Development of a Polarizable Force Field For Proteins via *Ab Initio* Quantum Chemistry: First Generation Model and Gas Phase Tests

GEORGE A. KAMINSKI,[1] HARRY A. STERN,[1] B. J. BERNE,[1] RICHARD A. FRIESNER,[1] YIXIANG X. CAO,[2] ROBERT B. MURPHY,[2] RUHONG ZHOU,[2] THOMAS A. HALGREN[2]
[1]*Department of Chemistry, Columbia University, New York, New York 10027* [2]*Schrödinger, Inc., 120 West 45th Street, Tower 45, 8th Floor, New York, New York 10036*

Received 17 November 2001; Accepted 26 April 2002

Abstract: We present results of developing a methodology suitable for producing molecular mechanics force fields with explicit treatment of electrostatic polarization for proteins and other molecular system of biological interest. The technique allows simulation of realistic-size systems. Employing high-level *ab initio* data as a target for fitting allows us to avoid the problem of the lack of detailed experimental data. Using the fast and reliable quantum mechanical methods supplies robust fitting data for the resulting parameter sets. As a result, gas-phase many-body effects for dipeptides are captured within the average RMSD of 0.22 kcal/mol from their *ab initio* values, and conformational energies for the di- and tetrapeptides are reproduced within the average RMSD of 0.43 kcal/mol from their quantum mechanical counterparts. The latter is achieved in part because of application of a novel torsional fitting technique recently developed in our group, which has already been used to greatly improve accuracy of the peptide conformational equilibrium prediction with the OPLS-AA force field.[1] Finally, we have employed the newly developed first-generation model in computing gas-phase conformations of real proteins, as well as in molecular dynamics studies of the systems. The results show that, although the overall accuracy is no better than what can be achieved with a fixed-charges model, the methodology produces robust results, permits reasonably low computational cost, and avoids other computational problems typical for polarizable force fields. It can be considered as a solid basis for building a more accurate and complete second-generation model.

© 2002 Wiley Periodicals, Inc. J Comput Chem 23: 1515–1531, 2002

Key words: polarizable force field; *ab initio* quantum chemistry; gas phase tests

Introduction

The explicit inclusion of polarization in molecular mechanics force field is a long-standing objective of molecular modeling. Over the past 5 years, significant progress has been made in developing polarizable models for small molecules,[2] particularly water;[3] such models now reproduce experimental gas phase and condensed phase data with a high degree of accuracy for many properties of interest. Although there is still much work to be done in the refinement of such models, as well as fundamental issues that must be addressed to better understand the comparisons between theory and experiment (e.g., influence of quantum effects on nuclear motion in predicting condensed phase dynamical properties, effects of Pauli exclusion upon polarizabilities and charges in the condensed phase), it is fair to say that success has been achieved in constructing an effective technology for small-molecule force field development.

However, interesting biological applications require treatment not only of small molecules but of larger medicinal compounds and peptides and also of macromolecules. Polarizable force field development for such systems is in a much less advanced state. There are few existing pubications in which systems with more

Correspondence to: R. A. Friesner; email: rich@chem.columbia.edu

Contract/grant sponsor: National Institutes of Health; contract/grant numbers: GM52018 (R.A.F.), GM19961 (G.A.K.), and GM43340 (B.J.B.)

Contract/grant sponsor: NSF; contract/grant number: CHE-00-76279 (B.J.B.)

Contract/grant sponsor: the National Computational Science Alliance; contract/grant number: MCA95C007N (utilizing the NCSA SGI/CRAY Origin 2000)

This article includes Supplementary Material available from the authors upon request or via the Internet at ftp://ftp.wiley.com/public/journals/jcc/suppmat/23/1515 or http://www.interscience.wiley.com/jpages/0192-8651/suppmat/v23.1515.html

© 2002 Wiley Periodicals, Inc.

than five heavy atoms per molecule are addressed,[2f,4] and none to our knowledge in which a complete force field suitable for macromolecular simulations has been presented. There are several reasons for this. First, parametrization of larger systems is a formidable problem both technically and because of a lack of suitable experimental data. Second, one has to be very careful about avoidance of a "polarization catastrophe" in which the variable part of the charge distribution grows without bound, leading to nonsensical structures and energies. Third, validation of such a force field is a difficult problem. Reliable testing requires an efficient simulation algorithm with appropriate boundary conditions and a model for aqueous solvation, which is compatible with the new force field.

In the present article, we take an initial step towards these objectives by presenting a methodology based primarily on *ab initio* quantum chemistry, and a first generation polarizable protein force field, which is tested in the gas phase. Althogh we have described some of the technology previously and demonstrated applications to small molecule systems,[5] we regard the scale-up to a complete protein force field as a nontrivial demonstration of the promise of our approach; the gas phase tests presented herein, while unable to definitively evaluate accuracy, provide substantial evidence that the model behaves reasonably under a variety of conditions. Accuracy of ca. 0.5 kcal/mol in evaluating the molecular interactions is similar to what was obtained in a previous article for the OPLS-AA fixed-charges force field,[1] and we consider such an accuracy to be a sufficiently good target in this project. Future work will involve improvement of the force field to incorporate liquid state simulation data into the fitting process and extensive testing in solvent for structural and energetic predictive capabilities.

The article is organized as follows. In the next section, we briefly review our methodology for construction of the electrostatic model of an arbitrary molecule (both the permanent and polarizable components of the model), and present a few examples examining accuracy when the method is applied to model dipeptides. We then explain how this model is then deployed to build an electrostatic model for a polypeptide chain, computing parameters for all 20 amino acids in various protonation states and then transferring the parameters to assemble the macromolecular electrostastic model. In the next section we discuss how van der Waals parameters are determined from fitting *ab initio* quantum chemical data. The Valence Part of the Force Field section presents methods for parametrization of the valence energy; we retain stretching and bending terms from the OPLS-AA force field, so emphasis is on fitting of torsional parameters. The Applications of the Polarizable Force Field section describes computational methods for evaluating energies and forces of the protein model for use in molecular dynamics and presents results for gas phase protein minimizations and molecular dynamics simulations, examining RMS deviations of the computed structure from the native structure and comparing with fixed charge force field simulations. Note that one does not expect these simulation results to be superior to a fixed charge force field because at this point we have not included an explicit or implicit solvent model, and have not carried out such an extensive testing of our parameters as was done, for example, for the OPLS-AA in a course of many years; however, we can establish whether the native structure remains a reasonable local minimum and whether the model exhibits a polarization catastrophe.

Polarizable Electrostatics

The Model

The electrostatics of a molecule are represented by a set of fixed bond-charge increments between pairs of bonded sites, and polarizable dipoles placed on sites. We use the term "site" to denote either an atom or an off-atom virtual site. Bond-charge increments are convenient in that site charges result only from the assignment of equal (in magnitude) and opposite partial charges on bonded neighbors, thus ensuring that the molecule is always neutral (or always has a fixed nonzero charge, if additional fixed charges are added). We may represent transfer of (positive) charge from site i to site j by a bond-charge increment q_{ij}, which contributes a charge $-q_{ij}$ to site i and $+q_{ij}$ to site j. The total charge on a site is then the sum of the contributions from all bond-charge increments containing that site. In this article we employ fixed bond charge increments, and polarization response only springs from the polarizable sites with inducible dipoles interacting with these bond charges, external field, and each other. In other work we have also allowed the bond charge increments to fluctuate in response to changing electrostatic environment (for representative examples, see refs. 3a and 9).

The expression for the energy of an induced dipole moment $\underline{\boldsymbol{\mu}}_i$ on a site i is

$$U(\underline{\mu}_i) = \underline{\chi}_i \cdot \underline{\mu}_i + \tfrac{1}{2}\, \underline{\mu}_i \cdot \underline{\alpha}_i^{-1} \cdot \underline{\mu}_i \qquad (1)$$

The quadratic term is the familiar self-energy of an induced dipole; $\underline{\boldsymbol{\alpha}}_i$ is the polarizability of site i. The linear coefficient $\underline{\boldsymbol{\chi}}_i$ represents (the negative of) an "intrinsic" electric field at site i—that is, an electric field that exists even in the absence of any other sites or external fields. We would expect $\underline{\boldsymbol{\chi}}_i$ to be nonzero only if the site were part of an asymmetric molecule. The parameter $\underline{\boldsymbol{\chi}}_i$ is really just a way to introduce a "permanent" nonzero dipole moment in an isolated molecule; by completing the squares we could have written eq. (1) up to a constant in the form; $\frac{1}{2}\,(\underline{\mu}_i - \underline{\mu}_i^0) \cdot \alpha_i^{-1} \cdot (\underline{\mu}_i - \underline{\mu}_i^0)$; where the permanent dipoles are $\underline{\boldsymbol{\mu}}_i^{\mathbf{0}} = -\underline{\boldsymbol{\alpha}} \cdot \underline{\boldsymbol{\chi}}_i$; however, eq. (1) is somewhat more convenient in that one need keep track of only one dipole moment on a site (rather than both a permanent and induced dipole moment).

The electrostatics of a system of molecules is represented by a collection of interacting bond-charge increments and dipoles. We introduce a scalar coupling $J_{ij,kl}$ between bond-charge increments on sites i,j and k,l; a vector coupling $\mathbf{S}_{ij,k}$ between a bond-charge increment on sites i,j and a dipole on site k; and a rank-two tensor coupling $\mathbf{T}_{i,j}$ between dipoles on sites i and j. Then the total energy is

$$U(\{q_{ij}\}, \{\underline{\boldsymbol{\mu}}_i\}) = \sum_i \left(\underline{\boldsymbol{\chi}}_i \cdot \underline{\boldsymbol{\mu}}_i + \frac{1}{2}\, \underline{\boldsymbol{\mu}}_i \cdot \underline{\boldsymbol{\alpha}}_i^{-1} \cdot \underline{\boldsymbol{\mu}}_i \right) + \frac{1}{2} \sum_{ij \neq kl} q_{ij} J_{ij,kl} q_{kl}$$

$$+ \sum_{ij,k} q_{ij} \mathbf{S}_{ij,k} \cdot \underline{\boldsymbol{\mu}}_k + \frac{1}{2} \sum_{i \neq j} \underline{\boldsymbol{\mu}}_i \cdot \mathbf{T}_{i,j} \cdot \underline{\boldsymbol{\mu}}_j \qquad (2)$$

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

$$J_{ij,kl} = \frac{1}{r_{ik}} - \frac{1}{r_{il}} - \frac{1}{r_{jk}} + \frac{1}{r_{jl}} \tag{3}$$

$$\mathbf{S}_{ij,k} = \frac{\mathbf{r}_{ik}}{r_{ik}^3} - \frac{\mathbf{r}_{jk}}{r_{jk}^3} \tag{4}$$

$$\mathbf{T}_{i,j} = \frac{1}{r_{ij}^3}\left(\mathbf{1} - 3\,\frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{\mathbf{r}_{ij}^2}\right) \tag{5}$$

The Coulomb interaction diverges as the distance between bond-charge increments and dipoles goes to zero, so will not be appropriate if they are too close. Physically, this represents the fact that a point multipole description is only accurate from far enough away. This can be remedied by omitting 1,2- and 1,3-interactions (between sites connected with each other directly or through a third one), as is commonly done in molecular-mechanics force fields.

