All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†

A. D. MacKerell, Jr.,*,‡,§ D. Bashford,‡,⊥ M. Bellott,‡,⊥ R. L. Dunbrack, Jr.,‡,⊥ J. D. Evanseck,‡,⊥ M. J. Field,‡,⊥ S. Fischer,‡,⊥ J. Gao,‡,⊥ H. Guo,‡,⊥ S. Ha,‡,⊥ D. Joseph-McCarthy,‡,⊥ L. Kuchnir,‡,⊥ K. Kuczera,‡,⊥ F. T. K. Lau,‡,⊥ C. Mattos,‡ S. Michnick,‡,⊥ T. Ngo,‡,⊥ D. T. Nguyen,‡,⊥ B. Prodhom,‡,⊥ W. E. Reiher, III,‡,⊥ B. Roux,‡,⊥ M. Schlenkrich,‡,⊥ J. C. Smith,‡,⊥ R. Stote,‡,⊥ J. Straub,‡,⊥ M. Watanabe,‡,⊥ J. Wiórkiewicz-Kuczera,‡,⊥ D. Yin,§ and M. Karplus*,‡,||

Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, Department of Pharmaceutical Sciences, University of Maryland, School of Pharmacy, Baltimore, Maryland 21201, and Laboratoire de Chimie Biophysique, ISIS, Institut Le Bel, Université Louis Pasteur, 67000 Strasbourg, France

Received: September 22, 1997; In Final Form: February 6, 1998

New protein parameters are reported for the all-atom empirical energy function in the CHARMM program. The parameter evaluation was based on a self-consistent approach designed to achieve a balance between the internal (bonding) and interaction (nonbonding) terms of the force field and among the solvent−solvent, solvent−solute, and solute−solute interactions. Optimization of the internal parameters used experimental gas-phase geometries, vibrational spectra, and torsional energy surfaces supplemented with ab initio results. The peptide backbone bonding parameters were optimized with respect to data for *N*-methylacetamide and the alanine dipeptide. The interaction parameters, particularly the atomic charges, were determined by fitting ab initio interaction energies and geometries of complexes between water and model compounds that represented the backbone and the various side chains. In addition, dipole moments, experimental heats and free energies of vaporization, solvation and sublimation, molecular volumes, and crystal pressures and structures were used in the optimization. The resulting protein parameters were tested by applying them to noncyclic tripeptide crystals, cyclic peptide crystals, and the proteins crambin, bovine pancreatic trypsin inhibitor, and carbonmonoxy myoglobin in vacuo and in crystals. A detailed analysis of the relationship between the alanine dipeptide potential energy surface and calculated protein ϕ, χ angles was made and used in optimizing the peptide group torsional parameters. The results demonstrate that use of ab initio structural and energetic data by themselves are not sufficient to obtain an adequate backbone representation for peptides and proteins in solution and in crystals. Extensive comparisons between molecular dynamics simulations and experimental data for polypeptides and proteins were performed for both structural and dynamic properties. Energy minimization and dynamics simulations for crystals demonstrate that the latter are needed to obtain meaningful comparisons with experimental crystal structures. The presented parameters, in combination with the previously published CHARMM all-atom parameters for nucleic acids and lipids, provide a consistent set for condensed-phase simulations of a wide variety of molecules of biological interest.

I. Introduction

Empirical energy calculations are of great utility in the study of the structure, dynamics, and thermodynamics of proteins, as well as of other macromolecules of biological interest.[1−3] An essential element is the simplicity of the potential energy function, which makes possible simulations of mesoscopic systems involving tens of thousands of atoms for time scales extending into the nanosecond range or longer. Although many potential functions are now in use, improvements in their accuracy continue to be important. This is of particular concern at present because most of the problems now being investigated by simulation methods require more quantitative results than did much of the earlier work,[2] where more qualitative features were of primary interest. Indeed, the need for more quantitative results from empirical energy calculations, ranging from structural and dynamic information to thermodynamic properties, motivated the development of the CHARMM22 force field for proteins presented in this paper; the "22" signifies that the present parametrization was first included in version 22 of CHARMM, which was released in 1992.

While improving the accuracy, it is desirable to limit the complexity of the potential function so as not to introduce unnecessary increases in the required computer time. The approach we take here is to optimize the parameters for the widely used CHARMM potential energy function without changing the functional form.[4−8] We present all-atom parameters for proteins that have been shown to yield good results in a variety of simulations.[9] The methodology used in the parameter optimization, which differs in certain aspects from that employed by others,[10−12] is consistent with recently published parameter sets for nucleic acids[13] and lipids.[14] Because of the important role of the solvent and explicit solvent

† Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; MbCO, carbonmonoxy myoglobin; NMA, *N*-methylacetamide; rms, root-mean-square.

* Corresponding authors.
‡ Harvard University.
§ University of Maryland.
|| Université Louis Pasteur.
⊥ Present address in Supporting Information.

S1089-5647(97)03084-8 CCC: $15.00
Published on Web 04/14/1998

representations now used in most simulations, considerable emphasis is placed on a balance among the protein−protein, protein−solvent, and solvent−solvent interactions. The approach for achieving such a balance is a refinement of that employed for the nonbonded interactions in the CHARMM 19 (Param 19) polar hydrogen potential energy function.[5,15]

Although the new parameter set uses the same functional form as employed previously in the CHARMM program, the resulting potential energy function is a considerable improvement over earlier functions.[4,7,8,15,16] In contrast to the CHARMM 19 polar hydrogen parameter set,[15] which uses extended atoms for carbons (e.g., a CH_3 group is treated as a single atom), the present potential function includes all atoms explicitly. Further, the parametrization is based on a much broader range of experimental and ab initio data. This yields parameters that are applicable to a wide variety of systems and reduces complications due to correlations among the parameters. The present parameter set was optimized for the protein main chain and for the individual side chains by detailed analyses of one or more small model compounds for each case. The backbone parameters used *N*-methylacetamide (NMA) and the alanine dipeptide; histidine parameters are based on imidazole, 4-methylimidazole, and imidazolium; valine, leucine, and isoleucine are based on small aliphatic compounds including ethane, propane, butane, and isobutane; and so on. Such a strategy ensures that the parameters for each amino acid are fully optimized with respect to the available data. All of the parameters were optimized by the same self-consistent procedures described here for the protein backbone, except those for the aromatic side chains, which were taken with some slight modifications from the values published by Jorgensen.[17a] It should be noted that the explicit representation of hydrogens in aromatic rings is necessary to produce the quadrapole moment required for reproduction of aromatic−aromatic interactions seen in small peptide crystals.[17b] Details of the studies made for parametrization of the individual amino acid side chains will be published elsewhere.

The present paper describes the philosophy used in the parametrization and gives the details of the parameter evaluation for the protein backbone. It also presents the results obtained by using the parameters for simulations of liquid NMA and of a number of peptides and proteins in solution and in a crystal environment. The resulting data elucidate a number of important physical effects, including the importance of nonbonding contributions in determining structures and vibrations and the need for balance of intramolecular and intermolecular terms and among the solvent−solvent, solvent−solute, and solute−solute contributions to the intermolecular terms. The present work shows that potential energy functions of the type used in the present study must be optimized, in part, with respect to condensed-phase properties; i.e., use of ab initio results by themselves is not sufficient.

The function and the parameters obtained are implemented in CHARMM 22 and subsequent versions of that program. The protein parameters, together with those for nucleic acid,[13] lipids,[14] and carbohydrates (in progress) form a consistent optimized set for a wide range of biomolecules. The CHARMM program includes the potential energy function we describe here and the other aspects of the molecular model that are required for a full description (e.g., cutoff values). Only with this information is it possible to repeat a calculation. The program is available to not-for-profit institutions at a nominal charge.[18]

Section II presents the potential function and the philosophy of the parameter development. Section III describes the methods used for minimization and dynamical simulations to determine and test the parameters. The results obtained for the various test cases are presented in Section IV. Section IV.a focuses on the parametrization of the peptide backbone, while Sections IV.b−IV.d present applications to tripeptide crystals, cyclic peptide crystals, and proteins, respectively. Conclusions and future directions are given in Section V. The full all-hydrogen parameter set is given in an Appendix that is presented as Supporting Information.

II. Parametrization Methodology

Calculations were performed with the simulation program CHARMM,[4] in which an empirical energy function that contains terms for both internal and external interactions was used. The energy function has the form

$$U(\vec{R}) = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{UB}} K_{UB}(S - S_0)^2 + \sum_{\text{angle}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\chi(1 + \cos(n\chi - \delta)) + \sum_{\text{impropers}} K_{\text{imp}}(\varphi - \varphi_0)^2 + \sum_{\text{nonbond}} \epsilon\left[\left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\epsilon_l r_{ij}} \quad (1)$$

where K_b, K_{UB}, K_θ, K_χ, and K_{imp} are the bond, Urey−Bradley, angle, dihedral angle, and improper dihedral angle force constants, respectively; b, S, θ, χ, and φ are the bond length, Urey−Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, with the subscript zero representing the equilibrium values for the individual terms. Coulomb and Lennard-Jones 6−12 terms contribute to the external or nonbonded interactions; ϵ is the Lennard-Jones well depth and $R_{\min}$ is the distance at the Lennard-Jones minimum, q_i is the partial atomic charge, ϵ_l is the effective dielectric constant, and r_{ij} is the distance between atoms i and j. The Lennard-Jones parameters between pairs of different atoms are obtained from the Lorentz−Berthelodt combination rules, in which ϵ_{ij} values are based on the geometric mean of ϵ_i and ϵ_j and $R_{\min_{ij}}$ values are based on the arithmetic mean between $R_{\min_i}$ and $R_{\min_j}$. Because of the role of electrostatic contributions in determining intramolecular, as well as intermolecular, energetics (as described below), the effective dielectric constant ϵ_1 must be set equal to unity in this potential energy function since otherwise an unbalanced parametrization will be obtained, particularly for the peptide group. This contrasts with the polar hydrogen parameter set, PARAM 19,[5,15] in which it is appropriate to introduce a distance-dependent dielectric parameter. For the CHARMM 22 set, neutralized charged groups can be introduced to mimic some aspects of the shielding from a high dielectric constant solvent.[120]

Given $\vec{R}$, the vector of the coordinates of the atoms, the various distances and angles required to evaluate $U(\vec{R})$ in eq 1 are readily determined. All possible bond angles and dihedral angles are included in $U(\vec{R})$, while a limited number of Urey−Bradley terms and improper dihedral angles are used to optimize the fit to vibrational spectra. As can be seen from eq 1, only the quadratic term is included in the Urey−Bradley function; this is in accord with an analysis[19] that shows the linear term can be omitted when Cartesian coordinates are used and the minimum energy structure is employed for determining the vibrational frequencies. Nonbonded interaction terms are

included for all atoms separated by three or more covalent bonds. No general scaling of the electrostatic or Lennard-Jones interactions for atoms separated by three bonds (the so-called 1−4 term) is used. In specific cases there is scaling of the 1−4 Lennard-Jones term; examples include the aliphatic carbons and the amide nitrogen and oxygen atoms. No explicit hydrogen-bond term is included because the Coulomb and Lennard-Jones terms can accurately represent the hydrogen-bonding interactions.[5,15] The water model used in all calculations is the TIP3P model[20] modified for the CHARMM force field.[5] The consistency of the protein and solvent interactions is based on the use of this water model; i.e., it forms part of the system description and other water models would be less appropriate.

II.a. Parametrization Strategy. Development of parameters for empirical potential energy functions, such as that in eq 1, requires a coherent strategy. The present work is an attempt to optimize the parameters by the use of a wide range of information in a consistent fashion. Self-consistency among the different terms in the potential energy function was achieved by iterative optimization of the parameters. Typically, initial values of the intermolecular parameters (Coulomb and Lennard-Jones) were chosen from previous CHARMM parameter sets[4,5,21] or based on the reproduction of ab initio interaction calculations on rigid monomers. Given these values, the intramolecular parameters (bond length, Urey−Bradley, bond angle, dihedral angle, and improper dihedral angle terms) were determined by using structural and vibrational data for the model compounds. The resulting structures were used for optimization of the intermolecular parameters relative to interaction energies and condensed-phase properties of model compounds. With the improved interaction parameters, the structures, vibrational spectra, and energy surfaces of the model compounds were reoptimized by adjusting the internal parameters. This iterative process was repeated until convergence of the parameters was achieved.

Intramolecular Terms. Geometries are dominated by the equilibrium values for the bond length and bond angle terms and by the dihedral term phase and multiplicity. These parameters were optimized by fitting to gas-phase structures from microwave and electron diffraction data or crystal structures from X-ray data. In the case of X-ray structures, care was taken in the interpretation of the individual crystal structures to account for the influence of intermolecular interactions on the intramolecular geometries. Ideally (e.g., for imidazole), both gas-phase and crystal data are used. Such a combination allows for parameter optimization in the gas phase followed by testing of the parameters with the crystal structure where intermolecular interactions, as well as the intramolecular parameters, influence the geometry. Ab initio data, especially for ionic species such as acetate, guanidinium, imidazolium, and methylammonium, were introduced to supplement the experimental geometric data. In many cases, survey results of crystal structures in the Cambridge Crystal Data Bank (CCDB)[22] were used to determine the range of the allowed geometric values (e.g., for indole, pyrrolidine); i.e., if a large number of fragment structures are available, the average geometries tend to diminish the distortions associated with crystal interactions. Such averages are, in fact, preferable to gas-phase data in some cases because they contain contributions to the geometry associated with condensed-phase effects. An example of particular importance for proteins arises in the determination of the peptide backbone parameters (see Section IV.b).

Adjustment of the parameters was performed manually, although in certain cases (e.g., for proline) automated procedures were employed. We have found that automated procedures must be used with great care owing to the extensive nature of parameter space, correlation among the parameters, and their underdetermined nature. An automated least-squares procedure often leads to a combination of "unphysical" parameters that reproduce the input data. More meaningful parameter values, which have a wider range of applicability, were obtained manually with "reasonable" parameter ranges for the optimization in the iterative refinement procedure described above.

Once satisfactory geometries were obtained, the force constants associated with the bond length, bond angle, dihedral angle, and improper torsion terms were adjusted by fitting vibrational data for the model systems. Gas-phase infrared and Raman data were the primary sources of such data. Solution and crystal data were used in certain cases, particularly for ionic species for which few gas-phase vibrational data are available. In the solution results, attention was paid to interactions that could influence the experimentally determined vibrational frequencies and efforts were made to account for condensed-phase contributions; an example is the NH stretch associated with the peptide backbone. The results of ab initio calculations were introduced where necessary to supplement the experimental data. One area where ab initio calculations were essential is in the assignments of experimental vibrational frequencies to internal coordinates. Only limited isotopic substitution data are available from many cases, and there are often ambiguities in the interpretation of the data because many normal modes contain contributions from the same internal coordinates. Ab initio results were also used to obtain values for low-frequency torsional modes that are difficult to observe experimentally. Finally, for crystal or solution measurements, isolated molecule ab initio results were used as an aid in determining the contribution of intermolecular interactions to the observed vibrational spectra. In particular, the optimization of force constants associated with the ionic side chains was significantly aided by the ab initio data.

Scaled HF/6-31G(d) ab initio values were used for the vibrational calculations. When feasible, the scaling factor was determined by comparison of known experimental frequencies with the ab initio results; the derived factor was then applied to the unobserved frequencies. Where this was not possible, a scale factor of 0.9 was introduced because it has been found to give good results in other studies.[23] Analysis of the vibrational spectra and potential energy distributions were made with the MOLVIB program (J. Wiórkiewicz-Kuczera and K. Kuczera, unpublished results). Availability of assignments from the potential energy distributions along with frequencies allowed both to be taken into account during optimization of force constants.

Following adjustment of force constants to fit the vibrational data, the minimized geometries were rechecked and adjustments were made to both the equilibrium parameters and the force constants in an iterative fashion, as pointed out above. Final optimization of the vibrational spectra was done by the addition of Urey−Bradley and improper terms in cases where the agreement between the calculated results and the available data was unsatisfactory. Urey−Bradley terms were important for the in-plane deformations as well as separating symmetric and asymmetric bond stretching modes (e.g., in aliphatic molecules). Improper dihedral terms aided mainly in the accurate reproduction of out-of-plane modes such as the wagging modes of the imidazole hydrogens, and in the amides, such as *N*-methylacetamide and acetamide.

In addition to the geometries and vibrational frequencies of specific structures, the relative energies of different backbone and side chain conformers are important. Examples include torsional surfaces for carbon−carbon bonds, methyl and ethyl rotations in model compounds such as 4-ethylimidazole and ethylbenzene, and the alanine dipeptide map. Although conformational analysis[24] can give an insightful description of the relative free energies of different side chain conformers, more quantitative results are needed for a potential energy function. Consequently, detailed calculations were made for the energies of side chain conformers. Also, in some cases large deviations from the minima can occur during molecular dynamics simulations so that it is necessary to have a more complete knowledge of the potential surface than that obtained from the relative energies of the minima and from their vibrational frequencies. An example is the out-of-plane distortion of aromatic hydrogens, including the $H_{\epsilon 1}$ atom in tryptophan, where deviations of 15° introduce strain energies of less than 1 kcal/mol. Thus, the intramolecular and intermolecular parameter optimization included information from adiabatic energy surfaces where appropriate. Such data made it possible to adjust the parameters so as to describe energy barriers and the positions of saddle points, as well as the minimum-energy structures used in the vibrational analysis (e.g., rotation of the side chain hydroxyl group in tyrosine). Experimental gas-phase data were used in many cases, and ab initio calculations were made to obtain surfaces for which no satisfactory experimental data for barriers were available (e.g., proline). Changes in the structure (e.g., bond elongation and angle opening) as a function of a dihedral angle can be important[25] and were obtained from the ab initio calculations. Such information was used for optimizing and testing the accuracy of the potential function in reproducing structural distortions, as well as energetic differences. Satisfactory agreement was obtained in most cases for both the vibrational frequency and the torsional barrier from the combined contributions of dihedral and nonbonded terms. In certain cases, compromises were made because a single dihedral term could not describe vibrational data and the energy barriers; e.g., the H−C−C−H torsional frequency in ethane was slightly elevated as compared to the experimental value to allow the rotational energy surface to be accurately modeled. In some cases that were regarded as particularly important (e.g., the dipeptide potential surface), more than one term was used for the dihedral angle potential in eq 1.

Intermolecular Terms. Intermolecular parameter optimization involves the van der Waals and electrostatic interactions. The objective was to obtain a set of parameters that result in balanced protein−protein, protein−water, and water−water interactions. Interaction energies and structural data for model dimer systems and macroscopic properties of pure liquids and solutions were used in the parameter determination.

The water model and water−water interactions were taken as the basis of the parametrization. As a first step, a number of published water models were tested, including the TIP3P,[20] TIP4P,[20] and extended SPC/E[26] models as well as alternate models (Gao, J.; MacKerell, A. D., Jr.; Karplus, M., unpublished results). On the basis of a comparison of the TIP3P and SPC/E models with ab initio calculations, a softer repulsive van der Waals term was examined. Although r^{-9} and r^{-10} repulsions yielded good results, there was no significant improvement, relative to r^{-12}, for either the energetics or the structure of liquid water. Consequently, it was decided to retain one of the previously published water models, and the TIP3P model was selected. This model[20] satisfactorily reproduces the first-shell hydration and the energetics of liquid water, although the tetrahedrality is too weak and the diffusion constant is somewhat too high. The SPC/E model, which also reproduces the first hydration shell and is somewhat better for the tetrahedrality, was not used because it has an inconsistency when applied to heterogeneous solutions. In the SPC/E model, a correction is made to account for the overestimation of the interaction energy due to the omission of electronic polarization. Such a correction is reasonable with respect to pure solvent properties but can lead to problems for solution simulations; i.e., the solute does not "know" that the energy correction is present in the water−water interactions. For a proper balance of water−water and water−solute interactions, the solute charges would have to be increased, thereby leading to a possible overestimation of the solute−solute interactions and incorrect results for the calculated properties of the solute itself. The TIP4P model, although it gives excellent results, was not used because it includes a virtual particle, which complicates the treatment because the forces have to be projected onto the "real" atoms. Moreover, because most of the simulation time of a solvated protein is spent on simulating the water molecules, the less costly TIP3P model was utilized. It should be noted that use of the CHARMM 22 parameter set with water models other than TIP3P may lead to inconsistencies because the water−protein and protein−protein intermolecular parametrization may not be well-balanced.

Given the water−water interaction for the TIP3P model, the solute-water interactions were optimized on the basis of ab initio results for interactions of complexes and experimental data for macrosopic systems, including thermodynamic parameters and molecular volumes. Ab initio calculations were performed to determine the minimum interaction energies and geometries between a water molecule and a model compound, primarily at sites involving hydrogen-bonding interactions with polar atoms. To determine the partial atomic charges, the interaction between water and all polar sites of the model compounds were examined. Typically, the isolated model compounds were optimized at the HF/6-31G(d) level. The optimized structures were then used for a series of supermolecular calculations involving the model compound and a single water molecule at each of the various sites. The HF/6-31G(d) optimized structure was replaced with an experimental gas-phase structure if available; the gas-phase water structure corresponding to the TIP3P model was used in all cases.[27a] The supermolecule structure was optimized at the HF/6-31G(d) level by varying the interaction distance and, in certain cases, a single angle, to find the local minimum for the water position with fixed monomer geometries. From the resulting structure, the interaction energy was calculated as the difference between the total supermolecule energy and the sum of the individual monomer energies; no corrections for basis-set superposition error were made. This approach is essentially that introduced by Reiher and Karplus.[5,15] It was subsequently adopted by Jorgensen and co-workers[8,28] and most recently used in the development of the MMFF energy function by Halgren.[29] In the present force-field development and in the work of Halgren[29] but not that of Jorgensen and co-workers,[8] we follow Reiher and Karplus[5,15] in scaling the calculated ab initio values used for the parametrization of the interactions between neutral polar molecules and water. This adjustment takes account of limitations in the level of the ab initio theory being employed and the neglect of many-body polarization in liquid water. Limitations in the HF/6-31G(d) level of theory include omission of the dispersive (attraction) term in the Lennard-Jones interaction, the use of fixed geometries, the relatively small size of the basis set and

the omission of corrections for basis-set superposition error. These limitations lead to a cancellation of errors so that the calculated minimum interaction energy and distance are −5.98 kcal/mol and 2.98 Å, respectively, for the gas-phase water dimer,[5] in satisfactory agreement with the experimental values of 5.4 ± 0.7 kcal/mol and 2.98 Å.[27b] Similar accuracy is obtained, probably fortuitously, with the HF/6-31G(d) level calculations for other polar systems and is the primary reason the relatively inexpensive level of theory was used as the basis for the parametrization. For the condensed phase, neglect of many-body polarization leads to the ab initio interaction energy being underestimated and the minimum distance overestimated as compared to the condensed phase, in agreement with previous work on a wide variety of molecules.[8,13,14] To overcome the underestimation of the condensed-phase interaction energy, a scaling factor was introduced. The scaling factor was obtained from the ratio of the empirical interaction energy of the TIP3P water dimer to the HF/6-31G(d) water dimer interaction energy. The resulting value, 1.16, was used to scale the ratio of the water to model compound ab initio interaction energies to be consistent with the TIP3P dimer model, so as to obtain a balance of the solute−water and water−water intermolecular interactions.[30] The use of a single scale factor makes the assumption that dispersion effects and polarizabilities are constant for the compounds being parametrized and that the aqueous solvent environment is being used for all calculations. Recent work,[31] using a combined quantum mechanical/molecular mechanical approach, has shown that the electronic polarization contribution to the electrostatic interaction energy varies linearly with the total interaction energies for solvated molecules; this supports the simple scaling model. When available, experimental data on interaction energies from mass spectrometry[32−35] were used in addition to the ab initio results. For charged species, no scaling was applied since the HF/6-31G(d) interaction energies themselves yield charge distributions that give satisfactory agreement with heats and free energies of solvation.[8,36] To compensate for the overestimation of the minimum interaction distances in the Hartree−Fock model due to the absence of the dispersion contribution and neglect of many-body effects, the minimum distances were assumed to be about 0.2 Å shorter than the HF/6-31G(d) values. Such an approach is consistent with the TIP3P water model[20] and pure liquid simulations for which the shorter distance is required to obtain the correct density.[37]

Once the interaction energies and minimum energy geometries for the model supermolecules had been determined from ab initio calculations, the partial charges on the atoms of the model compounds were adjusted to reproduce those values. For consistency, the water (TIP3P) and model compound geometries were kept fixed and only the one or two geometrical parameters used in the ab initio calculation were varied in the structural optimization of the complex with the CHARMM force field. The initial model compound geometries were those used in the ab initio calculations; the CHARMM optimized geometries were used during subsequent iterations. Use of the CHARMM optimized geometry ensured that the final partial atomic charges were consistent with the intramolecular portion of the force field. Initial partial charges were obtained from a Mulliken population analysis of the HF/6-31G(d) wave function. In addition to energies and distances, the magnitudes and directions of the dipole moments of the model compounds were used in the fitting procedure. If available, experimental gas-phase dipole moment values were used; if not, ab initio values at the HF/6-31G(d) level were adopted. As with the TIP3P water model, where the empirical dipole (2.35 D) is larger than the experimental gas-phase value (1.86 D), the charges of the model compounds were adjusted such that the empirical dipole moments were somewhat larger than the experimental or ab initio values. Once satisfactory agreement with all of these data had been obtained, condensed-phase simulations were performed to refine the van der Waals parameters. Pure solvent simulations of aliphatic and polar neutral compounds were used to calculate heats of vaporization and molecular volumes that could be compared with experimental data. Generally, only small adjustments in the van der Waals parameters were required to obtain satisfactory results. In certain cases, crystal simulations were performed to determine heats of sublimation and unit cell parameters; these were also used in refining the van der Waals parameters.[13] Following any adjustment of the van der Waals parameters, the supermolecule energies and distances were recalculated and adjustments made in the charges where necessary. In the present force field the CHARMM TIP3P van der Waals parameters are used for both the solvent−solvent and solvent−solute interactions. This is in contrast to PARAM19, where the TIP3P van der Waals parameters for solute−solvent interactions differ from the CHARMM TIP3P pure-solvent van der Waals parameters.[5]

To simplify the procedure and to allow for the transfer of the parameters from the model compounds to larger units such as amino acids, several assumptions were made in the parameter optimization. Charges were selected to yield "groups" of unit charge (0, ±1, as appropriate). As well as aiding in the transfer of the charges to larger molecules, this simplifies the treatment of long-range electrostatic interactions via multipole expansions.[38] The groups optimally contained five atoms or less; in certain instances larger groups were required (e.g., imidazole) to obtain satisfactory fits. Adjustment of the charges upon linking the model compounds to form larger entities was performed by adding the charge of the deleted hydrogen atom to the heavy atom to which it was previously attached. This approach maintains the unit charge groups from the original model compounds. Aliphatic intramolecular and intermolecular parameters were used without adjustment for all aliphatic moieties of amino acid side chains (e.g., for all C_β carbons), as well as for the nucleic acid and lipid parameter sets.[13,14] Simulations of the aqueous solvation of small aliphatic molecules have shown that the charge distribution has a negligible effect on pure solvent heats of vaporization and crystal heats of sublimation (S. Fischer and M. Karplus, unpublished results). Results from condensed-phase simulations of peptides and proteins presented below indicate that the use of unit charge groups does not have an adverse impact on the accuracy of the final intramolecular parameters.

