A New Force Field (ECEPP-05) for Peptides, Proteins, and Organic Molecules

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

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

Received: September 3, 2005; In Final Form: January 18, 2006

Parametrization and testing of a new all-atom force field for organic molecules and peptides with fixed bond lengths and bond angles are described. The van der Waals parameters for both the organic molecules and the peptides were taken from *J. Phys. Chem. B* **2003**, *107*, 7143 and *J. Phys. Chem. B* **2004**, *108*, 12181. First, the values of the 1−4 nonbonded and electrostatic scale factors appropriate to the new force field were determined by computing the conformational energies of six model molecules, namely, ethanol, ethylamine, propanol, propylamine, 1,2-ethanediol, and 1,3-propanediol with different values of these factors. The partial atomic charges of these molecules were obtained by fitting to the electrostatic potentials calculated with the HF/6-31G* quantum-mechanical method. Two different charge models (single- and multiple-conformation-derived) were also considered. We demonstrated that the charge model has a stronger effect on the conformational energies than the 1−4 scaling. The choice of a charge model affected the conformational energies of even the smallest molecules considered, whereas the effect of the 1−4 electrostatic or nonbonded scaling was apparent only for 1,3-propanediol. The best agreement with high-level ab initio data was obtained with the multiple-conformation-derived charges and with no scaling of the 1−4 nonbonded or electrostatic interactions (scale factors of 1.0). Next, the torsional parameters of a large number of neutral and charged organic molecules, assumed to be models of the side chains of the 20 naturally occurring amino acids, were computed by fitting to rotational energy profiles obtained from ab initio MP2/6-31G** calculations. The quality of the fits was high with average errors for torsional profiles of less than 0.2 kcal/mol. To derive the torsional parameters for the peptide backbone, the partial atomic charges of the 20 neutral and charged amino acids were obtained by fitting to the electrostatic potentials of terminally blocked amino acids using the HF/6-31G* quantum-mechanical method. Then, the $\phi-\psi$ energy maps of Ac-Ala-NMe and Ac-Gly-NMe were computed using MP2/6-31G**//HF/6-31G** quantum-mechanical methods. The $\phi-\psi$ energy map of Ac-Ala-NMe was used for refinement of the nonbonded parameters for the backbone nitrogen and hydrogen bonded to it. Subsequently, the main-chain torsional parameters were obtained by fitting the molecular mechanics energies to the $\phi-\psi$ energy maps of Ac-Ala-NMe and Ac-Gly-NMe. The transferability of the entire force field was demonstrated by reproducing the main energy minima of terminally blocked Ala_3 from the literature. The performance of the force field was also evaluated by simulating crystal structures of small peptides. By comparison of simulated and experimental data, examination of the torsional-angle and atom-positional root-mean-square deviations of the energy-minimized crystal structures from the corresponding X-ray model structures demonstrated high accuracy of the force field.

Introduction

Empirical all-atom force fields play an important role in theoretical studies of protein structure and of the thermodynamics and kinetics of protein reactions as well as of other macromolecules of biological interest.[1,2] Several force fields for proteins, including ECEPP/3,[3] AMBER,[4,5] CHARMM22,[6] GROMOS,[7] CVFF,[8] OPLS,[9] and others, are available in the literature. They have been used successfully for many applications including protein folding[10−16] and distinguishing native from non-native folds.[17−19] Although these force fields perform well for a wide variety of proteins, there are still limitations in their accuracy. For example, these limitations are seen in attempts to reproduce quantum-mechanical and experimental data for small peptides, such as terminally blocked alanine and alanine oligopeptides.[20,21] For larger systems, generalized-ensemble simulations[22] carried out for 13-residue α-helical and 16-residue β-hairpin peptides highlighted differences in the secondary-structure preferences of six popular force fields and led to the conclusion that none of the tested force fields can be used successfully for all types of folds (α, β, α/β, etc.).

The sources of the limitations of the existing force fields include the form of the potential functions commonly used in biomolecular simulations, the components of the force field (i.e., model systems and experimental and/or quantum-mechanical data) used in its parametrization, and the algorithms used for derivation of the parameters. The functional form of the potential energy is considered[7,23] to be a major problem with all-atom protein force fields. It has long been recognized that fixed partial atomic charges[23,24] cannot provide an adequate description of electrostatic interactions in both the gas and the condensed phases as well as at different stages of protein folding. Polarizable force fields that may represent an alternative to the fixed atomic-charge models are still too expensive computationally and cannot be applied to macromolecules of biological interest that contain thousands of atoms. Therefore, the attempts to improve empirical force fields have been focused mainly on better

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

[†] Current address: Georgia Institute of Technology, School of Biology, 250 14th St., Atlanta, GA 30332.

10.1021/jp054994x CCC: $33.50
Published on Web 02/17/2006

parametrization of the existing potential functions. The results of new parametrizations reported for several force fields[5−9] demonstrated improved accuracy for a variety of systems from terminally blocked amino acids to proteins and confirmed that this is still an important and useful direction to pursue.

The development of an all-atom force field has been one of the main directions of research carried out in our laboratory. The ECEPP algorithm (empirical conformational energy program for peptides), developed in this laboratory,[3,25−27] has been used for conformational energy computations of peptides and proteins, allowing only their torsional angles to be varied. The latest version of the force field (ECEPP/3)[3] has been applied to a variety of problems.[15,16,19,28] For example, the 46-residue protein A (a three-helix bundle)[15] and the 36-residue villin headpiece[16] were folded successfully using the all-atom ECEPP/3 energy function together with a continuum model of hydration and the electrostatically driven Monte Carlo (EDMC) method[29] as a search technique. It was also shown[19] that ECEPP/3 can be used as a scoring function to discriminate native from nonnative protein conformations when combined with an accurate solvation model.[19] Despite the good results obtained for some (mostly helical) proteins, the accuracy of the current force field is not high enough for modeling β and α/β proteins.

Since the parameters for the nonbonded, electrostatic, hydrogen-bonding, and torsional parts of the force field in the various versions of ECEPP[3,25−27] were derived, using experimental data, the amount of available experimental information has increased significantly. Also, the development of modern quantum-mechanical methods combined with greatly increased computer power can now provide a wealth of information for systems as large as polypeptides. All of these reasons motivated us to develop a new all-atom gas-phase force field applicable to isolated organic molecules and peptides and crystal structures thereof, utilizing ECEPP geometry (i.e., fixed bond lengths and bond angles). It is planned to augment this new all-atom force field with a hydration model to treat aqueous solutions of proteins.

The nonbonded and electrostatic parameters of the new force field (designated here as ECEPP-05) were developed previously.[30,31] For the functional form of the new force field, we chose a more accurate (but still a simple enough one to be used in global-search calculations) Buckingham (6-exp) potential for describing nonbonded interactions between all atom types including hydrogens involved in hydrogen bonds (i.e., no special potential function is used for hydrogen bonds).

It is reasonable to believe that some intermolecular interactions in crystals are essentially the same as the corresponding interactions in more flexible molecular aggregates (for example, the nonbonded interactions within a protein molecule). Therefore, experimental crystal data have been used not only for development of force fields for crystal calculations[32−39] but also for deriving van der Waals parameters of general all-atom potentials[25,40,41] for a wide range of molecules including proteins. In deriving the nonbonded parameters, we used both experimental data from organic crystals and quantum-mechanical (QM) calculations for dimers of organic molecules. We also employed our new global-optimization-based method to optimize the parameters.[42] The method uses information about the global shape of the potential and forces the condition that the experimental structure should correspond to the global minimum of the optimized force field. The potential was shown to predict crystal structures of a large set of organic molecules successfully, including aliphatic and aromatic hydrocarbons, alcohols, amines, imidazoles, amides, and carboxylic acids.[30,31] Although the nonbonded and electrostatic parameters were derived from crystal data and QM calculations, respectively, and tested on crystal data, they are also intended for ultimate use in protein structure prediction and refinement of low-resolution protein models. To facilitate transferability to proteins, the parameter set includes all types of atoms that occur in proteins.

Since the nonbonded and electrostatic parts of the force field were derived assuming rigid internal molecular geometry (i.e., fixed bond lengths and bond and dihedral angles), they have to be supplemented with the torsional potential to make the force field applicable to flexible molecules. Following the ECEPP methodology, we model molecular flexibility by allowing the torsional angles to vary while keeping bond lengths and bond angles fixed. In this paper, we report further development of a new all-atom force field including: (1) derivation of the parameters of torsional potentials of amino acid side chains with quantum-mechanical torsional profiles of small organic molecules used as target data for fitting; (2) derivation of peptide backbone torsional parameters using ab initio $\phi-\psi$ energy maps for terminally blocked alanine and glycine; and (3) computation of atomic point charges for all 20 naturally occurring amino acids by fitting ab initio electrostatic potentials of the corresponding terminally blocked amino acids. Since we are interested in ultimately obtaining an all-atom force field for proteins, the compounds used for deriving the torsional parameters include molecules with torsional degrees of freedom representing those of the amino amino acid backbone and side chains. All ab initio calculations correspond to the gas phase. In future work, we plan to test the performance of the force field with solvation included.

We also present the results of a series of tests carried out to assess transferability and the general performance of the force field. Transferability has been tested on a set of small flexible molecules with their calculated rotational barriers compared to quantum-mechanical data. The relative conformational energies computed for 10 conformations of terminally blocked Ala_3, using the new force field, were compared to ab initio results reported by Beachy et al.[20] Local energy minimization of experimental crystal structures of oligopeptides containing up to four amino acids was carried out to assess the performance of the entire force field including the torsional parameters and atomic partial charges.

Methods

1. Form of the Potential. Since we use the ECEPP approach to treat molecular flexibility (with fixed bond lengths and bond angles), the intramolecular energy is a function of torsional angles. The total energy of a molecule, E_{intra}, consists of nonbonded and electrostatic energies, $E_{\text{intra}}^{\text{nbe}}$, and torsional, E^{tor}, terms

$$E_{\text{intra}} = E_{\text{intra}}^{\text{nbe}} + E^{\text{tor}} \tag{1}$$

The nonbonded and electrostatic parts are computed as a sum of the Buckingham potential (instead of the Lennard-Jones potential used in the ECEPP/3 force field[3]) and the Coulomb contribution for pairwise interactions

$$E_{\text{intra}}^{\text{nbe}} = \sum_{ij(j>i)} [-A_{ij} r_{ij}^{-6} + (1/k_{14}) B_{ij} \exp(-C_{ij} r_{ij})] + \sum_{ij(j>i)} \frac{332 q_i q_j}{k_{14}^{\text{el}} \epsilon r_{ij}} \tag{2}$$

where r_{ij} is the distance between atoms i and j separated by at least three bonds; A_{ij}, B_{ij}, and C_{ij} are nonbonded parameters; q_i

TABLE 1: Nonbonded Parameters[a]

atom type	description	A (Å^6 kcal/mol)	B (kcal/mol)	C (Å^{-1})
HC	hydrogen bonded to aliphatic carbon	13.89	1567.7	3.81
HA	hydrogen bonded to aromatic carbon	41.75	3392.1	3.88
HO1	hydrogen in alcohol group	3.25	137.8	3.69
HO2	hydrogen in carboxyl group	13.43	360.2	3.73
HN1	hydrogen bonded to amine or sp^2 nitrogen	4.56	2376.8	4.75
HN2	hydrogen in primary amide group	11.31	269.3	3.50
HNB[b]	hydrogen in amide group (backbone)	12.56	190.2	3.50
C1	aliphatic carbon	650.65	30869.1	3.16
CA	aromatic carbon	445.65	50360.5	3.39
C2	carbon in carboxyl group	661.02	39788.8	3.41
C3	carbon in amide group	728.45	67230.2	3.60
N1	aliphatic nitrogen	769.30	51218.6	3.48
NA	sp^2 nitrogen	475.30	25836.2	3.48
N2	primary amide nitrogen	434.80	56066.0	3.48
NB[b]	amide nitrogen with one bonded hydrogen (backbone)	549.83	39657.2	3.48
OH1	hydroxy oxygen	332.72	20231.9	3.46
OH2	carboxyl oxygen bonded to two atoms as in C−O−H	374.99	40979.5	3.54
O1	carboxyl oxygen bonded to one atom as in C═O group	309.44	51363.0	3.99
O2	amide oxygen	310.15	36409.9	3.80
S[b]	sulfur (sulfides)	2658.96	38382.2	2.90

[a] Parameters were taken from ref 30 and 31. [b] Parameters derived in this work.

and q_j are point charges (in e.u.) localized on atoms. An additional point charge with zero nonbonded parameters assigned is used to model the lone-pair electrons of sp^2 nitrogen in histidine (see ref 31 for details). The summation runs over all of the pairs of atoms $i < j$. k_{14} and k_{14}^{el} are scale factors for 1−4 nonbonded and electrostatic interactions, respectively. The dielectric constant ϵ was taken as unity.

The following combination rules for the nonbonded parameters A_{ij}, B_{ij}, and C_{ij} were applied

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

The nonbonded and electrostatic interactions described by eq 2 are included for 1−4 or higher-order atom pairs. The 1−4 interactions are treated in a special way by introducing k_{14} and k_{14}^{el} scaling factors. The role of these factors and the choice of their values are discussed in section 1 of the Results and Discussion. Interactions between atoms separated by more than three bonds are not scaled ($k_{14} = 1$ and $k_{14}^{el} = 1$).

Unlike the ECEPP/3 force field,[3] there is no explicit hydrogen-bonding term in the potential function. This interaction is represented by a combination of electrostatic and nonbonded interactions with a hydrogen involved in a hydrogen bond treated as a separate atom type with parameters different from those of the other types of hydrogen.

The torsional energy term for each dihedral angle θ was computed as follows

$$E^{tor} = k_\theta{}^1[1 + \cos(\theta)] + k_\theta{}^2[1 - \cos(2\theta)] + k_\theta{}^3[1 + \cos(3\theta)] \quad (3)$$

where θ is a torsional angle varying from 0° to 180° and $k_\theta{}^n$ are the torsional parameters.

The atom types included in the force field are listed in Table 1. The values of the corresponding nonbonded parameters (Table 1) were taken from our earlier work[30,31] whereas the torsional energy Fourier coefficients $k_\theta{}^n$ were derived in this work. We also refined the "6-exp" parameters for the backbone nitrogen and hydrogen (see the Results and Discussion section for details).

Ab initio optimized geometry was used for all the model organic compounds, and ECEPP geometry was used for the peptides.

2. Atomic Partial Charges. A set of small model molecules was used for deriving torsional parameters for the side chains of all 20 naturally occurring amino acids (both charged and neutral ones). The atomic charges, q_i (eq 2), were fitted to reproduce the molecular electrostatic potential, calculated with the Hartree−Fock wave function and the 6-31G* basis set using the GAMESS program.[43] The fitting was carried out using the restrained electrostatic potential (RESP) method,[44a] implemented in the AMBER 6.0 program.[44b] The molecular electrostatic potentials were computed for all minimum-energy conformations of a given molecule containing different values of the dihedral angles that terminate in heavy atoms (C, N, O, S) or hydrogens bonded to N, O, or S atoms. The RESP method was also used to obtain a single set of charges using several conformations of a given molecule (multiple-conformation fitting). Point charges derived by this procedure were employed for calculating the intramolecular electrostatic interactions (eq 2) within a single model molecule in the gas phase.