For every spatial configuration of atoms, the dipole moments are determined by minimizing the total energy as given in eq. (2); that is, requiring that

$$\nabla_{\mu i} U(\{q_{ij}\}, \{\underline{\boldsymbol{\mu}}_i\}) = 0 \tag{6}$$

for all i. This is equivalent to the usual "self-consistent field" determination of induced dipole moments. Note that eq. (6) only specifies a minimum if the matrix of second derivatives of eq. (2) with respect to the dipole moments,

$$\nabla_{\mu i}\nabla_{\mu j} U = \underline{\boldsymbol{\alpha}}_i^{-1}\delta_{ij} + \mathbf{T}_{ij}(1 - \delta_{ij}), \tag{7}$$

is positive definite. If this matrix is not positive definite, no minimum of the total energy exists; the polarization energy can become arbitrarily large and negative. This is the so-called "polarization catastrophe" and again is due physically to the fact that the point multipole description is only accurate from far enough away. It should be noted here that, although we did not use any electrostatic screening of the inducible dipoles in this work, the Lennard–Jones part of the Hamiltonian was enough to prevent the molecular systems involved from approaching the "polarization catastrophe" regions mentioned above.

Equation (6) may be solved by matrix diagonalization or by iterative methods. Alternately, the dipole moments may be assigned fictitious masses and kinetic energies and integrated along with the spatial coordinates in the extended Lagrangian scheme.[6–9] The dynamics of inducible dipoles so generated is fictitious, and functions only as a way to keep the electronic degrees of freedom close to the minimum-energy "Born-Oppenheimer" surface.

Parameterization

Parameterizing the electrostatic model for a given molecule involves the following steps: (1) choosing virtual sites, (2) choosing sites on which a dipole moment will be placed, (3) fitting the polarizabilities—the parameters that specify the electrostatic response, and (4) fitting the "intrinsic fields" and fixed bond-charge increments—the parameters that describe the electrostatics of an isolated molecule.

Massless virtual sites representing lone pairs were attached to oxygen atoms at a distance of ca. 0.47 Å. For sp^3-hybridized oxygens (e.g., alcohols) the virtual sites mimic a tetrahedral geometry: they lie in the plane perpendicular to the plane containing the bisector of the angle between the oxygen and its two bond atoms, and make an angle of $\pm\tan^{-1}(2^{1/2}) \approx \pm 54.74°$ from the plane containing these three atoms. For sp^2-hybridized oxygen atoms (e.g., carbonyls) the virtual sites mimic a trigonal-planar geometry: they lie in the plane containing the oxygen, carbon, and atom bonded to the carbon and make an angle $\pm 120°$ with the O—C bond.

Computational experiments have indicated that the effect of oxygen lone pairs on structures and energies is in general greater than that of nitrogen, although it is certainly possible to find cases where nitrogen lone pairs are essential. We expect to add nitrogen lone pairs in the next generation of our polarizable force field effort.

Bond-charge increments were placed on all sites, but permanent dipoles are only induced on polarizable atoms.

The electrostatic parameters of the model were fit in a manner similar to that described in refs. 5b and 10. We applied a series of electrostatic perturbations to the target molecule, in the form of dipolar probes consisting of two opposite charges of magnitude 0.78 e, 0.58 Å apart (for a dipole moment of 2.17D—similar to that of nonpolarizable models for liquid water such as SPC/E[11]), placed at various locations. The outcome of the fitting procedure was relatively insensitive to the exact form of the perturbations, i.e., the magnitude or position of the probe charges. For each perturbation, the change in the electrostatic potential (ESP) at a set of gridpoints outside the van der Waals surface of the molecule was computed using density-functional theory (DFT) with the B3LYP method[12,13] and cc-pVTZ(-f) basis set. All calculations were performed with the Jaguar electronic structure code.[14] The polarizablities α_i are assumed to be isotropic and are chosen to minimize the mean-square deviation between the change in the ESP as given by model and by the DFT calculations. Next, $\underline{\boldsymbol{\chi}}_i$ and the fixed bond-charge q_{ij} are fit so as to best reproduce DFT calculations of the ESP of the charge distribution of the unperturbed target molecule. For each peptide molecule, the fitting was done using several conformers simultaneously. Numbers of conformations for each particular case are given in tables accompanying the Valence Part of the Force Field section. The vectors $\underline{\boldsymbol{\chi}}_i$ are expressed as a sum of vector parameters pointing along bonds connecting adjacent atoms, and as such, will change during the course of a simulation as a flexible molecule changes conformation.

The choice of electronic structure method (DFT/B3LYP functional, cc-pVTZ (-f) basis set) yields quite accurate permanent charge distributions, but underestimates the gas phase polarizability compared to experiment. Closer agreement with gas phase experiments could be obtained by including diffuse functions in the DFT calculations. However, our computational experiments with liquid state simulations strongly suggest that these diffuse function contributions are considerably damped in the condensed phase, and that ignoring them is in fact a much better approximation than fully including them. Briefly, the theoretical argument is that in the condensed phase, Pauli repulsion from neighboring molecules raises the energies of diffuse functions and so diminshes

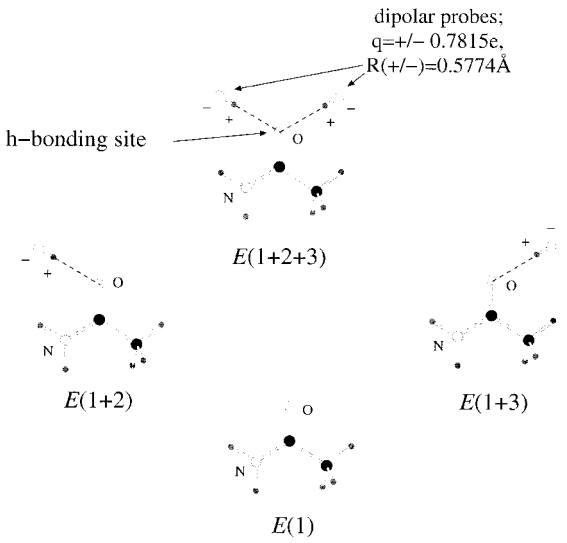

Figure 1. Definitions of two- and three-body energies used as the target/test of the polarizable force field quality.

their contribution to the polarization. Empirically, when diffuse functions are used to develop polarization responses for small molecules, liquid state simulations of these molecules manifest overpolarization of the solvent, in some cases leading to polarization catastrophes; quantitative properties such as the dielectric constant are also too large compared to experiment. We discuss this point further in a separate publication focused on liquid state simulation results.[15]

Results of Electrostatic Parameterization for Peptide Residues

The methodology described above was applied to all the residues in dipeptide form to produce the electrostatic part of the force field. Ability to reproduce two- and three-body energies of interaction of the dipeptides with the electrostatic probes described in the section above was used to validate the quality of the resulting parameters. The three-body energies were computed as follows:

$$E_{123} = E(1+2+3) - E(1+2) - E(1+3) - E(2+3) + E(1) + E(2) + E(3) \quad (8)$$

Here, $E(1+2+3)$ is the energy of all the three bodies put together (Fig. 1), $E(1+2)$, $E(1+3)$, and $E(2+3)$ represent two-body interaction energies, and finally $E(1)$, $E(2)$, and $E(3)$, are the energies of the molecule and the probes alone, respectively. The three-body response E_{123} is independent of the permanent charge distribution and is only influenced by the polarizable part of the electrostatics. First of all, the probes were placed at hydrogen-bonding positions. Then more probes were placed at random locations around the molecules. The total number of probes for each dipeptide was ca. 30–40 for each conformer. The distance of hydrogen-bonded probes from the acceptor or donor on the target molecule was fixed at 1.8 Å. The randomly positioned probes were also constrained to be located at 1.8 Å from the closest atomic site of the molecule involved. Table 1 shows RMS deviations/maximum deviations of the three-body energies for all the dipeptides from their DFT/cc-pVTZ(-f) counterparts, along with the *ab initio* average and maximum values of the three-body energies. It should be noted that, although the deviations are often similar in magnitudes to the averages, the accuracy of our fitting is still good because of their low absolute values. We have previously demonstrated that to capture the many-body effects one needs to use inducible dipoles; fluctuating charges alone do not suffice in some important cases.[5b] (It is conceivable that one could also attempt to solve this problem by placing additional fluctuating charge sites at locations other than atomic centers). The data in the Table 1 demonstrate that the polarization based on inducible dipoles alone (without the use of any fluctuating charges) provides an adequate representation of the many-body responses, with the greatest RMSD being only 0.450 kcal/mol. Magnitudes of the three-body energies themselves are typically within a couple of kcal/mol. The agreement can be further improved by including into the model both inducible dipoles and fluctuating charges.

To assess the quality of the permanent charges and dipoles, two-body energies $E(1–2)$–$E(1)$–$E(2)$ (see Fig. 1) were compared to their DFT counterparts. Table 2 presents RMS deviations for all the dipeptides involved as well as average values of the two-body energies (only those two-body energies with magnitudes under 30.0 kcal/mol were used in the comparison). In case of charged residues, formal charges were placed on appropriate atomic sites, followed by their redistribution through bond-charge increments in the process of fitting. It should be emphasized that, although the

Table 1. RMS Deviations of Three-Body Energies for Dipeptides Computed with Quantum Mechanics and Polarizable Force Field and Average Three-Body Energies.

