FORCE FIELDS FOR PROTEIN SIMULATIONS

By JAY W. PONDER* AND DAVID A. CASE†

*Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, and †Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037

I. Introduction

Molecular dynamics and Monte Carlo simulations of proteins, which began about 25 years ago, are by now widely used as tools to investigate their structure and dynamics under a wide variety of conditions, ranging from studies of ligand binding and enzyme-reaction mechanisms to problems of denaturation and protein re-folding. Fundamental to such simulations is the representation of the energy of the protein as a function of its atomic coordinates. The states expected to be populated at thermal equilibrium are the low-energy regions of this (potential) energy function, and forces on individual atoms are related to the gradient of this function, which is why such functions are also commonly referred to as "force fields."

Atomistic simulations of the properties of proteins commonly consider an average over the much-faster electronic motions, so that the energy surface on which the atoms move is the Born-Oppenheimer ground-state

ADVANCES IN PROTEIN CHEMISTRY, Vol. 66

0065-3233/03 $35.00

energy. It is not yet feasible to calculate directly such surfaces for macromolecules with high accuracy by means of quantum chemistry electronic structure calculations, so most practical simulations use a set of simple classical functions to represent the energy, adjusting a large number of parameters to optimize agreement with experimental data and with quantum calculations on smaller molecules. The design and parameterization of force fields for use in protein simulations is a complex task, involving many decisions concerning which data to emphasize in the fits, expectations of transferability to areas outside the "fit set," and computational efficiency. Our goal here is to provide an overview of the thinking that goes into the choices that must be made, a general description of the sorts of force fields that are most commonly used at present, and an indication of the directions of current research that may yield better functions in the near future. Although we hope the discussion can be read with profit by non-specialists, we will assume a general familiarity with molecular mechanics and dynamics and with their applications to proteins. Simulation methods for proteins are by now well-enough established that there are many good textbooks and monographs that cover the basics (Burkert and Allinger, 1982; Harvey and McCammon, 1987; Leach, 2001; Becker *et al.*, 2001; Cramer, 2002; Schlick, 2002).

It is impossible for a short review like this one to be comprehensive, and we have chosen to concentrate on a few areas, necessarily at the expense of others. We will restrict our discussion to simulations of proteins in water, leaving aside the interesting questions of how to deal with the large variety of small molecules that interact with proteins or with non-aqueous solvents. With perhaps less justification, we have chosen to focus on a few force fields that are very widely used, and even here it is not possible to give any real account of the differences among potentials. With regard to newer developments, we have chosen to concentrate on two main areas: the use of continuum methods to model the electrostatic effects of hydration, and the introduction of polarizability to model the electronic response to changes in the environment. These are not the only recent developments worthy of note, but we do expect them to be keys areas of research for the next few years.

There is much current development, and other good reviews of force fields (MacKerell, 2001; Hünenberger and van Gunsteren, 1997; Damm and Halgren, 2001). Hence, after a brief survey of current models, mostly generated during the 1990s, our review will focus on the general directions the field is taking in developing new models.

The most commonly used protein force fields incorporate a relatively simple (sometimes called "Class I") potential energy function:

$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi[\cos(n\phi + \delta) + 1] + \sum_{\substack{nonbond \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}\right] \quad (1)$$

The first three summations are over bonds (1-2 interactions), angles (1-3 interactions), and torsions (1-4 interactions). These are illustrated in Fig. 1. The torsion term can also include so-called "improper" torsions, where the four atoms defining the angle are not all connected by covalent bonds; such terms serve primarily to enforce planarity around sp^2 central atoms, and other functional forms can be used for the same purpose. The final sum (over pairs of atoms i and j) excludes 1-2 and 1-3 interactions and often uses separate parameters for 1-4 interactions as compared with those used for atoms separated by more than three covalent bonds. It describes electrostatics that use partial charges q_i on each atom that interact via Coulomb's law. The combination of dispersion and exchange repulsion forces are represented by a Lennard-Jones 6-12 potential; this is often called the "van der Waals" term.

Equation 1 is about the simplest potential energy function that can reproduce the basic features of protein energy landscapes at an atomic level of detail, and it has proved to give insight into a remarkably broad range of properties. The combination of a potential energy function (as in equation 1) and all the parameters that go into it (k_b, b_0, k_θ, θ_0, *etc.*) constitutes a "force field." In practice, there is a close connection between the force fields and the computer codes that implement them, although this is becoming less true now than it used to be.

In section II, we review some of the history and performance of widely used protein force fields based on equation 1, or closely related equations. Sections III and IV outline some promising developments that go beyond this, primarily by altering the way electrostatics interactions are treated.

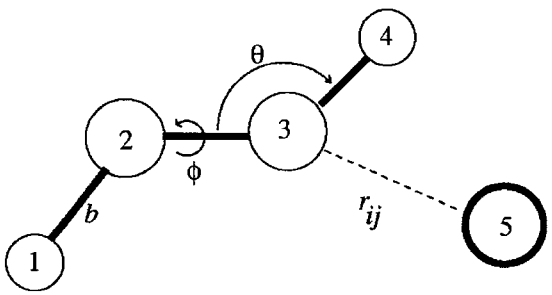


FIG. 1. Schematic view of force field interactions. Covalent bonds are indicated by heavy solid lines, nonbonded interactions by a light, dashed line.

Section III considers the use of atomic multipoles and off-center charge distributions, as well as attempts to incorporate electronic polarizability. Rather than making the electrostatic model more complex, Section IV considers continuum solvent models that simplify the problem by accounting for all solvent (water plus mobile counterions) effects in an averaged fashion.

II. Protein Force Fields, 1980 to the Present

There have been a large number of force fields used over the years for simulations of proteins, and it is not our intent to try to provide any sort of comprehensive review. Sections A through C provide an overview of three sets of parameters that have been widely used. The discussion emphasizes the decisions that went into their development. A very short summary of other important protein force fields is presented in section D, and a brief comparison of the results from different force fields is given in section E. Many of the more recent developments in protein force fields address electronic polarizability or the effects of solvation; these are discussed in sections III and IV, respectively.

Our historical discussion of protein force fields begins roughly in 1980, when molecular dynamics and Monte Carlo simulations of proteins were in their infancy. It is important to note, however, that the developments outlined below did not arise *sui generis* but rather were built on earlier force field developments in organic chemistry. Of particular importance were the ECEPP potentials from Scheraga and co-workers (Momany *et al.*, 1975; Némethy *et al.*, 1983) and the consistent force field (CFF) developments from the Lifson group (Lifson and Warshel, 1969; Hagler *et al.*, 1974; Hagler and Lifson, 1974; Niketic and Rasmussen, 1977). These provided important starting points for the work described here, as did efforts to develop potential energy functions in the general area of organic chemistry (Allinger, 1976; Burkert and Allinger, 1982).

We hope that our introduction below to three popular force fields will help to prepare readers to appreciate a literature that is much broader than we can cover here. Other good reviews exist (MacKerell, 2001; Hüenberger and van Gunsteren, 1997) that complement our presentation.

A. *The Amber Force Fields*

By the first half of the 1980s, enough experience had accumulated with earlier parameterizations for several groups to begin fairly systematic projects to develop a new generation of force fields. The earliest of these

efforts were still done at a time when the limited power of computers made it attractive to not include all hydrogen atoms as explicit force centers. The importance of hydrogen bonding, however, led many investigators to adopt a compromise whereby polar hydrogens were explicitly represented but hydrogens bonded to carbon were combined into united atoms. A widely used force field at this level was developed in 1984 in the Kollman group (Weiner *et al.*, 1984) and incorporated into the Amber molecular mechanics package, which was at an early stage of development as well (Weiner and Kollman, 1981). The key ideas in this initial work were to be used repeatedly in later efforts by this group. Charges were derived from quantum chemistry calculations at the Hartree-Fock STO-3G level, via fitting of partial atomic charges to the quantum electrostatic potential; these are generally called ESP (for electrostatic potential) charges. The van der Waals terms were adapted from fits to amide crystal data by Lifson's group (Hagler *et al.*, 1974; Hagler and Lifson, 1974) and from liquid-state simulations pioneered by Jorgensen (Jorgensen, 1981). Force constants and idealized bond lengths and angles were taken from crystal structures and adapted to match normal mode frequencies for a number of peptide fragments. Finally, torsion force constants were adjusted to match torsional barriers extracted from experiment or from quantum chemistry calculations. Since it is only the total potential energy, as a function of torsion angle, that needs to agree with the target values, and since (in this potential energy function) these barriers have significant electrostatic and van der Waals interactions between the end atoms (the so-called '1-4' interactions), the k_ϕ values are closely coupled to the nonbonded potentials used and are hardly transferable from one force field to another.

Three problems with this "polar hydrogen only" approach, along with improvements in the speed of available computers, led many groups to move to an all-atom approach. First, aromatic rings such as benzene have a significant quadrupolar charge distribution, with an effective positive charge near the hydrogens and an effective negative charge nearer to the middle of the ring. This effect can be crucial in determining the ways in which aromatic side chains in proteins interact with other groups. For example, "T-shaped" geometries between rings are stabilized relative to "stacked" geometries that optimize van der Waals interactions (Williams, 1991, 2001). Also important are π-cation interactions, where positive groups are found directly above the centers of aromatic rings (Dougherty, 1995; Zacharias and Dougherty, 2002). Second, the forces that affect the pseudorotation between conformations, or "pucker," of five-member aliphatic rings (Tomimoto and Go, 1995) are difficult to describe when only the heavy atoms are available as force centers. This affects only

proline residues in proteins, but analogous problems involving ribose and deoxyribose in nucleic acids (whose force fields were being parameterized at the same time) led momentum toward all-atom force fields. Finally, it is difficult with united atom models to make comparisons between computed and observed vibrational frequencies. An extension of the 1984 force field to an all-atom model was published in 1986, as a collaboration between the Kollman and Case groups (Weiner *et al.*, 1986). Both the 1984 and 1986 parameter sets were primarily developed based on experience with gas phase simulations.

The continued increase in the speed of computers led the Kollman group to decide in the early 1990s that a new round of force-field development was warranted; this came to be known as the "Cornell *et al.*" or *ff94* force field (Cornell *et al.*, 1995). In addition to improvements in the parameters, a more serious attempt was made to explicitly describe the algorithm by which the parameters were derived, so that consistent extensions could be made to molecules other than proteins (Fox and Kollman, 1998). This goal was not really achieved until the development almost a decade later of the *antechamber* program that completely automates all of the steps in the creation of an Amber-like force field for an arbitrary molecule or fragment.

A key motivation for this development was a desire to produce potentials suitable for condensed phase simulations, since the earlier work had concentrated in large part on gas phase behavior. In particular, the ways in which the OPLS potentials (discussed below) had been parameterized to reproduce the densities and heats of vaporization of neat organic liquids was very influential, along with recognition of the importance of having a balanced description of solute-solvent versus solvent-solvent interactions. A second point arose from the ability to use larger basis sets and fragment sizes to determine atomic charges that mimic the electrostatic potentials outside the molecule found from quantum mechanical calculations. Earlier work had established that fitting charges to the potentials at the Hartree-Fock 6-31G* level tended to overestimate bond-dipoles (compared with observed gas phase values) by amounts comparable to that in empirical water models such as SPC/E or TIP3P; such "overpolarization" is an expected consequence of electronic polarization in liquids. Hence, the use of fitted charges at the HF/6-31G* level appeared to offer a general procedure for quickly developing charges for all 20 amino acids in a way that would be roughly consistent with the water models that were expected to be used. Tests of this idea, with liquid-state simulations of amides and simple hydrocarbons, gave encouraging results.

The actual implementation of this scheme for developing charges had to deal with two complications, which continue to plague force field

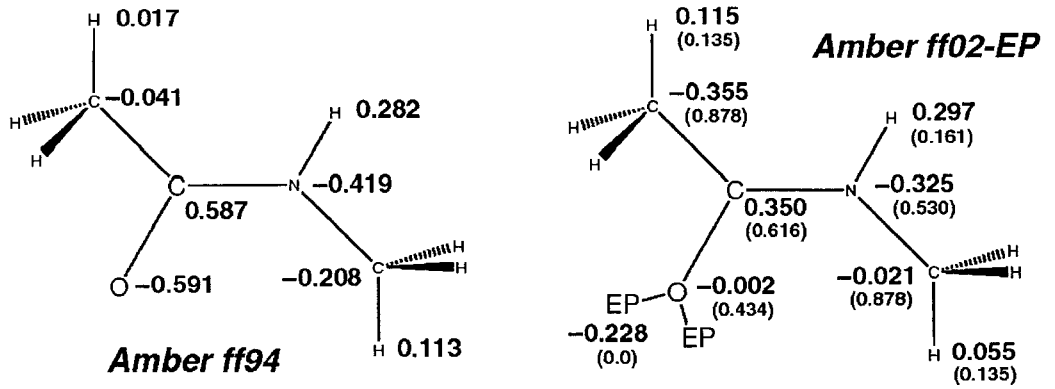


FIG. 2. Charge models for the Amber potentials. (Left) HF-6-31G*RESP charges in the style of *Amber ff94*. (Right) polarizable, extra-point charge model in the style of *Amber ff02-EP* with the atomic polarizabilities in $Å^3$ given in parentheses.

developers to the present day. First, the effective charges of the more-buried atoms are often underdetermined, so that charges for atoms in similar environments in different molecules might vary significantly. In effect, there are many combinations of atomic charges that will fit the electrostatic potential almost equally well. There are a variety of ways to overcome this problem, often involving statistical techniques based on singular-value decomposition, but Bayly *et al.* (Bayly *et al.*, 1993; Cornell *et al.*, 1993) chose to use a hyperbolic restraint term to limit the absolute magnitude of charges on non-hydrogen atoms. This is called RESP (for restrained electrostatic potential fit) and weakly favors solutions with smaller charges for buried atoms, yielding fairly consistent charge sets with little degradation in the quality of the fit to the electrostatic potential outside the molecule. As an example, the left-hand side of Fig. 2 shows the charges determined in this way for *N*-methylacetamide, modeling the peptide bond; the right-hand side of this figure is for a more complex electrostatic model, described below.

A second and more fundamental problem with the RESP procedure is that the resulting charges depend on molecular conformation, often in significant ways. This is a manifestation of electronic polarizability, which can only be described in a very averaged way if fixed atomic charges are to be used. Any real solution to this problem must involve a more complex model, such as those described in section III, below. The compromise chosen for the *ff94* force field was to fit charges simultaneously to several conformations, in the hopes of achieving optimal averaged behavior.

Once the charges and the "stiff" internal parameters for bonds and angles were available (the latter estimated in the same way as outlined above), the Lennard-Jones parameters could be established primarily by

reference to densities and heats of vaporization in liquid-state simulations. Only a small number of sets of 6-12 parameters were necessary to achieve reasonable agreement with experiment. A key expansion from earlier work was the notion that parameters for hydrogens should depend in an important way on the electronegativity of the atoms they are bonded to (Gough *et al.*, 1992; Veenstra *et al.*, 1992).

As with many other force-field projects, the final parameters to be fit were the "soft" torsional potentials about single bonds. It makes some sense to address these after the charges and Lennard-Jones parameters have been developed, since the energy profile for rotation about torsion angles depends importantly on the nonbonded interactions between the moving groups at the ends, as well as on whatever intrinsic torsional potential is assigned. The question of how best to partition torsional barriers into "bonded" versus "nonbonded" interactions is a thorny one, and many developers of force fields have adopted a strictly empirical approach, fitting k_ϕ, n, and δ so that the total profile (including the nonbonded terms) matches some target extracted from quantum mechanics or from experiment.

A key set of torsional parameters are those for the ϕ and ψ backbone angles, since these affect every amino acid residue and heavily influence the relative energies of helices, sheets, and turns in proteins. The *ff94* parameters were fit to representative points on the dipeptide maps for glycine and alanine, computed at the MP2 level with a triple-ζ + polarization (TZP) basis set. This is not an unreasonable choice for a target function, but it has a number of intrinsic difficulties. First, the α-helix region near $\phi,\psi = -60,-40$ is not a minimum for a gas-phase dipeptide, so fitting just a representative point (as was done) can lead to errors in the surface as a whole, compared to the full MP2/TZP surface. More importantly, the use of a gas phase dipeptide model as a target ignores both the non-local electronic structure contributions that would be seen in larger fragments (Beachy *et al.*, 1997) and the polarization effects inherent in a condensed phase environment (Ösapay *et al.*, 1996). Some account of the longer-range effects was provided in subsequent parameterizations, referred to as *ff96* (Kollman *et al.*, 1997) and *ff99* (Wang *et al.*, 2000), in which the ϕ and ψ potentials were fit to tetrapeptide as well as dipeptide quantum mechanical conformational energies. These later fits provided potential surfaces that were significantly different from those in *ff94*, but it was hard to tell if physical realism was really being improved.

In recent years it has become computationally feasible to test protein potentials (and especially their backbone torsion angle behavior) by carrying out converged or nearly converged simulations on short peptides

and comparing the resulting conformational populations to those derived from experiment (Damm and van Gunsteren, 2000; Mitsutake *et al.*, 2001; García and Sanbonmatsu, 2001). The experimental estimates, obtained mainly from circular dichroism or from NMR, are often only qualitative, but this can be enough to identify obvious errors in computed ensembles. For example, the *ff94* parameters appear to over-stabilize helical peptide conformers in many if not all instances. Computed melting temperatures for polyalanine helices are too high (García and Sanbonmatsu, 2002), and helical conformers can predominate in simulations of sequences that experimentally form other structures, such as β-hairpins. At least two modifications of the *ff94* ϕ and ψ potentials have been proposed and tested on large-scale peptide simulations (García and Sanbonmatsu, 2002; Simmerling *et al.*, 2002). It will be of interest to see how these ideas develop as a new generation of long time-scale peptide simulations becomes feasible.

B. The CHARMM Force Fields

As with Amber, the CHARMM program (Chemistry at HARvard using Molecular Mechanics) (Brooks *et al.*, 1983) was originally developed in the early 1980s and initially used an extended atom force field with no explicit hydrogens. By 1985, this had been replaced by the *CHARMM19* parameters, in which hydrogen atoms bonded to nitrogen and oxygen are explicitly represented, while hydrogens bonded to carbon or sulfur are treated as part of extended atoms (Reiher, 1985; Neria *et al.*, 1996). Key to the parameterization of this model were fits to quantum calculations at the HF/6-31G level of hydrogen bonded complexes between water and the H-bond donors or acceptors of the amino acids or fragments. This involves a series of supermolecular calculations of the model compound, such as formamide or *N*-methylacetamide and a single water molecule at each of several interaction sites. Before making the fits, the interaction energies are scaled by a factor of 1.16, which is the ratio of the water dimerization energy predicted by the TIP3P model to that predicted at the HF/6-31G level. As in the Amber parameterizations described above, the goal here was to obtain a balanced interaction between solute-water and water-water energies when the latter are represented by TIP3P. For peptides, it was found that fitting the peptide-water interactions in this way led to peptide-peptide hydrogen bonds that were also larger than HF/6-31G values by a factor very close to 1.16; in other cases, explicit fitting to solute-solute hydrogen bonded dimers may be needed for parameter generation (MacKerell, 2001).

As with the contemporaneous Amber 1984 united-atom parameterization, the *CHARMM19* values were developed and tested primarily on gas-phase simulations. However, the *CHARMM19* potential seems to do well (perhaps fortuitously) in solvated simulations and continues to be used for peptide and protein simulations; this is in contrast to the 1984 Amber force field, which is no longer widely used. In addition, the *CHARMM19* values have often been used in conjunction with a distance-dependent dielectric constant as a rough continuum solvation model.