To obtain the charges for all 20 naturally occuring amino acids, we followed the methodology[45] utilized with the AMBER force field. Terminally blocked amino acids X, Ac-X-NMe, were used as model molecules for each of the 20 amino acids. The charges on all the atoms of the molecule were derived using multiple-conformation fitting[45] and RESP fits, with constraints ensuring a net charge of 0 or ±1 on an amino acid and zero net charge on each acetyl and *N*-methyl terminal group.

The following conformations of the terminally blocked amino acids were considered. For each amino acid, at least one side-chain conformation with the highest occurrence probability, as determined by an analysis (using the rotamer library of Lovell et al.[46]) of the high-quality protein data from the Protein Data Bank[47] (PDB), was selected. If there was more than one conformation with a similar high probability, then the two most probable rotamers were used. For each molecule, the selected side-chain conformations were combined with three conformations of the backbone (C_5, α_R, and α_L) corresponding to the most populated areas[48] of the empirical Ramachandran map. The ECEPP/3 standard bond lengths and bond angles[3] were used for building the model molecules. The ab initio molecular electrostatic potential was calculated at the HF/6-31G* level. First, the charges were derived separately for each amino acid, allowing the charges on the atoms of the backbone, side chain, and terminal blocking groups to vary; i.e., we did not require

TABLE 2: Backbone Partial Atomic Charges (in e.u.) of Standard Neutral and Charged Amino Acids in ECEPP-05 and ECEPP/3 Force Fields

amino acid	N	HN	C^{α}	H^{α}	C	O
Gly	−0.3319	0.2538	−0.2696	0.1369	0.6284	−0.5545
Ala	−0.4181	0.2584	0.1665	0.0122	0.5177	−0.5405
Val	−0.3521	0.2244	−0.0845	0.0703	0.5194	−0.4900
Ile	−0.2982	0.1867	−0.1887	0.0849	0.5242	−0.4812
Leu	−0.3266	0.2227	−0.0463	0.0822	0.5088	−0.5038
Phe	−0.3313	0.2247	−0.0669	0.0875	0.5503	−0.5089
Pro	−0.2184		−0.0512	0.0784	0.5447	−0.5344
Met	−0.3179	0.2151	0.0204	0.0576	0.2885	−0.4701
Asp	−0.3127	0.2352	−0.1387	0.1294	0.5745	−0.5407
Asp^{-}	−0.4382	0.2516	0.0760	0.0642	0.4367	−0.5506
Glu	−0.3214	0.2295	−0.0998	0.0731	0.5717	−0.5295
Glu^{-}	−0.3995	0.2354	−0.0902	0.0334	0.5127	−0.5347
Lys	−0.3407	0.2279	−0.0374	0.0613	0.5485	−0.5298
Lys^{+}	−0.2882	0.2245	−0.0690	0.0896	0.5884	−0.5242
Arg	−0.2673	0.2103	−0.1517	0.1104	0.5451	−0.5463
Arg^{+}	−0.2666	0.2204	−0.1184	0.1121	0.5903	−0.5271
Ser	−0.3146	0.2339	−0.1258	0.0992	0.5782	−0.5171
Thr	−0.3285	0.2369	−0.1243	0.1084	0.5659	−0.5165
Tyr	−0.3164	0.2249	−0.1226	0.0987	0.5699	−0.5111
Tyr^{-}	−0.3971	0.2365	−0.0444	0.0755	0.4996	−0.5058
His ϵ	−0.2746	0.1928	−0.0076	0.0957	0.4857	−0.5140
His δ	−0.3073	0.2279	−0.0133	0.0643	0.5352	−0.5421
His^{+}	−0.2585	0.2230	−0.0866	0.0810	0.6244	−0.5213
Cys	−0.2973	0.2425	−0.0555	0.1475	0.5404	−0.5262
Asn	−0.3407	0.2371	−0.0586	0.1036	0.5672	−0.5439
Gln	−0.2713	0.2098	−0.0843	0.0710	0.5363	−0.5410
Trp	−0.3164	0.2249	−0.1226	0.0987	0.5699	−0.5111
			ECEPP/3 [a]			
C^{β}-containing residues (excluding Asp^{-})	−0.3563	0.1762	0.0636	0.0202	0.4502	−0.3842
Gly	−0.3438	0.1762	−0.0085	0.0551	0.4502	−0.3842
Pro	−0.2965		0.0567	0.0210	0.4564	−0.3795
Asp^{-}	−0.3563	0.1762	−0.0598	0.0241	0.4502	−0.3842

[a] ECEPP/3 partial atomic charges were taken from ref 25, except those for Pro were taken from ref 3.

the backbone charges to be the same for all the amino acids. The partial atomic charges on the atoms of the backbones and side chains of the 20 naturally occurring amino acids are given in Tables 2 and 3, respectively. The ECEPP/3 backbone charges[3] are included in Table 2 for comparison. After all the amino acid charges were computed, the charges of the acetyl and *N*-methyl end groups were obtained by the following procedure: with the charges of each amino acid fixed at the values just determined, a uniform set of charges for the atoms of the end groups (with the net charge of each end group constrained to zero) was computed by simultaneous RESP fitting of the electrostatic potentials of all conformations (described above) of all 20 terminally blocked amino acids. In this way, a single set of charges for the atoms of the acetyl and *N*-methyl groups, which reflected the combined influence of all 20 amino acids, was obtained.

3. Derivation of Parameters of the Torsional Potential. Our derivation of the torsional potential-energy terms relied on fitting the molecular-mechanical (MM) energy profiles for rotation around a specific bond against the corresponding quantum-mechanical (QM) profiles. The corresponding torsional potential-energy terms were obtained by fitting a cosine series (eq 3) to the difference between the QM and MM profiles (the latter consisting of nonbonded and electrostatic terms).

A set of model molecules containing the same types of torsional angles as those present in the naturally occuring amino acids and in the protein backbone was used. The four atoms (defining each type of torsional angle) with their covalently bound neighbors replaced by hydrogen atoms defined the molecules selected for the calculations. Thus, the torsional terms were parametrized to reproduce the properties of the simplest molecules possible and then applied to larger and more complex ones.

The QM and MM profiles of the model molecules were calculated adiabatically, i.e., by constraining the appropriate torsions for each of the torsional angles on a 10° grid while minimizing the energy with respect to all the other degrees of freedom. All the ab initio calculations were carried out at the MP2[49] level of theory with a 6-31G** basis set implemented in the Gaussian 98 program,[50] which also includes a minimization routine. The corresponding MM torsional profiles were computed using the SUMSL minimizer.[51] The molecular geometries (bond lengths and bond angles) were optimized by QM calculations, and the lowest-energy QM conformations were used for calculating the MM torsional profiles. Due to rotation, some functional groups such as methyl, phenyl, and amino groups can have higher symmetry than the ones obtained from QM calculations on fixed rotamers of these groups; hence, the corresponding bond lengths and bond angles of these groups were averaged to conform to the highest symmetry possible for a particular group.

4. QM and MM $\phi-\psi$ Maps for Terminally Blocked Alanine, Glycine, and Proline. Three model systems, namely, terminally blocked alanine, glycine, and proline, were used for deriving the ϕ and ψ backbone torsional parameters. We considered terminally blocked alanine as a model for deriving backbone torsional parameters for all the amino acids except glycine and proline. QM $\phi-\psi$ maps were computed for terminally blocked alanine and glycine. In the case of proline, where the ϕ angle is fixed due to the ECEPP rigid geometry,

TABLE 3: Side-Chain Partial Atomic Charges (in e.u.) of Standard Neutral and Charged Amino Acids

amino acid	C^{β}	H^{β}	C^{γ} (O^{γ}, S^{γ})	H^{γ}	C^{δ} (S^{δ}, O^{δ},N^{δ})	H^{δ} (X^{a})	$C^{\gamma 2}$	$H^{\gamma 2}$	$C^{\delta 2}$ ($N^{\delta 2}$,$O^{\delta 2}$)	$H^{\delta 2}$
Ala	−0.1197	0.0386								
Val	0.3744	−0.0451	−0.2466	0.0461						
Ile	0.2292	0.0345	0.0464	0.0066	−0.2273	0.0497	−0.2403	0.0643		
Leu	−0.0464	0.0150	0.3995	−0.0647	−0.3403	0.0709				
Phe	−0.0045	0.0346	0.0317		−0.1381	0.1359			−0.1381	0.1359
Pro	−0.0258	0.0342	0.0190	0.0224	−0.0126	0.0435				
Met	0.2535	0.0317	−0.0535	0.0446	−0.2857					
Asp	0.0054	0.0507	0.7137		−0.5989				−0.6344	0.4660
Asp^{-}	0.0670	−0.0324	0.7745		−0.8082				−0.8082	
Glu	0.0222	0.0319	0.0080	0.0330	0.7130					
Glu^{-}	0.2957	−0.0481	−0.0640	−0.0097	0.7693					
Lys	0.0727	−0.0036	−0.0008	−0.0017	−0.0749	0.0240				
Lys^{+}	−0.0224	0.0306	−0.0120	0.0160	0.0135	0.0262				
Arg	0.0626	0.0295	0.0611	−0.0097	−0.0537	0.0665				
Arg^{+}	−0.0205	0.0396	0.0719	−0.0021	−0.0209	0.0827				
Ser	0.1806	0.0261	−0.5874	0.4007						
Thr	0.3194	−0.0053	−0.6153	0.3936			−0.1350	0.0335		
Tyr	0.0087	0.0332	0.0402		−0.1856	0.1663				0.1663
Tyr^{-}	0.0720	−0.0110	0.0042		−0.2674	0.1416				0.1416
His ϵ	−0.0225	0.0553	0.1840		−0.0051	−0.2584			−0.2150	−0.0435
His δ	−0.0484	0.0642	−0.0318		−0.2541	0.3424			−0.1232	0.1857
His^{+}	−0.1138	0.1069	0.0579		−0.0862	0.3686			−0.2392	0.2610
Cys	−0.1509	0.1189	−0.3297	0.1913						
Asn	−0.0809	0.0462	0.7291		−0.5637				−0.9688	0.4136
Gln	0.0293	0.0468	−0.0121	−0.0148	0.7302					
Trp	0.0087	0.0332	0.0402		−0.1856	0.1663			−0.1856	0.1663

amino acid	$C^{\epsilon 1}$ ($O^{\epsilon 1}$,$N^{\epsilon 1}$)	$H^{\epsilon 1}$	$C^{\epsilon 2}$ ($N^{\epsilon 2}$,$O^{\epsilon 2}$)	$H^{\epsilon 2}$ (X^{a})	C^{ζ} (N^{ζ})	H^{ζ}	O^{η} ($N^{\eta 1}$)	$H^{\eta 1}$	$N^{\eta 2}$	$H^{\eta 2}$
Phe	−0.1782	0.1365	−0.1782	0.1365	−0.0836	0.1198				
Met	−0.0563	0.0652								
Glu	−0.6229		−0.6477	0.4743						
Glu^{-}	−0.8212		−0.8212							
Lys	0.3779	−0.0152			−1.0348	0.3615				
Lys^{+}	0.0345	0.0933			−0.3685	0.3338				
Arg	−0.5252	0.2989			0.8725		−1.0618	0.4346	−1.0418	0.4399
Arg^{+}	−0.4663	0.3659			0.6914		−0.7495	0.4424	−0.8932	0.4426
Tyr	−0.2708	0.1741	−0.2708	0.1741	0.3483		−0.5352	0.3602		
Tyr^{-}	−0.2441	0.1079	−0.2441	0.1079	0.4647		−0.8593			
His ϵ	0.1336	0.1488	−0.3598	0.3494						
His δ	−0.0319	0.1854	0.0759	−0.3931						
His^{+}	−0.0655	0.2657	−0.0969	0.3729						
Gln	−0.6116		−0.9722	0.4238						
Trp	−0.2708	0.1741	−0.2708	0.1741	0.3483		−0.5352	0.3602		

[a] X denotes the charge added to model lone-pair electrons of nitrogen in histidine. The additional charge is located 0.4 Å from the corresponding nitrogen, with quasi-trigonal symmetry.[31]

we computed one-dimensional sets of the QM energies corresponding to the different values of the ψ angle. Both trans and cis conformations of $\omega°$ ($\omega°$ pertains to the Ac-Pro peptide group) and "down" puckering of the pyrrolidine ring were considered. Two, i.e., trans and cis, conformations of Ac-Pro-NMe with "up" puckering of the ring were used for testing the resulting torsional parameters. All quantum-mechanical calculations were carried out using Gaussian 98 software.[50] The QM Ramachandran maps were calculated in two steps. First, all conformations generated in two-dimensional $\phi-\psi$ space on a 10° grid were geometry-optimized at the Hartree−Fock level with the 6-31G** basis set, with the ϕ and ψ angles constrained. Next, single-point energy calculations were carried out for each of the optimized geometries using the more accurate MP2 method with the 6-31G** basis set. The resulting $\phi-\psi$ energy maps were compared to the corresponding maps obtained with our new MM force field. The MM energy maps were computed using ECEPP geometries and minimizing the energy of each conformation with the main-chain torsional angles constrained at the designated values and all other (i.e., ω, side-chain, and end-group) torsionals allowed to vary.

For terminally blocked alanine, we considered only the $\phi-\psi$ region of $-180° < \phi < -70°$; $-40° < \psi < 180°$ plus α_L and α_D ($\phi \approx 53°$; $\psi \approx -133°$) minima. The energies of the other regions are higher and cannot be reproduced using rigid ECEPP geometry. For terminally blocked glycine, which does not exhibit such high-energy regions, the entire $\phi-\psi$ map was considered. There are two main conformations of Ac-Pro-NMe. For "down" puckering of the pyrrolidine ring, they correspond to the trans conformation with $\omega° = -172.8°$; $\psi = 75.9°$ and the cis conformation with $\omega° = 7.9°$; $\psi = -23.9°$. The ψ torsional parameters for proline were derived considering only the vicinity (±30° range of ψ) of each of the two minima. The QM energies of the trans and cis conformations with "up" puckering of the ring were used for testing the resulting parameters. Standard ECEPP geometry with "up" and "down" puckering of the pyrrolidine ring was used for MM calculations of the blocked proline.

5. Refinement of the Main-Chain Torsional and Nonbonded Parameters, Using QM $\phi-\psi$ Maps. Refinement of the MM parameters was carried out by minimizing the target function

$$F(k_\theta^n;A;B) = \sum_{i=1}^{N} w_i(\Delta E_i^{\mathrm{MM}} - \Delta E_i^{\mathrm{QM}})^2 \qquad (4)$$

with respect to the k_θ^n coefficients of the Fourier expansion (eq 3) and the A and B parameters of eq 2. The summation runs over all N points of the $\phi-\psi$ map taken into consideration.

ΔE_i^{MM} and ΔE_i^{QM} are the relative MM and QM energies, respectively, for a given point i; w_i are empirical weights. The weights were computed according to the formula $w_i = \exp(-\Delta E_i^{QM} c)$, where c is an empirical parameter to provide additional de-emphasis of high-energy regions. The value of c was chosen so as to give higher weights to those of the fitting points located at or near the energy minima.

The Metropolis Monte Carlo method was applied to minimize the target function $F(k_\theta^n;A;B)$, with k_θ^n varying in the range of ± 2 kcal/mol and A and B (eq 2) within $\pm 30\%$ of their initial values. These ranges yielded the best agreement between the QM and MM maps.