Dipeptide	E_{123} RMS deviations/ maximum deviations, kcal/mol	Average/maximum E_{123}, kcal/mol
Alanine	0.158/0.466	0.299/0.784
Serine	0.173/0.785	0.257/1.629
Phenylalanine	0.122/0.281	0.195/0.790
Cysteine	0.293/1.350	0.280/2.016
Asparagine	0.267/1.363	0.275/1.425
Glutamine	0.208/2.888	0.227/3.174
Histidine	0.280/1.614	0.249/4.267
Leucine	0.152/0.521	0.278/1.013
Isoleucine	0.258/1.984	0.307/2.592
Valine	0.136/0.340	0.286/1.154
Methionine	0.245/1.167	0.260/2.007
Proline	0.171/0.285	0.370/0.927
Tryptophan	0.276/0.949	0.196/4.178
Threonine	0.182/0.935	0.287/1.744
Tyrosine	0.450/1.775	0.179/1.402
Aspartic acid	0.333/1.155	0.376/3.206
Glutamic acid	0.244/1.146	0.282/2.161
Lysine	0.166/1.693	0.162/2.273
Protonated histidine	0.130/0.628	0.174/1.137
Arginine	0.120/1.253	0.196/1.798

Table 2. RMS Deviations of Two-Body Energies for Dipeptides Computed with Quantum Mechanics and Polarizable Force Field, and Average Two-Body Energies in kcal/mol[a]

Dipeptide	Number of points (N)	E_{12} RMSD	Average E_{12}
Alanine	225	1.017	3.236
Serine	221	1.332	4.664
Phenylalanine	130	1.181	16.023
Cysteine	198	1.442	4.543
Asparagine	77	2.902	5.710
Glutamine	509	2.732	7.413
Histidine	351	3.021	5.567
Leucine	364	3.187	4.418
Isoleucine	318	3.073	6.149
Valine	114	3.317	20.301
Methionine	296	2.053	5.238
Proline	35	1.000	3.834
Tryptophan	439	2.342	10.548
Threonine	307	1.456	6.837
Tyrosone	286	2.488	9.136
Aspartic acid	125	1.679	14.732
Glut. acid	327	1.809	15.515
Lysine	267	1.695	13.799
Histidine-H^+	258	2.181	11.653
Arginine	269	3.278	11.527

[a]Only those points with the magnitudes of the two-body energies below 30 kcal/mol are counted.

RMS deviations in Table 2 are greater than the target 0.5 kcal/mol, we are dealing with a totally different nature of energies in this case. The two-body energies are purely electrostatic interactions between a molecule and the bare charges of the dipole probes, with the distance to the closest charge of only 1.8 Å. Therefore, their magnitudes are much greater than those of relative conformational energies or even hydrogen-bonded dimers, and thus the deviation magnitudes are also bound to be significantly larger.

It is known that instabilities may arise in ESP fitting if charges are poorly determined by the set of gridpoints; for instance, in the case of charges on "buried" atoms far inside the van der Waals surface.[5] Instabilities might show up in unreasonable values for charges or dipole moments or small or negative eigenvalues in the matrix J. As in previous work,[5] we address this problem by zeroing poorly determined modes via singular value decomposition. In general, our protocol prescribes zeroing as many of the modes in the fitting as was possible without a significant increase in the two-body energy deviations from their *ab initio* values. In some cases there is an obvious point at which to stop cutting modes; in others, the behavior of the two body RMSD as a function of number of modes cut is smoother, in which case the results are relatively insensitive to the exact point of truncation. In this article we chose the number of modes cut for each dipeptide heuristically by examining the two body RMSD as a function of the number of modes cut; in future work, we intend to develop a more automated criteria.

One can see from the results in Table 2 that the RMS deviation of the two-body energies is no more that 3.317 kcal/mol (and less on average). Therefore, we have managed to reproduce both the many-body electrostatic effects, important in representing liquid-state properties, and the electrostatic potential itself, which is important in obtaining correct dimerization energies. It is not possible to reproduce the both with a fixed-charges model, with no explicit polarization included. Moreover, it should be pointed out that our electrostatic model has also shown a good degree of transferability. First, the parameters produced for the alanine dipeptide were transferred without any modifications to the alanine tetrapeptide, which, as will be shown in subsequent sections, gives results in close agreement with quantum mechanical data. Second and even more important, all the backbone parameters in all the amino acids were taken directly from the alanine dipeptide case, with no refitting or adjustments, and this still produced good agreement with the high-level *ab initio* two- and three-body energies as demonstrated by the data in Tables 1 and 2. Only parameters for the side chains were refitted, as well as those including both the backbone and the side chains. For example, when producing parameters for the serine dipeptide, we refitted every parameter for the —OH group plus the C—O(H) bond charge increments and permanent dipoles. Therefore, our parameters are transferable enough, which is crucial if one is to assemble and simulate actual protein systems out of the building blocks represented by the dipeptides.

Lennard–Jones Parametrization of the Force Field

The Target–**Ab Initio** *Energies*

Fitting the Lennard–Jones component of the force field was done with high accuracy *ab initio* results for intermolecular hydrogen bonding interactions as a target. We used the ability of our force field to reproduce gas-phase dimerization energies for model organic compounds, analogous to actual protein side-chain groups, as the criterion of the validity of the Lennard–Jones component of the Hamiltonian (with the electrostatic interactions assessed as described in the previous section). The functional form of this term is described by eq. (9):

$$E_{LJ} = \sum_{i \neq j} 4\varepsilon_{ij}[(\sigma_{ij}/R_{ij})^{12} - (\sigma_{ij}/R_{ij})^{6}]f_{ij} \qquad (9)$$

Geometric combining rules were employed for both σ and ε: $\sigma_{ij} = (\sigma_{ii} \cdot \sigma_{jj})^{1/2}$, $\varepsilon_{ij} = (\varepsilon_{ii} \cdot \varepsilon_{jj})^{1/2}$, and the scaling factor f_{ij} was set to 0.0 for atoms connected by a valence bond or angle (1,2- and 1,3- interactions, respectively), to 0.5 for the 1,4-interactions, and to 1.0 for the rest of i-j pairs.

In the development of a polarizable (as opposed to fixed charge) force field, the objective is to reproduce the true gas phase intermolecular binding affinities and geometries as accurately as possible. For the present efforts, we set a target of ~0.25 kcal/mole or better for the precision of the binding affinity. For hydrogen bonded dimers, this level of error can be attained via MP2 calculations extrapolated to the basis set limit, as has been demonstrated for example in recent work of Tsuzuki et al.,[16] where the contribution of higher level excitations [e.g., CCSD (T)] was shown to

Table 3. Comparison of *Ab Initio* Dimerization Energies.

Dimer	LMP2[a]	MP2[a]	CCSDT limit[b]
H_2O MeOH	4.80	4.99	4.90
H_2O Me_2O	5.74	5.70	5.51
H_2O H_2CO	5.15	5.21	5.17
MeOH MeOH	5.54	5.58	5.45
HCOOH — HCOOH	13.92	13.79	13.93

[a]Plus extrapolation.
[b]Ref. 16.

be negligible (note that there are intermolecular interactions, such as pi stacking of aromatic rings, where MP2 level calculations are not adequate to achieve the target accuracy). The methods used in this work to calculate binding energies are based on an MP2 extrapolation procedure that we have developed using our pseudospectral local MP2 (LMP2) approach.[17] The details of the method are summarized below and some representative examples provided.

The method is grounded in the LMP2 wave function, which is a form of canonical MP2 developed by Pulay[17] for computational efficiency as well as for removal of basis set superposition error (BSSE). Our pseudospectral implementation of LMP2[18] has a scaling with sytstem size of $\sim N^{2.5}$ allowing this method to be applied with large basis sets in reasonable CPU times. The elimination of BSSE effects with LMP2 implies that the method converges more quickly with basis set size,[18] which is important for the extrapolation to work with modest basis sets.

Dimer geometries were obtained by LMP2 optimizations with a cc-pVTZ(-f) basis set.[19] In the spirit of the extrapolation method of ref. 16, the empirical dimer binding energy consists of the LMP2 binding energy for a smaller cc-pVTZ(-f) basis set (E_{ccpvtz}) and the LMP2 binding energy with a larger cc-pVQZ(-g) basis set (E_{ccpvq}). The model binding energy E_{bind} takes the simple form:

$$E_{\mathrm{bind}} = C_1 \cdot E_{\mathrm{ccpvtz}} + C_2 \cdot E_{\mathrm{ccpvqz}} \tag{10a}$$

$$C_1 = a_1/(a_1 - a_2);\ C_2 = -a_2/(a_1 - a_2) \tag{10b}$$

$$a_1 = \exp(-2.7);\ a_2 = \exp(-1.8) \tag{10c}$$

The coefficients C_1 and C_2 were fit to the set of MP2 extrapolated dimer binding energies of ref. 1. In calculating binding energies the Hartree–Fock (HF) energies are corrected for BSSE at the HF level using the counterpoise method.

Table 3 presents a comparison of the binding energies obtained with the extrapolation above with those of Tsuzuki et al.,[16] including coupled cluster results. It shows that the extrapolation method we use allows one to produce accurate dimerization energies at relatively low computational cost. The extrapolated LMP2 dimerization energies are, on average, only 0.09 kca/mol away from the CCSD(T) limit results, with the maximum deviation of 0.23 kcal/mol. The average error is thus smaller than the error for the MP2 extrapolation (including cc-PV5Z basis set) reported in ref. 16, which was 0.12 kcal/mol.[16] Therefore, we adopted the LMP2/extrapolation technique for providing targets in the Lennard–Jones fitting procedure. A detailed discussion of the above extrapolation method, including validation on additional test molecules without further parameter adjustment, will be presented in a separate publication.

Fitting Dimerization Energies

To produce σ and ε values to be used in eq. (9), we performed a series of minimization of gas-phase dimers. CH_3OH and NH_2COCH_3 homodimers were considered, as well as heterodimers of a variety of organic molecules—analogs of peptide side chains—with the NH_2COCH_3. *Ab initio* geometry optimizations were run at the LMP2/6-31G** level, followed by single-point energy calculations performed with the extrapolation technique described in the previous subsection. Lennard–Jones parameter values were adjusted to reproduce both the dimerization energies and distances between the heavy atoms. Polar hydrogens had both σ and ε set to 0.0, just like in the case of the OPLS-AA force field. The electrostatic part of the molecular mechanics force field was produced as described in the previous section.

Table 4 shows the results. It can be seen that the agreement with the *ab initio* results is very good. The energy deviations are within 0.5 kcal/mol in all the cases except for the dimers of NH_2COCH_3 with two charged molecules—$CH_3NH_3^+$ and the protonated histidine analog. Even in the two latter cases, the error is only 0.6 kcal/mol. Distance between heavy atoms obtained through the molecular mechanics calculations agree with the quantum mechanical results with no worse than 0.15 Å error.

The above results were obtained with all the Lennard–Jones parameters on hydrogen and carbon atoms being the same as in the standard OPLS-AA.[20] We have found that it was enough to refit the heteroatoms parameters only. The resultant σ and ε values are listed in Table 5.

It should be mentioned that we have not investigated the performance of the present fitting protocol for liquid state simulations. In fitting two Lennard–Jones parameters to the binding

Table 4. Dimerization Energies (kcal/mol)/Distances (Å).