In the early 1990s, the CHARMM development group also recognized the need to refine parameters more explicitly pointed to obtaining a good balance of interaction energies in explicit solvent simulations. The resulting *CHARMM22* protein force field was first included in the corresponding version of CHARMM, released in 1992, and was fully described a few years later (MacKerell *et al.*, 1998; MacKerell, 2001). The key approach from *CHARMM19* was carried over by deriving charge models primarily from fits to solute-water dimer energetics (now calculated at the HF/6-31G* level). In addition to fitting the dimer interaction energies, charges for model compounds were adjusted to obtain dipole moments somewhat larger than experimental or *ab initio* values. This has the same goal as the RESP procedure described above: bonds are expected to be more polarized in condensed phases than in the gas phase. The use of empirical charges that yield enhanced dipoles both reflects this behavior and allows a reasonably balanced set of interactions with the TIP3P water model, which has a similarly enhanced dipole moment.

Once the charges were determined by these dimer studies, the Lennard-Jones parameters were refined to reproduce densities and heats of vaporization of liquids as well as unit cell parameters and heats of sublimation for crystals. As with the Amber parameterization, generally only small adjustments from earlier values were required to fit the empirical data. Nevertheless, because of the steep dependence of these forces, such adjustments may be crucial for a well-balanced and successful set of parameters.

As with the Amber *ff94* force field, the torsional parameters were finally adjusted to target data derived from vibrational spectra and from *ab initio* calculations. The torsional potentials for the ϕ and ψ torsions were initially fit to HF/6-31+G* calculations on an analog of the alanine dipeptide in which the terminal methyl groups are replaced by hydrogen. These were then refined in an iterative procedure to improve the agreement with experiment of the backbone angles in simulations of myoglobin. In principle at least, this latter adjustment provides a way of correcting the *ab initio* dipeptide energy map for effects caused by the protein environment.

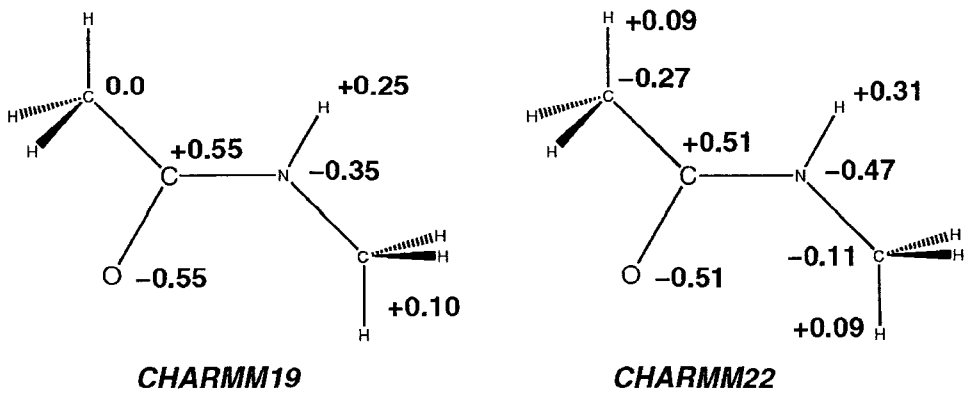


FIG. 3. Charge models for *N*-methylacetamide for two of the CHARMM force field parameterizations. In *CHARMM19*, the methyl groups are treated as united atoms, so that there is no breakdown into separate C and H charges.

As with the Amber parameterization, the question of how best to obtain good backbone torsional potentials is a vexing one, and studies are continuing, both at the dipeptide level and with solvated simulations of oligopeptides. Most recently, an extensive reworking of the nucleic acid parameters has resulted in the *CHARMM27* force field (Foloppe and MacKerell, 2000). However, the *CHARMM27* protein parameters are essentially identical to those from the *CHARMM22* force field.

One feature of the CHARMM parameterizations is the enforcement of *neutral groups*, which are small sets of contiguous atoms whose atomic charges are constrained to sum to zero. This is illustrated in Fig. 3, which shows partial atomic charges for *N*-methylacetamide in the *CHARMM19* and *CHARMM22* force fields. For example, charges for the C and O atoms of the peptide group form a small neutral group. These groups can be useful when truncating long-range electrostatic interactions: if an entire group is either included or ignored (or scaled by a smoothing factor), then there is never any splitting of dipoles. Ignoring charged side chains, each atom would then feel the electrostatic effects of a net neutral environment. The same behavior occurs with solvent molecules, if the interactions of a given water molecule are always treated as a group. Although it was long deemed plausible that such a group-based truncation scheme would yield better results than an atom-based scheme (wherein some members of a group might interact with a given atom, while others would not), this is probably not the case for most biomolecular simulations in water (Steinbach and Brooks, 1994; Darden *et al.*, 1998; Darden, 2001; Mark and Nilsson, 2002). In any event, such considerations are now much less important than in earlier times, since many current simulations use Ewald or fast multipole schemes to handle long-range electrostatics, where nothing is gained by having small neutral groups.

Comparison with the corresponding Amber charges from *ff94* (Fig. 2) shows that the peptide carbonyl group is somewhat less polar in *CHARMM22* than in Amber, whereas the opposite is true for the NH dipole. Still, the similarities in the charge models for these force fields are more striking than the differences.

C. *The OPLS Force Fields*

A third main development in the early 1980s (already alluded to above) involved potentials developed by Jorgensen and co-workers to simulate liquid state properties, initially for water and for more than 40 organic liquids. These were called OPLS (Optimized Potentials for Liquid Simulations) and placed a strong emphasis on deriving nonbonded interactions by comparison to liquid-state thermodynamics (Jorgensen, 1998). Indeed, the earliest applications of OPLS potentials were to rigid-molecule Monte Carlo simulations of the structure and thermodynamics of liquid hydrogen fluoride (Jorgensen, 1981). The reproduction of densities and heats of vaporization provides some confidence in both the size of the molecules and in the strengths of their intermolecular interactions. These early models (now called OPLS-UA) treated hydrogens bonded to aliphatic carbons as part of an extended atom but represented all other hydrogens explicitly.

The initial applications to proteins (Jorgensen and Swenson, 1985; Jorgensen and Tirado-Rives, 1988; Tirado-Rives and Jorgensen, 1990) used a polar-hydrogen–only representation, taking the atom types and the valence (bond, angle, dihedral) parameters from the 1984 Amber force field. This was called the AMBER/OPLS force field, and for some time was reasonably popular. As with Amber and CHARMM, an all-atom version (OPLS-AA) was developed later, but with much the same philosophy for derivation of charges and van der Waals parameters from simulations on pure liquids (Jorgensen *et al.*, 1996; Rizzo and Jorgensen, 1999; Kaminski *et al.*, 2001). Torsional parameters were developed in a consistent way by fits to HF/6-31G* energy profiles (Maxwell *et al.*, 1995), along with some recent modifications, especially for charged side chains (Kaminski *et al.*, 2001). Bond stretching and angle bending terms were standardized but were largely taken from the 1986 Amber all-atom force field. The parameter choices were intended to be "functional group friendly," so that they could be easily transferred to other molecules with similar chemical groupings. Although the parameters were principally derived with reference to condensed phase simulations, comparisons to gas-phase peptide energetics also show good results (Beachy *et al.*, 1997).

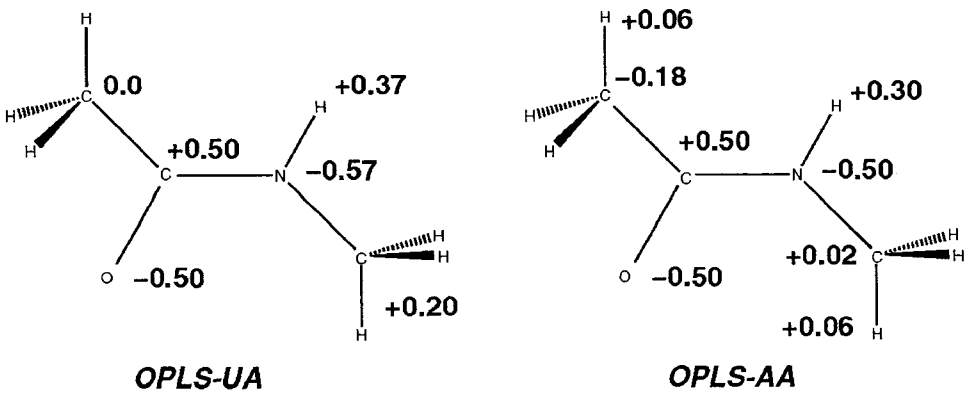


FIG. 4. Charge models for *N*-methylacetamide for two OPLS force field parameterizations. In OPLS-UA, the methyl groups are treated as united atoms, so that there is no breakdown into separate C and H charges.

Fig. 4 shows the fitted atomic charge models for the OPLS-UA and OPLS-AA force fields. Again, the polarity of the CO and NH bonds is similar to that in the Amber and CHARMM force fields. As with CHARMM, the OPLS charge fitting procedures yield small neutral groups.

D. Other Protein Force Fields

Our abbreviated discussion neglects many important advances in the development of protein force fields. In particular, there are several other protein potentials that have been widely used but will not be assessed in detail here. The GROMOS force fields (van Gunsteren *et al.*, 1998) were developed in conjunction with the program package of the same name (van Gunsteren and Berendsen, 1987; Scott *et al.*, 1999). The all-atom CEDAR and GROMACS force fields are largely derived from GROMOS. The Merck Molecular Force Field (MMFF) was developed by Halgren (Halgren, 1996a,b,c,d; Halgren and Nachbar, 1996; Halgren, 1999a,b), and has been aimed more at drug-like organic compounds than at proteins. MMFF was not derived for use in bulk phase simulations and performs poorly when used to model organic liquids (Kaminski and Jorgensen, 1996). This deficiency is not inherent in the buffered 14-7 function (Halgren, 1992) used in MMFF's van der Waals term, because this same functional form can be reparameterized to fit liquid data (Ren and Ponder, 2003). The DISCOVER force field (Maple *et al.*, 1998) has seen use primarily in conjunction with the commercial INSIGHT modeling package. The MM3 and MM4 potentials for amides (Lii and Allinger, 1991; Langley and Allinger, 2002) are an offshoot of

TABLE I
Other Force Fields for Peptide and Protein Modeling

Force field	Potential type	Key references
BUFF	All Atom	Carlson, 2000
CEDAR	All Atom	Hermans *et al.*, 1984; Hu *et al.*, 2003
CVFF	All Atom	Kitson and Hagler, 1988
DISCOVER	All Atom	Maple *et al.*, 1998
ECEPP/3	All Atom, Torsional	Némethy *et al.*, 1993
ENCAD	All Atom	Daggett and Levitt, 1993; Levitt *et al.*, 1995
GROMOS87	United Atom	van Gunsteren and Berendsen, 1987
GROMOS96	United Atom	Scott *et al.*, 1999
MM2	All Atom	Lii *et al.*, 1989
MM3	All Atom	Lii *et al.*, 1991
MM4	All Atom	Langley and Allinger, 2002
MMFF	All Atom	Halgren, 1996a,b,c,d
NEMO	Polarizable	Hermida-Ramón *et al.*, 2003
PROSA	Polarizable	Stern *et al.*, 1999
SCHRODINGER	Polarizable	Kaminski *et al.*, 2002
SDFF	Polarizable	Palmo *et al.*, 2003
SIBFA	Polarizable	Gresh, 1997; Guo *et al.*, 2000
SPASIBA	All Atom	Derreumaux and Vergoten, 1995
TRIPOS	All Atom	Clark *et al.*, 1989
UCSD-WILSON	All Atom	Mackay *et al.*, 1984
UFF	All Atom	Rappé *et al.*, 1992
UPJOHN	All Atom	Oie *et al.*, 1981
YETI	United, Torsional	Vedani, 1988

Allinger's highly respected molecular mechanics parameterizations and have been applied primarily to peptides. These MM methods use atomic charges only at formally charged groups, and rely on bond dipole moments to provide for most electrostatic interactions. A series of potentials refined over many years in Levitt's group (Levitt, 1983; Levitt and Sharon, 1988; Levitt *et al.*, 1995, 1997) are incorporated in the ENCAD *(EN*ergy *C*alculation *A*nd *D*ynamics) program and have been notably used to study protein folding and unfolding (Daggett, 2002). The ENCAD potential is unique in its use of group-based, rather than atom-based, neighbor exclusion of short-range electrostatic interactions. It also uses pairwise nonbonded potentials shifted to zero energy at short range, and specifically parameterized to reflect these small cutoff distances. The polarizable force fields listed in Table I will be discussed in section III of this review.

E. Comparisons Among Protein Force Fields

One might imagine that more than a quarter century of experience with force field methods for proteins would have led to some secure conclusions about their relative quality, at least for certain types of applications. It is very difficult, however, to generate "gold standards" by which different force fields may be compared. For one thing, experimental measurements reflect a great deal of thermal averaging of conformers in a complex condensed-phase environment. Until recently, few interesting simulations on peptides or proteins have been converged sufficiently well in their conformational sampling to allow deviations from experiment to be ascribed solely to deficiencies in the force field. Quantum chemistry calculations offer the attractive possibility of much simpler comparisons for fixed conformations or families of conformations depending on only a few degrees of freedom. For example, Halgren (1999b) has compared the predictions of inter- and intramolecular conformational energies for a variety of force fields to experiment and to quantum results. These comparisons offer useful insights into how force fields work but are not directly relevant to peptides and proteins. Furthermore, it is not clear that a force field optimized to fit such data for isolated molecules will work best in liquid-state simulations. This is especially true if electronic polarizability is ignored, as in the force fields described above, since the charge distributions appropriate for the gas phase will change significantly in a high-dielectric liquid such as water. Hence, simple class I–type force fields involve rather severe compromises between accuracy and simplicity, so that efforts to judge their "quality" are likely to depend strongly on what sorts of simulations are being considered and on the details of what is judged to be "correct" behavior.

With the passage of time, the fixed charge force fields appear to be converging toward a common electrostatic model. In Table II, it is seen that the most recent parameterizations of the Amber, CHARMM, and OPLS force fields use charge values that are more similar than those from previous sets—that is, compare the older (Amber *ff84*, *CHARMM19*, OPLS-UA) charges against the newer (*CHARMM22/27*, Amber *ff94/99*, OPLS-AA) values. Some of the convergence of these values represents movement toward an optimal fixed-charge model, but it also reflects the increasingly similar parameterization protocols and test sets (Beachy *et al.*, 1997) used by the Amber, CHARMM and OPLS developers. As the table indicates, other current force fields such as GROMOS96 and BUFF use very different charge values for amino acid residues.

After a lengthy period of refinement of partial charge values during the 1980s and 1990s, much effort has shifted toward improvement of the

TABLE II
Comparison of Fixed Partial Charge Models for Serine Taken from Current and Previous Generation Protein Force Fields

	Amber 84	Amber 94/99	CHARMM19	CHARMM22/27	OPLS-UA	OPLS-AA	GROMOS96	BUFF
N	−0.463	−0.4157	−0.35	−0.47	−0.57	−0.50	−0.28	−0.749
HN	0.252	0.2719	0.25	0.31	0.37	0.30	0.28	0.328
CA	0.035	−0.0249	0.10	0.07	0.20	0.14	0.00	0.189
HA	0.048	0.0843		0.09		0.06		0.048
C	0.616	0.5973	0.55	0.51	0.50	0.50	0.38	0.828
O	−0.504	−0.5679	−0.55	−0.51	−0.50	−0.50	−0.38	−0.679
CB	0.018	0.2117	0.25	0.05	0.265	0.145	0.15	0.296
HB	0.119	0.0352		0.09		0.06		0.006
OG	−0.55	−0.6546	−0.65	−0.66	−0.70	−0.683	−0.548	−0.764
HO	0.31	0.4275	0.40	0.43	0.435	0.418	0.398	0.491

protein torsional potentials. Each of the three major potentials discussed above has recently undergone revision of torsional parameters. The torsional parameters are traditionally the last values to be determined in the generation of a new protein force field, and these values have less theoretical underpinning than other force field terms. With the convergence of many of the other functional forms and parameters, the torsions are often used as a more general "error function" to correct the final force field results to agree with a desired set of conformational energy differences. For example, the recent OPLS-AA/L force field consists of a series of modifications of torsional parameters fit to *ab initio* results (Kaminski *et al.*, 2001). One important application for protein force fields with highly optimized torsional parameters is the prediction of side-chain rotamer preferences (Jacobson *et al.*, 2002), which is a critical component of accurate protein homology modeling protocols.

Even with the use of similar functional forms and parameter values that seem to be converging with the passage of time, recent versions of the traditional force fields still exhibit significant differences. Shown in Fig. 5 are Ramachandran free energy maps for solvated alanine dipeptide computed with the *CHARMM27*, Amber *ff94*, and OPLS-AA force fields. The simulation system consisted of one dipeptide molecule and 206 waters in a cubic box with periodic boundaries and with Ewald summation used to account for long-range electrostatics. Each map was determined by running 288 separate umbrella sampling MD trajectories restrained to small regions of ϕ-ψ space and totaling about 70 ns of simulation for each force field. The separate trajectories were then stitched together to generate full ϕ-ψ maps with a 2-D WHAM procedure (Kumar *et al.*, 1995). A similar comparison based on analysis of single lengthy MD trajectories has recently been published by Hermans' group (Hu *et al.*, 2003). In addition, this same group performed tight-binding DFT-based QM-MM simulations of a QM alanine dipeptide molecule solvated by MM water. Integrated conformer populations given in Table III show that all three force fields exhibit large differences from the DFT map: AMBER *ff94* strongly favors α-helical structures, OPLS-AA overpredicts β-sheet structure, and *CHARMM27* does not connect the helix and sheet structures via the expected bridge region (across $\psi = 0$) of the map.

