Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations

PENGYU REN, JAY W. PONDER
Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

Received 5 March 2002; Accepted 26 April 2002

Abstract: A protocol is described for the treatment of molecular polarization in force field calculations. The resulting model is consistent in that both inter- and intramolecular polarization are handled within a single scheme. An analytical formula for removing intramolecular polarization from a set of atomic multipoles for an arbitrary static structure or conformation is given. With the help of the intramolecular polarization, these permanent atomic multipoles can then be applied in modeling alternative conformations of a molecule. Equipped with this simple technique, one can derive transferable electrostatic parameters for peptides and proteins using flexible model compounds such as dipeptides. The proposed procedure is tested for its ability to describe the electrostatic potential around various configurations of the *N*-methylacetamide dimer. The effect of different intramolecular polarization schemes on the accuracy of a force field model of the electrostatic potential of alanine dipeptide is investigated. A group-based scheme for including direct intramolecular polarization is shown to be most successful in accounting for the conformational dependence of electrostatic potentials.

 J Comput Chem 23: 1497–1506, 2002

Key words: force fields; parameterization; empirical potential functions; polarizability; induction

Introduction

Traditional force fields lack the ability to adapt to environmental changes due to their use of a fixed electrostatic model and the absence of an explicit polarization term. While *ab initio* quantum theory has become a useful tool for deriving electrostatic parameters for small molecules in the gas phase,[1,2] these parameters are not directly suitable for the modeling of clusters, liquids, or solids. Instead, empirical potentials for bulk systems are usually parameterized against bulk properties with implicit inclusion of multibody effects such as polarization. In addition, nonpolarizable force field models are unable to describe the conformational dependence of electrostatic properties. A single set of fixed charges or multipoles is generally not accurate when applied to the wide variety of conformations available to a flexible molecule.

There have been encouraging results from several groups during the past several years in applying polarizable potentials to small molecule clusters and liquids.[3–8] Correct and accurate description of energetic effects related to molecular polarizability is critical for a potential model to be applied successfully to different environments. As interest extends beyond small rigid molecules such as water and small amides, it is important to develop polarizable force field models that incorporate intramolecular polarization in addition to having the correct molecular response to an external field. In an attempt to develop electrostatic parameters for peptides and proteins, Faerman and Price[9] demonstrated that dipeptides serve as much better model compounds than small amides. Nevertheless, as observed for other molecules with conformational flexibility, they found that the electrostatics of an alanine dipeptide depend strongly on conformation.[10] This dependency is mainly due to through-space polarization, and presents a great obstacle to parameterization of a fixed charge force field directly from data on flexible model compounds. One possible solution to the problem posed by conformational variability is to make the electrostatic model explicitly dependent on the local geometry.[11,12] Alternatively, a polarizable potential that captures the correct intramolecular polarization behavior should be able to model changes in electrostatic potential and energy as a function of conformation. There has been little attention paid in the literature to the energetic effects of intramolecular polarization, other than a consensus that the treatment used for intermolecular polarization can also be applied to the intramolecular case. In one of the few explicit discussions, Karlstrom's group[13] has proposed an intramolecular polarization model for dimethoxyethane. The importance

Correspondence to: J. W. Ponder; e-mail: ponder@dasher.wustl.edu

Contract/grant sponsor: Computational Biology Activity of the National Science Foundation; contract/grant number: CBA-9808317

of explicit polarization effects in the accurate simulation of large, flexible systems such as proteins is a subject of current debate.[14,15]

In the present work, we propose a unified scheme for the treatment of inter- and intramolecular polarization within a molecular mechanics model. Our method has four components: atomic multipoles derived from high-level *ab initio* calculations, an analytical method for removing intramolecular polarization from the atomic multipoles to expose the underlying permanent electrostatic components, use of Thole's interactive dipole polarizability model,[16] and group-based intramolecular polarization as a means of merging the polarization model with local valence terms.