Given the above assumptions, a hierarchical approach can be used for the extension of the parameter set to other molecules. Each parameter is optimized in the "best" possible model compound; "best" is defined by the nature of the compound and the available data. Once a specific parameter has been optimized, it is not changed when it appears in the corresponding groups of other compounds. As chemically similar molecules are introduced, the available parameters are employed as much as possible. Often, the connectivities of the new molecules (e.g., new bond angles) are such that additional parameters can be added without destroying the consistency. This allows some degree of flexibility in the parametrization for new systems. If the fit obtained from parameters to the data pertaining to the new molecule is not of sufficient accuracy, new atom types can be introduced. These new atom types allow for the introduction of new internal parameters, so that the optimization can be

improved, while ensuring that the results for other molecules are not compromised. Efforts are made to keep the number of atom types to a minimum. However, as the ultimate goal of this parameter set is the quality of fit to a wide range of data, the total number of atom types has increased over earlier parametrizations, i.e., the current protein parameter set contains 55 atom types as compared to 29 in CHARMM 19 (see Appendix for the list of atom types).[5,15]

III. Methods Used for Simulations of the Test Systems

Ab initio Hartree−Fock calculations were performed with various versions of Gaussian;[39] Gaussian 80, 88, 90, 92, and 94 were used. Optimizations of the molecular structures were performed by either the Berny or the Murtaugh−Sargent algorithm to the default tolerances. Interaction energies and geometries for optimized NMA and water in two different hydrogen-bonding positions (an NH as a donor and CO as an acceptor) and the NMA dimer were calculated on the basis of the fixed 6-31G(d) NMA geometry and the experimental geometry of water.[27] In the NMA−water interaction, the hydrogen-bond distance and a single angle were varied in the optimizations (see Figure 3 below); all other degrees of freedom were fixed. For the NMA dimer only the hydrogen-bond distance was optimized. The interaction energy was defined as the difference between the total energy of the supermolecular complex and the sum of the monomer energies; no basis set superposition error corrections were included.

Liquid NMA and NMA dissolved in water were simulated with the BOSS program[40] using Metropolis sampling in the NPT ensemble. The combination rules of the BOSS program were modified to the Lorentz−Berthelodt rules used in the CHARMM force field (see above). The pure solvent system consisted of 128 NMA molecules in a cubic cell with an edge length of approximately 26 Å, subjected to an external pressure of 1 atm and a temperature of 100 °C. Averages were obtained over 2 million configurations after an initial equilibration period of 1 million configurations. Both the heat of vaporization and the liquid density were calculated. For determining the heat of solution in water, the NMA molecule was placed in the center of a periodic box consisting of 267 water molecules and the water molecules that had interaction energies with NMA greater than 5×10^3 kcal/mol were removed. Equilibration was performed over 10^6 configurations followed by the evaluation of 1.5×10^6 or 3×10^6 configurations for averaging. The cutoffs used were 9.5 Å for solute−solvent interactions and 8.5 Å for solvent−solvent interactions with a 1.0 Å switch region for both van der Waals and electrostatic interactions; these cutoffs are default values in the BOSS program. The heat of solvation of NMA in water was calculated from the difference between the average energy of NMA in aqueous solution and the average energy of the same number of water molecules in the absence of NMA.

Several different types of systems were used for testing the parameters. Vacuum calculations were performed on crambin, BPTI, and carbonmonoxy myoglobin and crystal calculations were performed on tripeptides, cyclic peptides, crambin, BPTI and carbonmonoxy myoglobin. In all of the simulations, covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm.[41] For the condensed-phase calculations, truncation schemes for both the van der Waals and the electrostatic interactions were introduced. Hydrogens not present in the crystal structures were positioned on the basis of the default internal coordinates in the parameter set; water hydrogens and any other hydrogens that did not have unique positions (e.g., for many of the protein side chains) were placed using the CHARMM HBUILD facility.[42] Details of the setup, minimization, and simulation techniques are given below.

Crystal minimizations and simulations were performed with the CRYSTAL module in the CHARMM program.[43] Truncation of the nonbonded interactions was introduced by using an atom-based shifting function for the electrostatic interactions and a atom-based switching function for the van der Waals interactions with the IMAGE atom list cutoff set to be 1 Å larger than the nonbond list cutoff.[4] Comparisons of minimizations were made for the tripeptides and cyclic peptides with different cutoff distances; these ranged from 10 to 25 Å, as described below. In the peptide crystal minimizations, the lattice parameters and heavy atom positions were initially fixed for 50 adopted-basis Newton Raphson (ABNR)[4] minimization steps to optimize the hydrogen atom positions that were either poorly determined in the X-ray structures or placed in standard positions. All atoms were then allowed to relax with the lattice parameters fixed for an additional 200 ABNR steps. This was followed by a full minimization including the lattice parameters that was terminated when the rms gradient averaged over the minimization was 10^{-6} kcal/mol/Å or less or up to 1000 ABNR steps; the final rms gradients are reported for the various systems.

Constant volume, NVT, and constant pressure, NPT, simulations on the tripeptides and cyclic peptides were performed on the asymmetric unit using the temperature and pressure coupling scheme of Berendsen and co-workers[44] as implemented in CHARMM in conjunction with the leapfrog integrator. In the simulations, a time step of 1 fs was used with a temperature coupling constant of 0.1, a pressure coupling constant of 10, and an isothermal compressibility of 5×10^{-5} atm^{-1}. These values were selected to yield a stable temperature and pressure for the system during the simulations, while keeping the influence of the coupling to a minimum. Prior to the simulations, the crystals were submitted to a 50-step ABNR minimization of the hydrogen atoms followed by a 200-step ABNR minimization of all atoms with the lattice parameters fixed, as used in the minimization studies. Simulations were performed for 50 ps, and the coordinates were saved every 100 steps (0.1 ps) for analysis. Both internal and external pressures were monitored during the NVT and NPT simulations. Internal pressures were calculated from the forces on the primary atoms, and the external pressures were obtained from the difference between the total forces due to both image and primary atoms and the primary atom forces.[1] The internal and external pressures are expected to be approximately equal; a negative pressure indicates that the volume of the system would contract in a NPT simulation. The rms fluctuations of the pressures were up to 1 order of magnitude larger than the average pressure, as expected for a system of this size.

Molecular dynamics simulations of the proteins were performed using the leapfrog algorithm as implemented in CHARMM for the both the vacuum and crystal simulations. When specified, SHAKE was applied to all covalent bonds involving hydrogens.[41] Vacuum simulations were initiated with a 5-ps heating period in which the velocities were increased in increments of 6 K every 0.1 ps. This was followed by a 5-ps equilibration period in which a ± 5 K window was applied to the temperature and checked every 0.1 ps; if the temperature was out of range, velocity scaling was performed. The production run, unless specified, was continued for 300 ps without velocity scaling.

Crystal simulations on crambin, BPTI, and carbonmonoxy myoglobin were performed using the following protocol for generating the simulation system in the NVT ensemble. The starting configuration for the crambin crystal studies corresponded to the 0.94 Å X-ray coordinates with an *R*-factor of 0.104, including two ethanols and 86 water molecules.[45] Five additional water molecules had been added to fill vacuum points of the crystal, using the CHARMM19 parameter set (John Kuriyan, personal communication); this yielded a simulation system consisting of 933 atoms. BPTI calculations were initiated from the joint neutron and X-ray refined structure at resolutions of 1.8 and 1.0 Å and *R*-factors of 0.197 and 0.200, respectively.[46] Vacuum points in the BPTI crystal were filled by water molecules. This was performed by generating the primary atoms, as defined by the asymmetric unit, and all crystal image atoms within 13 Å of the primary atoms. The BPTI asymmetric unit cell was overlayed by a TIP3P box of dimensions $18.5 \times 23.4 \times 28.7$ Å. All water molecules whose oxygen atoms were within 2.8 Å of any of the primary or image non-hydrogen atoms were deleted. The resulting system was used as the starting configuration for the simulation. This system included 892 BPTI atoms, 6 atoms of a dianionic phosphate, 63 crystal waters, and 29 added water molecules for a total of 1174 atoms. The starting configuration was subjected to 50 steepest-descent (SD) steps followed by 5 steps of Powell minimization with all heavy atoms fixed to their initial positions in the presence of the crystal images; SHAKE was applied to bonds involving hydrogens. This was followed by a 50 SD step minimization of all atoms followed by 5 Powell steps with SHAKE, again in the presence of the crystal images. Carbonmonoxy myoglobin simulations started with the 1.5 Å crystal structure (*R*-factor = 0.171) obtained at 260 K.[47] The bound carbon monoxide and a sulfate ion present in the crystal were included in the simulation. As with BPTI, vacuum points in the crystal structure were filled with waters using a $20.7 \times 18.7 \times 18.7$ Å water box; the box size was chosen to cover one asymmetric unit following the methodology presented above for BPTI. This procedure added 208 water molecules in addition to the 137 water molecules identified in the X-ray study, yielding a total of 3574 atoms in the system. Following the addition of the waters the same minimization protocol as that applied to BPTI was used.

All crystal simulations were performed by gradually heating the system over a 5-ps period by increasing the temperature every 0.1 ps to final temperatures of 300, 285, and 260 K for crambin, BPTI, and carbonmonoxy myoglobin, respectively, in accord with the temperatures used for the structure determinations. This was followed by 5 ps of equilibration using a ± 5 K window with testing every 0.1 ps. If the temperature was outside of the window, the velocities were scaled to bring the temperature back to 300 K. Production trajectories were performed for 100 ps without velocity scaling. The integration time step was 1 fs. Coordinates were saved every 0.1 ps for analysis.

Analysis of the protein simulations was similar to that used previously with emphasis on the aspects of the results that test both the intramolecular and intermolecular contributions to the potential function. The rms differences were calculated for the specified non-hydrogen atoms following a least-squares fit of the backbone (C, N, C_α, O) atoms except where the rms difference associated with C_α atoms is reported. In that case only the C_α atoms were used in the least-squares fit. The rms fluctuations were calculated following reorientation of all non-hydrogen atoms in each time frame to the starting coordinates to ensure that translational and rotational motions of the protein did not contribute to the calculated fluctuations. This is done because the simulations were not long enough to provide a full sampling of the overall motion that does occur in the crystal. Thus, the calculated values are lower limits for the overall atomic fluctuations.[48]

Average values and fluctuations of internal coordinates were obtained by averaging over the individual time frames of the trajectories. Use of average differences, in addition to rms differences, exposes systematic trends introduced by the parameters. Time-averaged structures were obtained from the production portions of the simulations. Nonbonded interaction distances were calculated on the basis of a previously described approach for hydration.[13] The interactions are analyzed in terms of heavy atom-to-heavy atom (donor-to-acceptor) distances within a cutoff distance to avoid unphysical contributions. In the present study, a cutoff distance of 3.5 Å was employed. This distance corresponds to the first minimum in the TIP3P water model O-to-O radial distribution function[20] and is assumed to represent the outer limit of the first hydration shell. The same cutoff was used in the nucleic acid parametrization paper[13] and in a study of nonbonded interactions in proteins based on a survey of the Brookhaven Protein Data Bank.[49] Applying this cutoff distance allows for the average interaction distance and hydration or occupancy number (see legend of Table 23 below) to be obtained from the experimental X-ray structures, even though only a limited number of such interactions may be present.[50] For consistency, the same approach was used for analysis of the X-ray data and dynamics simulations. Average distances and hydration or occupancy numbers were obtained over the individual time frames in the trajectory and normalized with respect to the number of time frames and the types of atoms included in the analysis.

IV. Results and Discussion

The parametrization of the protein potential energy function was based on sets of small model compounds that are appropriate to represent the main chain and the amino acid side chains. The main chain parametrization and testing are presented in the first part of this section (Section IV.a). Details of the side chain parametrization are given in separate papers that are in preparation. The entire set of protein parameters is listed in the Appendix. The results obtained in testing the parameters on tripeptides, cyclic peptides, and proteins are presented in Section IV.b, IV.c, and IV.d, respectively. We have chosen the systems for their intrinsic interest, because extensive data were available for them, and/or because they have been used in testing other protein parameter sets (e.g., cyclic peptides, crambin). The analysis concentrates on comparisons with experimental results that test both the intramolecular and intermolecular portion of the potential function. Special attention is paid to intermolecular interactions involving water molecules. Crystal studies were performed on the noncyclical tripeptides Gly-Ala-Val·$3H_2O$ and Gly-Ala-Leu·$3H_2O$,[51] which are in helical conformations, and Ala-Ala-Ala, which is a parallel pleated sheet model.[52] These peptides include water molecules and ionic functional groups. Cyclic peptide crystal minimizations and simulations were performed on Ala-Ala-Gly-Gly-Ala-Gly·H_2O and Ala-Ala-Gly-Ala-Gly-Gly·$2H_2O$,[53] Gly-Gly-D-Ala-D-Ala-Gly-Gly·$4H_2O$,[54] $(Gly\text{-}Pro\text{-}Gly)_2$,[55] Gly-Pro-Gly-D-Ala-Pro,[56] and $(Cys\text{-}Gly\text{-}Pro\text{-}Phe)_2\cdot 4H_2O$.[57] Protein test calculations were made for crambin, the bovine pancreatic trypsin inhibitor (BPTI), and carbonmonoxy myoglobin in a vacuum and in a crystal environment.

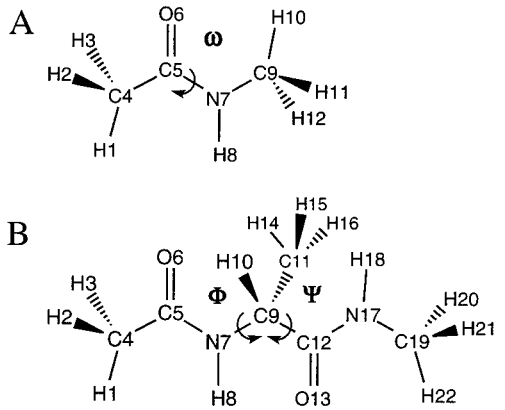

Figure 1. Structures of (A) *N*-methylacetamide (NMA) and (B) the alanine dipeptide. Atom names represent the nomenclature used in the text.

IV.a. Protein Backbone. Accurate parametrization of the protein backbone is essential for the overall quality of the potential energy function of peptides and proteins. Two molecules were selected as the model compounds for the parametrization of the peptide backbone. The first is *N*-methylacetamide (NMA), which is the simplest peptide model; it contains a single peptide bond that is methylated on the carbonyl carbon and the amide nitrogen. This results in a system that is closer to the interior peptide bond of a protein than models such as formamide or acetamide that have been used in previous studies. Experimental data for NMA include structural and vibrational measurements as well as thermodynamic data for liquid NMA and for NMA in aqueous solution. Most of the optimization of the peptide group interaction parameters were based on NMA. The second model system is the alanine "dipeptide", which contains two peptide linkages, again methylated. Detailed analysis of changes in structural and energetic properties associated with variations in the ϕ and ψ angles were made. Of particular interest are the $C7_{eq}$, $C7_{ax}$, and C5 conformations, which are the three minima for the dipeptide in a vacuum that are typically used for the study of protein backbone energetics. Although little is known experimentally about the conformational properties of the alanine dipeptide, reasonably high-level ab initio calculations are now available for it and for some closely related model systems. These theoretical results were used in the parametrization, instead of energy estimates based on the ϕ, ψ distributions observed in proteins that have been used in some other studies.[58,59] Diagrams of NMA and the alanine dipeptide, including the atom-naming convention, are shown in Figure 1. Because of its special covalent interactions, specific parameters were developed for the peptide bond of proline residues.

IV.a.1. Internal Parametrization. As mentioned in the Introduction, the internal parametrization of the peptide backbone is complicated by the important structural changes that occur in going from the gas-phase to the condensed-phase environment. The most important difference involves a significant shortening of the CN bond. This arises from the increased contribution of the resonance structure with a CN double bond when the C═O and N−H groups are involved in hydrogen-bonding interactions, as they generally are in proteins. The effect is most clearly demonstrated in the NMA crystal structure and in a comparison between theoretical calculations of NMA by itself and of NMA hydrogen bonded to water molecules. In fact, a theoretical prediction[60,61] that the standard NMA crystal structure[62] was incorrect has been confirmed recently by a new structure determination.[63] Such changes in electronic structure between the gas phase and the condensed phase are difficult to represent in empirical potentials without complicating the potential function by introducing polarization. Because the primary focus of the present parametrization is to develop a model for the peptide backbone for proteins and for condensed-phase simulations, in general, the optimization of the internal force field was done for NMA and the alanine dipeptide with their condensed-phase geometries. This assumes that in peptides and proteins the hydrogen-bonding propensities are generally satisfied either by internal hydrogen bonds or by hydrogen bonds to water.[49] Ab initio and experimental geometries for NMA and the alanine dipeptide were employed in the optimization, along with survey results on proteins.[64] Peptide backbone geometries in the Cambridge Crystal Data Base[22] were also used in the parameter development.

TABLE 1: Geometric Data on *N*-Methylacetamide[a]

		experimental			MP2/6-31G(d)[b]		
	CHARMM	gas[c]	crystal[d]	survey[e]	gas	$3H_2O$	H_2O, 2FM
			Trans				
			Bonds				
C4−C5	1.481	1.520(5)	1.515(3)	1.52(1)	1.514	1.510	1.512
C5−N7	1.339	1.386(4)	1.325(3)	1.33(1)	1.365	1.339	1.337
N7−C9	1.444	1.469(6)	1.454(3)	1.45(2)	1.448	1.454	1.454
C5═O6	1.223	1.225(3)	1.246(2)	1.23(1)	1.232	1.255	1.254
N7−H8	0.993				1.010	1.018	1.017
			Angles				
C4−C5−N7	116.4	114.1(15)	116.3(6)	116(2)	115.3	117.1	116.6
O6═C5−N7	122.6	121.8(4)	121.7(6)	123(1)	123.1	122.1	122.6
C4−C5═O6	121.0	124.1	121.9(6)	121(4)	121.6	120.9	120.9
C5−N7−C9	121.7	119.7(8)	121.3(6)	122(1)	122.1	121.1	121.3
C5−N7−H8	119.8	110.0(50)			118.9	119.9	119.5
			Cis				
			Bonds				
C4−C5	1.484				1.514		
C5−N7	1.338				1.369		
N7−C9	1.446				1.450		
C5═O6	1.222				1.233		
N7−H8	0.995				1.101		
			Angles				
C4−C5−N7	118.7				115.8		
O6═C5−N7	120.4				121.4		
C4−C5═O6	121.0				122.8		
C5−N7−C9	125.6				126.9		
C5−N7−H8	115.4				113.9		

[a] Distance in Å and angles in deg; values in parentheses represent the standard deviation error in the final digit(s). [b] From ref 60, $3H_2O$ indicates two water molecules hydrogen bonding to the carbonyl oxygen and one water molecule hydrogen bonding to the amide proton; H_2O, 2FM indicates one water molecule and one formamide hydrogen bonding to the carbonyl oxygen and one formamide hydrogen bonding to the amide proton; see original reference for the exact geometries. [c] Gas-phase electron diffraction data from ref 65. [d] Crystal values are from ref 63 for the 0.9 occupancy structure. [e] Survey of the Cambridge Crystal Data Bank[22] performed as part of the present study that involved 145 structures from which 133 peptide bonds were selected with *R*-factors less than 0.08.

Table 1 presents the internal geometries of NMA from the empirical force field, experiment,[63,65] and ab initio calculations;[60,61,66] for the structural definitions, see Figure 1. The NMA ab initio calculations include fully optimized structures for the isolated molecule in the gas phase and for the molecule with hydrogen-bonded water and/or formamide molecules. These structures indicate the nature of the changes in geometry expected in going to the condensed phase. There is a decrease in the peptide bond length and an increase in the carbonyl C═O bond length in going from the gas phase to the condensed phase. Analysis of Table 1 shows that the ab initio calculations

TABLE 2: Geometric Data on the Alanine Dipeptide[a]

	$C7_{eq}$		$C7_{ax}$		C5		
	emp.	ab initio	emp.	ab initio	emp.	ab initio	survey
ϕ	−81.3	−85.8	69.7	76.0	−151.4	−157.2	
ψ	70.6	79.0	−67.6	−55.4	170.6	159.8	
ω1	−178.7	180.0	179.4	174.0	178.0	179.9	
ω2	178.8	−174.4	−178.5	−177.8	−179.8	179.5	
			Bonds				
C4−C5	1.480	1.511	1.480	1.513	1.480	1.512	1.515(7)
C5−N7	1.339	1.349	1.343	1.348	1.335	1.348	1.330(7)
C5−O6	1.224	1.207	1.225	1.207	1.223	1.204	1.225(5)
N7−C9	1.449	1.457	1.456	1.463	1.442	1.442	1.450(8)
C9−C11	1.543	1.521	1.547	1.531	1.544	1.535	
C9−C12	1.529	1.535	1.527	1.535	1.517	1.526	1.514(6)
C12−O13	1.229	1.203	1.228	1.204	1.230	1.204	1.235(9)
C12−N17	1.346	1.345	1.345	1.340	1.348	1.345	1.331(3)
N17−C19	1.443	1.446	1.443	1.446	1.444	1.448	1.446(14)
N7−H8	0.992	0.993	0.992	0.992	0.996	0.994	
N17−H18	1.002	0.996	1.003	0.996	0.995	0.992	
			Angles				
C4−C5−N7	116.6	116.3	115.9	115.7	116.4	115.9	116(2)
C5−N7−C9	123.3	122.9	125.9	127.1	122.8	122.0	121(1)
N7−C9−C12	112.6	109.8	114.9	114.3	108.2	107.4	112(2)
C9−C12−N17	116.8	114.6	117.9	117.4	117.7	115.6	116(1)
C12−N17−C19	122.4	121.2	122.7	120.9	121.5	121.7	122(1)

[a] Bonds lengths in Å and angles in deg. Ab initio data from ref 70. Survey results from the Cambridge Crystal Data Base[22] performed as part of the present study; values in parentheses represent the standard deviation error in the final digit(s). The sample includes compounds containing dipeptides with terminal aliphatic carbons.

reproduce the experimentally observed trends. Comparison of the CHARMM structures shows satisfactory agreement for the C5−N7 and N7−C9 bonds and for the C4−C5−N7, O6═C5−N7, C4−C5═O6, C5−N7−C9, and C5−N7−H8 angles. Upon going from the trans to the cis conformer, CHARMM reproduces predicted changes in the ab initio MP2/6-31G(d) structures for the O6═C5−N7, C5−N7−C9, and C5−N7−H8 angles. The most obvious discrepancy occurs for the C4−C5 bond length; i.e., the empirical bond length is 1.481 Å versus values between 1.51 and 1.52 Å for the ab initio and experimental data. The alanine dipeptide results (see below) show that this difference is resolved in the larger compound.

