Published in final edited form as:
J Chem Theory Comput. 2013 ; 9(9): 4046–4063. doi:10.1021/ct4003702.

The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins

Yue Shi[1,+], **Zhen Xia**[1,+], **Jiajing Zhang**[1], **Robert Best**[2], **Chuanjie Wu**[3], **Jay W. Ponder**[3], and **Pengyu Ren**[1,*]

[1]Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712

[2]Laboratory of Chemical Physics, NIDDK, National Institute of Health, Bethesda, MD 20892

[3]Department of Chemistry, Washington University in St. Louis, St. Louis, MO 63130

Abstract

Development of the AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Simulation) force field for proteins is presented. The current version (AMOEBA-2013) utilizes permanent electrostatic multipole moments through the quadrupole at each atom, and explicitly treats polarization effects in various chemical and physical environments. The atomic multipole electrostatic parameters for each amino acid residue type are derived from high-level gas phase quantum mechanical calculations via a consistent and extensible protocol. Molecular polarizability is modeled via a Thole-style damped interactive induction model based upon distributed atomic polarizabilities. Inter- and intramolecular polarization is treated in a consistent fashion via the Thole model. The intramolecular polarization model ensures transferability of electrostatic parameters among different conformations, as demonstrated by the agreement between QM and AMOEBA electrostatic potentials, and dipole moments of dipeptides. The backbone and side chain torsional parameters were determined by comparing to gas-phase QM (RI-TRIM MP2/CBS) conformational energies of dipeptides and to statistical distributions from the Protein Data Bank. Molecular dynamics simulations are reported for short peptides in explicit water to examine their conformational properties in solution. Overall the calculated conformational free energies and *J*-coupling constants are consistent with PDB statistics and experimental NMR results, respectively. In addition, the experimental crystal structures of a number of proteins are well maintained during molecular dynamics (MD) simulation. While further calculations are necessary to fully validate the force field, initial results suggest the AMOEBA polarizable multipole force field is able to describe the structure and energetics of peptides and proteins, in both gas-phase and solution environments.

INTRODUCTION

Proteins are a ubiquitous class of biopolymers whose functionalities depend on the details of their 3D structures, which are in turn encoded by their specific amino acid sequences. Computer modeling and simulations are widely utilized in the study of protein structure, function, dynamics, and interactions with other synthetic or biological molecules. In molecular mechanics (MM) approaches, interactions among atoms are described by classical empirical potentials often referred to as force fields. Unlike *ab initio* quantum mechanical

*Corresponding Author: pren@mail.utexas.edu.
+These authors contributed equally.

Supporting Information

This information is available free of charge via the Internet at http://pubs.acs.org/.

The authors declare no competing financial interest.

(QM) methods, the classical MM models treat atoms as rigid particles with electronic degrees of freedom averaged out, thereby lowering the computational cost and allowing simulation of biological events for larger systems and longer time scales. On the other hand, high level *ab initio* theory is becoming more affordable and is now heavily utilized during the development of classic potentials for proteins such as Amber,[1] CFF,[2] CHARMM,[3] GROMOS,[4] MM3[5] and OPLS-AA.[6] This class of force field typically utilizes fixed atomic charges, point dispersion-repulsion, and simple empirical functions for valence interactions. The current generation of force field has enjoyed much success in many areas of biological and materials science; however, there remains significant room for future improvement.

Efforts to advance molecular mechanics force fields to the “next-generation” have largely focused on introducing explicit electronic polarization into the electrostatic model. A number of comprehensive reviews on the history and development of polarizable force fields have detailed the significance of polarization effects.[7-15] A wide range of studies on water,[16-18] organic molecules,[19-23] peptides,[24-25] protein-ligand binding,[26-33] ions,[34-38] and ion channels[39] using polarizable force fields have demonstrated various benefits of directly treating polarization effects. In addition to offering more accurate thermodynamic properties, a polarizable force field is more transferable, in principle, and can be more robustly parameterized by direct comparison with high-level *ab initio* quantum mechanical calculations in the gas phase.

Several different methods for incorporation of many-body effects have been explored. The fluctuating charge approach accounts for polarization by varying the magnitude of atomic charges based on electronegativity equalization.[40-47] It has been argued that fluctuating charge models fail for specific geometric situations, such as out of plane polarization and bifurcated hydrogen bonding, since charge flow is limited to bond directions.[48] Alternatively, the Drude oscillator or shell model, where a point charge moves about the nuclear position, has been applied to modeling of the induced dipole response in water and small molecule systems.[17, 49-53] Compared to the classical induced dipole method,[16, 19, 48, 54-57] these schemes involve less complex numerical algorithms since the point charge framework is retained. However, the interactive atomic induced dipole model [58-59] is superior in terms of reproducing anisotropy and nonadditivity of molecular polarization response across many different types of compounds. Moreover, intramolecular polarization assumes a critical role upon moving from small molecules to larger peptides and proteins possessing alternative conformational states. The conformational dependence of electrostatics can be significant,[60-61] and has received attention in the design and development of both polarizable and nonpolarizable force fields.[60, 62-64] As we have previously shown,[24] the interactive induction model used in AMOEBA can accurately reproduce this property.

In addition to polarization effects, we also want to stress the atomic partial charge-based representation of permanent electrostatics is itself inadequate. It has been shown that the error in a molecular electrostatic potential can be reduced by orders of magnitudes upon complementing atomic monopoles with dipole and quadrupole moments.[65-67] The incorporation of higher order atomic multipoles has been shown to greatly improve the quality of crystal structure predictions for simple organic molecules.[68-69] One may argue additional off-center charges are an alternative to point multipoles, as they can enable the same level of fidelity in description of electron density. For example, the use of charges at lone-pair sites of oxygen atoms can improve the ability of a water model to reproduce properties such as the density anomaly with temperature, and the dielectric constant.[70] Nevertheless, the determination of both position and magnitude of the charges at such extra sites is a nontrivial task, requiring fitting to the experimental density-temperature profile. In

contrast, the distributed multipole analysis of Stone[71-72] provides relatively unambiguous sets of atomic multipoles directly from molecular orbital calculations.

In this work, we present the development of a protein potential based upon a polarizable atomic multipole representation of electrostatics. The intramolecular polarization scheme formulated previously allows us to obtain the permanent atomic multipoles, directly from *ab initio* calculations on blocked dipeptides. A new protocol is applied to combine Distributed Multipole Analysis (DMA) and electrostatic potential fitting to derive conformation-independent "permanent" multipoles. The valence and van der Waals parameters have been derived from liquid simulations of small organic molecules containing similar functional groups. Merging of inter- and intramolecular interactions at short separation, including electrostatics, vdW, and torsional contribution, is determined via comparison to gas phase QM data for di- and tetra-peptides. The resulting potential is examined and validated by dynamics simulation of a number of peptides and proteins in solution, and comparison with experimental data. Compared to the preliminary AMOEBA protein force field parameters described in a previous article,[73] the current paper reports a completely new protein AMOEBA-2013 force field that is systematically developed using a newly improved strategy to parameterize permanent electrostatics and torsions, coupled with substantial validation against experimental data.

POTENTIAL ENERGY MODEL

The potential energy model has been explained in detail in several previous publications[18, 23]. Here we will briefly summarize the key features. The total energy of the system is given by

$$U = U_{bond} + U_{angle} + U_{cross} + U_{oop} + U_{torsion} + U_{vdW} + U_{ele}^{perm} + U_{ele}^{ind}. \quad (1)$$

The first five terms are the valence contributions corresponding to the bond, angle, bond-angle cross coupling, out-of-plane, and torsional energy, respectively. Common functional forms are used for these terms, and the detailed equations have been given previously.[73] Earlier versions of the AMOEBA protein force field included a torsion-torsion coupling term, implemented via a cMAP-style[74] two-dimensional bicubic spline. This term is essentially a grid-based correction to match the force field φ-ψ conformational energies to those of a target QM-based potential surface. In the current version, however, the "traditional" 3-term Fourier expansion function is used for all torsion angles, except for the backbone of glycine. The pairwise additive vdW interaction in AMOEBA is described by Halgren's buffered 14-7 function.[75] The buffered 14-7 function yields a slightly "softer" repulsive region than the Lennard-Jones 6-12 function, but a steeper repulsive wall than typical Buckingham exp-6 formulations. For a hydrogen atom connected to a heavy atom X, the vdW site is placed along the HX bond such that the distance between the atom X and vdW site of H is a percentage of the full bond length, referred to as the "reduction factor". This allows incorporation of some vdW anisotropy along the bond direction.

The permanent electrostatic energy in AMOBEA arises from atomic multipole-multipole interactions. Each atomic center consists of a point monopole (partial charge), a dipole vector and a quadrupole tensor. Note there are only five independent quadrupole components due to symmetry and use of traceless moments. The dipole and quadrupole are defined with respect to local reference frames formed by neighboring atoms. Examples of such local frames are illustrated in Figure 1. As the molecules rotate and diffuse over the course of a simulation, each atom's atomic multipoles maintain a constant orientation with respect to their local frame. The equation for calculating interaction energies and gradients

(forces and torques) between permanent multipoles, including a traditional Ewald summation method, was presented previously.[18]

Electronic polarization accounts for the majority of the many-body effects experienced in biomolecular systems although there are situations where many-body aspects of the dispersion and repulsion energies may be important as well.[76-77] AMOEBA utilizes an interactive atomic dipole induction scheme where the field produced by permanent multipoles and induced dipoles induces a dipole at each polarizable site, and each such induced dipole then further polarizes other atoms. As a result, anisotropic molecular polarizability can be effectively described by isotropic distributed atomic polarizabilities. Based on Thole's model,[59] polarization at very short range is damped, yielding energies in better agreement with *ab initio* results and avoiding the so-called polarization catastrophe. The same polarization model is used for both intermolecular and intramolecular polarization. A group-based scheme, where permanent multipoles do not polarize other atoms within their group, but induced-induced mutual polarization occurs between all atoms, allow us to merge molecular fragments to represent a larger chain molecule such as a protein. We have previously described the procedure used to extract intramolecular polarization from *ab initio* quantum mechanical calculations, resulting in "true" permanent atomic multipoles.[24] The computation of the polarization energy and gradient in molecules containing multiple polarization groups is detailed in our previous report on the AMOEBA force field for general organic molecules.[23]

The masking of short range intermolecular interactions is implemented by means of scale factors, which were determined by optimizing the transferability of conformational energies from alanine dipeptide to tetrapeptide, as discussed in the parameterization section. The final set of scale factors for interactions between permanent multipoles is 0.4 for the 1-4 interactions (separated by three bonds), 0.8 for 1-5 interactions, and zero (completely neglected) for any closer pair of atoms (1-2 and 1-3 interactions). The polarization energy between induced dipoles and permanent multipole moments is computed fully between atoms separated by three (1-4) or more bonds, and completely neglected for any closer separation. The analytical gradient of the polarization energy is nontrivial because the intramolecular "scaling" used in computing polarization interaction energies differs from the group-based scheme used during induced dipole generation. A derivation of the analytical polarization force resulting from the above masking scheme is detailed in the appendix our previous publication.[23]

A particle-mesh Ewald (PME) treatment of polarizable multipole interactions has been developed[78] and implemented in TINKER,[73] Amber[79] and OpenMM[80] for AMOEBA calculations. The latest version of TINKER has OpenMP shared-memory parallelization of AMOEBA simulations, AMBER (PMEMD) has the MPI parallel capability, while OpenMM is accelerated for GPU-based calculations. For typical systems, roughly half of the computational expense of AMOEBA is due to iterative calculation of the induced dipoles. For highly charged systems, the relative cost of dipole iteration can be higher (*e.g.* containing divalent ions). Recently we have implemented a new induced dipole solver modified from a conjugated-gradient method,[81] which significantly improves efficiency when tightly converged induced dipoles ($< 10^{-5}$ Debye RMS change) are required. In addition, we have introduced into TINKER a multiple time step (MTS) algorithm for AMOEBA molecular dynamics, in which nonbonded interactions including polarization are updated every 2-2.5 fs. This simple MTS algorithm allows use of longer MD time steps and yields improved computational throughput.

COMPUTATIONAL DETAILS

Ab initio calculations were performed using Gaussian 09[82] and Q-Chem 4.0.[83] Geometry optimization was carried out at the MP2/6-31G* level, which is computationally efficient and gives reliable molecular geometries. Initial atomic multipoles for dipeptide model compounds were derived at the MP2/6-311G** level using Stone's original DMA procedure.[71] Basis sets containing diffuse functions are avoided in the initial DMA determination as they lead to spurious multipole values, especially for buried atoms.[84] Note the original DMA procedure can be achieved in the current version of the GDMA program (v2.0 and above)[85] by setting "switch" to 0, and the "radius factor" to 0.65 for all atom types. The resulting atomic multipole values were then optimized against MP2/aug-cc-pVTZ electrostatic potential values computed on a grid of points around each model compound. MP2/aug-cc-pVTZ was used because our previous studies indicate inclusion of diffuse basis functions is important for capturing intermolecular interactions and reproducing experimental hydration free energy.[23] Partial charge values (monopoles) are held fixed during the potential optimization. All single point conformational energies were obtained with complete basis set (CBS) extrapolation from RI-TRIM MP2/cc-pVTZ and cc-pVQZ results.[86] These CBS calculations provide highly accurate conformational energies while still being computationally affordable. The TINKER v6 and Amber v10 molecular modeling packages were used for all the molecular mechanics calculations. Particle-Mesh Ewald summation[78, 87-88] was applied to treat electrostatic interactions, including polarization, with a real-space cutoff distance of 7.0 Å, a grid spacing of 0.8 Å, and 5^{th}-order B-splines. A cutoff with a switching window at 12 Å was applied to the vdW interactions. Induced dipoles were iterated to convergence, until the root mean square (RMS) change between interations fell below 0.01 Debye per atom. All molecular dynamics simulations were performed using a Velocity Verlet MTS integration algorithm[89] with a 2.5 fs time step. The system temperature was controlled via the Nose-Hoover chain thermostat method.[90]

For alanine, glycine, proline and some terminal capping groups (Ala-COOH, Gly-COOH, Gly-COO^-), the minimum energy map of the dipeptide was calculated on a uniform 15° grid in φ-ψ space. At each of the 576 points, MP2/6-31G* geometry optimization with constrained φ and ψ values was performed prior to a single point energy calculation at the RI-TRIM/MP2 CBS level. For proline, fewer grid points were available for QM calculations due to the limited conformational freedom. For the chain terminal amino acids, a single point *ab initio* energy was also calculated for each optimized structure using the Polarizable Continuum Model (PCM) solvation model. The torsion parameters for model compounds were initially fit to gas-phase *ab initio* conformational energies, and then adjusted to agree with statistical populations sampled from the Protein Data Bank (PDB). For the residue side chain torsions, geometry optimization was performed at the MP2/6-31G* level with each torsional angle constrained at 30° increments from 0° to 360°, followed by single point RI-TRIM MP2/CBS energy calculations.