One should not infer from these difficulties that there is a lack of interest in comparing one force field to another or in improving the existing parameterizations. It has been recognized for several years that many features seen in simulations are insensitive to the details of the force field parameterization. These are expected to include most qualitative features of the average structure of folded proteins and the nature and extent of fluctuations about the average structure. This expectation has

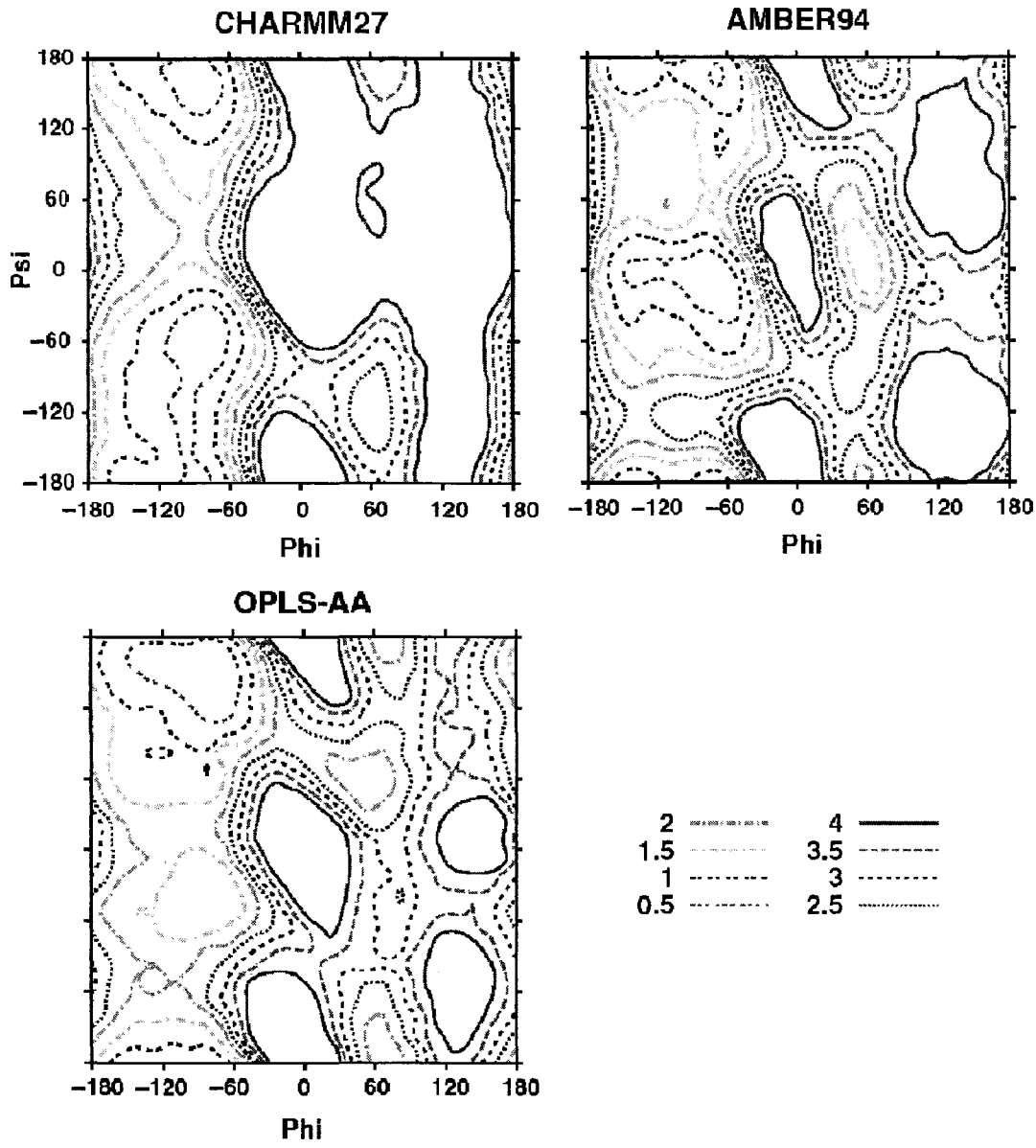


FIG. 5. Comparison of ϕ-ψ free energy plots from 2-D WHAM analysis of umbrella sampling trajectories for alanine dipeptide as computed with the AMBER *ff94*, CHARMM27 and OPLS-AA pairwise force fields. The contours shown are at 0.5 kcal/mol intervals from the global minimum of each Ramachandran map. (See Color Insert.)

been borne out in explicit comparisons of protein simulations with different force fields, concentrating on structure details (Roterman *et al.*, 1989; Ceccarelli and Marchi, 1997) or on fluctuation behavior (Teeter and Case, 1990; Price and Brooks, 2002). These results are often highly dependent on the strongly repulsive portions of the force field, which tend to be quite similar in different parameterizations, especially when

TABLE III
Alanine Dipeptide Population Percentages from QM-MM and Force Fields

	Alpha R	Bridge	Beta
DFT QM-MM	27	16	48
AMBER *ff94*	57	6	29
CHARMM27	46	3	49
OPLS-AA	14	10	70

The force field values were obtained by integrating data from Fig. 5 over the alpha R, bridge ("pass") and beta regions defined by Hu *et al.*, 2003.

liquid-state densities are a part of the fitting process. More variable are results that depend also on the softer, longer-range portions of the potential, where differences in the description of dispersion and electrostatic interactions become more important. Comparative tests on organic liquids (Kaminski and Jorgensen, 1996) show reasonable accord between popular force fields for those properties considered during the parameterization process. However, it is far from clear that this agreement would carry over, for example, to estimates of ligand-binding affinities or to helix-coil transition temperatures in peptides, which can be quite sensitive to energetic details.

III. BEYOND FIXED ATOMIC POINT-CHARGE ELECTROSTATICS

Unlike their biopolymer counterparts, empirical potentials for gas-phase, nonpolar organic molecules are exquisitely accurate. The widely respected MM3 force field for hydrocarbons has been used to compute the heat of formation of a set of 57 molecules, many of them very strained. The average estimated experimental error in ΔH_f was 0.40 kcal/mol for the full set of molecules. The standard deviation of the difference between the MM3 computed ΔH_f and the experimental values was 0.42 kcal/mol (Allinger *et al.*, 1989). The more recent MM4 (Allinger *et al.*, 1996) and Class II QMFF (Maple *et al.*, 1994) force fields give results at least as good as those for the original MM3. The quality of computed structures from any of these hydrocarbon force fields is equally high. We can conclude that, at least for hydrocarbons, a molecular mechanics computation is as trustworthy as the corresponding experimental result. For proteins and other biopolymers, the situation is currently much less satisfactory. Most of the effort to improve protein force fields has gone toward more accurate treatment of electrostatic and solvation effects, because these are thought to be the largest sources of error for these polar molecules in aqueous environments.

A. *Limitations of Fixed Atomic Point-Charges*

It is widely recognized that the use of an electrostatic model based on fixed atom-centered charges has two significant shortcomings. First, the restriction to only partial charges and to only the nuclear sites results in a model insufficiently flexible to describe certain features of molecular charge distributions. Second, the use of fixed charges means that the model is unable to respond directly to the molecular environment. For example, a carbonyl oxygen typically carries the same charge whether it is buried in the center of a folded protein or exposed to water on the protein surface. Both of these problems can be addressed via relatively simple extensions of the traditional fixed atomic point-charge model, but these modified force fields have yet to make their way into routine use.

Simple partial charge representations are intrinsically unable to accurately model the electrostatic potential around polar molecules. Algorithms to fit partial charges to the electrostatic potential defined by an *ab initio* wave function have been reported (PDM from Williams, 1988; CHELPG from Breneman and Wiberg, 1990). Williams' results indicate that optimally fit partial charges produce a potential with a relative RMS error of 5% to 15% from a target potential derived from quantum electronic structure calculations. Since partial charge electrostatic energies are usually expressed as a sum over pairs of interaction sites, the relative error in interaction energies will be even larger between two neighboring sets of sites. The key result is that atomic partial charge models lack the mathematical flexibility to describe the static, permanent electrostatics of general polar molecules to within "chemical accuracy." No amount of reparameterization can change this basic fact. A number of publications (Bayly *et al.*, 1993; Chipot *et al.*, 1993; Winn *et al.*, 1997) have proposed detailed schemes for arriving at atomic partial charges for biopolymers. In particular, Kollman's RESP model (Bayly *et al.*, 1993) solves some problems by applying penalty function restraints during the fitting process. Simultaneous fitting to multiple conformations of a flexible molecule can provide charges that are better determined statistically (Reynolds *et al.*, 1992). Stouch and Williams (1993) found similar improvement could be made by fixing ill-defined charges at chemically reasonable values. However, the inherent inflexibility of a nonpolarizable, fixed-charge model is not overcome by these methods.

Following on the work of Buckingham and Fowler (1985), atom- or bond-centered multipole expansions have been used to try to achieve a better fit to quantum-derived electrostatic potentials (for example, Sokalski *et al.*, 1993 and Colonna *et al.*, 1992). The relative RMS error in an optimal atomic multipole-based potential truncated after the quadrupole

TABLE IV
Energy and Structure of the K^+-Benzene Dimer

	ΔE_0	K^+-Centroid	ΔH_{298}
OPLS-AA	−9.32	2.90	
CHARMM27	−11.06	2.81	
Amber *ff94*	−12.55	2.74	
Amber *ff02*	−15.87	2.63	
AMOEBA	−19.27	2.81	−18.15
MP2/6-311+G(2d,2p)	−18.4	2.81	
MP2/aVQZ	−19.9	2.79	
CCSD(T)/CBS	−20.6	2.79	−20.1
Expt (HPMS)			−18.3
Expt (CID)			−17.7

Minimum energies and binding enthalples are given in kcal/mol, the K^+ to benzene centroid distance is given in Ångstoms. MP2/aVQZ and extrapolated CCSD(T)/CBS theoretical results are from Feller *et al.*, 2000; MP2/6-311+G(2d,2p) and CID experimental values are from Amicangelo and Armentrout (2000); the HPMS value, corrected for unimolecular dissociation, is from Sunner *et al.* (1981).

moments is usually less than 0.1% from the corresponding quantum-based target potential (Williams, 1988). The advantages of multipole methods are their functional flexibility and implicit inclusion of conformational effects (Dykstra, 1993; Price, 2000).

The need for explicit inclusion of polarization effects in protein force fields is a matter of current discussion in the literature (van der Vaart, 2000; Roux and Bernèche, 2002). A large body of work suggests that explicit polarization is required if a single set of parameters is to correctly describe both gas-phase cluster and bulk environments. However, the increased complexity and expense of polarizable force fields has seemed to require strong justification of their use in protein modeling.

A striking example of the importance of polarization in force-field calculations is provided by the interaction of a potassium cation with benzene. The benzene molecule is highly polarizable, and polarization effects are responsible for much of the large binding energy. As shown in Table IV, the simple partial charge force fields grossly underestimate the strength of this interaction. It is impossible to find reasonable sets of parameters that repair the simple force fields, especially in light of the concurrent need to fit water-ion and liquid benzene properties.

Table IV also shows results from two polarizable force fields, the Amber *ff02* parameter set and a polarizable force field being developed by one of us for the TINKER molecular modeling package (Ren and Ponder,

unpublished). Both of these models yield interactions closer to experimental and high-level quantum results, though the *ff02* binding energy is not tight enough.

Since truly accurate quantum calculations are finally becoming available for a variety of small molecule interactions (Dunning, 2000; Huang and MacKerell, 2002), polarizable protein force fields that are directly validated against such data will be parameterized in the near future. Optimal simple partial charge force fields for biopolymer modeling in water must implicitly include the effects of the aqueous environment. Without explicit polarization, current force fields should not be parameterized directly against high-level gas-phase quantum calculations. The ability to transfer quantum-derived electrostatics to bulk-phase modeling is a major practical advantage of polarizable force fields.

B. Flexible Models for Static Charge Distributions

Better static electrostatic models for polar molecules necessitate going beyond the atomic partial charge paradigm. This can be done by adding additional partial charges at sites other than nuclei or by placing a more elaborate model at the atomic centers. There is significant experience with both types of extensions to the basic atomic charge model.

Lone pairs sites have been used by various force fields on selected nitrogen, oxygen, or sulfur atoms to fine-tune hydrogen bond interactions. Both the MM methods of Allinger and some of the Amber parameterizations (Dixon and Kollman, 1997) have adopted lone pairs for some protein atoms. An example of off-center charge modeling in recent Amber force fields is given in the right-hand side of Fig. 2. A generalization of this method would be to use distributed off-site point charges that would in principle achieve any level of accuracy in reproducing a target potential (Brobjer and Murrell, 1982). The XED force field uses a charge on the nucleus and up to five nearby partial charges on "orbital points" to represent each atom, and it has advantages over traditional force fields in modeling aromatic interactions (Chessari *et al.*, 2002). Simple application of Coulomb's law then leads to an $O(N^2)$ CPU-time dependence on the total number of charges, which is inefficient when compared with atom-centered multipole approaches. In addition, off-site charges may result in "site fusion" during procedures that can generate large conformational changes between energy evaluations such as random or Monte Carlo searches. Finally, it is possible to replace simple point charges with isotropic atom-centered charge distributions. A method using a point charge and counterbalancing Gaussian charge distribution on each atom yields a novel water model claimed to mimic

some of the effects of the environment without the need for an explicit polarization calculation (Guillot and Guissani, 2001).

Another means of increasing the mathematical flexibility of a static electrostatic description beyond the limitations of partial atomic charges is to use atomic multipole moments. The derivation, use, and manipulation of higher-order moments in the study of intermolecular interactions are thoroughly covered in the important monograph by Stone (1996). Given the density matrix or electrostatic potential from a high-level electronic structure calculation, it is possible to derive atomic multipoles that provide an optimal least-squares fit to the target density. The PDM programs of Williams (1988) have been used to show that an atom-centered expansion through quadrupole moments can provide a potential within 0.1% of an MO-derived target. Alternatively, multipole expansions can be built by moving elements of overlap density from a quantum calculation to a distributed set of sites such as atomic nuclei. The similar distributed multipole analysis (DMA; Stone, 1981) and cumulative atomic multipole moments (CAMM; Sokalski and Poirier, 1983) protocols are examples of this approach. Spatial decomposition of the density, such as the atoms-in-molecules (AIM) procedure from Bader's lab (Bader, 1990; Popelier, 2000), has also been used to construct multipoles for protein atom types (Matta and Bader, 2000). It appears that atomic DMA values converge slightly more rapidly than AIM expansions (Popelier *et al.*, 2001), but the more important question for use in protein force fields relates to transferability of the multipole values.

If a limited set of lone pair partial charges are used, their motion relative to the rest of the structure can be governed by standard molecular mechanics bond and angle terms involving the lone pair sites. More generally, the use of off-site charges or atomic multipole moments requires definition of a local coordinate frame at each atom and motion of each atom's electrostatic model relative to its frame (Koch and Egert, 1995). There are two formulations available for computation of the electrostatic interactions between point multipoles: a polytensor scheme (Applequist, 1983; Dykstra, 1988) that becomes very efficient for higher-order derivatives and a spherical harmonic expansion (Price *et al.*, 1984; Popelier and Stone, 1994) leading to the energy, forces, and torques. In brief, the atomic multipoles at a site i can be represented as $M_i = [q_i, \mu_{i,1}, \mu_{i,2}, \mu_{i,3}, Q_{i,11}, Q_{i,12}, Q_{i,13}, \ldots Q_{i,33}]$ where q, μ and Q are the monopole, dipole, and quadrupole components, respectively. Then the interaction energy between two sites i and j is given by $M_j^T \cdot T \cdot M_i$ where the T tensors are constructed from extended derivatives of the inverse distance between the sites, $1/R$. Following Stone's notation (Stone, 1996), the multipole interaction T elements, with $4\pi\varepsilon_0$ omitted for clarity, are expressed as:

$$
\begin{aligned}
T &= \frac{1}{R} \\
T_\alpha &= \nabla_a T = -\frac{R_\alpha}{R^3} \\
T_{\alpha\beta} &= \nabla_\alpha T_\beta \\
T_{\alpha\beta\gamma} &= \nabla_\alpha T_{\beta\gamma} \\
&\cdots\cdots \\
(\alpha, \beta, \gamma, \ldots &= 1, 2, 3)
\end{aligned}
\tag{2}
$$

A concise description of the Ewald summation technique for multipole interactions is also available (Smith, 1998).

C. *Including Environmental Effects via Polarization*

As with the static electrostatic model, there are several possible methods available for including polarization in force-field calculations. An excellent recent review by Rick and Stuart (2002) covers the current methodologies for incorporating polarizability in simulations. An earlier summary of the status of polarizable force fields was provided by Halgren and Damm (2001). We will not repeat their treatments of the various polarization formalisms but focus instead on problems common to all polarizable force fields as they are extended from small, relatively rigid molecules such as water to large, flexible protein structures.

Three basic methods for including polarization have been studied: fluctuating charge, Drude oscillator, and induced dipole models. Fluctuating charge models use the principle of electronegativity equalization to produce a set of point charges that optimize the total electrostatic energy. In theory, intermolecular charge transfer is then handled by requiring conservation of charge for the whole system. In actual application, charge conservation is often enforced for each individual molecule. The original force field model of this class was the charge equilibration (QEq) method of Rappé and Goddard (1991) that is still used in their UFF force field (Rappé *et al.*, 1992) and its BUFF biopolymer counterpart (Carlson, 2000). Problems in computing analytical derivatives of QEq have led to the recent formulation of a "consistent" variation, CQEq (Kitao and Ogawa, 2003). The more recent dynamical fluctuating charge (FQ) method (Rick *et al.*, 1994) can be viewed as a perturbation formulation of QEq. In the typical use of FQ, the electrostatic forces required for dynamics simulation are propagated via an extended Lagrangian (Liu *et al.*, 1998), thus avoiding the need for a closed analytic form of the derivative.

Drude oscillator (DO) methods, also referred to as shell models, use a harmonic restraint potential to tether a mobile point charge to an interaction site. In a general Drude oscillator model, an atom will carry a charge fixed at the nucleus and a second restrained charge of variable position. The charge magnitudes and harmonic force constants are fit to atomic and molecular polarizability data and experimental energies. As with FQ, the DO model is most often used with an extended Lagrangian treatment of the variable charges during a molecular dynamics simulation (Mitchell and Fincham, 1993).

Perhaps the best-studied method for handling polarization is use of induced multipole moments. While higher-order multipole polarization and hyperpolarization can be included in force fields (Dykstra, 1989), only induced dipoles are usually considered. The use of induced dipoles in molecular dynamics dates back at least to Vesely's treatment of polarizable Stockmayer-type systems (Vesely, 1977). Point polarizabilities are generally assigned to either molecular centers of mass or distributed over some or all atomic sites. The direct, permanent electrostatic field felt at each polarizable site then results in a site of "direct" induced dipoles. But since the induced dipoles alter the field at each site, the procedure must be iterated to generate a self-consistent set of "mutual" induced dipoles arising from the mutual polarization.

Three methods are available for calculation of the induced dipole moments. The first is the iteration alluded to above. The induced dipole at each atomic site is computed as $\mu_{i,\alpha}^{ind} = \alpha_i E_{i,\alpha}$, where α_i is the atomic polarizability and $E_{i,\alpha}$is the sum of the fields generated by both permanent multipoles and induced dipoles:

$$\mu_{i,\alpha}^{ind} = \alpha_i \left(\sum_{\{j\}} T_\alpha^{ij} M_j + \sum_{\{j'\}} T_{\alpha\beta}^{ij'} \mu_{j',\beta}^{ind} \right) \quad \text{for } \alpha, \beta = 1, 2, 3 \tag{3}$$

where $M_j = [q_j, \mu_{j,1}, \mu_{j,2}, \mu_{j,3}, \ldots]^T$ contains the permanent multipole components and $T_\alpha^{ij} = [T_\alpha, T_{\alpha 1}, T_{\alpha 2}, T_{\alpha 3}, \ldots]$ is the interaction tensor between site i and j introduced in the previous section. The sets $\{j\}$ and $\{j'\}$ in equation 3 consist of all atomic sites except for separate small lists of omitted near-neighbor sites that may differ based on the details of the model. In fully interactive models (see below) the set $\{j'\}$ includes all atomic sites other than i itself. It can be shown that the solution of the above self-consistent equation can be written as:

$$\mu_{i,\alpha}^{ind}(n+1) = \mu_{i,\alpha}^{ind}(n) + \alpha_i \sum_{\{j'\}} T_{\alpha\beta}^{ij'} \mu_{j',\beta}^{ind}(n) \quad \text{for } n = 0, 1, 2, \ldots \tag{4}$$

where $\mu_{i,\alpha}^{ind}\ (0) = \alpha_i \sum_{\{j\}} T_{\alpha}^{ij} M_j$ is the "direct" induced dipole on site i due to the electric field from permanent multipoles of other molecules, and $\alpha_i \sum_{\{j'\}} T_{\alpha\beta}^{ij'} \mu_{j',\beta}^{ind}\ (n)$ is the "mutual" induced dipole further induced by dipoles on all the other sites. The iterative solution can generally be accelerated by well-known successive overrelaxation (SOR) techniques (Young, 1971) by using a value of $\omega = 0.75$ in equation 5, which converges rapidly to any reasonable level of accuracy.

$$\mu_{j,\alpha}^{ind}(n+1) = (1-\omega)\mu_{j,\alpha}^{ind}(n) + \omega \left[\mu_{j,\alpha}^{ind}(n) + \alpha_i \sum_{\{j'\}} T_{\alpha\beta}^{ij'} \mu_{j',\beta}^{ind}(n) \right] \qquad (5)$$

A second method for finding the induced dipoles is full-matrix direct solution of the set of coupled linear equations, avoiding the iterative procedure. Also, in the context of molecular dynamics, the dipoles can again be updated via an extended Lagrangian scheme. Ewald summation has been described for various polarizable multipole schemes (Nymand and Linse, 2000; Ren and Ponder, 2003), and particle mesh Ewald (PME) has been derived for induced dipole interactions (Toukmaji *et al.*, 2000).