A number of ab initio calculations indicate that the structure of NMA in a vacuum deviates slightly from planarity.[67,68] There is pyramidalization of the peptide nitrogen, leading to deviations in planarity of up to 10° for the O═C−N−H dihedral angle. Recent calculations show that the peptide bond is essentially planar when involved in hydrogen-bonding interactions.[60] Since the present parameter set is designed for condensed-phase simulations, the minimum energy geometry of NMA was parametrized to be planar. To treat peptide bond rotation, including pyramidalization of the peptide nitrogen, a modified force field is required.[69]

Table 2 contains the geometric data for the alanine dipeptide in the $C7_{eq}$, $C7_{ax}$, and the C5 conformations as calculated with CHARMM and by ab initio methods at the HF/6-31G(p,d) level.[70] Results from a survey of the CCDB of dipeptide-containing molecules are also included. The atom names for the alanine dipeptide are shown in Figure 1b. Overall, the empirical bond lengths and angles are in satisfactory agreement with the ab initio and survey data. Of note are the C4−C5 and C9−C12 bond lengths; these correspond to the C4−C5 bond in NMA for which poor agreement was obtained. For the alanine dipeptide, the C4−C5 bond is still too short; however, the C9−C12 bond length is in good agreement with both the ab initio and survey data. The C4−C5 and C9−C12 bonds are treated with the same parameters; the difference between the two is due to the influence of nonbonded interactions on the optimized distances. This emphasizes the importance of the iterative approach used for intramolecular and intermolecular parameters in the present study.

The remaining bonds and angles in Table 2 are in good agreement with the target data. This includes the C5−N7 and C12−N17 peptide bonds and the C5═O6 and C12═O13 carbonyl bonds. For the angles the agreement of the empirical and ab initio values is generally good for both the absolute values and the trends among the three minima (see Table 2). In some cases, including the C5−N7−C9 and N7−C9−C12 angles, the changes between the minima in the CHARMM structures are not as large as those predicted by the ab initio calculations. However, the changes in the CHARMM values are in the correct direction. The overall geometries of the three minima, as indicated by the ϕ and ψ dihedral angles, are in reasonable agreement with the ab initio values. The largest differences occur in the ψ values of the $C7_{ax}$ and C5 minima, where differences of −13.5° and −10.4° occur, respectively. No effort was made to reconcile these differences because of the overall success of the parameters in reproducing experimental values of ϕ and ψ in a number of peptides and proteins (see Table 22 and Tables 4 and 7 of the Supporting Information). The differences between the CHARMM and ab initio results may be due to the limitations in the form of the empirical energy function in CHARMM. However, it is not clear that the ab initio values have converged to the correct results since the ϕ, ψ dihedrals are sensitive to the level of the ab initio calculation.

Parametrization of the force constants for the peptide backbone was based on the vibrational spectra and the relative energies of different conformers of NMA and the alanine dipeptide. Vibrational data for NMA are obtained from gas-phase and Ar matrix IR[71a,b,72] and RAMAN[73] studies and ab initio results.[74,75a] Solid and liquid-phase studies indicate that certain frequencies are shifted due to hydrogen bonding; most noticeable are the NH stretching[76,77] and in-plane and out-of-plane bending modes.[75b] In the optimization, parameters associated with the methyl groups were transferred directly from the CHARMM aliphatic parameter set. Table 3 lists the

TABLE 3: Vibrational Data for *N*-Methylacetamide[a]

	experimental/ab initio[b]		CHARMM	
mode	frequency	assignment	frequency	assignment
1	VLF		64	τCCH3(101)
2	VLF		89	τNCH3(1001)
3	171[c]	ωN7H[d] τC5−N7	200	τC5−N7(107)
4	279	βCNC βCCN	271	βCNC(62) βCCN(25)
5	391	τC5−N7 ωN7H[d]	431	βCCN(50)
6	431	βCCN βC5═O	579	βC5═O(50) νC5−C4(29)
7	628	βC5═O νC5−C4	652	ωC5═O(67) ωN7H(30)
8	718[c]	βC5═O rCH3	776	νC5−N7(34) νC5═O(20)
9	812	νC5−N7 rCH3 νC5−C4	797	ωN7H(66) rCH3(15)
10	973	rCH3 νN7−C9 νC5−C4	949	rCH3(36) νN7−C9(34)
11	1042	rCH3 βC5═O	996	rCH3(47) νN7−C9(26)
12	1092	νN7−C9 rCH3	1056	rCH3(83)
13	1176	rCH3	1087	rCH3 (72)
14	1263	νN7−C9 βC5═O βN7H	1093	rCH3(67) ωC5═O(17)
15	1279[c]	rCH3	1267	βN7H(44) νC5−C4(24)
16	1374	δCH3s	1384	δCH3s(94)
17	1410	δCH3s	1413	δCH3as(89)
18	1430	δCH3as	1416	δCH3as(88)
19	1430	δCH3as	1418	δCH3as(91)
20	1430	δCH3as	1426	δCH3as(87) rCH3(15)
21	1430	δCH3as	1481	δCH3s(50) βN7H(21)
22	1494	βN7H βN7−C9	1587	δCH3s(39) βN7H(20) νN7−C9(17)
23	1723	νC5═O	1683	νC5═O(66)
24	2830	νCH3s	2852	νCH3s(100)
25	2830	νCH3s	2914	νCH3as(100)
26	2940	νCH3as	2915	νCH3as(100)
27	2940	νCH3as	2917	νCH3s(100)
28	2940	νCH3as	2975	νCH3as(100)
29	2940	νCH3as	2975	νCH3as(100)
30	3495	νN7H	3326	νN7H(99)

[a] Frequencies in cm^{-1}. Potential energy distributions determined with the MOLVIB module in CHARMM. Only modes contributing greater than 12% are included. VLF indicates unobserved very low frequencies. ω indicates wagging modes, ν indicates stretching modes, τ indicates torsional rotations, r indicates rocking, δ indicates deformations, and β indicates bends. [b] Experimental data from refs 71a,b as reported and supplemented with ab initio data in ref 74. [c] Frequencies estimated from the cited ab initio calculations. [d] In ref 74 these modes are assigned as NH deformations. On the basis of more recent studies[75] we have assigned these as wagging modes.

CHARMM vibrational frequencies and the experimental gas-phase NMA frequencies reported by Sugawara et al.[74] Examination of Table 3 shows satisfactory agreement for modes 3−15, which are dominated by the internal parameters describing the peptide bond. Of note is the agreement for modes 7 and 9, which contain significant contributions from the N−H and C═O out-of-plane wagging modes. This agreement was obtained by use of improper dihedral angle force constants for the peptide bond (see Appendix). Modes 16−29 are associated with the methyl groups and are in good agreement with the experimental data, confirming the validity of the direct transfer of the aliphatic parameters to the present system.

Experimental and ab initio studies[75a,76,77] show modes associated with the N−H group to change upon going from the gas to condensed phase. Optimization of the force constants associated with the N−H stretching, bending, and wagging modes, therefore, emphasized the reproduction of condensed-phase vibrations rather than gas-phase data. This is based on the assumption that the N−H group always participates in hydrogen-bonding interactions and is consistent with the optimization of the bond and angle equilibrium parameters discussed above; i.e., an attempt is made to provide a parameter set that mirrors the condensed-phase environment. This approach leads to the N−H stretch mode being lower than the gas-phase experimental value, as shown in Table 3. Recent studies have shown aqueous hydrogen-bonded N−H bending modes to occur at 1313 and 1580 cm^{-1},[75b] values that are higher than those present in Table 3. The CHARMM values of 1267, 1481, and 1587 cm^{-1} are in satisfactory agreement with the condensed-phase values. Similarly, gas-phase N−H wags occur at 171 and 391 cm^{-1}, while the aqueous-phase frequency is calculated to occur at 745 cm^{-1}.[75b] The CHARMM values of 652 and 797 cm^{-1} are in good agreement with the latter value. This approach is used to obtain better dynamic properties of the protein backbone for condensed-phase simulations within the limitation that harmonic bond stretching, angle bending, and improper terms are used.

The parametrization of NMA also accounts for the relative cis/trans energies and the barrier to rotation about the ω dihedral angle (see Figure 1A). Analysis of NMR line shapes has been used to determine an enthalpic barrier to rotation of 19.8 ± 1.8 kcal/mol and a free energy barrier of 21.3 ± 0.3 kcal/mol.[78] Ab initio calculations at the MP2/6-31G(d)//HF/6-31G(d) level predict that the cis conformer is 2.07 kcal/mol above the trans conformer.[66] These data were used for the optimization of the dihedral angle parameters associated with the peptide bond. Values of 21.0 kcal/mol for the energetic barrier to rotation and 1.74 kcal/mol for the cis−trans energy difference were obtained with the present parameter set. This required inclusion of a 1-fold and a 2-fold term for the C4−C5−N7−C9 dihedral angle (see Appendix).

The internal parameter optimization for the alanine dipeptide was based on the transfer of the parameters from NMA. As the majority of internal parameters were determined in this way, only a few terms remained to be adjusted. These include the dihedral terms associated with the ϕ and ψ dihedral angles and the angle term associated with the central angle N7−C9−C12; this angle is often referred to as τ and was one of the few angular degrees of freedom that were adjusted in early crystal structure determinations. Adjustments of internal parameters associated with the peptide backbone have previously been based on experimental geometric data for the variation of the angle τ and with ϕ and ψ on relative energies of alanine dipeptide conformers from ab initio studies[10,79,80] or on the free energies obtained from the ϕ, ψ distributions observed in protein structures.[58,59,80] Use of energetics from survey data is appropriate when the goal of the force field is to obtain condensed-phase free energy information based on gas-phase calculations alone. However, for force fields to be used for simulations with explicit solvent models, parameters should be based primarily on potential energy rather than free energy data.[59] In the present work, optimization was initially based on ab initio results for the relative energies of certain conformations of the alanine

TABLE 4: Ab Initio Results on the Alanine Dipeptide and the Alanine Dipeptide Analogue[a]

level	$C7_{eq}$	$C7_{ax}$	C5	helical
		Alanine Dipeptide		
3-21G[d]	0.0(−85.8, 69.0)	2.81(74.4, −58.2)	1.13(−193.0, 190.6)	
4-21G[b]	0.0(−84.6, 73.0)	2.6(74.6, −62.0)	1.4(−165.7, 167.3)	
DZP[c]	0.0(−85.9, 79.1)	2.99(75.8, −58.9)	0.50(−156.0, 161.0)	3.22(65.9, 33.5)
6-31G[f]	0.0(−86.4, 72.7)	2.55(74.1, −58.6)	0.48(−159.8, 160.5)	
6-31G(d,p)[h]	0.0(−85.8, 79.0)	2.82(76.0, −55.4)	0.40(−157.2, 159.8)	4.35(−60.7, −40.7)
MP2/TZVP//HF/6-31G(d,p)[h]				
	0.0	2.05	1.47	3.91
LMP2/cc-pVTZ(−f)//MP2/6-31G(d)[g]				
	0.0(−83.1, 77.8)	2.48(74.4, −64.2)	1.11(−158.4, 161.3)	
MP2/cc-pVTZ(−f)//MP2/6-31G(d)[g]				
	0.0	2.41	1.61	
MP2/TZP//MP2/6-31G(d)[g]				
	0.0	2.13	1.86	
"MP4"/cc-pVTZ(−f)//MP2/6-31G(d)[g]				
	0.0	2.48	1.39	
"MP4−BSSE"/cc-pVTZ(−f)//MP2/6-31G(d)[g]				
	0.0	2.55	0.89	
		Alanine Dipeptide Analogue		
3-21G[d]	0.0(−84.5, 67.3)	2.53(74.1, −57.3)	1.26(−191.6, 189.4)	
4-21G[e]	0.0(−84.7, 67.3)		1.39(−166.6, 169.9)	
4-31G[e]	0.0(−85.5, 69.4)		0.45(−161.5, 164.5)	
6-31G[e]	0.0(−85.2, 69.8)		0.33(−160.9, 164.0)	
6-31G(d,p)[e]	0.0(−85.3, 76.0)		0.30(−157.9, 162.6)	
6-311G(d,p)[e]	0.0(−85.5, 78.3)		0.25(−156.8, 162.2)	
MP2/6-311G(d,p)[e]	0.0(−81.5, 82.5)		1.66(−159.8, 162.1)	
6-31+G(d)[i]	0.0(−85.8, 78.1)	2.56(75.1, −54.2)	0.19(−155.6, 160.0)	
HF/6-31+G(d,p)//HF/6-31+G(d)[i]				
	0.0	2.53	0.14	
MP2/6-31+G(d,p)//HF/6-31+G(d)[i]				
	0.0	2.19	1.13	
MP2/6-31+G(d,p)[f]				
	0.0(−83.0, 79.2)	2.20(74.3, −60.2)	1.27(−156.2, 160.5)	

[a] Energies in kcal/mol. Values in parentheses represent the ϕ and ψ angles. Alanine dipeptide analogue is the alanine dipeptide with the two terminal methyl groups omitted. [b] Reference 25. [c] Reference 81. [d] Reference 83. [e] Reference 82. [f] Guo, H.; Karplus, M. Unpublished results. [g] Reference 85. [h] Reference 70. [i] Reference 84.

dipeptide that correspond to different ϕ and ψ values and on the change in τ as a function of conformation.

Application of the above parameters for minimization and molecular dynamics simulations on crambin, BPTI, and MbCO gave reasonable results (see Section IV.d). However, the MbCO calculated structures showed significant deviations from experiment concerning the backbone ϕ, ψ angles. Since these deviations, which concerned mainly the α-helical region, were systematic, an iterative procedure based on the average differences between calculated and experimental MbCO backbone geometries and the conformational energetics of an alanine dipeptide analogue in which the terminal methyl groups are omitted was undertaken to obtain the final parameter set. The resulting parameter set had satisfactory behavior in the α-helical regions of MbCO, BPTI, and crambin and in the β-sheet regions of crambin and BPTI; there are 7 and 15 amino acids in β-sheets in crambin and BPTI, respectively. From the crystal simulations, the average differences of the ϕ, ψ values in the β-sheet regions with the final parameter set are −7.1, 6.0° and −4.5, 0.8° for crambin and BPTI, respectively, while the rms deviations are 11.7, 7.8° and 9.5, 6.6° for crambin and BPTI, respectively. The average deviations suggest that there remains a small, possibly systematic, deviation in the β-sheet region, although the sample is rather small. However, it should be noted that the rms differences for the β-sheet dihedral angles are significantly smaller than those occurring for all residues in the two proteins (see Table 22). The adjustment based on MbCO provides a way of correcting the ab initio dipeptide energy map for energetic effects due to the protein environment. Since the geometric parameters were determined for NMA in a solution environment, the ab initio map is being fitted to alanine dipeptide conformers whose internal geometries (particularly the CN and CO bond lengths) differ significantly from the ab initio values (see Table 2).

Table 4 gives the relative energies as a function of ϕ and ψ for the alanine dipeptide and an alanine dipeptide analogue from a variety of published ab initio studies.[25,70,81−85] In the analogue the terminal CH_3 groups are replaced by a hydrogen. The study of Head-Gordon et al.[84] included energies for 15 stationary points for the alanine dipeptide analogue at the HF/6-31+G(d) level, and the study by Gould et al.[70] contained energies for 7 alanine dipeptide structures at the HF/6-31G(p,d) and MP2/TZVP//HF/6-31G(p,d) levels. As the Gould et al. study was published following completion of the present work, the energies of Head-Gordon et al.[84] were used in the optimization of the ϕ and ψ dihedral parameters. This was performed by adjusting the dihedral parameters, optimizing the full alanine dipeptide with ϕ and ψ fixed at the values reported by Head-Gordon et al., and determining the sum of the squares of the difference between the ab initio and empirical relative energies. In the calculations the empirical alanine dipeptide energies were compared directly with the ab initio results for the alanine dipeptide analogue; i.e., no CHARMM calculations were made for the alanine dipeptide analogue because it contains an aldehyde functional group (see above) not included in the protein parametrization.

Tables 5 and 6 show the results of the optimization procedure used for the dipeptide parameters to give satisfactory results for the relative energies of the various conformers of the alanine dipeptide and simultaneously to remove the systematic deviation

TABLE 5: Relative Energy of the Empirical α_R Conformer and the Sum of the Squares Difference between ab Initio and Charmm22 Energies for the Alanine Dipeptide or the Alanine Dipeptide Analogue for Six-Parameter Sets[a]

parameters		sum of the squares	
	α_R energy[b]	HF/6-31+G(d) (dipeptide analog)[c]	MP2/TZVP// HF/6-31G(p,d)[d]
set 1, no cutoff	6.0	65.0	33.6
set 2, no cutoff	5.5	83.3	39.0
set 3, no cutoff	4.9	94.8	41.6
set 4, no cutoff	4.5	112.8	48.1
set 5, no cutoff	4.0	145.6	61.2
set 6, no cutoff	3.7	140.7	56.4

[a] Energies in kcal/mol and dihedrals in deg. Sum of the squares of the relative energy differences between the ab initio energies and the empirical energies not including the $C7_{eq}$ conformer, where the energy is 0.0 for all levels of theory. The empirical energies were obtained following full-geometry optimization with the ϕ and ψ dihedrals constrained at the ab initio values as reported in the cited studies. [b] α_R (right-handed helix) energy is the empirical energy relative to the $C7_{eq}$ conformer following minimization with ϕ and ψ constrained to -65 and $-41°$, respectively. [c] The 15 conformers used for the sum of the squares determination are those listed in Table 2 of ref 84. [d] The six conformers used for the sum of the squares determination are those listed in Table 2 of ref 70.

TABLE 6: Average Difference in the ϕ and ψ Values between the Myoglobin−CO Minimized and Crystal Structures[a]

	minimization		MD simulation	
parameters	ϕ	ψ	ϕ	ψ
set 1, no cutoff	-3.8 ± 1.4	3.4 ± 1.4	-10.0 ± 2.1	9.5 ± 2.5
set 1, 13−12−10	-3.9 ± 1.4	3.5 ± 1.5	-8.0 ± 2.0	7.1 ± 2.5
set 2, 13−12−10	-2.6 ± 1.5	2.4 ± 1.5	-4.6 ± 2.0	5.0 ± 2.4
set 3, 13−12−10	-1.5 ± 1.4	1.5 ± 1.5	-4.1 ± 1.8	4.6 ± 2.0
set 4, 13−12−10	-1.0 ± 1.4	1.0 ± 1.4	-1.0 ± 1.7	2.4 ± 2.2
set 5, 13−12−10	0.7 ± 1.7	-0.4 ± 1.4	-6.5 ± 1.7	6.0 ± 6.0
set 6, 13−12−10	-2.1 ± 1.4	1.9 ± 1.4	-11.8 ± 1.9	11.5 ± 2.4

[a] Dihedrals in deg. Minimizations involved 100 steepest descent steps followed by 500 ABNR steps, and MD simulations involved 20-ps vacuum simulations with the analysis performed using the 18−20-ps time-averaged structure.

in the ϕ, ψ values in MbCO. Only the dihedral angle parameters for ϕ and ψ were changed in the different sets; the values are given in the Appendix. Table 5 lists the sum of the square differences of the energies relative to $C7_{eq}$ of the CHARMM calculations and the alanine dipeptide analogue for the different parameter sets. Also included in Table 5 are the sum of the square differences between empirical and ab initio relative energies for seven conformers of the alanine dipeptide,[70] which were published after the parameters had been determined (see above). The comparison provides a posteriori verification of the use of the alanine dipeptide analogue ab initio results as the basis for the parameter optimization. The analogue and full sum of the squares values have a parallel behavior, indicating that the alanine dipeptide analogue results were appropriate as a basis for the optimization procedure. In fact, better agreement is obtained with the higher-level full dipeptide calculations than with those for the analogue.

The initial set of dihedral parameters associated with the peptide backbone are identified as set 1. They were used only to perform a molecular dynamics simulation for carbonmonoxy myoglobin.[47] They give the best agreement with the ab initio values, to which they were fitted (see Table 5). However, as indicated in Table 6, the structures from the vacuum calculations in both the presence and absence of an atom truncation scheme yielded significant deviations for the average values of ϕ and ψ for carbonmonoxy myoglobin. In particular, the results show that the ϕ angles were consistently too small and the ψ angles were consistently too large in comparison with the experimental values. Use of average deviations, rather than rms values, shows the direction of the deviations. Such systematic differences indicate that the parameters are introducing a systematic bias into the backbone conformation. This bias has been observed in other force fields[10] and is likely to be present in general. Such a decrease in ϕ and increase in ψ means that residues in a helical conformation, which dominates the MbCO structure, are being shifted to a more extended conformation (i.e., in the direction of the $C7_{eq}$ dipeptide minimum). Accordingly, the dihedral parameters associated with ϕ and ψ were adjusted to lower the energy of the α_R ($-61°$, $-41°$) conformer relative to $C7_{eq}$ (see Table 5). Care was taken during the adjustments of the dihedral angle parameters to ensure that the α_R conformer did not become a local minimum, since it is not a minimum in the ab initio calculations. As can be seen by comparing Tables 4 and 5, the α_R energy of set 1 is significantly too high, while the values for sets 4−6 are in a reasonable range. As the α_R-conformer energy is lowered by altering the parameters, poorer agreement between the relative energies for the empirical and ab initio values for other alanine dipeptide conformers is obtained (Table 5). The vacuum minimizations and molecular dynamics simulations for MbCO were repeated using the five-parameter sets (sets 2−6). The agreement between the calculated and experimental structures shows significant improvement. Set 4 yielded the best agreement of the molecular dynamics simulations and the carbonmonoxy myoglobin of ϕ, ψ values. It was selected as the final parameter set and used in subsequent tests.

The influence of the six-parameter sets used in determining ϕ and ψ on the adiabatic potential energy surface for the alanine dipeptide is shown in Figure 2; six alanine dipeptide maps corresponding to the six ϕ, ψ parameter values are presented. As the energy of the α_R conformer decreases the path between the $C7_{eq}$ and α_R conformers becomes more well-defined and a relatively narrow channel forms between the two. Analysis of the sum of the squares of the energy differences with respect to ab initio data in Table 5 shows these values increase, indicating a lower-quality surface with respect to the ab initio data. While the discrepancy between the relative energies of the empirical parameter sets and the ab initio data in Table 5 may be attributed in part to limitations in the level of theory in the ab initio calculations, the results reinforce the view that gas-phase data should be used with care in parametrization of force fields designed for use in the condensed phase. This is consistent with changes observed in the internal geometries of NMA and the alanine dipeptide upon going from the gas to condensed phases (Tables 1 and 2). The ability of force-field calculations to treat both the gas and condensed phases accurately may require alteration of the potential function by the addition of electronic polarization, as already mentioned.

During adjustment of the parameters, emphasis was placed on changing the relative energies of $C7_{eq}$ and α-helical structures; however, in no instance was the α-helical structure a true minimum. Analysis of the maps in Figure 2 shows the presence of a "channel" leading from the $C7_{eq}$ region to the α-helical region. Upon going from parameter set 1 to set 5, the channel becomes narrower and the $C7_{eq}$ to α-helical energy difference becomes smaller, in agreement with the ab initio values. Verification of the validity of the surface beyond that outlined above is difficult, especially considering that exact reproduction of gas-phase ab initio data may not yield the best

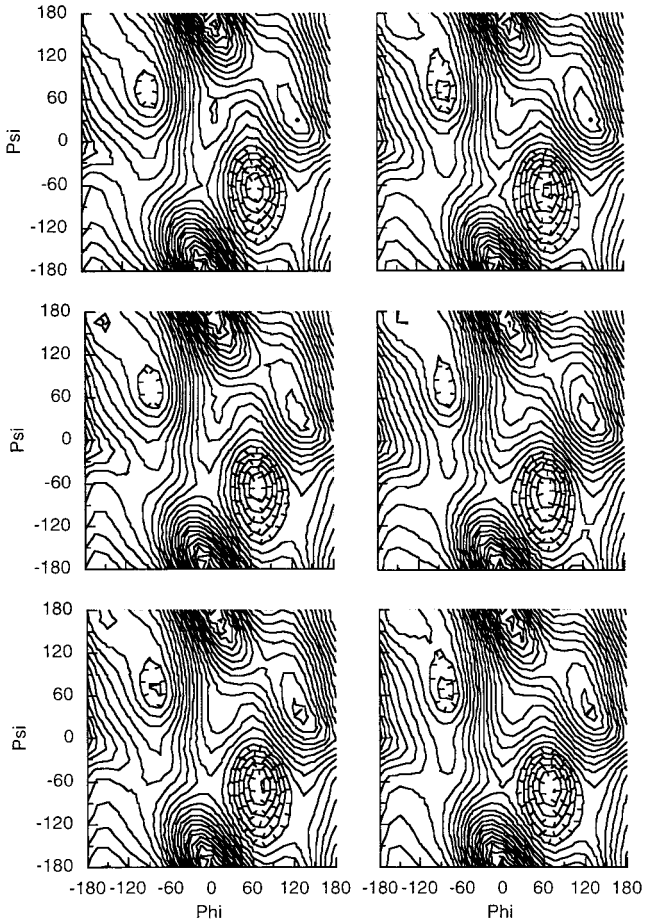

Figure 2. Adiabatic alanine dipeptide potential energy surfaces for parameters sets 1−6 (see Table 6). The left-hand column of surfaces going top to bottom corresponds to sets 1, 2, and 3, and the right-hand column of surfaces going top to bottom corresponds to sets 4, 5, and 6. Contours represent 1 kcal/mol.

condensed-phase properties with the present form of the potential energy function. Comparison with the HF/3-21G alanine dipeptide analogue map of Head-Gordon et al.[83] is limited owing to the numerous minima on that map; certain ones appear to be associated with the use of the truncated molecule since they do not occur in the full alanine dipeptide. To further validate the ϕ, ψ maps, the empirical and ab initio energies for the seven conformations studied by Gould et al.[70] are presented in Table 7. For the $C7_{ax}$ and C5 conformers the empirical data fall in the range of the ab initio values. Concerning the shape of the region in the upper left quadrant of the surface, the empirical α_R structure is slightly higher in energy while the β conformer is slightly lower than the ab initio values. The differences, however, are within 1 kcal/mol, suggesting this region of the map to be in reasonable agreement with the ab initio predicted results. Interestingly, decreasing the α_R empirical energy and increasing the β-conformer energy would yield a surface more similar to that of parameter set 5 in Figure 2. The $\beta 2$ conformer is 2 kcal/mol or more above the ab initio value, suggesting that the barrier in that region may be too large. This effect occurs to a greater extent for the empirical α_L conformer, which is overestimated by approximately 6 kcal/mol. This further suggests that the energies of the barrier regions may be overestimated by the present force field. Additional ab initio calculations on different conformers of the full alanine dipeptide will allow further verification of the maps.