The potential of mean force (PMF) of a solvated alanine tripeptide, NH_3^+-Ala-Ala-Ala-COO^-, with respect to φ and ψ angles of the central Ala residue was computed using the two-dimensional weighted histogram analysis method (2D WHAM).[91-93] A total of 576 independent molecular dynamics simulations of an alanine dipeptide and 206 water molecules in a 26.6 Å truncated octahedral box were carried out at 298K, using the same grid as for the gas phase map. In each simulation, the φ and ψ angles were restrained to a grid point on the Ramachandran map using weak harmonic potentials with a force constant of 0.01 kcal/mol-deg^2. The resulting conformer populations, sampled from the 576 × 70 ps production trajectories were utilized to construct the PMF and the relative free energy map via the 2D WHAM procedure.

An experimentally derived PMF was calculated from $-ln(P)$ where P is the torsion distribution as sampled from PDB statistics.[94] For the alanine backbone, the PDB PMF was obtained by averaging the data for tripeptides with a central alanine and either the right- or left-neighboring residue being alanine (*i.e.*, Ala-Ala-X or X-Ala-Ala, where X represents any residue type). For proline, PDB data with either a right- or left-neighbor residue being glycine (*i.e.*, Gly-Pro-X or X-Pro-Gly) was averaged. Similarly, the glycine PMF was calculated by averaging the PDB data for Pro-Gly-X and X-Gly-Gly.

For peptide systems, including unblocked and protonated $(Ala)_5$, NH_3^+-Gly-Pro-Gly-Gly-COO^-, and Ac-(Ala-Ala-Glu-Ala-Ala)$_3$–NH_2, replica exchange molecular dynamics (REMD)[95-96] simulations were performed using 36 replicas at temperatures between 278 K and 620 K. The $(Ala)_5$ peptide was unblocked, protonated at the N-terminus (-NH_3^+) and neutral at the C-terminus (-COOH), corresponding to the experimental conditions of pH 2.[97-98] For each system, the peptide was placed in a truncated octahedral box with approximately 800 water molecules. REMD simulations were performed under the NVT ensemble for at least 30 ns/per replica, using the PMEMD program from Amber 10.[79] with snapshots saved every 0.5 ps for analysis.

Protein molecular dynamics simulations were performed using the Amber10 software package. The proteins simulated include: crambin (PDB:1EJG, 46 residues),[99] Trp cage (PDB:1L2Y, 20 residues),[100] villin headpiece (PDB:1VII, 35 residues),[101] ubiquitin (PDB: 1UBQ, 76 residues),[102] GB3 domain (PDB:2OED, 56 residues),[103] RD1 antifreeze protein (PDB:1UCS, 64 residues),[104] SUMO-2 domain (PDB:1WM3, 72 residues),[103] BPTI (PDB: 1BPI, 58 residues),[105] FK binding protein (PDB:2PPN, 107 residues),[106] and lysozyme (PDB:6LYT, 129 residues).[107] For each system, a single protein molecule was placed in a truncated octahedral water cell chosen to provide a minimum distance of 8 Å between protein and cell wall. Simulations were performed in the NPT ensemble at a temperature of 298K. Approximately 10 ns of MD trajectory were generated for each system. The initial 500 ps portion of each trajectory was discarded corresponding to equilibration, and structural information was saved every 0.5 ps during the subsequent production phase. Analysis of the dynamics trajectories, including computation of RMSD values and average structures, was performed via the Amber Ptraj module.[79]

PARAMETER DERIVATION

Electrostatic Parameters

Permanent atomic multipoles for glycine, alanine and proline residues were derived from capped Ac-X-NMe dipeptides with X = Gly, Ala and Pro, with the goal of deriving conformationally independent electrostatic parameters for the central residue X from QM. A critical first step is definition of intramolecular direct polarization groups since the intramolecular polarization contribution needs to be extracted from the DMA multipole values as described in the Potential Energy Model section. Recall that an atom's permanent atomic multipoles (PAMs) only polarize atoms outside its polarization group. Figure 2 shows the group definition for alanine. For side chains of other residues, the groups are chosen in the same spirit, *i.e.*, no freely rotatable bonds within the group. For example, the -CH_2-, phenyl ring, and hydroxyl group are each classified as a single group for purposes of direct polarization. Next, the initial multipole parameters were derived by via DMA at the MP2/6-311G** level, and these parameters were then optimized against the MP2/aug-cc-pVTZ electrostatic potential computed on a set of grid points around the dipeptide compounds. Multiple grid layers were chosen at 0.35 Å intervals, beginning at a 1.0 Å distance from the vdW surface. Multipole parameters are iterated until the RMS difference from the target QM potential falls below 0.5 kcal/mol, and the RMS gradient is less than 0.5 kcal/mol/Å. Note the point charges are fixed at the original DMA values during the

optimization process, while the dipole and quadrupole moments are allowed to relax. The point charges at the boundary between the middle and capping residues, and between the backbone and side chain atoms are then manually adjusted to ensure an integral net charge for each residue. This adjustment is usually very small and the ESP optimization process ensures that charge adjustments are compensated by dipole and quadrupole changes. The resulting PAMs for the central residue are the "conformation independent" PAMs included in the AMOEBA-2013 parameter file. When connected to other residues in proteins, intramolecular polarization due to the PAMs takes place according to the polarization group definition. For alanine, we selected five local minima (α_L, α', C5, C7a and C7e conformers) and used all five conformers simultaneously in the above procedure. The multipoles from this multiple-structure potential fitting were used as the final set of permanent multipole parameters, and one additional conformer (β2) was used as validation by computing its electrostatic potential and dipole moments with the final parameters and comparing to QM results. PAMs for glycine and proline were obtained via an analogous protocol.

For all other residues, capped (Ac-X-NMe) dipeptides were chosen as the model compounds. For each amino acid, three conformations were selected for use in permanent multipole (monopole, dipole and quadrupole) parameterization, and another three for validation. The conformers have either an extended backbone, or represent compact structures with internal hydrogen bonds corresponding to gas phase minima. The backbone atoms were constrained to have the same PAM values as alanine, while side chain PAMs were derived using the procedure outlined above. When the side chain parameters were merged with alanine backbone parameters, charge neutralization (typically on the order of one-hundredth of an electron) was performed at the C_β atom. Dipole and quadrupole moments were fit to MP2/aug-cc-pVTZ electrostatic potentials for the three conformers in the parameterization set, allowing only the side chain values to vary. ESP fitting and validation results are discussed in detail in the Simulation and Validation section. In addition to Ac (CH_3CO-) and NMe (-$NHCH_3$) terminal groups, other N-termini (-NH_3^+ and -NH_2) and C-termini (-COOH and -COO^-) were also parameterized. Note glycine requires a separate parameterization for -NH_3^+ and -COOH termini because of its use of a different C_α atom type. Due to the amount of data involved, most parameters, including the PAMs and direct polarization group definitions, are not given in the text, but are included in the AMOEBA-2013 parameter files, which are publicly available from the TINKER web site at http://dasher.wustl.edu/tinker/.

vdW and Valence Parameters

The van der Waals and valence parameters were transferred from the AMOEBA parameters for small organic molecules, as reported in a previous publication.[23] The parameterization for these terms follows essentially the approach used for water[18] and ions,[34] where the vdW parameters were optimized to gas phase cluster structures and energetics as well as condensed-phase properties. A critical strategy in deriving the vdW parameters, due to their empirical nature, is to ensure chemical consistency among different elements. This was achieved by simultaneously parameterizing multiple compounds sharing the same vdW "classes" to improve transferability. The "atom classes" (super set of "atom types" used for electrostatics) in AMOEBA are subdivided beyond the simple chemical element types, largely to take into account different hybridizations such as sp^2 *vs.* sp^3. A selection containing the most common vdW parameters is provided in Table. The valence parameters, including bond, angle, bond-angle cross term and out-of-plane parameters, were transferred from small organic molecules with minor modifications to improve agreement with *ab initio* (MP2/6-31G*) geometries and protein PDB structures.

Torsional Parameters

The valence parameters (bond stretching, angle bending, bond-angle cross terms, out-of-plane bending) were transferred from organic small molecules, which were derived by matching the QM geometry and vibrational frequency of small moelcules.[108] The last step was then to derive the torsional parameters by comparing AMOEBA (everything but the to-be-fit torsion) and *ab initio* conformational energy values. Note that the molecular mechanics conformational energy depends not only on the torsional energy term, but also on the treatment of nonbonded intramolecular interactions – in particular on the treatment of 1-4 interactions across each torsion. In this work the scaling factors for the intramolecular electrostatic and vdW interactions have been chosen in order to minimize the contribution of the explicit torsional terms and ensure maximal transferability of parameters between dipeptides and tetrapeptides.

The alanine dipeptide is used to parameterize backbone torsional terms for all amino acids, with the exception of glycine and proline. The *ab initio* (RI-TRIM MP2/CBS) energy of alanine dipeptide was systematically evaluated at different backbone torsion angles over a 24 × 24 grid (15° intervals in both φ and ψ) as described in the computational details. The AMOEBA energy without the φ/ψ torsional contribution was computed for the same conformation while enforcing torsion constraints. The difference between AMOEBA and RI-TRIM MP2/CBS energy is taken as the fitting target for the torsional parameters, using a standard three-term Fourier expansion. Subsequently 2D WHAM simulations of an Ala-Ala-Ala tripeptide in explicit water were performed to obtain the Ramachandran potential of mean force (PMF) for the middle residue. The backbone torsion parameters were further improved by comparing the AMOEBA PMF in solution to a statistically-derived alanine backbone PMF derived from the PDB database.[94] We note that the torsion parameters were not directly fit to the PDB PMF. Instead, the parameter refinement was achieved by assigning relatively higher weight factors to the QM energy of gas-phase conformers located in the polyproline II (PII), α-helical and β-sheet regions while still fitting to the whole QM gas phase energy map. Torsional parameters were manually refined over 3-4 iterations to balance the relative populations of solvated peptide in these minimum energy regions. The gas phase Ramachandran potential energy from RI-TRIM MP2/CBS and AMOEBA with the final torsion parameters are compared in Figure 3. The simulated solution phase AMOEBA PMF and the PDB-derived statistical PMF maps for the alanine backbone are shown in Figure 4a.

The parameterization of proline backbone torsions followed essentially the same procedure as for alanine, except that fewer grid points were used due to the limited conformational freedom. For glycine, a torsion-torsion spline term is introduced in addition to the Fourier torsional terms for φ and ψ. After the Fourier torsional parameters were fit to the gas phase *ab initio* RI-TRIM MP2/CBS energy, the difference between the *ab initio* and AMOEBA energy was used to determine the torsion-torsion grid based correction table. (See Potential Energy Model for definition). These 2D values were fixed in the subsequent refinement of the Fourier torsional parameters. Use of the torsion-torsion term improves the fit to both *ab initio* data and solution phase properties. Similar to the parameterization of alanine backbone torsions, the torsional parameters for proline and glycine backbones were refined to match the statistical PMF for proline and glycine residues in the PDB database, respectively. REMD simulations were performed with a model tetrapeptide, GPGG (NH_3^+-Gly-Pro-Gly-Gly-COO^-), to obtain the simulated backbone torsion angle distribution of proline and glycine. All other residues share the same backbone torsion parameters (together with valence, vdW and electrostatic parameters) as alanine. Parameterization of the -COOH C-terminal form of alanine (and other non-Gly/Pro residues), and the −COOH and -COO^- terminal forms of glycine were fit to RI-TRIM MP2/CBS energies on a 12 × 12 torsional grid. Since there are limited PDB data available for terminal residues, we optimized these

parameters by matching AMOEBA energies in implicit solvent (based on Generalized Kirkwood model[109]) with QM PCM energies at the MP2/6-311G(2d,2p) level.

In addition, conformation energies for a benchmark set of 27 alanine tetrapeptides[110] have been assessed. This comparison was made to validate transferability and to adjust the scale factors for short-range intramolecular nonbonded interactions. The AMOEBA results are compared with those from MP2, LMP2, DFT and RI MP2 calculations in Table 3. All the *ab initio* calculations are single point energy evaluations at the same HF/6-31G** geometries. AMOEBA calculations were performed with both full geometry optimization, and with φ and ψ angles constrained to the HF geometry. All comparison is made against RI MP2/CBS results. While the AMOEBA-optimized structures deviate only slightly from those at HF/6-31G** (average SRMS = 0.47 Å), the RMS difference between AMOEBA and RI MP2/CBS energies is 1.15 kcal/mol, similar to those of LMP2/cc-pVTZ(-f) and MP2/6-311+G2d2p. Note that the relative conformational energies of the first two conformers (extended) *vs.* the third conformer (a compact structure) given by RI MP2/CBS lies between the canonical MP2 and LMP2 results, as do the AMOEBA results. The relative energy of extended and folded conformations is an important measure of the strength of intramolecular dispersion attraction is such systems.

For the side chain torsions for all other residues, the parameters were obtained by fitting to the RI-TRIM MP2/CBS conformational energy of the corresponding dipeptides. To derive parameters independent of the backbone conformation and to verify the transferability of the backbone torsional parameters for alanine, two or more dipeptide conformations with backbone φ/ψ values fixed at α-helix and β-sheet values were used simultaneously to fit the parameters of side chain χ torsions. For each conformer, the side chain torsional angle of interest was rotated in increments of 30°. At each point, the dipeptide structure was optimized with φ, ψ and χ angles constrained to the same values in both *ab initio* level and AMOEBA calculations. Examples of conformational energy profiles for isoleucine and serine side chains are plotted in Figure 5. The former has an aliphatic side chain, and a single set of torsional parameters performs well across both conformers. The energy order of the local minima is also captured by AMOEBA. The situation for serine is a slightly worse due to the additional complication of intramolecular hydrogen bonding between the side chain and backbone. In both cases, the torsional energy contribution accounts for about 15% of the total conformational energy. In Table, the average RMS error from the conformational energy fitting for each residue is listed.

SIMULATION AND VALIDATION

For small molecules, it is possible to systematically compare force field calculations over a range of *ab initio* data and experimental properties, which allows us to optimize the individual components in the potential energy model, such as the relative contributions from electrostatics and vdW interactions. For larger biomolecules, however, one can only compare detailed simulations with limited data such as database-derived statistical populations, NMR *J* coupling constants, and atomic structures of native proteins from X-ray or NMR experiments. Such comparison, while absolutely necessary, is most useful for inspecting the torsional parameters, and provides less feedback for other force field components (*e.g.* vdW or electrostatics). It will require more extensive investigation across a wide range of proteins research areas, such as protein folding, ligand binding, pKa shifts and mutant stability, to fully validate the different components and aspects of our proposed force field. Here we provide initial results through examination of the electrostatic parameters against QM for multiple conformations of each amino acid in the gas phase, and assessment of peptide conformational properties and protein structures in solution.