To determine whether better results could be achieved by optimizing some of the 6-exp parameters, additional runs, including optimization of A and B of the 6-exp potential, using the subsequently refined torsional parameters, were carried out. The reason for carrying out these additional calculations was that the 6-exp parameters for the amide nitrogen and hydrogen were computed for a set of primary amides, whereas the backbone nitrogen of a polypeptide has only one hydrogen attached and, therefore, may have slightly different properties. (Some force fields, e.g., W99,[39] distinguish between nitrogens with two, one or no bonded hydrogens.)

Parameter optimizations were carried out for three systems, namely, terminally blocked alanine, glycine, and proline.

A total of 221 conformers of terminally blocked alanine corresponding to $-180° < \phi < -70°$; $-40° < \psi < 180°$ were used for the parameter refinement. The range was augmented by inclusion of the α_L and α_D conformations. Since the rest of the map has high energies, it was not considered for parameter refinement. Because the energies in the region around the C_7^{ax} minimum were shown[52] to be very high with the use of ECEPP geometry, i.e., fixed bond lengths and bond angles, it is not possible to reproduce the QM values. The whole $\phi-\psi$ map (1145 conformations) was studied in the case of terminally blocked glycine. In the case of proline, 14 conformers generated by starting from the lowest-energy trans ($\omega° = -172.8°$; $\psi = 75.9°$) and cis ($\omega° = 7.9°$; $\psi = -23.9°$) conformations (with "down" puckering of the pyrrolidine ring) by varying the ψ angle in 10° steps were used for fitting the parameters.

The root-mean-square deviation (RMSD) between the QM and the MM energies over all the points of the maps, as well as the relative energies of the minima, were used as criteria for assessing the accuracy of the force field.

6. Crystal Calculations. The accuracy of the new force field was tested by lattice energy minimizations carried out for a set of crystal structures of small peptides.

The lattice energy of a crystal structure was considered as a function of torsional angles, unit cell parameters, and positions and orientations (Euler angles) of the molecules in the unit cell. It was calculated as a sum of inter- and intramolecular terms with the latter described by eq 1

$$E_{lattice} = E_{inter} + E_{intra} \tag{5}$$

The intermolecular term was calculated as a sum of atom−atom interactions

$$E_{inter} = \frac{1}{2}\sum_i \sum_j E_{ij}^{inter} = \frac{1}{2}\sum_i \sum_j (E_{ij}^{nbond} + E_{ij}^{electr}) \tag{6}$$

where the summation is carried out over atoms in different molecules.

The intermolecular electrostatic energy was calculated with the Coulomb term of eq 2 and the Ewald summation[53] without including a dipole moment correction term.[54] The charges derived for the atoms of the 20 naturally occurring amino acids and the atoms of the Ac− and NMe− end groups were used in the crystal calculations of terminally blocked peptides. The partial atomic charges of the $-NH_3^+$, $-COO^-$, and $-COOH$ end groups for peptides (not for amino acids) were computed by fitting the ab initio electrostatic potential of each of the peptides, keeping the charges in the rest of the molecule fixed. To achieve better accuracy of fitting, the charges on C^α and H^α atoms of the residues of long peptides connected to the end groups were allowed to vary as well.

The details for calculating lattice energies are given in ref 55. The molecular geometries, i.e., bond lengths and bond angles, were taken as those in the ECEPP package.[3]

Local energy minimizations of the experimental crystal structures were carried out using the CRYSTALG program.[56] The experimental space group symmetry was used to generate coordinates of all the atoms in the unit cell. However, no symmetry constraints were used during the minimization; i.e., the positions and orientations of all the molecules in the unit cell, as well as the torsional angles, were allowed to vary for each molecule independently. In all cases but one, the starting symmetry was obtained at the end of the calculations.

The following three measures were used to provide a more stringent test of the force field.

(1) An average percent deviation of the unit cell parameters from their experimental values (Δ_{cell}) was calculated using the formula

$$\Delta_{cell} = 100\% \times \sum_{i=1}^{N} \left| \frac{x_{exp}^i - x_{m.exp}^i}{x_{exp}^i} \right| (1/N) \tag{7}$$

where x_{exp}^i and $x_{m.exp}^i$ are the unit cell parameters of the experimental and the minimized experimental structures, respectively; N is the number of unit cell parameters that changed as a result of energy minimization.

(2) The root-mean-square deviation of the torsional angles ($RMSD_{tor}$) provided information about the change in molecular conformation caused by the minimization with the new force field.

(3) The heavy-atom RMSD ($RMSD_{cryst}$) was used to evaluate the differences between the crystal structures before and after the energy minimization. The $RMSD_{cryst}$ values were computed using the COMPACK program.[57] In this program, each of the compared structures (the experimental and the minimized experimental) is represented by a molecule and its coordination shell, which was taken as the closest 12−16 molecules. The relationships between these molecules are then represented by a set of interatomic distances and angles. If the distances and angles in the experimental and the minimized experimental structures match to within specified tolerances, then the coordination spheres are overlaid, and the RMSD for the atomic positions is calculated for all 12−16 molecules. For RMSD calculations, hydrogen atoms were ignored because of the uncertainties in the X-ray determination of their positions, but they were included in the energy calculations.

7. Experimental Data. X-ray data were used for assessing the ability of the new force field to model the experimental crystal structures of small peptides. A search of the Cambridge Structural Database[58] (CSD), carried out for structures with a maximum R-factor of 7.5% and containing no ions or water

molecules, yielded several structures. The following criteria were applied to choose crystal structures for our calculations:

(1) The coordinates of all the atoms including hydrogen have to be provided.

(2) The observed structures should have no disorder.

(3) If several structures are available for a given molecule, then the one with the lowest *R*-factor was used.

(4) If several structures obtained at different temperatures are available, then the one corresponding to the lowest temperature was used to minimize the errors due to the neglect of temperature effects.

(5) Only structures with one molecule in the asymmetric unit were considered.

The 29 crystal structures considered in this work consisted of terminally blocked and unblocked amino acids and di-, tri-, and tetrapeptides.

Results and Discussion

1. Scaling 1−4 Interactions. Before deriving parameters of the torsional potential (eq 3), we decided to consider the issue of scaling 1−4 nonbonded interactions in detail. In many force fields, the interactions between atoms separated by three bonds (1−4 interactions) are treated specially. Empirical factors are introduced to scale down electrostatic and nonbonded interactions between 1,4 atoms. For example, a factor of 2 is commonly used to scale down the repulsion (or both repulsion and attraction) part of the potential. Electrostatic interactions are also scaled down, e.g., by 2 in the ECEPP/3[3] and OPLS,[9] and by 1.2 in the AMBER03[5] force fields. The values of the scaling factors are determined empirically by considering small model molecules.

The accuracy of a force field depends on the proper balance of all its components and on an accurate description of all the interactions in a molecule. The 1−4 interactions should also be described accurately because they affect the relative contributions of the nobonded and torsional terms to the total energy and, as a consequence, the local conformational preferences of a molecule. Sorin and Pande[59] demonstrated that, in the case of peptides, variation of the scale factor for 1−4 interactions can lead to changes in the helix-forming propensity of a given force field.

In the case of nonbonded interactions, a scale factor is necessary if the r^{-12} dependence is used for modeling nonbonded repulsion because it overestimates short-range repulsion compared with the more correct exponential term. As for 1−4 electrostatic interactions, some redistribution of charge along the bonds connecting 1,4 atoms, which would affect interaction energies, may be expected; therefore, a scaling factor is usually introduced to account for this effect. However, there is no physical basis for scaling 1−4 electrostatic interactions. Originally the 1−4 electrostatic scale factor of 2 was introduced in the AMBER force field to allow the united-atom model to more closely mimic the results of the all-atom simulations. Later the same value of the scale factor was adopted in the all-atom version of the AMBER force field. Different optimum values of the 1−4 electrostatic scaling factor have been proposed.[4,44b,60] For example, it was found[61] to be necessary to scale the 1−4 nonbonded interactions by the factor of 2 to permit use of the same parameters of the OPLS force field for intra- and intermolecular interaction. The 1−4 scale factor of 1.2 is used in the recent versions[4,5] of the AMBER force field. Billeter et al.[62] and Smith and Karplus[63] noted that the 1−4 electrostatic scaling can lead to artifacts in the conformational energies if relatively large charges are used and suggested that a scale factor of 1 should be used. Unfortunately, there is no strong evidence in favor of any one of these values. Moreover, a value of the 1−4 electrostatic scaling factor appears to depend on the particular force field.[59] This motivated us to study in detail the effect of scaling (k_{14}^{el} and k_{14}) on the accuracy of our force field and to determine values appropriate to our potential model.

In this work, the 6-exp nonbonded term, which should not require scaling down the repulsion energy, is used; nevertheless, the 1−4 nonbonded scaling may still be necessary since we use the rigid ECEPP-like geometry, which means that the bond lengths and bond angles cannot adjust to the changes in torsional angles that may lead to higher than real rotational barriers.

Since the effect of scaling electrostatic interactions is more pronounced when at least one of the 1, 4 atoms is electronegative, we selected ethanol, *n*-propanol, ethylamine, *n*-propylamine, 1,2-ethanediol, and 1,3-propanediol as model molecules. Two charge models were also considered: one including charges derived for a single lowest-energy conformation of each molecule and the second one including charges obtained from multiple-conformation fitting. According to eq 1, the torsional energy is the difference between the total intramolecular energy and the sum of the nonbonded and electrostatic terms. This means that, if the torsional parameters are derived by fitting this difference (with the total energy obtained from ab initio calculations), they will depend on the values of the 1−4 scaling factors employed to compute the nonbonded and electrostatic contributions. Therefore, to obtain a consistent set of torsional parameters for the torsional angles in the model molecules, the fitting was carried out for each combination of the parameters considered (including the values of the 1−4 scaling factors and charge model). Combinations of the parameters examined for each of the molecules are presented in Tables 4−7.

The torsional parameters for rotation of −OH and $-NH_2$ groups were derived from QM data for ethanol and ethylamine, respectively, and then used for deriving torsional parameters for rotation about the $-CH_2-CH_2-$ bonds in propanol, propylamine, 1,2-ethanediol, and 1,3-propanediol and for computing all rotational profiles for all the molecules. The parameters for the CC−CX (X = O, N) torsional angles were derived with the CC−XH and methyl groups in the trans and staggered conformations, respectively. The rotational barriers of the methyl groups in ethane and propane were found to be well transferable to larger molecules, such as butane, ethanol, propanol, etc., and are not sensitive to the scaling of the 6-exp or electrostatic interactions. Therefore, the same torsional parameters for rotation of the methyl group were used in all the calculations without recalculating them for each value of a scaling factor.

The accuracy of each scaling factor was evaluated by comparing the empirical torsional profiles with the ab initio ones using two criteria: (i) the RMSDs calculated for all minima and (ii) the RMSDs for a given torsional profile.

The results of the calculations carried out for ethanol and ethylamine show that the QM energy barriers and the relative conformational energies,[64,65] as well as the torsional profiles, are reproduced very well using all the MM models considered (i.e., different values of k_{14}, k_{14}^{el}, q_s, and q_m) with negligible (<2.5%) differences in the energies and RMSDs between them.

As we consider larger molecules, i.e., propanol and propylamine (Tables 4 and 5, respectively), the differences in performance of different models become more noticeable. The multiconformational charge model is clearly superior to the model based on a single conformation. Through the use of the single-conformation charge model for propanol, the conformational energy differences are quite sensitive to the electrostatic

TABLE 4: Relative Conformational Energies of the Minima and Energy RMSDs for Propanol[a]

conformation		ΔE^{QM}		ΔE^{MM}							
				k_{14}^{el} (q_s, $k_{14}=1$)[c]			k_{14}^{el} (q_m, $k_{14}=1$)[c]			k_{14}^{el} (q_m, $k_{14}=2$)[c]	
CC−CO	CC−OH	MP3[b]	MP2/6-31G**	1.0	1.2	2.0	1.0	1.2	2.0	1.2	2.0
trans	g^+	0.19	0.08	1.12	0.81	0.08	−0.18	−0.10	−0.67	−0.28	−0.28
g^+	g^-	0.33	0.37	0.03	−0.30	−1.07	−0.18	−0.11	0.02	−0.15	−0.02
g^+	g^+	0.17	0.07	0.89	0.54	−0.30	−0.25	−0.18	−0.02	−0.20	−0.17
trans	trans	0.18	0.06	−0.01	−0.01	0.00	−0.02	−0.02	−0.73	−0.07	−0.07
g^+	trans	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.00
				RMSD							
minima[d]				0.61	0.49	0.66	0.31	0.26	0.51	0.32	0.28
rotation about C−O bond (CC−CO trans)[e]				1.15	0.78	0.57	0.27	0.21	0.10	0.35	0.23
rotation about C−O bond (CC−CO g^+)[e]				0.53	0.60	0.58	0.58	0.48	0.43	0.51	0.39

[a] All values are energies in kcal/mol. [b] The MP3/6-31+G**//HF/6-31G* energy was taken from ref 65. [c] k_{14}^{el} and k_{14} are scale factors for 1−4 electrostatic and nonbonded interactions, respectively; q_s and q_m denote sets of partial atomic charges derived using a single (lowest-energy) and several low-energy conformations, respectively. [d] RMSD calculated for the energy minima of propanol. [e] RMSD (in kcal/mol) between QM and MM energies for an ensemble of conformations, calculated at 10° intervals of the given torsional angle.

TABLE 5: Relative Conformational Energies of the Minima and Energy RMSDs for Propylamine[a]

conformation		ΔE^{QM}		ΔE^{MM}							
				k_{14}^{el} (q_s, $k_{14}=1$)			k_{14}^{el} (q_m, $k_{14}=1$)			k_{14}^{el} (q_m, $k_{14}=2$)	
CC−CN	CC−NX[b]	MP3[c]	MP2/6-31G**	1.0	1.2	2.0	1.0	1.2	2.0	1.2	2.0
trans	g^+	0.30	0.07	0.30	0.48	0.83	0.00	0.00	0.45	0.00	0.31
g^+	g^-	0.25	0.30	1.52	1.70	2.05	0.62	0.63	1.08	1.54	1.52
g^+	g^+	0.61	0.72	1.10	1.28	1.64	0.62	0.62	1.08	1.52	1.54
trans	trans	0.00	0.00	0.00	0.00	0.00	0.29	0.04	0.00	0.19	0.00
g^+	trans	0.32	0.55	0.60	0.60	0.61	0.90	0.65	0.61	1.69	1.17
				RMSD							
minima				0.58	0.70	0.95	0.26	0.16	0.42	0.87	0.72
rotation about C−N bond (CC−CN trans)				0.27	0.40	0.68	0.24	0.09	0.36	0.18	0.31
rotation about C−N bond (CC−CN g^+)				0.54	0.58	0.75	0.86	0.75	0.58	0.34	0.45
rotation about C−C bond (CC−NX trans)				0.09	0.09	0.09	0.08	0.08	0.08	0.48	0.27
rotation about C−C bond (CC−NX gauche)				0.38	0.38	0.38	0.26	0.26	0.26	0.39	0.24

[a] For symbols and units, see footnotes to Table 4. [b] X denotes the position of the lone-pair electrons of the amine nitrogen. X was used to define the conformation of the amino group; however, no additional lone-pair charge was used for the amine nitrogen. [c] The MP3/6-31+G**//HF/6-31G* energies were taken from ref 65.