Values of the angle τ are also included in Table 7 for the conformers that were studied. Overall, the empirical force field mimics the ab initio results reasonably well. To obtain this level of agreement for the change in τ with conformation, as well as for the adjustment of the relative energies of the dipeptide conformers, it was necessary to introduce altered van der Waals (1, 4) interactions for the peptide bond nitrogen and oxygen atoms (see Appendix); i.e., the $R_{\min_{1,4}}$ values on the nitrogen and oxygen atoms were set to 1.55 and 1.40 Å, respectively, allowing closer approach of those atoms. These terms were essential, in particular, for obtaining satisfactory values of τ for the C5 conformer. In that conformer, the nitrogen of the first peptide bond is within van der Waals contact of the oxygen of the second peptide bond. Use of the van der Waals parameters obtained from the optimization of the interaction parameters (see below) leads to an overestimation of the van der Waals repulsion and an opening of τ by 3.7° in the C5 conformer (not shown). Since the nitrogen and oxygen are in a 1, 4 configuration, the introduction of the van der Waals (1, 4) terms allowed for this problem to be overcome. In many other force fields, such 1,4 van der Waals scaling is used to varying degrees; e.g., in the CHARMM polar hydrogen parameter set (1, 4) scaling is used only for the carbon atoms while the AMBER force fields use a (1, 4) scaling factor of $^1/_2$ for all the van der Waals interactions.[7,79,86] In the present CHARMM all-atom parameter set, scaling is used only for the peptide oxygen and nitrogen interaction and for the aliphatic carbons; (1, 4) van der Waals scaling in the latter case is required for the proper treatment of cyclic structures, such as cyclohexane. The results obtained here for the alanine dipeptide τ values are comparable to those of Momany et al.,[25] who used similar (1, 4) scaling of the peptide N and C atoms.

Analysis of the vibrational spectra of the three alanine dipeptide minima was performed. Comparisons with recent ab initio calculations[87] and experimental solution studies based on vibrational Raman optical activity (VRAO)[87] provide another test of the force-field parameters. Table 8 shows the vibrational spectra obtained with the CHARMM parameters for the $C7_{eq}$, $C7_{ax}$, and C5 conformers, including the potential energy distributions. For the $C7_{eq}$ and C5 structures the HF/6-31G(d) vibrational frequencies below 1800 cm^{-1} are also presented.[87] Only the empirical assignments are included; the ab initio data were assigned on the basis of the published potential energy distributions.[87] Overall comparison of the empirical and ab initio data, excluding modes 22 and 21 for the $C7_{eq}$ and C5 structures, respectively (see below), shows average differences of 11 and 19 cm^{-1} and rms differences of 33 and 56 cm^{-1} for the $C7_{eq}$ and C5 structures, respectively. Both average differences are positive, indicating the empirical values are consistently larger than the ab initio data. This may be partially due to the use of a scale factor of 0.88 for the ab initio results;[87] other studies suggest a value closer to 0.9.[23] Use of the latter value yields average differences of −9 and −1 cm^{-1}. In general, the empirical and ab initio data are in satisfactory agreement for both the frequencies and assignments. The largest differences occur for modes 22 and 21 for the $C7_{eq}$ and C5 structures, respectively, which are both associated with wagging of the N−H protons. The empirical force field predicts these wags to have values significantly higher than the ab initio calculations; however, other N−H wags as well as C═O wags in the region of 600−800 cm^{-1} are in reasonable agreement, consistent with the N−H wag frequency calculated for NMA (see Table 3, mode 9). For the low frequencies, represented as modes 1−7, the agreement is generally good. Mode 2,

TABLE 7: Energies, Conformations, and Dipole Moments of Fixed Conformations of the Alanine Dipeptide[a,b]

	energy			τ		dipole	
(ϕ,ψ)	emp.	6-31G(p,d)	MP2/TZVP[c]	emp.	6-31G(p,d)	emp.	6-31G(p,d)
$C7_{eq}$(−85.8, 79.0)	0.00	0.00	0.00	111.6	109.8	2.50	2.87
$C7_{ax}$(76.0, −55.4)	2.44	2.82	2.05	114.5	114.3	3.56	3.91
C5(−157.2, 159.8)	1.01	0.40	1.47	108.0	107.4	2.83	2.56
α_R(−60.7, −40.7)	4.50	4.35	3.91	115.7	113.8	7.44	6.59
α_L(67.0, 30.2)	10.76	4.76	4.42	115.3	113.4	7.12	6.26
β(−57.6, 134.4)	3.91	4.90	4.08	114.7	109.3	4.15	2.36
β_2(−130.9, 22.3)	5.36	2.58	3.25	110.9	112.6	5.49	4.94

[a] Energies in kcal/mol, angles in deg, and dipole moments in D. [b] The conformations selected are those in Table 2 of ref 70. Empirical energies were determined following full optimization with the ϕ and ψ dihedral angles constrained to the values in Table 2 of ref 70; the values are given in parentheses in column one. [c] The MP2/TZVP energy for the HF/6-31G(p,d) optimized geometry.[70]

representing the methyl torsional rotations, has the largest disagreement, with the empirical values being approximately 40 cm^{-1} higher than the ab initio values. However, the empirical frequency of mode 4 in the C5 conformer, which contains significant contribution from the methyl torsion, is lower than the ab initio values, indicating the requirement for a compromise in the optimization with the present energy function. The frequencies for the ϕ and ψ torsions are well-reproduced by the empirical force field, including the decrease in the ϕ frequency upon going from the $C7_{eq}$ to the C5 conformer. Differences between the conformers occur also for the modes associated with τ(dN−CT−C), as expected due to the large changes in τ with conformation. These results suggest that certain modes could be used to characterize the different conformers in experimental studies.

Comparison with the experimental data is limited because only a few frequencies were assigned.[87] Furthermore, the assignments rely on the ab initio data included in that study and shown in Table 8. The C═O stretch is suggested to occur in the experimental regime at 1654 cm^{-1}. Empirical modes 47 and 48 for both the $C7_{eq}$ and C5 conformers range from 1677 to 1684 cm^{-1}, slightly higher, but still in good agreement with experiment. The experimental studies indicate that the C═O deformations occur in the range 300−370 cm^{-1}. Empirical values occur in this region and up to approximately 640 cm^{-1}. The C−N stretch of the peptide bond with contributions from deformations of the N−H and C_α−H groups occur in the region of 1298 cm^{-1} from experiment. A corresponding vibration occurs at 1184 cm^{-1} of the $C7_{eq}$ empirical spectrum. Experimental modes at 1370, 1445, and 1503 cm^{-1} are associated with in-plane bending of the N−H moieties. Corresponding frequencies occur at 1265, 1574, and 1598 cm^{-1} of the $C7_{eq}$ and 1218, 1273, 1572, and 1609 cm^{-1} of the C5 empirical spectra; i.e., the empirical values bracket the experimental data. The C−CT stretch is suggested to occur at 963 cm^{-1} in the experimental spectra. Empirical vibrations with significant C−CT contributions occur at 569, 662, 819, and 1265 cm^{-1} for the $C7_{eq}$ and 640, 766, and 1273 cm^{-1} for the C5 conformers, again bracketing the experimental value. Overall, the empirical force field produces vibrational spectra for the alanine dipeptide that are in satisfactory agreement with both ab initio and experimental data. Additional experimental assignments would allow more detailed comparisons.

A comparison of the $C7_{eq}$, $C7_{ax}$, and C5 minima from several empirical force fields is presented in Table 9. Comparison can be made also with a recent review of various aspects of the properties of the alanine dipeptide.[88] On the basis of the ab initio results presented in Table 4, the $C7_{ax}$ conformer is from 2.0 to 2.8 kcal/mol above the $C7_{eq}$ conformer, while the C5 conformer ranges from 0.4 to 1.5 kcal/mol above the $C7_{eq}$ conformer. The wide variation found in the ab initio results makes clear that the required level of ab initio theory has not yet been reached for these molecules. The present CHARMM parameters yield energies 2.05 and 0.92 kcal/mol for $C7_{ax}$ and C5, respectively, relative to $C7_{eq}$. While the C5 energy falls in the middle of the range of ab initio values, the $C7_{ax}$ lies at the lower end. Results on the alanine dipeptide and the alanine dipeptide analogue indicate that the inclusion of electron correlation leads to a lowering of the $C7_{ax}$ energy and an elevation of the C5 relative to $C7_{eq}$.[88] The CHARMM values, thus, are consistent with the ab initio data when electron correlation is taken into account. Of the empirical parameter sets listed in Table 9, the AMBER/OPLS and MSI CHARMm sets are in reasonable agreement with ab initio data; MM3 yields satisfactory agreement for the C5 conformer, and no value of the $C7_{ax}$ is available. In AMBER (all atom) the $C7_{ax}$ and C5 are similar, with the $C7_{ax}$ being underestimated and the C5 overestimated.[79] The opposite trend occurs with ECEPP/2, which was parametrized to reproduce protein distributions. Also included in Table 9 are the values of ϕ and ψ for the three minimum conformers. Comparison of the ab initio and empirical values show differences of 10° or more. The magnitude of these differences may be of relatively minor importance considering the rather flat character of the energy surfaces in the vicinity of the minimum-energy conformations (see Figure 2). Also, the addition of electron correlation via MP2 theory leads to a significant shift in the $C7_{ax}$ minimum conformation,[85] again suggesting that convergence has not been achieved in the ab initio calculations.

In considering the present parametrization, it is useful to refer also to several studies that have been published recently on the alanine dipeptide and other models of the protein backbone. A study by Dudek and Ponder[89] explored the role of electrostatics on the energetics of the alanine dipeptide in a number of molecular mechanics models. Ab initio relative energies were determined for a series of structures of the alanine dipeptide in which only the ϕ and ψ values were changed; i.e., the various conformers were obtained with rigid rotations and did not allow for adiabatic relaxation of the other degrees of freedom. The geometric changes that occur in the alanine dipeptide between the $C7_{ex}$ and C5 conformers, for example (Tables 2 and 7), indicate that such ab initio rigid-rotation results are only of limited value. It was shown more than 20 years ago in a study of acetylcholine with empirical energy functions[90] that it was essential to include conformational flexibility to obtain meaningful relative energies for different conformers. Also, comparison of other ab initio energy calculations based on rigid structures[58] with a variety of ab initio calculations that included full relaxation show that the differences in the energies of the alanine dipeptide conformers are significantly overestimated when rigid geometries are used.[91]

TABLE 8: Alanine Dipeptide Charmm22 and ab Initio Vibrational Spectra[a]

	$C7_{eq}$			$C7_{ax}$		C5		
mode	emp.	a.i.	assign.	emp.	assign.	emp.	a.i.	assign.
1	51	42	ψ(95)	58	ψ(83)	33	36	ψ(64) ϕ(33)
2	62	22	tCH3(94)	66	tCH3(90)	57	22	tCH3(69) ϕ(17)
3	84	70	tCH3(62) tC−N(29)	82	tCH3(71) tC−N(22)	70	68	ϕ(44) tCH3(35) ψ(17)
4	89	81	tC−N(42) tCH3(32) ϕ(17)	93	tC−N(62) tCH3(18) dN−CT−C(17)	91	53	tCH3(64) tC−N(21)
5	110	90	ϕ(82)	132	ϕ(101)	96	88	dN−CT−C(26) tC−N(25) tCH3(24)
6	179	172	tC−N(42) dC−N−CT(24)	174	tC−N(74)	151	133	tC−N(40) dC−N−CT(16) dN−CT−C(16)
7	190	141	tC−N(57)	198	dCT−C−N(31) dC−N−CT(23) dN−CT−C(18)	166	124	tC−N(83)
8	230	205	dC−N−CT(40) dCT−C−N(33)	242	dC−N−CT(58) dCT−C−N(21)	228	217	dC−N−CT(41)
9	281	231	tCH3(76)	270	tCH3(81)	253	228	tCH3(36) dCT−C−N(25) dC−N−CT(17)
10	283	306	dN−CT−CT(25) dC−CT−CT(22)	285	dC−CT−CT(31) tCH3(16)	265	256	tCH3(52)
11	308	269	dN−CT−C(33) tC−N(19) sC−CT(18)	320	dC−N−CT(45)	304	289	dC−C_α−C_β(31) dC−N−CT(21)
12	332	320	dC−N−CT(32) dC═O(31) dC−CT−CT(19)	357	dN−CT−CT(53) dC═O(19)	349	343	dCT−C−N(36) dC−N−CT(18) dC═O(16)
13	431	404	dCT−C−N(25)	402	dCT−C−N(31)	397	369	dN−C_α−C_β(43)
14	466	472	dCT−C−N(50) dN−CT−CT(21)	519	dC═O(53)	520	399	dCT−C−N(32) dC═O(29)
15	569	549	dC═O(54) sC−CT(23)	577	dCT−C−N(19) sC−CT(16)	572	490	dC═O(44) sC−CT(21)
16	636	578	dC═O(33) sC−CT(18)	645	sC−CT(28) dC═O(20)	640	584	dC═O(23) sC−CT(22) dC−N−CT(17)
17	662	612	wC═O(70) wN−H(24)	656	wC═O(73) wN−H(23)	669	620	wC═O(74) wN−H(18)
18	738	658	wC═O(41) wN−H(36)	742	sC−N(21)	705	484	wN−H(54) wC═O(29)
19	775	766	sC═O(18) sC−N(15) wN−H(15)	755	wN−H(41) wC═O(33)	766	825	sC−N(24) sC═O(17) sC−CT(17)
20	819	873	sC−N(32) sC═O(21) sC−CT(18)	805	sC−N(28) sC═O(18)	823	869	sC−N(24) sC═O(17)
21	837	831	wN−H(64) dCH3(20)	833	wN−H(59) dCH3(15)	833	484	wN−H(63)
22	887	422	wN−H(41) wC═O(22)	893	wN−H(50) wC═O(26)	884	735	wN−H(28) wC═O(28) dCH3(16)
23	911	920	dCH3(28) sN−CT(26) sCT−CT(20)	909	dCH3(34) sN−CT(18) sCT−CT(15)	908	943	dCH3(33) sN−CT(21)
24	946	977	dCH3(48) sN−CT(37)	951	dCH3(46) sN−CT(38)	954	996	sN−CT(48) dCH3(32)
25	979	1004	dCH3(84)	976	dCH3(78)	975	978	dCH3(78)
26	1022	1024	dCH3(41)	1005	dCH3(32) dC_α−H(21) sN−CT(17)	1003	1045	dCH3(41) dC_α−H(21)
27	1038	1032	dCH3(60) sC_α−C_β(18)	1044	dCH3(73)	1034	1065	dCH3(71)
28	1073	1087	dCH3(88)	1072	dCH3(84)	1072	1111	dCH3(89)
29	1086	1109	dCH3(77)	1082	dCH3(53)	1083	1031	dCH3(72) wC═O(19)
30	1087	1135	dCH3(77)	1086	dCH3(76)	1087	1137	dCH3(83)
31	1130	1153	sN−CT(34) dCH3(24)	1090	dCH3(58)	1121	1154	dCH3(29) sN−CT(27)

TABLE 8: (Continued)

	$C7_{eq}$			$C7_{ax}$		C5		
mode	emp.	a.i.	assign.	emp.	assign.	emp.	a.i.	assign.
			$dC_\alpha-H(16)$					
32	1184	1207	$dC_\alpha-H(33)$ dN−H(23) sC−N(19)	1244	dN−H(26) dCT−HA(22) sC−CT(15)	1218	1196	dN−H(33)
33	1265	1236	dN−H(44) sC−CT(21)	1270	dN−H(43) sC−CT(23)	1273	1232	dN−H(35) sC−CT(22)
34	1350	1307	$dC_\alpha-H(37)$ dCH3(15)	1325	$dC_\alpha-H(41)$ $sC_\alpha-C_\beta(18)$	1340	1309	$dC_\alpha-H(48)$ sCT−CT(15)
35	1386	1352	dCH3(94)	1386	dCH3(95)	1384	1337	dCH3(95)
36	1406	1370	dCH3(88)	1407	dCH3(81)	1406	1369	dCH3(85)
37	1413	1382	dCH3(92)	1413	dCH3(98)	1411	1373	dCH3(86)
38	1416	1413	dCH3(99)	1416	dCH3(99)	1415	1415	dCH3(98)
39	1418	1421	dCH3(100)	1418	dCH3(100)	1418	1422	dCH3(100)
40	1425	1434	dCH3(98)	1425	dCH3(98)	1423	1434	dCH369) $dC_\alpha-H(18)$
41	1428	1438	dCH3(68) $dC_\alpha-H(18)$	1429	dCH3(62) $dC_\alpha-H(22)$	1426	1436	dCH3(100)
42	1437	1439	dCH3(80)	1438	dCH3(95)	1433	1439	dCH3(57) $dC_\alpha-H(21)$
43	1441	1441	dCH3(66)	1442	dCH3(60) $dC_\alpha-H(20)$	1442	1448	dCH3(98)
44	1492	1458	dCH3(50)	1482	dCH3(53) dN−H(23)	1480	1458	dCH3(42) dN−H(22)
45	1574	1501	dN−H(26) sN−CT(15)	1552	dN−H(32) sC−N(22) sN−CT(17)	1572	1496	dCH3(37) dN−H(20) sN−CT(17)
46	1598	1533	dN−H(22) dCH3(22) sC−N(18) sN−CT(16)	1590	dCH3(34) dN−H(25) sN−CT(17) sC−N(15)	1609	1523	dN−H(24) $dC_\alpha-H(15)$
47	1680	1694	sC═O(64)	1685	sC═O(66)	1677	1701	sC═O(63)
48	1684	1720	sC═O(65)	1692	sC═O(59)	1684	1715	sC═O(66)
49	2852		sCH3(100)	2852	sCH3(100)	2852		sCH3(100)
50	2902		sCH3(92)	2902	sCH3(86)	2902		sCH3(92)
51	2905		$sC_\alpha-H(91)$	2903	$sC_\alpha-H(86)$	2905		$sC_\alpha-H(92)$
52	2914		sCH3(100)	2914	sCH3(100)	2914		sCH3(100)
53	2915		sCH3(100)	2915	sCH3(100)	2914		sCH3(100)
54	2917		sCH3(100)	2917	sCH3(100)	2917		sCH3(100)
55	2959		sCH3(100)	2958	sCH3(100)	2958		sCH3(100)
56	2960		sCH3(100)	2960	sCH3(100)	2961		sCH3(100)
57	2975		sCH3(100)	2975	sCH3(100)	2975		sCH3(100)
58	2975		sCH3(100)	2975	sCH3(100)	2976		sCH3(100)
59	3319		sN−H(99)	3319	sN−H(99)	3325		sN−H(99)
60	3328		sN−H(100)	3325	sN−H(100)	3328		sN−H(99)

[a] Frequencies in cm^{-1}. Ab initio data from ref 87. Potential energy distributions determined with the MOLVIB module in CHARMM.

In a recent study by Beachy et al.,[85] extensive ab initio calculations were performed on the alanine dipeptide and the alanine tetrapeptide. These calculations employed geometries fully optimized at the HF/6-31G(p,d) level for the alanine dipeptide and tetrapeptide, respectively, and significantly extended the level of theory used to calculate the relative energies of the analyzed conformers, including electron correlation, based on single-point calculations of the HF/6-31G(p,d) optimized structures. The alanine dipeptide results were generally consistent with data from lower levels of theory (see Table 4), and the new alanine tetrapeptide results were used to test a number of available force fields, including the present CHARMM potential. A developmental version of the all-atom OPLS force field,[92] MMFF,[93] and MM3[10] were shown to best reproduce the ab initio energies of the different conformers obtained from unrestrained minimizations, while the present parameters performed somewhat worse (e.g., 3.78 vs 1.21 kcal/mol for the rms difference for CHARMM22 and MMFF, respectively, for optimized geometries with the ϕ and ψ dihedral angles restrained to the HF/6-31G(d,p) values). In terms of the actual energy values, the difference appears to be significant because the uncertainty in the ab initio values of the relative energies is probably not much worse than 1 kcal/mol. However, the comparison is of limited value because the objective of the present force field is to produce accurate results in solution simulations of peptides and proteins, rather than to reproduce ab initio calculations of isolated peptides. This is of particular relevance here because some of the conformers of the alanine tetrapeptide that were studied by Beachy et al.[85] correspond to ones only rarely observed for alanine-containing regions in proteins, and presumably in peptides, in solution. In particular, several of the local minima have ϕ, ψ values in the α_L region ($\phi \cong 60$, $\psi \cong 60$), which are rarely found for alanine. While the α_L region is significantly populated in proteins, more than 50% of the residues observed in this region are glycines (R. L. Dunbrack, Jr., personal communication).

The conformations of the "tetrapeptide", which consist of three alanines plus terminal blocking groups (in analogy to the "dipeptide"), can be described in terms of the conformations of alanine units; i.e., the total conformational energy of the tetrapeptide, relative to the minimum energy structure, can be approximated by the sum of the relative energies of the alanine units. The three tetrapeptide conformers for which there are large deviations in the empirical CHARMM energies from the

TABLE 9: Relative Energies and Conformations of the Alanine Dipeptide $C7_{eq}$, $C7_{ax}$, and C5 Minima from Empirical and ab Initio Calculations[a]

source	$C7_{eq}$	$C7_{ax}$	C5
	Energies		
MP2/TZVP[b]	0.00	2.05	1.47
CHARMM	0.00	2.05	0.92
MM3[c]	0.00		1.0
AMBER (all-atom)[d]	0.00	1.5	1.5
AMBER/OPLS[e]	0.00	2.5	1.5
ECEPP/2[e]	0.00	7.3	0.7
MSI CHARMm[f]	0.00	2.7	1.3
	Conformations (ϕ, ψ)		
HF/6-31G**	−86, 78	75, −54	−156, 160
CHARMM	−81, 71	70, −68	−151, 171
MM3	−83, 66		−164, 154
AMBER(all-atom)			
AMBER/OPLS	−84, 70	67, −56	−150, 162
ECEPP/2	−80, 76	76, −65	−155, 157
MSI CHARMm	−79, 72	70, −68	−161, 171

[a] Energies in kcal/mol and dihedrals in deg. Fully optimized empirical geometries. [b] Reference 70, MP2/TZVP energy at the HF/6-31G(p,d) optimized geometry. [c] Reference 10. [d] Reference 79. [e] Reference 8. [f] Reference 25.

ab initio results have at least one of three ϕ, ψ values in the α_L region; one has two ϕ, ψ in the α_L region (relative energy difference 9.92 kcal), and the other two (relative energy differences 6.92 and 4.82 kcal) each have one set of ϕ, ψ values in the α_L region. The α_L energy of the empirical energy function is too high by 6.0−6.3 kcal in the alanine dipeptide (see Table 7). To examine the importance of the α_L configurations in the tetrapeptide results, we calculated the rms difference between the CHARMM and LMP2/cc-pVTZ(-f) relative energies (with respect to structure 3 of Beachy et al.). For the restrained geometries without applying any scaling procedure (see Table 5 of Beachy et al.) only relative energies have a meaning in the empirical energy function since the zero of energy is arbitrary. Our calculated rms difference is 4.39 kcal/mol, somewhat larger than that reported by Beachy et al. using their comparison method. The rms difference without the three conformers that contained one or more ϕ, ψ pairs in the α_L region of the dipeptide was found to be 0.77 kcal/mol. This result is consistent with results from the Beachy et al. study, where omission of conformations in the α_L region lead to a decrease in the reported CHARMM rms difference from 3.78 to 0.95 kcal/mol. The high energy of the CHARMM α_L conformer is consistent with a generally too-high relative energy of the upper right quadrant of the alanine dipeptide map (see Table 7 and Figure 2). Because of the rarity of the α_L conformer in proteins (other than for glycines), the error in the energy was ignored in the parametrization. It should be noted that the α_R conformer of the glycine dipeptide, which is equivalent to the α_L alanine dipeptide conformer, is also overestimated (see Table 10). Simulations of cyclic peptides, however, show the ϕ, ψ angles of glycines in the α_R conformation to be well-maintained by the present force field (see Table 7 of Supporting Information). These results and those presented above on the relationship between energetics of the alanine dipeptide and calculated ϕ, ψ values of MbCO point to the importance of not parametrizing protein and peptide force fields simply on the basis of limited gas-phase ab initio data, as already pointed out in the paper by Beachy et al. Another consideration in evaluating the comparison is that the geometry of the peptide bond in the CHARMM energy function was parametrized for solution and crystal structures. This leads to CO bond lengths in the alanine dipeptide in the range 1.223−1.230 Å (Table 2). These values

TABLE 10: Relative Energies and Conformations for the Glycine Dipeptide from ab Initio and the Empirical Calculations[a]

conformer	ϕ	ψ	empirical[b]	HF/6-31G(p,d)[c]	MP2/TZVP[d]
C7	−85.8	72.0	0.00	0.00	0.00
C5	180.9	180.5	0.89	−0.27	1.99
α_R	−60.7	−40.7	7.43	4.03	3.95
β_2	−116.2	19.9	4.24	1.90	3.25
			Fully Optimized Empirical Results[e]		
C7	−83.0	67.9	0.00		
C5	180.0	180.0	0.94		
α_R					
β_2					

[a] Energies in kcal/mol and dihedral angles in deg. [b] Empirical energies determined with the ϕ, ψ values constrainted to the HF/6-31G(p,d) (listed) values and the remainder of the molecule fully optimized. [c] Fully optmized values from ref 70. [d] MP2/TZVP (Dunning's triple-ζ basis set plus polarization functions) energy for the HF/6-31G(p,d) optimized geometry from ref 70. [e] Full optimizations performed following partial optimizations with the ϕ, ψ values constrained to the HF/6-31G(p,d) (listed) values. Both the α_R and β_2 conformers converted to the C7 during the full optimizations.