Use of higher order atomic multipole moments has been shown to drastically improve the description of the electrostatic potential around small molecules.[17] For a database of small molecules, a least-squares fit atom-centered monopole+dipole model typically reduces the error in the resulting electrostatic potential by a factor of 10 versus a simple partial charge model. Further addition of atom-centered quadrupoles reduces the average error in the potential by another order of magnitude.[17] While higher order multipoles, such as atom-centered quadrupole moments, obviously add to computational cost, this expense could be mitigated by only considering explicit higher order interactions at short range. Also, inclusion of atomic quadrupole components has been known for some time to be critical in obtaining correct geometries of van der Waals and hydrogen bonded complexes.[18] The Distributed Multipole Analysis (DMA) protocol introduced by Stone[19,20] provides atomic multipoles via a convenient and rigorous redistribution of the electron density associated with overlap integral products.

The atomic multipoles may be considered as a sum of a permanent electrostatic component plus the contribution from intramolecular polarization as described by the specific empirical model. This contribution must be removed from the multipoles to avoid double counting when the force field polarization model is subsequently applied. In general, this correction of the raw quantum-derived electrostatics has been done by an empirical procedure.[13,21] In the following, we propose a simple analytical method for exactly correcting atomic multipoles to yield a permanent electrostatic model consistent with our polarization scheme.

A number of atomic polarizability models, including additive[22,23] and interactive models,[16,24,25] have been proposed for treatment of molecular polarizability. Among these models the scheme originally suggested by Thole[16,26,27] exhibits several advantages: it avoids the polarization catastrophe at short range by replacing point dipole interactions with interactions between smeared dipoles; it produces anisotropic responses to an external field using only isotropic atomic polarizabilities; and 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.

All molecular mechanics methods face the challenge of combining a description of long-range nonbonded interactions (vdW, electrostatics, polarization) with short-range valence interactions (bonds, angles, torsions). Traditionally, force fields ignore or scale intramolecular vdW and permanent electrostatic interactions between atoms separated by three or fewer bonds. Polarizable force fields developed to date apply identical schemes for scaling of short-range intramolecular polarization. A goal of the current work is to investigate alternative atom- and group-based methods for merging polarization effects with the valence portion of a force field.

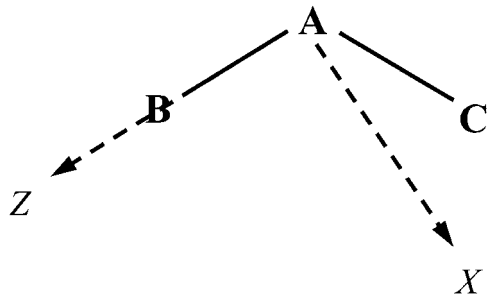

Figure 1. Local coordinate frame used to define atomic multipoles. A coordinate frame with origin at atom A is generated by placing the positive *z*-axis along the direction of a directly bonded atom B. The *x*-axis is defined to lie in the A-B-C plane where atom C is an atom either directly bonded or adjacent to atom A. The positive *x*-axis lies in the direction of atom C. The positive *y*-axis is then chosen perpendicular to the *z*- and *x*-axes to give a right-handed local coordinate system for atom A.

Methods

In order to transfer atomic multipoles among different conformations and molecules, local frames are defined for each atomic multipole site *i*. The *"z*-then-*x"* convention employed by TINKER[28] uses an atom *j* covalently bonded to atom *i* to define the positive *z*-axis. A second noncolinear reference atom *k* is selected such that the positive *x*-axis lies in the *ijk* plane and forms an acute angle with $\vec{ik}$. Finally, the *y*-axis is chosen to give a right-handed coordinate system. An example of such a local frame definition is provided in Figure 1. Reference sites for each unique atom type considered in this study are given in Table 1 together with the corresponding atomic multipole parameters.

The induced dipole at each atomic site is computed as

$$\mu_{i,\alpha}^{\text{ind}} = \alpha_i E_{i,\alpha} \tag{1}$$

where E is the field experienced by atom i and the subscript $\alpha \in \{x,y,z\}$. In an interactive polarizability model such as Thole's, the induced dipole on each atom will further polarize all other atoms both within and outside the molecule such that E becomes the sum of the fields generated by both permanent multipoles and induced dipoles at sites other than atom i:

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

where T represents the interaction matrix elements, and M is the vector of permanent atomic multipole components in Cartesian polytensor form.[29] The above equation may be solved iteratively for all atomic sites at the same time.

Table 1. Multipole Parameters.

