Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from ab Initio Quantum Chemistry

George A. Kaminski,† Harry A. Stern, B. J. Berne, and Richard A. Friesner*

Department of Chemistry, and Center for Biomolecular Simulation, Columbia University, New York, New York 10027

Received: January 27, 2003; In Final Form: May 30, 2003

We have produced a polarizable force field for a series of small molecules, representative of functional groups found in organic and biochemical systems. We have used high-level ab initio results for fitting values of all the parameters except for the dispersion-term coefficient B in the $-B/r^6$ energy term, which, although obtained from comparison with experimental condensed-phase data, depended only on atomic number of the site in hand. Heats of vaporization and densities of the pure liquids, computed with molecular dynamics, agreed with experiment within ca. 0.5 kcal/mol and 5%, respectively.

I. Introduction

A great deal of progress has been made over the past several decades in the development of molecular mechanics force fields for use in condensed-phase simulations. When parameters are extensively fit to condensed-phase experimental data, quantitatively accurate models can be produced. However, for the overwhelming majority of important biological and materials science applications, only limited experimental data is available. For such applications, present force fields often provide a qualitatively correct picture of the details of atomic motions and energetic interactions, but the level of quantitative accuracy is uncertain and in many cases may not be adequate for predictive, as opposed to retrospective, studies. For example, in a structure-based drug-design project, one would like to predict protein−ligand binding affinities to better than 0.5 kcal/mol, yet it is far from clear that current force fields are capable of achieving this precision for an individual hydrogen-bonding interaction.

In this paper, we demonstrate for the first time that it is possible to design force-field parameters, which achieve high accuracy in the condensed phase without any explicit fitting to experimental data for the specific system under consideration. The key idea is the development of a systematic protocol for explicitly fitting a polarizable molecular mechanics model to high-quality ab initio quantum chemical data for the electrostatic charge distribution, polarization response, and intermolecular interactions. Universal empirical parameters developed for a small training set of molecules, dependent only upon atomic number, are used to represent the long-range dispersive ($1/r^6$) component of the atom−atom pair potential which is difficult to obtain accurately from quantum chemical calculations. We achieve our target accuracy of ∼0.5 kcal/mol for the liquid-state heat of vaporization and no more than ∼3−5% error in the density for a set of small molecule test cases, results that are comparable in quality to what is obtained by adjusting parameters to directly fit the experimental results. The advantage of our protocol is that it can be immediately applied to new molecules for which experimental data is sparse or nonexistent.

The rest of this article is organized as follows. Section II contains a description of the methodology employed. Results are presented in Section III. Conclusions are given in Section IV.

II. Method

A. General Protocol. The explicit incorporation of electronic polarization effects into molecular-modeling calculations has been the subject of intensive effort over the past decade.[1,2] While a great deal of success has been achieved using fixed-charge force fields, there are two important fundamental limitations of such an approach which have not to date been overcome:

(1) In a homogeneous environment, such as a pure liquid, the assignment of fixed charges based on the polarization in the average liquid-state environment has been quite successful. However, if the goal is to develop a parametrization that accurately represents intermolecular interactions in a wide range of environments—gas phase or liquid state, hydrophobic or hydrophilic—a single set of fixed-charge parameters cannot accomplish this task.[3] In practice, this is highly relevant to modeling complex biological systems such as proteins and nucleic acids and their interactions with small molecule ligands, in which many different types of local environments are present.

(2) Over the past several years it has become possible on a routine basis to accurately compute intermolecular hydrogen-bonding interactions from quantum chemical calculations.[4] However, parameters for a fixed-charge force field cannot be fitted to reproduce the results of these calculations because the effects of "average" polarization must be added in empirically. While it is possible to estimate such effects (e.g., by scaling the quantum chemical data), this procedure is subject to considerable uncertainty from a quantitative point of view. In contrast, a polarizable model is supposed to reproduce the gas-phase intermolecular interaction energies, and so direct fitting to accurate quantum chemical data is unambiguously the correct protocol.

It is well known that linear response provides an excellent approximation for the many-body component of the electronic energy; from this, the functional form of the polarizable potential-energy function is straightforward to derive.[1,2] Hence, the major difficulty in developing a polarizable force field applicable to a wide range of organic compounds has been the determination of an appropriate set of parameters. Construction

* To whom correspondence should be addressed. E-mail: rich@chem.columbia.edu.

† Current address: Department of Chemistry, Central Michigan, University, Mt. Pleasant, MI 48859.

10.1021/jp0301103 CCC: $27.50
Published on Web 12/30/2003

of a force field with broad coverage of chemical space is a highly labor-intensive activity; furthermore, many additional parameters are needed in a polarizable model, particularly if one is aspiring to higher quantitative accuracy and robustness than has heretofore been possible in fixed-charge models (due to the functional form limitations). For this reason, while there have been a large number of polarizable models for water (many of which display quite reasonable performance in both the gas phase and the condensed phase),[1a,f,i] there has been much less success in accurate reproduction of properties of bigger and more diverse molecular systems.[1e,2]

In this paper, we describe a protocol for constructing a polarizable force field for an arbitrary organic compound that involves no fitting to experimental condensed phase data to treat a new molecule yet yields excellent accuracy for thermodynamic properties when condensed phase simulations are carried out. In broad outline, this protocol is based on the following key components:

(1) All electrostatic parameters, describing both the fixed-charge distribution and polarization response of the target molecule, are obtained from ab initio quantum chemical calculations.

