that of W99 except for the slightly weaker repulsion at short interatomic distances. The O1⋯O1 curve for *final*0 shows that oxygen−oxygen interactions are significantly stronger for this set of parameters (the well depth is approximately 2 times larger than that for W99 and *final*). It is also stronger than the repulsion characteristic of the interactions between aliphatic or aromatic hydrogens. This behavior is in contrast to the results of Williams[26] who concluded that reduction of the normally expected exchange repulsion is essential for an accurate description of hydrogen bonding. As seen from Figure 4, a shallow minimum (−0.002 kcal/mol at 2.9 Å) occurs for the *final* parameters of the HO⋯HO potential.

The results of the global- (columns 12 and 13 of Table 5) and local-energy (columns 3−11) minimizations indicate that the *final*0 and *final* sets of nonbonded parameters satisfy all of the requirements discussed in the Introduction and are accurate

enough to predict the crystal structures of alcohols correctly. No refinement of the parameters by the VMC procedure was considered to be necessary.

The final values of the nonbonded parameters for the atom types considered in this work are presented in the Table 2. In general, all components of a potential should be used in a consistent way; thus, the use of the resulting dispersion−repulsion parameters, together with a completely different description of electrostatic interactions or mixing them with potential parameters taken from a different force field, may influence the accuracy of the crystal calculations. When a more elaborate electrostatic model is used, the nonbonded parameters may require an additional adjustment to provide sufficient accuracy of the potential model. In this case, the VMC method represents a very useful tool because it can be applied to the refinement of nonbonded parameters in conjunction with any electrostatic and polarization models.

5. Conclusion

In this work, a new procedure was proposed for deriving nonbonded potential parameters. It makes use of both ab initio and classical crystal calculations to obtain the best possible set of parameters for a given form of the potential. The procedure includes two main steps. First, an initial set of potential parameters is derived by fitting to ab initio interaction energies of dimers. Second, the initial set is refined to satisfy the following criteria: the parameters should reproduce the observed crystal structures and sublimation enthalpies accurately; the observed crystal structure should correspond to the global-energy minimum or one of the lowest-energy minima of the potential. The results show that the ab initio calculations carried out in this work provide a set of parameters which, in many cases, already satisfies both criteria or can be used as a good starting point for the refinement. The VMC procedure enables one to refine potential parameters and to evaluate limitations of the chosen form of the potential.

This work is considered to be the first step in the development of nonbonded parameters for a general all-atom force field. Here, we presented parameters derived for several common atom types, such as aliphatic and aromatic carbon, aliphatic and aromatic hydrogen, alcohol hydrogen, and alcohol oxygen. The resulting parameters were tested on a set of small organic molecules. For aromatic hydrocarbons and alcohols, the new potential reproduces the experimental data with high accuracy and enabled us to predict the crystal structures as global minima for all of the molecules tested. The potential developed for aliphatic hydrocarbons describes the observed crystal structures with sufficient accuracy. The observed crystal structures were also predicted as the global or one of the lowest-energy minima. However, for ethane and butane, the observed relative stabilities of different crystal polymorphs were not reproduced, which is consistent with results obtained by van de Streek et al.[51,52] and indicates that static calculations with a simple "6-exp-1" potential are not completely adequate for this type of molecule.

References and Notes

(1) Fifth Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP5), Asilomar Conference Center, Pacific Grove, Asilomar, CA, Dec 1−5, 2002.

(2) 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.

(3) 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.

(4) Scheraga, H. A. *Adv. Phys. Org. Chem.* **1968**, *6*, 103.

(5) Warshel, A.; Lifson, S. *J. Chem. Phys.* **1970**, *53*, 582.

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

(7) Momany, F. A.; Carruthers, L. M.; McGuire, R. F.; Scheraga, H. A. *J. Phys. Chem.* **1974**, *78*, 1595.

(8) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361.

(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) Williams, D. E.; Houpt, D. J. *Acta Crystallogr.* B **1986**, *42*, 286.

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

(15) Filippini, G.; Gavezzotti, A. *Acta Crystallogr. B* **1993**, *49*, 868.