TABLE 11: Minimum Interaction Energies and Geometries of NMA with Water and the NMA Dimer[a]

	HF/6-31G(d)[c]			empirical		
interaction[b]	E_{min}	R_{min}	angle	E_{min}	R_{min}	angle
(1) C═O···HOH	−7.67	1.98	146	−7.69(0.90)	1.76	146
(2) N−H···OHH	−6.29	2.13	174	−6.30(0.47)	1.93	171
(3) parallel dimer	−7.75	2.08		−7.88(−0.59)	1.84	

[a] Energies in kcal/mol, distances in Å, and angles in deg. Values in parentheses with the empirical E_{min} data are the Lennard-Jones contributions to the interaction energies. [b] See Figure 3 for the interaction geometries. [c] HF/6-31G(d) energies have been scaled by 1.16 (see text).

are significantly longer than the values found in the gas-phase ab initio calculations; e.g., values of 1.203−1.207 Å are obtained in ref 70.

IV.a.2. Interaction Parameters. Partial atomic charges and Lennard-Jones parameters for the protein backbone were optimized using NMA as the model compound. Data that were used included the interaction energies and geometries of the complexes of NMA with water and the NMA dimer from ab initio calculations, the dipole moment of NMA, the heat of vaporization and molecular volume of pure NMA, and the heat of solvation of NMA. Additional testing of the parameters was performed via crystal simulations of NMA and the alanine dipeptide. Initial interaction parameters were obtained from the CHARMM19 parameter set[5,15] for the polar atoms. The charges of the methyl groups treated were determined using the standard charge of 0.09 for the hydrogens[21] and methyl carbon charges selected to yield a neutral total charge. The aliphatic hydrogen charge was previously determined on the basis of the electrostatic contribution to the trans−gauche energy difference of *n*-butane.[21] Lennard-Jones parameters were obtained from the CHARMM22 all-hydrogen nucleic acid parameters for the amide group[13] and from the CHARMM22 all-hydrogen alkane parameters for the methyl groups. Optimization of the van der Waals parameters was limited, therefore, to adjustment of the peptide bond carbonyl carbon radius and well depth. The adjustment of the partial atomic charges and van der Waals parameters was performed in an iterative fashion, as outlined in Section II.a.

Table 11 lists the interaction energies and geometries for the NMA−water and NMA dimer complexes at the minimum-energy geometry from the empirical and ab initio calculations

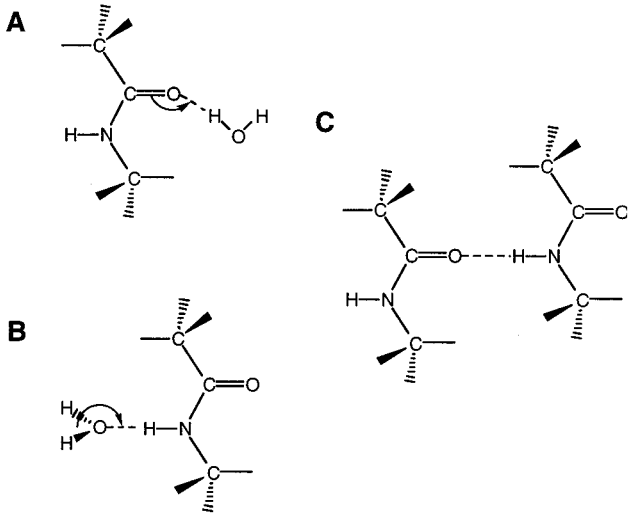

Figure 3. Interaction orientations of *N*-methylacetamide with water (A and B) and the *N*-methylacetamide parallel dimer (C).

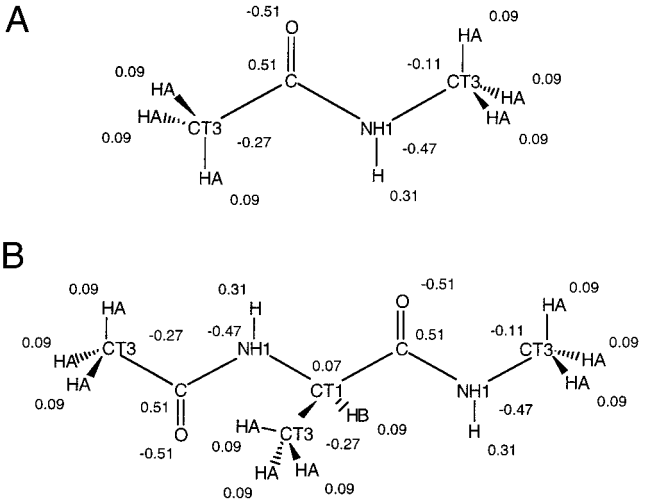

Figure 4. CHARMM partial atomic charges and atom types for (A) *N*-methylacetamide and (B) the alanine dipeptide.

for the orientations shown in Figure 3. As discussed in Section II.a, the ab initio interaction energies have been scaled by 1.16. Comparison of the ab initio and empirical energies shows excellent agreement. The geometries are also in good agreement, with the empirical distances approximately 0.2 Å less than the ab initio values (see Section II.a), while the angles for the interactions with water are in good agreement. The largest disagreement occurs for the NMA dimer, for which the empirical interaction energy is slightly too favorable. Included in Table 11 are the contributions of the Lennard-Jones term to the interaction energies. As may be seen, the magnitude of these contributions is significant, emphasizing the need to balance the Lennard-Jones and electrostatic parameters.[30] It is of interest that a developmental version of the CHARMM force field reproduces the cooperativity of binding of multiple water molecules to NMA.[94]

The dipole moment was monitored while adjusting the charges to reproduce the interaction energies. The final charge distribution is shown in Figure 4; it yields a dipole moment of 4.12, somewhat larger than the experimental dipole of 3.7.[95] Such an overestimation is due to the requirement that protein polarization effects be included implicitly in the force field to be consistent with the charges of TIP3P water (see Section II.a).

TABLE 12: Condensed-Phase Calculated and Experimental Data for *N*-Methylacetamide[a]

Pure Solvent			
calculated		experimental	
ΔH_{vap}	mol vol.	ΔH_{vap}	mol vol.
13.85 ± 0.02	133.7 ± 0.2	14.2	135.9
Aqueous Solvent[b,c]			
ΔH_{solv}	mol vol.	ΔH_{solv}	mol vol.
−18.8(−19.4)	75(65)	−19.2	122.7

[a] Energies in kcal/mol and molecular volumes in Å^3. [b] Both the heat of solvation and molecular volume are determined from the difference between two large fluctuating numbers; on the basis of the statisitical error in the individual values the errors are estimated to be ±3 kcal/mol and ±20 Å^3, respectively. [c] The calculated values are based on 6 M configurations; the values in parentheses are from 4.5 M configurations.

Additional analysis of the validity of the charges was performed by calculating the dipole moment of the alanine dipeptide as a function of conformation. Shown in Table 7 are the calculated dipoles from the empirical force field along with those from ab initio calculations at the HF/6-31G(p,d) level.[70] Overall, the empirical atomic charges reproduce the trends seen in the ab initio calculations. With the exception of the $C7_{eq}$ and $C7_{ax}$ conformers, the empirical values are larger than the HF/6-31G(p,d) values, as expected owing to the implicit inclusion of polarization in the force field. The largest dipole occurs in the α_R conformer, consistent with the ab initio result. The magnitude of the dipole moment of this conformer is important for the proper treatment of the α_R helix.

Table 12 presents results for liquid NMA from Monte Carlo calculations and experiment.[96] The calculated heat of vaporization is somewhat underestimated. However, other experimental data have indicated a value of 13.3 kcal/mol,[97] suggesting the present value is reasonable. Comparison of the molecular volumes shows that the calculated value (133.7 Å^3) is within 2% of experiment. As an additional test of the NMA parameters in the condensed phase, the heat of solvation and molecular volume in infinitely dilute aqueous solution were calculated. The results are included in Table 12. Monte Carlo calculations of NMA in a box of 262 TIP3P water molecules yielded a heat of solution of −18.8 kcal/mol and a molecular volume of 75 Å^3, as compared to experimental values of −19.2 kcal/mol and 122.7 Å^3, respectively. Comparison with previously reported values of −25.5 kcal/mol and −9 Å^3 for the heat of solution and molecular volume of NMA based on the OPLS force field and a Monte Carlo calculation shows that the present parameters represent a significant improvement;[98] in the Monte Carlo calculations the internal geometry of NMA was constrained to the optimized gas-phase geometry. The agreement of the solution results, combined with the NMA pure-liquid properties, confirms that the interaction parameters are a good representation of the nonbonded interactions of NMA in different environments; i.e., there is the appropriate balance between the solute−solute and solute−solvent interactions of the protein backbone in the present parameter set.

Recently, Kaminski and Jorgensen[99] have stressed the importance of the correct representation of liquid properties by molecular mechanics force fields. They made comparisons of a new OPLS all-atom force field,[100] AMBER 94,[69] and MMFF.[93] The best agreement for liquid NMA was obtained for the OPLS force field; the calculated values ΔH_{vap} = 13.61 kcal/mol and a molecular volume of 133.8 Å^3 are almost identical to those reported here.

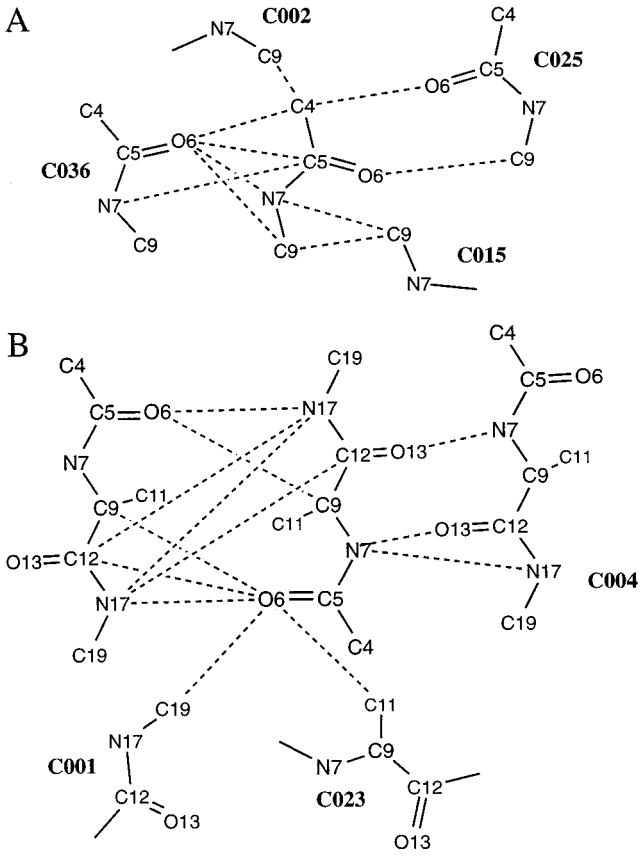

Figure 5. Diagrams of the interactions between the primary and images atoms for the (A) *N*-methylacetamide and (B) alanine dipeptide crystals. Bold characters identify images based on the CHARMM image nomenclature.

IV.a.3. N-Methylacetamide and Alanine Dipeptide Crystal Calculations. Crystal calculations were performed on NMA and the alanine dipeptide as an additional test of the validity of the backbone parameters in the condensed phase. The two crystals are shown in Figure 5. NMA crystallizes in an orthorhobmic *Pnma* space group at 238 K with four molecules per unit cell.[62] The asymmetric unit corresponds to "half" of the NMA molecule based on a mirror plane through the heavy atoms of the molecule. One symmetry operation generates the methyl hydrogen across the mirror plane. In the present calculations, we use this operation to generate the full NMA molecule and then use the CRYSTAL facility to create the remaining four molecules in the unit cell as well as other unit cells, such that the primary atoms represent a single NMA molecule. L-Alanine dipeptide crystallizes in an orthorhombic $P2_12_12_1$ space group at room temperature.[101] Constant volume, constant temperature (NVT), and constant pressure, constant temperature (NPT) molecular dynamics simulations were performed for both crystals.

Energy minimizations of the crystals were performed as a function of the nonbonded interaction truncation distances to determine an appropriate value. Table 1 of the Supporting Information presents the unit cell parameters and energies from the minimizations of both NMA and the alanine dipeptide. Other than the shortest distances (10−9−7), the results are reasonably behaved. The truncation scheme 22−21−19 was chosen since the crystal parameters appear to be well-converged. For both crystals the total volume decreased upon minimization, as expected because the experiments are performed at finite temperatures while the minimization corresponds to 0 K. In NMA all three lattice parameters contract by between 2.1 and 4.1%, indicating that the crystal is well-represented by the empirical model. In the alanine dipeptide, the *B*- and *C*-lattice parameters contract while the *A*-term increases; the reason for this is not evident. This trend is also found in the NPT simulations (see below).

Simulations on NMA and the alanine dipeptide were performed in both the NVT and NPT ensembles with the 22−21−19 truncation scheme. Table 13 shows the pressures and unit cell parameters obtained from the simulations. In the NVT ensemble simulations NMA yielded a negative external pressure of approximately −2000 atm while the alanine dipeptide yielded a positive pressure of approximately 3000 atm (see Section IV.b). Upon going to the NPT ensemble, the average pressures approach unity. The deviation from unity is due to the large pressure fluctuations; in fact, values of less than 1000 atm or so have only a small effect on the structure and energy. As expected, there is an overall contraction of the NMA crystal (corresponding to the negative pressure in the NVT simulation) and an expansion of the alanine dipeptide crystal. Comparison of the unit cell parameters from the minimizations and the NPT simulations shows an increase in the simulations, again as expected because of the kinetic energy corresponding to 300 K in the system. In NMA this yields unit cell parameters that are in good agreement with the experimental values; the *B*-axis is 0.15 Å shorter than in the crystal structure. In the alanine dipeptide crystal the *A*-axis is significantly larger than the crystal value. The expansion of the *A*-axis appears to be associated with interactions of the terminal methyl groups of the alanine dipeptide (see below), but the values for *B*- and *C*-axes are in good agreement with experiment.

Additional analysis of the NMA and alanine dipeptide crystal calculations was based on examination of the structural details. The comparison with the NMA crystal is complicated by the fact that it is disordered with a major and minor occupancy of 0.9 and 0.1, respectively.[63] We use the structural data as given, which presumably refers to the major conformer. Table 14 shows the rms deviation between the calculated and the crystal structure of the primary cell non-hydrogen atoms; corresponding results for the dihedral angles are given in Table 15 and for the nonbonded interaction distances in Table 3 of the Supporting Information. The rms deviations indicate that there are only minor changes in the internal structures. This is supported by a comparison of the calculated and experimental dihedral angles. The excellent agreement for NMA is expected as the 2-fold dihedral term for rotation about the peptide bond combined with the high energy barrier (see Appendix) leads to only small fluctuations in the vicinity of the minimum. In the alanine dipeptide, rotation about ϕ and ψ is relatively unrestrained, but there is good agreement between the calculated and experimental values. The quality of the agreement is encouraging considering that the ϕ, ψ values in the crystal are not minima on the alanine dipeptide map but shifted approximately 2 kcal/mol above the C5 minima. Although there is significant deviation of the two peptide bonds from planarity in the experimental crystal structure, the calculated values are close to planarity. This may suggest that the empirical potential function is somewhat too steep near planarity, although the rms fluctuations of these dihedrals are approximately 9°. A recent survey of the CCDB indicates that deviations from planarity of the peptide bond do occur; the standard deviation from planarity is 6°.[102] For the simulation results, the differences between the calculated and experimental structures are less than the rms fluctuations of the dihedral angles in all cases.

TABLE 13: Results from the Crystal Simulations of *N*-Methylacetamide and the Alanine Dipeptide Using the 22−21−19 Truncation Scheme[a]

	pressure		unit cell parameters			
system	P_{ext}	P_{int}	*A*	*B*	*C*	vol
			NMA			
expt	1	1	9.61	6.52	7.24	454
NVT	−2265 ± 3127	−1367 ± 9596				
NPT	−310 ± 3162	551 ± 11915	9.54 ± 0.06	6.37 ± 0.07	7.17 ± 0.05	436
			Alanine Dipeptide			
expt	1	1	13.87	6.98	16.29	1579
NVT	2892 ± 6702	2884 ± 7273				
NPT	239 ± 5795	227 ± 7372	14.62 ± 0.19	7.02 ± 0.04	16.17 ± 0.07	1660

[a] Pressures in atm, lengths in Å, and volumes in Å^3; the unit cell parameters are fixed in the NVT ensemble at the experimental values.

TABLE 14: Rms Differences from the *N*-Methylacetamide and Alanine Dipeptide Crystal Calculations[a]

system	minimized	NVT	NPT
NMA	0.036	0.039	0.037
alanine dipeptide	0.139	0.125	0.152

[a] Rms differences in Å for all non-hydrogen atoms following a least-squares fit of the non-hydrogen atoms to the crystal structures. For the simulations the time-averaged structures were used.

Since the internal geometries of NMA and the alanine dipeptide in the crystal calculations are very close to the experimental values, the changes in the unit cell parameters are associated with the nonbonded interaction distances. Table 2 of the Supporting Information lists various distances for both NMA and alanine dipeptide. For NMA the differences between the experimental and calculated distances are small and similar. This is consistent with the isotropic changes in the unit cell parameters (see Table 13 and Table 1 of the Supporting Information). In the minimization, the majority of distances become slightly shorter than experiment. In the simulations in the NVT ensemble, the majority of distances are slightly longer than experiment with all the differences well within the rms fluctuations of the simulation values. In the NPT ensemble, the majority of distances again contract and the deviations from the experimental values are all smaller than those from the minimization. In the alanine dipeptide, most distances increase in the minimization, as well as in the NVT and NPT calculations. There was a significant expansion of the *A*-axis (see Tables 13 and Table 1 of the Supporting Information), while the agreement of axes B and C with experiment was satisfactory. The diagram of the crystal structure in Figure 5B shows the interactions in the crystal. Hydrogen bonds involving the peptide bonds are aligned with the *B*- and *C*-axes. As may be seen in Table 2 of the Supporting Information, the interaction distances between the nitrogens and oxygens in the alanine dipeptide are generally too long; for example, the N17 to O5 distances increase by approximately 0.2 Å in both the NVT and NPT simulations. These differences, however, were not observed in the crambin, BPTI, and MbCO crystal simulations (see below), so no additional optimization of the parameters was performed. Analysis of the remaining interactions, many of which are associated with the *C*-axis, shows a trend in the simulations for the nonbonded interaction distances to increase in the simulations. This is true, in particular, for the primary-to-primary interactions between the peptide bonds of molecules 1 and 2 (atoms C12 and N17) and the primay-to-image interactions involving O6 to C023 1 C11 and O6 to C001 1 C19. All of these interaction distances increase significantly in both the minimization and the NPT simulation. Such expansion may be due to limitations in the parameters for the interaction between polar atoms and aliphatic moieties. Limits in the potential energy function related to the interaction of the dipoles of the peptide bonds containing atoms C12 and N17 in molecules 1 and 2 of the primary atoms may lead to the associated increased distances. In addition, the spherical model for the atomic van der Waals surfaces may be insufficient to reproduce the interactions of the π orbitals of the peptide bonds.[103] Despite these limitations, the current parameters adequately reproduce the NMA crystal structure and lead to a reasonable reproduction of the L-alanine dipeptide crystal, although areas for improvement are evident.

IV.b. Tripeptide Crystal Simulations. Previous parametrization studies of proteins have focused on cyclic peptides as test systems (see Table 16).[6,8,10,104] Although we also consider cyclic peptides (see Section IV.c), their constrained structures and the lack of ionic groups limits their applicability as model systems for proteins. Consequently, we also used three noncyclic tripeptide crystals in testing the present parameter set. They are Gly-Ala-Leu·3H_2O (GAL), Gly-Ala-Val·3H_2O (GAV),[51] and Ala-Ala-Ala (AAA).[52] GAL and GAV represent conformers that are nearly α-helical and have been suggested to correspond to nucleation structures for helices, while AAA has an extended parallel β-pleated sheet conformation. Diagrams of the three tripeptides are shown in Figure 6. All of these structures are zwitterions, which allows for testing of the present parameters on nonbonded interactions involving ionic groups.

Crystal minimizations as a function of different cutoffs were performed to test the influence of the truncation scheme on the resulting structures. The minimization results are presented in Table 3 of the Supporting Information. As in the NMA and alanine dipeptide crystal minimizations presented above (Section IV.a.3), there are significant fluctuations in the unit cell parameters and the energies as the cutoff distances change for the shorter cutoff distances. For the longer cutoff distances, the fluctuations decreased. The 22−21−19 cutoff regime was again selected for more detailed studies, although the unit cell parameters and energies have not fully converged. The GAL and AAA crystals contract in an isotropic fashion. There is some asymmetry in the contraction in the GAV crystal, with the *A*-axis contracting, the *B*-axis relatively unchanged, and the *C*-axis expanding. In all cases the minimizations lead to the expected decrease in the total volumes of the crystals, as discussed in Section IVa.3.

For the GAL crystal, which has an orthogonal space group, both NVT and NPT molecular dynamics simulations were performed; only NVT simulations were performed for GAV and AAA. Table 17 gives the global crystal properties and Table 18 presents the rms differences between the simulation results and the crystal structures. In the NVT simulations, positive pressures were obtained for GAV and GAL, while a negative pressure was obtained for AAA. Correspondingly, the GAL NPT simulation yielded a small expansion of the unit cell. The

TABLE 15: *N*-Methylacetamide and Alanine Dipeptide Intramolecular Dihedral Angles[a]

	X-ray	min	diff	NVT	diff	NPT	diff
			NMA				
Cm−C−N−C	0	0	0	0 ± 6	0	0 ± 5	0
			Alanine Dipeptide				
			Molecule 1				
C4−C5−N7−C9, ω	−168	−178	−10	−177 ± 8	−9	−177 ± 8	−9
C5−N7−C9−C12, ϕ	−84	−90	−6	−89 ± 12	−4	−88 ± 13	−4
N7−C9−C12−N17, ψ	159	158	−1	153 ± 11	−6	157 ± 12	−2
C9−C12−N17−C19, ω	173	−179	8	−179 ± 9	8	179 ± 9	6
			Molecule 2				
C4−C5−N7−C9, ω	−178	−178	1	−176 ± 8	2	−177 ± 8	1
C5−N7−C9−C12, ϕ	−88	−89	−1	−91 ± 11	−3	−90 ± 12	−2
N7−C9−C12−N17, ψ	155	159	5	160 ± 10	6	158 ± 12	4
C9−C12−N17−C19, ω	172	−179	9	−180 ± 9	9	180 ± 9	8

[a] Dihedral angles in deg. The variations shown for the NVT and NPT simulation results correspond to the rms fluctuations.