(2) The electrostatic model is supplemented with an atom−atom pair potential. We have found that the traditional Lennard-Jones 6−12 potential is inadequate to properly describe both the energetics and structure of hydrogen-bonded interactions; however, a combination of a Lennard-Jones and exponential form provides enough functional flexibility to properly represent the nonelectrostatic component of the pair interaction. All terms in this pair potential other than the B parameter associated with the long-range dispersive interaction $-B/r^6$ in the Lennard-Jones potential are obtained by fitting to quantum chemical binding-energy calculations for molecular pairs (in this paper, homodimers of the target molecule).

(3) We have hypothesized that the atomic B parameter described above in (2) should depend only upon atomic number and vary relatively little with chemical environment. To test this hypothesis, which enormously simplifies the construction of parameters for a general organic molecule, we have produced B parameters for C, O, N, H, and S atomic sites for a small set of model compounds by fitting to experimental condensed-phase data and then applied these parameters, without adjustment, to a small test set of additional molecules. The force field for the test set is constructed without any input of experimental data and using a well defined and uniform protocol; its performance in reproducing gas-phase and liquid-state properties then represents an unbiased measure of the quality of results to be expected from this modeling methodology when applied to a new molecule, for which condensed-phase data would likely not be available.

The protocol described above can readily be applied to develop parameters for amino acid side chains or pharmaceutical compounds in the 50−100-atom range without unacceptable human or computational effort and is based on an energy function that is computationally efficient enough to allow simulations of large condensed-phase systems containing thousands of atoms to be carried out. The question is then what accuracy is achievable within the limitations of this functional form, given that no explicit fitting to experimental data is to be permitted. To evaluate the accuracy of the force field, we rely on liquid-state simulations in which the heat of vaporization and density of the liquid are compared with experimental data. Our target accuracy is drawn from the work of Jorgensen and co-workers,[5a] whose OPLS-AA force-field models typically achieve agreement (via direct optimization of charges and van der Waals parameters) of the heat of vaporization to within ~0.5 kcal/mol and density of 2−3%. If accuracy at this level can be obtained routinely for an arbitrary organic molecule in an arbitrary environment, this will constitute a true next-generation molecular mechanics force field. We note that in this paper we focus primarily upon thermodynamics, the accurate prediction of which would be adequate for many of the biology and materials science problems mentioned above. We expect dynamical quantities such as the diffusion constant, dieletric constant, and rotational tumbling time to be predicted reasonably well (see ref 1j for results for a water model constructed in a similar fashion) but have not made an attempt here to quantify the size of the errors.

B. Electrostatic Interactions. Density-functional theory (DFT) is used to calculate the molecular charge distribution and polarization responses to external probes that provide the data used to fit the electrostatic parameters. The polarization response is modeled via atom-centered dipoles; we have found empirically that the addition of fluctuating charges provides only a marginal improvement in accuracy, as compared to our objectives in the present paper.[6a] We note that it is possible that the use of a model containing exclusively fluctuating charges, but augmented by locations at positions other than the atomic centers, may be able to achieve similar accuracy; investigation of this type of model is reserved for another publication. The charge distribution is modeled using a combination of point charges and permanent dipoles, in general centered on atomic sites, although lone pairs are used for oxygen, and we have found that the use of a charge at the bond midpoint improves the description of the aliphatic C−H moiety. Introducing lone pairs on nitrogen atoms would also be beneficial but less so than for the oxygen sites, as has been pointed out elsewhere.[6b] A brief description of the electrostatic model is given below, followed by a more detailed discussion of some important aspects of its implementation. A more comprehensive description of the model itself has been published recently.[6b]

The total electronic energy of the model is given by

$$U(\{q_{ij}\},\{\mu_i\}) = \sum_i \left(\chi_i\mu_i + \frac{1}{2}\mu_i\alpha_i^{-1}\mu_i\right) + \frac{1}{2}\sum_{ij\neq kl} q_{ij}J_{ij,kl}q_{kl} + \sum_{ij,k} q_{ij}\mathbf{S}_{ij,k}\mu_k + \frac{1}{2}\sum_{i\neq j}\mu_i\mathbf{T}_{i,j}\mu_j \quad (1)$$

Here, $J_{ij,kl}$ is introduced as a scalar coupling between bond−charge increments on sites i,j and k,l, q_{ij} and q_{kl}; $\mathbf{S}_{ij,k}$ is a vector coupling between a bond−charge increment on sites i,j and a dipole on site k, μ_k; a rank-two tensor-coupling $\mathbf{T}_{i,j}$ describes interactions between dipoles on sites i and j. α_i is the polarizability of site i. Parameters χ_i describe the "dipole affinity" of site i.