System	*Ab initio*[a]	PFF
MeOH-MeOH	−5.6/2.89	−5.6/2.85
$(NH_2COCH_3)_2$	−7.6/2.05	−8.1/1.93
NH_2COCH_3-MeOH	−14.5/1.76	−14.9/1.78
NH_2COCH_3-MeSH	−5.0/3.66	−5.1/3.58
NH_2COCH_3-phenol	−9.9/1.87	−10.1/1.82
NH_2COCH_3-$CH_3CO_2^-$	−24.7/2.79	−24.6/2.71
NH_2COCH_3-$CH_3NH_3^+$	−29.5/1.56	−28.9/1.41
NH_2COCH_3-histidine$^+$ analog	−23.8/1.63	−23.2/1.53
NH_2COCH_3-arginine$^+$ analog	−25.6/1.92	−25.2/1.85
NH_2COCH_3-tryptophan	−9.4/1.97	−9.4/1.96
NH_2COCH_3-histidine analog	−8.9/1.95	−8.7/1.96

[a]LMP2/cc-pVQZ(-f) with extrapolation.

Table 5. New and Old (OPLS-AA) σ and ε Values.

Atom	σ/ε, PFF	σ/ε, OPLS-AA[a]
O, sp^2, backbone and side chains	3.16/0.280	2.96/0.210
NH, backbone and side chains	3.30/0.280	3.25/0.170
OH, serine	3.25/0.280	3.12/0.170
OH, tyrosine	3.20/0.190	3.07/0.170
S, cysteine and methionine	3.60/0.425	3.60/0.425
N heterocycle, histidine	3.25/0.170	3.25/0.170
N3, lysine	3.0/0.080	3.25/0.170
O2, aspartic and glutamic acid	2.96/0.210	2.96/0.210
N2, arginine	3.20/0.100	3.25/0.170

[a]Refs. 1 and 20.

affinity and hydrogen bonding distance of molecular dimers, there are no parameters remaining to independently adjust the long range dispersive part of the pair potential. Although the values we obtain for Lennard–Jones B coefficients [$B_{ij} = 4\varepsilon_{ij}\sigma_{ij}$,[12] see eq. (9)] are qualitatively reasonable, lack of quantitative precision could affect quantities such as the liquid state heat of vaporization or density, which can be quite sensitive to the long range behavior of the potential function.

The rigorous solution to this problem is to incorporate additional degrees of freedom in the pair potential (such as an exponential term in van-der-Waals energy expression), and to couple the development of parameters with liquid state simulations. We are pursuing this direction in other work, which will be reported in subsequent publications. For the present article, we do not believe that the precise value of the dispersive interactions (given that, as argued above, they are in the correct ballpark) will have a large effect on local hydrogen bonded structure or packing interactions. Thus, we believe that the calibrations below of dipeptide conformational energetics, gas-phase protein minimizations and short molecular dynamics simulations would be relatively unaffected by the pair potential modifications suggested above.

Finally, a principal hypothesis of the above approach is that the Lennard–Jones parameters depend principally on the local chemical functional group, and thus can be transferred from small molecule models to larger systems without noticeable sacrifice of accuracy. With this assumption, a prescription is in place for completely specifying the nonbonded component of the force field, and what remains is to determine the valence part of the force field. We discuss the fitting methodology and results in the next section.

Valence Part of the Force Field—Refitting the Torsional Potential

The Method

We can now discuss the remaining parts of the force field, namely the bond stretching, angle bending, and torsional contributions into the total Hamiltonian in eq. (11):

$$E_{\text{total}} = E_{\text{electrostatics}} + E_{LJ} + E_{\text{bonds}} + E_{\text{angles}} + E_{\text{torsion}} \quad (11)$$

For all the calculations presented in this work, we retained the harmonic form of the stretching and bending potential:

$$E_{\text{bonds}} = \sum_{\text{bonds}} K_r(r - r_{eq})^2 \quad (12)$$

$$E_{\text{angles}} = \sum_{\text{angles}} K_\Theta(\Theta - \Theta_{eq})^2 \quad (13)$$

where K_r and K_Θ represent the force constants; r, and Θ are actual values of bond lengths and angles; r_{eq} and Θ_{eq} are their equilibrium magnitudes, taken directly from the OPLS-AA fixed charge force field.[20] On one hand, this approach is validated by the fact that the OPLS-AA is a technique carefully built over more than two decades and tested on a wide variety of organically and biophysically relevant systems. On the other hand, the OPLS-AA does not include interactions between atomic sites in the same covalent bond or covalent angle into the total energy, and so the above harmonic terms are the only ones to represent the corresponding part of the Hamiltonian. The same is also true for the presented polarizable force field, and thus we believed that we could do without reparametrization of bond stretching and angle bending. The results we obtained have proved that this approach was correct, as the accuarcy of our results presented below is no worse than that achieved with the OPLS-AA force field, partly reparameterized to better reproduce gas-phase conformational energies of proteins.[1]

$$E_{\text{torsion}} = \frac{1}{2}\sum_{\text{dihedrals}} V_1^i[1 + \cos(\phi_i)] + V_2^i[1 - \cos(2\phi_i)] + V_3^i[1 + \cos(3\phi_i)] \quad (14)$$

On the other hand, the torsional part of the Hamiltonian, with the Fourier expansion functional form presented in eq. (14), was built in a different way. First of all, nonkey torsions, such as methyl group-rotation parameters and out-of-plane bending improper tosions, were taken directly from the OPLS-AA. But all parameters pertaining to the peptide backbone ϕ and ψ, as well as side-chain χs, were fitted from scratch with a fitting technique described in detail in ref. 1. The new parameters were produced and tested on all the possible dipeptides and alanine tetrapeptide.

A concise summary of the torsional fitting technique is as follows: (1) first of all, the fitting was done with high-level *ab initio* data as the target. We ran LMP2/cc-pVTZ(-f)//HF-6-31G** calculations with Jaguar software suite.[14] (2) Our choice of the fitting subspace is illustrated on Figure 2 (for the alanine dipeptide). Out of the six alanine dipeptide local minima previously found,[21] only two are shown on Figure 2 for the sake of clarity. (3) We used a non-Boltzmann weighting scheme for the error at the fitting points:

$$W_i = A \cdot \exp(-b \cdot G_i) \quad (15)$$

Here G_i stands for the absolute value of the torsional surface gradient at the point i, for which the weight W_i is to be produced. The coefficient A is adjusted to change the maximum weight/

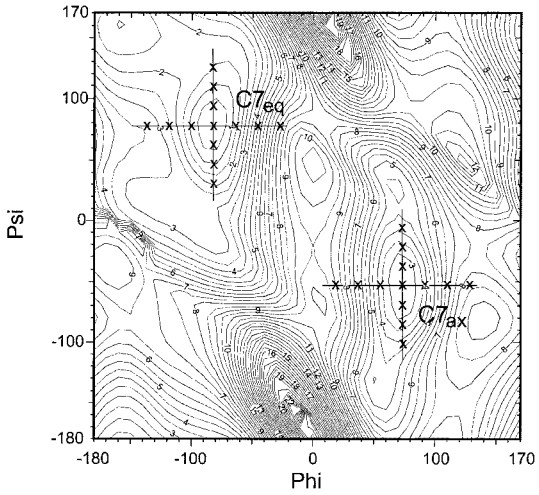

Figure 2. Crosslike torsional fitting subspase, examplified on the alanine dipeptide ϕ/ψ potential energy surface. The crosses were placed at each minima and each arm contained four fitting points. Some crosses and points are omitted for clarity on this figure.

minimum weight ratio for the fitting and is chosen independently for each particular dipeptide fitting. (4) Treating charged residues required a special approach to sample the part of the conformational space relevant in the liquid state, while using gas-phase calculations. Liquid-phase SCRF runs at HF/6-31G** level were used to find the solvated energy minimum structures. Then liquid-phase restrained *ab initio* geometry optimizations were carried out, in the same way as for the uncharged dipeptides to obtain the data for the cross-shaped fitting subspaces. Finally, gas-phase single point LMP2/cc-pVTZ(-f) calculations were carried out to find the final target energies. Polarizable molecular mechanics runs were also performed in the gas phase, with all the principal dihedral angles restrained to their positions found in the hydrated *ab initio* minimizations.

The above torsional fitting technique has been tested previously by developing a new set of torsional parameters for the OPLS-AA di- and tetra-peptides.[1] It allowed to reduce energy RMS deviations of conformational energies of all the electrostatically neutral dipeptides by ca. 40%, from 0.81 kcal/mol down to 0.47 kcal/mol. For the five charged residues, the conformational energy RMSD dropped from 2.20 to 0.94 kcal/mol. The result was achieved without changes in the nonbonded parameters, except for the cases of sulphur-containing dipeptides.

Results

Alanine

Alanine has a special place in this work. On one hand, this is the only system for which not only the dipeptide, but also the tetrapeptide was studied to ensure the transferability of the torsional parameters. On the other hand, all parameters obtained for the alanine dipeptide were then transferred to the rest of the residues without any changes at all. The method worked well, which confirms the transferability of the toriosnal fitting results. The same is also true for the alanine tetrapeptide—the dipeptide parameters were used in it with no further modifications.

Table 6 shows errors in conformational energies of the alanine dipeptide obtained with the presented polarizable force field (PFF), compared to the *ab initio* results at the LMP2/cc-pVTZ(-f)//6-31G** level. For the purpose of benchmarking, we also include results obtained with the OPLS-AA force field, to which the same torsional refitting procedure had been applied in a previous work.[1] The choice of the OPLS-AA is further justified by the fact that it was found to be the best molecular force field available in reproducing the *ab initio* conformational energies of the alanine tetrapeptide (even before modifying the torsional part or it).[21] Also given in Table 6 are RMS deviations of the key dihedral angles values from their *ab initio* counterparts. The dihedrals for the alanine dipeptide are ϕ and ψ, and they are ϕ_{1-3} and ψ_{1-3} for the tetrapeptide. In cases of the other dipeptides, both ϕ and ψ of the backbone and χs of the side chains are counted.

It can be seen from Table 6 that the presented force fields perform adequately in reproducing the alanine dipeptide conformational energies. The energy RMS deviation is only 0.35 kcal.mol. It is greater than the 0.27 kcal/mol for the OPLS-AA, but the both numbers are quite satisfactory in the view of the accuracy of the *ab initio* methodology. The same is true for the dihedral angles RMSD with the polarizable force field (PFF) and OPLS-AA, 7.1° and 6.5°, respectively. Moreover, the greatest discrepancy in the energy is contributed by the high-energy conformer α', with the three lower energy minima in excellent agreement with the LMP2 data.

Only four out of the six *ab initio* minima can be obtained with either force field. The quantum mechanics barriers for the missing minima are very low, and it is probably not ultimately important to reproduce them exactly as true minima on the torsional energy surface.