TABLE 6: QM and MM Relative Conformational Energies of the Stable Conformations of 1,2-Ethanediol[a]

	ΔE^{QM} (kcal/mol)		ΔE^{MM}, kcal/mol								
			k_{14}^{el} (q_s, $k_{14}=1$)		k_{14}^{el} (q_m, $k_{14}=1$)				k_{14}^{el} (q_m, $k_{14}=0.5$)		
conformation[b]	MP3/MP2[c]	MP2/6-31G**	1.2	2.0	1.0	1.1	1.2	2.0	1.0	1.2	2.0
ttt	2.69	1.89	5.31	6.23	3.55	3.95	4.03	4.39	3.78	4.17	4.54
tgt	4.45	3.55			4.98	5.33	5.41		5.18		
ttg	3.41	2.25	6.48	6.89	4.77	4.93	4.95	5.04	4.82	4.98	5.07
ggt	4.65	4.00									
gtg	3.49	2.75	7.92	7.55	6.22	6.09	6.04	5.79	6.06	5.92	5.66
ggg	3.56	0.73									
gtg^-	3.15	2.37	6.73	6.42	5.17	5.05	5.01	4.79	5.02	4.90	4.67
ggg^-	0.65	0.72	1.06	0.42	1.26	0.93	0.85	0.49	1.00	0.65	0.28
tgg^-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
gg^-g	1.22	1.31	1.03	0.89	1.21	1.18	1.16	1.06	1.22	1.23	1.14
			RMSD[d]								
			3.29	3.38	1.98	2.03	2.03	2.05	1.96	2.04	2.03

[a] For symbols and units, see footnotes to Table 4. [b] A three-letter code with each of the letters denoting the conformations of one of the torsional angles is used to define the conformation of the 1,2-ethanediol molecule. [c] This set uses the MP2/6-31+G**//HF/6-31G* energies for the ttt, ggg^-, tgg^-, and gg^-g conformations, and the MP2/6-31G*//HF/6-31G* energies for the remaining conformations.[65] [d] RMSD calculated for the energy minima of 1,2-ethanediol.

scale factor (columns 5−7, Table 4). This sensitivity is reduced with the multiple-conformation charges (columns 8−10, Table 4). In general, use of the multiple-conformation charge model leads to lower RMSD values. All three multiple-conformation charge models with $k_{14}=1.0$ (columns 8−10, Table 4) yielded similar results with slightly lower RMSDs for the minima and the energy profiles obtained with the 1−4 electrostatic scaling factor of $k_{14}^{el}=1.2$.

None of the models was able to reproduce the ab initio relative stabilities (ΔE) of different conformers of propanol; this also happened when the AMBER94[4] or MM3[66] force fields were applied.[65] However, the propanol conformations have very similar energies, and the differences between our MM and the QM results are on the order of the differences between the energies obtained using different QM methods.[65]

TABLE 7: QM and MM Relative Conformational Energies of the Stable Conformations of 1,3-Propanediol[a]

		ΔE^{MM}, kcal/mol						
	ΔE^{QM} (kcal/mol)	k_{14}^{el} (q_m, k_{14} = 1)				k_{14}^{el} (q_m, k_{14} = 0.5)		
conformation[b]	MP2/6-31G**	1.0	1.1	1.2	2.0	1.0	1.2	2.0
gg$^-$g$^-$g	3.86	3.24	3.32	3.38	3.57	3.24	3.42	3.60
gg$^-$gg$^-$	0.00[c]							
ggg$^-$g$^-$	0.09							
ggg$^-$g	0.09	−0.06	−0.01	0.04	0.22	−0.06	−0.01	0.17
gggg$^-$	1.45	0.60	0.80	0.97	1.66	0.66	1.05	1.60
gggg	2.43	2.82	2.86	2.89	2.89	2.81	2.88	2.95
gtg$^-$g$^-$	2.31	1.84	1.52	1.26	−0.46	1.86	1.27	0.13
gtg$^-$g	3.42	3.37	3.02	2.72	0.75	3.37	2.72	1.40
gtgg$^-$	3.15	2.61	2.27	1.99	0.13	2.62	2.03	0.77
gtgg	2.46	1.77	1.43	1.15	−0.70	1.78	1.19	−0.06
gttg$^-$	3.20	3.27	2.49	1.84	−2.21	3.27	1.83	−0.93
gttg	2.90	2.67	1.88	1.24	−2.82	2.67	1.23	−1.54
tgg$^-$g	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
tgg$^-$g$^-$	0.00[c]							
tggg$^-$	1.50	0.60	0.75	0.86	1.38	0.66	1.07	1.41
tggg	1.75	1.62	1.62	1.62	1.52	1.61	1.73	1.65
tgg$^-$t	6.12							
tggt	1.29	0.68	0.66	0.63	0.49	0.68	0.86	0.69
tgtg$^-$	2.37	2.27	1.91	1.60	−0.32	2.30	1.73	0.38
tgtg	1.98	1.15	0.77	0.46	−1.53	1.17	0.59	−0.81
ttgg$^-$	3.34	3.37	2.96	2.63	0.52	3.37	2.79	1.28
ttgg	1.97	1.23	0.85	0.54	−1.43	1.24	0.67	−0.72
ttgt	2.02	1.60	1.18	0.83	−1.26	1.62	1.10	−0.45
tttg	2.92	2.96	2.12	1.43	−2.82	2.96	1.53	−1.45
tttt	3.08	3.67	2.76	2.01	−2.48	3.66	2.27	−0.96
RMSD[d]		0.50	0.71	1.00	3.17	0.49	0.92	2.42

[a] For symbols and units, see footnotes to Table 4. [b] Each letter of a four-letter code denotes the conformation of one of the four torsional angles of the 1,3-propanediol molecule. [c] These conformations are not stable and transform to tgg$^-$g upon the energy minimization. [d] RMSD calculated for the energy minima of 1,3-propanediol.

For propylamine (Table 5), the multiple-conformation charge models also perform better than the models based on a single conformation. The lowest RMSD calculated for the minima was found for this model with the electrostatic scale factor of k_{14}^{el} = 1.2 and no scaling for nonbonded interactions (k_{14} = 1.0). There is no single best model, as judged by the RMSD for the energy profiles; i.e., different values of the electrostatic scale factor lead to low RMSD values for different energy profiles.

When the models including scaling of repulsion interactions (k_{14} = 2.0) are compared with the models with k_{14} = 1.0 (columns 11−12, Tables 4 and 5), it should be noted that they behave similarly. In general, the results obtained for propanol and propylamine using the models with different values of the 1−4 scaling factors do not enable one to make a definite conclusion regarding the best values of k_{14}^{el} and k_{14}. Propanol and propylamine are too small to be good model molecules for studying 1−4 interactions.

Two larger molecules, also containing an electronegative atom (oxygen), namely, 1,2-ethanediol and 1,3-propanediol, were considered. As was discussed previously,[65] 1,2-ethanediol is a good model system for studying conformational energies as a function of the charge model and the 1−4 electrostatic scale factor. This molecule has three important torsional angles and two electronegative atoms in the 1 and 4 positions. 1,3-Propanediol is larger and has four torsional angles. 1,2-Ethanediol and 1,3-propanediol can assume 10^{65} (Table 6) and 25 (Table 7) different nonsymmetric conformations, respectively. In each of these conformations, the torsional angles are either trans or gauche (+ or −). We use a three- and four-letter code to define the conformations (trans, gauche$^+$, and gauche$^-$) of the 1,2-ethanediol and 1,3-propanediol molecules, respectively, with each of the letters denoting the conformations of one of the torsional angles.

Table 6 presents QM and MM relative conformational energies obtained by full energy minimization of the 10 conformations of 1,2-ethanediol (i.e., all torsional angles were allowed to vary). It should be mentioned that the rigid geometry used in our simulations results in some conformations of 1,2-ethanediol, such as gtg$^-$, gtg, ttg, and ttt, having energies much higher than the corresponding ab initio values, while others, such as ggt and ggg, are not stable minima of the potential. To study the influence of the 1−4 electrostatic scaling factor and the charge model on the conformational energy of 1,2-ethanediol, we considered both the single- and multiple-conformation charge models (the latter derived using ttt, ggg$^-$, tgg$^-$, and gg$^-$g conformations). As in the case of other molecules considered so far, the multiple-conformation charge models (columns 6−9, Table 6) perform noticeably better (lower RMSDs) than the single-conformation ones (columns 10−12, Table 6). As the next step, we compared performance of the models characterized by the different values of the 1−4 repulsion scaling factor (k_{14}). Here, two values of k_{14} (1.0 and 0.5) were used. The models with (k_{14} = 0.5) and without (k_{14} = 1.0) scaling of the repulsive interactions appear to behave similarly when the same value of k_{14}^{el} is used. For example, almost identical lowest RMSD values were obtained for both types of models (1.98 and 1.96 kcal/mol for k_{14} = 1.0 and k_{14} = 0.5, respectively).

As for the scaling of the 1−4 electrostatic interaction, all the values of the scale factor yield similar RMSDs for the stable minima; however, the number of these minima is different for k_{14}^{el} = 1.0, 1.2, and 2.0, with the additional tgt conformation being stable for the electrostatic scale factors of 1.0 and 1.2. All the models identify the tgg$^-$ conformation as the lowest-energy minimum, in agreement with the QM data. The lowest

RMSD was obtained for $k_{14}^{el} = 1.0$; however, it is only slightly lower than the deviations obtained for other values of k_{14}^{el}. These results suggest that molecules larger than 1,2-ethanediol have to be considered to observe an effect of the 1−4 scaling; hence, we considered 1,3-propanediol.

QM and MM relative conformational energies obtained by full energy minimization of the 25 conformations of 1,3-propanediol using different values of k_{14}^{el} and k_{14} are given in Table 7. Since it has been shown that the single-conformation charge model performs significantly worse than the multiconformation one, it was not used for calculations of 1,3-propanediol. Partial atomic charges were computed using five low-energy conformers (tttg, ttgg, tggt, tgg⁻g, and ggg⁻g). For all the models considered, the gg⁻gg⁻ and tgg⁻g⁻ conformations are not stable minima of the MM force field.

As in the case of 1,2-ethanediol, the models with different scaling of the 1−4 repulsive interactions yield similar results for the same values of k_{14}^{el}. For instance, RMSDs of 1.0 and 0.92 kcal/mol were obtained for $k_{14} = 1.0$ (column 5, Table 7) and 0.5 (column 8, Table 7), respectively. The effect of scaling the 1−4 repulsion becomes noticeable only for $k_{14}^{el} = 2$ where it leads to a lower RMSD (2.42 vs 3.17 kcal/mol for $k_{14} = 1.0$).

By contrast to the results reported for the other molecules considered in this section, the data in Table 7 demonstrate a strong effect of the 1−4 electrostatic scaling on the relative conformational energies. The lowest RMSD of 0.5 kcal/mol was obtained for $k_{14}^{el} = 1.0$ and is 2 times lowers than the corresponding value for $k_{14}^{el} = 1.2$.

Additionally, to make sure that $k_{14}^{el} = 1.0$ is the optimum value, we carried out energy minimizations for all conformers of 1,3-propanediol using the intermediate k_{14}^{el} value of 1.1. The resulting RMSD (0.71 kcal/mol) was still higher than the value obtained for $k_{14}^{el} = 1.0$.

In summary, the procedure for calculating partial atomic charges has a strong effect on both the conformational energies and the torsional profiles of all the model molecules. At the same time, the effect of the 1−4 scaling depends on the size of the molecule and becomes really pronounced only for 1,3-propanediol, the largest molecule considered here.

Scaling the nonbonded interactions does not lead to a significant improvement in RMSDs, and therefore, $k_{14} = 1.0$ was used in all the calculations described in the rest of the paper. As for the electrostatic scale factor, the value of 1.0 was found to provide the best results and was adopted in the rest of the work.

2. Torsional Barriers of Small Molecules. The list of small molecules considered in this work is given in Tables 8−12. It includes both the molecules used for deriving torsional parameters appropriate for the side-chain groups found in proteins and the related molecules used for testing the transferability of the parameters.

All torsionals were divided into four categories according to the type of the central bond for each torsional angle: C−C, C−N, C−O, and C−S.

We began the development of the torsional part of the force field with hydrocarbons. Ethane was selected as the simplest model for methyl torsionals. For the calculations on the molecules with methyl groups bonded to sp^3 carbons, the aim was to see if the potential function is sufficiently versatile to reproduce widely varying methyl rotational barriers. The torsional parameters derived for ethane yield almost exact agreement with ab initio data (RMSD = 0.01 kcal/mol) and are well transferable to other alkanes (Table 8). New types of torsional parameters had to be introduced for alcohols and amines (Table 8) where one of the 1 and 4 atoms involved in methyl group rotation is electronegative. The parameters derived for alcohols and amines using ethanol and ethylamine, respectively, perform well when applied to the related molecules (RMSD ≤ 0.2 kcal/mol, Table 8).

TABLE 8: Calculated and Experimental Torsional Barriers (kcal/mol) for Rotation of the Methyl Group (C−CH$_3$)[a]

molecule	barrier (MP2)	barrier (MM)	RMSD[b] (kcal/mol)
ethane	2.99	3.00	0.01
propane	3.35	3.28	0.05
n-butane (CC−CC t)	3.24	3.29	0.03
n-butane (CC−CC g)	3.03	3.23	0.13
2-methylpropane	3.71	3.80	0.17
2,2-dimethyl-propane	4.22	4.04	0.09
ethylamine (CC−CN t)	3.71	3.72	0.11
ethylamine (CC−CN g)	3.73	3.79	0.03
methyl-ethyl sulfide (CC−SC t)	3.32	3.32	0.01
tert-butylamine	4.19	4.10	0.11
ethylmethylamine	3.55	3.57	0.13
	3.16	3.24	0.15
ethanol (CC−CO t)	3.72	3.72	0.00
ethanol (CC−CO g)	3.80	3.74	0.19
1-propanol (CC−CO t; CC−OH t)	2.98	3.28	0.19
1-propanol (CC−CO t; CC−OH g)	3.11	3.27	0.12
1-propanol (CC−CO g; CC−OH g)	3.20	3.24	0.20
2-propanol	3.68	3.49	0.20
2-methyl-2-propanol		3.87	
propanoic acid	2.67	2.67	0.01
ethylbenzene	3.64	3.64	0.02
acetone	0.74	0.75	0.03
acetic acid	0.59	0.60	0.11
alanine	3.53	3.50	0.10
acetamide	0.24	0.14	0.07
N-methylacetamide	0.08	0.06	0.01

[a] Torsional profiles for a given angle were computed with the rest of the angles in the lowest-energy conformation; however, for some molecules we considered higher-energy conformations as well. The t and g conformations considered are specified in parentheses. [b] The RMSDs between QM and MM energy profiles are calculated at 10° intervals of the XC−CH$_3$ torsional angle (X is any atom).

New types of torsionals were used for acetone and acetic acid, two other molecules with electronegative atoms involved in 1−4 interactions. The barriers calculated for these molecules are in good agreement with ab initio data.

The torsional parameters appear to be somewhat sensitive to the type of 1,4 carbons for CC−CC angles. For example, better agreement with the ab initio data (Table 8) is achieved when the torsional parameters were derived specifically for propanoic acid and ethylbenzene rather than transferred from ethane.