A natural choice for coupling of bond−charge increments and dipoles that are well separated in space is the Coulomb interaction

$$J_{ij,kl} = \frac{1}{r_{ik}} - \frac{1}{r_{il}} - \frac{1}{r_{jk}} + \frac{1}{r_{jl}} \quad (2)$$

$$\mathbf{S}_{ij,k} = \frac{\mathbf{r}_{ik}}{r_{ik}^3} - \frac{\mathbf{r}_{jk}}{r_{jk}^3} \quad (3)$$

$$\mathbf{T}_{i,j} = \frac{1}{r_{ij}^3}\left(\mathbf{1} - 3\frac{\mathbf{r}_{ik}r_{ik}}{r_{ij}^2}\right) \quad (4)$$

We have discussed, in detail in previous publications,[6,7] how the permanent and polarization parameters are determined, via

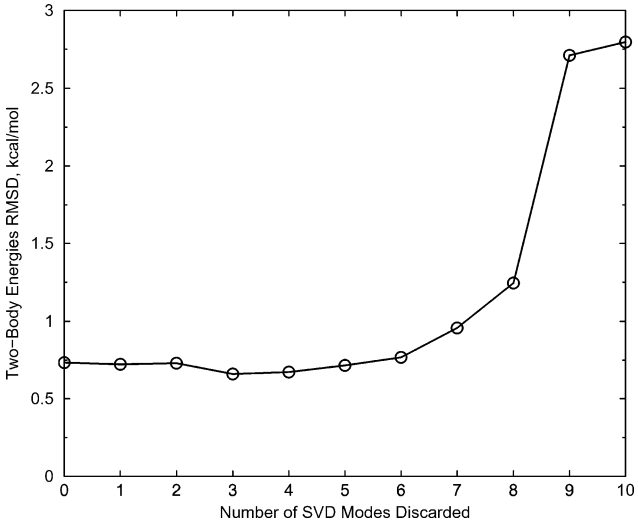

Figure 1. RMS deviations of two-body energies of interaction of a phenol molecule with dipolar probes as a function of the number of singular-value decomposition modes discarded in electrostatic potential surface fitting of fixed charges and dipoles. We chose to cut seven modes in this case.

least-squares fitting, from the quantum chemical data; the most sensitive aspect of this methodology for the present work is the removal of unstable polarization modes by the use of singular value decomposition (SVD). Modes are removed one by one until a noticeable jump in the error in the electrostatic energy of test charges interacting with the molecular charge distribution and polarization is observed. Figure 1 displays a typical dependence of the root mean square (RMS) deviation on the number of the modes discarded.

There are a number of other important physical parameters of the model that are not directly determined from the least-squares fitting to the quantum chemical electrostatic potential and response data. The lone-pair distance for oxygen, 0.47 Å, was determined by optimization of condensed-phase results for one of the molecules in the training set, methanol, and this value was then used without adjustment for all other molecules (except as explicitly noted below). The lone-pair distance for sulfur was set to 0.9 Å to achieve the best condensed-state and gas-phase results. Dipole−dipole and charge−dipole screening is used to damp the polarization response when the perturbing site is at short distances; a screening length, the site−site distance below which electrostatic interactions are damped, is determined for each functional group by reducing RMS errors in three-body energies (see ref 7 for a detailed description of how three-body energies are computed and compared with quantum chemical results) across the molecular database; the values of these parameters are presented in Table 1. Note that optimization of the screening lengths employs only quantum chemical data (no iterative improvement of these parameters based on condensed phase results was permitted), and hence, the development of screening length parameters for new functional groups does not require experimental input.

Finally, a crucial component of the methodology is the precise level of quantum chemical theory used to generate the fitting data. The experience of our group and others is that hybrid DFT methods such as B3LYP (which is what is used here) provide excellent quantitative results for charge distributions and polarization responses as compared to experimental data.[8] However, accurately reproducing gas-phase polarizabilities requires the use of very large basis sets, particularly a substantial number of diffuse functions. This is computationally tractable with modem computers and algorithms, but there is a fundamental question as to whether the use of such functions yields optimal results in the condensed phase. The spatial extent of many of the functions that contribute nontrivially to the polarization response is substantially larger than a molecular diameter so that, in a liquid or solid at normal densities, these functions would have large overlaps with neighboring molecules. One would then expect the Pauli exclusion principle to qualitatively raise the energies of these basis functions, thus reducing their mixing into the ground-state wave function.

TABLE 1: Screening Lengths, below Which Charge−Dipole and Dipole−Dipole Interactions Are Scaled in Order to Avoid Unreasonably Large Interaction Magnitudes for Short Distances (when Point−Dipole and Point−Charge Approximations Are No Longer Accurate), in Å

atom		length for charges	length for dipoles
$-CH_3$ and $-CH_2-$	C	0.8	2.0
$-CH_3$ and $-CH_2-$	H	0.8	
CH_3OH	O	0.4	1.8.
CH_3OH	H(O)	0.4	1.0
CH_3OCH_3	O	0.4	1.8
CH_3COCH_3	C	0.8	2.0
CH_3COCH_3	O	0.4	2.5
amides	N	0.8	2.0
amides	H(N)	0.8	
amides	C(O)	0.8	2.0
amides	O	0.4	2.5
C_6H_6	C	0.8	3.2
C_6H_6	H	0.8	
C_6H_5OH	O	0.4	1.8
C_6H_5OH	H(O)	0.4	1.0
CH_3SH	S	0.8	2.5
CH_3SH	H(S)	0.4	1.0
lone pairs, midpoints	X	0.8	