TABLE 16: Peptides in the Present Study

peptide	identifier	space group	*Z*	ref
Tripeptides				
Gly-Ala-Val·3H_2O	GAV	$P2_1$	2	51
Gly-Ala-Leu·3H_2O	GAL	$P2_12_12_1$	4	51
Ala-Ala-Ala	AAA	$P2_1$	4	52
Cyclic Peptides				
cyclo-(Ala-Ala-Gly-Gly-Ala-Gly)·H_2O	CP1	$P2_1$	2	53
cyclo-(Ala-Ala-Gly-Ala-Gly-Gly)·2H_2O	CP2	$P2_12_12_1$	4	54
cyclo-(Gly-Gly-D-Ala-D-Ala-Gly-Gly)·3H_2O	CP3	$P2_12_12_1$	4	55
cyclo-(Gly-Pro-Gly-Gly-Pro-Gly)·4H_2O	CP4	$P2_1$	4	56
cyclo-(Gly-Pro-Gly-D-Ala-Pro)	CP5	$P2_12_12_1$	4	57
cyclo-(Cys-Gly-Pro-Phe-Cys-Gly-Pro-Phe)·4H_2O	CP6	$P2_1$	4	58

expansion of 2.7% corresponds to the positive pressure of approximately 4000 atm in the NVT simulation. It involves a small contraction of the *A*-axis and a significant expansion of the *C*-axis, while the *B*-axis remains unchanged. The changes in the unit cell differ from those that occurred in the minimizations, again emphasizing the importance of molecular dynamics simulations for the analysis of crystal properties. The rms differences of the nonhydrogen atoms were 0.5 Å or less. This indicates that there are only minimal changes in the internal geometries. The rms differences found on minimization were significantly smaller (0.15 Å or less) than those resulting from the simulations.

To obtain more detailed information on the changes in the internal geometries of the tripeptides, the dihedral angles of the X-ray and the simulation structures were compared (see Table 4 of the Supporting Information). The minimizations led to only minor changes in the dihedral angles; the largest change was −15° for the C-terminal dihedral angle of GAL. Although in most cases the simulations also show deviations of less than 15°, certain of the dihedral angles undergo a transition from one minimum to another. The largest differences occurred in the N-terminal dihedral angles of GAL and GAV. The environment of the terminal ionic groups of GAL and GAV is dominated by water (see Figure 6 and Table 5 of the Supporting Information and below). This is especially true for the two amino termini where there are no interactions with any peptide atoms. The interactions with water are such that small changes in their positions can lead to reorientation of the terminal amino group. The temperature factors of the GAL and GAV peptides[51] are large for the N-terminal nitrogen and the C-terminal oxygens. This indicates that there may be disorder in the crystal so that the structural changes found in the simulations are not unreasonable. In the AAA crystal, the terminal ionic groups interact with the oppositely charged ionic groups of neighboring peptides so that no significant changes in the dihedral angles occur. The other dihedral angle that changed significantly is $C_\alpha-C_\beta-C_\gamma-C_{\delta1}$ of the leucine side chain in the GAL NVT simulation. Since this change was not observed in the NPT simulation, it suggests that there are two minima that are very close in energy.

The nonbonded distances involving nitrogen and oxygen atoms in the tripeptide crystals are shown in Table 5 of the Supporting Information. In the minimizations, there are only small changes in distances, most of which decreased, as expected. For the NVT simulations of AAA, the changes in the nonbonded distances are similar to those found from the minimization; as in the latter, most of the distances decrease although certain ones increase a little. This is in accord with the tight packing and absence of water in the crystal. Significantly greater changes in the distances occur in the NVT molecular dynamics simulations of GAL and GAV. As can be seen from the contacts included in Table 5 of the Supporting Information and in Figure 6A,B, most of the nonbonded contacts are between the peptides and water rather than between peptides; of the contacts listed in Table 5 of the Supporting Information only four for GAL and three for GAV do not involve water. Thus, the tripeptides appear to be able to undergo displacements with relatively small changes in energy owing to the coupled rearrangement of water molecules; e.g., the overall rms for GAL is 0.5 Å in the NVT ensemble. When the interactions involve a water molecule (see Table 5 of the Supporting Information), the differences in the distances are particularly large, with many of them greater than 0.5 Å. This trend is maintained in the GAL NPT simulation, where even larger average differences occur. Analysis of the rms fluctuations of the distances in which large changes occurred reveals values of 0.5 Å or greater. The isotropic *B*-values for the three waters in GAV correspond to rms fluctuations of 0.40, 0.43, and 0.37 Å. The water molecules in the tripeptide crystal simulations undergo large shifts in position that lead to the change in the nonbonded interaction distances. This may indicate limitations in the force field with respect to interactions between water molecules and charged species. Calculations on ionic model compounds in solution with the present parameters[14] show that they reproduce both microscopic interactions with water and macroscopic thermodynamic properties. In the GAL and GAV crystals, individual water molecules interact with more than one charged group. In these structures polarization effects are enhanced and the omission of polarization in the present potential energy function could be important. However, it is also possible that there is a range of structures of similar energies that are sampled in the simulations. Clearly, the three tripeptide crystals are a good test for the validity of potential energy functions. It would be

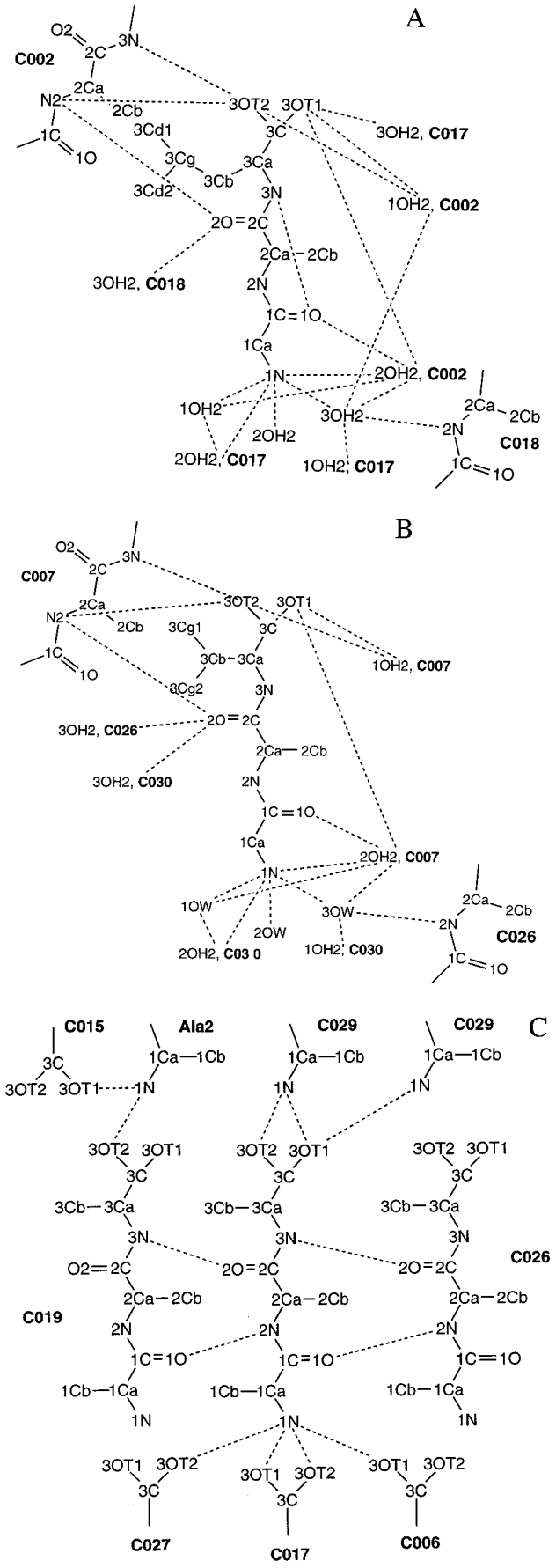

Figure 6. Diagrams of the interactions between the primary and images atoms for the (A) Gly-Ala-Leu·$3H_2O$ (GAL), (B) Gly-Ala-Val·$3H_2O$ (GAV), and (C) Ala-Ala-Ala (AAA) tripeptides. Bold characters identify images based on the CHARMM image nomenclature.

interesting to have results for these systems obtained with other force fields in current use. Useful data for such comparisons can be found in Tables 3, 4, and 5 of the Supporting Information.

IV.c. Cyclic Peptide Crystal Simulations. The cyclic peptide crystals studied with the present potential function are listed in Table 16. They are the peptides that have been used in tests of other empirical potential energy functions.[6,8,10,104] All are relatively flexible because they contain Gly and Ala residues and there are varying numbers of waters in the crystals.

Table 6 of the Supporting Information lists the minimized unit cell parameters and energies for the cyclic peptides as a function of cutoff distance. The observed trends are similar to those found for the tripeptides. Large fluctuations in the results occur for the shorter cutoff distances, with the values converged at the longer distances. As before, the 22−21−19 truncation scheme was selected for NVT simulations. The rms differences of the non-hydrogen atoms with respect to the crystal structure from the minimizations and from the NVT time-averaged structures are shown in Table 19, along with the pressures obtained in the simulation. In all cases the pressures are small and positive. They are of magnitudes similar to those seen for GAL and GAV (see Table 17). The rms differences are 0.45 Å or less, demonstrating that the parameter set adequately reproduces the internal geometries of the molecules. In five of the six cyclic peptides the rms difference is smaller in the average structure from the NVT simulation than in the minimized structure. This again points to the need to perform crystal studies using molecular dynamics simulations, rather than energy minimizations. Previously published rms differences from energy minimizations using other parameter sets are also included in Table 19; they show that all the listed parameter sets give similar values. Rigorous comparisons between parameter sets are not possible because of differences in minimization methodologies; no molecular dynamics simulations were performed in tests of the other parameter sets.

The changes in the dihedral angles following energy minimization and from the time averages of the NVT simulations are presented in Table 7 of the Supporting Information. The average differences and rms fluctuations for all the ϕ and ψ dihedrals in the cyclic peptides ($n = 37$) were 1.7 ± 16.3 and $-1.5 \pm 15.3°$, respectively. Thus, the backbone parameters do not bias the conformation of the backbone in a systematic manner. The individual dihedral angles show small differences, in general. However, in accord with the sizable rms fluctuations about the averages, there are some large differences. Some of these (e.g., the change of the 2C−3N−3C$_\alpha$−3C and 3N−3C$_\alpha$−3C−4N dihedral angles on minimization of CP1) occur in an anticorrelated fashion so that there result only small changes in the overall structures of the peptides.[105] For many of the dihedral angles listed in Table 7 of the Supporting Information there is a decrease in the differences upon going from the minimized structures to the simulation averages, in accord with that found for the rms differences (Table 19). The largest difference in the dihedral angles occurs for 2C−3N−3C$_\alpha$−3C in CP2. This change is partially compensated for by an anticorrelated change in the 2N−2C$_\alpha$−2C−3N dihedral angle that precedes the peptide bond prior to the 2C−3N−3C$_\alpha$−3C dihedral; however, significant displacements of the atoms in residues 1 and 2 do occur (see Table 8 of the Supporting Information and below). Though still significant, the magnitudes of the changes in these two dihedral angles decrease in the NVT simulation.

The nonbonded interaction distances are presented in Table 20. Average differences and rms fluctuations of the differences

TABLE 17: Tripeptide Crystal Simulation Results[a]

simulation	P_{ext}	P_{int}	*A*	*B*	*C*	β	volume
			GAL				
expt	1	1	6.024	8.171	32.791		403.5
NVT	3943 ± 27072	3909 ± 8069					
NPT	52 ± 25643	52 ± 7767	5.899 ± 0.054	8.175 ± 0.031	34.39 ± 0.36		414.5 ± 3.9 (2.7%)
			GAV				
expt	1	1	8.052	6.032	15.779	98.52	379.0
NVT	4919 ± 13496	5478 ± 7173					
			AAA				
expt	1	1	11.849	10.004	9.862	101.3	573.2
NVT	−2094 ± 7127	−1594 ± 5656					

[a] Distance in Å, angles in deg, volume in Å^3, and pressures in atm. The variations shown for the NVT and NPT simulation results correspond to the rms fluctuations, and the value in parantheses for the volume is the percent change in the simulation with respect to the experimental value. Volumes represent the asymmetric unit on which the calculations were performed. It should be noted that the volumes in Table 3 of the Supporting Information are for the total unit cell.

TABLE 18: Rms Differences of the Tripeptide Crystals[a]

	minimized		simulation NVT		simulation NPT	
system	protein	water	protein	water	protein	water
GAL	0.131	0.126	0.498	0.327	0.298	0.411
GAV	0.134	0.153	0.285	0.339		
AAA	0.110		0.130			

[a] Rms differences in Å for all non-hydrogen atoms following a least-squares fit of the non-hydrogen atoms to the crystal structures. For the simulations, time-averaged structures were used.

TABLE 19: Cyclic Peptide Rms Differences of All Non-Hydrogen Atoms Excluding Water and Pressure[a]

Minimized Rms Differences				
compd	CHARMM22	MM3[b]	AMBER/Weiner[c]	AMBER/OPLS[c]
cp1	0.31	0.10	0.11	0.09
cp2	0.45	0.09	0.23	0.14
cp3	0.13	0.11		
cp4	0.15	0.15	0.36	0.25
cp5	0.21	0.08	0.17	0.18
cp6	0.19	0.09		

NVT simulation	rms difference	PresE	PresI
cp1	0.17	2381 ± 2903	2516 ± 4663
cp2	0.33	1177 ± 8858	1174 ± 4751
cp3	0.18	4061 ± 6047	4057 ± 7226
cp4	0.14	2806 ± 3294	3279 ± 5967
cp5	0.19	5581 ± 6800	5567 ± 7820
cp6	0.16	3185 ± 3268	3247 ± 5169

[a] Distance in Å, and pressures in atm. Rms differences are for all non-hydrogen atoms with respect to the crystal structure following a least-squares fit of those atoms. Errors for the pressures represent the rms fluctuations. [b] See ref 10; no convergence criteria were reported, all unit cells were constrained to the experimental values, and the unit cells of CP1 and CP2 were expanded by 1% prior to the energy minimizations. [c] See ref 8; no convergence criteria were reported, unit cells were included in the energy minimizations, and rms differences include all atoms. AMBER refers to the parameters reported in ref 6.

for specific types of interacting atoms are given. The first six interactions listed are standard hydrogen bonds. For those involving solute-to-solute hydrogen bonds (N to O and O to O), there is a tendency for the distances to increase in the calculations, while hydrogen-bond interactions involving water molecules tend to decrease. The NVT simulations agree slightly better with experiment than the minimization results. The solute−solute interactions tend to be slightly too long, although the differences are less than the rms fluctuations, while the hydrogen-bond lengths involving water molecules are in excellent agreement with experiment. For the remaining interactions listed in Table 20, the agreement with experiment is generally satisfactory, although the trend toward increased distances is evident. In some cases the behavior is due to the fact that a small number of interactions have relatively large increases in distance (see below). Overall, the nonbonded interaction distances in the cyclic peptides are adequately reproduced by the force field.

TABLE 20: Average Difference and Rms Fluctuations of the Difference for the Nonbonded Interaction Distances between the Cyclic Peptide X-ray Structures and the Crystal Minimizations and Simulations[a]

atom pair		number	crystal	minimized difference	dynamics difference
N	O	43	3.11 ± 0.18	0.14 ± 0.30	0.13 ± 0.21
O	O	9	3.32 ± 0.13	0.18 ± 0.18	0.14 ± 0.20
OH2	OH2	6	2.82 ± 0.03	−0.09 ± 0.04	−0.04 ± 0.04
OH2	O	25	2.91 ± 0.21	−0.06 ± 0.17	0.01 ± 0.15
OH2	N	22	3.19 ± 0.23	−0.06 ± 0.18	0.02 ± 0.15
C	O	27	3.19 ± 0.25	0.18 ± 0.18	0.17 ± 0.13
C	N	17	3.24 ± 0.10	0.14 ± 0.12	0.14 ± 0.10
CA	O	20	3.35 ± 0.08	0.14 ± 0.20	0.15 ± 0.12
CA	OH2	16	3.32 ± 0.08	0.10 ± 0.08	0.15 ± 0.05
CA	N	1	3.37	0.01	0.13
CA	CE2	1	3.32	0.10	0.17
CA	CZ	1	3.49	0.20	0.17
CB	O	11	3.35 ± 0.13	0.04 ± 0.13	0.08 ± 0.14
CB	OH2	2	3.48	0.14	0.08 ± 0.04
CG	O	3	3.36 ± 0.08	0.18 ± 0.17	0.20 ± 0.14
CD	O	4	3.34 ± 0.14	−0.06 ± 0.11	0.05 ± 0.08
CD	C	1	3.41	0.19	0.22
CD	N	1	3.10	0.26	0.24
CD1	N	1	3.31	0.05	0.06
CD2	C	1	3.43	0.09	0.12
CD2	N	1	3.22	0.00	0.00
CE2	N	1	3.38	0.09	0.14

[a] All non-hydrogen to non-hydrogen nonbonded interaction distances less than 3.5 Å in the X-ray structures are included. Distances in Å. Crystal data represents the experimental averages and the fluctuations. *N* is the number of interactions between specific atom types. No fluctuations are given if a specified interaction appeared only once.

Table 8 of the Supporting Information lists the individual nonbond interaction distances in the cyclic peptide crystals. In accord with the results in Table 20, the differences are generally less than 0.2 Å. In the minimizations, decreases of up to 0.4 Å and some increases greater than 1 Å occur, while in the NVT simulations, the maximum decrease is 0.34 Å and the maximum increase is 0.85 Å. The largest increases occur in CP2; the cause of these changes is not evident, although they are associated with the large shift in the backbone dihedrals in CP2 (see Table 7 of the Supporting Information). This asymmetric behavior is due to a skewing of the distribution toward larger values by

the steep van der Waals repulsions at short distances. Analysis of the rms differences and the average differences for all interactions listed in Table 8 of the Supporting Information show values of 0.23 and 0.04 ± 0.14 Å for the minimizations, respectively, and 0.19 and 0.04 ± 0.07 Å for the NVT simulations. There is a small improvement upon going from the minimized structures to the simulation averages, as in the other results.

IV.d. Protein Simulations. The final test of the parameters was performed by using them for NVT simulations of crambin, bovine pancreatic trypsin inhibitor (BPTI), and carbonmonoxy myoglobin. For comparison, the results of vacuum simulations of these three proteins are also presented. Since the CHARMM 22 force field is designed for the condensed phase, the solvated results are expected to be significantly better than the vacuum calculations. This is borne out by the comparisons.

Table 21 presents an overall comparison of the crystal and vacuum simulations with the experimental data. Pressures from the crystal simulations range from −2000 to 1250 atm. The pressures are sensitive to the exact number of waters included in the simulations, which have to be determined by an overlay procedure since the total number of waters observed in the crystal is smaller than the actual number (see methods section). For the GAL tripeptide (see Table 17), a pressure of 3900 atm in the NVT calculation corresponded to an expansion of the unit cell volume by 2.7% in the NPT ensemble. This suggests that the number of waters is satisfactory.

In the simulation protocol, the systems were initially heated over 5 ps to temperatures of 300, 285, and 260 K for the crambin, BPTI, and carbonmonoxy myoglobin crystal simulations, respectively, and 285 K for all vacuum simulations; the crystal temperatures correspond to those used in the structure determinations. Heating was followed by 5 ps of equilibration. For the crambin and BPTI crystal simulations the temperature remained in the initial range, while a slight rise occurred in the carbonmonoxy myoglobin simulation. Increases of 10° or more in temperature occurred in the BPTI and carbonmonoxy myoglobin vacuum simulations. This was associated with additional relaxation of the potential energy of the systems following the 10 ps of heating and equilibration; both relaxations occurred after approximately 150 ps of simulation, and the energies in the final 100 ps of the simulations were stable. The temperature of all systems was stable over the final 100 ps (see methods section) and the total energy of the systems was well-conserved with rms fluctuations of less than 0.5 kcal/mol.

The structural changes resulting from the simulations can be evaluated in terms of the rms differences and the change of the radius of gyration with respect to the crystal structures. The values reported in Table 21 correspond to the time-averaged structures from the final 100 ps of the simulations. The rms differences and radius of gyration reached stable values in 10 ps for the crystal simulations and 30 ps for the vacuum simulation of crambin and fluctuated about those values for the remainder of the simulations. In the BPTI and carbonmonoxy myoglobin vacuum simulations, a second structural adjustment occurred at approximately 150 ps, in accord with the relaxation of the energies (see above). This was followed by stable behavior over the final 100 ps of the simulation. In the crystal the rms differences were less than 1.0 Å for all non-hydrogen atoms. Such agreement with experiment is satisfactory. The rms differences in the vacuum simulations were significantly larger than those of the crystal calculations, as expected.

Also included in Table 21 are the rms fluctuations averaged over various groups of non-hydrogen atoms from both experiment and the molecular dynamics simulations. The experimental results are obtained from the crystallographic isotropic *B*-factors, without any corrections.[106] Comparison of the experimental and crystal simulation results show that the calculated values are consistently smaller than the experimental values. This is as expected since crystal lattice disorder and rigid-body motions contribute to the *B*-factors,[48] in addition to the internal motions, which are calculated in the simulations. For all three systems the rms fluctuations increase upon going from the crystal environment to vacuum. It is clear that the

TABLE 21: Overall Protein Crystal Simulation Results[a]

property	exptl	crystal	vacuum
		Crambin	
internal pressure	1	1254 ± 1659	
temp	room	304 ± 7	286 ± 7
total energy		−810.2 ± 0.4	328.4 ± 0.1
rms difference			
backbone[d]		0.63	1.70
side chain[d]		0.94	2.16
non-hydrogen[d]		0.76	1.91
radius of gyration			
backbone	9.594	9.564	9.469
non-hydrogen	9.667	9.644	9.513
rms fluctuations			
C_α	0.46	0.32	0.50
backbone	0.47	0.34	0.51
side chain	0.55	0.45	0.68
non-hydrogen	0.50	0.39	0.58
		BPTI	
internal pressure	1	−2010 ± 1362	
temp	room	287 ± 6	295 ± 8
total energy		−2221.4 ± 0.1	−502.0 ± 0.2
rms difference			
C_α[c]		0.86	2.63
backbone[d]		0.82	2.58
side chain[d]		1.09	3.73
non-hydrogen[d]		0.96	3.19
radius of gyration			
backbone	10.607	10.838	10.348
non-hydrogen	10.944	11.222	10.562
rms fluctuations			
C_α	0.71	0.37	0.46
backbone	0.70	0.39	0.47
side chain	0.80	0.53	0.62
non-hydrogen	0.75	0.46	0.54
		MBCO	
internal pressure	1	−357 ± 828	
temp	260	268 ± 4	297 ± 5
total energy		−5331.7 ± 0.4	−173.4 ± 0.4
rms difference			
C_α[c]			1.98
backbone[d]		0.72	1.97
side chain[d]		1.16	2.59
non-hydrogen[d]		0.97	2.30
radius of gyration			
backbone	15.052	15.242	15.178
non-hydrogen[b]	15.047	15.279	15.139
rms fluctuations			
C_α	0.56	0.37	0.48
backbone	0.55	0.39	0.49
side chain	0.62	0.54	0.63
non-hydrogen	0.59	0.46	0.56

[a] Internal pressure (atm), temperature (K) and total energies (kcal/mol) are over the final 200 ps for the crambin crystal, 100 ps for the BPTI and myoglobin crystals, and 290 ps for all vacuum simulations. The rms differences (Å), radii of gyration (Å), and rms fluctuations (Å) are calculated from the time-averaged structures from the final 100 ps of the respective simulations. Errors represent the rms fluctuations of the respective values. [b] Includes the heme. [c] Following least-squares fit to all C_α atoms. [d] Following least-squares fit to all backbone (C, N, C_α, O) atoms.

presence of the condensed-phase environment is important for both the structural and dynamic properties of these systems.