Next we considered torsionals pertaining to the rotation about the C1−C′ bond (CC−C′X torsional angle), where C′ denotes any type of carbon and X is C, N, O, or S (Table 9). Here, we begin with the torsionals describing the rotation about the C1−C1 bond. Both the rotational barrier and the relative conformational energy of *n*-butane are reproduced well (RMSD of 0.1 kcal/mol). However, the torsional parameters are not transferable to 2-methylbutane, a model molecule for the leucine and isoleucine side chains. When the torsional parameters fitted to the ab initio data for 2-methylbutane were applied to this molecule, the agreement with the ab initio data improved. There is still a small deviation (~0.14 kcal/mol) for the t → g$^+$ barrier, whereas the g$^+$ → g$^-$ barrier and the relative conformational energy are reproduced almost exactly (2-methylbutane′ in Table 9).

New types of torsional parameters (beyond those already established for propanol, methylamine, and 1,2-ethanediol) were introduced for each of the C1C1−C1X (X = N, O, S) dihedral

TABLE 9: Calculated and Experimental Torsional Barriers and Relative Conformational Energies (in kcal/mol) for Rotation about C−C Bonds[a]

molecule	barrier (MP2)[b]	barrier (MM)	ΔE (MP2)	ΔE (MM)	RMSD[c]
		C1−C1			
n-butane	3.67 (t → g$^+$)	3.55	0.00 (t)	0.00 (t)	0.10
	6.39 (g$^+$ → g$^-$)	6.48	1.13 (g)	1.04 (g)	
2-methyl butane[d]	2.73 (g$^+$ → g$^-$)	2.73	0.00 (t)	0.00 (t)	0.47
	5.72 (t → g$^+$)	5.86	1.02 (g)	1.10 (g)	
n-propylamine	4.01 (t → g$^+$)	4.01	0.00 (t)	0.00 (t)	0.14
(CC−NH t)	6.06 (g$^+$ → g$^-$)	6.12	0.74 (g)	0.50 (g)	
n-propylamine	4.14 (t → g$^+$)	3.93	0.00 (t)	0.00 (t)	0.23
(CC−NH g$^+$)	5.87 (g$^+$ → g$^-$)	6.46	0.61 (g)	0.46 (g)	
1-propanol	3.96 (t → g$^+$)	3.93	0.00 (g)	0.00 (t)	0.04
	5.48 (g$^+$ → g$^-$)	5.54	0.04 (t)	0.04 (g)	
methyl propyl	3.49 (t → g$^+$)	3.28	0.00 (t)	0.00 (t)	0.15
sulfide (CC−SC t)	6.43 (g$^+$ → g$^-$)	6.49	0.89 (g)	0.82 (g)	
methyl propyl	3.69 (g$^-$ → t)	3.05	0.00(t)	0.00 (t)	0.36
sulfide (CC−SC g$^+$)	6.73 (g$^+$ → g$^-$)	6.05	0.59 (g$^+$)	0.57 (g$^+$)	
	3.47 (t → g$^+$)	3.05	1.72 (g$^-$)	0.78 (g$^-$)	
butyl amide	3.03 (t → g$^+$)	2.74	0.00(t)	0.00(t)	0.23
	6.07 (g$^+$ → g$^-$)	6.32	0.17(g)	0.24 (g)	0.86 (butane)[e]
		(3.42/4.72, butane)[e]		0.04(t) (butane)[e]	
butyrate	3.07 (t → g$^+$)	3.06	0.00 (g)	0.00(g)	0.11
	6.19 (g$^+$ → g$^-$)	6.18	0.46 (t)	0.36 (t)	
1-propylamine cation	3.82 (t → g$^+$)	3.67	0.00 (t)	0.00 (t)	0.11
	5.41 (g$^+$ → g$^-$)	5.54	0.50 (g)	0.43 (g)	(0.24 butane)[e]
		(3.42/5.62, butane)[e]		(0.26 butane)[e]	
		C1−C2			
propanoic acid	0.89 (t → g$^+$)	0.85	0.0 (t)	0.00 (t)	0.05
	2.09 (g$^+$ → g$^-$)	2.12	0.61 (g)	0.61 (g)	
propionate	1.08	0.84			0.17
butyrate (CC−CC t)	1.81	1.02			0.58
		C1−C3			
propylamide	1.27 (t → g$^+$)	1.30			0.15
	1.85 (g$^+$ → g$^-$)	1.84			
		C1−CA			
ethylbenzene	1.31	1.37			0.23
4-ethyl-imidazole (δ)	0.85 (t → g$^+$)	0.84	0.00 (g)	0.00 (g)	0.01
	1.62 (g$^+$ → g$^-$)	1.61	0.19 (t)	0.18 (t)	
5-ethyl-imidazole (ϵ)	1.18 (t → g$^+$)	1.18	0.00 (g)	0.00 (g)	0.02
	1.56 (g$^+$ → g$^-$)	1.54	0.82 (t)	0.80 (t)	
4(5)-ethylimidazole cation (+)	0.87 (t → g$^+$)	0.81	0.00 (t)	0.00 (t)	0.05
	0.96 (g$^+$ → g$^-$)	0.94	0.16 (g)	0.10 (g)	

[a] Atom types are as defined in Table 1. [b] More than one value indicates that there are several nonequivalent barriers in the torsional energy profiles. Values of the barriers are defined relative to the most stable conformation. [c] The RMSDs (kcal/mol) between QM and MM energy profiles are calculated at 10° intervals of the torsional angle corresponding to rotation about the C−C bond. [d] The trans conformation corresponds to a H_3C-CH_2-CH angle of 180°. [e] On the basis of torsional parameters derived for butane.

angles (Table 9). These parameters perform very well (low RMSD) for the torsional profiles used for fitting, namely, for the trans torsional profile (amino group in the trans conformation) of *n*-propylamine and 1-propanol and the trans profile of methyl propyl sulfide. Larger deviations from the ab initio data were observed when the parameters were applied to calculate the *n*-propylamine and methyl propyl sulfide gauche profiles (RMSDs of 0.23 and 0.36 kcal/mol, respectively). For these profiles, there are noticeable differences between the calculated and ab initio barriers, as well as between the relative conformational energies; however, these results are acceptable, taking into account the size of the molecules and the rigid geometry used.

Two sets of torsional parameters were considered for *n*-butyl amide. First, the corresponding parameters obtained for *n*-butane were used to examine their transferability. The second set was produced by fitting the QM data for *n*-butyl amide. The deviations of the calculated torsional barriers from the ab initio values are larger for the butane parameter set, and the corresponding RMSD over the entire energy profile is almost 3 times greater (0.23 vs 0.86 kcal/mol) than for the parameters derived from the QM data for *n*-butyl amide.

The parameters for the rotation about the C1−C′ bond (where C′ stands for the amide, carboxyl, or aromatic carbon) were calculated by fitting the ab initio data for butyl amide, propionic acid, ethylbenzene, 4-ethyl-imidazole, and 5-ethyl-imidazole. Very good agreement with the ab initio profiles was obtained for all five molecules (RMSD $<$ 0.25 kcal/mol).

The torsional potential consisting of the three-term Fourier expansion with the coefficients derived using ab initio data also works well for torsionals associated with rotation about the C−O bond (Table 10). The rotational energy profiles of methanol and phenol are reproduced almost exactly (RMSDs of 0.00 and 0.07 kcal/mol, respectively). The set of parameters derived for ethanol yields accurate values of the barriers and the conformational energy of this molecule (relative to the lowest-energy trans conformation) and is transferable to 1-propanol and 2-methyl-2-propanol. The ethanol torsional parameters do not perform well for 2-propanol; therefore, a new set of parameters was obtained from the QM data for this molecule. As can be seen, the use of the new parameters leads to good agreement with the ab initio results (RMSD = 0.08 kcal/mol).

To obtain torsional parameters for rotation of the −OH group in carboxylic acids, we considered acetic and propanoic acid

TABLE 10: Calculated and Experimental Torsional Barriers and Relative Conformational Energies (in kcal/mol) for Rotation about C−O Bonds[a]

molecule	barrier (MP2)	barrier (MM)	ΔE (MP2)	ΔE (MM)	RMSD
		C1−OH1			
methanol	1.39	1.39			0.00
ethanol	1.29 (t → g^+)	1.26	0.00 (t)	0.00 (t)	0.02
	1.77 (g^+ → g^-)	1.77	0.09 (g)	0.08 (g)	
1-propanol (CC−CO t)	1.18 (t → g^+)	1.38	0.00 (t)	0.16 (t)	0.27
	1.73 (g^+ → g^-)	1.61	0.01 (g)	0.00 (g)	
1-propanol (CC−CO g^+)	1.22 (t → g^+)	1.15	0.00 (t)	0.25 (t)	0.58
	2.25 (g^+ → g^-)	1.29	0.17 (g^+)	0.00 (g^+)	
	0.98 (g^- → t)	1.01	0.33 (g^-)	0.07 (g^-)	
2-propanol	1.27 (g^+ → g^-)	1.02	0.00 (g^+)	0.00 (g^+)	0.08
	1.59 (g^- → t)	1.62	0.24 (t)	0.43 (t)	
2-methyl-2-propanol	1.50 (t → g^+)	1.22	0.00 (t = g)	0.24 (t)	0.14
	1.50 (g^+ → g^-)	1.36		0.00 (g)	
		CA−OH1			
phenol	3.58	3.58			0.07
		C2−OH2			
acetic acid	13.79	13.85	0.00 (t)	0.00 (t)	0.22
			7.14 (c)	7.17 (c)	
propanoic acid	13.57	13.61	0.00 (t)	0.00 (t)	0.15
		(14.14 acetic acid)[b]	6.92 (c)	6.96 (c)	(0.78 acetic acid)[b]
				(8.12 acetic acid)[b]	

[a] For symbols and units, see footnotes to Table 9. [b] On the basis of torsional parameters derived for acetic acid.

TABLE 11: Calculated and Experimental Torsional Barriers and Relative Conformational Energies (in kcal/mol) for Rotation about C−N Bonds[a]

molecule	barrier (MP2)	barrier (MM)	ΔE (MP2)	ΔE (MM)	RMSD
		C1−N1			
methylamine	2.28	2.27			0.02
ethylamine	2.09 (t → g^+)	2.06	0.00 (t)[b]	0.00 (t)[b]	0.05
	2.72 (g^+ → g^-)	2.77	0.03 (g)	0.06 (g)	
n-propylamine (CC−CN t)	1.99 (t → g^+)	2.21	0.00 (t)[b]	0.29 (t)[b]	0.19
	2.75 (g^+ → g^-)	2.83	0.08 (g)	0.00 (g)	
n-propylamine (CC−CN g^+)	3.36 (t → g^+)	2.56	0.42 (g^+)	0.00 (g^+)	0.45
	1.75 (t → g^-)	2.17	0.25 (t)	0.28 (t)	
	3.05 (g^+ → g^-)	2.74	0.00 (g^-)	0.00 (g^-)	
ethylamine cation	2.65	2.65			0.00
1-propylamine cation (CC−CN t)	2.55	2.65			0.06
1-propylamine cation (CC−CN g)	2.55	2.62			0.22
		C1−NB			
N-methylacetamide	0.09	0.08			0.05
		C3−N2			
acetamide	22.54	22.29			1.11
propionamide	22.27	22.05			1.13
		(22.35 acetamide)			(1.25 acetamide)
		C3−NB			
N-methylacetamide	25.65	25.89	0.00 (t)	0.00 (t)	0.32
			2.33 (c)	2.61 (c)	
		CA−NA			
methylguanidine	22.41	22.51	0.00 (c)	0.00 (c)	0.98
			0.30 (t)	0.43 (t)	
		N1−CA			
methylguanidine	12.54	12.59	0.00 (t)	0.00 (t)	0.10
			0.85 (c)	0.88 (c)	
methylguanidine (NH_2)	6.96	7.12			0.82
1-methylguanidine cation	11.48	9.83			0.82
		C1−N1			
1-ethylguanidine	6.60	7.09			0.59

[a] For symbols and units, see footnotes to Table 9. [b] The initial conformation of CC−NH_2 corresponds to a CC−NX angle of 180°, where X denotes the position of lone-pair electrons of amine nitrogen.

(Table 10). The resulting parameters derived for acetic acid perform well for this molecule but are not really transferable to propanoic acid, for which they yield the relative conformational energy ~1.2 kcal/mol greater than the ab initio value (6.92 kcal/mol).

Three groups of molecules, namely, amines, amides, and guanidine, were used for deriving torsional parameters for rotations about C−N bonds (Table 11). The MM torsional model (eq 3) with the parameters derived using QM data reproduces the QM rotational profiles of methylamine and ethylamine

TABLE 12: Calculated and Experimental Torsional Barriers and Relative Conformational Energies (kcal/mol) for Rotation about C−S and S−S Bonds[a]

molecule	barrier (MP2)	barrier (MM)	ΔE (MP2)	ΔE (MM)	RMSD
		C1−S			
dimethyl sulfide	2.08	2.08			0.01
methyl-ethyl sulfide (CC−SC t)	1.96	1.95			0.02
		(2.09 dimethyl sulfide)[b]			(0.07 dimethyl sulfide)[b]
methyl-ethyl sulfide (CC−SC g)	1.67	1.99			0.20
		(2.12 dimethyl sulfide)			(0.28 dimethyl sulfide)[b]
methyl-ethyl sulfide (CC−SC)	2.03 ($t \rightarrow g^+$)	1.54	0.00 (t)	0.00 (t)	0.43
	4.47 ($g^+ \rightarrow g^-$)	4.87	0.43 (g)	0.39 (g)	
methyl-ethyl disulfide (SS−CH)	1.60	1.60			0.13
methyl-ethyl disulfide (CC−SS)	1.58 ($t \rightarrow g^+$)	1.59	0.00 (g^+)	0.00 (g^+)	0.23
	3.98 ($g^+ \rightarrow g^-$)	4.39	0.02 (t)	0.19 (t)	
	1.76 ($g^- \rightarrow t$)	1.53	0.46 (g^-)	0.78 (g^-)	
		S−S			
methyl-ethyl disulfide	5.64 ($g^+ \rightarrow g^-$)	5.60			0.29
	11.34 ($g^- \rightarrow g^+$)	11.52			

[a] For symbols and units, see footnotes to Table 9. [b] On the basis of torsional parameters derived for dimethyl sulfide.

almost exactly. The results for *n*-propylamine have been discussed in the previous subsection, and agreement with the ab initio data obtained for this molecule can be considered as acceptable.

The torsional parameters for rotation of the $-NH_2$ group in amides were calculated from the QM results for acetamide and were subsequently applied to modeling the rotation of the NH_2 group in propionamide. In both cases, the barrier heights are reproduced quite well (Table 11) with just one term (the second) of the Fourier expansion (eq 3); however, the RMSDs are unusually high (1.11 and 1.25 kcal/mol for acetamide and propylamide, respectively). Such large deviations arise from the difference in the shapes of the ab initio and the empirical barriers. The ab initio barrier is narrower; i.e., the energy decreases faster than for the empirical potential. This difference is no problem for our purposes, as long as the barrier height and the general shape of the profile are correct.