It is extremely difficult to determine the quantitative magnitude of the Pauli exclusion effects, although in principle one should be able to address this problem via careful cluster calculations. We have instead approached the problem heuristically. We have found consistently that liquid-state simulations carried out with polarization parameters fitted to quantum calculations employing a large number of diffuse functions lead to overpolarization of the liquid, as manifested in excessive dielectric constants, enhanced heats of vaporization, and, in extreme cases, polarization catastrophes. Therefore, we have chosen to employ the cc-pVTZ(-f) basis set of Dunning[9] as a standard basis set for computation of quantum chemically derived electrostatic data. This basis set is dense enough to converge the polarization response and charge distribution at relatively short distances from the atoms of the molecule but does not contain any basis set exponents, which appear to grossly violate Pauli exclusion under typical condensed-phase conditions. The results that follow provide a measure of how well this heuristic approximation works; in the future, a more extensive exploration of this issue is clearly going to be necessary if we are to move to a higher level of accuracy. The exclusion of diffuse functions does, of course, have a significant advantage of lowering the computational cost of the calculations as well. We note however that this implies that the polarization response of the model may not accurately reproduce experiment in the gas phase, small clusters, or at the liquid−vapor interface; these defects are not important for our targeted applications.

C. Nonelectrostatic Interactions. This part of the potential includes all effects not explicitly represented by electrostatics, e.g., exchange, Pauli repulsion, dispersion, and charge transfer; the validity of such a representation depends on whether these terms are indeed accurately modeled as two-body effects. Our

previous work,[6] and that of others,[2] suggests that this is quite a good approximation.

The most important pair interactions for organic molecules are hydrogen-bonding interactions (although we shall see below that several other attractive atom pair interactions play a significant role in some liquids). The pair potential between the donor and acceptor heavy atoms (we do not explicitly include parameters for polar hydrogens, following the OPLS-AA force field) should then be able to reproduce the quantum chemical hydrogen-bonding distance and the binding energy of the hydrogen-bonded molecular pair. Since the *B* parameter of the pair term is fixed at the universal values discussed above, a traditional Lennard-Jones 6−12 form is left with only one parameter (A in A/r^{12}) to fit *both* of these quantities, a task we have in general found impossible to carry out. At least one more parameter is required. We have chosen to employ a composite function in which an exponential term of the form $C\exp(-r/\alpha)$ is added to the Lennard-Jones form, leading to the overall nonelectrostatic pair potential

$$E_{nb} = \sum_{i<j} A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6} + C_{ij}\exp(r_{ij}/\alpha_{ij}) \qquad (5)$$

The use of a $1/r^6$ term to represent dispersive interactions at medium to long range is a standard approximation in molecular mechanics; while other functional forms have been investigated, it is clear that a term of this type has been highly successful in reproducing condensed-phase properties for a wide range of organic liquids and solids, e.g., as evidenced by results obtained with the OPLS-AA force field. Both the $1/r^{12}$ and the exponential terms have been successfully used in the literature to represent the short-range repulsive component (arising due to Pauli exclusion as well as other quantum mechanical effects) of the atom−atom pair potential. If the parameters are to be adjusted to reproduce a small number of thermodynamic condensed-phase properties at room temperature only, the details of exactly how the functional form rises at the repulsive wall are not going to be terribly important. Our employment of a combination of these terms is based on the desire to accomplish three objectives at short distances: (1) avoidance of penetration into nonphysical regions of phase space, this is particularly important for a polarizable force field, as is discussed below; (2) reproduction of accurate quantum mechanical binding energies for molecular dimers; and (3) reproduction of interatomic distances for these same dimers. Simultaneously achieving all three of these objectives is not possible with a functional form such as the Lennard-Jones 6−12 potential if only the *A* coefficient is to be adjusted. The proposed functional form has sufficient flexibility to enable reasonable satisfaction of all objectives while presenting a physically plausible rise in the potential as the repulsive region is entered. Of course, other functional forms with equal to or greater flexibility could be used; this issue will be investigated further in subsequent publications.

The parameters are determined as follows. The *B* parameter depends only upon atomic number as discussed above. The *A* parameter is set so that the $1/r^{12}$ term is close to zero in the hydrogen-bonding region but is large enough to prevent another atom from penetrating the nonphysical region of the phase space. This automatically avoids any sort of polarization catastrophe or other problem of electrostatics overwhelming the repulsive wall while allowing the exponent 6 terms to be adjusted solely with the objective of reproducing the quantum chemical binding energy and hydrogen-bond distance. As there are now two parameters, *C* and α, with which this can be accomplished, obtaining a suitable parametrization is unproblematic in the vast majority of cases. Problems do arise when not considering true hydrogen-bonded pairs for which the potential surface is very flat (as in the case of the methyl hydrogen of one dimethyl ether molecule interacting with the oxygen of a second), and future versions of the model likely will be augmented to treat such cases more accurately; for the present, however, we accept a larger degree of error in the structure of such pairs (which has a rather small effect on the energetics).

Fitting of the *C* and α parameters is accomplished via a grid search in which the model pair potential is geometry optimized at each (C, α) point on the grid, and the best simultaneous fit to the binding energy and hydrogen-bond length is chosen.