Fitting the Fourier coefficients for the backbone involved a variety of adjustments. First, gradient weighting was applied in according with eq. (15). The value of parameter b was adjusted to produce 1000.0 ratio of the highest and lowest weights. Second,

Table 6. Alanine Dipeptide, Energy of the Conformers, RMS Deviations in ϕ, ψ from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
$C7_{eq}$	0.00	−0.23/9.3	−0.11/9.1
C5	0.95	0.77/1.2	0.82/4.5
$C7_{ax}$	2.67	2.48/6.5	2.46/0.5
β_2	2.75	—	—
α_L	4.31	—	—
α'	5.51	6.11/8.6	5.97/8.0
RMS error[c]	—	0.35/7.1	0.27/6.5

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 21.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results. The RMS computed for the C7, C5, and α' minima only.

Table 7. Alanine Tetrapeptide, Energy of the Conformers, RMS Deviations in ϕ_{1-3}, ψ_{1-3} from the *Ab Initio* Data.

Conformer	*Ab Initio*[a]	PFF	OPLS-AA/L[b]
1	2.71	3.31/1.0	3.19/4.4
2	2.84	2.87/4.7	3.19/6.5
3	0.00	0.14/7.8	−0.32/8.4
4	4.13	3.85/4.0	4.40/5.8
5	3.88	3.24/16.7	3.14/9.3
6	2.20	0.80/13.9	0.96/12.7
7	5.77	6.91/16.0	5.82/6.6
8	4.16	4.12/47.2	4.83/18.8
9	6.92	7.69/8.8	7.14/8.2
10	6.99	6.69/23.0	7.25/14.2
RMS error[c]	—	0.69/19.1[d]	0.56/10.4

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 21.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.
[d]12.0° without the "runaway" conformer number 8.

conformers $C7_{eq}$, C5, $C7_{ax}$, α_L, and α' were given an extra weight multiplier of 50.0. Finally, points on the target surface for the conformers C5 and $C7_{ax}$ were moved up by 0.3 kcal/mol.

Results of applying the complete PFF to the alanine tetrapeptide are shown in Table 7. Here, the energy and dihedral angles RMS deviations from the *ab initio* are 0.69 kcal/mol and 19.1°, respectively. The latter number is relatively high due to the "runaway" minimum number 8. Excluding this minimum results in the dihedral RMS of 12.0°. The conformer numebr 8 has a low potential energy barrier, and the torsional energy surface is rather flat, as can be seen from the final energy for that case. The rest of the conformers are reproduced well (again, there was no special refitting for the tetrapeptide case). The OPLS-AA results for the energy and angular RMSD are 0.56 kcal/mol and 10.4°, respectively. We conclude that the alanine force field is adequate to our aims, and we adopt it without changes for the backbones of all the other peptides in this work.

Serine

Serine dipeptide conformational study results are presented in Table 8. It can be seen that not only we have managed to obtain a rather low energy and dihedrals RMS deviations of 0.34 kcal/mol and 8.1°, respectively, but also the order of the conformers is correct. This is also the case with the OPLS-AA results (the errors or 0.34 kcal/mol and 4.9°). The original OPLS-AA yielded an energy RMS deviation of 0.47 kcal/mol, which is reasonably low, but the minima were out of order, and only the torsional refitting procedure, introduced earlier and also used in this work, allowed to produced the correct order of the minima energies.[1] The torsional refitting for the PFF involved the gradient weighting with the maximum/minimum ration of 1000.0, assigning the conformers 1 and 3 an extra weight factor of 10.0, and constraining the V_1 Fourier coefficient in the C(O)-CT1-CT-O(H) χ_1 torsion to −4.7 kcal/mol, to avoid its magnitude being over 5.0 kcal/mol (such an increase in the magnitude would produce little positive effect and could distort unrelated areas of the potential surface).

Table 8. Serine Dipeptide, Energy of the Conformers, RMS Deviations in ϕ, ψ, χ_1, χ_2 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L 1[b]	OPLS-AA/L 2[b]
1	0.00	−0.26/5.8	0.49/7.9	0.30/7.8
2	2.76	2.67/8.7	3.30/1.3	2.83/1.6
3	3.75	3.72/12.9	3.08/1.7	3.45/2.2
4	3.95	4.50/3.5	4.12/6.7	3.69/6.7
5	5.13	5.44/7.3	4.90/4.0	4.76/3.7
6	7.43	6.95/7.0	7.13/4.2	7.98/4.0
RMS error[c]	—	0.34/8.1	0.44/4.9	0.34/4.9

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Phenylalanine

Torsional fitting for the phenylalanine dipeptide involved no gradient weighting (maximum/minimum weights ratio set to 1.0). One Fourier coefficient (CA-CT-CT1-N V_1) was manually changed from −1.221 to −2.221 kcal/mol to allow a better overall agreement with the *ab initio* data. The results are shown in Table 9. The very low energy RMS deviation of 0.02 kcal/mol, at the angular RMS error of 9.5°, is explained partly by the fact that only three conformers are found for the system, and partly by the fact that the aromatic ring is represented well by both our method and by the OPLS-AA, which allows energy and dihedral angles RMS deviations of 0.15 kcal/mol and 7.5°, respectively.

Cysteine

In this case, both the conformational energies and key dihedral angles values RMS deviations from the quantum mechanics, obtained with the PFF, were not only quite low, but also lower than

Table 9. Phenylalanine Dipeptide, Energy of the Conformers, RMS Deviations ϕ, ψ, χ_1, χ_2 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	−0.02/8.4	−0.19/6.2
2	0.88	0.91/10.0	0.90/10.6
3	1.65	1.63/10.2	1.82/3.8
RMS error[c]	—	0.02/9.5	0.15/7.5

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Table 10. Cysteine Dipeptide, Energy of the Conformers, RMS Deviations ϕ, ψ, χ_1, χ_2 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	−0.29/7.0	0.15/6.2
2	1.72	1.96/2.2	1.82/3.7
3	2.26	2.43/6.2	2.79/6.7
4	3.18	2.83/3.5	2.84/6.9
5	4.79	5.03/3.5	4.36/4.9
RMS error[c]	—	0.27/4.8	0.35/5.8

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

those produced by the OPLS-AA (0.27 kcal/mol and 4.8° vs. 0.35 kcal/mol and 5.8°). The order of the minima is correct (Table 10). This is a case when the excellent agreement of our model with the LMP2-level quantum mechanics was obtained by setting the highest/lowest weights ratio in the gradient-based weighting in accordance with the eq. (15) to 1000.0. No other adjustments was made.

Asparagine

RMS deviations from the *ab initio* conformational energies and dihedral angles are shown in Table 11. They are 0.02 kcal/mol and 8.7°, respectively. The OPLS-AA results are 0.16 kcal/mol and 19.5°. Thus, the second conformers is represented much better with the PFF than with the OPLS-AA. The maximum weight ratio was set to 1000.0 in this case to obtain the best result, once again confirming the validity of the gradient-based approach to weighting.

Glutamine

Glutamine results are presented in Table 12. The resultant RMD deviations of the conformational energies and the key dihedral angles from *ab initio* data are 0.92 kcal/mol amd 18.0° for the PFF and 0.96 kcal/mol and 13.9° for the OPLS-AA. The numbers are somewhat high, but we are trying to reproduce 11 conformers, more than in any of the previous cases. To produce the parameter set, we only did fitting on the crosslike subspaces for the conformers number 1, 2, 3, 4, and 6. This was done to concentrate on several main motifs in the torsional behavior and avoid unnecessary conformers, which are close to those already on the short list geometrically. The gradient-based weighting was employed, with the maximum weights ratio of 1000.0. In addition, conformers 2 and 6 were given an extra factor of 10.0 in the weights of the points around them.

The results are satisfactory, but a similar trend can be noticed in both the PFF and OPLS-AA results for the lowest conformer—it is too low compared to the conformers 4–11. We could not avoid such a feature while producing a reasonable whole picture for the glutamine dipeptide. This case should and will be investigated further, while the developed force field is applied to realistic biomolecular systems, and relative importance of the conformers in such big systems can be assessed.

Table 11. Asparagine Dipeptide, Energy of the Conformers, RMS Deviations ϕ, ψ, χ_1, χ_2 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	0.02/9.4	−0.16/8.8
2	3.49	3.46/8.0	3.64/26.2
RMS error[c]	—	0.02/8.7	0.16/19.5

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Table 12. Glutamine Dipeptide, Energy of the Conformers, RMS Deviations in ϕ, ψ, χ_1, χ_2, χ_3 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.19	−0.30/8.5	0.30/6.1
2	0.46	−1.03/3.1	0.73/9.7
3	0.00	−0.72/4.5	−0.60/10.4
4	1.07	1.53/13.6	0.52/9.0
5	0.92	1.10/10.8	0.16/21.7
6	1.80	3.53/21.9	1.29/9.8
7	2.83	4.17/12.4	3.91/8.8
8	4.02	3.70/43.0	5.90/8.8
9	5.29	4.66/10.2	5.83/7.9
10	5.32	5.91/22.5	5.72/5.6
11	8.54	7.90/7.6	6.68/31.6
RMS error[c]	—	0.92/18.0	0.96/13.9

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Histidine

Conformational energies and dihedral energy RMS deviations from the quantum mechanical counterparts are given in Table 13. The RMSD in energies is 0.83 kcal/mol (vs. 0.96 kcal/mol for the OPLS-AA) and 18.2° for the key dihedrals, with the 19.3° result for the OPLS-AA. The gradient weighting was employed in the fitting, with the maximum to minimum weight ratio of 1000.0. To keep the last conformer from drifting away even more, an artificial barrier was added to the *ab initio* potential energy surface, and the χ_2 direction for that conformer received an additional weighting factor of 10.0. Still, the results for the first two conformers are worse than we expected, and additional research, similar to the one described above for the glutamine dipeptide, will be conducted at a later time.

Table 13. Histidine Dipeptide, Energy of the Conformers, RMS Deviations in ϕ, ψ, χ_1, χ_2 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	−0.14/4.5	−0.63/5.0
2	0.19	0.78/14.6	0.31/4.1
3	2.41	1.31/8.9	0.77/7.3
4	2.95	3.96/10.1	4.18/10.7
5	3.26	3.96/5.6	4.18/3.9
6	3.45	3.54/11.9	3.88/10.4
7	4.90	5.25/7.1	4.48/49.6
8	5.48	3.96/45.0	5.45/4.3
RMS error[c]	—	0.83/18.2	0.85/18.7

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Leucine, Isoleucine, and Valine

It would seem logical to fit a single set of torsional parameters for the three dipeptides with purely aliphatic side chains, leucine, isoleucine, and valine. However, we could not produce a single set that would allow uniformly good results for the all three of them, and separate ones are presented instead. Tables 14–16 show the conformational energies compared to the LMP2 results for the PFF force field and the OPLS-AA (for the benchmarking purpose), as well as RMS deviations of the key dihedral angles.