Electrostatic Properties in the Gas Phase

One of the important advantages of a polarizable force field is its transferability from gas phase to solution. Therefore, it is to our advantage that the gas phase electrostatic properties can be rigorously compared to QM *ab initio* results. It is especially important to ensure transferability of the alanine backbone electrostatic multipoles to other residues (except glycine and proline), as well as transferability among different conformations of each amino acid. We have validated the electrostatic properties of the AMOEBA protein model by computing the *x*-, *y*- and *z*-components of the dipole moment and the electrostatic potential of dipeptide model compound for each amino acid. For each amino acid dipeptide, three conformations were chosen as the validation set, and these structures were excluded from use in the parameterization process. In Figure 6, we compare the AMOEBA and QM molecular dipole moments of the dipeptide conformers in the validation set. The dipole moment xyz components of all the dipeptides are accurately reproduced, regardless of conformation and residue type, resulting in a correlation coefficient of 0.998. To the best of our knowledge, no other force field has demonstrated to be able to represent peptide electrostatic properties at different conformations with one set of electrostatic parameters to such accuracy. The comparison between *ab initio* and AMOEBA electrostatic potential for all dipeptide model compounds is detailed in the supporting information, for both the parameterization and validation sets. The average RMSE is 0.45 kcal/mol per unit charge on a grid surrounding the neutral amino acids, and only slightly higher (0.64) for charged dipeptides, with the absolute value of the potential for the latter being orders of magnitude higher. Thus, thanks to the intramolecular polarization model in AMOEBA, the transferability of backbone and side chain electrostatic multipoles of AMOEBA is quite satisfactory. We believe the test performed here, while not commonly used in force field development, is important and necessary for validating the electrostatics of a candidate force field model before other components such as the torsional parameters are empirically adjusted.

Polyalanine Conformational Free Energy in Solution

Several recent studies have used oligopeptide conformational properties in solution to calibrate force field torsional parameters.[98, 111-117] Simulation results can be directly compared to experimental nuclear magnetic resonance (NMR) data for the corresponding peptides. Building on this prior work, we have performed a series of simulations on Ala/Gly/Pro-based peptides using the AMOEBA protein force field.

For alanine, we have first examined the solvation of the unblocked and protonated $(Ala)_5$ peptide using REMD. Conformational preference is presented as a potential of mean force with respect to φ and ψ angles in Figure 4b, which is calculated from the averaged torsional population distributions of Ala-2, Ala-3 and Ala-4 residues. A distinct global minimum is located around the PII conformation. Two other basins with energies approximately 0.5 kcal/mol higher in energy lie in the β-sheet and α-helix regions. The energy barriers separating the global basin from the two local minima are about 1~2 kcal/mol. Overall, the upper left region of the Ramachandran map is relatively flat compared to the rest of the conformational space. The distribution within this highly populated region of the map agrees well with the statistical PMF drawn from the PDB database (Figure 4c),[94] suggesting transferability of the overall model from $(Ala)_3$ to $(Ala)_5$.

The φ/ψ torsional angle distributions for $(Ala)_5$ have been probed experimentally by NMR.[98] The spin-spin coupling (*J*-coupling) constants reflect the ensemble character of the conformational distribution, and were compared with those calculated form REMD simulation trajectories of $(Ala)_5$ via Karplus relations.[116, 118] In total, eight NMR *J*-coupling constants were reported: five for the backbone angle φ

[$^3J(H_N,H_\alpha)$, $^3J(H_N,C')$, $^3J(H_\alpha,C')$, $^3J(C,C')$, $^3J(H_N,C_\beta)$], two for the backbone angle ψ [$^1J(N,C_\alpha)$, $^2J(N,C_\alpha)$], and one value for both φ and ψ, [$^3J(H_N,C_\alpha)$].[98] The trajectory at 298K from REMD simulation of $(Ala)_5$ was extracted, and used to compute predicted *J*-coupling values. Twenty-seven predicted *J*-coupling values are compared to those measured by NMR in Table 4. The *J*-coupling constants involving the N- and C-termini were also included. The predicted values are in excellent agreement with those probed by experiment. The chi-square (χ^2) difference between simulation and experiment, computed using the experimental uncertainties,[98] is about 0.994, while the overall RMS difference is 0.33 (Table 4). Note when using torsional parameters directly fit to the gas-phase QM energy of alanine dipeptide, the χ^2 is 3 or 4 times higher. Torsion refinement based on both QM (gas-phase) and the PDB PMF (in solution) thus plays a significant role in improving the calculated *J*-coupling constants. A notable consequence of the refinement is that the location of the α-helix population from simulations shifts lower and to the right, toward the average (ϕ, ψ) angles in the PDB distribution. In contrast, with torsional terms fit to QM gas-phase energies alone, the simulated α-helical population was broader than that from the PDB and centered at a more negative ϕ and less negative ψ value. A similar effect has recently been discussed for the CHAMRMM22/CMAP, CHARMM36-MP2 and CHARMM36 force fields, and it was suggested empirical correction to CMAP approach is important.[113-114] Improving agreement with PDB ϕ/ψ angle distribution in terms of shape and location, especially for residues not in actual helices, led to thermodynamic properties and helix-coil transition cooperativity more consistent with experimental data.

Proline and Glycine Conformational Free Energies in Solution

The φ/ψ torsional angle distributions for proline and glycine backbones were validated through REMD simulations of the GPGG tetrapeptide (NH_3^+-Gly-Pro-Gly-Gly-COO^-). For both proline (Pro-2 residue) and glycine (Gly-3 residue), the simulated PMF maps with respect to the φ/ψ angles show good agreement with the PDB statistical PMF maps as shown in Figure 7. For Pro-2, the relative free energy of the α-helix and PII regions from the PDB data are well reproduced in our simulations, with the α-helical local minimum about 1 kcal/mole higher in energy than the global minimum in the PII region (Figure 7a). The simulated torsional distributions for the Gly-3 residue are also consistent with PDB data, with global minima located at the right- and left-handed α-helix regions. Two other local minima, about 1 kcal/mol higher in energy, are located at the PII structure and its mirror reflection (Figure 7c).

Similar to alanine, the *J*-coupling constants were also calculated for the Pro-2 and Gly-3 residues. Three *J*-coupling constants, $J(H_\alpha,C')$ for Pro-2, $J(H_\alpha,H_N)$ and $J(H_\alpha,C')$ for Gly-3 were evaluated using the Karplus coefficients obtained from B972/EPR-III and B3LYP/EPR-III (methods for DFT investigations of electron paramagnetic resonance) level calculations.[119] Table 5 compares the *J*-coupling values obtained from AMOEBA simulation and experiment for the GPGG tetrapeptide. The RMS difference between the calculated and experimental *J*-constants is 0.44 (with B972) and 0.39 (with B3LYP).

Secondary Structure Distribution for the Ac-$(AAQAA)_3$-NH_2 Peptide

It is important for a protein force field to accurately reproduce φ/ψ structural distributions and correctly balance peptide secondary structure, as these metrics are directly related to phenomena such as folding, misfolding, aggregation and conformational change. The REMD simulation of the small peptide $(Ala)_5$, described above, provides the sensitive test of the "intrinsic" secondary structure preferences of the AMOEBA force field. However, the $(Ala)_5$ peptide is too short to form a stable α-helix, and a more extended PII backbone conformation is found more favorable than the helix in the Ramachandran free energy map.[97-98, 116, 120-124] We have further simulated a longer helix-forming peptide, Ac-

$(AAQAA)_3$-NH2,[125-126] to investigate the helix-coil transition. Because the helical population of Ac-$(AAQAA)_3$-NH2 has been determined from NMR chemical shift data,[127] we can directly compare the AMOEBA helical-fractions calculated from REMD simulations. We use the same definition of the helical state as a previous study,[97] namely ϕ in the range [-160°, -20°] and ψ in [-120°, 50°]. The fraction of helix was evaluated as the frequency of the individual residues existing in the helical states during the whole trajectory at 303 K. Figure 10 shows the fraction of helix, $\langle h_i \rangle$, for each residue from our simulation and NMR chemical shift data. Reasonable agreement is seen between the simulated values and those obtained from NMR experiments at 303K, where both fluctuated between 10 to 30% for most residues. The error was estimated using block averages derived from six 5-ns simulation windows. In addition, the lower helical propensity trend at the C-terminus was well captured, while AMOEBA seems to underestimate the helix fraction for residues 3, 4 and 7.

Molecular Dynamics Simulations of Protein Systems

Ten well-studied proteins were chosen as the validation set to evaluate the new parameters. While limited, this set is representative of several different types of protein structures. For example, Trp cage (1L2Y) and villin headpiece (1VII) mainly consist of alpha helices; GB3 domain (2OED) and FK binding protein (2PPN) contain a significant proportion of beta sheet structure; crambin (1EJG) and lysozyme (6LYT) are disulfide bond-rich domains; the remaining proteins are mixtures of different motifs.

The stability of each protein was characterized by its backbone RMSD value relative to the PDB structure over 10 ns MD simulation, as summarized in Figure 8. The overall average RMSD of the ten simulated protein structures is 1.33 Å, and seven of them are close to 1.0 Å. The final MD trajectory snapshot for each protein is superposed on its PDB structure in Figure 9. We can see that helical and sheet motifs are well maintained, and do not significantly drift from the experiment structures. For many of the proteins, the large deviations arise from the terminal regions. For example, inspection of the PDB entry for ubiquitin shows the final five residues have significantly higher temperature factors than the rest of the molecule. Flexibility near the C-terminus is also observed in the MD results, particularly for the two glycine residues at positions 75 and 76. Elimination of just these two residues from the data analysis results in a backbone structural RMSD of 1.16 Å against the PDB structure.

Calculation of NMR Order Parameters

The order parameter (S^2) indicates the flexibility of each residue, with lower S^2 values corresponding to greater flexibility. We have compared simulated order parameters for ubiquitin and hen egg white lysozyme to NMR relaxation experiments[128-129] which measure the relaxation of amide ^{15}N-^{1}H dipolar interaction.[130-131] Relaxation is caused by fluctuations of interaction energies as the internuclear interaction vectors are reoriented by thermal motion, and therefore allows the observation of anisotropic local residue motions.[132] The isotropic reorientational eigenmode dynamics (iRED) matrix was extracted from the MD trajectories, and order parameters (S^2) were then computed from the eigenvalues of this matrix.[133]

Figure 11 compares of S^2 values computed from the 10 ns MD simulations with the corresponding NMR measurements. The calculated and experimental S^2 curves are generally well-correlated; the RMSDs between the two sets of values are 0.04 and 0.12 for ubiquitin and lysozyme, respectively. For ubiquitin, the high flexibility near turn 1 (residues 7-10), turn 3 (residues 37-40), turn 6 (residues 62-65), and the C-terminus was nicely reproduced. For lysozyme, flexibility at turn 1 (residues 46-49), long loop 2 (residues

61-78), loop 3 (residues 85-89), loop 4 (residues 100-107), loop 5 (residues 116-119) and the termini are in general agreement, although the computed values are somewhat higher than experiment. All the large deviations come from loop residues exposed to water. Specifically, the greater flexibility than what the experimental order parameters have indicated seems to suggest that the interactions between the 48D/61R and 101D/103N were underestimated by the model. However, the same types of residues give reasonable S2 values at other sites in lysozyme and ubiquitin. Additional evidence is necessary in order to decide whether we need to and how to improve the parameters for these residues. But overall, the AMOEBA force field appears able to reproduce protein structures and flexibilities.

Calculation of Side Chain J-Couplings

Amino acid side chain distributions were evaluated by comparing the simulated side chain *J*-couplings to the experimental NMR values for four protein systems: BPTI, GB3 domain, ubiquitin and hen egg lysozyme.[129, 134-137]

Table 6 shows the side-chain RMSD between the experimentally derived and simulation-based scalar couplings. Also listed for comparison are the RMSD values obtained from the Amber ff99SB force field and Amber ff99SB-ILDN,[138] which were refined against the side chain NMR data of the four proteins. Note the side chain parameters for AMOEBA were directly from fitting to QM energy profile and were not optimized against the experimental data. Overall, the performance of AMOEBA force field is comparable to Amber ff99SB-ILDN. Although a few outliers are evident for each protein system, most of the experimental scalar couplings are reasonably reproduced. The correlation between the calculated and experimental *J*-couplings of all the four protein systems is summarized in Figure 12. The overall R^2 correlation coefficient is 0.75. Details of the *J*-coupling values for each protein can be found in the Supporting Information. Within high scalar-coupling regions, AMOEBA failed to distinguish subtle differences observed in experiments. For example, most AMOEBA-predicted *J*-coupling values were ~10.8 Hz, while experimental data range from 8.0 to 15.0 Hz. We further subdivided the RMSD values by amino acid residue. The values ranged from ~0.7 to 3.3 (Figure 12). For residues with relatively high RMSDs, like Ile and Gln, we did not attempt force field modifications since the sample size for these residues is small in the four proteins studied. The side chain *J*-coupling data suggests the AMOEBA force field, derived from gas phase QM energy profiles, reasonably captures side chain conformational properties in solution. Further improvement would require additional reliable experimental data for individual residue types.

CONCLUSIONS

The development and parameterization of the AMOEBA-2013 protein force field is been reported. A distinct feature of the force field is its use of atomic multipole-based electrostatics and explicit treatment of dipole polarization. A mutual induction model with Thole damping is applied to describe both intra- and intermolecular polarization in a consistent manner. The polarization among permanent multipoles is handled via a group-based scheme while induced-induced dipole polarization occurs amongst all polarizable sites. By extracting the intramolecular polarization as defined by the force field, we are able to derive conformation-independent electrostatic multipole parameters from high-level *ab initio* calculations using a combination of Distributed Multipole Analysis and electrostatic potential fitting. This rigorously derived electrostatic model will be important for accurate description of protein interactions with other biomolecules as well as electrostatic forces within protein molecules.

After determination of vdW parameters, largely obtained by transferred from liquid simulations of small organic molecules,[23] the AMOEBA model was applied to simulation of several peptides and proteins. In addition to the nonbonded electrostatic and vdW forces, the "torsional" energy also plays an important role in setting relative conformational energies. The torsion term is essentially an error function in most classical force fields, and yet it plays a crucial role in determining the detailed conformational properties of peptides and proteins. Recent improvements to the Amber [111, 138] and CHARMM force fields[113-114] have demonstrated that conformational populations of small peptides are extremely sensitive to subtle changes (involving fractions of a kcal/mol) in the torsional parameters. In developing the AMOEBA force field, we have resorted to both high-level *ab initio* (MP2/CBS) peptide energetics and PDB structural statistics in deriving the backbone torsion parameters. The resulting force field performs well in the reproduction of NMR *J* coupling constants and S^2 order parameters of several peptide and protein systems. Nonetheless, these are limited validations focusing on conformational properties and torsional parameters. Further extensive investigations involving more proteins and a broad range of thermodynamic properties will be necessary to fully explore various aspects of the potential energy model, and to determine the overall success and possible deficiencies of the force field.