We have found that it is essential to model the dimer interactions (in this paper we consider homodimers of the target molecules) with a high level of quantum chemical theory if liquid-state thermodynamic properties are to be accurately reproduced; DFT calculations are insufficient to generate results in the range of our target accuracy. We use the Jaguar ab initio electronic structure package,[10] employing pseudospectral (PS) numerical methods,[11] to efficiently perform the large basis set, wave function based quantum chemical calculations that are required. Geometry optimizations are carried out at the local MP2 (LMP2)/6-31G** level of theory, which is sufficient to adequately determine the hydrogen bonding distance. Then, single-point energies are determined using an extrapolation of LMP2 calculations with increasingly large basis sets at the cc-pVTZ(-f) and cc-pVQZ(-g) levels. Comparison of this protocol with ultralarge CCSD(T) calculations presented in ref 12 demonstrate agreement to within ∼0.1−0.2 kcal/mol for a significant number of test cases; this is within the limits of our target accuracy. At the same time, the PS-LMP2[11] calculations are several orders of magnitude more efficient than alternative conventional MP2 approaches for basis sets of this size, a ratio that increases dramatically if molecules substantially larger than those considered here are to be studied.

Another question that arises in construction of the pair term is how to define the combining rules for the atomic parameters to generate the pair potential. We adopt geometric combinations for parameter C ($C_{ij} = (C_{ii}C_{jj})^{1/2}$). Arithmetic combining rules were employed for α, so that $\alpha_{ij} = (\alpha_{ii} + a_{jj})/2$. Parameters *A* and *B* converted to σ and ϵ first, in accordance with eq 6

$$\epsilon = \frac{B^2}{4A} \qquad \sigma = \left[\frac{A}{B}\right]^{1/6} \qquad (6)$$

Then arithmetic and geometric combining rules are applied to σ and ϵ, respectively, and they converted back to *A* and *B*

$$A = 4\epsilon\sigma^{12} \qquad B = 4\epsilon\sigma^{6} \qquad (7)$$

Torsional energy is represented by a Fourier series for each torsion, with the values of the parameters taken directly from the OPLS-AA force field.[5a] The same approach was used with harmonic bond-stretching and angle-bending terms. For small molecules such as those considered here, this appears to be a reasonable approximation; refitting of torsions for larger, more flexible molecules is clearly required. All the molecules were completely flexible, both in the gas-phase and in the liquid-state simulations.

D. Liquid-State Simulations. Condensed-phase simulations are carried out using the molecular dynamics technique. All runs were performed with $N = 216$ molecules and used a time step of 0.8 fs for a total of 80 ps. Constant pressure and temperature (NPT) simulations with the Andersen−Hoover-type barostat of

TABLE 2: Values of Parameters *A* and *B*, Depending on Atomic Number Only, in kcal/mol Å^{12} and kcal/mol Å^6, Respectively

atom	A	B
C	7500.0	740.0
H (nonpolar)	600.0	20.0
H (polar)	0.0	0.0
O	3500.0	950.0
N	4000.0	900.0
S	10000.0	4100.0

Martyna et al.[14] were used. We used cubic periodic boundary conditions and Ewald summation for the electrostatics.[15] Long-range corrections to the energy and pressure (due to the B/r^6 portion only) were applied.[16] The cutoff was scaled along with the box length in order for the long-range corrections to the energy and pressure to be thermodynamically consistent.[17] The "electronic" degrees of freedom (the fluctuating dipole moments) were propagated using the extended Lagrangian method,[18] that is, assigned masses and integrated along with the spatial coordinates. Using the constant-pressure ensemble allowed us to compute densities with simple averaging, and heats of vaporization ΔH_{vap} were calculated in the standard fashion

$$\Delta H_{vap} = E(\text{gas}) - E(\text{liq}) + RT \quad (8)$$

Here E(liq) and E(gas) are liquid- and gas-state energies, the former produced in the course of the molecular dynamics simulations and the letter obtained via running gas-phase Monte Carlo calculations, 10^5 configurations for equilibration followed by 10^6 configurations of averaging. Liquid-state equilibration was carried out for 40 ps with a 0.8-fs time step. The following example is intended to illustrate the sufficiency of such an equilibration time length for the purpose of this work. Liquid methanol heat of vaporization and density, averaged over the last 8 ps of the equilibration time, were found to be 8.86 kcal/mol and 0.790 g/cm^3. The same parameters averaged over the immediately preceding 8 ps of the equilibration were 8.85 kcal/mol and 0.792 g/cm^3, which clearly suggests that an equilibrated state was achieved.

III. Results and Discussion

After we produced the electrostatic parameters, the next phase of our work was to determine the universal dispersion parameters. Gas-phase dimerization energies were used to obtain C and α values for each system, then molecular dynamics simulations were carried out for pure liquids. Dispersion parameters for carbon and aliphatic hydrogen were obtained first by modeling a series of hydrocarbons. The parameters A and B and thermodynamic results obtained are presented in Tables 2 and 3, respectively. Polarization energy is relatively unimportant for hydrocarbons, so the parameters obtained are similar to those of OPLS-AA; however, there are small differences, which arise from the alternative description of the electrostatics (including the use of bond charges as well as the incorporation of polarization). Overall, the results are comparable in quality to those obtained from OPLS-AA.