For the leucine dipeptide, the fitting was done with no gradient reweighting and the conformers 1 and 5 having a weight of 50.0 for points around them. Both isoleucine and valine calculations were carried out with the gradient fitting, the maximum/minimum weights ration equal to 1000.0. In addition, the conformers 1 and 6 of the isoleucine dipeptide were weighted heavier by a factor of 50.0.

Table 14. Leucine Dipeptide, Energy of the Conformers, RMS Deviations in ϕ, ψ, χ_1, χ_2 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L 1[b]	OPLS-AA/L 3[b]
1	0.00	−0.45/3.0	0.41/2.5	0.53/2.5
2	0.81	1.25/4.1	0.15/10.8	0.25/10.5
3	0.77	0.65/5.0	0.38/6.2	0.14/6.0
4	1.23	1.03/3.6	1.33/5.1	1.32/4.1
5	1.28	1.05/7.1	1.00/3.4	1.23/2.7
6	2.01	2.02/4.8	2.05/4.6	1.83/4.5
7	2.91	2.59/7.4	3.16/8.8	2.95/8.8
8	3.27	3.42/5.1	3.60/2.5	3.70/1.4
9	3.63	4.32/4.3	3.80/5.6	3.93/5.6
RMS error[c]	—	0.35/5.1	0.34/6.1	0.38/5.9

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Table 15. Isoleucine Dipeptide, Energy of the Conformers, RMS Deviations ϕ, ψ, χ_1, χ_2 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	0.56/29.7	0.26/4.9
2	0.69	1.18/5.5	0.58/5.7
3	0.88	1.18/4.2	0.75/2.8
4	1.00	0.67/9.3	0.40/8.8
5	1.11	0.11/3.9	0.80/3.0
6	1.80	1.54/4.8	2.19/6.6
7	2.18	1.69/5.3	2.84/6.0
8	3.49	4.21/6.0	3.32/3.6
RMS error[c]	—	0.88/11.8	0.38/5.5

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

All the results are rather good, demonstrating that the method performs adequately for alkane-type side-chains. The RMSD in conformational energies and the key dihedral angles was 0.35 kcal/mol and 5.1° for leucine, 0.57 kcal/mol and 11.8° for isoleucine, and 0.01 kcal/mol and 5.1° for valine. The accuracy is similar to the OPLS-AA case, for which the numbers are 0.34 kcal/mol and 6.1°, 0.38 kcal/mol and 5.5°, and 0.08 kcal/mol and 8.4°.

Methionine

For the methionine dipeptide, torsional fitting was carried out at the maximum/minimum ratio in the gradient weighting equal to 1000.0. The results are given in Table 17. The accuracy (0.53 kcal/mol for the energies and 5.4° for the key dihedral angles) is very good and close to the OPLS-AA results (0.59 kcal/mol and 5.2°).

Proline

Proline represents a special case. Because no side-chain rotation can happen, we did not refit any torsions. Instead, we took the

Table 16. Valine Dipeptide, Energy of the Conformers, RMS Deviations in ϕ, ψ, χ_1 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L 2[b]	OPLS-AA/L 3[b]
1	0.00	0.00/3.9	0.06/6.2	−0.20/6.5
2	0.35	0.36/2.1	0.24/3.2	0.36/3.3
3	0.69	0.67/7.6	0.74/12.8	0.87/12.9
RMS error[c]	—	0.01/5.1	0.08/8.4	0.16/8.6

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Table 17. Methionine Dipeptide, Energy of the Conformers, RMS Deviations ϕ, ψ, χ_1, χ_2, χ_3 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	−0.31/4.6	0.64/3.5
2	2.95	3.46/3.6	2.26/5.1
3	2.49	2.47/5.8	2.47/3.9
4	1.88	0.81/9.3	1.28/3.9
5	3.06	3.07/4.1	2.64/4.3
6	2.07	2.35/2.8	2.17/3.1
7	3.56	4.17/5.0	4.55/5.3
RMS error[c]	—	0.53/5.4	0.59/5.2

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

alanine parameters derived above (the same we used for the backbones in all the other cases) and did proline dipeptide geometry optimizations. After that we performed energy minimizations with the N-CT1-C(O)-N dihedral angle constrained at +60°, −60°, and 180° from the position of the minimum. These minimizations were done with both the high-level *ab initio* and molecular mechanics methods. The resultant energies are reported in Table 18. Their magnitudes (as well as the magnitudes of the errors) are, of course, greater than in the other cases, as we are dealing not with actual minima, but with rather constrained restrained ones. The RMS error in energy is 1.27 kcal/mol for the PFF, compared with the 1.54 kcal/mol for the refitted OPLS-AA.

Tryptophan

The tryptophan dipeptide side-chain torsional fitting was run with the gradient weighting employed, the maximum/minimum weight ratio of 1000.0. The results are given in Table 19. It can be seen that both the presented PFF and the OPLS-AA produce good conformational energy accuracy (0.49 and 0.50 kcal/mol, respectively), while the geometric one is 19.4° and 24.2°, worse than in the most cases, because of a couple of "runaway" minima, which

Table 18. Proline Dipeptide, Energy of the Rotamers, kcal/mol.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
Minimum	0.00	1.72	1.78
+60°[c]	3.18	3.65	2.86
−60°[c]	2.99	2.56	3.87
+180°[c]	12.45	10.69	10.12
RMS error[d]	—	1.27	1.54

[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Along the N—C—C(O)—N, constrained minimizations.
[d]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Table 19. Tryptophan Dipeptide, Energy of the Conformers, RMS Deviations ϕ, ψ, χ_1, χ_2 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	0.19/6.9	−0.01/7.2
2	0.15	0.56/14.2	0.38/6.5
3	1.30	1.68/16.4	2.16/11.6
4	1.65	2.01/2.1	1.72/4.7
5	2.18	0.94/38.4	2.56/9.8
6	2.22	2.43/6.2	2.05/5.2
7	3.26	2.94/34.9	2.19/49.0
8	2.91	2.94/11.0	2.56/48.2
9	3.41	3.39/3.5	3.48/13.7
RMS error[c]	—	0.49/19.4	0.50/24.2

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

did not fall into the same vicinities in the conformational space as their *ab initio* counterparts. However, no new artificial minima were introduced. Optimizations starting with the conformer number 7 ended up near the conformer number 8. As to the conformer 5, it finally converged near a minimum, which we had found previously to be present on the tryptophan conformational surface together with the nine others.[1] And the OPLS-AA situation for the "runaway" conformers is very similar.

Threonine

The threonine dipeptide torsional parameters were fitted with the maximum/minimum weight ratio in the gradient-based scheme equal to 1000.0. The results are given in Table 20. The accuracy of the PFF results (0.75 kcal/mol and 8.9°) is close to that of the OPLS-AA ones (0.87 kcal/mol and 6.9°) and is rather satisfactory in the view of the accuracy of the *ab initio* calculations.

Table 20. Threonine Dipeptide, Energy of the Conformers, RMS Deviations in ϕ, ψ, χ_1, χ_2 from the *Ab Initio* Data.

Conformer	*Ab Initio*[a]	PFF	OPLS-AA/L 1[b]	OPLS-AA/L 2[b]
1	0.00	0.77/4.3	−0.22/8.1	0.20/8.2
2	2.81	3.11/7.6	2.46/3.1	2.48/3.5
3	3.72	3.20/13.9	2.05/2.7	2.00/3.3
4	5.25	6.14/8.5	6.64/11.5	6.26/11.5
5	5.45	5.88/8.9	5.69/4.0	5.82/4.2
6	5.99	5.11/7.6	6.03/6.4	5.49/6.1
7	7.52	6.61/8.9	8.10/7.7	8.60/8.7
RMS error[c]	—	0.75/8.9	0.87/6.9	0.87/7.1

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Table 21. Tyrosine Dipeptide, Energy of the Conformers, RMS Deviations ϕ, ψ, χ_1, χ_2, χ_6 from the *Ab Initio* Data.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	−0.07/4.8	−0.09/4.4
2	0.34	0.81/13.4	1.13/5.7
3	0.39	0.12/7.0	0.07/9.2
4	1.67	1.88/1.8	1.73/3.2
5	2.17	1.86/2.4	1.78/5.5
6	2.64	2.61/14.8	2.30/14.9
RMS error[c]	—	0.27/8.9	0.39/8.1

Energies in kcal/mol, angles in degrees.
[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Tyrosine

The fitting was done with the maximum/minimum gradient weighting ratio of 1000.0. The PFF accuracy in the conformational energies (0.27 kcal/mol RMSD) and in reproducing the quantum mechanical values of the key dihedrals (8.9°) is basically similar to the OPLS-AA case (0.39 kcal/mol and 8.1°). However, the relative energies of the three lowest conformers are reproduced significantly better by the PFF, as can be seen from Table 21.

Aspartic Acid

Aspartic acid dipeptide is the first one presented here that possesses a net electrostatic charge. This is why electrostatic interactions in the system are stronger, and thus magnitudes of the conformational energies and error are greater as well. This, of course, is also true for all the other charged residues.

Table 22 shows the results. Because we were performing gas-phase optimizations with all the key dihedral angles constrained at their positions obtained in the liquid-state runs (as described in the previous subsection), no deviations of the dihedral could exist, and none reported. The RMS deviation in conformational energies is 0.77 kcal/mol, while the result for the OPLS-AA is 0.16 kcal/mol.

Table 22. Aspartic Acid Dipeptide, Energies of the Restrained Conformers Compared with the *Ab Initio* Data, kcal/mol.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L, Ver. 1[b]	OPLS-AA/L, Ver. 2[b]
1	5.40	6.41	5.63	7.74
2	0.00	−0.84	−0.08	2.43
3	3.72	3.54	3.57	3.80
RMS error[c]	—	0.77	0.16	1.95

[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Table 23. Glutamic Acid Dipeptide, Energies of the Restrained Conformers Compared with the *Ab Initio* Data, kcal/mol.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	0.58	−1.28
2	7.89	8.67	7.89
3	3.68	4.23	3.19
4	14.09	13.27	13.62
5	7.20	4.93	6.05
6	12.79	11.47	12.60
7	10.95	13.45	14.55
RMS error[c]	—	1.47	1.53

[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

The latter seems to be much better, but it was achieved by introducing a rather high-magnitude Fourier coefficient.[1] No gradient-based reweighting was done in this case.

Glutamic Acid

Results for the glutamic acid dipeptide are presented in Table 23. One again, no gradient-based weight adjustment was made. The PFF and OPLS-AA allow similar accuracy of the results, 1.47 kcal/mol and 1.53 kcal/mol, respectively.