D. Consistent Treatment of Electrostatics

In this section, we discuss the features necessary for a consistent and effective treatment of polarization that is intended for use in modeling flexible, polar molecules such as proteins. Discussion will be limited to induced dipole models, since these are the most commonly used type at present, but most of the points below also apply and can be adapted for fluctuating charge or Drude oscillator formulations.

At the most basic level, the polarization method used in a force field should be able to compute molecular polarizabilites with reasonable fidelity. A number of atomic polarizability models, including additive (Miller, 1990; Stout and Dykstra, 1998) and interactive models (Applequist *et al.*, 1972; Bode ande Applequist, 1996; Thole, 1981; van Duijnen *et al.*, 1998), have been proposed for treatment of molecular polarizability. Additive models allow the polarizable sites to respond to an external field but not to other sites within the molecule. Interactive models include the mutual effects of polarizable sites within a molecule on each other, as well as polarization induced by an external field. Simple additive models are sufficient for use with a force field if intramolecular polarization can be neglected, as is often the case for very small molecules. However, for larger peptides and proteins the distinction between intra- and intermolecular

polarization is obviously blurred, and an interactive model that treats both on the same basis is required.

The successful Dang-Chang polarizable model for water (Dang and Chang, 1997) uses only a single polarizable site that carries the experimental molecular polarizability of 1.444 $Å^3$. Other models have used distributed polarizabilities on oxygen and the hydrogens that linearly sum to the correct water value. In these models, intramolecular polarization is neglected and is implicitly included within the overall parameterization scheme.

Alternatively, the POL3 water model (Caldwell and Kollman, 1995) and the Amber *ff02* protein force field (Cieplak *et al.*, 2001) use a modified form of Applequist's interactive polarization. The Applequist model uses generally smaller distributed atomic polarizabilities than most additive models. When the sites interact via mutual polarization, the total polarizability is increased and correct molecular polarizabilities are generated. For example, the POL3 model for water contains point polarizabilities of 0.528 $Å^3$ on oxygen and 0.17 $Å^3$ on hydrogen. When these sites interact via full mutual induction, a molecular polarizability of 1.46 $Å^3$ is generated in good agreement with the experimental water value. However, the Amber *ff02* energy model calls for neglect of mutual induction between atoms that are 1-2 or 1-3 bonded in analogy with the molecular mechanics tradition of omitting electrostatic interactions between near neighbor atoms. The result is that the atomic polarizabilities in the POL3 model do not interact, and the molecular polarizability of POL3 water is reduced to the additive value of 0.87 $Å^3$. Based on this analysis, we might predict that the POL3 model will be underpolarized, and indeed this appears to be the case. The POL3 water dimer interaction energy and O-O separation distance are 5.44 kcal/mol and 2.79 Å, respectively. These values are intermediate between "fully polarized" high-level quantum results of 5.0 kcal/mol and 2.91 Å and a typical nonpolarizable pairwise model such as TIP3P with dimer values of 6.54 kcal/mol and 2.75 Å. The analysis extends to proteins as well. Neglect of neighboring atom mutual polarization leads to a systematic underestimation of polarization effects in *ff02*. The remaining polarization can, of course, be implicitly included in the overall parameterization as is the case with the simple pairwise protein force fields.

Unfortunately, it is not possible to simply "turn on" the 1-2 and 1-3 mutual polarization to regain the full molecular polarizability. Attempts to include full mutual polarization within a modified POL3 model lead to a "polarization catastrophe" in the liquid wherein the induced dipole moments fail to converge and become infinite in magnitude. The origin of the catastrophe resides in the non-physical approximation of using a

point polarizability model. The diffuse nature of the true charge distribution and the fact that polarizability is more realistically spread over corresponding regions of space suggest that polarization should be damped at short range. The lack of damping in the Applequist model leads to failure of energetic models derived from it. The inability of the original Applequist scheme to adequately model the molecular polarizability of very polar, anisotropic molecules such as benzene is another symptom of the same problem (Applequist, 1993).

Among the interactive models, the damped induction suggested by Thole (1981) exhibits several advantages: (1) it avoids the polarization catastrophe at short range by replacing point dipole interactions with interactions between smeared dipoles, (2) it produces anisotropic responses to an external field by using only isotropic atomic polarizabilities, and (3) atomic polarizabilities derived within the model are highly transferable. For example, a single atomic polarizability value for each of the elements C, N, O, and H gives an excellent fit to a large number of experimental molecular polarizabilities (van Duijnen and Swart, 1998).

To avoid a "polarization catastrophe" at very short range, Thole introduced a modification scheme in which dipole interactions are damped as though one of the point dipoles in each pairwise interaction is replaced by a smeared charge distribution. As a result, the dipole interaction energy approaches a finite value instead of becoming infinite as the separation distance approaches zero. Several charge distributions yield roughly similar results, but the one most commonly used to date has the form

$$\rho = \frac{3a}{4\pi} \exp(-au^3) \tag{6}$$

where $u = R_{ij}/(\alpha_i\alpha_j)^{116}$ is the effective distance as a function of atomic polarizabilities of sites i (α_i) and j (α_j). The factor a is a dimensionless width parameter of the smeared charge distribution and controls the strength of damping. Using the charge distribution given by equation 6, the damped T matrix elements corresponding to equation 3 can be derived. It has been shown that the damped first-order T element is (Kong, 1997)

$$T_\alpha^D = -[1 - \exp(-au^3)]\frac{R_\alpha}{R^3} \tag{7}$$

The modified higher-order T matrix elements can be obtained successively by taking the derivative of the preceding lower rank elements:

$$
\begin{aligned}
T^D_{\alpha\beta} &= \lambda_5 \frac{3R_\alpha R_\beta}{R^5} - \lambda_3 \frac{\delta_{\alpha\beta}}{R^3} \\
T^D_{\alpha\beta\gamma} &= -\lambda_7 \frac{15R_\alpha R_\beta R_\gamma}{R^7} - \lambda_5 \frac{3(R_\alpha \delta_{\beta\gamma} + R_\beta \delta_{\alpha\gamma} + R_\gamma \delta_{\alpha\beta})}{R^5} \\
T^D_{\alpha\beta\gamma\eta} &= \lambda_9 \frac{105R_\alpha R_\beta R_\gamma R_\eta}{R^9} \\
&\quad -\lambda_7 \frac{15(R_\alpha R_\beta \delta_{\gamma\eta} + R_\alpha R_\gamma \delta_{\beta\eta} + R_\alpha R_\eta \delta_{\beta\gamma} + R_\beta R_\gamma \delta_{\alpha\eta} + R_\beta R_\eta \delta_{\alpha\gamma} + R_\gamma R_\eta \delta_{\alpha\beta})}{R^7} \\
&\quad +\lambda_5 \frac{3(\delta_{\alpha\beta}\delta_{\gamma\eta} + \delta_{\alpha\gamma}\delta_{\beta\eta} + \delta_{\alpha\eta}\delta_{\beta\gamma})}{R^5}
\end{aligned}
\tag{8}
$$

where the λ_i for $i \in \{3,5,7,9\}$ are the damping coefficients that modify the standard interactions

$$
\begin{aligned}
\lambda_3 &= 1 - \exp(-au^3) \\
\lambda_5 &= 1 - (1 + au^3)\exp(-au^3) \\
\lambda_7 &= 1 - (1 + au^3 + \frac{3}{5}a^2u^6)\exp(-au^3) \\
\lambda_9 &= (1 - [1 + au^3 + (18a^2u^6 + 9a^3u^9)/35]\exp(-au^3)
\end{aligned}
\tag{9}
$$

After replacing the original interaction matrix elements with the above damped ones, the energy, force, and electric field are computed in the usual fashion. Some groups apply damping to all electrostatic interactions (Burnham *et al.*, 1999). Other workers choose to damp only those energy terms involving induced moments, leaving the interactions between permanent multipoles unmodified (Ren and Ponder, 2003).

Under a Thole-style scheme, atomic polarizabilities are generally larger than those from simple additive fits because of the reduction in polarization caused by the short-range damping. For example, a modified Thole water model devised by the Ponder group (Ren and Ponder, 2003) uses atomic polarizabilities of 0.837 Å^3 on oxygen and 0.496 Å^3 on each hydrogen but yields a molecular polarizability of only 1.41 Å^3. The Thole interaction model was introduced into energy calculations by the Groningen group (de Vries *et al.*, 1997) and later adopted by Burnham *et al.* (1999) for their initial "Thole-type" model (TTM) for water. It has also recently been used by Karlstrom's group for their NEMO polarizable potential (Brdarski *et al.*, 2000).

Once a polarization model gives satisfactory results for molecular polarizabilities, the next test is generation of accurate electrostatic potentials in the region surrounding a polarizable molecule. A key point is that permanent electrostatics taken from a quantum calculation should not be used directly as the permanent model in a polarizable force field. If mutual induction is allowed as in the Applequist or Thole models, then

some intramolecular polarization must be removed from the quantum results, where it is already implicitly included, to avoid "double counting" of induction effects. Various groups have devised procedures to modify molecular orbital-derived charges or multipoles for use with their intramolecular polarization schemes (Cieplak *et al.*, 2001; Ren and Ponder, 2002). The multipoles on each atom taken directly from a quantum calculation may be considered as a sum of intrinsic "permanent" and "induced" moments:

$$M_i = M_i^{perm} + M_i^{ind}, \tag{10}$$

where M_i^{ind} is produced by direct and mutual induction from all sites in the absence of an external field:

$$M_{i,\alpha}^{ind} = \alpha_i \left(\sum_{\{j\}} T_\alpha^{ij} M_j^{perm} + \sum_{\{j'\}} T_\alpha^{ij'} M_{j'}^{ind} \right). \tag{11}$$

This is the same relation as in equation 3, except that the induced dipole is replaced by generalized induced moments. Substitution of equation 10 into the above expression yields

$$M_{i,\alpha}^{ind} = \alpha_i \left(\sum_{\{j\}} T_\alpha^{ij} (M_j - M_j^{ind}) + \sum_{\{j'\}} T_\alpha^{ij'} M_{j'}^{ind} \right). \tag{12}$$

If the same scaling factors are employed for mutual and direct induction using identical groups $\{j\}$ and $\{j'\}$, then equation 12 reduces to a surprisingly simple expression:

$$M_{i,\alpha}^{ind} = \alpha_i \sum_{\{j\}} T_\alpha^{ij} M_j. \tag{13}$$

In the case where $\{j\}$ and $\{j'\}$ are not the same, or when different fractional scaling factors are desired for short-range direct and mutual interactions, a more complex formulation requiring iterative solution is necessitated. The underlying "permanent" moments, M_i^{perm}, calculated via equation 10 will return the full quantum mechanical result when combined with the force field polarization model defined by equation 11. Following this protocol, "permanent" moments from small quantum calculated fragments can be combined in a consistent fashion to build electrostatic models for larger structures. Finally, we should point out that it may be convenient (or necessary) to differentiate the short-range treatment of direct induction caused by the permanent field from the treatment of mutual induction. A number of empirical models for inclusion of direct

induction have been tested (Ren and Ponder, 2002), and the best choice will depend on other details of the particular polarization implementation being considered.

Finally, a polarizable force field has to achieve a balance between inter- and intramolecular energetics. All molecular mechanics methods face the challenge of combining a description of nonbonded interactions (vdW, electrostatics, polarization) with short-range valence interactions (bonds, angles, torsions). As noted above, force fields usually ignore or scale intramolecular vdW and permanent electrostatic interactions between atoms separated by three or fewer bonds. Such interactions are thought to be handled by the valence terms. Polarizable force fields developed to date apply similar schemes for scaling of short-range intramolecular polarization. However, there is no reason that the scaling scheme used for the induction calculation will be optimal, or even appropriate, for computing energetic terms arising from polarization. Use of separate scaling schemes provides a useful flexibility during parameterization of a polarizable force field but at the cost of a somewhat more complex calculation. The felicitous merging of valence and nonbonded terms is a critical consideration in any successful force field for flexible proteins, and polarizable models are no exception.

E. Current Status of Polarizable Force Fields

The systems studied via polarizable potentials fall into two separate classes, each exposing an additional layer of challenges for the force-field developer. The simpler class contains water and most other small molecules. A large number of water models have been proposed, many of which use potentials that would be difficult to extend to more general molecular systems. As noted above, several workers have found that polarization in water is treated adequately by a single isotropic polarizable site. Force fields for small organics require potentials applicable for several elements or atom types and a distributed polarization model. However, as long as the systems considered are monofunctional or lacking rotatable bonds, it is usually sufficient to ignore explicit intramolecular polarization by including it implicitly in the parameterization. The second class would include polarizable force fields for arbitrarily large, flexible molecules such as proteins. In this case, the force field developer has the additional problem of maintaining consistency and balance in the treatment of intra- vs. intermolecular polarization. Also, there is the hope that the final model will account for at least some of the conformational effects of polarization in flexible systems.

1. *Polarizable Models for Small Molecules*

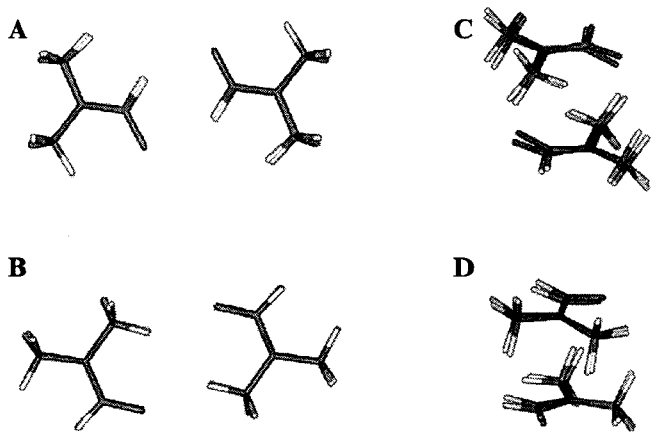
There have undoubtedly been more polarizable simulations reported for liquid water than for any other substance. Previous attempts to parameterize a water model applicable across multiple phases date back to the "polarizable electropole" model of Barnes *et al.* (1979) that treated electrostatics via a single site carrying the experimental dipole and quadrupole moments and isotropic dipole polarizability. An interesting polarizable and dissociable potential was suggested by Stillinger and David (1978) and applied to water clusters and ion monohydrates. The original MCY model (Matsuoka *et al.*, 1976) based on a fit to points on an *ab initio* dimer surface has evolved into the NCC-vib potential (Corongiu, 1992) that includes induced dipole polarization and vibrational flexibility. Sprik and Klein (1988) made an early attempt at modifying the existing TIP4P potential to include polarization within a Drude oscillator framework particularly suitable for molecular dynamics simulation. Another early water model from the Levy group (Bernardo *et al.*, 1994) uses partial charges and induced dipoles on each atomic center, along with a Thole-like polarization damping scheme.

Recent work on polarizable water includes a simple, rigid, non-iterative 3-site model (Yu *et al.*, 2003) parameterized to include only direct induction and an elaborate Drude oscillator model (MCDHO; Saint-Martin *et al.*, 2000) that was compared with high-quality *ab initio* calculations. The POL5/TZ and POL5/QZ water models (Stern *et al.*, 2001), which combine fluctuating charge and induced dipole polarization, yield good agreement with a range of structural and thermodynamic properties. The AMOEBA water model (Ren and Ponder, 2003) also gives excellent cluster and liquid phase results and has been designed for compatibility with a protein force field built around a polarizable atomic multipole electrostatic description.

The family of second-generation TTM models, TTM2-R and TTM2-F (Burnham and Xantheas, 2002a,b), exhibit excellent agreement with *ab initio* structures and energies for water clusters through $(H_2O)_6$ but perform less well in liquid simulations. The TTM2-F model points up an interesting problem not solved by including polarization in the potential. It is known experimentally that the bond angle in water increases from 104.52° in the gas phase to a value near 106° in the liquid. Essentially all flexible water models, pairwise or polarizable, exhibit the opposite trend in that the average angle value decreases on moving from gas to liquid. As Burnham and Xantheas show with TTM2-F, the problem is that the electrostatic model does not respond correctly to changes in valence geometry. In particular, the dipole moment derivative vector with respect

to bond stretching lies in the wrong direction in all empirical potentials, including those with explicit polarization. The solution lies in explicitly coupling the electrostatic model to the valence terms. The complex TTM2-F model uses a numerical correction based on a theoretical dipole moment surface for the water molecule. A potentially similar, but simpler, method based on bond charge fluxes has been proposed by Palmo and Krimm (1998). It seems likely that analogous problems will be exposed for the backbone amide group of proteins. Many studies suggest that typical protein force fields are too stiff with respect to pyramidalization at the amide nitrogen. A very recent publication, also from the Krimm group (Mannifors *et al.*, 2003), provides an explanation and fix based again on electrostatic-valence coupling and charge flux.

A major impetus for "next generation" empirical potentials is the hope that accurate prediction of ligand and drug-binding energies with proteins will become a reality. The specific electrostatic interactions encoded in a polarizable force field are certainly capable of describing effects completely missing in current generation force fields. For example, Fig. 6 shows a comparison of the OPLS-AA and AMOEBA force fields for



Energy Model	A	B	C	D
MP2/aug-cc-pVTZ (BSSE)	−5.35	−4.14	−8.34	−8.90
OPLS-AA	−3.45 (0.25)	−3.54 (0.64)	−5.42 (0.41)	−5.18 (0.69)
AMOEBA	−4.97 (0.08)	−5.37 (0.09)	−7.80 (0.24)	−8.79 (0.15)

FIG. 6. Comparison of *ab initio*, OPLS-AA and AMOEBA structures and energies for dimer configurations of dimethylformamide (DMF). The MP2 values are taken from Vargas *et al.* (2000), who suggested the DMF dimer as a model for possible C—H to O=C hydrogen bonding interactions in proteins. All energies are in kcal/mol. RMS values in Å from the *ab initio* minima are given in parentheses. The upper panel shows the superposition of the AMOEBA minima on the Vargas *et al.* structures.

four configurations of the dimethylformamide (DMF) dimer. According to correlated large-basis molecular orbital calculations (Vargas *et al.*, 2000), configurations C and D exhibit much tighter binding than either A or B because of the presence of non-traditional C-H to O=C hydrogen bonds. This structural feature is not included in the parameterization of current generation force fields, and OPLS-AA fares poorly on all of the DMF configurations, especially C and D. Such specific, but subtle, energetic effects almost certainly play an important role in regulating binding at protein receptor sites.

2. *Polarizable Protein Force Fields*

As of early 2003, there is no polarizable force field that sees wide use in protein modeling. Several efforts are under way, but parameter sets are still under development, and published applications are limited to those from the development groups. This is partly a reflection of the difficulty of the problems involved and the fact that the behavior of polarizable force fields for flexible molecules is not yet completely understood. Also, the methods and software used to treat polarization are not as standardized as for the current generation of pairwise protein potentials.

The Amber *ff02* potential (Cieplak *et al.*, 2001) has been mentioned above. It represents an initial polarizable member of the Amber family of force fields and is the result of Peter Kollman's long-standing interest (Lybrand and Kollman, 1985) in non-additive force fields. Unlike other developing polarizable protein models, *ff02* was intentionally derived via perturbation of an existing pairwise force field, a fact that should make it easier to judge the overall effect of including *ff02*-style polarization in protein simulations.