As previously noted,[113, 139] a CMAP-style splined 2D torsional potential allows a force field to reproduce gas-phase *ab initio* conformational energies exactly. Such CMAP terms may also exacerbate unphysical features in the rest of the force field and therefore are not transferable. For example, the QM and force field difference could result from an incorrect valence energy bin classical force field at extremely stressed conformations, an effect which has little to do with the "torsion" energy. We plan to revisit use of CMAP correction terms in the future. However, with the exception of Gly, we have removed such terms from the AMOEBA-2013 protein model. While we have strived to derive a balanced and physical force field, better understanding of the limitations of molecular mechanics force fields is essential for continued systematic improvement. Accurate modeling of short-range electrostatic and vdW components, as well as their coupling to valence interactions, will be the key to achieving still greater accuracy and transferability in a future protein force field that performs well in various physical and chemical environments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to the support by the National Institute of General Medical Sciences (R01GM079686 and R01GM106137) and Robert A. Welch Foundation (F-1691). JWP wishes to acknowledge support from National Science Foundation awards NSF CHE-1152823.

References

1. Cornell WD, Cieplak P, et al. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. J Am Chem Soc. 1995; 117(19):5179–5197.
2. Maple JR, Hwang MJ, et al. Derivation of Class-Ii Force-Fields. 1. Methodology and Quantum Force-Field for the Alkyl Functional-Group and Alkane Molecules. J Comput Chem. 1994; 15(2): 162–182.
3. MacKerell AD, Bashford D, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B. 1998; 102(18):3586–3616.
4. Wang DQ, Freitag F, et al. Validation of the GROMOS 54A7 Force Field Regarding Mixed alpha/beta-Peptide Molecules. Helv Chim Acta. 2012; 95:2562–2577.

5. Allinger NL, Yuh YH, et al. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. J Am Chem Soc. 1989; 111(23):8551–8566.
6. Jorgensen WL, Maxwell DS, et al. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J Am Chem Soc. 1996; 118(45): 11225–11236.
7. Rick SW, Stuart SJ. Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. Rev Comp Ch. 2002; 18:89–146.
8. Ponder JW, Case DA. Force Fields for Protein Simulations. Adv Protein Chem. 2003; 66:27–85. [PubMed: 14631816]
9. Warshel A, Kato M, et al. Polarizable Force Fields: History, Test Cases, and Prospects. J Chem Theory Comput. 2007; 3(6):2034–2045.
10. Lopes PEM, Roux B, et al. Molecular Modeling and Dynamics Studies with Explicit Inclusion of Electronic Polarizability: Theory and Applications. Theor Chem Acc. 2009; 124(1-2):11–28. [PubMed: 20577578]
11. Cieplak P, Dupradeau FY, et al. Polarization Effects in Molecular Mechanical Force Fields. J Phys-Condens Mat. 2009; 21(33):333102 1–21.
12. Gresh N, Cisneros GA, et al. Anisotropic, polarizable molecular mechanics studies of inter- and intramoecular interactions and ligand-macromolecule complexes. A bottom-up strategy. J Chem Theory Comput. 2007; 3(6):1960–1986. [PubMed: 18978934]
13. Soderhjelm P, Ryde U. How Accurate Can a Force Field Become? A Polarizable Multipole Model Combined with Fragment-wise Quantum-Mechanical Calculations. J Phys Chem A. 2009; 113(3): 617–627. [PubMed: 19093829]
14. Soderhjelm P, Ohrn A, et al. Accuracy of typical approximations in classical models of intermolecular polarization. J Chem Phys. 2008; 128(1)
15. Holt A, Bostrom J, et al. A NEMO Potential that Includes the Dipole-Quadrupole and Quadrupole-Quadrupole Polarizability. J Comput Chem. 2010; 31(8):1583–1591. [PubMed: 20222056]
16. Stern HA, Rittner F, et al. Combined Fluctuating Charge and Polarizable Dipole Models: Application to a Five-Site Water Potential Function. J Chem Phys. 2001; 115(5):2237–2251.
17. Lamoureux G, MacKerell AD, et al. A Simple Polarizable Model of Water based on Classical Drude Oscillators. J Chem Phys. 2003; 119(10):5185–5197.
18. Ren PY, Ponder JW. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. J Phys Chem B. 2003; 107(24):5933–5947.
19. Kaminski GA, Stern HA, et al. Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from ab Initio Quantum Chemistry. J Phys Chem A. 2004; 108(4):621–627.
20. Anisimov VM, Vorobyov IV, et al. Polarizable Empirical Force Field for the Primary and Secondary Alcohol Series Based on the Classical Drude Model. J Chem Theory Comput. 2007; 3(6):1927–1946. [PubMed: 18802495]
21. Lopes PEM, Lamoureux G, et al. Polarizable Empirical Force Field for Aromatic Compounds based on the Classical Drude Oscillator. J Phys Chem B. 2007; 111(11):2873–2885. [PubMed: 17388420]
22. Harder E, Anisimov VM, et al. Understanding the Dielectric Properties of Liquid Amides from a Polarizable Force Field. J Phys Chem B. 2008; 112(11):3509–3521. [PubMed: 18302362]
23. Shi Y, Wu C, et al. Multipole Electrostatics in Hydration Free Energy Calculations. J Comput Chem. 2011; 32(5):967–977. [PubMed: 20925089]
24. Ren PY, Ponder JW. Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations. J Comput Chem. 2002; 23(16):1497–1506. [PubMed: 12395419]
25. Jiang JL, Wu YB, et al. Assessing the Performance of Popular Quantum Mechanics and Molecular Mechanics Methods and Revealing the Sequence-Dependent Energetic Features Using 100 Tetrapeptide Models. J Chem Theory Comput. 2010; 6(4):1199–1209.
26. Jiao D, Golubkov PA, et al. Calculation of Protein-Ligand Binding Free Energy by Using a Polarizable Potential. P Natl Acad Sci USA. 2008; 105(17):6290–6295.

27. Jiao D, Zhang JJ, et al. Trypsin-Ligand Binding Free Energies from Explicit and Implicit Solvent Simulations with Polarizable Potential. J Comput Chem. 2009; 30(11):1701–1711. [PubMed: 19399779]
28. Shi Y, Zhu CZ, et al. Probing the Effect of Conformational Constraint on Phosphorylated Ligand Binding to an SH2 Domain Using Polarizable Force Field Simulations. J Phys Chem B. 2012; 116(5):1716–1727. [PubMed: 22214214]
29. Zhang, J.; Shi, Y., et al. Protein-Ligand Interactions. Wiley-VCH Verlag GmbH & Co. KGaA; 2012. Polarizable Force Fields for Scoring Protein–Ligand Interactions; p. 99-120.
30. Harder E, MacKerell AD, et al. Many-Body Polarization Effects and the Membrane Dipole Potential. J Am Chem Soc. 2009; 131(8):2760–2761. [PubMed: 19199514]
31. Bauer BA, Lucas TR, et al. Water Permeation Through DMPC Lipid Bilayers using Polarizable Charge Equilibration Force Fields. Chem Phys Lett. 2011; 508(4-6):289–294. [PubMed: 21647243]
32. de Courcy B, Piquemal JP, et al. Polarizable Water Molecules in Ligand-Macromolecule Recognition. Impact on the Relative Affinities of Competing Pyrrolopyrimidine Inhibitors for FAK Kinase. J Am Chem Soc. 2010; 132(10):3312–3320. [PubMed: 20178314]
33. Gresh N, de Courcy B, et al. Polarizable Water Networks in Ligand-Metalloprotein Recognition. Impact on the Relative Complexation Energies of Zn-Dependent Phosphomannose Isomerase with D-Mannose 6-Phosphate Surrogates. J Phys Chem B. 2011; 115(25):8304–8316. [PubMed: 21650197]
34. Grossfield A, Ren PY, et al. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. J Am Chem Soc. 2003; 125(50):15671–15682. [PubMed: 14664617]
35. Jiao D, King C, et al. Simulation of Ca2+ and Mg2+ Solvation Using Polarizable Atomic Multipole Potential. J Phys Chem B. 2006; 110(37):18553–18559. [PubMed: 16970483]
36. Yu HB, Whitfield TW, et al. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. J Chem Theory Comput. 2010; 6(3):774–786. [PubMed: 20300554]
37. Wu JC, Piquemal JP, et al. Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field. J Chem Theory Comput. 2010; 6(7):2059–2070. [PubMed: 21116445]
38. Zhang J, Yang W, et al. Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential. J Chem Theory Comput. 2012; 8:1314–1324. [PubMed: 22754403]
39. Patel S, Davis JE, et al. Exploring Ion Permeation Energetics in Gramicidin A Using Polarizable Charge Equilibration Force Fields. J Am Chem Soc. 2009; 131(39):13890–1. [PubMed: 19788320]
40. Rappé AK, Goddard WA III. Charge Equilibration for Molecular Dynamics Simulations. J Phys Chem. 1991; 95:3358–3363.
41. Rick SW, Stuart SJ, et al. Dynamical Fluctuating Charge Force-Fields - Application to Liquid Water. J Chem Phys. 1994; 101(7):6141–6156.
42. Banks JL, Kaminski GA, et al. Parametrizing a Polarizable Force Field from ab Initio Data. I. The Fluctuating Point Charge Model. J Chem Phys. 1999; 110(2):741–754.
43. Ando K. A Stable Fluctuating-Charge Polarizable Model for Molecular Dynamics Simulations: Application to Aqueous Electron Transfers. J Chem Phys. 2001; 115(11):5228–5237.
44. Yoshii N, Miyauchi R, et al. A Molecular-Dynamics Study of the Equation of State of Water Using a Fluctuating-Charge Model. Chem Phys Lett. 2000; 317(3-5):414–420.
45. Patel S, Brooks CL. CHARMM Fluctuating Charge Force Field for Proteins: I Parameterization and Application to Bulk Organic Liquid Simulations. J Comput Chem. 2004; 25(1):1–15. [PubMed: 14634989]
46. Patel S, Brooks CL. Fluctuating Charge Force Fields: Recent Developments and Applications from Small Molecules to Macromolecular Biological Systems. Mol Simulat. 2006; 32(3-4):231–249.
47. Bauer BA, Patel S. Recent Applications and Developments of Charge Equilibration Force Fields for Modeling Dynamical Charges in Classical Molecular Dynamics Simulations. Theor Chem Acc. 2012; 131(3)
48. Stern HA, Kaminski GA, et al. Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from ab Initio Quantum Chemistry. J Phys Chem B. 1999; 103:4730–4737.

49. Sprik M, Klein ML. A Polarizable Model for Water Using Distributed Charge Sites. J Chem Phys. 1988; 89(12):7556–7560.
50. van Maaren PJ, van der Spoel D. Molecular Dynamics Simulations of Water with Novel Shell-Model Potentials. J Phys Chem B. 2001; 105(13):2618–2626.
51. Yu HB, Hansson T, et al. Development of a Simple, Self-Consistent Polarizable Model for Liquid Water. J Chem Phys. 2003; 118(1):221–234.
52. Anisimov VM, Lamoureux G, et al. Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. J Chem Theory Comput. 2005; 1(1):153–168.
53. Zhu X, Lopes PE, et al. Recent Developments and Applications of the CHARMM Force Fields. Wiley Interdiscip Rev Comput Mol Sci. 2012; 2(1):167–185. [PubMed: 23066428]
54. Caldwell JW, Kollman PA. Structure and Properties of Neat Liquids Using Nonadditive Molecular Dynamics: Water, Methanol, and N-Methylacetamide. J Phys Chem. 1995; 99:6208–6219.
55. Dang LX, Chang TM. Molecular Dynamics Study of Water Clusters, Liquid, and Liquid-Vapor Interface of Water with Many-Body Potentials. J Chem Phys. 1997; 106(19):8149–8159.
56. Brdarski S, Åstrand P-O, et al. The Inclusion of Electron Correlation in Intermolecular Potentials: Applications to the Formamide Dimer and Liquid Formamide. Theor Chem Acc. 2000; 105(1):7–14.
57. Kaminski GA, Stern HA, et al. Development of a Polarizable Force Field for Proteins via ab Initio Quantum Chemistry: First Generation Model and Gas Phase Tests. J Comput Chem. 2002; 23(16): 1515–1531. [PubMed: 12395421]
58. Applequist J, Carl JR, et al. An Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities. J Am Chem Soc. 1972; 94(9):2952–2960.
59. Thole BT. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. Chem Phys. 1981; 59:341–350.
60. Faerman CH, Price SL. A Transferable Distributed Multipole Model for the Electrostatic Interactions of Peptides and Amides. J Am Chem Soc. 1990; 112:4915–4926.
61. Price SL, Faerman CH, et al. Toward Accurate Transferable Electrostatic Models for Polypeptides: A Distributed Multipole Study of Blocked Amino Acid Residue Charge Distributions. J Comput Chem. 1991; 12(10):1187–1197.
62. Rogalewicz F, Ohanessian G, et al. Interaction of neutral and zwitterionic glycine with Zn2+ in gas phase: Ab initio and SIBFA molecular mechanics calculations. J Comput Chem. 2000; 21(11): 963–973.
63. Tiraboschi G, Fournie-Zaluski MC, et al. Intramolecular chelation of Zn(2+) by alpha- and beta-mercaptocarboxamides. A parallel ab initio and polarizable molecular mechanics investigation. assessment of the role of multipole transferability. J Comput Chem. 2001; 22(10):1038–1047.
64. Gresh N, Shi GB. Conformation-dependent intermolecular interaction energies of the triphosphate anion with divalent metal cations. Application to the ATP-binding site of a binuclear bacterial enzyme. A parallel quantum chemical and polarizable molecular mechanics investigation. J Comput Chem. 2004; 25(2):160–168. [PubMed: 14648615]
65. Williams DE. Representation of the Molecular Electrostatic Potential by Atomic Multipole and Bond Dipole Models. J Comput Chem. 1988; 9(7):745–763.
66. Vignemaeder F, Claverie P. The Exact Multicenter Multipolar Part of a Molecular Charge-Distribution and Its Simplified Representations. J Chem Phys. 1988; 88(8):4934–4948.
67. Day PN, Jensen JH, et al. An effective fragment method for modeling solvent effects in quantum mechanical calculations. J Chem Phys. 1996; 105(5):1968–1986.
68. Mooij WTM, Leusen FJJ. Multipoles versus Charges in the 1999 Crystal Structure Prediction Test. Phys Chem Chem Phys. 2001; 3(22):5063–5066.
69. Lommerse JPM, Motherwell WDS, et al. A Test of Crystal Structure Prediction of Small Organic Molecules. Acta Crystallogr B. 2000; 56:697–714. [PubMed: 10944263]
70. Mahoney MW, Jorgensen WL. A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. J Chem Phys. 2000; 112(20): 8910–8922.