Lysine

This is another case when no weights were assigned to the points of fitting subspace. As can be seen from Table 24, the PFF allows energies of the conformers to fall within 0.59 kcal/mol RMS from the quantum mechanical data. The OPLS-AA result is 0.88 kca/mol, which is also a great result, given the charged nature of the dipeptide.

Table 24. Lysine Dipeptide, Energies of the Restrained Conformers Compared with the *Ab Initio* Data, kcal/mol.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	17.16	17.65	16.87
2	21.45	20.68	20.08
3	16.70	16.10	17.48
4	0.00	0.38	1.39
5	15.21	16.01	15.05
6	13.25	12.92	12.90
RMS error[c]	—	0.59	0.88

[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Table 25. Protonated Histidine Dipeptide, Energies of the Restrained Conformers Compared with the *Ab Initio* Data, kcal/mol.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	1.66	0.94
2	4.86	5.50	5.40
3	0.31	−1.08	0.56
4	7.20	7.13	6.92
5	4.48	3.78	5.03
6	4.67	4.54	2.68
RMS error[c]	—	0.97	0.97

[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Protonated Histidine

For this dipeptide, both OPLS-AA and PFF give exactly the same RMS deviation of the energies of the constrained conformers from the *ab initio* results—0.97 kca/mol. No gradient-based weights were assigned, and the complete results are shown in Table 25.

Arginine

It can be seen from Table 26, that the PFF conformational energy RMSD for the arginine dipeptide are in better agreement (RMS equals to 0.79 kcal/mol) with the quantum mechanical data than the OPLS-AA (1.15 kcal/mol). The torsional fitting was done using the gradient weighting scheme, with 1000.0 maximum weights ratio. In addition, energies of the fitting point at and around the first conformer were raised by 2.0 kcal/mol to prevent the conformer energy from being too low. Finally, the CT-CT-CT-N2 V_3 was manually changed to −2.0 kcal/mol to improve the accuracy of the final conformational energies.

Summary

Tables 27 and 28 demonstrate summaries of the conformational energies and dihedral energy errors for the neutral and charged dipeptides, respectively. Several observations can be made. First, with the average RMS deviations of only 0.41 and 0.92 kcal/mol for the neutral and charged residues, we can conclude that the goal of creating an adequate force field for our purposes has been achieved. Second, it has been achieved for both the presented polarizable force field and the OPLS-AA.[1] Therefore, the torsional fitting methodology can be applied with success to creating force fields with different methods of computing the nonbonded part of the energy . Finally, the average errors of the PFF are not very far from (and slightly better than) those of the OPLS-AA (0.47 and 0.94 kcal/mol). In fact, it is interesting to examine the individual RMS errors for the various amino acids; there is a remarkably close correspondence between the errors in the polarizable and fixed charge models for most of the amino acids on an individual basis. This suggests that the remainder of the error lies beyond the scope of torsional refitting and, therefore, one has to examine other

Table 26. Arginine Dipeptide, Energies of the Restrained Conformers Compared with the *Ab Initio* Data, kcal/mol.

Conformer	*Ab initio*[a]	PFF	OPLS-AA/L[b]
1	0.00	−0.88	−0.80
2	10.76	11.78	9.72
3	3.29	2.29	1.95
4	13.87	13.82	15.73
5	8.58	9.65	9.21
6	4.25	3.47	4.93
RMS error[c]	—	0.79	1.15

[a]LMP2 cc-pVTZ(-f)//HF6-31G**, ref. 1.
[b]Ref. 1.
[c]Positions of the minima shifted uniformly to achieve the lowest RMS deviation from the *ab initio* results.

Table 27. Summary of RMS Deviations of Energies from LMP2/cc-pVTZ(-f)//HF/6-31G** for Peptides, kcal/mol.

Peptide	PFF	OPLS-AA/L[a]	MMFF94[a]
Tetrapeptide			
Alanine	0.69	0.56	
Dipeptides			
Alanine	0.35	0.27	
Serine	0.35	0.44/0.34	0.97
Phenylalanine	0.02	0.15	0.21
Cysteine	0.27	0.35	1.21
Asparagine	0.02	0.16	2.25
Glutamine	0.92	0.96	1.00
Histidine	0.83	0.85	1.60
Leucine	0.35	0.34/0.38	1.27
Isoleucine	0.57	0.38	0.66
Valine	0.01	0.08/0.16	1.01
Methionine	0.53	0.59	1.05
Proline	1.27	1.54	
Tryptophan	0.49	0.50	0.83
Threonine	0.76	0.87	1.15
Tyrosine	0.27	0.39	0.28
Average[b]	0.43	0.47	1.04

[a]Ref. 1.
[b]Proline not included.

Table 28. Summary of RMS Deviations of Energies from LMP2/cc-pVTZ(-f)//HF/6-31G** for Charged Dipeptides, kcal/mol.

Peptide	PFF	OPLS-AA/L[a]
Aspartic acid	0.77	0.16/1.95
Glutamic acid	1.47	1.53
Lysine	0.59	0.88
Protonated histidine	0.97	0.97
Arginine	0.79	1.15
Average	0.92	0.94/1.29

[a]Ref. 1.

sources of error (e.g., the remainder of the valence part of the force field) to improve the results beyond the present level of accuracy.

It should be noted that we were able to keep the errors in hydrogen bonding energies (Table 4) and conformational energies (Tables 6–27) within the 0.5 kcal/mol target value (as they are compared with the high-level *ab initio* results). The question as to how large the errors are in the condensed phase will be addressed in our subsequent work on the second generation of the polarizable force field.

Applications of the Polarizable Force Field to Realistic Systems

Simulation Methods

Thirty-nine realistic protein structures from the Protein Data Bank (listed in Tables 29 and 30) were used as initial geometries for gas-phase energy minimizations and molecular dynamics runs. The calculations were performed with both the standard OPLS-AA, which is a fixed-charges force field, and the polarizable force field presented in this article. The minimizations were carried out with the conjugate gradient algorithm. The initial step size was set to 0.05, the maximum step size was 1.0. Maximum number of iterations for line search was 3, maximum number of cycles was 10,000. Criterion for convergence of the RMS gradient was set to 0.05. Criterion for convergence of the change in energy for each atom, averaged over the whole system, was set to 10^{-5} kcal/mol. Molecular dynamics runs were done with the NVT ensemble. All molecular dynamics runs used a timestep of 1 fs and had a length of 1 ps. Relaxation time for velocity scaling was 0.01 ps. The target temperature was 298 K, with the initial temperature of 10 K.

In the molecular dynamics simulations, the "electronic" degrees of freedom (the inducible dipole moments) were propagated using the extended Lagrangian method[6–9]—that is, assigned masses and integrated along with the spatial coordinates. The dynamics so generated is fictitious and functions only as a scheme to keep the electronic degrees of freedom close to the minimum-energy "Born-Oppenheimer" surface, without doing expensive iterative solves or matrix inversion. We used the following simple method for choosing the fictitious masses of the fluctuating dipole moments: given a single frequency ω, the mass was set to $1/\alpha^2\omega$.[2] In this way, if the coupling between different fictitious degrees of freedom is weak, all the fictitious degrees of freedom will be in resonance. Arguably this is beneficial because any leaks of energy from the real system will be quickly distributed throughout the entire fictitious system rather than building up a "hotspot," which could make the fictitious dynamics unstable. More importantly, if ω is chosen to be much larger than the frequencies of nuclear motion, then the fictitious degrees of freedom will be far from resonance with the nuclear degrees of freedom, little energy will be transferred from the "real" system to the fictitious system, and the electronic degrees of freedom will remain close to the minimum-energy surface as desired. In practice, the choice of $\omega = 1800$ ps^{-1} (9556 cm^{-1}) worked well: for all simulations, the temperature of the fictitious subsystem remained below ca. 5 K.

Table 29. Geometry RMSD, in Å, of Protein Structures Optimized with the Fixed-Chares (OPLS-AA) Force Field and the Presented Polarizable Force Field from PDB Geometries.

	No H atoms		Backbone only	
Molecule	OPLS	PFF	OPLS	PFF
155c	2.22	2.37	1.86	2.07
1bp2	1.93	1.90	1.53	1.55
1cc5	1.92	2.22	1.65	1.91
1crn	2.00	1.83	1.60	1.63
1ctf	2.26	2.18	1.76	1.65
1fdx	2.67	2.43	2.33	1.94
1fx1	2.46	2.21	1.71	1.47
1gcn	4.14	3.77	2.88	2.00
1gcr	1.53	1.54	1.07	1.11
1paz	1.73	1.71	0.96	1.04
1pcy	1.78	1.77	1.05	1.05
1pgx	4.10	4.02	2.43	2.15
1ppt	2.20	1.90	1.07	1.70
1r69	1.68	1.70	1.12	1.13
1rnt	2.32	2.30	1.27	1.30
1sn3	2.56	2.28	1.52	1.17
1ubq	2.08	1.97	1.35	1.16
2cdv	3.41	3.53	2.25	2.37
2fxb	2.32	2.01	1.87	1.61
2gn5	2.57	2.53	2.10	2.06
2lzm	1.65	1.81	1.21	1.37
2ovo	1.74	1.76	1.53	1.55
2prk	1.26	1.33	1.03	1.10
2rn2	1.92	1.54	1.34	1.09
2sns	2.09	2.27	1.37	1.53
2ssi	1.93	1.89	1.49	1.48
351c	1.64	1.77	1.20	1.38
3adk	1.65	1.62	1.26	1.28
3c2c	2.11	2.02	1.29	1.16
3fxc	3.66	3.92	2.67	3.11
3icb	2.24	1.82	1.90	1.47
3wrp	1.93	1.98	1.30	1.28
4fd1	2.51	2.48	1.82	1.76
4fin	2.89	2.50	2.20	1.79
4fxn	2.25	1.91	1.38	1.07
4pti	2.41	2.30	1.66	1.58
5cpv	1.97	1.97	1.11	1.29
5fin	3.83	2.87	3.32	2.40
7rxn	1.73	1.74	1.26	1.31
Average	2.29	2.20	1.66	1.57

Results

The results of the energy minimizations and molecular dynamics runs are shown in Tables 29 and 30, respectively. For each of the proteins, their Protein Data Bank (PDB) ID is given, along with the geometry RMS deviations of the systems from their native PDB structure, computed with the OPLS-AA[1] and the polarizable force field. These RMSD are shown for the whole molecules (with hydrogen atoms excluded), as well as for the backbones only.

One immediate conclusion from the presented data is that the PFF results demonstrate uniformly good quality, with the average

Table 30. Geometry RMSD, in Å, of Protein Structures after a Molecular Dynamics Run with the Fixed-Chares (OPLS-AA) Force Field and the Presented Polarizable Force Field from PDB Geometries.