Site	z Ref	x Ref	q	μ_x	μ_y	μ_z	Q_{xx}	Q_{yx}	Q_{yy}	Q_{zx}	Q_{zy}	Q_{zz}
N-Methylacetamide (NMA)												
C(O)	O	N	0.64220	0.22706	0.00000	0.31459	−0.50085	0.00000	0.23573	0.02285	0.00000	0.26613
O	C(O)	N	−0.66524	−0.05964	0.00000	−0.00965	−0.38021	0.00000	−0.05967	−0.12838	0.00000	0.43987
N	C(O)	O	−0.18752	0.20427	0.00000	−0.34027	1.06373	0.00000	−0.82316	0.04814	0.00000	−0.24057
H(N)	N	C(O)	0.07726	−0.01582	0.00000	−0.17196	0.02821	0.00000	−0.05415	0.00463	0.00000	0.02594
C(C)	C(O)	O	−0.24387	0.00000	0.00000	0.07193	0.28325	0.00000	0.19307	0.29847	0.00000	−0.47632
H(CC)	C(C)	C(O)	0.07693	0.00000	0.00000	−0.08930	−0.00778	0.00000	−0.03133	0.00708	0.00000	0.03910
C(N)	N	C(O)	−0.03623	0.00000	0.00000	0.23240	−0.39078	0.00000	−0.13890	−0.04627	0.00000	0.52968
H(CN)	C(N)	N	0.06087	−0.00898	0.00000	−0.08765	−0.00712	0.00000	−0.03030	−0.02713	0.00000	0.03741
Alanine dipeptide Acetyl *N*-terminal												
C	C(O)	O	−0.19495	−0.05933	0.00000	0.30647	−0.12000	0.00000	−0.18289	0.06006	0.00000	0.30289
H	C	C(O)	0.07797	−0.00148	0.00000	−0.10232	0.00077	0.00000	−0.01358	0.00508	0.00000	0.01281
C(O)	O	C	0.69878	−0.28625	0.00000	0.40485	0.09631	0.00000	−0.19283	0.05482	0.00000	0.09652
O	C(O)	C	−0.74615	0.10832	0.00000	−0.15915	−0.51576	0.00000	0.23754	0.12638	0.00000	0.27822
Alanine residue												
N	Cα	HN	−0.14206	0.03550	0.00000	0.44301	0.09170	0.00000	−1.11491	−0.24348	0.00000	1.02321
Cα	N	C(O)	−0.21238	0.15677	0.10446	0.14545	−0.17425	−0.22519	−0.40549	−0.00044	−0.22429	0.57974
Hα	Cα	Cβ	0.08921	0.014105	0.00000	0.04185	−0.00099	0.00000	−0.00099	0.00801	0.00000	0.00198
C(O)	O	Cα	0.85846	−0.01685	0.00000	0.27745	0.29463	0.00000	−0.41598	−0.00743	0.00000	0.12135
O	C(O)	Cα	−0.77770	−0.01897	0.00000	−0.21786	−0.61542	0.00000	0.33936	0.00866	0.00000	0.27606
H(N)	N	Cα	0.12992	−0.00883	0.00000	−0.14169	0.04124	0.00000	−0.02604	−0.00466	0.00000	−0.01521
Cβ	Cα	Hα	−0.15440	−0.00010	0.00000	0.36287	−0.29475	0.00000	−0.29287	0.00762	0.00000	0.58762
Hβ	Cβ	Cα	0.07484	−0.00111	0.00000	−0.10010	−0.00304	0.00000	−0.01538	0.01240	0.00000	0.01842
N-Methyl amide *C*-terminal												
N	C	H(N)	−0.28294	0.12868	0.00000	0.29428	0.13167	0.00000	−0.77946	−0.26698	0.00000	0.64780
H(N)	N	C	0.12524	0.04469	0.00000	−0.11589	0.03868	0.00000	−0.05683	0.02734	0.00000	0.01816
C	N	H(N)	−0.00902	0.04962	0.00000	0.26123	−0.42853	0.00000	−0.34055	0.01113	0.00000	0.76909
H(C)	C	N	0.05319	−0.00200	0.00000	−0.11501	0.00409	0.00000	−0.00544	−0.00819	0.00000	0.00135

Permanent atomic multipoles, as described in the text, are given in atomic units for *N*-methylacetamide (NMA) and for alanine dipeptide.