The Columbia groups of Friesner and Berne, in collaboration with researchers at Schrödinger, have produced an evolving series of polarizable protein models. The first, sometimes referred to as the PROSA force field, is based on a fluctuating charge formalism (Banks *et al.*, 1999). This model was later extended via the addition of induced dipoles on some sites (Stern *et al.*, 1999). Recently, many of these same workers have published a new "first generation" protein potential that abandons the FQ methodology in favor of exclusive use of induced atomic dipoles. The latest model claims to reproduce quantum mechanical energies of di- and tetrapeptide test cases to within an average RMSD of 0.5 kcal/mol. The same torsional fitting techniques currently being used for the OPLS-AA parameter set were applied by the Columbia and Schrödinger workers in an attempt to improve side chain rotamer prediction.

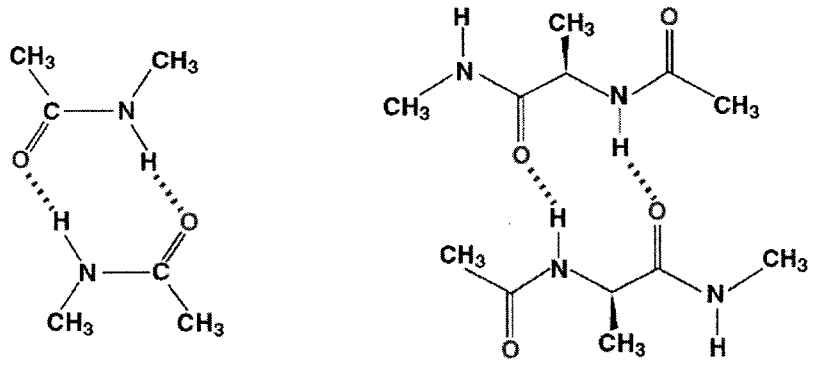
The SIBFA force field developed by Gresh (1997) has been applied to a series of protein structure and molecular recognition problems.

The energy function is based on a rather direct decomposition of *ab initio* SCF computations, including multipole electrostatics, repulsion, and dispersion-like terms and separate terms for polarization and charge transfer. The latter term is considered critical for ion interactions but also contributes a substantial fraction of the energy of polar hydrogen bonds as computed by SIBFA (Gresh, 1997; Masella *et al.*, 1998). The SIBFA potential was combined with a continuum reaction field solvation model to analyze the conformational preferences of alanine dipeptide and other oligopeptides (Gresh *et al.*, 1998).

Very recently, a NEMO potential for a capped glycine residue has been reported (Hermida-Ramón *et al.*, 2003). This study devoted considerable attention to the intramolecular potential, and presented comparisons with several nonpolarizable protein force fields. The authors conclude that their model accounts for most of the intramolecular polarization, but that errors on the order of 1-2 kcal/mol should be expected in conformational energies.

The SDFF (spectroscopically determined force field) from Krimm's group at Michigan (Palmo *et al.*, 2003) is unique among current polarizable protein force-field efforts in its extensive use of cross-terms and emphasis on valence potential functions. In particular the use of charge flux methods, as described earlier, appears to provide an important next step with its flexible coupling of electrostatics to bond stretching (Palmo and Krimm, 1998; Mannfors *et al.*, 2003). At present the SDFF has been described for amides and peptide backbone components, but it remains to be elaborated into a full protein force field.

In 1997 Beachy *et al.* reported an extensive comparison of nonpolarizable protein force fields and uncovered a systematic difference in alternative hydrogen bonding arrangements between fixed charge potentials and *ab initio* quantum calculations. Fig. 7 highlights the difference in dimerization energy between the simple amide *cis*-NMA and a larger protein β-sheet model based on alanine dipeptide. These systems were subsequently modeled with the SIBFA force field (Gresh *et al.*, 1999). The SIBFA results are shown in the figure and compared with some nonpolarizable models, a preliminary AMOEBA protein force field, and quantum calculations. The current generation of fixed-charge force fields predicts the β-sheet model to be more stable than the *cis*-NMA structure, whereas more elaborate force fields such as AMOEBA and SIBFA predict the opposite, in agreement with the quantum results. The most reasonable conclusion is that fixed charge force fields are missing cooperative effects in hydrogen bonding networks that can lead to significant energetic errors.



Energy Model	*cis*-NMA	β-sheet	ΔE
OPLS-AA	−11.5	−16.9	−5.4
CHARMM27	−11.6	−16.9	−5.3
AMBER *ff94*	−11.3	−14.8	−3.5
AMBER *ff02*	−13.5	−14.8	−1.3
AMOEBA	−18.5	−12.6	+5.9
SIBFA	−18.7	−17.1	+1.6
MP2/(CEP)4-31G+(2d)	−20.5	−17.5	+3.0
BP/DZVP *(BSSE corrected)*	−16.2	−8.4	+7.8

FIG. 7. Comparison of interaction energies for dimers of *cis*-N-methylacetamide (NMA) and alanine dipeptide (β-Sheet) as computed via quantum and force field methods. All energies are in kcal/mol. Force field values for the NMA dimer are fully optimized geometries, while the β-sheet structures were restrained to $\phi = -139^{\circ}$ and $\psi = +135^{\circ}$. The SIBFA and quantum values are taken from Gresh *et al.* (1999).

IV. Modeling the Solvent Environment

A. *Explicit Water Models*

The original atomic-scale computational model for liquid water was proposed by Bernal and Fowler (1933). At the time of the development of the earliest protein force fields, the ST2 model of Stillinger and Rahman (1974) was in wide use. The ST2 model was adopted for some of the first protein simulations done in the Karplus group at Harvard. The early 1980s saw the introduction of the SPC (Berendsen *et al.*, 1981) and TIP3P (Jorgensen *et al.*, 1983) potentials, two similar, rigid 3-site water models parameterized to reproduce the basic bulk phase structure and thermodynamics of liquid water. At present, these two potentials are still the most commonly cited solvent models for use in protein modeling and

simulation. Still other water potentials have added intramolecular flexibility and interpolate the parameter space between the SPC and TIP3P models (for example, the F3C model of Levitt *et al.*, 1997 and a model from Ferguson, 1995). The addition of a fourth site along the H-O-H bisector leads to the TIP4P model (Jorgensen *et al.*, 1983). This potential yields improvements in the gas-phase dimer structure and liquid O-O radial distribution function and has been generalized to the flexible TIP4F model (Mahoney and Jorgensen, 2001).

The recently developed TIP5P potential (Mahoney and Jorgensen, 2000) exhibits excellent agreement with the experimental internal energy, density, and O-O radial distribution at room temperature. This 5-site model is essentially a modernization of the ST2 potential. It uses only fixed partial charges to model electrostatics and includes polarization response to the environment only in an averaged, mean-field sense. As a result, TIP5P provides an excellent description of the homogeneous bulk phase but is a poor model for gas-phase clusters and for nonpolar solutes in polar solvents. For example, the gas-phase binding energy of the water dimer is overestimated by more than 30% in the TIP5P model.

In application to large biomolecular systems, there is concern that such models cannot correctly account for situations where the same nonpolarizable moiety is exposed to different electrostatic environments, either within a single large static structure or during a course of simulation. In addition, there is an inherent inconsistency in most nonpolarizable models related to their static inclusion of average bulk polarization within the potential. This results in internal energies and other properties that are derived against a gas-phase reference state that is already "prepolarized" for the liquid phase. While it is possible to correct for the resulting self-energy of the reference state, as in the SPC/E water model (Berendsen *et al.*, 1987), such corrections are not routinely used for heterogeneous systems. This suggests that use of SPC/E with uncorrected protein potentials will lead to a mismatch in the balance of water-water and protein-water electrostatics. Of course, this reference state problem disappears if fully polarizable potentials are used for both the protein and water components.

Protein simulations containing explicit water have obvious advantages over continuum solvent treatments in some respects. For example, detailed bound solvent motifs, such as water bridges, require solvent molecules as an integral part of the structural model. A recent report of potentials of mean force (PMFs) for side chain pairs in explicit and various continuum solvents (Masunov and Lazaridis, 2003) shows a dramatic lack of fine structure in the continuum versus explicit PMFs because of continuum averaging of specific water interactions.

B. *Continuum Solvent Models*

There are many circumstances in molecular modeling studies in which a simplified description of solvent effects has advantages over the explicit modeling of each solvent molecule. One of the most popular models, especially for water, treats the solvent as a high dielectric continuum, interacting with charges that are embedded in solute molecules of lower dielectric. The solute charge distribution, and its response to the reaction field of the solvent dielectric, can be modeled either by quantum mechanics or by partial atomic charges in a molecular mechanics description. In spite of the severity of the approximation, this model often gives a good account of equilibrium solvation energetics and is widely used to estimate pKs, redox potentials, and the electrostatic contributions to molecular solvation energies (Tomasi and Persico, 1994; Cramer and Truhlar, 1999; Beroza and Case, 1998; Schaefer *et al.*, 1998; Roux and Simonson, 1999). For molecules of arbitrary shape, the Poisson-Boltzmann equations that describe electrostatic interactions in a multiple-dielectric environment are typically solved by finite-difference or boundary element numerical methods. These can be efficiently solved for small molecules but may become expensive for proteins or nucleic acids. Although progress continues to be made in numerical solutions, there is a clear interest in exploring more efficient, if approximate, approaches to this problem.

Most of the simulation techniques to be used here rest on models in which the protein degrees of freedom are treated explicitly but the solvent degrees of freedom are not. This requires that the energy surface used for the protein degrees of freedom be a potential of mean force (PMF) in which the solvent degrees of freedom are implicitly averaged over (Roux and Simonson, 1999). Assuming that the full potential energy function consists of a term, U_{vac} for the interactions within the protein, depending only on the protein degrees of freedom, $\boldsymbol{r}$, and additional terms for the protein-solvent and solvent-solvent interactions, the PMF is ideally,

$$U_{pmf}(\boldsymbol{r}) = U_{vac}(\boldsymbol{r}) + \Delta G_{sol}(\boldsymbol{r}) \tag{14}$$

where $\Delta G_{sol}(\boldsymbol{r})$ is the free energy of transferring the protein from vacuum to the solvent with its internal degrees of freedom fixed at $\boldsymbol{r}$. A very common model divides $\Delta G_{sol}(\boldsymbol{r})$ into electrostatic and non-polar contributions, which are defined in terms of a hypothetical thermodynamic cycle in which the process of vacuum to solvent transfer is carried out in three steps: in vacuum, set to zero all atomic partial charges; then transfer the resulting apolar molecule to solvent; and finally, in solvent, restore the atomic partial charges to their original values. ΔG_{np} is then the

free energy change for the apolar transfer step, and ΔG_{elec} is the difference of the electrostatic work of charging the system in solvent versus vacuum.

The non-polar term should be very similar to the free energy of transferring from vacuum to solvent an aliphatic compound with the same steric form as the solute molecule of interest. For small, fairly rigid molecules, this can be taken from experimental measurements on hydrocarbons (Chen *et al.*, 1994) or from a variety of theories for solvation of non-polar molecules. A common estimate is that aliphatic solvation energies in water have linear dependence on solvent accessible surface area,

$$\Delta G_{np} = \gamma(SA) + b \tag{15}$$

where SA is the solvent accessible surface area and γ and b are empirical parameters fit to the solvation free energy of alkanes (Sitkoff *et al.*, 1994; Simonson and Brünger, 1994). This simple model is certainly not fully correct, and there have been a number of attempts to extend it (Sharp *et al.*, 1996; Gallichio *et al.*, 2002), but a full discussion of this topic is beyond the scope of this review. For the rest of this section, we will consider models for the electrostatic contribution to hydration.

1. The COSMO Model

The COSMO approach (Klammt and Schüürmann, 1993) is a very simple yet effective model for continuum solvents that allows the basic ideas to be developed in a few lines of math. COSMO stands for Conductor-like Screening Model and treats the situation illustrated in Fig. 8, where the solvent dielectric becomes infinite (as in a conductor). As with other models, we imagine the solute to be a low-dielectric object (with embedded charges) that is immersed in this conductor-like solvent environment.

Upon immersion, charge carriers in the solvent can move freely to the surface of the solute, creating a non-uniform charge distribution at the molecular surface. The electrostatic energy of this set of charges will be:

$$E = E_{gas} + \sum_i \int z_i \frac{1}{|r_i - r_q|}\, qdS + \frac{1}{2}\int q \frac{1}{|r_q - r_{q'}|}\, q' dSdS' \tag{16}$$

Here E_{gas} is the electrostatic energy of the molecular charge distribution in the absence of solvent. The z_i are partial atomic charges in the molecule, at positions r_i, and q represents the surface charge in an element of surface area, dS. Hence, the second term in equation 16 represents the interaction energy between the molecular charge distribution and the induced surface charges, whereas the final term represents the mutual repulsion among these induced charges. Anticipating the fact that the

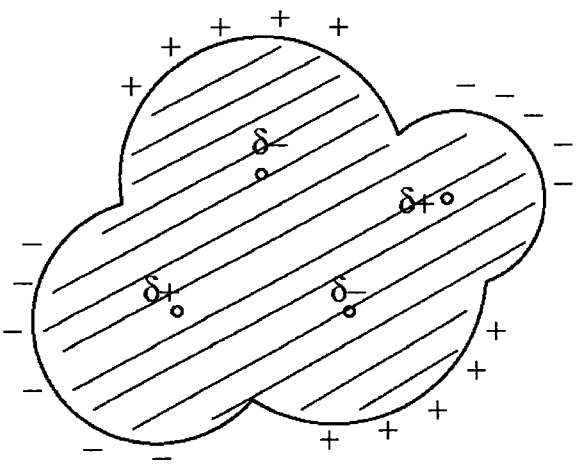


FIG. 8. Schematic diagram of a solute embedded in a medium of infinite dielectric. The shaded area represents the molecule, which consists of low dielectric material with embedded partial atomic charges. The "+" and "−" signs along the molecular surface represent the accumulation of charges to balance the atomic charges. This surface charge distribution can be determined by minimizing as the electrostatic energy of the system, as described in the text.

integrals above eventually will be estimated by some sort of quadrature, we can re-write this as a matrix equation:

$$E = E_{gas} + \mathbf{z}^T\mathbf{B}\mathbf{q} + \frac{1}{2}\mathbf{q}^T\mathbf{A}\mathbf{q} \tag{17}$$

Here **z** and **q** are vectors representing the molecular and surface charges, and **A** and **B** are matrices of inverse inter-charge distances. For the present discussion, the details of how the discretization is carried out are not relevant.

To determine the induced charge distribution, we simply minimize the energy expression in equation 17 by setting $(\partial E/\partial \mathbf{q}) = 0$, which yields

$$\mathbf{A}\mathbf{q} = -\mathbf{B}\mathbf{z} \tag{18}$$

or

$$\mathbf{q} = -\mathbf{A}^{-1}\mathbf{B}\mathbf{z} \tag{19}$$

Substituting **q** from equation 19 into equation 17 gives expressions for the solute-solvent interaction energy:

$$E_{solute\text{-}solvent} = -\mathbf{z}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{B}\mathbf{z} \tag{20}$$

and for the solvent-solvent repulsion energy:

$$E_{solvent\text{-}solvent} = \frac{1}{2}\mathbf{z}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\mathbf{B}\mathbf{z} = \frac{1}{2}\mathbf{z}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{B}\mathbf{z} \tag{21}$$

There are two very useful features of these models that can be seen from this very simple, symbolic, approach. First, the solvent-solvent energy is proportional to the squares of the atomic charges, since **z** appears twice in equation 20. The is known to be approximately true for simple ions (e.g., the hydration free energy of divalent ions such as Mg^{2+} is about four times that of monovalent ions such as Na^{+}), and the same ideas hold for *partial* atomic charges as well: solvation effects for partial charges near ± 0.5 (such as the C or O atoms in a ketone or amide group) are much more important (by something like a factor of 25) than are solvation effects for most CH groups, which have partial atomic charges near ± 0.1. Furthermore, in this model, the sign of the charge is not important, so that both $+0.5$ and -0.5 charges have the same (favorable) solvation free energy; this aspect of the model is only approximately true for water, where negative charges (of equivalent size) are somewhat better solvated than the corresponding positive charges.

The second general feature of continuum solvent models that is evident from the above derivation is that the solvent-solvent repulsion or penalty is exactly minus one-half of the (favorable) solute-solvent interaction. This is a general feature of linear-response theories, and again appears to be approximately valid for real solutes as well, at least as judged by microscopic simulations (Simonson, 2002).

Of course, biological solvents are not immersed in infinite dielectrics, and so the model as presented above would be of limited usefulness: for relatively high dielectrics like water, it gives solvation energies that are roughly correct, but electrostatic interactions between solute molecules would be completely screened by the surface charges, completely removing all long-range electrostatic interactions. To a remarkable extent, these deficiencies can be ameliorated by increasing the electrostatic interactions between the surface charge sites (i.e., the second term in equation 16, or equivalently the **A** matrix) by a factor $\varepsilon/(\varepsilon - 1)$. This has the effect of making the screening charge opposite in sign, but slightly smaller in magnitude, than the corresponding atomic charge, so that charge screening by the solvent is not complete. These ideas were originally developed in the context of a quantum mechanical description of the solute charge density and are still widely used for such calculations (Andzelm *et al.*, 1995; Truong and Stefanovich, 1995; Baldridge and Klammt, 1997; Barone and Cossi, 1998; Dolney *et al.*, 2000), primarily because energy gradients can be expressed in a simple fashion. The model can also be used to describe the electrostatic effects of solvation in a force field context (York and Karplus, 1999); here, however, the COSMO model has mostly been overshadowed by Poisson-Boltzmann models (which are more general) and generalized Born models (which are

faster). Nevertheless, the basic physics present in the conductor-like model are present as well in these alternative approaches, which we describe next.

2. *The Poisson-Boltzmann Models*

For an approach that is more general and fundamental than COSMO, but which preserves its continuum character, one can continue to treat with the two-dielectric model illustrated in Fig. 5 but then formulate and solve the corresponding electrostatic model without resorting to any approximations concerning the magnitude of the exterior dielectric constant. An additional advantage of this model is that the effects of mobile co- and counterions in the solvent can be included as in Debye-Hückel theory. For this model (Sharp and Honig, 1990; Tomasi and Persico, 1994; Cramer and Truhlar, 1999), the electrostatic potential ϕ is determined by the Poisson or linearized Poisson-Boltzmann (PB) equation,

$$\nabla\varepsilon(r)\,\nabla\phi(r) - \kappa^2(r)\varepsilon(r)\phi(r) = -4\pi\rho(r) \tag{22}$$

where $\varepsilon(r)$ has the interior or exterior value depending on whether r lies inside or outside the molecular surface (Connolly, 1983); $\kappa(r)$ is the usual Debye-Huckel parameter in the solvent region but is zero in the interior and within an ion exclusion radius from the molecular surface, and ρ is the solute charge distribution. Because of the linearity of equation 22, the work of creating a charge distribution, ρ is simply $\int\rho\phi dV/2$, and the difference in the work of charging required for the calculation of ΔG_{es} is

$$\Delta G_{es} = \frac{1}{2}\sum_i q_i[\phi_{sol}(r_i) - \phi_{vac}(r_i)] \tag{23}$$

where the q_i and r_i are the solute atomic partial charges and atomic coordinates, respectively, and ϕ_{sol} and ϕ_{vac} are the solutions of equation 22 for exterior conditions corresponding to solvent and vacuum, respectively.