Rotation about the C−N peptide bond in *N*-methylacetamide represents a special case because the MP2/6-31G** method predicts that the barrier is more than 2 kcal/mol higher than the experimental value (22.6−23.2 kcal/mol[67]). Since ab initio data were used for deriving all other torsional parameters including backbone torsionals, we decided to use the ab initio results (instead of the experimental data) for the C−N peptide bond as well to make the parameters consistent with the rest of the force field. The resulting parameters reproduce the ab initio barrier in *N*-methylacetamide accurately (Table 11). We considered *N*-methylacetamide as a model molecule for deriving torsional parameters for the peptide bond in peptides and proteins.

The barriers and the relative conformational energies for rotation of the −NH and $-NH_2$ groups in methylguanidine are in good agreement with the ab initio data, but as in the case of the amides, the RMSDs are quite high (0.98 and 0.82 kcal/mol, respectively) due to the difference in the shape of the energy barriers. The parameters for the methylguanidine torsional potential corresponding to rotation about the CN−CN bond yield correct values of the barrier and the relative conformational energy.

According to the ab initio calculations, the barrier to rotation of the $N-CH_3$ group in *N*-methylacetamide is very low (~0.1 kcal/mol) and is modeled well by a single third term of the Fourier expansion. The torsional parameters derived here will be used later to describe the rotation of the methyl group in the *N*-methyl end group of peptides.

Finally, we considered sulfur-containing compounds with C1S−C1C1 and C1S−SC1 dihedral angles (Table 12). First, the MM parameters were fitted to the QM data for dimethyl sulfide and then adopted to model rotation of the methyl group in methyl-ethyl sulfide. Both trans$^-$ and gauche$^-$ (configuration of the C1C1−SC1 (methyl-ethyl sulfide) fragment) ab initio profiles are reproduced well with RMSDs of 0.07 and 0.28 kcal/mol, respectively. Since the barrier to the methyl rotation is somewhat overestimated for the gauche conformation (0.13 kcal/mol higher than the ab initio value), we calculated another set of torsional parameters, this time by fitting the QM data for methyl-ethyl sulfide rather than dimethyl sulfide. The calculations that were repeated using these new parameters did not show any significant improvement of the barrier values or the RMSD (the latter decreased by only ~0.08 kcal/mol). The relatively high RMSD (0.43 kcal/mol) was obtained for the methyl-ethyl sulfide torsional profile corresponding to rotation about the middle C−S bond (CC−SC torsional angle). The parameters derived for rotation of the $S-CH_3$ methyl group in methyl-ethyl sulfide by fitting QM data for methyl ethyl sulfide appeared to be transferable (RMSD = 0.13 kcal/mol) to the rotation of the corresponding methyl group in methyl-ethyl disulfide.

The ab initio torsional profile for the CS−SC rotation in methyl-ethyl disulfide was used for deriving the corresponding torsional parameters and was reproduced accurately (RMSD = 0.29 kcal/mol) with the resulting empirical potential. According to the QM data obtained in this work, methyl-ethyl disulfide can have three stable conformations corresponding to rotation about the C−S bond (CC−SS torsional angle). The stability order for these three energy minima is gauche$^-$ > trans > gauche$^+$ with the gauche$^+$ minimum being only marginally (0.02 kcal/mol) lower than the trans one. The empirical torsional parameters derived by fitting to the QM data yielded the correct order of the minima and a low RMSD (0.23 kcal/mol) between the whole QM and MM torsional profiles (computed in 10° steps).

The resulting parameters for the torsional potential for all types of torsional angles are given in Table 13.

3. $\phi-\psi$ Maps for Blocked Alanine and Glycine: Backbone Parameter Refinement. The torsional parameters for side-chain torsional angles and the backbone peptide bond were derived in the previous section; therefore, in this section we focus on obtaining the ϕ and ψ torsional parameters. We considered terminally blocked alanine as a model molecule for deriving

TABLE 13: Parameters of the Torsional Potential (Equation 3, in kcal/mol)

torsional angle	molecule	k_θ^1	k_θ^2	k_θ^3
HC−C1−C1−C1	propane	0.000	0.000	1.427
C1−C1−C1−C1	butane	−1.155	0.759	1.083
C1−C1−C1−C1	2-methyl butane	1.051	−0.745	0.697
C1−C1−C1−C3	*n*-butylamide	−0.288	0.106	1.034
C1−C1−C1−C2	butyrate	−0.656	−0.412	1.124
C1−C1−C1−N1	aliphatic amines	−0.707	0.491	1.499
C1−C1−C1−OH1	aliphatic alcohols	−1.216	0.428	1.503
HC−C1−C1−NB	alanine	0.000	0.000	1.200
C1−C1−CA−CA	tryptophan	0.000	0.000	−0.679
C1−C1−CA−NA	4-ethyl-imidazole (δ)	0.134	0.425	0.432
	5-ethyl-imidazole (ϵ)	−0.566	0.433	0.508
	4(5)-ethylimidazole cation (+)	0.023	0.799	0.307
HC−C1−CA−CA	toluene	0.000	0.000	0.001
C1−C1−CA−CA	ethylbenzene	0.274	0.782	0.114
HO1−OH1−C1−C1	ethanol	−0.314	0.026	0.633
HO1−OH1−CA−CA	tyrosine	0.000	2.251	0.000
C1−C1−N1−HN1	ethylamine	−0.260	0.201	1.146
HN1−N1−C1−HC	methylamine	0.000	0.000	1.088
C1−C1−C2−OH2	propanoic acid	−0.397	0.796	0.126
C1−C2−OH2−HO2	acetic acid	−1.545	4.712	0.226
C1−C2−OH2−HO2	propanoic acid	−2.638	4.926	0.181
C1−C3−NB−C1	backbone ω	0.372	13.724	−0.168
HC−C1−NB−C3	*N*-methylacetamide	0.000	0.000	−0.171
C3−NB−C1−C3	backbone ϕ	−1.430	1.405	0.186
HC−C1−C3−NB	*N*-methylacetamide	0.000	0.000	−0.229
NB−C1−C3−NB	backbone ψ	−1.702	1.952	−0.460
C3−NB−C1−C3	backbone ϕ (glycine)	−1.986	1.987	−0.505
NB−C1−C3−NB	backbone ψ (glycine)	−1.571	1.850	−0.382
NB−C1−C3−NB	backbone ψ (proline)	1.363	0.725	−0.099
C1−C3−N2−HN2	propionamide	0.001	10.650	0.000
C1−C3−N2−HN2	acetamide	0.000	10.804	−0.010
C1−C1−C3−N2	propionamide (zero: C−C−C=O cis)	−0.177	0.086	−0.582
C1−C1−N1−HN1	ethylamine cation	0.000	0.000	1.166
C1−C1−C2−O1	propionate	0.000	1.027	0.000
C1−N1−CA−N1	1-methylguanidine	0.096	7.642	−0.182
N1−CA−N1−HN1	1-methylguanidine (−NH_2)	−0.020	4.375	−0.020
N1−CA−NA−HN1	1-methylguanidine (−NH)	−0.841	11.711	0.220
C1−C1−N1−CA	1-ethylguanidine	−5.355	2.049	−1.045
N1−CA−NA−HN1	1-methylguanidine cation	0.000	6.015	0.000
HC−C1−S−C1(S)	methyl-ethyl sulfide	0.000	0.000	0.796
C1−C1−S−C1	methyl-ethyl sulfide	−3.080	1.881	−0.059
C1−C1−C1−S	methyl-propyl sulfide	−1.142	0.781	0.950
C1−C1−S−S	methyl-ethyl disulfide	−1.557	0.759	0.463
C1−S−S−C1	methyl-ethyl disulfide	−1.127	−2.514	0.119

ϕ,ψ torsional parameters for all the amino acids except glycine. Glycine ϕ and ψ torsional parameters were obtained by fitting the ab initio map of terminally blocked glycine with the corresponding MM map. The set of QM energies computed according to the procedure described in the Methods section was used for fitting the ψ torsional parameters for the blocked proline.

First, we attempted to derive the ϕ,ψ torsional parameters (coefficients of the three terms of the Fourier expansion (eq 3) for ϕ and ψ torsional angles, respectively) by fitting an MM (ϕ,ψ) map of Ac-Ala-NMe to the corresponding ab initio (ϕ,ψ) map. To assess the quality of the fit, both the RMSDs computed for all the points of the map and the eight energy-minimized conformations of Ac-Ala-NMe (Table 14, the α_D, α', and β_2 nomenclature was taken from ref 68). General agreement with the ab initio energies of the main minima was not satisfactory, and the RMSD was quite high. These computations showed that four of the main minima, namely, C_5, β_2, α_R, and α_D, were not true minima of the new potential even though they are minima on the QM map. It should be noted that the majority of the molecules used for deriving the nonbonded parameters for amide nitrogen and hydrogen bound to it were primary amides (i.e., the amides containing nitrogen with two bonded hydrogens) that might have biased the resulting parameters. To check if this was the reason for the moderate performance of the new force field, we carried out a two-stage fitting with refinement of nonbonded parameters *A* and *B* (eq 2) for the backbone nitrogen and hydrogen and optimization of the ϕ and ψ torsional

TABLE 14: Conformers of Terminally Blocked Alanine: Relative Energies and RMSDs of ϕ and ψ Angles from the Ab Initio Data

		new force field	
conformer	ΔE (ab initio[a])	ΔE (kcal/mol)	RMSD[b] (deg)
C_7^{eq}	0.0	0.00	7.2
$C5$	1.80	0.66	5.6
C_7^{ax}	2.52	4.64	4.6
β_2	3.19	3.47[c]	
α_R	3.92	3.15[c]	
α_L	4.38	5.53	8.0
α_D	5.13	7.78[c]	
α'	6.15	5.30	7.4
RMS error		0.8	6.7

[a] MP2/6-31G**. [b] The RMSDs were computed according to the formula RMSD = $\sqrt{[\sum_i^N (x_i^{QM} - x_i^{MM})^2]/N}$, where x_i is both ϕ and ψ angles, and *N* is the sum of the number of ϕ and ψ values. [c] Results of constrained minimizations with ϕ and ψ angles kept fixed; no true minima.

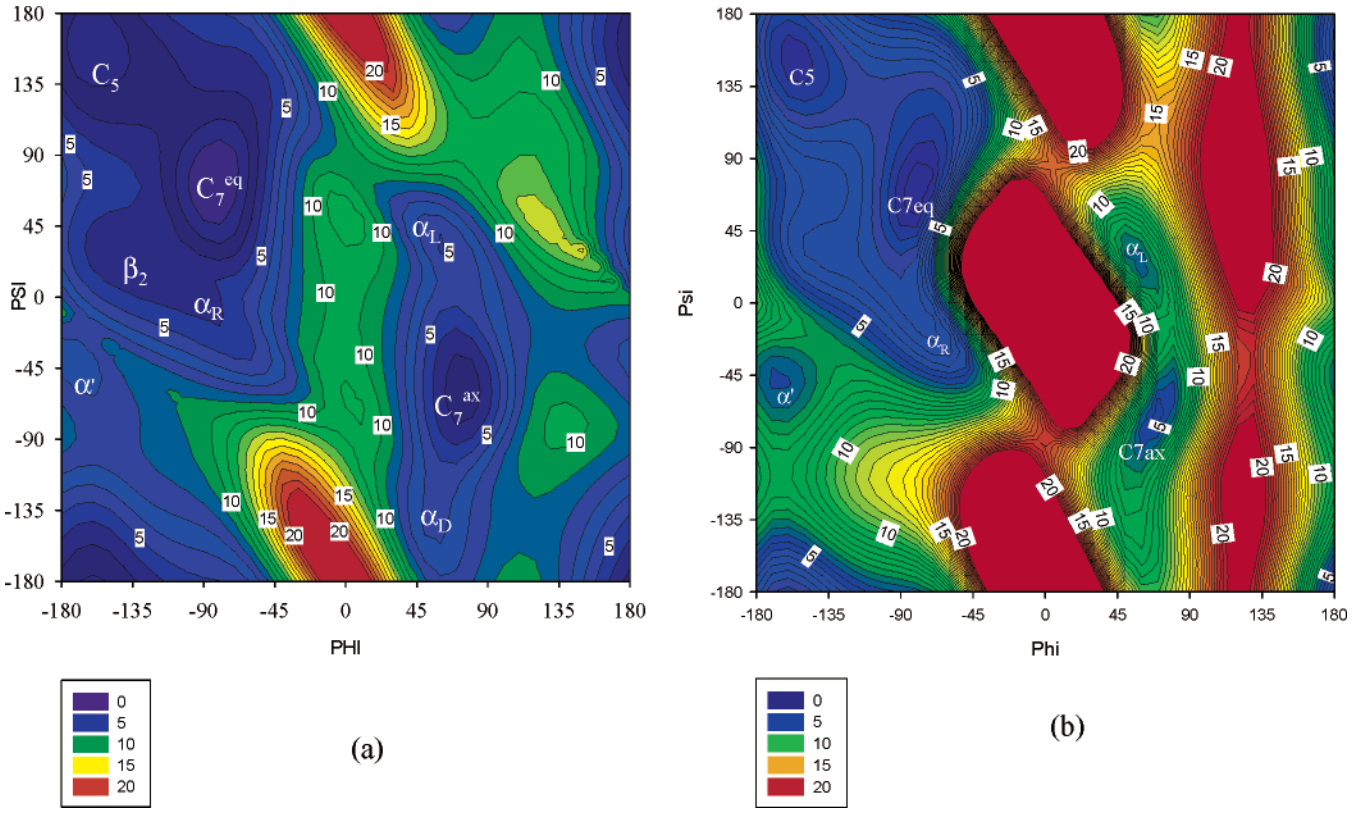

Figure 1. (a) QM and (b) MM maps for terminally blocked alanine. The QM energies were calculated using the MP2/6-31G**//HF/6-31G** methods. The α_D, α', and β_2 nomenclature was taken from ref 68. The color codes are in kcal/mol.

parameters as the first and second stage, respectively (see Methods). The nonbonded parameters were allowed to vary within ±30% of the initial values and the torsional parameters within ±2 kcal/mol. Several values of the weighting parameter c ranging from 0 to 2 were considered. The best results in terms of RMSD and agreement with relative stabilities of the main conformations of Ac-Ala-NMe were obtained for $c = 0.7$. The final RMS difference between the QM and MM energies in the region used for fitting ($-180° < \phi < -70°$; $-40° < \psi < 180°$ plus the α_L and α_D conformations) was 1.1 kcal/mol (the weighted RMSD was 0.3 kcal/mol with the weights calculated using the formula $w_i = \exp(-\Delta E_i^{QM} c)$).

Figure 1 shows the QM and MM energy maps of Ac-Ala-NMe computed using the backbone torsional and nonboned *A* and *B* parameters optimized by the two-stage fitting. Although only a part of the ab initio (ϕ,ψ) map was used for fitting, all the main features of the entire map are reproduced accurately. Notable differences include the shape and higher energy of the $\phi > 0$ low-energy region (containing C_7^{ax}, α_L, and α_D ab initio minima) in the MM map, which is a result of the rigid internal geometry used. Conformations from the $C_7^{ax}-\alpha_D$ region do not occur frequently in proteins;[48] therefore, a less accurate description of this region with the new force field should not cause concern. However, the α_L region corresponding to the backbone conformations observed for many residues in proteins is reproduced well.