71. Stone AJ. Distributed Multipole Analysis: Methods and Applications. Mol Phys. 1985; 56(5): 1047–1064.
72. Stone AJ. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. Chem Phys Lett. 1981; 83(2):233–239.
73. Ponder JW, Wu CJ, et al. Current Status of the AMOEBA Polarizable Force Field. J Phys Chem B. 2010; 114(8):2549–2564. [PubMed: 20136072]
74. MacKerell AD, Feig M, et al. Improved Treatment of the Protein Backbone in Empirical Force Fields. J Am Chem Soc. 2004; 126(3):698–699. [PubMed: 14733527]
75. Halgren TA. Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. J Am Chem Soc. 1992; 114:7827–7843.
76. Stone, AJ. The Theory of Intermolecular Forces. 2. Oxford University Press; Oxford: 2013.
77. Chaudret R, Gresh N, et al. Many-Body Exchange-Repulsion in Polarizable Molecular Mechanics. I. Orbital-Based Approximations and Applications to Hydrated Metal Cation Complexes. J Comput Chem. 2011; 32(14):2949–2957. [PubMed: 21793002]
78. Sagui C, Pedersen LG, et al. Towards an Accurate Representation of Electrostatics in Classical Force Fields: Efficient Implementation of Multipolar Interactions in Biomolecular Simulations. J Chem Phys. 2004; 120(1):73–87. [PubMed: 15267263]
79. Pearlman DA, Case DA, et al. Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. Comput Phys Commun. 1995; 91(1-3):1–41.
80. Friedrichs MS, Eastman P, et al. Accelerating Molecular Dynamic Simulation on Graphics Processing Units. J Comput Chem. 2009; 30(6):864–872. [PubMed: 19191337]
81. Wang W, Skeel RD. Fast evaluation of polarizable forces. J Chem Phys. 2005; 123(16)
82. Frisch, MJ.; T, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, GA.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, HP.; Izmaylov, AF.; Bloino, J.; Zheng, G.; Sonnenberg, JL.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, JA., Jr; Peralta, JE.; Ogliaro, F.; Bearpark, M.; Heyd, JJ.; Brothers, E.; Kudin, KN.; Staroverov, VN.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, JC.; Iyengar, SS.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, JM.; Klene, M.; Knox, JE.; Cross, JB.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Martin, RL.; Morokuma, K.; Zakrzewski, VG.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Dapprich, S.; Daniels, AD.; Farkas, Ö.; Foresman, JB.; Ortiz, JV.; Cioslowski, J.; Fox, DJ. Gaussian 09. Gaussian, Inc.; Wallingford CT: 2009.
83. Kong J, White CA, et al. Q-Chem 2.0: A High-Performance ab Initio Electronic Structure Program Package. J Comput Chem. 2000; 21(16):1532–1548.
84. Stone AJ. Distributed Multipole Analysis: Stability for Large Basis Sets. J Chem Theory Comput. 2005; 1(1128-1132)
85. Stone, AJ. GDMA. Cambridge University Technical Services; Cambridge, England: 1998.
86. DiStasio RA, Jung YS, et al. A Resolution-of-the-Identity Implementation of the Local Triatomics-in-Molecules Model for Second-Order Moller-Plesset Perturbation Theory with Application to Alanine Tetrapeptide Conformational Energies. J Chem Theory Comput. 2005; 1(5):862–876.
87. Essmann U, Perera L, et al. A Smooth Particle Mesh Ewald Method. J Chem Phys. 1995; 103(19): 8577–8593.
88. Darden T, York D, et al. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. J Chem Phys. 1993; 98(12):10089–10092.
89. Verlet L. Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. Phys Rev. 1967; 159(1):98–104.
90. Martyna GJ, Tuckerman ME, et al. Explicit Reversible Integrators for Extended Systems Dynamics. Mol Phys. 1996; 87:1117–1157.
91. Roux B. The Calculation of the Potential of Mean Force Using Computer-Simulations. Comput Phys Commun. 1995; 91(1-3):275–282.

92. Kumar S, Bouzida D, et al. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. J Comput Chem. 1992; 13(8):1011–1021.

93. Kumar S, Rosenberg JM, et al. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. J Comput Chem. 1995; 16(11):1339–1350.

94. Ting D, Wang GL, et al. Neighbor-Dependent Ramachandran Probability Distributions of Amino Acids Developed from a Hierarchical Dirichlet Process Model. Plos Comput Biol. 2010; 6(4):e100763.

95. Patriksson A, van der Spoel D. A Temperature Predictor for Parallel Tempering Simulations. Phys Chem Chem Phys. 2008; 10(15):2073–2077. [PubMed: 18688361]

96. Sugita Y, Okamoto Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. Chem Phys Lett. 1999; 314(1-2):141–151.

97. Best RB, Hummer G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. J Phys Chem B. 2009; 113(26):9004–9015. [PubMed: 19514729]

98. Graf J, Nguyen PH, et al. Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study. J Am Chem Soc. 2007; 129(5):1179–1189. [PubMed: 17263399]

99. Jelsch C, Teeter MM, et al. Accurate Protein Crystallography at Ultra-high Resolution: Valence Electron Distribution in Crambin. P Natl Acad Sci USA. 2000; 97(7):3171–3176.

100. Neidigh JW, Fesinmeyer RM, et al. Designing a 20-Residue Protein. Natural Structural Biology. 2002; 9(6):425–430.

101. McKnight CJ, Matsudaira PT, et al. NMR Structure of the 35-Residue Villin Headpiece Subdomain. Natural Structural Biology. 1997; 4(3):180–184.

102. Vijaykumar S, Bugg CE, et al. Structure of Ubiquitin Refined at 1.8 a Resolution. J Mol Biol. 1987; 194(3):531–544. [PubMed: 3041007]

103. Ulmer TS, Ramirez BE, et al. Evaluation of Backbone Proton Positions and Dynamics in a Small Protein by Liquid Crystal NMR Spectroscopy. J Am Chem Soc. 2003; 125(30):9179–9191. [PubMed: 15369375]

104. Ko TP, Robinson H, et al. The Refined Crystal Structure of an Eel Pout Type III Antifreeze Protein RD1 at 0.62-Ang Resolution Reveals Structural Microheterogeneity of Protein and Solvation. Biophys J. 2003; 84(2):1228–1237. [PubMed: 12547803]

105. Parkin S, Rupp B, et al. Structure of Bovine Pancreatic Trypsin Inhibitor at 125 K: Definition of Carboxyl-Terminal Residues Gly57 and Ala58. Acta Crystallogr D. 1996; 52:18–29. [PubMed: 15299722]

106. Szep S, Park S, et al. Structural Coupling between FKBP12 and Buried Water. Proteins. 2009; 74(3):603–611. [PubMed: 18704951]

107. Young ACM, Dewan JC, et al. Comparison of Radiation-Induced Decay and Structure Refinement from X-Ray Data Collected from Lysozyme Crystals at Low and Ambient-Temperatures. J Appl Crystallogr. 1993; 26:309–319.

108. Ren P, Wu C, et al. Polarizable Atomic Multipole-based Molecular Mechanics for Organic Molecules. J Chem Theory Comput. 2011; 7(10):3143–3161. [PubMed: 22022236]

109. Schnieders MJ, Ponder JW. Polarizable atomic multipole solutes in a generalized Kirkwood continuum. J Chem Theory Comput. 2007; 3(6):2082–2097.

110. Beachy MD, Chasman D, et al. Accurate ab Initio Quantum Chemical Determination of the Relative Energetics of Peptide Conformations and Assessment of Empirical Force Fields. J Am Chem Soc. 1997; 119(25):5908–5920.

111. Hornak V, Abel R, et al. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins. 2006; 65(3):712–725. [PubMed: 16981200]

112. Hegefeld WA, Chen SE, et al. Helix Formation in a Pentapeptide Experiment and Force-field Dependent Dynamics. J Phys Chem A. 2010; 114(47):12391–12402. [PubMed: 21058639]

113. Best RB, Mittal J, et al. Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of alpha-Helix and beta-Hairpin Formation. Biophys J. 2012; 103(5):1045–1051. [PubMed: 23009854]

114. Best RB, Zhu X, et al. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. J Chem Theory Comput. 2012; 8(9):3257–3273. [PubMed: 23341755]
115. Best RB, Mittal J. Balance between Alpha and Beta Structures in ab Initio Protein Folding. J Phys Chem B. 2010; 114(26):8790–8798. [PubMed: 20536262]
116. Best RB, Buchete NV, et al. Are Current Molecular Dynamics Force Fields Too Helical? Biophys J. 2008; 95(1):L7–L9.
117. Lindorff-Larsen K, Maragakis P, et al. Systematic Validation of Protein Force Fields against Experimental Data. Plos One. 2012; 7(2):e32131. [PubMed: 22384157]
118. Karplus M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. J Chem Phys. 1959; 30(1):11–15.
119. Aliev AE, Courtier-Murias D. Experimental Verification of Force Fields for Molecular Dynamics Simulations Using Gly-Pro-Gly-Gly. J Phys Chem B. 2010; 114(38):12358–12375. [PubMed: 20825228]
120. Shi ZS, Olson CA, et al. Polyproline II Structure in a Sequence of Seven Alanine Residues. P Natl Acad Sci USA. 2002; 99(14):9190–9195.
121. Chen K, Liu Z, et al. Spin Relaxation Enhancement Confirms Dominance of Extended Conformations in Short Alanine Peptides. Angew Chem Int Edit. 2007; 46(47):9036–9039.
122. Schweitzer-Stenner R, Measey TJ. The Alanine-rich XAO Peptide Adopts a Heterogeneous Population, Including Turn-like and Polyproline II Conformations. P Natl Acad Sci USA. 2007; 104(16):6649–6654.
123. Schweitzer-Stenner R, Gonzales W, et al. Conformational Manifold of alpha-Aminoisobutyric Acid (Aib) Containing Alanine-based Tripeptides in Aqueous Solution Explored by Vibrational Spectroscopy, Electronic Circular Dichroism Spectroscopy, and Molecular Dynamics Simulations. J Am Chem Soc. 2007; 129(43):13095–13109. [PubMed: 17918837]
124. Mukhopadhyay P, Zuber G, et al. Characterizing Aqueous Solution Conformations of a Peptide Backbone Using Raman Optical Activity Computations. Biophys J. 2008; 95(12):5574–5586. [PubMed: 18805935]
125. Marqusee S, Robbins VH, et al. Unusually Stable Helix Formation in Short Alanine-Based Peptides. P Natl Acad Sci USA. 1989; 86(14):5286–5290.
126. Spek EJ, Olson CA, et al. Alanine is an Intrinsic Alpha-Helix Stabilizing Amino Acid. J Am Chem Soc. 1999; 121(23):5571–5572.
127. Shalongo W, Dugad L, et al. Distribution of Helicity within the Model Peptide Acetyl(AAQAA) (3)Amide. J Am Chem Soc. 1994; 116(18):8288–8293.
128. Buck M, Boyd J, et al. Structural Determinants of Protein Dynamics: Analysis of 15N NMR Relaxation Measurements for Main-Chain and Side-Chain Nuclei of Hen Egg White Lysozyme. Biochemistry. 1995; 34(12):4041–4055. [PubMed: 7696270]
129. Tjandra N, Feller SE, et al. Rotational Diffusion Anisotropy of Human Ubiquitin from N-15 NMR Relaxation. J Am Chem Soc. 1995; 117(50):12562–12566.
130. Lipari G, Szabo A. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. J Am Chem Soc. 1982; 104(17):4559–4570.
131. Lipari G, Szabo A. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. J Am Chem Soc. 1982; 104(17):4546–4559.
132. Fischer MW, Zeng L, et al. Characterizing semilocal motions in proteins by NMR relaxation studies. P Natl Acad Sci USA. 1998; 95(14):8016–9.
133. Prompers JJ, Bruschweiler R. General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation. J Am Chem Soc. 2002; 124(16):4522–4534. [PubMed: 11960483]
134. Smith LJ, Sutcliffe MJ, et al. Analysis of Phi and Chi 1 Torsion Angles for Hen Lysozyme in Solution from 1H NMR Spin-Spin Coupling Constants. Biochemistry. 1991; 30(4):986–996. [PubMed: 1989688]

135. Grimshaw, SB. Doctoral Thesis, University of Oxford. 1999. Novel Approaches to Characterizing Native and Denatured Proteins by NMR.

136. Miclet E, Boisbouvier J, et al. Measurement of Eight Scalar and Dipolar Couplings for Methine-Methylene Pairs in Proteins and Nucleic Acids. J Biomol NMR. 2005; 31(3):201–216. [PubMed: 15803394]

137. Berndt KD, Guntert P, et al. Determination of a High-Quality Nuclear Magnetic Resonance Solution Structure of the Bovine Pancreatic Trypsin Inhibitor and Comparison with Three Crystal Structures. J Mol Biol. 1992; 227(3):757–775. [PubMed: 1383552]

138. Lindorff-Larsen K, Piana S, et al. Improved Side-chain Torsion Potentials for the Amber ff99SB Protein Force Field. Proteins. 2010; 78(8):1950–1958. [PubMed: 20408171]

139. Mackerell AD, Feig M, et al. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. J Comput Chem. 2004; 25(11):1400–1415. [PubMed: 15185334]

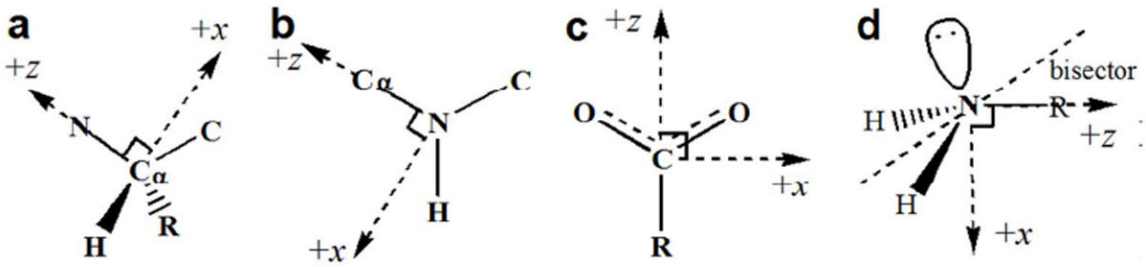

Figure 1.

Local frame definitions for (a) a protein backbone Cα, (b) backbone amide N, (c) carboxylate carbon, and (d) amine nitrogen. The Cα and amide N use the "Z-then-X" convention, where a first neighboring atom is selected to define the Z-axis, and a second neighbor defines the ZX-plane and the positive X-direction. The carboxylate example uses the "Bisector" convention, where the bisector of two neighboring atoms defines Z-axis. This is typically used in structures with local 2-fold symmetry. The amine N is represented by the "Z-Bisector" convention, where the N-R bond defines the Z-axis, and the bisector between the two N-H bonds defines the ZX-plane. In all cases, the Y-axis is then defined according to the right hand rule.