Because of the complex geometry of the dielectric and ion-exclusion boundaries, equation 22 must generally be solved numerically. A variety of programs have been developed to carry out such calculations that are beyond the scope of this review. There have been several efforts to develop charge and radius parameters for use in these models based on fitting small molecule solvation energies (Lim *et al.*, 1991; Sitkoff *et al.*, 1994; Nina *et al.*, 1997, 1999). There have also been a number of studies comparing potential energy profiles and surfaces computed with the continuum solvent models to potentials of mean force calculated with explicit solvent for both hydrogen bonding and ion-ion interactions (Ösapay *et al.*, 1996, Luo *et al.*, 1999; Masunov and Lazaridis, 2003). The qualitative features of

the continuum results are generally in good agreement with the explicit solvent results, but the details and fine structure can depend critically on parameterization. When a continuum solvent model is being used in conjunction with a molecular mechanics potential energy function, it is important that the continuum model use the same atomic partial charges and dielectrics as the molecular mechanics potential, or artifacts related to the over-screening or under-screening of charge-charge interactions can occur (Ösapay *et al.*, 1996). Since almost all protein force field charge distributions were designed to be used in a $\varepsilon = 1$ environment, troublesome complications arise when these are combined with continuum solvent models that other dielectrics (such as 2, 4, or 20) for the molecular interior. These difficulties are not insuperable, but a discussion of them is beyond the scope of this review.

Although the Poisson-Boltzmann model has a much lower computational cost than an analogous explicit solvent model, the direct use of the model in molecular dynamics simulations has only recently been shown to be feasible. First, the cost of solving equation 22 at each time step is significantly greater than the cost of evaluating U_{vac}. Second, it is difficult, although not impossible, to calculate the energy gradients that are needed for molecular mechanics (Sharp, 1991; Gilson *et al.*, 1995; Cortis *et al.*, 1996; Ripoll *et al.*, 1996; Luo *et al.*, 2002). Finally, the possibility that solvent-sized and thus high-dielectric-filled voids can appear in the protein interior during a simulation of unfolding creates discontinuities of the energy surface that would be very problematic in dynamics. These considerations have led many workers to consider approximate models, such as the generalized Born theories considered next.

3. The Generalized Born Model

The Born (1920) model computes the electrostatic work required to move a charged sphere from a vacuum environment into a continuous dielectric region. The result is proportional to the square of the charge (as in the COSMO model discussed above), and is inversely proportional to the size of the ion. The basic *ansatz* of generalized Born theory (Still *et al.*, 1990; Bashford and Case, 2000) is to extend these ideas to non-spherical molecules by casting the electrostatic contribution to solvation into the following form:

$$\Delta G_{es} = -\frac{1}{2}\sum_i \sum_j \frac{q_i q_j}{f_{GB}} \left(1 - \frac{1}{\varepsilon_{out}}\right) \tag{24}$$

A simple plausibility argument for the generalized Born model is the following. Imagine a molecule consisting of charges q_i embedded in spheres

of radii, a_i; if the separation r_{ij} between any two spheres is sufficiently large in comparison to the radii, then the solvation free energy can be given by a sum of individual Born terms, and pairwise Coulombic terms:

$$\Delta G_{es} = -\sum_i \frac{q_i^2}{2a_i}\left(1-\frac{1}{\varepsilon_{out}}\right) - \frac{1}{2}\sum_i \sum_{j\neq i} \frac{q_i q_j}{r_{ij}}\left(1-\frac{1}{\varepsilon_{out}}\right) \tag{25}$$

where the factor $(1 - 1/\varepsilon_{out})$ appears in the pairwise terms because the Coulombic interactions are re-scaled by the change of dielectric constant on going from vacuum to solvent.

The goal of generalized Born theory can be thought of as an effort to find a relatively simple analytical formula resembling equation 25 that for real molecular geometries will capture as much as possible the physics of the Poisson equation. Hence, to obtain a formula such as equation 24, we seek a function f_{GB}, such that in the self terms (the first summation in equation 25), f_{GB} acts as an "effective Born radius" while in the pairwise terms (second summation in equation 25) it becomes an effective interaction distance. A variety of functional forms for f_{GB} have been tested (Onufriev *et al.*, 2002), but the most common form chosen is (Still *et al.*, 1990):

$$f_{GB}(r_{ij}, R_i, R_j) = \left[r_{ij}^2 + R_i R_j \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right)\right]^{1/2} \tag{26}$$

Here, the R_i are the effective Born radii of the atoms, which generally depend not only on a_i, the intrinsic atomic radii, but also on the radii and relative positions of all other atoms. Ideally, R_i should be chosen so that if one were to solve the Poisson equation for a single charge placed at the position of atom i, and a dielectric boundary determined by all of the molecule's atoms and their radii, then the self-energy of charge q_i in its reaction field, $q_i\phi(r_i)/2$, would be equal to

$$\frac{q_i^2}{2R_i}\left(1-\frac{1}{\varepsilon_{out}}\right)$$

Such a scheme leads to what have been called "perfect" radii (Onufriev *et al.*, 2002). These, in conjunction with the functional form for the cross terms, equation 26, then yields very accurate cross terms when compared with numerical solutions of the Poisson equation. Thus the key to accurate GB models is the estimation of the effective radii (Onufriev *et al.*, 2002; Lee *et al.*, 2002).

In practice, of course, obtaining "perfect" radii requires an expensive solution to the Poisson-Boltzmann equation, and a faster algorithm for computing the R_i values is needed for efficient molecular dynamics simulations. In the classical electrostatics of a linearly polarizable medium (Jackson, 1999), the work required to assemble a charge distribution can be formulated either in terms of a product of the charge distribution with the electric potential, as in equation 23 above, or in terms of the scalar product of the electric field **E** and the electric displacement **D**:

$$W = \frac{1}{8\pi}\int \mathbf{E}\cdot\mathbf{D}dV \tag{27}$$

We now introduce the essential approximation used in most forms of generalized Born theory: that the electric displacement is Coulombic in form and remains so even as the exterior dielectric is altered from 1 to ε_{out} in the solvation process. In other words, the displacement that is due to the charge of atom i (which for convenience is here presumed to lie on the origin) is,

$$\mathbf{D}_i \approx \frac{q_i\mathbf{r}}{r^3}$$

This is called the *Coulomb field approximation.* In the spherically symmetric case (as in the Born formula) it is exact, but in more complex geometries, there can be substantial deviations. The work of placing a charge q_i at the origin within a molecule whose interior dielectric constant is ε_{in}, surrounded by a medium of dielectric constant ε_{out} and in which no other charges have yet been placed is then,

$$W = \frac{1}{8\pi}\int (\mathbf{D}/\varepsilon)\cdot\mathbf{D}dV \approx \frac{1}{8\pi}\int_{in}\frac{q_i^2}{r^4\varepsilon_{in}}dV + \frac{1}{8\pi}\int_{out}\frac{q_i^2}{r^4\varepsilon_{out}}dV \tag{28}$$

The electrostatic component of the solvation energy is found by taking the difference in W when ε_{out} is changed from 1 to 80:

$$\Delta G_{es} = \frac{-1}{8\pi}\left(1-\frac{1}{80}\right)\int_{out}\frac{q_i^2}{r^4}dV \tag{29}$$

where the contribution that is due to the interior region has canceled in the subtraction.

Comparing equation 29 to the Born formula (equation 16), we conclude that the effective Born radius should be,

$$R_i^{-1} = \frac{1}{4\pi}\int_{out} r^{-4}dV \tag{30}$$

It is often convenient to re-write this in terms of an integration over the interior region, excluding a radius a_i around the origin,

$$R_i^{-1} = a_i^{-1} - \frac{1}{4\pi}\int' r^{-4} dV \tag{31}$$

Here the integral is over the interior of the molecule, excluding regions within a sphere of radius a_i of atom i, and we have used the fact that the integration of r^{-4} over all space outside radius a is simply $4\pi a^{-1}$. In the case of a monatomic ion, where the molecular boundary is simply the sphere of radius a_i, the Born formula is recovered exactly.

The original formulation (Still *et al.*, 1990), as well as some more recent methods (Lee *et al.*, 2002), use a numerical evaluation of this integral. In alternative "pairwise" approaches, the integral is estimated via a sum over pairs of atoms. There are a variety of ways to implement a pairwise formula. In the ACE (analytical continuum electrostatics) model (Schaefer and Karplus, 1996; Schaefer *et al.*, 1998; Calimet *et al.*, 2001), Voronoi volumes for various atom types are determined from a database of protein structures, and these are used to estimate the contribution of each atom to the integral in equation 30. An alternative approach considers the molecule as a set of (overlapping) spheres. The integrals over spheres can then be obtained analytically, including the case that atom j overlaps atom i (Schaefer and Froemmel, 1990; Hawkins *et al.*, 1995, 1996). A straightforward pairwise summation using these ideas would over-count the solute region, since neighboring atoms j themselves overlap with each other. Hawkins *et al.* proposed scaling the neighboring values of a_j as an empirical correction to compensate for this neglect of overlap (Hawkins *et al.*, 1995, 1996); the original correction factors were subsequently optimized in the TINKER package to fit small molecule data. The expression for the generalized Born radii then takes the form,

$$R_i^{-1} = a_i^{-1} - \sum_j H(r_{ij}, S_j a_j) \tag{32}$$

where H is a fairly complex expression and the S_j scaling factors are the additional empirical parameters that account for overlaps. Several groups have adopted this idea, using different training sets to determine the intrinsic radii and scaling factors (Hawkins *et al.*, 1996; Jayaram *et al.*, 1998; Dudek *et al.*, 1998; Tsui and Case, 2000, 2001).

Ghosh *et al.* (1998) have proposed an alternative approach in which the Coulomb field is still used in place of the correct field and Green's theorem is used to convert the volume integral in equation 31 to a surface integral. At this level, the "S-GB" (surface-GB) model is formally identical

to the model outlined above, although there are potential computational advantages in the surface integral approach, especially for large systems and for evaluating gradients. In practice, empirical short-range and long-range corrections are added to improve agreement with numerical Poisson theory. The combination of S-GB with novel models for the non-electrostatic component of solvation appears to offer excellent prospects for many applications, including distinguishing misfolded decoys from more native conformations (Gallicchio *et al.*, 2002; Felts *et al.*, 2002).

4. *Incorporation of Salt Effects*

Generalized Born models have not traditionally considered salt effects, but the model can be extended to low-salt concentrations at the Debye-Huckel level by the following arguments (Srinivasan *et al.*, 1999). The basic idea of the generalized Born approach can be viewed as an interpolation formula between analytical solutions for a single sphere and for widely separated spheres. For the latter, the solvation contribution becomes

$$\Delta G_{el} = -\left(1 - \frac{\exp\left[-\kappa r_{ij}\right]}{\varepsilon_{out}}\right)\frac{q_i q_j}{r_{ij}} \tag{33}$$

where κ is the Debye-Huckel screening parameter. The first term removes the gas-phase interaction energy, and the second term replaces it with a screened Coulomb potential. On the other hand, at short distances, the generalized Born formula is used to reduce to the result for a single spherical ion (Kirkwood, 1934; Tanford and Kirkwood, 1957):

$$\Delta G_{el} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon_{out}}\right)\frac{q^2}{a} - \frac{q^2\kappa}{2\varepsilon_{out}(1 + \kappa b)} \tag{34}$$

where a is the radius of the sphere and b is the radial distance to which salt ions are excluded, so that b-a is the ion exclusion radius. To a close extent, these two limits can be obtained by the simple substitution

$$\left(1 - \frac{1}{\varepsilon_{out}}\right) \Rightarrow \left(1 - \frac{\exp\left[-\kappa f_{GB}\right]}{\varepsilon_{out}}\right) \tag{35}$$

in equation 24. This reduces directly to equation 33 for large distances and to the correct result as r_{ij} goes to zero to terms linear in κ. However, this model does not contain an ion exclusion radius and hence tends to overestimate salt effects (as compared with the usual PB model) by allowing counterions to approach the solute more closely than they should. A simple, *ad hoc* modification that leads to acceptable results can be obtained by scaling κ by 0.73 (Srinivasan *et al.*, 1999).

5. *"Second Generation" Generalized Born Models*

As outlined above, the key to making GB calculations more accurate, in the sense of agreement with PB calculations, is improved estimation of the effective Born radii. In the conventional models, two levels of approximation are involved, the Coulomb field approximation (CFA) used to derive equation 30 and approximations used to estimate the integral in that expression. Recent efforts to improve generalized Born models have looked at both of these issues.

Some work has gone beyond the CFA. Lee *et al.* (2002) have looked at replacing equation 30 with a formula that involves both volume integrals of r^{-4} and also r^{-5}, to provide a correction for deviations from spherical geometry. Requirements of dimensional consistency and reduction to the CFA form in the spherical case (where the CFA is exact) place strong constraints on the functional form. The most promising approach uses the formula

$$R_i^{-1} = P\left(\frac{1}{2a_i^2} - \frac{1}{4\pi}\int r^{-5}dV\right)^{1/2} - a_i^{-1} + \frac{1}{4\pi}\int r^{-4}dV \qquad (36)$$

which has one empirical parameter P, and where the integrals cover the same space as in equation 31. One can estimate the anticipated value for P from the asymptotic limit of a single spherical cavity, where $P = 2\sqrt{2}$ gives the exact Born result. Empirically, values slightly greater than 2.83 appear optimal. When the integrals are evaluated with an accurate grid-based numerical method, equation 36 provides remarkably good agreement with effective Born radii calculated from Poisson equation solutions. Further experimentation with the parameter P, for different dielectric environments, and molecular sizes and shapes is planned.

A second track for improving the GB model involves more accurate estimation of integrals such as that in equation 30. Onufriev *et al.* noted that models that estimate the integral based on a fused-spheres picture (and parameterized for small molecules) tend to understimate the effective radii of atoms deeply buried in proteins (Onufriev *et al.*, 2002). They argued that this happens because interstices between spheres are not counted as part of the low-dielectric interior, even when they are much too small to accommodate a water molecule. In their new method, an adjustable parameter λ is introduced into equation 32:

$$R_i^{-1} = a_i^{-1} - \lambda \sum_j H(r_{ij}, S_j a_j) \qquad (37)$$

This has the effect of keeping the effective radii for atoms near the surface the same as in conventional GB models (where the approximations are

quite good) and to increase the R_i values for more buried atoms. Tests on protein structures showed substantially improved agreement with PB solutions not only for the total solvation energies but for the individual interaction terms that go into a pK_a calculation. Further elaborations of this basic idea appear to give considerable improvement in protein molecular dynamics simulations and folding landscape characterization (Onufriev, Bashford and Case, unpublished).

Another "second track" method (Lee *et al.*, 2002) takes a more systematic approach to improving the accuracy of these integrals. Developing some ideas from the ACE model (Schaefer *et al.*, 1998, 2001), this approach approximates the molecular volume by a set of atom-based, overlapping weight functions; these are chosen so that their sum is nearly constant inside the molecular surface and decays to zero very quickly outside it. In this way, the effective Born radii (and their derivatives) can be computed analytically in a sum-over-atoms approach.

C. Molecular Dynamics Simulations with the Generalized Born Model

It should be emphasized that adding a GB/SA solvation model to an existing protein molecular mechanics potential creates a new combination whose characteristics and quality can be difficult to predict. In principle, all of the parameters of the combined model should be reoptimized, following a protocol similar to that used for force fields designed for explicit solvent simulations. This time-consuming task has not yet been carried out, and early reports of results of molecular dynamics simulations are mostly still in a mode of testing the accuracy of the models.

Because GB simulations can be faster than explicit water simulations, and because the lack of solvent friction allows conformational space to be sampled more quickly, continuum solvent simulations are very attractive in studies of peptide and protein folding, where the qualitative features of the landscapes (including likely folding pathways and intermediates) are still unclear. Models with continuum solvation have reasonably good performance in distinguishing native folds from misfolded "decoys" (Vorobjev *et al.*, 1998; Dominy and Brooks, 2002; Felts *et al.*, 2002; Feig and Brooks, 2002), and searches for native conformations and folding pathways have also yielded promising results (Pande and Rokhsar, 1999; Bursulaya and Brooks, 2000; Mitsutake *et al.*, 2001; Zagrovic *et al.*, 2001, 2002a,b; Simmerling *et al.*, 2002; García and Sanbonmatsu, 2002). Simulations of folded proteins have also been reported (Wagner and Simonson, 1999; Dominy and Brooks, 1999; Calimet *et al.*, 2001; Cornell, *et al.*, 2001; Tsui and Case, 2001). The general aspects of the dynamics, as measured by NMR order parameters or by fluctuations about the mean,

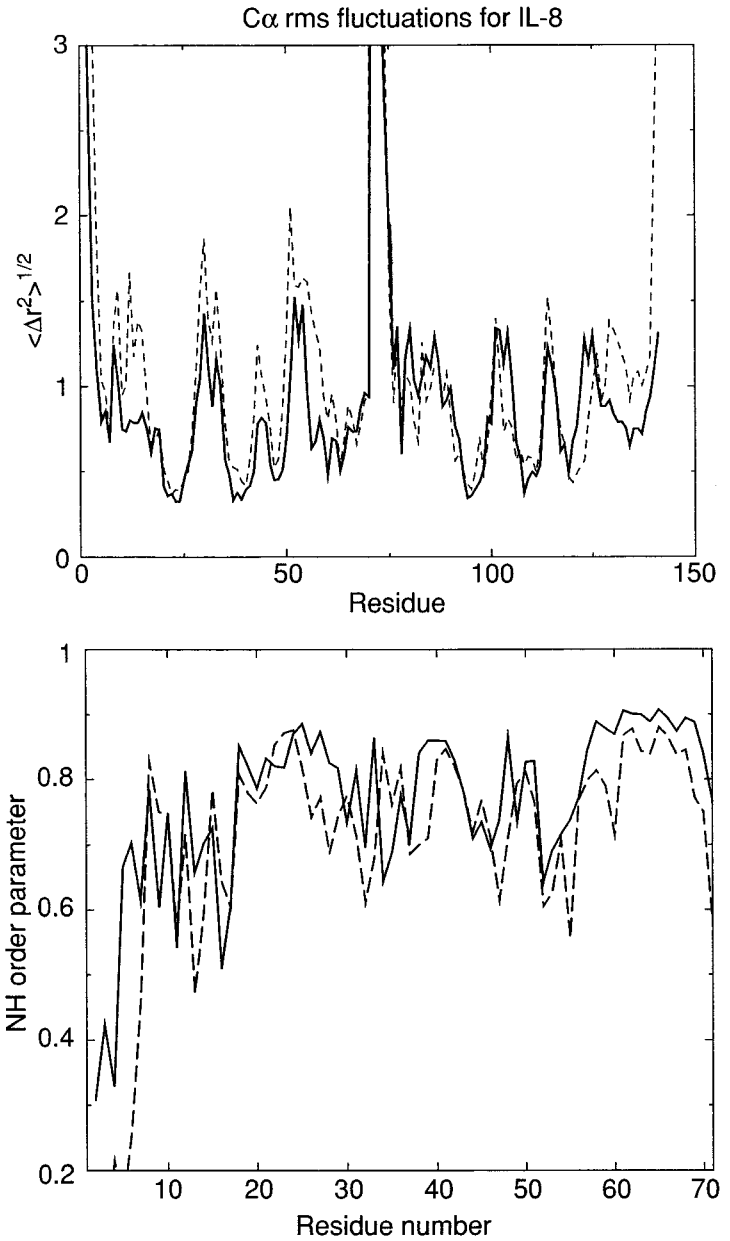


FIG. 9. (Top) Atomic fluctuations of the Cα atoms of interleukin-8 about the mean. Solid line: explicit water simulation; dashed line: generalized Born simulation. Residues 1-71 are from the first monomer, 72-142 from the second monomer. (Bottom) Order parameters (S^2) for N-H motion in interleukin-8, with the same meaning for solid and dashed lines. Adapted from Cornell *et al.* (2001).

are nearly the same as in more conventional, explicit water simulations. This is illustrated in Fig. 9 for interleukin-8. However, deviations from the crystallographic or NMR starting structures have sometimes been noticeably larger than in explicit solvent simulations of comparable length.