Comparison of the results in Table 21 can be made with previous studies on the same systems. The only published crystal simulation for BPTI included the full unit cell (four individual BPTI molecules) and was 20 ps in duration.[107] The rms differences for all non-hydrogen atoms and C_α atoms, respectively, were 1.31 and 0.88 Å for a structure obtained by averaging over the last 2 ps and all four molecules. The present crystal values show smaller deviations, despite the extended simulation time and no averaging over different structures. The vacuum rms differences for BPTI are larger than the previous values.[108] This is not unexpected since the parameters were optimized explicitly for condensed-phase simulations, while previous parameter sets were often adjusted for use in a vacuum by inclusion of a distance-dependent dielectric function and neutralization of the charged side chains. Most previous studies on carbonmonoxy myoglobin are vacuum simulations. As an example, a 96-ps simulation of a carbonmonoxy myoglobin−xenon complex using the AMBER extended atom force field[86] yielded rms differences of approximately 2.1 Å for all atoms and 1.6 for the C_α, C, N atoms.[109] These are somewhat smaller than the vacuum values found here, though significantly larger than the crystal simulation results reported in Table 21. The rms fluctuations were 0.72 and 0.71 Å for all atoms and the main chain atoms, considerably larger than the vacuum fluctuations found here. A series of simulations on carbonmonoxy myoglobin hydrated with 349 water molecules was performed for comparison with neutron-scattering data[110] using an all-atom representation from Polygen Corporation.[111] At 260 K rms differences for the entire protein and the backbone atoms with respect to the crystal structure were 1.82 and 1.62 Å, respectively, significantly larger than those for the crystal with the present parameter set. The rms fluctuations were 0.55 and 0.44 Å for all protein and backbone atoms, respectively, which is comparable to the values in Table 21 for the vacuum simulations, while the present crystal values are smaller. In another study a vacuum simulation of carbonmonoxy myoglobin was performed at 298 K for 300 ps using an extended atom representation with the exception of certain methyl groups on which hydrogens were included.[112] For the final 100-ps window in that study, rms differences of 2.82 and 2.45 Å were obtained for all atoms and the backbone atoms and rms fluctuations of 0.707 Å were obtained for the N, C, and C_α atoms. Recently, a 1.4-ns simulation of carbonmonoxy myoglobin in solution with periodic boundary conditions has been performed using the present parameters (B. K. Andrews and B. M. Pettitt, personal communication). Rms differences of the final time frame with respect to the crystal structure[113] were 2.05 and 1.49 Å for the protein non-hydrogen and backbone atoms, respectively. Analysis of the backbone conformation showed average differences, determined over the final 1 ns, of 0.10 ± 1.53 and -2.38 ± 1.49 for ϕ and ψ, respectively, with respect to the crystal values. This confirms that the parameters do not induce systematic variations in the backbone conformation in extended simulations. Rms fluctuations from the final 1 ns of the simulation averaged over the residues were 0.90 ± 0.33, 1.04 ± 0.43, and 0.73 ± 0.25 Å for all, side chain, and main chain non-hydrogen protein atoms, respectively. These compare well with the estimates from *B*-factors of 0.80 ± 0.14, 0.82 ± 0.18, and 0.76 ± 0.12 Å for all, side chain, and main chain non-hydrogen protein atoms, respectively. The larger values for the all-atom and side chain atom fluctuations obtained in the simulation are suggested to be due to the absence of crystal contacts. This is supported by the good agreement of the experimental and calculated rms fluctuations as a function of residue, except for selected regions where crystal contacts occur.[113]

TABLE 22: ϕ, ψ, and ω Averages and Rms Fluctuations from the MD Simulations[a]

	average differences			rms differences			rms fluctuations		
	ϕ	ψ	ω	ϕ	ψ	ω	ϕ	ψ	ω
				Crambin					
crystal	2.4	−1.5	−2.0	16.3	17.7	4.8	20.2	21.5	8.2
vacuum	−0.5	−1.0	−0.9	26.0	29.6	6.9	29.9	41.0	10.2
				BPTI					
crystal	−0.6	1.1	−1.2	15.9	20.9	4.4	19.4	23.9	8.2
vacuum	−0.3	5.2	0.4	40.8	36.2	8.7	42.4	38.3	11.4
				Myoglobin					
crystal	−0.1	1.0	−2.4	13.0	14.5	6.9	16.1	17.3	9.1
vacuum	−4.2	5.1	−2.5	20.8	22.9	8.3	27.1	28.7	11.1

[a] Values in deg. Dihedral angles and rms fluctuations were calculated from the time average of the final 100 ps of the simulations. Average difference is the sum of the differences between the time average and crystal dihedrals divided by the total number of dihedrals in each protein, rms difference is the rms of all the differences in each protein, and rms fluctuations represents the rms of all the individual dihedral rms fluctuations calculated from the simulations.

The radius of gyration for the present parameter set shows a consistent decrease upon going from the crystal to a vacuum environment, which is due to the loss of crystal interactions and solvent contributions. In crambin the crystal simulation radius of gyration is in good agreement with experiment, while the BPTI and carbonmonoxy myoglobin values are too large by 2.5 and 1.5%, respectively. In the vacuum simulations the radius of gyration of the crambin and BPTI time-averaged structures is 1.6 and 3.5% smaller than experiment, while the MbCO vacuum result is 0.6% larger. Previous calculations on hydrated carbonmonoxy myoglobin[110] yielded a radius of gyration of 14.67 Å, representing a 2.5% contraction, while results from a vacuum simulation yielded a radius of gyration for all atoms of 13.88 Å,[112] a 7.7% contraction.

An essential element of a protein empirical force field is that the protein backbone is represented accurately. The optimization of the backbone parameters was described in Section IV.a. To validate the backbone parameters, the backbone geometries obtained from the protein simulations were examined. Table 22 presents the calculated average and rms differences from the crystal structure for the ϕ, ψ and ω dihedral angles; the overall rms fluctuations of the respective dihedral angles are also listed. Use of the average difference, as well as the rms results, makes it possible to find systematic trends in the deviations that could indicate a problem with the parameters. No such trends are observed, as is evident from Table 22. Both positive and negative values are obtained for the average deviation of the three dihedral angles, depending on the protein and the simulation. Rms differences in the crystal simulations are 20° or less, with larger values in the vacuum simulations. Comparison of the rms differences and fluctuations show that the differences are smaller than the fluctuations in all cases. These results show that the parameter set is satisfactory in its treatment of the protein backbone.

No direct comparison of the present protein backbone results with those from other force fields can be made because no corresponding studies have been published. For several force fields, the rms differences in the backbone dihedrals based on energy minimizations of crambin have been reported. For ϕ, ψ, and ω, respectively, the rms differences are 10.3, 12.1, and 4.5° from a MM3 vacuum minimization;[10] also, values of 8.2°,

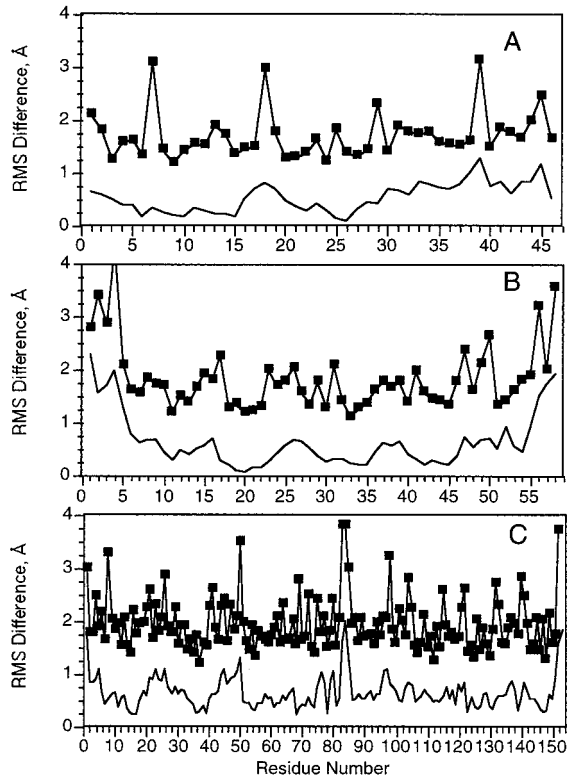

Figure 7. Rms differences versus residue between the 100-ps time-averaged crystal simulation structures and the experimental structures for (A) crambin, (B) BPTI, and C) myoglobin. The line represents the backbone non-hydrogen atoms (C, C_α, N, O), and the squares represent the side chain non-hydrogen atoms. Side chain rms differences have been offset by 1.0 Å. For glycine residues the side chain difference correspond to those of the carbonyl oxygen.

8.1°, and 3.4° from MM3, 7.2°, 7.9°, and 4.1 from AMBER (1984), and 6.1°, 5.6°, and 4.6° from AMBER/OPLS[8] for crystal minimizations have been reported. In the MM3 study,[10] plots of the difference in ϕ and ψ with respect to residue number showed trends where ϕ decreased for the majority of residues and the ψ values increased; such systematic deviations are undesirable, as discussed in Section IV.a. With the present parameters, crambin was minimized in the crystal environment for 1000 ABNR steps with SHAKE constraining all covalent bonds involving hydrogens. Rms differences of 6.6, 6.7, and 4.4° for ϕ, ψ, and ω, respectively, were obtained. These values are similar to those of the other parameter sets. However, it should be noted that in the present study a final rms gradient of the forces of 0.004 kcal/mol/Å was obtained as compared to values of 0.79−0.80 for AMBER and AMBER/OPLS; if the minimization is terminated at a rms gradient of 0.8, the rms deviations are 4.2°, 4.3°, and 3.1°. Thus, the other force-field results do not correspond to fully minimized structures. For the average difference, the present minimization yields values of 2.1, −1.5, and −2.3° for ϕ, ψ, and ω, respectively, which can be compared with values of −4.4, 3.7, and 0.0 for the MM3 crystal minimization. Such comparisons, although limited, indicate that the present parameters are equivalent to or better than published sets, at least for this property.

Figure 7 shows the rms differences as a function of residue numbers for the crambin, BPTI, and carbonmonoxy myoglobin crystal simulations versus the X-ray results. For the three proteins the main chain deviations are generally less than 1 Å, with most values below 0.5 Å. The side chains show larger deviations with certain ones as large as 3 Å. In crambin, for example, there are five conspicuous side chain deviation peaks for residues Ile7, Leu18, Tyr29, Thr39, and Ala45. In the crambin crystal, conformational heterogeneity occurs at positions 7 and 25 and compositional heterogeneity occurs at residues 22 and 25; residue 22 is either proline or serine, and position 25 is leucine or isoleucine. Proline and isoleucine, respectively, were used in those positions in the present study in accord with the more probable amino acid in the crystal. The compositional heterogeneity at position 22 has been suggested to lead to disorder at position 29.[114] The large differences at positions Leu18, Thr39, and Ala45 correspond to regions of high flexibility in the protein and result from the reorientation of the side chain. For example, residue Leu18 is at the surface of the protein and in the simulation rotations about the $C_\beta-C_\gamma$ and $C_\gamma-C_{\delta1}$ bonds lead to a change in the orientation of the residue, although a similar region of space is occupied in the two conformations. In such cases, it is likely that there are two positions of nearly equal free energy, and the simulation finds one of them while the other is observed in the particular crystal structure. In BPTI, the central region of the protein main chain was well-maintained while larger structural changes occurred at the termini. Analysis of the experimental rms fluctuations (see below) indicates that the terminal regions are very flexible. The large difference in residues 1−4 occurs such that the rings of Pro2 and Tyr4 and the guanidinium group of Arg1 are in approximately the same position as in the crystal structure. Since there is a salt bridge between the N- and C-termini, changes in one region are coupled to changes in the other. This probably explains the larger deviation at the C-terminal end. As in crambin and BPTI, the majority of residues in carbonmonoxy myoglobin shifted position by relatively small amounts. A number of the larger structural changes correspond to regions with large rms fluctuations (see below). Small structural changes are associated with the relatively rigid helices. Overall, the crystal simulations show the structures to be well-maintained by the present force field, although in certain cases, most notably the N-terminus of BPTI, significant structural changes do occur. Recent work (Caves, Evanseck, and Karplus, manuscript in preparation) involving comparison of crystal data with results from several simulations that differ only in the random number seed used to select the initial velocities shows that averages from the multiple simulations give better agreement with experiment than any of the individual simulations. This suggests that some of the larger differences observed in the present study may be associated with incomplete sampling of configurational space.

Figure 8 shows the rms fluctuations as a function of residue number for all non-hydrogen atoms in crambin, BPTI, and carbonmonoxy myoglobin. Values obtained from the experimental *B*-factors and the crystal simulations are also presented. In all three systems the experimental fluctuations are consistently larger than the calculated results, as expected (see the discussion of the overall values in Table 21). For crambin and BPTI, the pattern of fluctuations with respect to residue number in the experiments is reasonably reproduced in the calculations. The only exceptions are Thr36 in crambin and Lys15 in BPTI. Both of these residues are solvent-exposed, which allows for the increased mobility. Further, it is likely that such large fluctuations are underestimated by the X-ray *B*-factors.[115] For carbonmonoxy myoglobin the relative mobility of the residues from the calculations generally reflects the experimental data, although the correlation is significantly worse than in crambin and BPTI. Large deviations are found in the regions centered around residues 20 and 120 in which the calculated fluctuations do not show the marked increase seen in the experimental

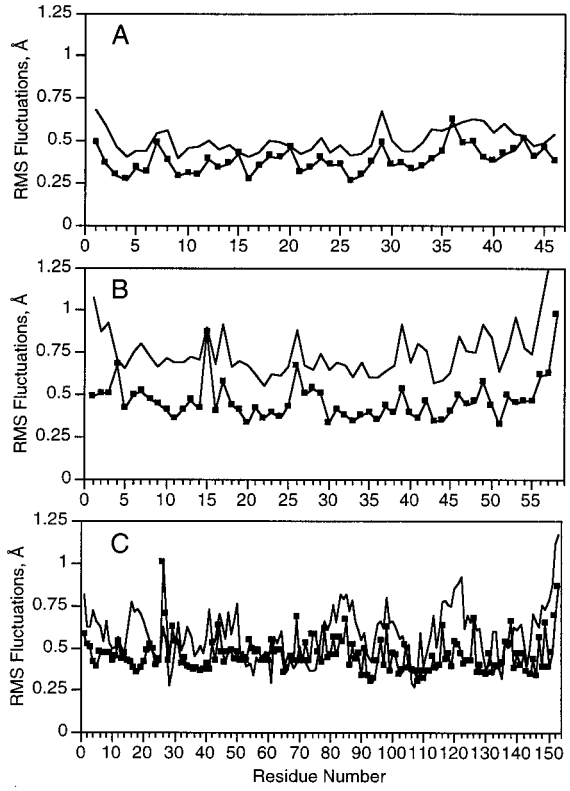

Figure 8. Rms fluctuations versus residue for all heavy atoms from the crystal simulations (squares) and experiment (lines) for (A) crambin, (B) BPTI, and (C) carbonmonoxy myoglobin.

B-factors. Both regions are centered in loops between helices; residue 20 is the first residue in helix B and residue 120 is between the final residue of helix G (residue 118) and the first residue of helix H (residue 124). It is likely that longer simulations (or multiple simulations) are needed to properly sample the helix motions that contribute to fluctuations in this region.[116] Limitations in agreement between molecular dynamics and experimental fluctuations may also occur owing to crystal contacts in the crystal that significantly alter the experimental *B*-factors.[117] Such effects, however, should not contribute significantly to the present comparison since the molecular dynamics calculations were performed in the crystal environment.

The results for the nonbonded distances of a variety of protein−protein and protein−water interactions are presented in Table 23. It lists the averages, rms deviations, and occupancies for protein−protein interaction distances involving oxygen, nitrogen, and sulfur atoms. The backbone nitrogen-to-backbone oxygen distances for helical and sheet regions of the proteins are given first; these regions were defined as in the Brookhaven PDB files for crambin (1crn), BPTI (5pti), and carbonmonoxy myoglobin (1mbc). Comparison of the experimental backbone N···O distances in crambin and BPTI show the average distance to be 0.25 Å or more shorter in the sheets, as compared with the helices. The average distance for the helical N···O interactions in carbonmonoxy myoglobin are similar to those in crambin and BPTI. This is in accord with a survey of hydrogen bonds in a variety of proteins, which showed that N···O distances are shorter in sheets than helices.[49] In the simulations the average distances are maintained, including the helix-sheet difference observed in the experimental structures. In the crambin X-ray results, the rms fluctuations of the distances are smaller for the sheets as compared with the helices; this trend is found in the simulation. For the occupancies, which represent the normalized number of such interactions per atom within 3.5 Å, the simulation and experimental values for the sheets are almost identical. For the helical regions, the calculated values are smaller than experiment in all three structures. This is attributable to the average distance and its fluctuations, which lead to distances greater the 3.5 Å for certain time frames and a resulting decrease in the occupancy by the criterion used. Since the average distance of the helix N···O interactions is larger than that for sheets, there is an increase of

TABLE 23: Protein-to-Protein Interaction Distances in the Protein Crystal Simulations[a]

	crystal		simulation		difference	
interaction	distance	occup.	distance	occup.	distance	occup.
Crambin						
Backbone-to-Backbone Interactions						
N···O helix	3.22 ± 0.19	1.76	3.20 ± 0.21	1.24	−0.02	−0.52
N···O sheet	2.89 ± 0.04	0.50	2.89 ± 0.14	0.49	0.00	−0.01
Side Chain-to-Backbone Interactions						
acid O···N	2.84 ± 0.07	0.33	2.88 ± 0.19	0.45	0.04	0.12
amide O···N	3.27 ± 0.06	1.00	3.24 ± 0.16	0.85	−0.03	−0.15
hydroxy O···N	3.24 ± 0.10	0.50	3.22 ± 0.18	0.49	−0.02	−0.01
amide N···O	3.40	0.33	3.16 ± 0.23	0.46	−0.24	0.12
amino N···O	2.93	1.00	2.89 ± 0.19	1.04	−0.05	0.04
Arg N···O	2.90 ± 0.08	0.33	2.85 ± 0.20	0.36	−0.05	0.02
hydroxy O···O	3.06 ± 0.26	0.80	3.06 ± 0.29	0.71	0.00	−0.10
BPTI						
Backbone-to-Backbone Interactions						
N···O helix	3.17 ± 0.22	1.31	3.19 ± 0.20	0.89	0.02	−0.42
N···O sheet	2.92 ± 0.19	0.67	3.00 ± 0.21	0.69	0.07	0.02
Side Chain-to-Backbone Interactions						
acid O···N	3.18 ± 0.12	0.50	3.19 ± 0.22	0.40	0.01	−0.10
amide O···N	3.16 ± 0.17	1.25	3.15 ± 0.25	0.64	−0.01	−0.61
hydroxy O···N	3.18 ± 0.08	0.38	3.27 ± 0.15	0.39	0.09	0.01
amide N···O	3.02 ± 0.17	0.50	3.20 ± 0.19	0.51	0.18	0.01
amino N···O	NI < 3.5		3.38 ± 0.07	0.00		
Arg N···O	2.57	0.06	3.38 ± 0.07	0.00	0.81	−0.05
hydroxy O···O	2.98 ± 0.27	0.50	3.12 ± 0.25	0.51	0.14	0.01
MbCO						
Backbone-to-Backbone Interactions						
N···O helix	3.19 ± 0.21	1.90	3.18 ± 0.22	1.44	−0.01	−0.46
Side Chain-to-Backbone Interactions						
acid O···N	2.91 ± 0.26	0.27	3.09 ± 0.26	0.23	0.18	−0.05
amide O···N	3.41	0.17	3.39 ± 0.09	0.03	−0.03	−0.14
His N···N	3.19	0.08	3.18 ± 0.20	0.18	−0.01	0.09
hydroxy O···N	3.12 ± 0.15	0.57	3.19 ± 0.20	0.44	0.07	−0.14
amide N···O	3.16 ± 0.19	0.33	3.08 ± 0.21	0.26	−0.08	−0.08
amino N···O	3.21 ± 0.11	0.15	2.85 ± 0.20	0.10	−0.36	−0.05
Arg N···O	NI < 3.5		3.23 ± 0.17	0.11		
His Np(0)···O	2.99 ± 0.36	0.50	3.10 ± 0.24	0.66	0.11	0.16
hydroxy O···O	3.03 ± 0.29	0.93	2.98 ± 0.30	0.76	−0.05	−0.17
Trp N···O	NI < 3.5		3.34 ± 0.12	0.11		

[a] Distance, in Å, represents the average distance of all interactions less than 3.5 Å with the error representing the rms fluctuations about the average. Occupancy is the number of interactions less than 3.5 Å per site, and the normalization procedure is based on the number of specified atom types in the different proteins; values of 0.00 indicates that the occupancy numbers based on the number of interactions less 3.5 Å yield values <0.005. NI < 3.5 indicates no interaction distance <3.5 Å in the experimental structure. Backbone-to-backbone interactions only included the peptide bond carbonyl oxygen and nitrogen atoms. Side chain-to-backbone interactions involve the side chain atoms as follows with the peptide bond carbonyl oxygen and nitrogen atoms. Atom types associated with the different types of specifications are as follows: Arg N, three nitrogens in the Arg guanidinium group; amide N, side chain nitrogens in Asn and Gln; amide O, side chain oxygens in Asn and Gln; acid O, side chain oxygens in Asp and Glu and the C-terminal carboxylate; His N, unprotonated nitrogen in neutral histidine; His Np(0), protonated nitrogen in neutral histidine; His Np(1), nitrogens in protonated histidine; amino N, nitrogen of Lys side chain and the N-terminal amino group; hydroxy O, the oxygen in the Ser, Thr, and Tyr side chains.

the likelihood that alterations of those interactions will yield distances longer than 3.5 Å resulting in lower calculated occupancy values. The ability of the present parameters to differentiate between the hydrogen bonds in helices and sheets suggests that the protein−protein backbone nonbonded interactions are being treated satisfactorily.

Some protein−protein nonbonded distances between the backbone and side chains are also shown in Table 23. For the majority of cases the differences in the average distances between the crystal and simulation values were less than 0.1 Å. Discrepancies include the amide N···O interactions in crambin, the ArgN···O interactions in BPTI, and the amino N···O interactions in carbonmonoxy myoglobin. The crambin amide N···O and BPTI ArgN···O interactions correspond to a single interacting pair. It is Asn12 to the carbonyl O of residue 8 in crambin. Figure 8A shows Asn12 to be relatively mobile and the residue is solvent-exposed; both of these may contribute to the calculated decrease in the interaction distance. The BPTI ArgN···O experimental distance is 2.57 Å (Arg1 NH1 to Thr55 O) and the Arg1 NH1 atom is 2.93 Å from the Tyr23 OH atom. Arg1 is one of the most mobile residues in the protein on the basis of the experimental temperature factors (see Figure 8B). Additionally, the Arg1 side chain moves approximately 1.7 Å from its starting position in the simulation (see Figure 7B). Thus, the deviation is probably not caused by the nonbonded parameters per se, but by the significant structural change. The shorter carbonmonoxy myoglobin amino N···O distance is likely to be associated with the flexibility of the lysine side chains and an error in its position in the X-ray structure; e.g., the crambin amino N···O distance is 2.93 Å, much closer to the calculated value. Occupancies for all the side chain-to-backbone interactions were well-maintained in the simulations. Only for the BPTI amide O···N interaction did a significant decrease occur. This decrease is due to the methodology used for the determination of the occupancy numbers, as discussed in the preceding paragraph. For most interactions where a decrease in the occupancy occurred relative to the crystal value, the average distance is relatively long. Importantly, the largest differences occurred for different interaction types in the various proteins. This suggests that specific effects are involved and that the nonbonded interactions parameters for all the amino acid side chain are not inducing systematic errors.

Since the BPTI and MbCO crystal structures include the positions of water molecules that were observed in the crystal, protein−water distances were also examined. It should be noted that the simulations contain many more crystal waters than those identified in the X-ray structures. In BPTI there are 63 waters in the neutron structure[46] and 92 waters in the simulation, including those identified in the experimental structure. For MbCO there are 137 experimentally identified waters[47] and a total of 345 waters in the simulation. The crambin crystallographic data entry deposited in the PDB does not contain any water molecules. Table 24 lists the experimental and calculated average distances, fluctuations, and hydration numbers for various protein hydrogen-bond acceptors and donors for the two systems analyzed. As with the protein−protein interactions, the protein−water interactions are well-reproduced in the calculations. Most of the discrepancies are related to the inclusion of all water molecules in the calculated values versus only the explicitly identified waters in the experimental studies. This leads to larger hydration numbers for the majority of interactions observed in the simulations. Concerning the interaction distances, the differences are greater than the fluctuations of the calculated values for BPTI and carbonmonoxy myoglobin in only three cases. The decrease in the BPTI amino N···OH2 interaction distance and increase in the hydration number in the simulation is due to the mobility and solvent exposure of the lysine side chains, which leads to only a minimal number of Lys NZ-to-water interactions that are identified in the crystal. In the simulations, these groups are fully solvated, leading to the shorter distance and the significant increase in the hydration number. A large increase in the hydration number of the amino N was also observed in carbonmonoxy myoglobin, although the experimental and calculated average distances are in good agreement. With carbonmonoxy myoglobin significant increases occur in the amide N···OH2 and HisNp(0)···OH2 interaction distances, where Np(0) represents the unprotonated nitrogen on the neutral histidine side chain. Analysis of the experimental crystal structure shows an amide N···OH2 distance of 2.32 Å and a HisNp(0)···OH2 distance of 2.49 Å. Such short interaction distances, which are probably wrong, lead to the average experimental distances that are too short. As with the protein−protein interactions, the protein−water interactions involving both the backbone and the side chains appear to be satisfactorily treated by the present force field.