(16) Hwang, M. J.; Stockfisch, T. P.; Hagler, A. T. *J. Am. Chem. Soc.* **1994**, *116*, 2515.

(17) Sun, H.; Mumby, S. J.; Maple, J. R.; Hagler, A. T. *J. Am. Chem. Soc.* **1994**, *116*, 2978.

(18) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(19) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.

(20) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.;, Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., Jr.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem.* B **1998**, *102*, 3586.

(21) Ewig, C. S.; Thacher, T. S.; Hagler, A. T. *J. Phys. Chem.* B **1999**, *103*, 6998.

(22) 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.

(23) Tsui, H. H. Y.; Price, S. L. *CrystEngComm* **1999**, *1*, 24.

(24) (a) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269. (b) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 6*; University of California: San Francisco, CA, 1999.

(25) Williams, D. E. *J. Mol. Struct.* **1999**, *485−486*, 321.

(26) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1.

(27) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1154.

(28) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205.

(29) Mooij, W. T. M.; van Eijck, B. P.; Kroon, J. *J. Phys. Chem*. A **1999**, *103*, 9883.

(30) Pillardy, J.; Wawak, R. J.; Arnautova, Y. A.; Czaplewski, C.; Scheraga, H. A. *J. Am. Chem. Soc.* **2000**, *122*, 907.

(31) van Eijck, B. P. *J. Comput. Chem.* **2002**, *23*, 456.

(32) Arnautova, Y. A.; Pillardy, J.; Czaplewski, C.; Scheraga, H. A. *J. Phys. Chem. B* **2003**, *107*, 712.

(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

(34) Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, 864. Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133.

(35) Scheiner, A. C.; Baker, J.; Andzelm, J. W. *J. Comput. Chem.* **1997**, *18*, 775.

(36) Chałasiński, G.; Szczęśniak M. M. *Chem. Rev.* **1994**, *94*, 1723.

(37) Gutowski, M.; van Duijneveldt, F. B.; Chałasinski, G.; Piela, L. *Chem. Phys. Lett.* **1986**, *129*, 325. Gutowski, M.; van Duijneveldt, F. B.; Chałasinski, G.; Piela, L. *Mol. Phys.* **1987**, *61*, 233. Gutowski, M.; Chałasinski, G. *J. Chem. Phys.* **1993**, *98*, 5540.

(38) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(39) Gay, D. M. *ACM Trans. Math. Software* **1983**, *9*, 503.

(40) Pillardy, J.; Arnautova, Y. A.; Czaplewski, C.; Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12351.

(41) Dzyabchenko, A. V. *Acta Crystallogr.* B **1994**, *50*, 414.

(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, U.K., 1999; p 507.

(44) *International Tables for Crystallography*; Hahn, T., Ed.; Kluwer Academic Publishers: Dordrecht/Boston/Lancaster/Tokyo, 1987; Vol. A, p 25.

(45) Tauer, K. J.; Lipscomb, W. N. *Acta Crystallogr.* **1952**, *5*, 606.

(46) Torrie, B. H.; Weng, S.-X. *Mol. Phys*. **1989**, *67*, 575.

(47) Allan, D. R.; Clark, S. J.; Brugmans, M. J. P.; Ackland, G. J.; Vos, W. L. *Phys. Rev. B* **1998**, *58*, R11809.

(48) Bondi, A. *J. Chem. Eng. Data* **1963**, *8*, 371.

(49) Gavezzotti, A., Ed. *Theoretical Aspects and Computer Modeling of the Molecular Solid State*; John Wiley & Sons: Chichester, U.K., 1997; p 97.

(50) Pertsin, A. J.; Kitaigorodsky, A. I. *The Atom-Atom Potential Method. Applications to Organic Molecular Solids*; Springer-Verlag: Berlin, 1987; Vol. 43.

(51) van de Streek, J.; Verwer, P.; Bennema, P.; Vlieg, E. *J. Comput. Chem.* **2002**, *23*, 365.

(52) van de Streek, J.; Verwer, P.; Bennema, P.; Vlieg, E. *Acta Crystallogr. B* **2002**, *58*, 677.