We next determine the B parameter for oxygen, as well as the oxygen lone-pair distance, by carrying out a simulation for methanol. Parameters and results are shown in Tables 2 and 3; the fit is quite satisfactory. Note that the B value for oxygen is not very different from that for carbon; in fact, the use of a single universal value for first-row atoms would not be terribly inaccurate. The fact that the oxygen value is slightly larger is consistent with the physically intuitive idea that lone-pair electrons will be more polarizable than those in bond pairs. On the basis of this reasoning, we then proceed to assign the B parameter for nitrogen by interpolating between the carbon and oxygen values. After carrying out gas-phase dimerization and liquid-state calculations for acetamide, the value of the nitrogen B parameter was slightly adjusted (by ca. 5%).

With parameter values assigned for C, O, N, and H, we can now carry out a series of tests in which no adjustable B parameters are employed. We have chosen to initially investigate the molecules acetone, dimethyl ether, formamide, and NMA. Dimers of both acetone and dimethyl ether involve unusual intermolecular interactions, as shown in Figure 2. The C···O interaction in acetone is strongly attractive (roughly half the strength of a hydrogen bond) and leads to the two dimer structures shown in Figure 2. The polarizable model accurately reproduces their structures, absolute energies, and relative energies quite well, in contrast to OPLS-AA, which, while the liquid-state properties of acetone are well described, does not contain a highly precise depiction of the intermolecular interactions. The dimethyl ether dimer is characterized, as discussed above, by a very flat energy surface, which leads to a rather large deviation of the dimer intermolecular equilibrium distance from the quantum chemical value; however, the energetics are reasonably well described. We have obtained a very good level of agreement with quantum chemical data of our gas-phase dimerization results for the amides. Deviation of the PFF energies from their ab initio counterparts are below 0.1 kcal/mol for formamide and acetamide and only about 0.3 kcal/mol for NMA, and the hydrogen bond length is off by only 0.02, 0.07, and 0.02 Å, respectively. All the gas-phase results are presented in Table 4. It is also worth mentioning that, although values of the C and α parameters were adjusted to produce correct gas-phase structures and energetics, an independent test with no further parameter changes has also been carried out. We compared PFF methanol−acetamide dimerization with the high-level ab initio results, produced as described above. For this heterodimer, the quantum mechanical association energy is 8.92 kcal/mol, with two hydrogen bonds, O(MeOH)··· H(acetamide) and H(MeOH)···O(acetamide), 2.055 and 1.961 Å, respectively. The PFF results were 9.19 kcal/mol and 1.979 and 2.057 Å, in a very good agreement with ab initio calculations. The liquid-state thermodynamic results for all the systems are shown in Table 3. As can be seen, the accuracy is close to the level which was originally targeted. The results obtained by OPLS-AA are shown for comparison. These are somewhat more precise, but one should remember that this was achieved by explicit fitting to the condensed-phase data. For the vast majority of organic compounds, such fitting is not possible; in such cases, however, there is no reason to believe that our present protocol would not produce accuracies comparable to those shown in Table 3.

It should be noted that we employed a value of virtual site−oxygen distance of 0.35 Å in the formamide and NMA calculations. This value is different from 0.47 Å used for all the other oxygen-containing molecules, including acetamide. We plan to overcome this problem in the next generation of our polarizable force field and to develop uniformly usable values of this distance for all the atoms in similar chemical groups.

As was mentioned above, the accuracy of the liquid-state results is critically dependent upon the accuracy of the quantum chemical computation of pair distances and binding energies. In the final section of this paper, we consider several cases where the reliability of the quantum chemical pair calculations is less clear than for the molecules discussed above. The first is aromatic systems. Convergence of the stacking energy of the

TABLE 3: Heats of Vaporization in kcal/mol, Molecular Volumes in Å^3, and Densities in g/cm^3 for the Pure Liquids

	ΔH_{vap}			V_{mol}			d		
liquid	PFF	OPLS	exp	PFF	OPLS	exp	PFF	OPLS	exp
CH_4	1.89	2.19[a]	1.96[a]	62.2	57.2[a]	62.8[a]	0.428	0.466[a]	0.424[a]
C_2H_6	3.32	3.44[a]	3.62[a]	94.4	92.5[a]	91.5[a]	0.529	0.540[a]	0.546[a]
C_3H_8	4.79	4.55[a]	4.49[a]	123.9	125.2[a]	126.0[a]	0.591	0.585[a]	0.581[a]
C_4H_{10}	5.62	5.43[a]	5.35[a]	157.2	161.3[a]	160.3[a]	0.614	0.598[a]	0.602[a]
CH_3OH	8.84	8.95[b]	8.95[b]	67.0	68.3[b]	67.7	0.794	0.779[b]	0.786
CH_3OCH_3	5.68	5.15	5.14	105.8	106.5[b]	104.1[b]	0.723	0.717	0.735[b]
CH_3COCH_3	7.92	7.24	7.48	127.5	121.2	123.0	0.756	0.795	0.784
NH_2COH	15.5	14.8[c]	15.5[d]	63.76	66.8[c]	66.3	1.175	1.120[c]	1.129[c]
NH_2COCH_3	13.2	13.7	13.4	112.1	109.3		0.876	0.897	
NMA	13.9	13.6[b]	13.3[b]	128.3	133.9	135.9	0.947	0.907[b]	0.894[b]
C_6H_6	8.20	8.05[e]	8.09[c]	153.7	148.6[e]	148.4[c]	0.844	0.873[e]	0.874[e]
C_6H_5OH	13.8	14.1[b]	13.8[b]	146.8	148.8[b]	147.8	1.065	1.050[b]	1.058[b]
CH_3SH	5.96	6.05	5.87[b]	92.5	89.5	90.0[b]	0.864	0.892[b]	0.888[b]
C_2H_5SH	6.73	6.79[b]	6.58[b]	121.9	120.7[b]	123.8[b]	0.846	0.855[b]	0.833[b]

[a] Reference 19d. [b] Reference 5. [c] Reference 19b. [d] Reference 19a. [e] Reference 19c.