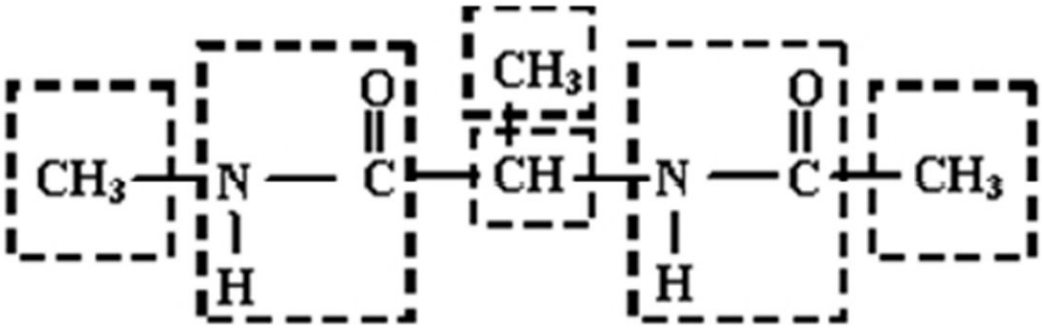

Figure 2.
Illustration of the intramolecular polarization group definition. Each group consists of a functional group or cluster of atoms with limited conformational flexibility. The permanent multipole moments on each atom only polarize atoms of other groups, while induced dipoles on all atoms polarize all other atoms regardless of group membership.

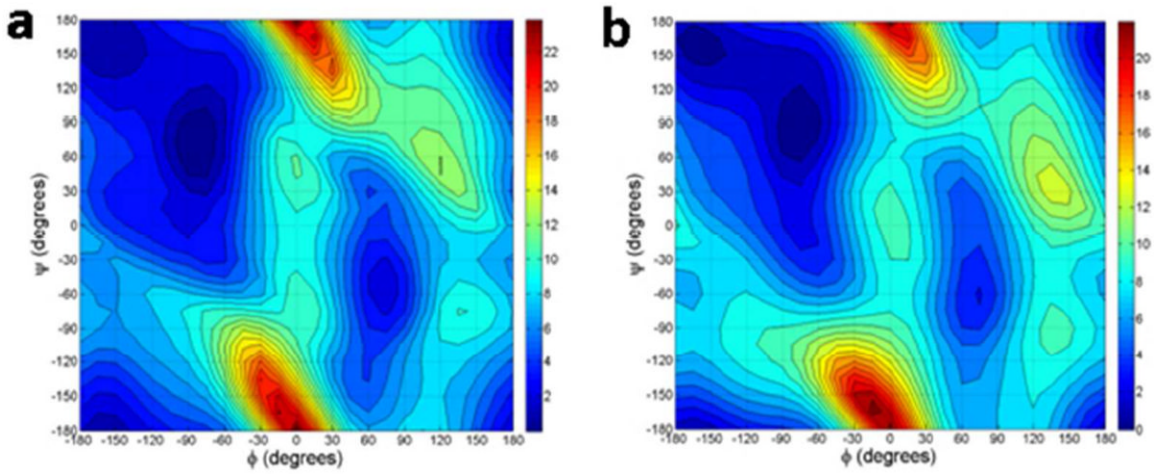

Figure 3.
Gas-phase energy contours for alanine dipeptide from RI-TRIM MP2/CBS calculations (a) and the AMOEBA protein force field (b). The energies were computed on a 24 × 24 grid and then contoured.

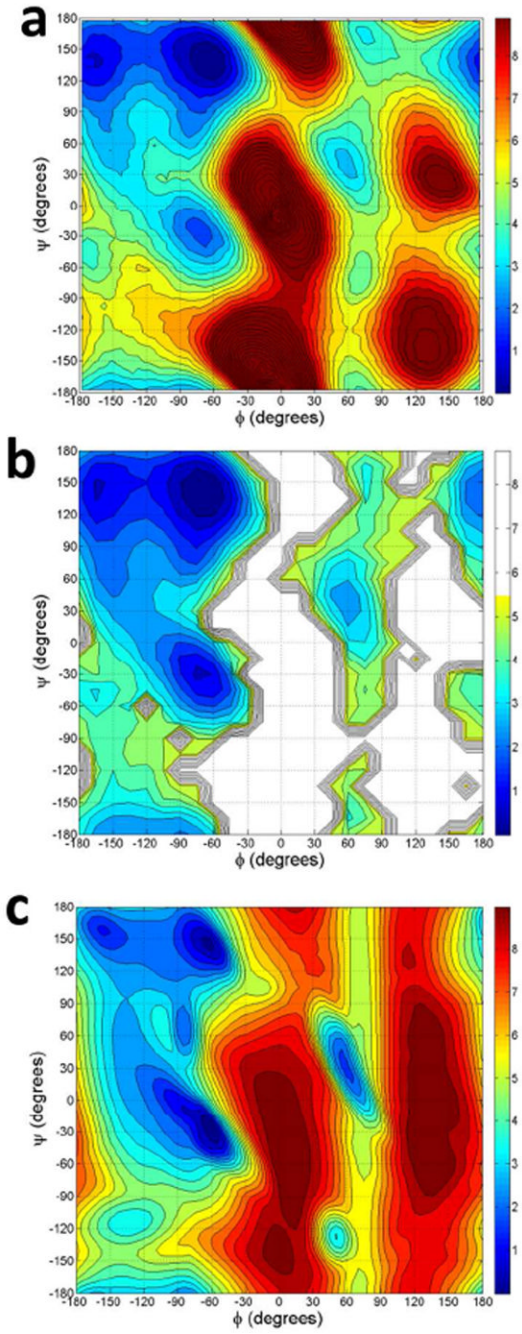

Figure 4.

Comparison of Ramachandran map potentials of mean force for alanine. (a) Ala-2 residue of (Ala)3 as predicted by 2D-WHAM simulations. (b) Average of ala-2, ala-3, and ala-4 residues in replica exchange molecular dynamics simulation of the (Ala)5 peptide. The trajectory at 298 K was used. (c) The PDB data are ref 83.

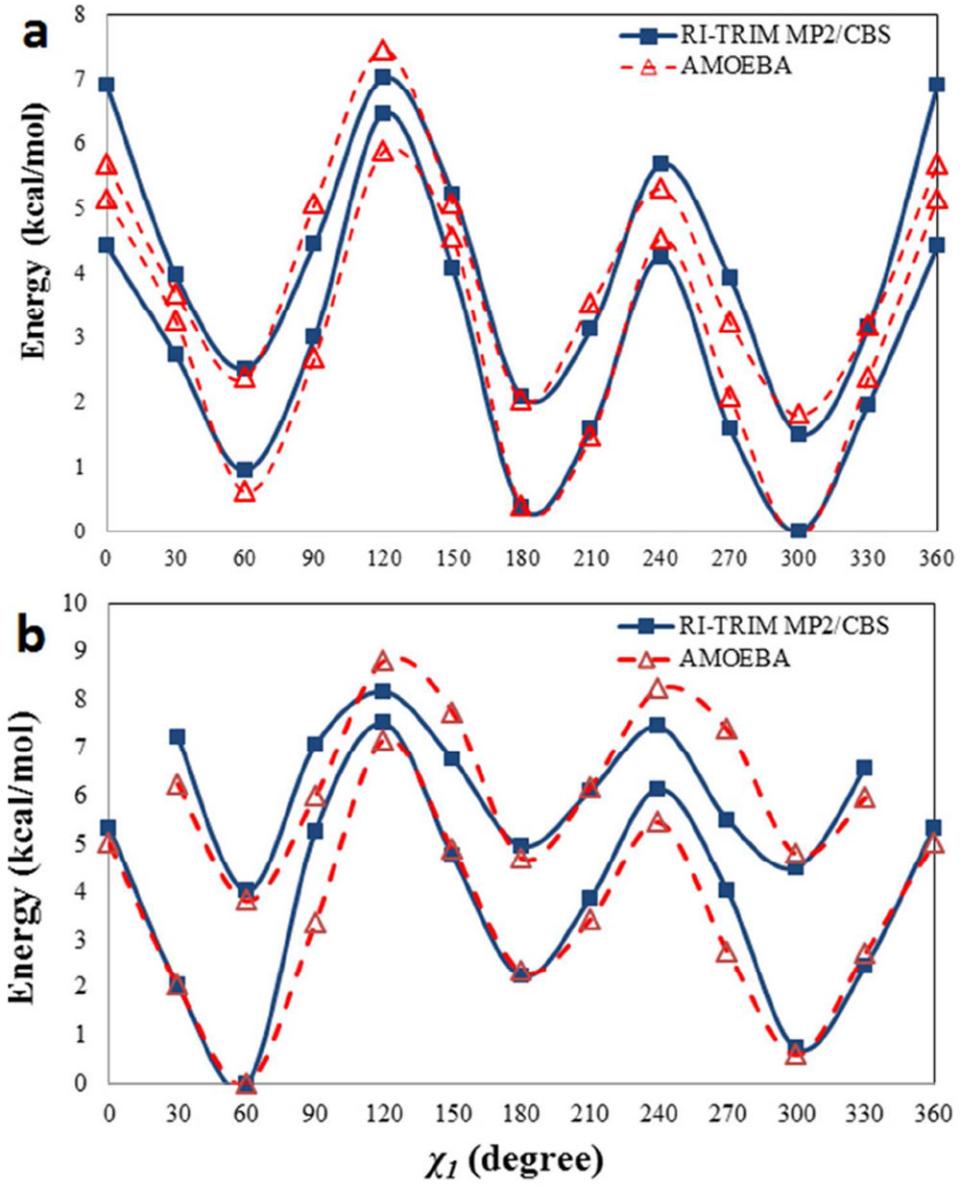

Figure 5.

Comparison of (a) isoleucine and (b) serine conformational energy about χ1 angle. The solid lines are RI-TRIM MP2/CBS energy while the dashed lines are AMOEBA values. The AMOEBA curve is shifted to minimize the overall RMS difference between AMOEBA and QM. The top set of curves (with higher energy at 0 degree) corresponds to a backbone conformation of (-60.0, -45.0), and the other corresponds to (-140.0, 135.0).

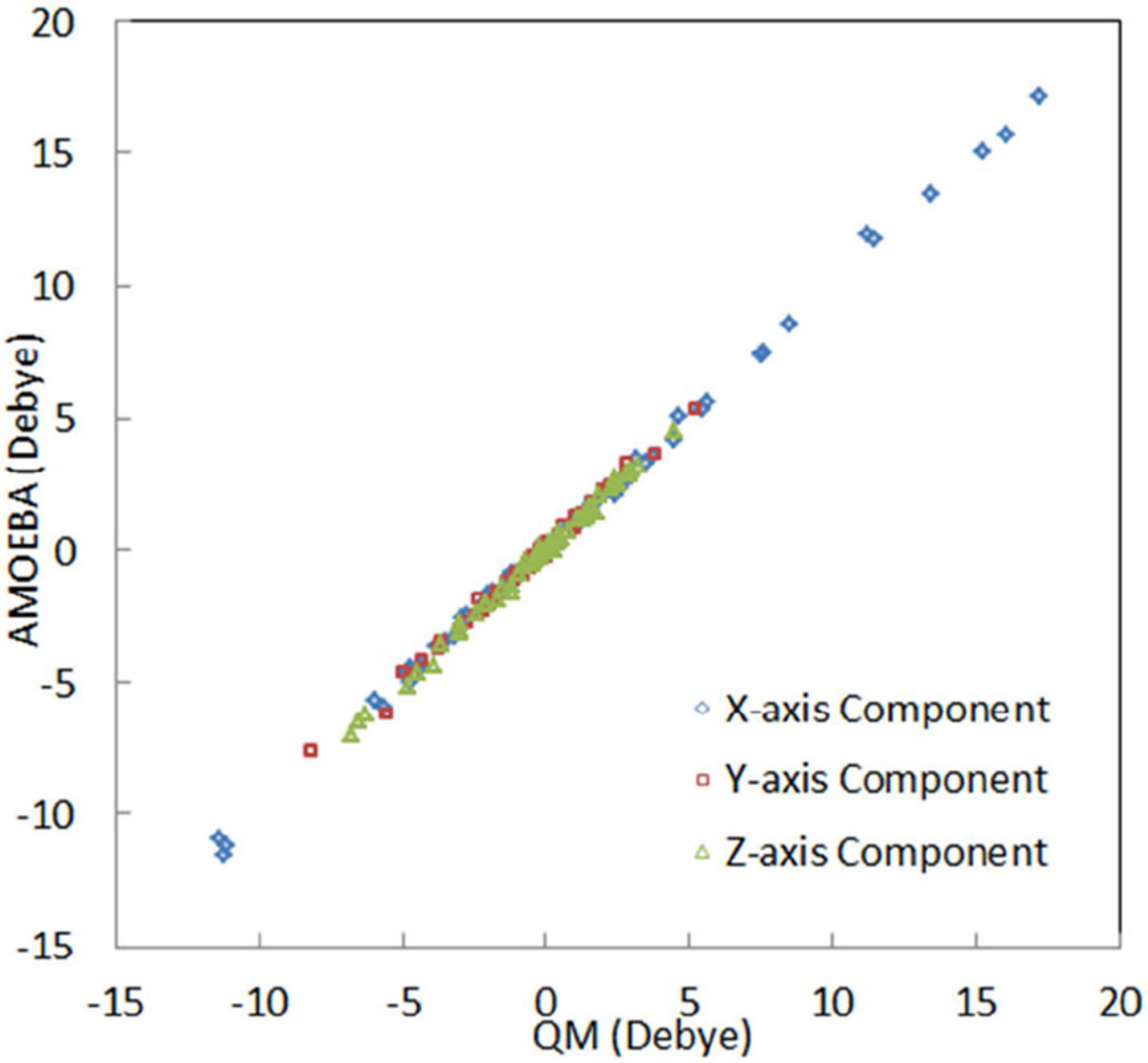

Figure 6.
Comparison of amino acid molecular dipole moments predicted by AMOEBA and QM (MP2/aug-cc-pVTZ). The AMOEBA permanent atomic multipoles were derived from a set of dipeptides and validated on additional conformations (3 for each amino acid). Only the results for the validation sets are shown. The actual data can be found in the supporting information.