	No H atoms		Backbone only	
Molecule	OPLS	PFF	OPLS	PFF
155c	2.95	2.70	2.67	2.46
1bp2	2.79	2.48	2.16	2.02
1cc5	2.71	2.49	2.23	2.14
1crn	2.83	2.29	2.28	1.93
1ctf	2.82	2.33	2.08	1.75
1fdx	3.44	2.94	2.90	2.48
1fx1	3.21	2.89	2.48	2.13
1gcn	4.30	4.21	2.58	2.33
1gcr	2.37	1.88	1.82	1.40
1paz	2.70	2.09	2.13	1.52
1pcy	2.62	2.27	1.96	1.65
1pgx	4.21	4.05	2.00	1.97
1ppt	2.90	2.82	2.38	2.32
1r69	2.45	2.27	1.79	1.40
1rnt	3.35	2.90	2.50	2.07
1sn3	2.80	2.73	1.73	1.77
1ubq	2.72	2.38	1.90	1.59
2cdv	3.60	3.50	2.29	2.12
2fxb	3.55	3.21	3.04	2.73
2gn5	3.15	2.89	2.47	2.25
2lzm	2.66	2.34	2.28	1.93
2ovo	2.56	2.30	1.93	1.82
2prk	2.13	1.91	1.82	1.64
2rn2	2.58	2.17	1.81	1.70
2sns	2.89	2.73	2.01	1.79
2ssi	2.39	2.31	2.02	1.90
351c	2.47	2.14	2.01	1.69
3adk	2.35	2.19	1.89	1.75
3c2c	2.83	2.48	2.06	1.63
3fxc	4.20	3.97	2.99	2.74
3icb	3.17	2.50	2.73	2.19
3wrp	3.20	2.69	2.54	2.02
4fd1	3.56	3.17	2.85	2.49
4fin	3.15	2.99	2.53	2.35
4fxn	2.90	2.64	2.17	1.96
4pti	2.98	2.70	2.32	2.04
5cpv	2.83	2.60	2.15	1.93
5fin	3.20	2.83	2.72	2.32
7rxn	2.75	2.72	2.20	2.17
Average	2.98	2.68	2.27	2.00

geometry RMSD from the native forms of 2.20 Å and 2.68 Å for the minimizations and the molecular dynamics calculations, respectively. This result is actually slightly better than the OPLS-AA (2.29 and 2.98 Å). At present, we do not assign any significance to these differences, given their small magnitudes, the fact that the simulations are in the gas phase, and short duration of the dynamics trajectories.

It should be understood that the purpose of these gas phase simulations is simply as follows. The native structure forms a reasonable local minimum in the gas phase as does OPLS-AA. *A priori*, there is no reason to expect that the native structure is a global minimum in the gas phase in nature; indeed, it almost certainly is not. However, the solvation forces (when viewed as a potential of mean force, a rigorous theoretical approach) tend to be rather long range and slowly varying (although they can of course be quite large when comparing two qualitatively different structures) so that one expects the native structure to be very close to a local minimum even in the gas phase. The new methodology is no worse than the standard OPLS-AA in reproducing gas phase potential energy minima of the systems and the above calculations can be performed without the polarization catastrophe arising. From this point of view, the results obtained indicate that the model behaves in a reasonable fashion. An assessment of the quantitative accuracy of the PFF, compared to OPLS-AA or any other protein force field, will require much longer simulations in a solvated environment.

The computational efficiency of our minimization and molecular dynamics protocols has not yet been highly optimized, as our principle goal at this point is to assess accuracy and provide proof of concept with regard to the robustness and stability of the model. Protein minimizations on average require ca. 30 time more CPU time than corresponding gas phase calculations with fixed charges, while molecular dynamics simulations require ca. 20 times more CPU time. For both types of calculations, the use of permanent and polarizable dipoles leads to the long range electrostatic interaction of a substantially larger number of sites than are present in a fixed, point charge model (the benefit is of course greater accuracy). The additional computational effort associated with these extra sites (approximately one order of magnitude) can be reduced by a variety of techniques including multipoles, multiple time scale methods, and treatment of only a limited region of the protein at this level of detail (e.g., if one was studying protein-ligand binding, localized around the active site). The molecular dynamics simuilations avoid an iterated solve for the polarization vectors at each time step by the use of the extended Lagrangian formalism; however, no analogous protocol has been implemented for minimizations, thus explaining why the latter has a larger ratio of CPU time with equivalent fixed charge calculations than the former. All of these issues can be addressed in relatively straightforward fashion, and we defer a realistic assessment of the performance of the methodology until these strategies have been implemented.

Conclusions

We have developed a polarizable force field for protein modeling based on fitting parameters of a linear response model to an extensive set of high quality *ab initio* quantum chemical data, and have tested the performance of this force field in the gas phase for both dipeptides and for entire proteins. Gas-phase many-body effects for the dipeptides were captured within the average RMSD of 0.22 kcal/mol from their *ab initio* values. A previously developed highly efficient torsional fitting technique allowed conformational energies for the dipeptides and alanine tetrapeptide to be reproduced within the average RMSD of 0.43 kcal/mol from their quantum mechanical counterparts. The behavior of the force field for protein simulations was examined via minimization and short molecular dynamics simulations for 39 proteins from the PDB. Geometry deviations from the native protein structure, as com-

puted with the new model, were slightly lower than those given by the standard OPLS-AA force field (2.20 and 2.68 Å for the PFF minimizations and the molecular dynamics calculations, respectively, vs. the 2.29 and 2.98 Å for the OPLS-AA).

A deficiency of the present development protocol is that it is not based on a parametrization scheme that has been shown to yield accurate liquid state thermodynamic properties. In particular, the values of the Lennard–Jones *B* coefficients (dispersive tails) for the atom–atom pair potentials could not be adjusted separately, and hence, were not optimize to reproduce liquid state properties. As was briefly discussed above, independent optimization of these terms requires a modification of the functional form of the atom–atom pair potential to incorporate a greater degree of functional flexibility. Work on a second generation force field along these lines is ongoing, and will address the major uncertainty in the present effort. However, the descriptions of valence energetics, electrostatics, and hydrogen bonding in the second generation force field will be very similar to that in the present first generation model; thus, our expectation is that the results reported here provide a good approximation to what will be obtained from the second generation force field for structures and energies controlled by these terms. For example, the contribution of the dispersion terms to the dipeptide relative energetics is modest, so it would be surprising if the RMS errors in fitting conformational energies in the second generation model were very different than what is reported above.

To make meaningful comparisons with condensed phase experimental data, a solvation model to complement the force field is required. Our intention is to develop both explicit and implicit water descriptions and to carry out simulations using these models. Deviations from crystallographic coordinates of such simulations should be considerably smaller than those for existing fixed charge force fields if the inclusion of polarizability as described above has really allowed a substantial advance in the overall accuracy of the potential energy surface. Similarly, improvements in energetics must be assessed by looking at mutational and binding free energy experimental studies. The present work, while not achieving these goals, is nevertheless in our view a useful first step. Our model can be further improved by employing both fluctuating charges and inducible dipoles, but our quest here was to develop an accurate minimalist model.

Finally, a major objective of the present project is to develop and validate a method that is not only applicable to proteins but to arbitrary organic molecules. There is nothing in the above protocol that restricts our approach in this regard. At present, our philosophy is to regenerate electrostatic parameters for each new molecule, thus avoiding the problem of constructing transferable permanent and fixed charge parameters. Further tests will be required to ascertain whether such recomputation provides better accuracy due to a more reliable treatment of inductive effects; if this is not the case, it will be possible to build up a database of transferable parameters from small molecule calculations. Although we have some preliminary results suggesting that inductive effects are important (e.g., one obtains rather different charges for the amide group in a dipeptide than in small molecule analogues such as formamide), further exploration is required to reliably answer this question.

Acknowledgments

We thank Dr. Mike Beachy for *ab initio* data.

Supporting Information Available

Values of all the peptide parameters used in this work.

References

1. Kaminski, G. A.; Friesner, R. A.; Tirado–Rives, J.; Jorgensen, W. L. J Phys Chem B 2001, 105, 6474.
2. See for example: (a) Ramon, J. M. H.; Rios, M. A. Chem Phys 1999, 250, 155; (b) Gonzalez, M. A.; Enciso, E.; Bermejo, F. J.; Bee, M. J Chem Phys 1999, 110, 8045; (c) Soetens, J. C.; Jansen, G.; Millot, C. Mol Phys 1999, 96, 1003; (d) Dang, L. X. J Chem Phys 2000, 113, 266; (e) Ribeiro, M. C. C. Phys Rev B 2001, 6309, 4205; (f) Cieplak, P.; Caldwell, J.; Kollman, P. J Comp Chem 2001, 22, 1048.
3. For representative publications see: (a) Liu, Y. P.; Kim, K.; Berne, B. J.; Friesner, R. A.; Rick, S. W. J Chem Phys 1998, 108, 4739; (b) Chen, B.; Xing, J. H.; Siepmann, J. I. J Phys Chem B 2000, 104, 2391; (c) Jedlovszky, P.; Vallauri, R. J Chem Phys 2001, 115, 3750.
4. (a) Kolafa, J.; Ratner, M. Mol Sim 1998, 21, 1; (b) Kaminski, G. A.; Jorgensen, W. L. J Chem Soc Perkins Trans 2, 1999, 11, 2365.
5. (a) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. J Chem Phys 1999, 110, 741; (b) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. J Phys Chem B 1999, 103, 4730.
6. Car, R.; Parinello, M. Phys Rev Lett 1985, 55, 2471.
7. Van Belle, D.; Froeyen, M.; Lippens, G.; Wodak, S. J. Mol Phys 1992, 77, 239.
8. Spirk, M. J Phys Chem 1991, 95, 2283.
9. Rick, S. W.; Stuart, S. J.; Berne, B. J. J Chem Phys 1994, 101, 6141.
10. Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. J Chem Phys 2001, 115, 2237.
11. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. J Phys Chem 1987, 91, 6269.
12. Becke, A. D. Phys Rev A 1988, 38, 3098.
13. Lee, C.; Yang, W.; Parr, R. G. Phys Rev B 1988, 37, 785.
14. Jaguar v3.5, Schrödinger, Inc. Portland, OR, 1998.
15. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A., in preparation.
16. Tsuzuki, S.; Uchimaru, T.; Matsumara, K.; Mikami, M.; Tanabe, K. J Chem Phys 1999, 110, 11906.
17. Saebo, S.; Pulay, P. Annu Rev Phys Chem 1993, 44, 213.
18. Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. J Chem Phys 1995, 103, 1481.
19. Dunning, T. H. J Chem Phys 1989, 90, 1007.
20. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. J Am Chem Soc 1996, 113, 11225.
21. Beachy, M. D.; Chasman, D.; Murphy, R .B.; Halgren, T. A.; Friesner, R. A. J Am Chem Soc 1997, 119, 5908.