Table 14 shows the energies obtained for terminally blocked alanine conformers using the optimized parameters and the ab initio MP2/6-31G** method. It can be seen that the relative MM energies of the most important conformations agree fairly well with the ab initio results. The β_2, α_R, and α_D conformations are not stable minima of the new force field and transform to the C_7^{eq} and C_7^{ax} conformations, respectively, as a result of energy minimization. The same behavior (Table 15) can be observed for the β_2 and α_L conformations minimized with the OPLS force field,[9] for the α_L, α_D, α_R, and α' conformations with AMBER(parm99)[4] and for the β_2, α_L, α_D, and α_R conformations with AMBER03.[5] The RMS difference in energy, calculated (excluding C_7^{ax} and α_D minima) using the new force field is 0.8 kcal/mol.

To assess the structural differences between the QM and MM conformations, we calculated the RMSDs of the ϕ and ψ angles (for the stable minima only) from the corresponding ab initio values. The average RMSD was 6.7°, and the largest deviation was 8.0° (for α_L) (Table 14). It should be noted that different QM values of the relative energies and the torsional angles have been reported[9,20] depending on the ab inito method and the basis set used. These angular differences can be more than 10° for the ϕ and ψ torsional angles.

For comparison, we also considered the results obtained for the eight conformers of Ac-Ala-NMe using the ECEPP/3, OPLS-AA/L, AMBER(parm99), and AMBER03 force fields (Table 15). Energy minimizations with the ECEPP/3 and AMBER force fields were carried out using the ECEPPAK and AMBER 8 program packages, respectively, while conformational energies and RMSDs for the OPLS-AA/L force field were taken from ref 9. The C_7^{eq} conformation was used as a reference for all the force fields. It can be seen that the new force field represents a significant improvement over ECEPP/3. Although all eight conformations correspond to minima of the ECEPP/3 force field, the agreement between their relative energies and the ab initio results is worse than for the force field reported in this work (1.6 vs 0.8 kcal/mol). The angular RMS differences are also higher for ECEPP/3 (RMSD of 14.7° compared to 8.8° for the new force field). It has to be noted that, although the ECEPP/3 force field employs the same geometry (with a different functional form of the force field), its parameters were not fitted to the high-level ab initio data

TABLE 15: Relative Energies and RMSDs of ϕ,ψ Angles (from the Ab Initio Data) Computed for Terminally Blocked Alanine with Different Force Fields

		new force field		ECEPP/3		OPLS-AA/L[a]		AMBER03		AMBER99	
conformer	ΔE (MP2/6-31G**)	ΔE[b]	RMSD[c]	ΔE[b]	RMSD[c]	ΔE[b]	RMSD[c]	ΔE[b]	RMSD[c]	ΔE[b]	RMSD[c]
C_7^{eq}	0.00	0.00	7.2	0.00	2.3	0.00	9.1	0.00	18.4	0.00	29.8
C_5	1.80	0.66	5.6	1.71	4.0	0.93	4.5	−0.19	5.0	0.90	7.4
C_7^{ax}	2.52	4.64	4.6	8.28	2.4	2.57	0.5	1.94	11.0	2.23	25.3
β_2	3.19	3.47[d]		0.99	17.9	6.68[d]		4.14[d]		2.58	5.9
α_R	3.92	3.15[d]		0.85	10.8			2.49[d]		2.53[d]	
α_L	4.38	5.53	8.0	2.21	10.1	3.27[d]		7.89[d]		6.33[d]	
α_D	5.13	7.78[d]		4.93	30.2			5.85[d]		8.56[d]	
α'	6.15	5.30	7.4	1.77	16.3	6.08	8.0	5.35	10.6	5.17[d]	
RMS error		0.82	6.7	1.60	14.7	0.43[a]	10.3	0.70	12.2	0.56	20.1

[a] The results were taken from ref 9. [b] Energies in kcal/mol. [c] RMSDs (in deg) of ϕ and ψ angles from the ab initio values (see footnotes to Table 14). [d] Results of constrained minimizations with ϕ and ψ angles kept fixed; no true minima.

TABLE 16: Terminally Blocked Ala_3: Relative Energies of the Conformers and RMSDs in ϕ_{1-3} and ψ_{1-3} from the Ab Initio Data

conformer		ΔE (ab initio[a])	ΔE (ab initio[b])	ΔE (ECEPP-05)	$RMSD_{\phi\psi}$[c]
$\phi < 0$	1	2.71	0.51	2.09	4.4
	2	2.84	0.64	2.38	5.7
	4	4.13	1.93	2.21	21.5
	5	3.88	1.68	0.71	9.6
	α_R[d]	8.40	6.20	6.94	
$\phi > 0$	3	0.00	−2.20	1.99	24.4
	6	2.20	0.00	0.00	4.8
	7	5.77	3.57	3.08	7.5
	8	4.16	1.96	5.04	9.2
	9	6.92	4.72	14.59	25.8
	10	6.99	4.79	2.89	33.6

[a] Ab initio (LMP2 cc-pVTZ(-f)//HF6-31G**) relative energies (in kcal/mol) computed with conformer number 3 as a reference (ref 20). [b] Ab initio relative energies (in kcal/mol) from the previous column recomputed with conformer number 6 as a reference. [c] RMSDs of ϕ_{1-3} and ψ_{1-3} angles, in deg (see footnotes to Table 14). [d] Ab initio energy of the regular α_R conformer of Ala_3 taken from ref 5. The corresponding MM energy was calculated with the torsion angles restrained to the QM geometry.

employed in this work, and thus, the value of this particular comparison is somewhat limited. Taking into account the rigid geometry used in this work, the new force field performs very well when compared to the flexible-geometry AMBER and OPLS-AA/L force fields, which were also parametrized recently[5,9] using QM energy maps. Both the energy and the angular RMS differences are comparable for all three force fields.

To test the transferability of the backbone torsional parameters, we also used them to carry out local energy minimizations of 10 conformations of Ac-$(Ala)_3$-NMe. In addition to the ab initio $\phi-\psi$ map of blocked alanine (Figure 1a), the QM energies of these 10 conformations reported by Beachy et al.[20] have been a popular test for all-atom force fields.[5,9,20] The conformational energies computed using the new force field and the corresponding QM energies are given in Table 16. All the conformers with $\phi > 0$ except conformer number 6 have at least one of the residues with ϕ and ψ angles in the $C_7^{ax}-\alpha_D$ region. As can be expected, these conformations have higher empirical energies and/or change significantly as a result of energy minimization. Conformations from the $C_7^{ax}-\alpha_D$ region do not occur frequently in proteins;[48] therefore, we focused our attention on the rest of the structures (conformers 1, 2, and 4−6). Since conformation 3, which has the lowest QM energy, is one of those high-energy conformers, we used conformation number 6 (the second lowest according to the QM calculations) as the reference. When conformers 3 and 7−10 were excluded, the RMSD of the energy was 1.1 kcal/mol, and general agreement with the ab initio data was found acceptable. We also considered the regular α-helical conformer[5] (for which the main-chain torsionals were restrained to the QM values during the minimization) to estimate the energy difference between α- and β-conformations. This difference is ∼5.6 kcal/mol for the new force field, which agrees well with the quantum-mechanical difference of 5.7 kcal/mol.[5] The resulting ϕ,ψ torsional parameters are given in Table 13.

The backbone torsional parameters for proline (coefficients of the three terms of the Fourier expansion (eq 3) for the ψ torsional angle) were computed by fitting the QM energies of the points lying in the vicinity of the two main minima (trans ($\omega° = -172.8°$; $\phi = -86.2°$; $\psi = 75.9°$) and cis ($\omega° = 7.9°$; $\phi = -76.5°$; $\psi = -23.9°$)) of Ac-Pro-NMe (with down puckering of the pyrrolidine ring). The best values of the parameters (Table 13) were obtained for the weight factor $c = 0.5$. The quality of the resulting parameters was assessed by energy minimization carried out by starting from the four conformations, namely, trans-down, cis-down, trans-up, and cis-up, with the ab initio optimized values of the ψ and $\omega°$ angles. These calculations yielded the order of relative stabilities, i.e., trans-down > trans-up > cis-down > cis-up, which is in line with the QM data. The MM energies of each conformation compare to the corresponding ab initio values as follows: $\Delta E_{QM} = 0.0$ and $\Delta E_{MM} = 0.0$ for trans-down, $\Delta E_{QM} = 1.9$ and $\Delta E_{MM} = 1.3$ for trans-up, $\Delta E_{QM} = 3.1$ and $\Delta E_{MM} = 3.6$ for cis-down, $\Delta E_{QM} = 4.3$ and $\Delta E_{MM} = 5.9$ kcal/mol for cis-up. Except for the noticeably higher MM energy for the cis-up conformation, the MM energies computed using the derived torsional parameters agree very well with the QM data obtained in this work as well as with those available from the literature.[69]

The ab initio $\phi-\psi$ map of terminally blocked glycine was first used to test the transferability of the alanine MM backbone torsional parameters. The RMS difference in energy over the entire map was high, and what is more important, the lowest-energy region was not reproduced. Therefore, we introduced additional ϕ,ψ torsional parameters derived especially for glycine by fitting the corresponding QM map. Again, several values of the weighting parameter c (0.0−2.0) were considered with the best RMSD in energy of 5.3 kcal/mol obtained for $c = 0.0$. The QM and the MM $\phi-\psi$ maps of blocked glycine are shown in Figure 2. As can be seen, the new force field reproduces the main features of the ab initio map accurately. ϕ,ψ Torsional parameters for glycine are given in Table 13.

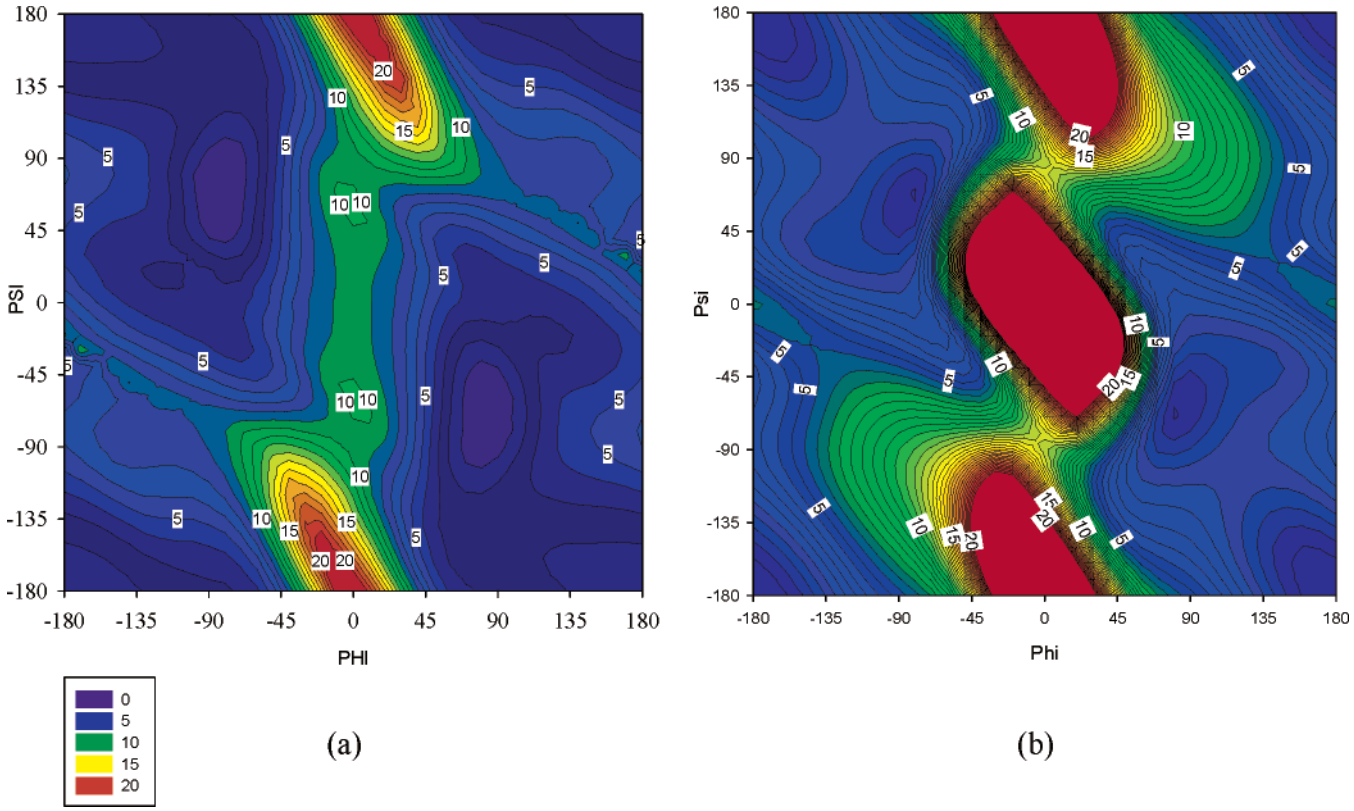

Figure 2. (a) QM and (b) MM maps for terminally blocked glycine. The QM energies were calculated using the MP2/6-31G**//HF/6-31G** methods. The α_D, α', and β_2 nomenclature was taken from ref 68. The color codes are in kcal/mol.

4. Performance of the Force Field in Modeling Crystal Structures of Amino Acids and Small Peptides. For a further test of the resulting force field, we carried out energy minimization of the experimental crystal structures of amino acids and small peptides. The results of the calculations are given in Table 17.

For all the crystal structures except Val-Gly-Gly, the experimental symmetry was preserved after the minimization. The average deviation of the unit cell parameters from their experimental values (Δ_{cell}, %) is traditionally used to evaluate the accuracy of a force field when applied to modeling crystal structures. A force field is considered as accurate when it leads to deviations of 5% or less. According to this measure, the new force field performs very well. On average, Δ_{cell} is about 3% with Δ_{cell} of all the structures except Val-Gly-Gly below 5% (a maximum of 4.8% was obtained for Ac-Phe-Tyr). It should be noted that the values of Δ_{cell} may be misleading in the case of flexible molecules. A low value of Δ_{cell} may be a result of significant changes in both molecular conformation and orientation within the unit cell that balance each other. Therefore, the values of two other measures, namely, $RMSD_{tor}$ and $RMSD_{cryst}$, are given in Table 17. The first of these reflects changes in the molecular conformation, whereas the second one provides an estimate of the changes in the crystal structure. The values of $RMSD_{cryst}$ are on average low (0.3−0.5 Å), which means that the crystal structures are reproduced with high accuracy. The RMS differences calculated for the torsional angles ($RMSD_{tor}$) show that, in general, molecular conformations are reproduced well. High $RMSD_{tor}$ values were obtained only for Gly-D,L-Ala (12.6°), Val-Gly-Gly (13.6°), Leu-Tyr (13.4°), and Ac-Phe-Tyr (12.7°). One of the ω angles of Ac-Phe-Tyr and Val-Gly-Gly has an unusually low experimental value (167.8° and 170° for Ac-Phe-Tyr[70] and Val-Gly-Gly,[71] respectively); hence, the large deviations obtained for these molecules may possibly be an indication of problems with their experimental structures.

The deviations obtained for tri- and tetrapeptides are comparable to those obtained for the smaller peptides. The $RMSD_{cryst}$ values are slightly higher, whereas the $RMSD_{tor}$ values are very similar. Considering all three measures presented in Table 17, the new force field performs very well.