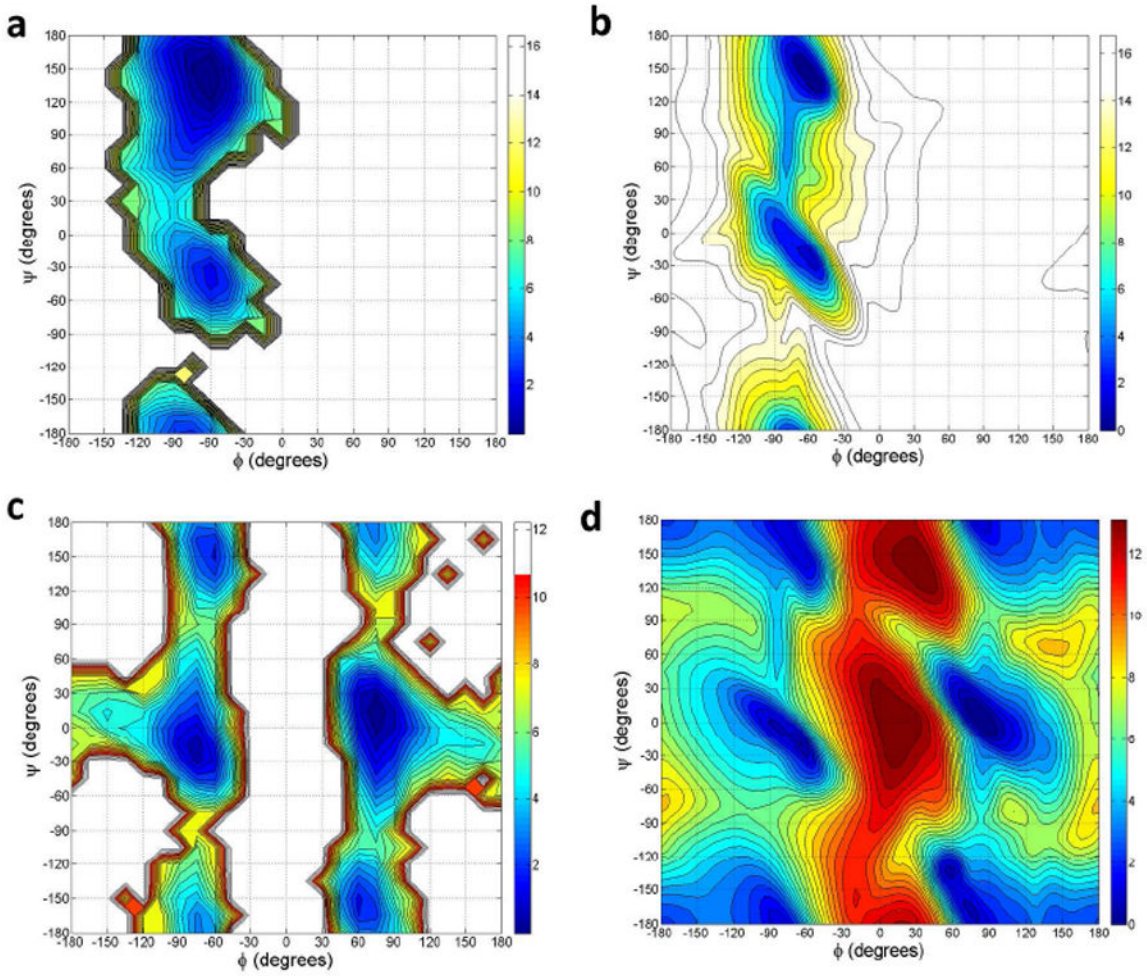

Figure 7.

Comparison of Ramachandran potential of mean force maps for proline and glycine. (a) Pro-2 residue of GPGG from AMOEBA simulations. (b) The PDB data for proline. (c) Gly-3 residue of GPGG from AMOEBA simulations. (d) PDB data for glycine. All PDB PMFs were computed using data from Dunbrack *et al.*83

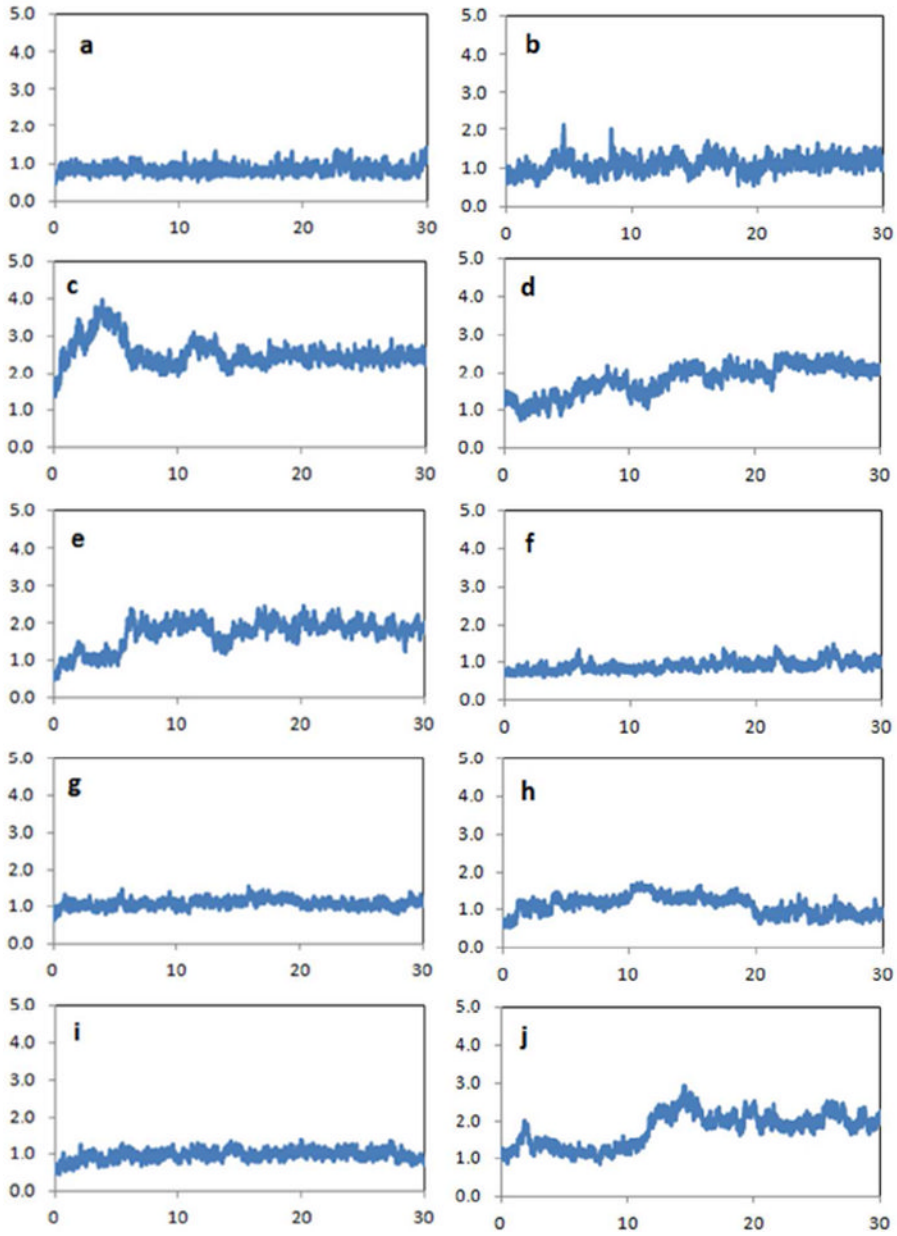

Figure 8.

The time evolution of backbone RMSDs from the X-ray structures for ten simulated proteins. For each protein, 30 ns MD simulations were performed with the AMOEBA force field in explicit water. The X-axis represents time (ns) and the Y-axis is the RMSD values in Å. (a) crambin (PDB:1EJG), (b) Trp cage (PDB:1L2Y), (c) villin headpiece (PDB:1VII), (d) ubiquitin (PDB:1UBQ), (e) GB3 domain (PDB:2OED), (f) RD1 antifreeze protein (PDB: 1UCS), (g) SUMO-2 domain (PDB:1WM3), (h) BPTI (PDB:1BPI), (i) FK binding protein (PDB:2PPN), (j) and lysozyme (PDB:6LYT).

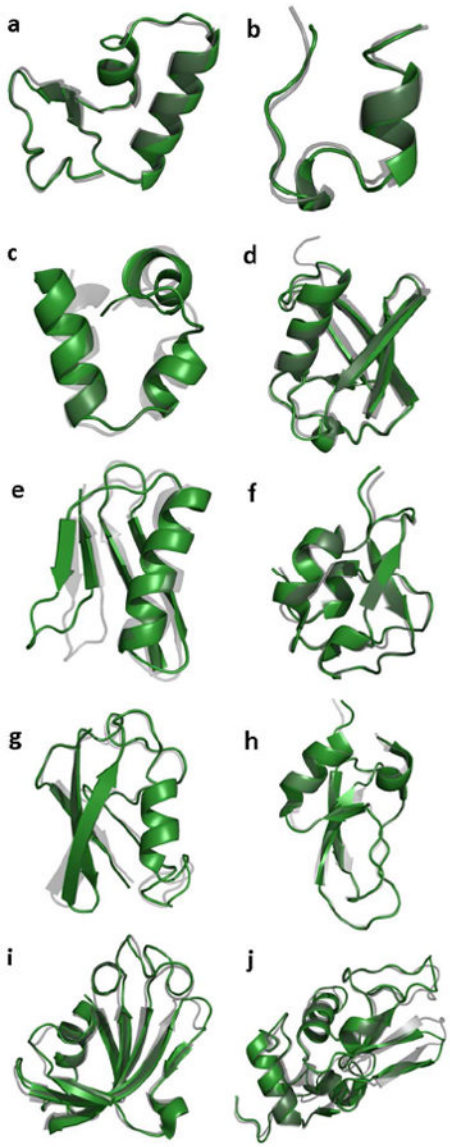

Figure 9.

Superimposition of the final structures from AMOEBA simulations (green) and the experimental X-ray crystal structures (grey). (a) crambin (PDB:1EJG), (b) Trp cage (PDB: 1L2Y), (c) villin headpiece (PDB:1VII), (d) ubiquitin (PDB:1UBQ), (e) GB3 domain (PDB: 2OED), (f) RD1 antifreeze protein (PDB:1UCS), (g) SUMO-2 domain (PDB:1WM3), (h) BPTI (PDB:1BPI), (i) FK binding protein (PDB:2PPN), (j) and lysozyme(PDB:6LYT).

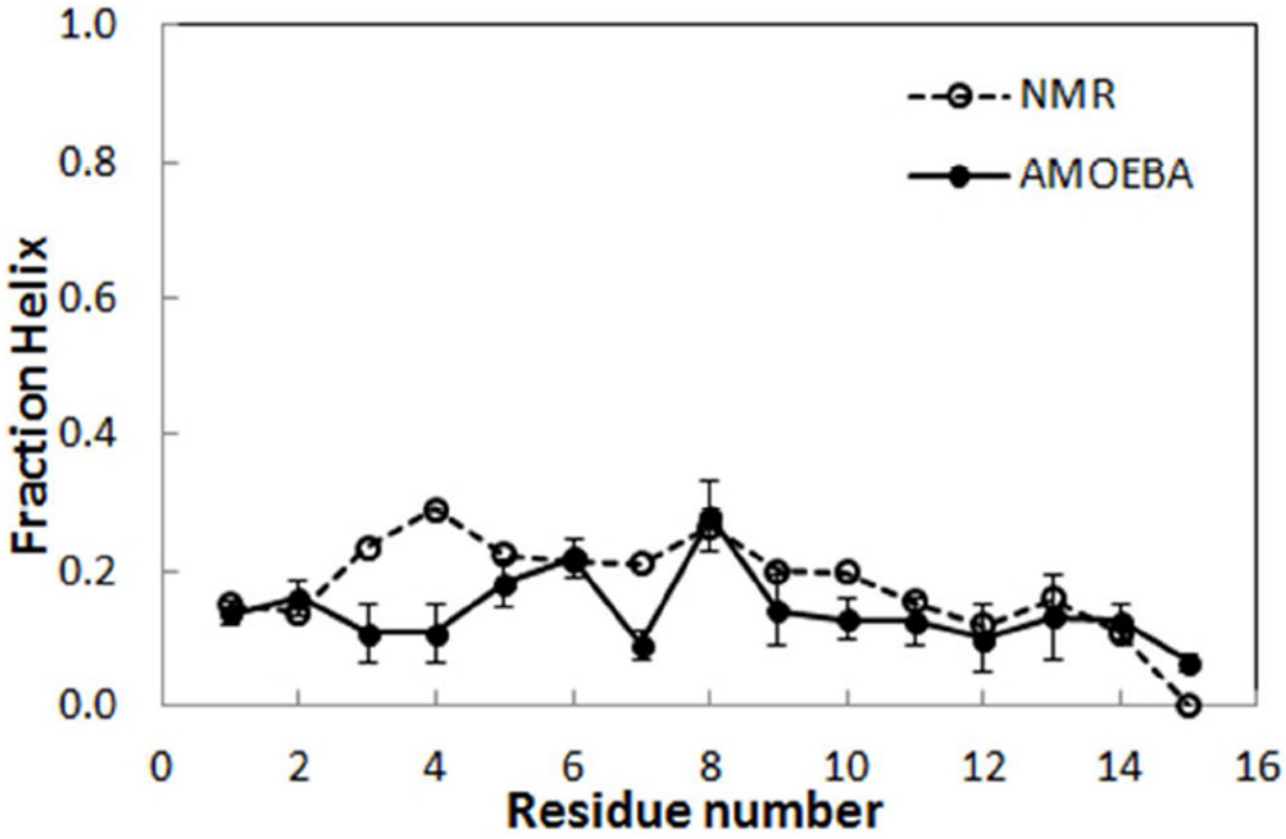

Figure 10.
Fraction of helix ⟨hi⟩ for each residue in Ac-(AAQAA)3-NH2 from replica exchange MD simulations and NMR chemical shifts at 303 K.