To some extent, this might be a consequence of the improved conformational sampling, which lets the simulation more quickly find non-native structures that are energetically favored by the particular force field. But it also seems likely that the current generation of GB models do not have as good a balance between protein-protein and protein-solvent interactions as do the more widely tested explicit solvent models. Further experience and reparameterization will clearly be very useful in sorting this out, but users of the current generation of GB models should be aware that they have definite limitations, many of which are not yet fully understood.

V. Conclusions

An increase in computer power of at least two orders of magnitude should occur over the next decade. Without further research into the accuracy of force-field potentials, future macromolecular modeling may well be limited more by validity of the energy functions, particularly electrostatic terms, than by technical ability to perform the computations. For many calculations related to ligand binding, drug design, and protein structure prediction, accuracy of the underlying potential functions is critical. For example, imagine a hypothetical protein of $N = 100$ residues for which the lowest potential energy basin, representing the family of "native" folds, lies $\Delta E = 5$ kcal/mol below either a compact "misfolded" structure or the massive regime of unfolded states. Then the average acceptable error per residue if we are to be able to distinguish the correct fold is approximately $\Delta E/N^{1/2}$, or 0.5 kcal/mol. The above example has been discussed by Dill (1997) and Abagyan (1997), with a similar analysis advanced by many others. Note that the 0.5 kcal/mol value assumes that error is randomly distributed. Systematic errors in the potential function can further reduce the acceptable error per residue. However, contrary to the pessimistic dogma held by some in the modeling community, this simple error analysis indicates that further work on potential function accuracy is valuable.

Calculation of ligand and drug binding energetics would benefit greatly from potentials of "chemical accuracy"—errors of less than 0.5 kcal/mol in relative binding energies. The promise of structure-based drug design remains largely unmet as a direct result of the inability of current modeling protocols to generate reliable answers. In contrast to gas-phase hydrocarbon potentials, protein potentials have not reached the required level of accuracy, but they are close enough to suggest that improved treatment of the major sources of error will lead to success. In addition,

the argument should be made that we do not adequately understand the basics of protein structure and energetics until quantitative force-field modeling methods become truly predictive tools.

In this article we have sought to provide a description of the current standard practice of protein simulations, providing some of the details of how three of the currently popular protein force fields were developed and tested. Along with this, we have highlighted two directions in which force fields are currently moving, incorporating models of electronic polarizability and simplifying solvent interactions with continuum ideas. This selection of topics is of course heavily influenced by our interests and experience and necessarily leaves out a discussion of other important aspects of protein force fields. Nevertheless, we hope that even this restricted view of the successes and limitations of current force fields will provide a useful perspective on ways in which "chemical accuracy" may be more closely approached in practical protein simulations in the near future.

Acknowledgments

This work was supported by NIH grant GM56531 (to DAC) and NSF grant 9808317 (to JWP). We thank Dr. Alan Grossfield for the WHAM calculations presented in Fig. 5 and Table III and Dr. Pengyu Ren for the AMOEBA polarizable force field results reported in Table IV and Figs. 6 and 7.

References

Abagyan, R. A. (1997). *In* "Computer Simulation of Biomolecular Systems" Vol. 3, (W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkerson, Eds.), pp. 363–394. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Allinger, N. L. (1976). *Adv. Phys. Org. Chem.* **13,** 1.

Allinger, N. L., Yuh, Y. H., and Lii, J.-H. (1989). *J. Am. Chem. Soc.* **111,** 8551–8566.

Allinger, N. L., Chen, K., and Lii, J.-H. (1996). *J. Comput. Chem.* **17,** 642–668.

Amicangelo, J. C., and Armentrout, P. B. (2000). *J. Phys. Chem. A* **104,** 11420–11432.

Andzelm, J., Kölmel, C., and Klamt, A. (1995). *J. Chem. Phys.* **103,** 9312–9320.

Applequist, J. (1983). *J. Math. Phys.* **24,** 736–741.

Applequist, J. (1993). *J. Phys. Chem.* **97,** 6016–6023.

Applequist, J., Carl, J. R., and Fung, K.-K. (1972). *J. Am. Chem. Soc.* **94,** 2952–2960.

Bader, R. F. W. (1990). "Atoms in Molecules: A Quantum Theory." Oxford University Press, New York.

Baldridge, K., and Klamt, A. (1997). *J. Chem. Phys.* **106,** 6622–6633.

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

Barnes, P., Finney, J. L., Nicholas, J. D., and Quinn, J. E. (1979). *Nature* **282,** 459–464.

Barone, V., and Cossi, M. (1998). *J. Phys. Chem. A* **102,** 1995–2001.

Bashford, D., and Case, D. A. (2000). *Annu. Rev. Phys. Chem.* **51,** 129–152.

Bayly, C., Cieplak, P., Cornell, W., and Kollman, P. A. (1993). *J. Phys. Chem.* **97,** 10269–10280.

Beachy, M. D., Chapman, D., Murphy, R. B., Halgren, T. A., and Friesner, R. A. (1997). *J. Am. Chem. Soc.* **119,** 5908–5920.

Becker, O., MacKerell, A. D., Jr., Roux, B., and Watanabe, M. (2001). "Computational Biochemistry and Biophysics." Marcel Dekker, New York.

Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987). *J. Phys. Chem.* **91,** 6269–6271.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981). *In* "Intermolecular Forces." (B. Pullmann, Ed.), pp. 331–342. D. Reidel Publishing Company, Dordrecht, The Netherlands.

Bernal, J. D., and Fowler, R. H. (1933). *J. Chem. Phys.* **1,** 515–548.

Bernardo, D. N., Ding, Y., Krogh-Jespersen, K., and Levy, R. M. (1994). *J. Phys. Chem.* **98,** 4180–4187.

Beroza, P., and Case, D. A. (1998). *Methods Enzymol.* **295,** 170–189.

Bode, K. A., and Applequist, J. (1996). *J. Phys. Chem.* **100,** 17820–17824.

Born, M. (1920). *Z. Phys.* **1,** 45–48.

Brdarski, S., Astrand, P. O., and Karlström, G. (2000). *Theoret. Chem. Acc.* **105,** 7–14.

Breneman, C. M., and Wiberg, K. B. (1990). *J. Comput. Chem.* **11,** 361–373.

Brobjer, J. T., and Murrell, J. N. (1982). *J. Chem. Soc. Faraday Trans. 2* **78,** 1853–1870.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). *J. Comput. Chem.* **4,** 187–217.

Buckingham, A. D., and Fowler, P. W. (1985). *Can. J. Chem.* **63,** 2018–2025.

Burkert, U., and Allinger, N. L. (1982). "Molecular Mechanics." American Chemical Society, Washington, D.C.

Burnham, C. J., Li, J., Xantheas, S. S., and Leslie, M. (1999). *J. Chem. Phys.* **110,** 4566–4581.

Burnham, C. J., and Xantheas, S. S. (2002a). *J. Chem. Phys.* **116,** 1479–1510.

Burnham, C. J., and Xantheas, S. S. (2002b). *J. Chem. Phys.* **116,** 5115–5124.

Bursulaya, B. D., and Brooks, C. L., III. (2000). *J. Phys. Chem. B* **104,** 12378–12383.

Caldwell, J. W., and Kollman, P. A. (1995). *J. Phys. Chem.* **99,** 6208–6219.

Calimet, N., Schaefer, M., and Simonson, T. (2001). *Proteins* **45,** 144–158.

Carlson, M. J. (2000). "BUFF: A Biological Universal Forcefield Derived from Quantum Mechanics." Ph.D. thesis, California Institute of Technology.

Ceccarelli, M., and Marchi, M. (1997). *J. Phys. Chem. B* **101,** 2105–2108.

Chen, J. L., Noodleman, L., Case, D. A., and Bashford, D. (1994). *J. Phys. Chem.* **98,** 11059–11068.

Chessari, G., Hunter, C. A., Low, C. M. R., Packer, M. J., Vinter, J. G., and Zonta, C. (2002). *Chem. Eur. J.* **8,** 2860–2867.

Chipot, C., Angyan, J. G., Maigret, B., and Scheraga, H. A. (1993). *J. Phys. Chem.* **97,** 9788–9796.

Cieplak, P., Caldwell, J., and Kollman, P. (2001). *J. Comput. Chem.* **22,** 1048–1057.

Clark, M., Cramer, R. D., and Van Opdenbosch, N. (1989). *J. Comput. Chem.* **10,** 982–1012.

Colonna, F., and Evleth, E. (1992). *J. Comput. Chem.* **13,** 1234–1245.

Connolly, M. L. (1983). *Science* **221,** 709–713.

Cornell, W., Abseher, R., Nilges, M., and Case, D. A. (2001). *J. Mol. Graph. Model.* **19,** 136–145.

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., and Kollman, P. A. (1995). *J. Am. Chem. Soc.* **117,** 5179–5197.

Cornell, W. D., Cieplak, P., Bayly, C. I., and Kollman, P. A. (1993). *J. Am. Chem. Soc.* **115,** 9620–9631.

Corongiu, G. (1992). *Int. J. Quant. Chem.* **42,** 1209–1235.

Cortis, C. M., Langlois, J.-M., Beachy, M. D., and Friesner, R. A. (1996). *J. Chem. Phys.* **105,** 5472–5484.

Cramer, C. J. (2002). "Essentials of Computational Chemistry: Theory and Models." John Wiley and Sons, New York.

Cramer, C. J., and Truhlar, D. G. (1999). *Chem. Rev.* **99,** 2161–2200.

Daggett, V. (2002). *Acc. Chem. Res.* **35,** 422–429.

Daggett, V., and Levitt, M. (1993). *Annu. Rev. Biophys. Biomol. Struct.* **22,** 353–380.

Damm, W., and van Gunsteren, W. F. (2000). *J. Comput. Chem.* **21,** 774–787.

Dang, L. X., and Chang, T.-M. (1997). *J. Chem. Phys.* **106,** 8149–8159.

Darden, T. A. (2001). *In* "Computational Biochemistry and Biophysics." (O. Becker, A. D. MacKerell, Jr., B. Roux, and M. Watanabe, Eds.), pp. 91–114. Marcel Dekker, New York.

Darden, T., Pearlman, D., and Pedersen, L. G. (1998). *J. Chem. Phys.* **109,** 10921–10935.

Derreumaux, P., and Vergoten, G. (1995). *J. Chem. Phys.* **102,** 8586–8605.

de Vries, A. H., van Duijnen, P. T., Zijlstra, R. W. J., and Swart, M. (1997). *J. Electron Spect. Related Phenom.* **86,** 49–55.

Dill, K. A. (1997). *J. Biol. Chem.* **272,** 701–704.

Dixon, R. W., and Kollman, P. A. (1997). *J. Comput. Chem.* **18,** 1632–1646.

Dolney, D. M., Hawkins, G. D., Winget, P., Liotard, D. A., Cramer, C. J., and Truhlar, D. G. (2000). *J. Comput. Chem.* **21,** 340–366.

Dominy, B. N., and Brooks, C. L., III. (1999). *J. Phys. Chem. B* **103,** 3765–3773.

Dominy, B. N., and Brooks, C. L., III. (2002). *J. Comput. Chem.* **23,** 147–160.

Dougherty, D. A. (1995). *Science* **271,** 163–168.

Dudek, M. J., Ramnarayan, K., and Ponder, J. W. (1998). *J. Comput. Chem.* **19,** 548.

Dunning, T. H. (2000). *J. Phys. Chem. A* **104,** 9062–9080.

Dykstra, C. E. (1988). *J. Comput. Chem.* **9,** 476–487.

Dykstra, C. E. (1989). *J. Am. Chem. Soc.* **111,** 6168–6174.

Dykstra, C. E. (1993). *Chem. Rev.* **93,** 2339–2353.

Feig, M., and Brooks, C. L., III. (2002). *Proteins* **49,** 232–245.

Felts, A. K., Gallicchio, E., Wallqvist, A., and Levy, R. M. (2002). *Proteins* **48,** 404–422.

Ferguson, D. M. (1995). *J. Comput. Chem.* **16,** 501–511.

Foloppe, N., and MacKerell, Jr., A. D. (2000). *J. Comput. Chem.* **21,** 86–104.

Fox, T., and Kollman, P. A. (1998). *J. Phys. Chem. B* **102,** 8070–8079.

Gallicchio, E., Zhang, L. Y., and Levy, R. M. (2002). *J. Comput. Chem.* **23,** 517–529.

García, A. E., and Sanbonmatsu, K. Y. (2001). *Proteins* **42,** 345–354.

García, A. E., and Sanbonmatsu, K. Y. (2002). *Proc. Natl. Acad. Sci. USA* **99,** 2782–2787.

Ghosh, A., Rapp, C. S., and Friesner, R. A. (1998). *J. Phys. Chem. B* **102,** 10983–10990.

Gilson, M. K., McCammon, J. A., and Madura, J. D. (1995). *J. Comput. Chem.* **16,** 1081–1095.

Gough, C., DeBolt, S., and Kollman, P. A. (1992). *J. Comput. Chem.* **13,** 963–970.

Gresh, N. (1997a). *J. Chim. Phys. PCB* **94,** 1365–1416.

Gresh, N. (1997b). *J. Phys. Chem. A* **101,** 8680–8694.

Gresh, N., Guo, H., Salahub, D. R., Roques, B. P., and Kafafi, S. A. (1999). *J. Am. Chem. Soc.* **121,** 7885–7894.

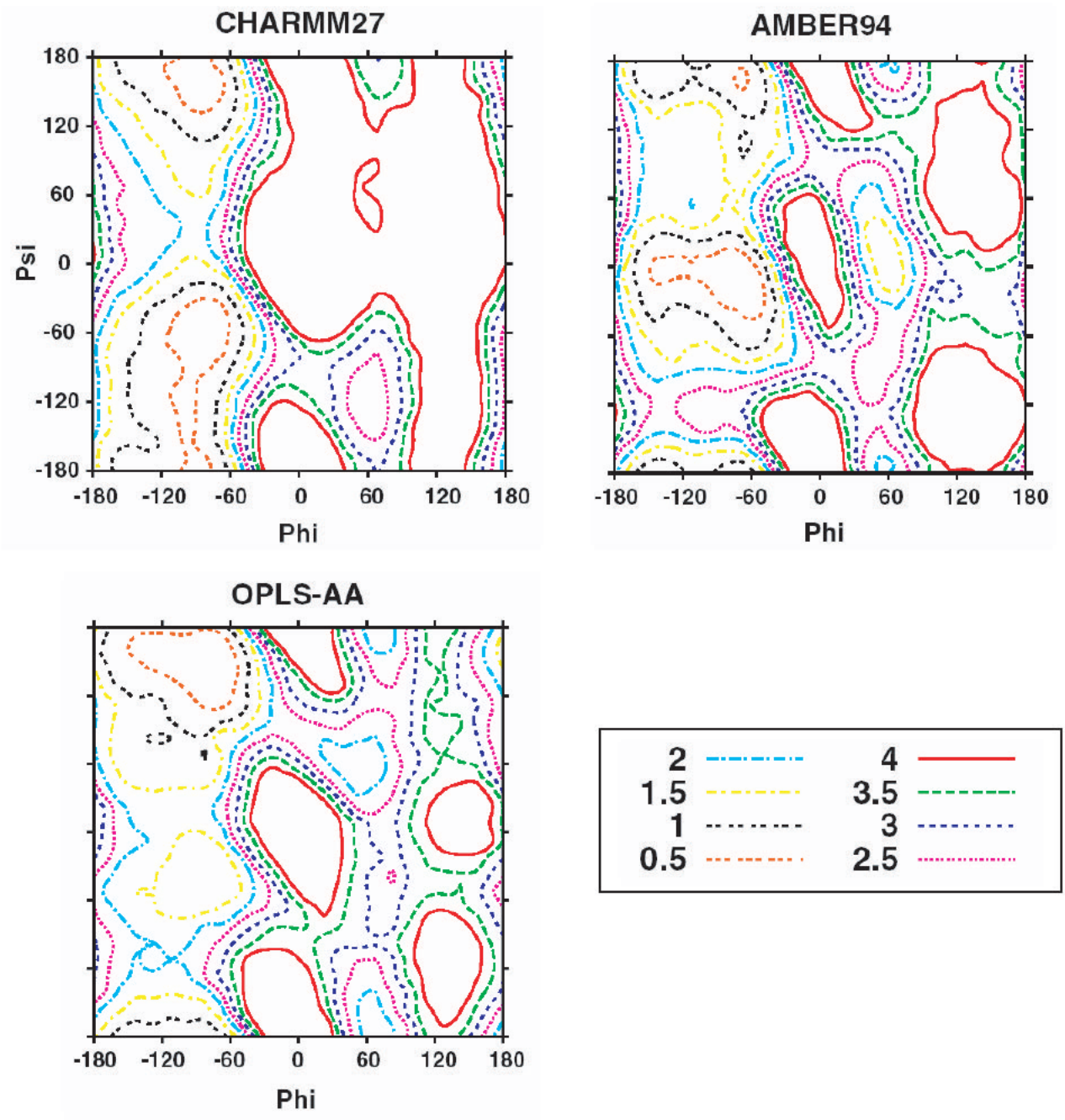
Gresh, N., Tiraboschi, G., and Salahub, D. R. (1998). *Biopolymers* **45,** 405–425.
Guillot, B., and Guissani, Y. (2001). *J. Chem. Phys.* **114,** 6720–6733.
Guo, H., Gresh, N., Roques, B. P., and Salahub, D. R. (2000). *J. Phys. Chem. B* **104,** 9746–9754.
Hagler, A., Euler, E., and Lifson, S. (1974). *J. Am. Chem. Soc.* **96,** 5319–5327.
Hagler, A., and Lifson, S. (1974). *J. Am. Chem. Soc.* **96,** 5327–5335.
Halgren, T. A. (1992). *J. Am. Chem. Soc.* **114,** 7827–7843.
Halgren, T. A. (1996a). *J. Comput. Chem.* **17,** 490–519.
Halgren, T. A. (1996b). *J. Comput. Chem.* **17,** 520–552.
Halgren, T. A. (1996c). *J. Comput. Chem.* **17,** 553–586.
Halgren, T. A. (1996d). *J. Comput. Chem.* **17,** 616–641.
Halgren, T. A. (1999a). *J. Comput. Chem.* **20,** 720–729.
Halgren, T. A. (1999b). *J. Comput. Chem.* **20,** 730–748.
Halgren, T. A., and Damm, W. (2001). *Curr. Opin. Struct. Biol.* **11,** 236–242.
Halgren, T. A., and Nachbar, R. B. (1996). *J. Comput. Chem.* **17,** 587–615.
Harvey, S., and McCammon, J. A. (1987). "Dynamics of Proteins and Nucleic Acids." Cambridge University Press, Cambridge.
Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1995). *Chem. Phys. Lett.* **246,** 122–129.
Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996). *J. Phys. Chem.* **100,** 19824–19839.
Hermans, J., Berendsen, H. J. C., van Gunsteren, W. F., and Postma, J. P. M. (1984). *Biopolymers* **23,** 1513–1518.
Hermida-Ramón, J. M., Brdarski, S., Karlström, G., and Berg, U. (2003). *J. Comput. Chem.* **24,** 161–176.
Hu, H., Elstner, M., and Hermans, J. (2003). *Proteins* **50,** 451–463.
Huang, N., and MacKerell, A. D., Jr. (2002). *J. Phys. Chem. A* **106,** 7820–7827.
Hünenberger, P. H., and van Gunsteren, W. F. (1997). *In* "Computer Simulations of Biomolecular Systems" Vol. 3, (W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, Eds.), pp. 3–82. Kluwer Academic Publishers, Dordrecht, The Netherlands.
Jackson, J. D. (1999). "Classical Electrodynamics." 3rd edition. John Wiley and Sons, New York.
Jacobson, M. P., Kaminski, G. A., Friesner, R. A., and Rapp, C. S. (2002). *J. Phys. Chem. B* **106,** 11673–11680.
Jayaram, B., Liu, Y., and Beveridge, D. L. (1998). *J. Phys. Chem.* **109,** 1465–1471.
Jorgensen, W. L. (1981). *J. Am. Chem. Soc.* **103,** 335–340.
Jorgensen, W. L. (1998). *In* "Encyclopedia of Computational Chemistry" (P. von R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, and H. F. Schaefer III, Eds.), pp. 1986–1989. John Wiley and Sons, Chichester.
Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). *J. Chem. Phys.* **79,** 926–935.
Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). *J. Am. Chem. Soc.* **118,** 11225–11236.
Jorgensen, W. L., and Swenson, C. J. (1985). *J. Am. Chem. Soc.* **107,** 569–578.
Jorgensen, W. L., and Tirado-Rives, J. (1988). *J. Am. Chem. Soc.* **110,** 1657–1671.
Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001). *J. Phys. Chem. B* **105,** 6474–6487.
Kaminski, G., and Jorgensen, W. L. (1996). *J. Phys. Chem.* **100,** 18010–18013.
Kaminski, G. A., Stern, H. A., Berne, B. J., Friesner, R. A., Cao, Y. X., Murphy, R. B., Zhou, R., and Halgren, T. A. (2002). *J. Comput. Chem.* **23,** 1515–1531.
Kirkwood, J. G. (1934). *J. Chem. Phys.* **2,** 351–361.