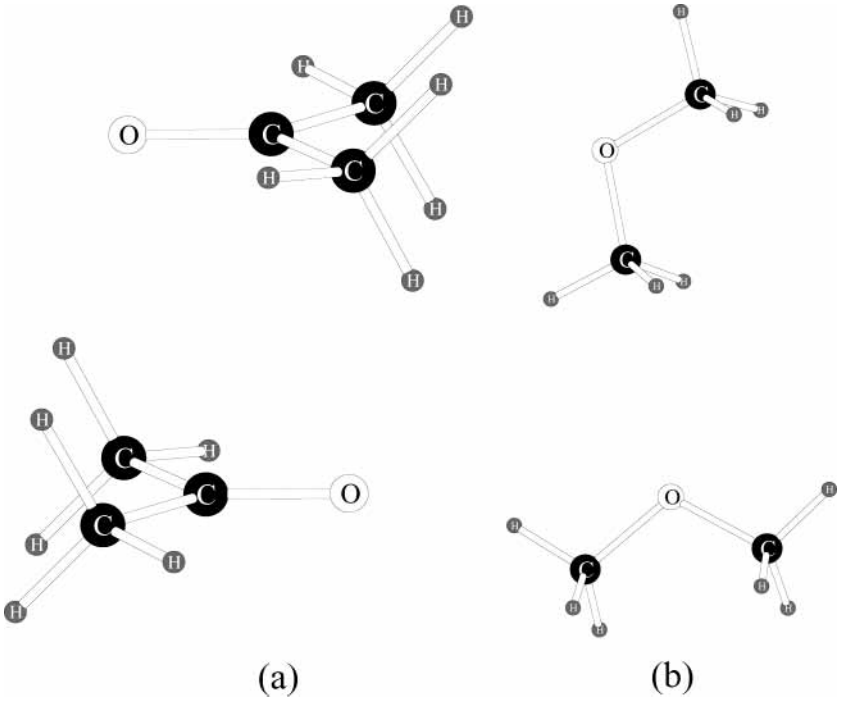

Figure 2. Gas-phase acetone (a) and dimethyl ether (b) dimers, examples of weak bonds.

TABLE 4: Gas-Phase Dimerization Energies in kcal/Mol and Distances between Heavy Atoms in Å

	E_{dimer}			R		
molecule	PFF	OPLS	ab initio	PFF	OPLS	ab initio
CH_4	−0.44	−0.48	−0.50	3.86	3.77	4.06
CH_3OH	−5.63	−6.41	−5.59	2.81	2.78	2.80
CH_3OCH_3	−1.45	−1.69	−1.46	3.46	3.09	3.09
CH_3COCH_3	−5.74	−4.65	−5.76	3.38	3.45	3.28
NH_2COH	−14.0	−14.1[22]	−14.0	2.82	2.81[22]	2.84
NH_2COCH_3	−12.9	−13.9	−12.8	3.01	2.78	2.94
NMA	−14.1	−11.3	−14.4	2.90	2.83	2.92
C_6H_6	−3.14	−2.11	*b*	3.57		*b*
C_6H_5OH	−5.87	−7.99	−5.68	3.10	−2.74	2.98
CH_3SH	−2.77	−2.31	*b*	−3.77	3.74	*b*

[a] Reference 19c. [b] Reliable ab initio data not available at this time. See text for a recent estimate for the benzene−benzene dimerization energy.

benzene dimer with basis set and quantum chemical methodology has not yet rigorously been determined in the published literature; recent work of Tsuzuki et al.[20] suggests substantial differences between CCSD(T) and MP2 results, in contrast to classical hydrogen-bonding energetics where the difference was shown to be negligible. We therefore studied benzene and phenol with the objective of determining the gas-phase binding affinity required to reproduce the liquid-state thermodynamic properties. We use the parameters determined for benzene for the ring carbons in phenol, so the phenol calculation actually has no adjustable parameters, as the phenolic O−H group interaction is modeled by fitting to our usual extrapolated MP2 calculations. As can be seen from Tables 3 and 4, both benzene and phenol PFF models demonstrate excellent agreement with the available experimental liquid-state heats of vaporization and densities. The phenol gas-phase dimerization energy and interatomic distance is also in a very good agreement with the ab initio results, as the dimer we considered was not a stacked one but had an O···H hydrogen bond, while the liquid-state heat of vaporization and density both deviate by less than 1% from the experimental data. Recent unpublished work on the benzene dimer, which combines complete basis-set results for the MP2 component of the energy with large basis-set CCSD(T) calculations, has obtained the result of 2.92 kcal/mol for the stacking energy,[21] which is in remarkably good agreement with our value of 3.14 kcal/mol reported in Table 4.