TABLE 24: Water-to-Protein Interaction Distances in the Protein Crystal Simulations[a]

	crystal		simulation		difference	
interaction	distance	hyd. no.	distance	hyd. no.	distance	hyd. no.
			BPTI			
bkb N···OH2	3.13 ± 0.22	0.26	3.13 ± 0.21	0.29	0.00	0.03
bkb O···OH2	2.84 ± 0.27	0.68	2.90 ± 0.24	0.74	0.06	0.06
Arg N···OH2	3.12 ± 0.19	0.56	3.05 ± 0.23	1.06	−0.08	0.50
amide N···OH2	3.11 ± 0.22	0.75	3.12 ± 0.21	0.87	0.00	0.12
amide O···OH2	3.10 ± 0.18	0.75	2.92 ± 0.24	0.64	−0.18	−0.11
acid O···OH2	2.91 ± 0.42	1.50	2.84 ± 0.26	2.33	−0.07	0.83
Cys S···OH2	NI < 3.5		3.35 ± 0.11	0.10		
amino N···OH2	3.17 ± 0.18	1.60	2.91 ± 0.21	2.77	−0.26	1.17
Met S···OH2	NI < 3.5		3.37 ± 0.09	0.02		
hydroxy O···OH2	3.10 ± 0.20	1.13	2.95 ± 0.22	1.11	−0.16	−0.01
			MBCO			
bkb N···OH2	3.13 ± 0.30	0.08	3.19 ± 0.20	0.22	0.06	0.14
bkb O···OH2	2.93 ± 0.25	0.30	2.96 ± 0.26	0.56	0.04	0.25
Arg N···OH2	2.91 ± 0.36	0.75	3.06 ± 0.24	1.74	0.15	0.99
amide N···OH2	2.73 ± 0.51	1.17	3.13 ± 0.20	1.33	0.41	0.16
amide O···OH2	3.13 ± 0.44	1.00	2.98 ± 0.27	1.44	−0.16	0.44
acid O···OH2	2.83 ± 0.35	0.70	2.84 ± 0.26	2.09	0.01	1.38
His N···OH2	3.12 ± 0.34	0.67	3.00 ± 0.21	1.09	−0.12	0.42
His Np(0)···OH2	2.83 ± 0.27	0.33	3.14 ± 0.22	0.46	0.31	0.12
amino N···OH2	2.99 ± 0.37	0.70	2.98 ± 0.23	2.30	−0.01	1.60
Met S···OH2	NI < 3.5		3.48	0.01		
hydroxy O···OH2	3.13 ± 0.28	0.43	2.99 ± 0.23	1.03	−0.14	0.60
Trp N···OH2	2.91 ± 0.03	1.00	3.07 ± 0.19	0.37	0.16	−0.63

[a] Distance, in Å, represents the average of all interactions less than 3.5 Å with the error representing the rms fluctuations about the average. Hydration number is the number of interactions less than 3.5 Å per site. NI < 3.5 indicates no interaction distances <3.5 Å.

V. Concluding Discussion

Parameters for proteins have been developed for the empirical energy function used with the CHARMM program. The model treats all atoms, including hydrogens, explicitly. This is referred to as an all-atom model, in contrast to a model that treats only polar hydrogens explicitly and represents nonpolar hydrogens as part of extended atoms. The parameters were determined by fitting an extended set of experimental and ab initio results. A self-consistent approach was employed to obtain a proper balance between the intramolecular (bonding) and intermolecular (nonbonding) portions of the potential energy function. Emphasis was placed on a balance of the solvent−solvent, solvent−

solute, and solute−solute portions of the intermolecular portion of the potential energy. This is essential for accurate condensed-phase simulations from which both structural and thermodynamic information can be obtained. With the already published parameters for nucleic acids and lipids, the all-atom protein parameters form a consistent set that can be used for simulations of a wide range of molecules of biological interest. A list of the parameter values is included as an Appendix in the Supporting Information.

The parametrization was based on results for a wide variety of model compounds that represent the protein backbone and the individual side chains. Internal parametrizations (bond length, bond angle, Urey−Bradley, dihedral, and improper dihedral terms) were chosen to reproduce geometries from crystal structures, infrared and Raman spectroscopic data, and ab initio calculations. Interaction parameters (electrostatic and van der Waals terms) were chosen to fit 6-31G* ab initio interaction energies and geometries for water molecules bonded to polar sites of the model compounds and experimental condensed-phase properties such as heats of vaporization and molecular volumes.

The CHARMM22 parameters for proteins give satisfactory results when compared with test data. These include results for cyclic and acyclic peptides and several proteins. The details are provided in the body of the paper and we do not repeat them here. Instead we describe some of the aspects of the present work that distinguish it from other parametrization efforts and outline certain insights that are of interest in themselves and suggest points of general significance for macromolecular parametrization.

An essential element in the parametrization is the peptide group and the connectivity between the peptide groups that represent the repeating structural element of the protein backbone. *N*-Methylacetamide and the alanine "dipeptide" were chosen for optimization of the parameters that determine the geometries, vibrations and torsional potentials for the soft dihedral angles ϕ, ψ. The experimental data that were used for comparison were obtained mainly from the condensed phase rather than from the gas phase. Thus, the present approach is to be contrasted with one based purely on ab initio calculations for isolated molecules. This was found to be essential for the relatively simple potential energy function (without polarization) used here to avoid increasing the time required for computations. An illustrative example is the use of the condensed-phase CN peptide bond length, which equals 1.33 Å, in contrast to the gas-phase value of 1.386 Å; the contraction is due to conjugation resulting from hydrogen bonding of the carbonyl group in solution and in crystals. The harmonic nature of the bond and angle terms causes the selected equilibrium parameters (see eq 1) to dominate the geometry. In the present model, the condensed-phase environment is of interest and the assumption is made that the most common hydrogen-bonded geometries are always appropriate. This can introduce errors for special cases, but they are expected to be rare. Nevertheless, this limitation of the potential function and its parametrization should be kept in mind.

Extrapolation of parameters optimized for small molecules to macromolecules is often difficult. The alanine dipeptide parametrization illustrate the types of problems that can occur. Experimental macromolecular data were used in determining, as well as in testing, the parameters. Specifically, the results of protein simulations were used for the adjustment of the dihedral parameters associated with the peptide backbone. This eliminated systematic deviations of the calculated ϕ and ψ dihedral angles from the experimental values, a shortcoming seen in the MM3[10] and AMBER[118] empirical potential functions. The analysis also led to a better understanding of the relationship of the alanine dipeptide map to the behavior of the peptide backbone in proteins. This will be useful in future efforts to improve and extend potential energy functions.

Evaluation of the agreement between calculated and experimental data was based on the observed deviations and their relation to the magnitudes of the fluctuations found in the simulations. If the difference between experiment and calculations was smaller than the calculated root-mean-square fluctuations, the level of agreement was deemed to be satisfactory. While more rigorous criteria may be appropriate in specific cases, the present method provides an approach for identifying limitations in the potential function and correcting them after detailed examination.

The testing of the nonbonded parameters made considerable use of crystal calculations because they provide the most detailed information on intermolecular interactions. Although other force fields[8,10,11] have used crystal data, the comparisons were mainly restricted to energy-minimization results. In the present work it was found that molecular dynamics simulations were needed for meaningful comparisons. For a number of observables the agreement between calculations and experiment is improved with molecular dynamics simulations, as compared with the minimizations that yield results corresponding to absolute zero, rather than the temperature of the experiment. For example, minimization led to a decrease in the unit cell volume of the GAL crystal by −5.3%, while the NPT molecular dynamics simulation resulted in a small increase (2.7%). The simulations also yielded better agreement with experiment in the overall rms structural deviations, the differences in dihedral angles, and the nonbonded interaction distances.

Most optimizations and comparisons of empirical potential functions parameters have concentrated on the cyclic peptides. This rather arbitrary choice appears to be a consequence of the early paper of Hall[104] that used three cyclic peptides to compare potential functions. Clearly, the cyclic peptides are useful test systems for peptide bond parameters, but the extent of the comparisons is somewhat limited because no charged groups are present. Moreover, the ϕ, ψ values in the cyclic peptides differ significantly from those that are most common in proteins; e.g., cyclic peptides CP3 and CP5 each have a ϕ, ψ conformer in the vicinity of 130°, −60°, values that are very rare in proteins. To overcome these limitations, three noncyclic tripeptides were added to the test set. Two of them have α-helical structures, and the third is a parallel β sheet. They all have terminal charged groups, and for two of them, several water molecules are present in the unit cell. The present parameter set was shown to reproduce the experimental properties of these peptides satisfactorily. Of particular interest are the charged terminal group interactions with water molecules. The largest deviations in nonbonded distances occurred for the peptide−water distances. Many of these involved a single water molecule interacting with two or more protein hydrogen-bonding sites. It was found that there is some reorientation of the charged groups coupled with rearrangements in the positions of certain waters. This suggests that alternative minima of similar energy are present in the crystal, in accord with the large temperature factors found experimentally. It would be of interest to determine how well other published force fields represent these peptides, which are not included in the optimization process.

An important point for condensed-phase simulations is the balance between protein-to-protein and protein-to-solvent nonbonded interactions. The corresponding distances observed in crystals were examined and shown to be well-described by the potential function. Of note are the average interaction distances between the hydrogen-bonded backbone N and O atoms in helices versus sheets. In crambin and BPTI, the experimental structures show that the distances in sheets are shorter than those in helices. This was reproduced in the crystal simulations of those proteins. It results from the correct balance of the nonbonded Coulomb and Lennard-Jones interactions.

Although the details of the side chain parametrization were not described in this paper, the quality of the side chain parameters may be inferred from the relatively small rms differences for the majority of the side chains in the protein crystals and the maintenance of the protein−protein and protein−water nonbonded interaction distances. Analysis of side chain-to-backbone interactions in Table 23 shows that the calculated results are within 0.1 Å of the crystal data for the majority of interaction distances. Larger discrepancies do occur. However, they are either associated with questionable experimental values (e.g., the BPTI Arg N···O distance), or they do not occur consistently for the three proteins (e.g., the amino N···O or amide N···O interaction distances). Water−side chain interactions, shown in Table 24, are reasonably accurate, although many of the differences in the interaction distances are greater than 0.1 Å. Again, the larger differences (e.g., amino N···OH2 or amide N···OH2 interaction) are not consistent for the two proteins studied; this indicates that the parameters are not unfavorably biasing the results. As with the protein backbone, it is evident that the interactions of the protein side chains are satisfactorily treated by the parameter set.

The present all-atom parameter set, designed for the simulation of peptides and proteins in the condensed phase with an effective dielectric constant equal to unity, has a broad range of applicability. It should be noted that the CHARMM22 parametrization does not replace the CHARMM19 polar hydrogen parameter set that has been widely used in the past and continues to used in problems where speed is essential and the highest accuracy is not required. Recent examples include protein-folding studies with explicit[119] and implicit solvent.[120] Beyond studies of protein structure and dynamics, interactions with small molecules and other macromolecules, such as nucleic acids and lipids, are now accessible. Calculations on ribonuclease T1 with the inhibitors 2′-guanosine monophosphate and guanylyl-2′,5′-guanosine yield good agreement with experimental crystal structures for the protein-to-nucleic acid interactions (A. D. MacKerell, Jr., work in progress). The present parameters have also been used in combination with the lipid parameters for the study of gramicidin in a phospholipid bilayer.[121] The parametrization approach outlined in the present paper can be extended to small molecules as needed. Examples are parameters for NAD and pyrophosphate groups that have been published recently.[122] While the approach used in the optimization procedure may not be necessary for certain applications, for others, high-quality parameters are essential for reliable results. This is true in particular for free energy simulations.

It is hoped that the formal publication of this already widely used parameter set for the CHARMM22 all-atom protein energy function provides a useful service to the scientific community.

Acknowledgment. We thank John Kuriyan for the equilibrated crambin crystal coordinates and B. Kim Andrews and B. Montgomery Pettitt for results from their MbCO solution simulation. This work was supported in part by grants from the National Science Foundation and the National Institutes of Health to M.K. and a grant from the Mildred Mindell Cancer Foundation to A.D.M.

Supporting Information Available: The present addresses of the authors whose address is not given in the published paper, one figure and eight tables, and an Appendix of two figures and seven tables that lists the parameters developed in the present study (41 pages). Included are results from the *N*-methylacetamide and alanine dipeptide crystal calculations, including the effects of truncation distance and details of the nonbond interactions, as well as the effects of truncation distance on minimizations, calculated dihedral angles, and nonbonded interactions for both the tripeptides and the cyclic peptides. The Appendix contains a complete list of the bond, bond angle, dihedral angle, improper, Lennard-Jones parameters, and partial atomic charges used in the CHARMM22 force field.[123] Ordering and accessing information is given on any current masthead page.

References and Notes

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

(2) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins, A Theoretical Perspective Dynamics, Structure, and Thermodynamics*; John Wiley and Sons: New York, 1988; Vol. LXXI.

(3) McCammon, J. A.; Harvey, S. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: 1987.

(4) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(5) Reiher, W. E., III. Ph.D. Thesis, Harvard University, 1985.

(6) Weiner, S. J.; Singh, U. C.; O'Donnell, T. J.; Kollman, P. A. *J. Am. Chem. Soc.* **1984**, *106*, 6243.

(7) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.

(8) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(9) (a) Feng, M.-H.; Philippopoulos, M.; MacKerell, A. D., Jr.; Lim, C. *J. Am. Chem. Soc.* **1996**, *118*, 11265. (b) Schmidt, J. M.; Brüschweiler, R.; Ernst, R. R.; Dunbrack, R. L., Jr.; Joseph, D.; Karplus, M. *J. Am. Chem. Soc.* **1993**, *115*, 8747−8756.

(10) Lii, J.-L.; Allinger, N. L. *J. Comput. Chem.* **1991**, *12*, 186.

(11) Hagler, A. T.; Maple, J. R.; Thacher, T. S.; Fitzgerald, G. B.; Dinur, U. In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: Leiden, 1989; p 149.

(12) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. *J. Phys. Chem.* **1990**, *94*, 8897.

(13) MacKerell, A. D., Jr.; Wiórkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.

(14) Schlenkrich, M.; Brickmann, J.; MacKerell, A. D., Jr.; Karplus, M. *Biological Membranes: A Molecular Perspective from Computation and Experiment*; Birkäuser: Boston, MA, 1996; p 31.

(15) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902.

(16) van Gunsteren, W. F.; Berendsen, H. J. C. Groningen molecular simulation (GROMOS) library manual; Biomos, Nijenborgh, 4, 9747 AG Groningen, The Netherlands, 1987.

(17) (a) Jorgensen, W. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768. (b) Burley, S. K.; Petsko, G. A. *Adv. Protein Chem.* **1988**, *39*, 125.

(18) The CHARMM program can be obtained by not-for-profit institutions under a license agreement by writing to the CHARMM Development Program, Department of Chemistry & Chemical Biology, Harvard University, Cambridge, MA 02138 (email: marci@tammy.harvard.edu).

(19) Pettitt, B. M.; Karplus, M. *J. Am. Chem. Soc.* **1985**, *107*, 1166.

(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(21) Smith, J. C.; Karplus, M. *J. Am. Chem. Soc.* **1992**, *114*, 801.

(22) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday: A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. *Acta Crystallogr.* **1979**, *B35*, 2331.

(23) Florián, J.; Johnson, B. G. *J. Phys. Chem.* **1994**, *98*, 3681.

(24) Dunbrack, R. L., Jr.; Karplus, M. *Nat. Struct. Biol.* **1994**, *1*, 334.

(25) Momany, F. A.; Klimkowski, V. J.; Schäfer, L. *J. Comput. Chem.* **1990**, *11*, 654.

(26) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, *J. Phys. Chem.* **1987**, *91*, 6269.
(27) (a) Hirato, E.; Sugisali, R.; Nielsen, C. J.; Sórensen, G. O. *J. Mol. Spectrosc.* **1974**, *53*, 62. (b) Dyke, T. R.; Mack, K. M.; Muenter, J. S. *J. Chem. Phys.* **1977**, *66*, 498.
(28) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810.
(29) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 520.
(30) MacKerell, A. D., Jr.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 10559.
(31) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.
(32) Moet-Ner, M. *J. Am. Chem. Soc.* **1988**, *110*, 3858.
(33) Moet-Ner, M. *J. Am. Chem. Soc.* **1988**, *110*, 3071.
(34) Moet-Ner, M. *J. Am. Chem. Soc.* **1988**, *110*, 3075.
(35) Moet-Ner, M. *J. Am. Chem. Soc.* **1988**, *110*, 3854.
(36) Gao, J. Ph.D. Thesis, Department of Chemistry, Purdue University, 1987.
(37) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276.
(38) Stote, R. H.; States, D. J.; Karplus, M. *J. Chim. Phys.* **1991**, *88*, 2419.
(39) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *Gaussian*; Gaussian Inc.: Pittsburgh, PA, 1992.
(40) Jorgensen, W. L. *BOSS*; Yale University: New Haven, CT, 1994.
(41) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(42) Brünger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148.
(43) Field, M. J.; Karplus, M. Harvard University, Cambridge, MA. Unpublished results, 1992.
(44) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(45) Hendrickson, W. A.; Teeter, M. M. *Nature (London)* **1981**, *290*, 107.
(46) Wlodawer, A.; Walter, J.; Huber, R.; Sjölin, L. *J. Mol. Biol.* **1984**, *180*, 301.
(47) Kuriyan, J.; Wilz, S.; Karplus, M.; Petsko, G. *J. Mol. Biol.* **1986**, *192*, 227.
(48) Kuriyan, J.; Weis, W. I. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2773.
(49) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97.
(50) Burling, F. T.; Weis, W. I.; Flaherty, K. M.; Brünger, A. T. *Science* **1996**, *271*, 72.
(51) Chaturvedi, S.; Go, K.; Parthasarathy, R. *Biopolymers* **1991**, *31*, 397.
(52) Hempel, A.; Camerman, N.; Camerman, A. *Biopolymers* **1991**, *31*, 187.
(53) Hossain, M. B.; van der Helm, D. *J. Am. Chem. Soc.* **1978**, *100*, 5191.
(54) Karle, I. L.; Gibson, J. W.; Karle, J. *J. Am. Chem. Soc.* **1970**, *92*, 3755.
(55) Kostansek, E. C.; Thiessen, W. E.; Schomburg, D.; Lipscomb, W. N. *J. Am. Chem. Soc.* **1979**, *101*, 5811.
(56) Karle, I. L. *J. Am. Chem. Soc.* **1978**, *100*, 1286.
(57) Kopple, K. D.; Wang, Y.-S.; Cheng, A. G.; Bhandary, K. K. *J. Am. Chem. Soc.* **1988**, *110*, 4168.
(58) Grant, J. A.; Williams, R. L.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 929.
(59) Kollman, P. A.; Dill, K. A. *J. Biomol. Struct. Dyn.* **1991**, *8*, 1103.
(60) Guo, H.; Karplus, M. *J. Phys. Chem.* **1992**, *96*, 7273.
(61) Guo, H.; Karplus, M. *J. Phys. Chem.* **1994**, *98*, 7104.
(62) Katz, J.; Post, B. *Acta Crystallogr.* **1960**, *13*, 624.
(63) Hamzaoui, F.; Baert, F. *Acta Crystallogr.* **1994**, *C50*, 757−579.
(64) Benedetti, E. In *Peptides: Proceedings of the 5th American Peptide Symposium*; John Wiley & Sons Inc.: New York, 1979; p 257.
(65) Kitano, M.; Fukuyama, T.; Kuchitsu, B. *Bull. Chem. Soc. Jpn.* **1973**, *46*, 384.
(66) Jorgensen, W. L.; Gao, J. *J. Am. Chem. Soc.* **1988**, *110*, 4212.
(67) Evanseck, J. D.; Karplus, M. Unpublished work, 1994.
(68) Sulzbach, H. M.; Schleyer, P.v. R.; Schaefer, H. F., III. *J. Am. Chem. Soc.* **1995**, *117*, 2632.
(69) Fischer, S.; Dunbrack, R. L., Jr.; Karplus, M. *J. Am. Chem. Soc.* **1994**, *116*, 11931.
(70) Gould, I. R.; Cornell, W. D.; Hiller, I. H. *J. Am. Chem. Soc.* **1994**, *116*, 9250.
(71) (a) Jones, R. *J. Mol. Spectrosc.* **1963**, *11*, 411. (b) Fillaux, F.; DeLoze, C. J. *J. Chim. Phys.* **1976**, *73*, 1004.
(72) Venkatachalapathi, Y.; Mierke, D.; Taulane, J.; Goodman, M. *Biopolymers* **1987**, *26*, 763.
(73) Kohlrausch, K.; Seka, R. *Z. Phys. Chem.* **1939**, *B42*, l-480.
(74) Sugawara, Y.; Hirakawa, A.; Tsuboi, M.; Kato, S.; Morokuma, K. *J. Mol. Spectrosc.* **1986**, *115*, 21.
(75) (a) Mirkin, N.; Krimm, S. *J. Mol. Struct.* **1991**, *242*, 143. (b) Chen, X. G.; Schweitzer-Stenner, R.; Asher, S. A.; Mirkin, N. G.; Krimm, S. *J. Phys. Chem.* **1995**, *99*, 3074.
(76) Miazawa, T.; Shimanouchi, T.; Mizushima, S. *J. Chem. Phys.* **1958**, *29*, 611.
(77) Itoh, K.; Shimanouchi, T. *Biopolymers* **1967**, *5*, 921.
(78) Drakenberg, T.; Forsén, S. *Chem. Commun.* **1971**, 1404.
(79) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(80) Vásquez, M.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 1043.
(81) Böhm, H.-J.; Brode, S. *J. Am. Chem. Soc.* **1991**, *113*, 7129.
(82) Frey, R. F.; Coffin, J.; Newton, S. Q.; Ramek, M.; Cheng, V. K. W.; Momany, F. A.; Schäfer, L. *J. Am. Chem. Soc.* **1992**, *114*, 5369.
(83) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C., III; Pople, J. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1989**, *16*, 311.
(84) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L. I.; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989.
(85) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908.
(86) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(87) Deng, Z.; Polavarapu, P. L.; Ford, S. J.; Hecht, L.; Barron, L. D.; Ewig, C. S.; Jalkanen, K. *J. Phys. Chem.* **1996**, *100*, 2025.
(88) Brooks, C. L., III; Case, D. A. *Chem. Rev.* **1993**, *93*, 2487.
(89) Dudek, M. J.; Ponder, J. W. *J. Comput. Chem.* **1995**, *16*, 791.
(90) Gelin, B.; Karplus, M. *J. Am. Chem. Soc.* **1975**, *97*, 6996.
(91) Kang, Y. K.; No, K. T.; Scheraga, H. A. *J. Phys. Chem.* **1996**, *100*, 15588.
(92) Maxwell, D. S.; Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **1995**, *16* 984.
(93) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
(94) Baudry, J.; Smith, J. C. *J. Mol. Struct.: THEOCHEM* **1994**, *308*, 103.
(95) Pangali, C.; Rao, M.; Berne, B. J. *Mol. Phys.* **1980**, *60*, 661.
(96) Riddick. J. A.; Bunger, W. B.; Sakano, T. K. *Organic Solvents: Physical Properties and Methods of Purification*; John Wiley & Sons: New York, 1986.
(97) Lemire, R. H.; Sears, P. G. *Top. Curr. Chem.* **1978**, *78*, 45.
(98) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 1489.
(99) Kaminski, G.; Jorgensen, W. L. *J. Phys. Chem.* **1996**, *100*, 18010.
(100) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(101) Harada, Y.; Iitaka, Y. *Acta Crystallogr.* **1974**, *B30*, 1452.
(102) MacArthur, M. W.; Thorton, J. M. *J. Mol. Biol.* **1996**, *264*, 1180.
(103) Stone, A. J.; Price, S. L. *J. Phys. Chem.* **1988**, *92*, 2: 3325.
(104) Hall, D.; Pavitt, N. *J. Comput. Chem.* **1984**, *5*, 441.
(105) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.
(106) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631.
(107) van Gunsteren, W. F.; Berendsen, H. J. C.; Hermans, J.; Hol, W. G. J.; Postma, J. P. M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4315.
(108) van Gunsteren, W. F.; Karplus, M. *Biochemistry* **1982**, *21*, 2259.
(109) Tilton, R. F., Jr.; Singh, U. C.; Kuntz, I. R., Jr.; Kollman, P. A. *J. Mol. Biol.* **1988**, *199*, 195.
(110) Longcharich, R. J.; Brooks, B. R. *J. Mol. Biol.* **1990**, *215*, 439.
(111) Polygen All-atom Protein Parameter Set; Polygen Corp.: Waltham, MA, 1988.
(112) Levy, R. M.; Sheridan, R. P.; Keepers, J. W.; Dubey, G. S.; Swaminathan, S.; Karplus, M. *Biophys. J.* **1985**, *48*, 509.
(113) Quillin, M. L.; Arduini, R. M.; Olson, J. S.; Phillips, G. N., Jr. *Proteins*, **1990**, *7*, 358.
(114) Teeter, M. M.; Roe, S. M.; Heo, N. H. *J. Mol. Biol.* **1993**, *230*, 292.
(115) Kuriyan, J.; Petsko, G. A.; Levy, R. M.; Karplus, M. *J. Mol. Biol.* **1986**, *190*, 227.
(116) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318.
(117) Phillips, G. N., Jr. *Biophys. J.* **1990**, *57*, 381.
(118) Whitlow, M.; Teeter, M. M. *J. Am. Chem. Soc.* **1986**, *108*, 7163.
(119) Boczko, E. M.; Brooks, C. L., III. *Science* **1995**, *269*, 393.
(120) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928.
(121) Woolf, T. B.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11631.
(122) Pavelites, J. J.; Gao, J.; Bash, P. A.; MacKerell, A. D., Jr. *J. Comput. Chem.* **1997** *18*, 221.
(123) The parameters may be obtained also from A.D.M.'s Web page at www.pharmacy.ab.umd.edu/∼alex.