Kitao, O., and Ogawa, T. (2003). *Mol. Phys.* **101,** 3–17.
Kitson, D. H., and Hagler, A. T. (1988). *Biochemistry* **27,** 5246–5257.
Klamt, A., and Schüürmann, G. (1993). *J. Chem. Soc. Perkin Trans.* **2,** 799–805.
Koch, U., and Egert, E. (1995). *J. Comput. Chem.* **16,** 937–944.
Kollman, P., Dixon, R., Cornell, W., Fox, T., Chipot, C., and Pohorille, A. (1997). *In* "Computer Simulations of Biomolecular Systems," Vol. 3 (W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, Eds.), pp. 83–96. Kluwer Academic Publishers, Dordrecht, The Netherlands.
Kong, Y. (1997). "Multipole Electrostatic Methods for Protein Modeling with Reaction Field Treatment." Ph.D. thesis, Washington University School of Medicine.
Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1995). *J. Comput. Chem.* **16,** 1339–1350.
Langley, C. H., and Allinger, N. L. (2002). *J. Phys. Chem. A* **106,** 5638–5652.
Leach, A. R. (2001). "Molecular Modelling. Principles and Applications." Second Edition. Prentice-Hall, Harlow, England.
Lee, M. S., Salsbury, F. R., Jr., and Brooks, C. L. III. (2002). *J. Chem. Phys.* **116,** 10606–10614.
Levitt, M. (1983). *J. Mol. Biol.* **168,** 595–620.
Levitt, M., Hirshberg, M., Sharon, R., and Daggett, V. (1995). *Comp. Phys. Commun.* **91,** 215–231.
Levitt, M., Hirschberg, M., Sharon, R., Laidig, K. E., and Daggett, V. (1997). *J. Phys. Chem. B* **101,** 5051–5061.
Levitt, M., and Sharon, R. (1988). *Proc. Natl. Acad. Sci.* **85,** 7557–7561.
Lifson, S., and Warshel, A. (1969). *J. Chem. Phys.* **49,** 5116–5129.
Lii, J.-H., and Allinger, N. (1991). *J. Comput. Chem.* **12,** 186–199.
Lim, C., Bashford, D., and Karplus, M. (1991). *J. Phys. Chem.* **95,** 5610–5620.
Liu, Y.-P., Kim, K., Berne, B. J., Friesner, R. A., and Rick, S. W. (1998). *J. Chem. Phys.* **108,** 4739–4755.
Luo, R., David, L., and Gilson, M. K. (2002). *J. Comput. Chem.* **23,** 1244–1253.
Luo, R., David, L., Hung, H., Devaney, J., and Gilson, M. K. (1999). *J. Phys. Chem. B* **103,** 727–736.
Lybrand, T. P., and Kollman, P. A. (1985). *J. Chem. Phys.* **83,** 2923–2933.
Mackay, D. H. J., Berens, P. H., Wilson, K. R., and Hagler, A. T. (1984). *Biophys. J.* **46,** 229–248.
MacKerell, A. D., Jr. (2001). *In* "Computational Biochemistry and Biophysics" (O. Becker, A. D. MacKerell, Jr., B. Roux, and M. Watanabe, Eds.), pp. 7–38. Marcel Dekker, New York.
MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunback, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E. III, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998). *J. Phys. Chem. B* **102,** 3586–3616.
Mahoney, M. W., and Jorgensen, W. L. (2000). *J. Chem. Phys.* **112,** 8910–8922.
Mahoney, M. W., and Jorgensen, W. L. (2001). *J. Chem. Phys.* **115,** 10758–10768.
Mannfors, B. E., Mirkin, N. G., Palmo, K., and Krimm, S. (2003). *J. Phys. Chem. A.* **107,** 1825–1832.
Maple, J. R., Hwang, M.-J., Jalkanen, K. J., Stockfisch, T. P., and Hagler, A. T. (1998). *J. Comput. Chem.* **18,** 430–458.

Maple, J. R., Hwang, M.-J., Stockfisch, T. P., Dinur, U., and Hagler, A. T. (1994). *J. Comput. Chem.* **15,** 162–182.
Mark, P., and Nilsson, L. (2002). *J. Comput. Chem.* **23,** 1211–1219.
Masella, M., Gresh, N., and Flament, J.-P. (1998). *J. Chem. Soc. Faraday Trans.* **94,** 2745–2753.
Masunov, A., and Lazaridis, T. (2003). *J. Am. Chem. Soc.* **125,** 1722–1730.
Matsuoka, O., Clementi, E., and Yoshimine, M. (1976). *J. Chem. Phys.* **64,** 1351–1361.
Matta, C. F., and Bader, R. F. W. (2000). *Proteins* **40,** 310–329.
Maxwell, D. W., Tirado-Rives, J., and Jorgensen, W. L. (1995). *J. Comput. Chem.* **16,** 984–1010.
Miller, K. J. (1990). *J. Am. Chem. Soc.* **112,** 8533–8542.
Mitchell, P. J., and Fincham, D. (1993). *J. Phys.: Condens. Matter* **5,** 1031–1038.
Mitsutake, A., Sugita, Y., and Okamoto, Y. (2001). *Biopolymers* **60,** 96–123.
Momany, F. A., McGuire, R. F., Burgess, A. W., and Scheraga, H. A. (1975). *J. Phys. Chem.* **79,** 2361–2381.
Némethy, G., Pottle, M. S., and Scheraga, H. A. (1983). *J. Phys. Chem.* **87,** 1883–1887.
Neria, E., Fischer, S., and Karplus, M. (1996). *J. Chem. Phys.* **105,** 1902–1921.
Niketic, S. R., and Rasmussen, K. (1997). "The Consistent Force Field: A Documentation." Springer-Verlag, New York.
Nina, M., Beglov, D., and Roux, B. (1997). *J. Phys. Chem. B* **101,** 5239–5248.
Nina, M., Im, W., and Roux, B. (1999). *Biophys. Chem.* **78,** 89–96.
Nymand, T. M., and Linse, P. (2000). *J. Chem. Phys.* **112,** 6152–6160.
Oie, T., Maggiora, G. M., and Christoffersen, R. E. (1981). *Int. J. Quant. Chem., Quant. Biol. Symp.* **8,** 1–47.
Onufriev, A., Bashford, D., and Case, D. A. (2000). *J. Phys. Chem. B* **104,** 3712–3720.
Onufriev, A., Case, D. A., and Bashford, D. (2002). *J. Comput. Chem.* **23,** 1297–1304.
Ösapay, K., Young, W. S., Bashford, D., Brooks, C. L. III, and Case, D. A. (1996). *J. Phys. Chem.* **100,** 2698–2705.
Palmo, K., and Krimm, S. (1998). *J. Comput. Chem.* **19,** 754–768.
Palmo, K., Mannfors, B., Mirkin, N. G., and Krimm, S. (2003). *Biopolymers* **68,** 383–394.
Pande, V. S., and Rokhsar, D. S. (1999). *Proc. Natl. Acad. Sci. USA* **96,** 9062–9067.
Popelier, P. (2000). "Atoms in Molecules: An Introduction." Pearson Education Ltd., Essex, England.
Popelier, P. L. A., Joubert, L., and Kosov, D. S. (2001). *J. Phys. Chem. A* **105,** 8254–8261.
Popelier, P. L. A., and Stone, A. J. (1994). *Mol. Phys.* **82,** 411–425.
Price, D. J., and Brooks, C. L. III. (2002). *J. Comput. Chem.* **23,** 1045–1057.
Price, S. L. (2000). *Rev. Comput. Chem.* **14,** 225–289.
Price, S. L., Stone, A. J., and Alderton, M. (1984). *Mol. Phys.* **52,** 987–1001.
Rappé, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A., and Skiff, W. M. (1992). *J. Am. Chem. Soc.* **114,** 10024–10035.
Rappé, A. K., and Goddard, W. A. (1991). *J. Phys. Chem.* **95,** 3358–3363.
Reiher, W. E., III, (1985). "Theoretical Studies of Hydrogen Bonding." Ph.D. thesis, Harvard University.
Ren, P., and Ponder, J. W. (2002). *J. Comput. Chem.* **23,** 1497–1506.
Ren, P., and Ponder, J. W. (2003). *J. Phys. Chem. B* **107,** 5933–5947.
Reynolds, C. A., Essex, J. W., and Richards, W. G. (1992). *J. Am. Chem. Soc.* **114,** 9075–9079.
Rick, S. W., and Stuart, S. J. (2002). *Rev. Comput. Chem.* **18,** 89–146.
Rick, S. W., Stuart, S. J., and Berne, B. J. (1994). *J. Chem. Phys.* **101,** 6141–6156.

Ripoll, D. R., Vorbjev, Y. N., Liwo, A., Vila, J. A., and Scheraga, H. A. (1996). *J. Mol. Biol.* **264,** 770–783.
Rizzo, R. C., and Jorgensen, W. L. (1999). *J. Am. Chem. Soc.* **121,** 4827–4836.
Roterman, I. K., Lambert, M. H., Gibson, K. D., and Scheraga, H. A. (1989). *J. Biomol. Struct. Dyn.* **7,** 421–453.
Roux, B., and Bernèche, S. (2002). *Biophys. J.* **82,** 1681–1684.
Roux, B., and Simonson, T. (1999). *Biophys. Chem.* **78,** 1–20.
Saint-Martin, H., Hernández-Cobos, J., Bernal-Uruchurtu, M. I., Ortega-Blake, I., and Berendsen, H. J. C. (2000). *J. Chem. Phys.* **113,** 10899–10912.
Schaefer, M., Bartels, C., and Karplus, M. (1998). *J. Mol. Biol.* **284,** 835.
Schaefer, M., Bartels, C., Leclerc, F., and Karplus, M. (2001). *J. Comput. Chem.* **22,** 1857–1879.
Schaefer, M., and Froemmel, C. (1990). *J. Mol. Biol.* **216,** 1045–1066.
Schaefer, M., and Karplus, M. (1996). *J. Phys. Chem.* **100,** 1578–1599.
Schaefer, M., van Vlijmen, H. W. T., and Karplus, M. (1998). *Adv. Protein Chem.* **51,** 1–57.
Schlick, T. (2002). "Molecular Modeling and Simulation: An Interdisciplinary Guide." Springer-Verlag, New York.
Scott, W. R., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krüger, P., and van Gunsteren, W. F. (1999). *J. Phys. Chem. A* **103,** 3596–3607.
Sharp, K. (1991). *J. Comput. Chem.* **12,** 454–468.
Sharp, K. A., and Honig, B. (1990). *Annu. Rev. Biophys. Biophys. Chem.* **19,** 301–332.
Sharp, K. A., Kumar, S., Rossky, P. J., Friedman, R. A., and Honig, B. (1996). *J. Phys. Chem.* **100,** 14166–14177.
Simmerling, C., Strockbine, B., and Roitberg, A. E. (2002). *J. Am. Chem. Soc.* **124,** 11258–11259.
Simonson, T. (2002). *Proc. Natl. Acad. Sci. USA* **99,** 6544–6549.
Simonson, T., and Brünger, A. T. (1994). *J. Phys. Chem.* **98,** 4683–4694.
Sitkoff, D., Sharp, K. A., and Honig, B. (1994). *J. Phys. Chem.* **98,** 1978–1988.
Smith, W. (1998). *CCP5 Newsletter* **46,** 18–30.
Sokalski, W. A., Keller, D. A., Ornstein, R. L., and Rein, R. (1993). *J. Comput. Chem.* **14,** 970–976.
Sokalski, W. A., and Poirier, R. A. (1983). *Chem. Phys. Lett.* **98,** 86–92.
Sprik, M., and Klein, M. L. (1988). *J. Chem. Phys.* **89,** 7556–7560.
Srinivasan, J., Trevathan, M. W., Beroza, P., and Case, D. A. (1999). *Theor. Chem. Acc.* **101,** 426–434.
Steinbach, P. J., and Brooks, B. R. (1994). *J. Comput. Chem.* **15,** 667–683.
Stern, H. A., Kaminski, G. A., Banks, J. L., Zhou, R., Berne, B. J., and Friesner, R. A. (1999). *J. Phys. Chem. B* **103,** 4730–4737.
Stern, H. A., Rittner, F., Berne, B. J., and Friesner, R. A. (2001). *J. Am. Chem. Phys.* **115,** 2237–2251.
Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990). *J. Chem. Phys.* **115,** 6127–6129.
Stillinger, F. H., and David, C. W. (1978). *J. Chem. Phys.* **69,** 1473–1484.
Stillinger, F. H., and Rahman, A. (1974). *J. Chem. Phys.* **60,** 1545–1557.
Stone, A. J. (1981). *Chem. Phys. Lett.* **83,** 233–238.
Stone, A. J. (1996). "The Theory of Intermolecular Forces." Oxford University Press, New York.
Stouch, T. R., and Williams, D. E. (1993). *J. Comput. Chem.* **14,** 858–866.
Stout, J. M., and Dykstra, C. E. (1998). *J. Phys. Chem. A* **102,** 1576–1582.

Sunner, J., Nishizawa, K., and Kebarle, P. (1981). *J. Phys. Chem.* **85,** 1814–1820.
Tanford, C., and Kirkwood, J. G. (1957). *J. Am. Chem. Soc.* **79,** 5333–5339.
Teeter, M. M., and Case, D. A. (1990). *J. Phys. Chem.* **94,** 8091–8097.
Thole, B. T. (1981). *Chem. Phys.* **59,** 341–350.
Tirado-Rives, J., and Jorgensen, W. L. (1990). *J. Am. Chem. Soc.* **112,** 2773–2781.
Tomasi, J., and Persico, M. (1994). *Chem. Rev.* **94,** 2027–2094.
Tomimoto, M., and Go, N. (1995). *J. Phys. Chem.* **99,** 563–577.
Toukmaji, A., Sagui, C., Board, J., and Darden, T. (2000). *J. Chem. Phys.* **113,** 10913–10927.
Truong, T. H., and Stefanovich, E. V. (1995). *J. Chem. Phys.* **103,** 3709–3717.
Tsui, V., and Case, D. A. (2000). *J. Am. Chem. Soc.* **122,** 2489–2498.
Tsui, V., and Case, D. A. (2001). *Biopolymers (Nucl. Acid. Sci.),* **56,** 275–291.
van der Vaart, A., Bursulaya, B. D., Brooks III, C. L., and Merz, K. M. J. (2000). *J. Phys. Chem. B* **104,** 9554–9563.
van Duijnen, P. T., and Swart, M. (1998). *J. Phys. Chem. A* **102,** 2399–2407.
van Gunsteren, W. F., and Berendsen, H. J. C. (1987). "Groningen Molecular Simulation (GROMOS) Library Manual." University of Groningen, The Netherlands.
van Gunsteren, W. F., Daura, X., and Mark, A. E. (1998). *In* "Encyclopedia of Computational Chemistry" (R. von, P. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, and H. F. Schaefer III, Eds.), pp. 1211–1216. John Wiley, Chichester.
Vargas, R., Garza, J., Dixon, D. A., and Hay, B. P. (2000). *J. Am. Chem. Soc.* **122,** 4750–4755.
Vedani, A. (1988). *J. Comput. Chem.* **9,** 269–280.
Veenstra, D., Ferguson, D., and Kollman, P. A. (1992). *J. Comput. Chem.* **13,** 971–978.
Vesely, F. J. (1977). *J. Comput. Phys.* **24,** 361–371.
Vorobjev, Y. N., Almagro, J. C., and Hermans, J. (1998). *Proteins* **32,** 399–413.
Wagner, F., and Simonson, T. (1999). *J. Comput. Chem.* **20,** 322–335.
Wang, J., Cieplak, P., and Kollman, P. A. (2000). *J. Comput. Chem.* **21,** 1049–1074.
Weiner, P. K., and Kollman, P. A. (1981). *J. Comput. Chem.* **2,** 287–303.
Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. Jr., and Weiner, P. (1984). *J. Am. Chem. Soc.* **106,** 765–784.
Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986). *J. Comput. Chem.* **7,** 230–252.
Williams, D. E. (1988). *J. Comput. Chem.* **9,** 745–763.
Williams, D. E. (1991). *Rev. Comput. Chem.* **2,** 219–271.
Williams, D. E. (2001). *J. Comput. Chem.* **22,** 1154–1166.
Winn, P. J., Ferenczy, G. G., and Reynolds, C. A. (1997). *J. Phys. Chem. A* **101,** 5437–5445.
York, D. M., and Karplus, M. (1999). *J. Phys. Chem. A* **103,** 11060–11079.
Young, D. M. (1971). "Iterative Solution of Large Linear Systems." Academic Press, New York.
Yu, H., Hansson, T., and van Gunsteren, W. F. (2003). *J. Chem. Phys.* **118,** 221–234.
Zacharias, N., and Dougherty, D. A. (2002). *Trends Pharm. Sci.* **23,** 281–287.
Zagrovic, B., Snow, C. D., Khaliq, S., Shirts, M. R., and Pande, V. S. (2002a). *J. Mol. Biol.* **323,** 153–164.
Zagrovic, B., Snow, C. D., Shirts, M. R., and Pande, V. S. (2002b). *J. Mol. Biol.* **323,** 927–937.
Zagrovic, B., Sorin, E. J., and Pande, V. (2001). *J. Mol. Biol.* **313,** 151–169.



PONDER AND CASE, FIG. 5. Comparison of ϕ-ψ free energy plots from 2-D WHAM analysis of umbrella sampling trajectories for alanine dipeptide as computed with the AMBER *ff94*, CHARMM27 and OPLS-AA pairwise force fields. The contours shown are at 0.5 kcal/mol intervals from the global minimum of each Ramachandran map.