Finally, we have extended our force field to simulate sulfur-containing compounds, methane and ethane thiols. We have had some difficulty in obtaining converged results for sulfur-containing dimers with our quantum chemical methods due to the fact that sulfur is a second-row atom (and hence the MP2 extrapolation protocol likely needs to be recalibrated) and that the potential surface is very flat, leading to problems in optimizing the geometry with large basis sets at the MP2 level. Therefore, we have optimized parameters empirically in this case to fit the condensed-phase experimental data. Excellent agreement is obtained for both systems with the density and heat of vaporization; moreover, parameters for the ethane thiol were exactly the same as for the methane thiol case, with no further adjustments. We present the gas-phase dimerization energy and geometry obtained from the model as a prediction, which can be compared with accurate quantum chemical results when they appear (as was done in the case of benzene, cited above).

IV. Conclusions

We have succeeded in creating a polarizable force field, which accurately reproduces both gas-phase and condensed-phase results, using ab initio data as a target for the fitting of all the paramreters, except the dispersion interaction factor *B*, which can be made to depend on atomic number only. This result is of great importance, as it opens the road for parametrizing systems of arbitrary size based on quantum miechanical data alone, once the dispersion coefficients are obtained for all the necessary elements from liquid-state simulations of small model compounds.

Acknowledgment. This work was funded by the National Institutes of Health under Grants GM52018 (R.A.F.), GM19961 (G.A.K.), and GM43340 (B.J.B.) and by the National Science Foundation Grant No. CHE-00- 76279 (B.J.B.). This work was partially supported by the National Computational Science Alliance (NCSA) under Grant Number MCA95C007N and utilized the NCSA SGI/CRAY Origin 2000.

Supporting Information Available: Full list of the exponential parameters C and α (eq 5) and all the electrostatic parameters used in this work. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) For example, see: (a) Liu, Y. P.; Kim, K.; Berne, B. J.; Friesner, R. A.; Rick, S. W. *J Chem. Phys.* **1998**, *108*, 4739−4755. (b) Ramon, J. M. H.; Rios, M. A. *Chem. Phys.* **1999**, *250*, 155−69. (c) Gonzalez, M. A.; Enciso, E.; Bermejo, F. J.; Bee, M. *J. Chem. Phys.* **1999**, *110*, 8045−8059. (d) Soetens, J. C.; Jansen, G.; Millot, C. *Mol. Phys.* **1999**, *96*, 1003−1012. (e) Dang, L. X. *J. Chem. Phys.* **2000**, *113*, 266−273. (f) Chen, B.; Xing, J. H.; Siepmann, J. I. *J. Phys. Chem. B* **2000**, *104*, 2391−2401. (g) Ribeiro, M. C. C. *Phys. Rev. B* **2001**, *63*, *4205*. (h) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048−1057. (i) Jedlovszky, P.; Vallauri, R. *J. Chem. Phys.* **2001**, *115*, *3750−3762*. (j) Stem, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, *115*, 2237−2251.

(2) (a) Kolafa, J.; Ratner, M. *Mol. Simul.* **1998**, *21*, 1−26. (b) Kaminski, G. A.; Jorgensen, W. L. *J. Chem. Soc., Perkin Trans. 2* **1999**, *11*, 2365−2375.

(3) For example, see: Tables 3 and 4. While the OPLS-AA works well in condensed state, it does not allow accurate reproduction of gas-phase dimerization.

(4) For example, see: Vargas, R.; Garza, J.; Friesner, R. A.; Stern, H. A.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* **2001**, *105*, 4963−4968.

(5) . (a) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236. (b) For example, see: Williams, D. W. *J. Comput. Chem.* **2001**, *22*, *1*.

(6) (a) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, *103*, 4730−4737. (b) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R. H.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515−1531.

(7) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741−754.

(8) For example, see: Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. X. *J. Phys. Chem. A* **1999**, *103, 1913−1928*.

(9) Dunning, T. H. J. *Chem. Phys.* **1989**, *90*, 1007−1023.

(10) Jaguar v3.5, Schrödinger, Inc.: Portland, OR, 1998.

(11) Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. *J. Chem. Phys.* **1995**, *103*, 1481−1490.

(12) Tsuzuki, S.; Uchimaru, T.; Matsumara, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11906−11910.

(13) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(14) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177−4189.

(15) Smith, W. *CCP5 Newsletter* **1998**, *46*, 18−46.

(16) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Claredon Press: Oxford, 1987.

(17) Martyna, G. J.; Hughes, A.; Tuckerman, M. E. *J Chem. Phys.* **1999**, *110*, 3275−3290.

(18) (a) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474. (b) Van Belle, D.; Froeyen, M.; Lippens, G.; Wodak, S. J. *Mol. Phys.* **1992**, *77*, 239−255. (c) Spirk, M. *J. Phys. Chem.* **1991**, *V3*, 2283−2291. (d) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, 6141−6156.

(19) (a) Bauder, A.; Gunthard, H. H. *Helv. Chim. Acta* **1958**, *41*, 670−673. (b) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 569−578. (c) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768−4774. (d) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L.; *J. Phys. Chem.* **1994**, *98*, 13077−13082.

(20) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K.; *Chem. Phys. Lett.* **2000**, *319*, *547−554*.

(21) David Sherrill of Georgia Institute of Technology, private communications.

(22) Gao, J.; Pavelites, J. J.; Habibollazadeh, D. *J. Phys. Chem.* **1996**, *100*, 2689−2697.