The ability to reproduce the experimental geometry is only one of the requirements for an accurate force field. Another much more important requirement is the ability to locate nativelike structures as the lowest in energy. In general, crystal structure predictions carried out for small peptides might be used for assessment of the part of a force field describing nonbonded, electrostatic, and torsional interactions. Of course, this test will be valid only if the observed crystal structure is thermodynamically the most stable one. In contrast to small organic molecules, crystals of peptides are much less studied experimentally. They also possess a high degree of flexibility and contain many hydrogen-bond-forming groups that are the factors usually favoring formation of different crystal polymorphs. As a consequence, the CSD structure of a given peptide may not be the thermodynamically most stable one and, therefore, cannot be used for evaluation of a force field. It should also be mentioned that there are no reliable algorithms for crystal structure prediction of large flexible molecules. The Crystal Structure Prediction (CSP04) exercises[72] organized by the Cambridge Crystallographic Data Centre demonstrated that, although some progress has been achieved in the field in general, prediction of crystal structures of even small flexible molecules (with up to six torsional angles) represents a difficult and yet unsolved problem.

For proteins, the accepted paradigm is Anfinsen's thermodynamic hypothesis[73] according to which the native structure is the global minimum of the free energy of a protein plus the surrounding solvent. We have developed the all-atom force field to reproduce and predict the crystal structures of small organic molecules and the energy surfaces of peptides without consider-

TABLE 17: Results of Local Energy Minimization for Crystal Structures of Small Peptides

compound	CSD ref code[a]	space group	Δ_{cell}[b]	$RMSD_{tor}$[c]	$RMSD_{cryst}$[d]
		Dipeptides			
Gly-Gly	glygly04	$P2_1/c$, $Z = 4$	2.3	4.1	0.37
Gly-DL-Ala	glydla	$P2_1/c$, $Z = 4$	4.6	12.6	0.53
Gly-Ala	glyalb	$P2_12_12_1$, $Z = 4$	2.6	3.4	0.32
Ser-Gly	sergly	$P2_12_12_1$, $Z = 4$	3.6	5.1	0.64
Asp-Ala	fumtai	$P2_1$, $Z = 2$	1.8	8.6	0.36
Glu-Gly	glugly	$P2_12_12_1$, $Z = 4$	2.3	8.2	0.47
Gly-Leu	glyleu10	$P2_1$, $Z = 2$	2.6	8.1	0.75
Gly-DL-Phe	gldlpa	*Pbca*, $Z = 8$	1.5	4.9	0.25
Leu-Glu	bofzol	$P2_12_12_1$, $Z = 4$	1.1	4.7	0.18
Val-Glu	cijgux	$P2_12_12$, $Z = 4$	1.9	5.0	0.22
Leu-Tyr	jukmen	$P2_12_12_1$, $Z = 4$	3.9	13.4	0.91
Ac-Ala-Ala-NH_2	acaaam	$P2_12_12_1$, $Z = 4$	0.8	4.4	0.15
Ac-Phe-Tyr	acfy	$P2_1$, $Z = 2$	4.8	12.7	1.15
Ac-Gly-Ala-NH_2	acgaam	$P2_1$, $Z = 2$	1.3	6.9	0.21
Ac-Val-Leu	acvl	$P2_1$, $Z = 2$	3.2	7.9	0.54
		Blocked Amino Acids			
Ac-DL-Phe-NMe	acdlfnm	$P2_1$/c, $Z = 4$	1.9	4.0	0.28
Ac-Phe-NMe	acfnm	$P2_1$, $Z = 2$	3.1	5.8	0.34
Ac-DL-Trp-NMe	acdlwnm	$P2_1$, $Z = 2$	3.8	9.5	0.96
Ac-Trp-NMe	acwnm	$P2_12_12_1$, $Z = 4$	2.7	4.3	0.30
Ac-Tyr-NMe	acynm	$P4_1$, $Z = 4$	0.5	6.1	0.23
Ac-Trp-OMe	acwme	$P2_12_12_1$, $Z = 4$	3.6	5.8	0.36
Ac-Tyr	actyr	$P2_1$, $Z = 2$	2.3	7.7	0.36
Ac-DL-Val	acdlval	*Pbca*, $Z = 8$	2.6	5.8	0.31
Ac-DL-Met	acdlmet	$P2_1$/c, $Z = 4$	3.4	5.7	0.46
		Tri- and Tetrapeptides			
Phe-Gly-Phe-Gly	bajpim	$P2_12_12_1$, $Z = 4$	2.2	5.2	0.44
Val-Gly-Gly	copbis10	*C*2, $Z = 4$	6.2	13.6	0.57
Leu-Gly-Gly-Gly	duvfij	$P2_12_12_1$, $Z = 4$	2.5	7.2	0.65
Tyr-Ala-Phe	iwanid	*C*2, $Z = 4$	1.7	6.0	0.46
Ala-Gly-Ala	wirniu	$P2_12_12_1$, $Z = 4$	2.0	5.6	0.24

[a] Cambridge Structural Database reference code.[58] [b] Average percent deviation of the unit cell parameters from their experimental values, computed using eq 7. [c] RMSDs of the torsional angles (in deg). [d] Heavy-atom RMSDs (in Å).

ing interactions with water. To make it applicable to protein simulations, we will supplement it with a solvation model implemented in the ECEPPAK[29,74] package, i.e., a solvent-accessible surface area model,[75] a generalized Born model of Hawkins et al.,[76] or a model based on a solution of the Poisson equation as given by the multigrid boundary element method.[19] The resulting all-atom force field will be tested as a scoring function on a large set of structures including native and non-native conformations of several proteins. Depending on the outcome of these computations, the parameters of the force field (including those of the solvation model of choice) may be refined to enable distinguishing nativelike from non-native structures of proteins.

Conclusions

The development and testing of a new force field (ECEPP-05) based on ECEPP geometry (fixed bond lengths and bond angles) has been described. The derivation of the charges and torsional parameters was based on high-level quantum-mechanical data. The partial atomic charges for the 20 naturally occuring amino acids were derived using multiple-conformation fitting to the ab initio electrostatic potential.

We studied the influence of scaling of the 1−4 nonbonded and electrostatic interactions on the accuracy of the force field. It was shown that scaling of the nonbonded interactions is not necessary for our potential. The results are not very sensitive to a particular value of the electrostatic scale factor when small molecules, such as propanol, propylamine, and 1,2-ethanediol are considered. However, the effect of the scaling becomes evident for the larger molecule (1,3-propanediol). We found that our force field performs best when no scaling of 1−4 electrostatic interactions is applied. It should be emphasized that the charge model has a stronger effect on both the conformational energies and the torsional profiles of the smaller model molecules than the scaling of electrostatic interactions.

The torsional energy parameters were derived to reproduce gas-phase structures and conformational energetics from ab initio MP2/6-31G** calculations. Transferability of the torsional parameters was tested on a large set of model molecules. In the majority of cases, the quality of fits was very high, with an average RMSD for torsional profiles of ~0.2 kcal/mol. The main-chain backbone torsional parameters of peptides were obtained by fitting to the $\phi-\psi$ MP2/6-31G**//HF-6-31G** energy surfaces of terminally blocked alanine and glycine. The ab initio $\phi-\psi$ map of alanine was also used for refitting the nonbonded parameters for backbone nitrogen and hydrogen. The resulting overall energy deviations from the MP2/6-31G**//HF/6-31G** data for the regions of the $\phi-\psi$ maps used in fitting were 1.1 and 5.3 kcal/mol for terminally blocked alanine and glycine, respectively. Fitting QM data for blocked proline enabled us to obtain the backbone torsional parameter that accurately describes the relative stability of the main conformations of this terminally blocked amino acid.

Comparison with other force fields employing either rigid (ECEPP/3) or flexible (OPLS-AA/L, AMBER) geometry demonstrated that the new force field performs much better than ECEPP/3 and provides an accuracy comparable to that of AMBER or OPLS-AA/L when applied to energy minimization of the conformers of terminally blocked alanine.

We have demonstrated transferability of the Fourier coefficients obtained by the fitting procedure involved. This was done by using the backbone torsional parameters derived for alanine to compute the conformational energies of the terminally blocked Ala_3 peptide. Moreover, applying the new torsional parameters and partial atomic charges to the energy minimizations of a large set of peptide crystals enabled us to reproduce the observed structures accurately including both the unit cell parameters and the molecular conformations.

The success of our efforts to produce an accurate ECEPP-type force field was ensured by combining accurate nonbonded parameters derived using a novel global-optimization-based method[30,31,42] and a torsional fitting technique, which utilized high-level ab initio data as a target. Our future goal is to optimize our force field for protein structure prediction. In the previous work,[30,31] we demonstrated that the current force field was successful in crystal structure prediction; i.e., it was able to predict experimental crystal structures as the global energy minima. In this work, the new force field with rigid geometry was able to satisfy the experimental crystal structures of amino acids and peptides (average heavy-atom RMSD = 0.42 Å). When applied to calculations of peptides in the gas phase, it provided accuracy similar to that of force fields with flexible geometry. Therefore, we consider the force field presented in this work as a good starting point for its future optimization for proteins.

Acknowledgment. We thank Dr. J. A. Chisholm for providing us with the COMPACK program. This research was supported by a grant from the National Science Foundation (MCB00-03722). This work was carried out using computational resources provided in part by (a) the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center and (b) our own array of 338 dual-processor PC computers.

References and Notes

(1) Brooks, C. L., III.; Karplus, M.; Pettitt, B. M. *Proteins, A Theoretical Perspective Dynamics, Structure, and Thermodynamics*; Advances in Chemical Physics 71; John Wiley and Sons: New York, 1988.

(2) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: New York, 1989.

(3) Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472.

(4) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(5) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput.Chem.* **2003**, *24*, 1999.

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

(7) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulations: The GROMOS96 Manual and User Guide*; BI-OMOS: Zürich, 1996.

(8) Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J.-R.; Hwang, M.-J.; Li, H.; Liang, C.; Maple, J.; Peng, Z.; Stockfisch, T. P.; Thacher, T. S.; Yan, L.; Ni, X.; Hagler, A. T. *J. Comput. Chem.* **2001**, *22*, 1782.

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

(10) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740.

(11) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258.

(12) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102.

(13) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548.

(14) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711.

(15) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14812.

(16) Ripoll, D. R.; Vila, J. A.; Scheraga, H. A. *J. Mol. Biol.* **2004**, *339*, 915.

(17) Felts, A. K.; Gallicchio, E.; Wallqvist, A.; Levy, R. M. *Proteins: Struct., Funct., Bioinf.* **2002**, *48*, 404.

(18) Lee, M. C.; Duan, Y. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 620.

(19) Vila, J. A.; Ripoll, D. R.; Arnautova, Y. A.; Vorobjev, Y. N.; Scheraga, H. A. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 56.

(20) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908.

(21) Gnanakaran, S.; Garcia, A. E. *J. Phys. Chem. B* **2003**, *107*, 12555.

(22) Yoda, T.; Sugita, Y.; Okamoto, Y. *Chem. Phys.* **2004**, *307*, 269.

(23) MacKerell, A. D., Jr. *J. Comput.Chem.* **2004**, *25*, 1584.

(24) Ripoll, D. R.; Vorobjev, Y. N.; Liwo, A.; Vila, J. A.; Scheraga, H. A. *J. Mol. Biol.* **1996**, *264*, 770.

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

(26) Némethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 1883.

(27) Sippl, M. J.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1984**, *88*, 6231.

(28) Scheraga, H. A. *Rev. Comput. Chem.* **1992**, *3*, 73.

(29) Ripoll, D. R.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1283.

(30) Arnautova, Y. A.; Jagielska, A.; Pillardy, J.; Scheraga, H. A. *J. Phys. Chem. B* **2003**, *107*, 7143.

(31) Jagielska, A.; Arnautova, Y. A.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 12181.

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

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

(34) Williams, D. E.; Starr, T. L. *Comput. Chem.* **1977**, *1*, 173.

(35) Lifson, S.; Hagler, A. T.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5111.

(36) Cox, S. R.; Hsu, L.-Y.; Williams, D. E. *Acta Crystallogr., Sect. A* **1981**, *37*, 293.

(37) Williams, D. E.; Cox, S. R. *Acta Crystallogr., Sect. B* **1984**, *40*, 404.

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

(39) Williams, D. E. *J. Mol. Struct.* **1999**, *485*, 321.

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

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

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

(43) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. GAMESS6. *J. Comput. Chem.* **1993**, *14*, 1347.

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

(45) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357.

(46) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 389.

(47) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235.

(48) Hovmöller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr., Sect. D* **2002**, *58*, 768.

(49) (a) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503. (b) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 275.

(50) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.;

Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

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

(52) Roterman, I. K.; Lambert, M. H.; Gibson, K. D.; Scheraga, H. A. *J. Biomol. Struct. Dyn.*. **1989**, *7*, 421

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

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

(55) Gibson, K. D.; Scheraga, H. A. *J. Phys. Chem.* **1995**, *99*, 3752.

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

(57) Chisholm, J. A.; Motherwell, S. *J. Appl. Crystallogr.* **2005**, *38*, 228.

(58) Allen, F. H.; Kennard, O.; Taylor, R. *Acc. Chem. Res.* **1983**, *16*, 146.

(59) Sorin, E.; Pande, V. S. *J. Comput. Chem.* **2005**, *26*, 682.

(60) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

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

(62) Billeter, M.; Howard, A. E.; Kuntz, I. D.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 8385.

(63) Smith, J. C.; Karplus, M. *J. Am. Chem. Soc.* **1992**, *29*, 801

(64) Durig, J. R.; Bucy, W. E.; Wurrey, C. J.; Carreira, L. A. *J. Phys. Chem.* **1975**, *79*, 988.

(65) Cornell, W. D.; Cieplak, P.; Bayly, C.; Kollman, P. A.. *J. Am. Chem. Soc.* **1993**, *115*, 9620.

(66) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(67) Drakenberg, T.; Forsén, S. *J. Chem. Soc. D* **1971**, 1404.

(68) Wang, Z.-X., Duan, Y. *J. Comput. Chem.* **2004**, *25*, 1699.

(69) Kang, Y. K. *J. Phys. Chem.* **1996**, *100*, 11589.

(70) Stenkamp, R. E.; Jensen, L. H. *Acta Crystallogr., Sect. B* **1973**, *29*, 2872.

(71) Lalitha, V.; Murali, R.; Subramanian, E. *Int. J. Pept. Protein Res.* **1986**, *27*, 472.

(72) (a) Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Lommerse, J. P. M.; Mooij, W. T. M.; Price, S. L.; Scheraga, H.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. *Acta Crystallogr., Sect. B* **2002**, *58*, 647. (b) Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko A.; Dunitz, J. D.; van Eijck, B. P.; Erk, J.C.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Torrissi, A.; Nowell, H.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Schweizer, B.; Verwer, P. *Acta Crystallogr., Sect. B* **2005**, *61*, 511.

(73) Anfinsen, C. B. *Science.* **1973**, *181*, 223.

(74) (a) Ripoll, D. R.; Pottle, M. S.; Gibson, K. D.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **1995**, *16*, 1153. (b) Ripoll, D. R.; Liwo, A.; Czaplewski, C. *T.A.S.K. Q.* **1999**, *3*, 313.

(75) Vila, J.; Williams, R. L.; Vasquez, M.; Scheraga, H. A. *Proteins: Struct., Funct., Genet.* **1991**, *10*, 199.

(76) (a) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122. (b) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.