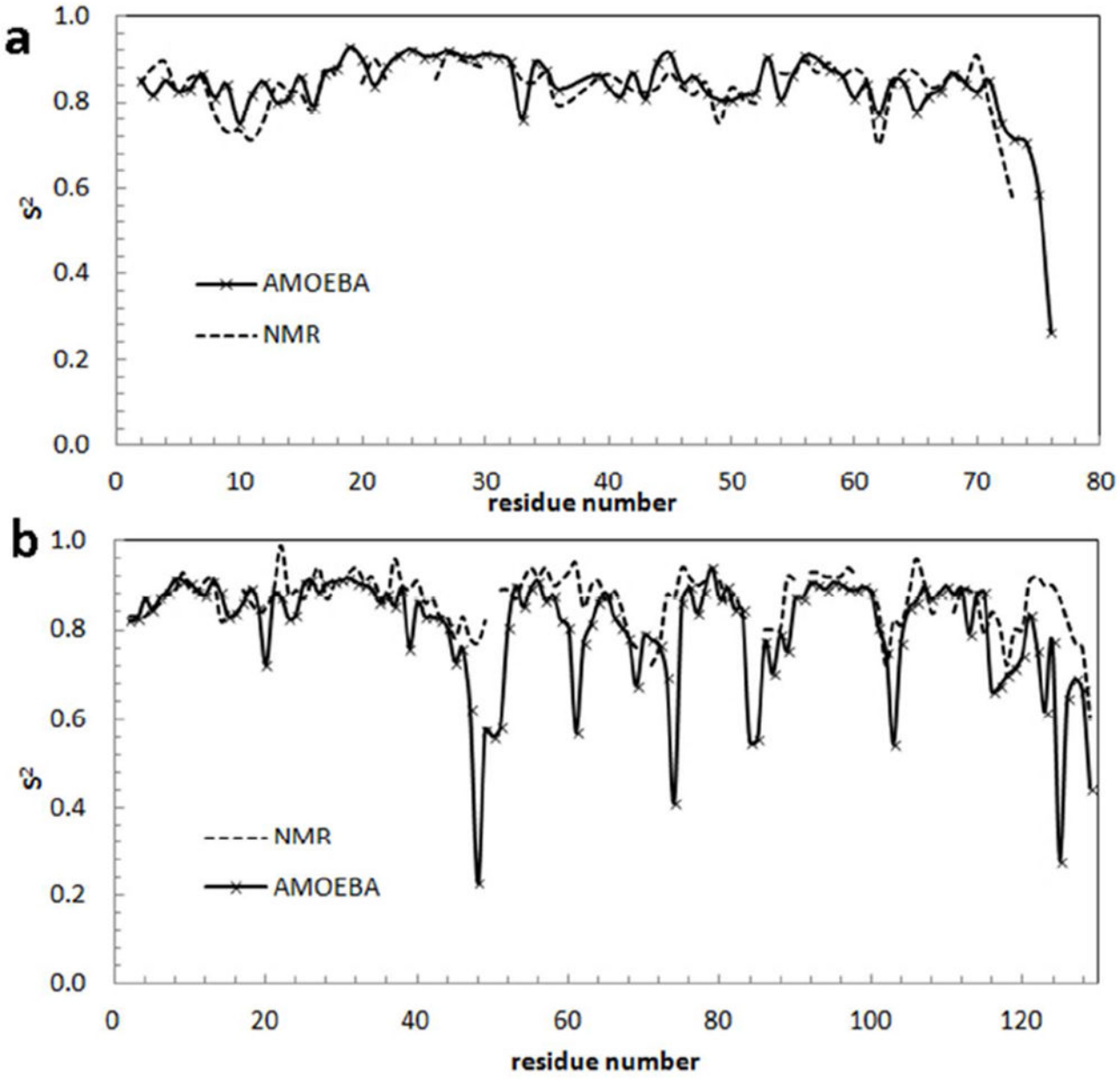

Figure 11.

Order parameters (S2) derived from experimental NMR115, 116 (dash lines) and calculated from MD simulations in explicit water using AMOEBA. (a) Ubiquitin, (b) Lysozyme.

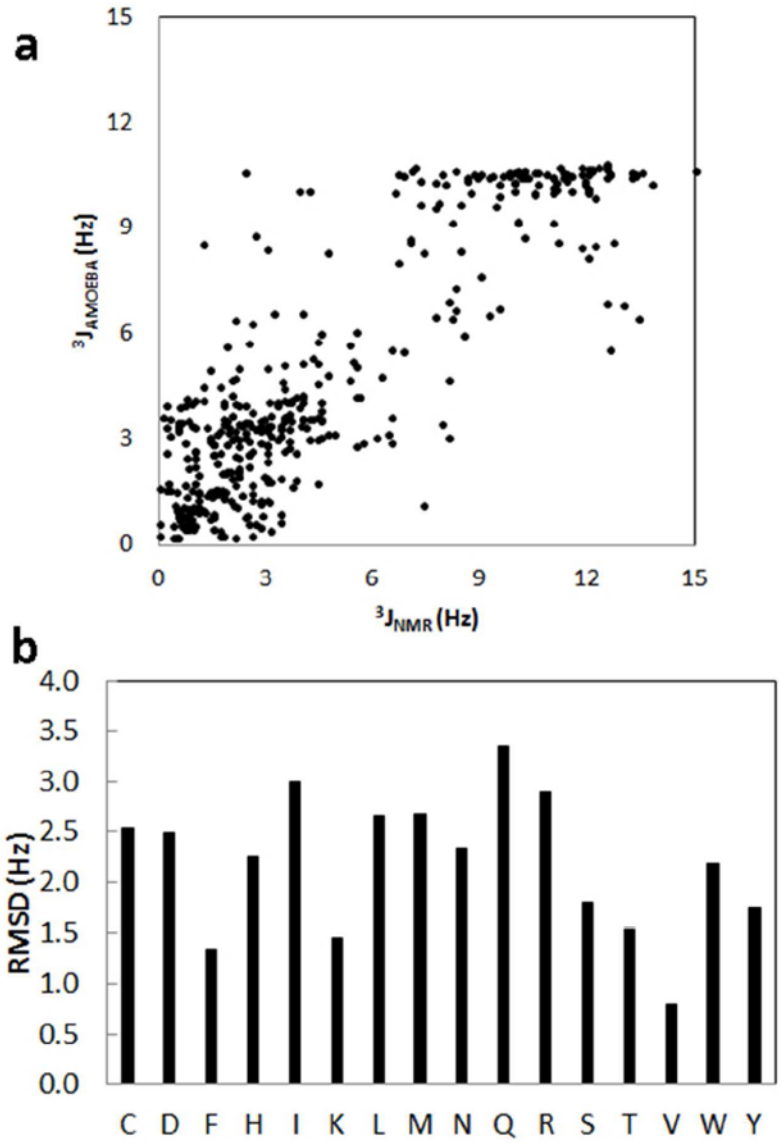

Figure 12.

(a) Correlation of the experimental NMR J-couplings and the calculated J-coupling values from the MD simulations of BPTI, GB3 domain, ubiquitin and lysozyme. (b) The RMSDs between the experimental and AMOEBA calculated J-coupling constants for each residue.

Table 1

The vdW parameters for protein backbone atoms.

Atom class	$2r$ (Å)	ε (kcal/mol)	H-reduction factor
C_α	3.650	0.101	
H_α	2.940	0.026	0.91
C_β	3.820	0.101	
H_β	2.980	0.024	0.92
N (amide)	3.710	0.110	
H (amide)	2.590	0.022	0.90
C (carbonyl)	3.820	0.106	
O (carbonyl)	3.300	0.112	
O^- (in $-COO^-$)	3.700	0.129	
S	4.005	0.355	
S^-	4.200	0.355	

AMOEBA uses “atom classes” to define vdW and valence parameters while more finely-grained “atom types” are used for electrostatic parameters. Several different atom types may belong to the same atom class. r is the atomic radius; ε is the potential well depth; H-reduction factor accounts for the distance reduction that moves H vdW sphere off its nuclear center toward the heavy atom (X) it attaches to. The number represents the percentage of H-X bond length where the H atom vdW sphere is located at. Note that $2r$ (diameter) is used here to be consistent with TINKER parameter format.

Table 2

Comparison of side chain conformational energy (kcal/mol) as calculated by AMOEBA and QM (RI-TRIM MP2/CBS).

	Side chain torsion	RMSD		Side chain torsion	RMSD
Cys	χ_1	0.67	Asn	χ_1	1.35
	χ_2	0.10		χ_2	1.06
Met	χ_1 same as Glu and Gln	0.40	Glu	χ_1 from Met	0.79
	χ_2	0.52		χ_2 from Gln	1.86
	χ_3	0.10		χ_3 from Asp χ_2	0.42
Ser	χ_1	0.84	Leu	χ_1	0.32
	χ_2	0.22		χ_2 averaged over Leu/ Ile	0.31
Thr	χ_1 same as Ser and Val	0.96	Val	χ_1	0.58
His	χ_1	1.24	Phe	χ_1	0.65
	χ_2	1.31		χ_2	0.54
Hid	χ_1	0.83	Tyr	χ_1 from Phe	0.55
	χ_2	0.67		χ_2	0.44
Hie	χ_1	1.10	Trp	χ_1	0.54
	χ_2	1.00		χ_2	0.55
Ile	χ_1 from Val	0.53	Asph	χ_1	1.39
	χ_2 averaged over Leu/Ile	0.34		χ_2	0.97
Arg	χ_1 from Met	1.63	Lys	χ_1 from Asn	0.74
	χ_2 same as Lys	1.08		χ_2	0.95
	χ_3	1.76		χ_3	0.96
	χ_4	0.80		χ_4	0.73
Asp	χ_1 from Asph	1.22	Gln	χ_1 from Met	0.79
	χ_2	1.01		χ_2	1.86

J Chem Theory Comput. Author manuscript; available in PMC 2014 July 25.

Side chain torsion	RMSD	Side chain torsion	RMSD
		χ_3 from Asn χ_2	0.42

The conformational energies were obtained by rotating each listed side chain torsion from 0 to 360° at 30° intervals. For all RMSD calculations, data from two backbone conformations, α-helical and β-sheet, were combined.

Table 3

Comparison of alanine tetrapeptide conformational energy (kcal/mol).

RI MP2 CBS	LMP2/cc-pVTZ(-f)	MP2/6-311+G2d2p	DFT B3LYP/6-31G*	AMOEBA	Struct. RMS (Å)	AMOEBA (φ/ψ restrained)
4.13	2.50	4.61	1.62	3.07	0.30	2.54
4.19	2.60	4.21	1.71	3.62	0.42	0.74
0.57	0.00	-0.70	-1.00	0.00	0.21	0.33
5.73	3.87	5.50	3.61	4.07	0.37	3.82
5.26	3.88	5.14	4.25	3.96	0.30	2.27
2.90	2.19	2.10	2.10	2.45	0.53	0.14
6.67	5.73	5.61	6.56	7.64	0.45	0.65
4.64	4.17	3.32	4.99	5.45	0.44	1.06
7.92	6.93	6.98	5.20	10.01	0.25	3.14
7.79	6.99	6.57	7.24	6.34	0.34	0.62
0.00	-0.19	-1.41	0.14	0.75	0.68	0.58
0.29	0.50	-1.07	1.73	0.75	0.91	0.22
3.66	1.77	3.20	1.14	3.56	0.62	0.02
4.68	3.68	4.14	3.89	4.66	0.71	0.00
2.19	2.07	0.65	3.47	2.28	0.59	0.08
3.55	2.83	2.33	3.31	2.93	0.48	0.24
3.42	2.78	2.02	2.00	2.32	0.28	1.09
1.91	0.52	1.15	-0.87	2.19	0.56	0.20
3.82	2.83	2.90	1.13	4.25	0.56	0.19
1.76	0.87	0.88	0.80	3.18	0.47	2.91
2.92	2.11	1.59	1.78	0.00	0.92	8.51
5.82	4.82	4.59	4.84	6.87	0.59	1.60
5.82	4.82	4.57	4.84	6.84	0.33	1.46
3.98	2.98	2.89	3.59	4.11	0.30	0.19
2.50	1.59	1.54	1.92	2.87	0.50	0.35
0.67	0.18	-0.41	1.40	1.60	0.37	1.51
4.02	3.18	3.04	3.53	6.26	0.38	5.57
RMS deviation	1.05	1.06	1.54	*1.15*	*0.47*	*1.22*

J Chem Theory Comput. Author manuscript; available in PMC 2014 July 25.

This set conformers were used in previous studies.[110,86] The RI MP2/CBS and other QM results are taken from taken from ref [86]. Two sets of AMOEBA results are listed. Both are from energy minimization of QM structures, one with backbone ϕ/ψ torsional angles restrained at QM optimal values and the other fully relaxed in the AMOEBA model. The RMSD was computed using RI MP2/CBS energies as the reference.

Table 4

Comparison of *J*-coupling values (Hz) from AMOEBA simulations with experimental data for the $(Ala)_5$ peptide.

Rsidue index	*J*-coupling type	*J*-simulation	*J*-expt.[98]
Ala-2	$^1J(N,C_\alpha)$	11.066	11.36
Ala-3	$^1J(N,C_\alpha)$	10.923	11.26
Ala-4	$^1J(N,C_\alpha)$	10.922	11.25
Ala-2	$^2J(N,C_\alpha)$	8.448	9.20
Ala-3	$^2J(N,C_\alpha)$	8.170	8.55
Ala-4	$^2J(N,C_\alpha)$	8.232	8.40
Ala-5	$^2J(N,C_\alpha)$	8.250	8.27
Ala-2	$^3J(C',C')$	0.866	0.19
Ala-2	$^3J(H_\alpha,C')$	1.729	1.85
Ala-3	$^3J(H_\alpha,C')$	1.705	1.86
Ala-4	$^3J(H_\alpha,C')$	1.713	1.89
Ala-5	$^3J(H_\alpha,C')$	1.929	2.19
Ala-2	$^3J(H_N,C')$	1.087	1.13
Ala-4	$^3J(H_N,C')$	1.315	1.15
Ala-5	$^3J(H_N,C')$	1.216	1.16
Ala-2	$^3J(H_N,C_\beta)$	1.819	2.30
Ala-3	$^3J(H_N,C_\beta)$	1.833	2.24
Ala-4	$^3J(H_N,C_\beta)$	1.743	2.14
Ala-5	$^3J(H_N,C_\beta)$	1.584	1.96
Ala-2	$^3J(H_N,H_\alpha)$	6.269	5.59
Ala-3	$^3J(H_N,H_\alpha)$	5.988	5.74
Ala-4	$^3J(H_N,H_\alpha)$	6.079	5.98
Ala-5	$^3J(H_N,H_\alpha)$	6.607	6.54
Ala-2	$^3J(H_N,C_\alpha)$	0.421	0.67
Ala-3	$^3J(H_N,C_\alpha)$	0.614	0.68
Ala-4	$^3J(H_N,C_\alpha)$	0.648	0.69
Ala-5	$^3J(H_N,C_\alpha)$	0.663	0.73
χ^2 = 0.994	**RMS=0.33**		

The peptide was unblocked and protonated at both the N- and C-termini, corresponding to experimental conditions of pH=2. Replica exchange molecular dynamics (REMD) simulations were performed using 32 replicas at temperatures between 278 K and 620 K (30-ns of MD for each replica). The trajectory at 298 K was extracted and used for the *J*-coupling calculation.

Table 5

Comparison of *J*-coupling values (Hz) from AMOEBA simulations and NMR experimental data for the GPGG tetrapeptide.

Residue index	J-coupling type	*J*-simulation (B972 EPR-III)	*J*-simulation (B3LYP EPR-III)	*J*-expt.[119]
Pro-2	$J(H_\alpha,C')$	1.75	1.88	1.30
Gly-3	$J(H_\alpha,H_N)$	4.94	3.67	4.10
Gly-3	$J(H_\alpha,C')$	6.07	6.76	6.30
		RMS=0.44	RMS=0.39	

Two sets of simulated *J*-coupling using different Karplus coefficients are shown. Replica exchange molecular dynamics (REMD) simulations were performed using 32 replicas at temperatures between 278 K and 620 K (30-ns of MD for each replica). The trajectory at 298 K was extracted and used for the *J*-coupling calculation.

Table 6

Comparison (RMSD) of *J*-coupling values (Hz) from AMOEBA simulations and experiments for BPTI, GB3, ubiquitin and lysozyme.

	BPTI	GB3	Ubiquitin	Lysozyme
AMOEBA	1.741	1.44	1.41	2.23
AMBER FF99SB[138]	1.779	1.48	1.89	2.60
AMBER FF99SB-ILDN[138]	1.448	0.89	1.43	2.12

Results from the Amber ff99SB force field, and the refined Amber ff99SB-ILDN force field are included for comparison.