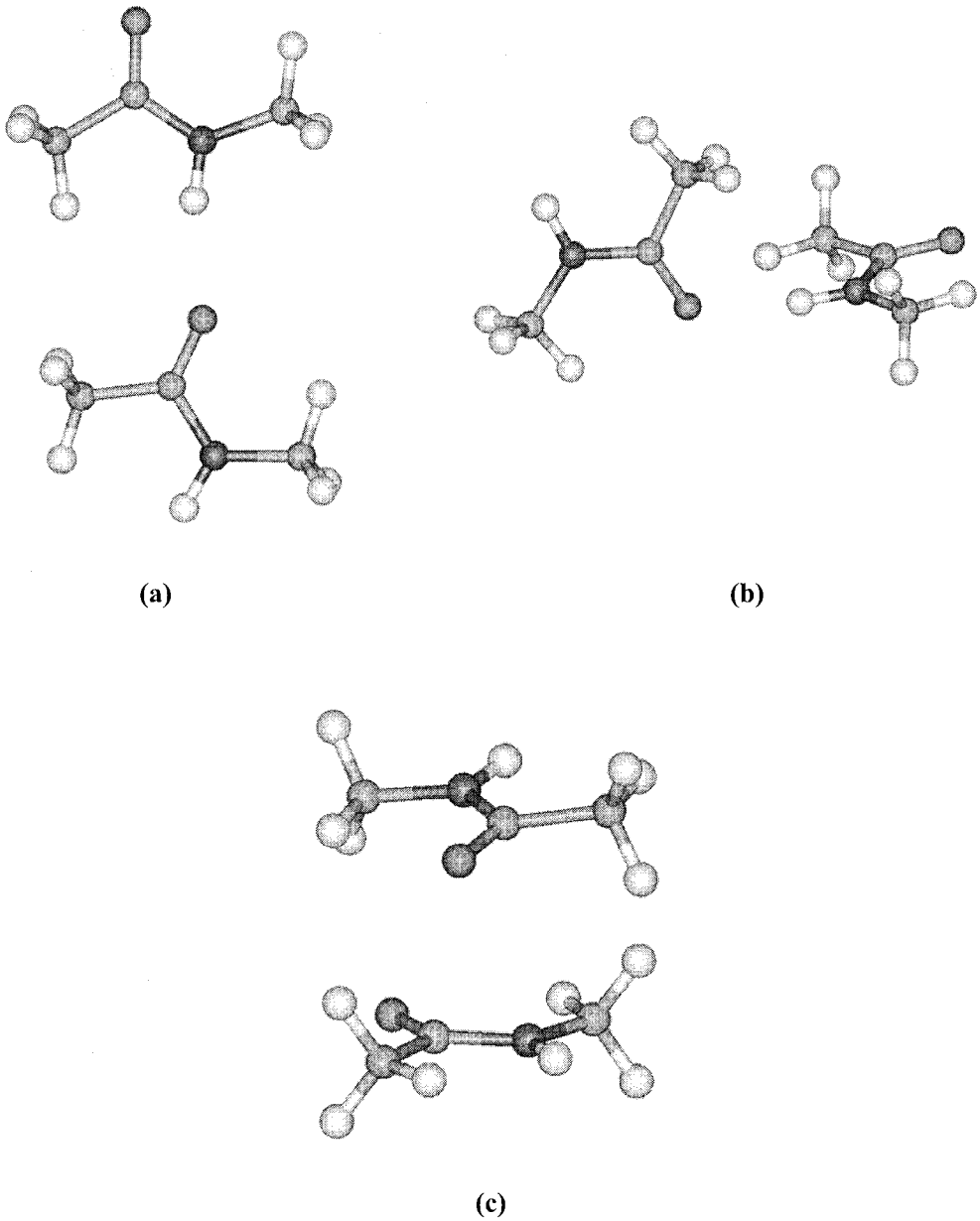

Figure 2. Configurations of the *N*-methylacetamide (NMA) dimer. The structures considered in this work, as described in the text, are (a) parallel, (b) perpendicular, and (c) stacked. Both dimers (a) and (b) contain intermolecular amide-amide hydrogen bonds, while configuration (c) does not.

The first term in eq. (2) corresponds to induction due to the permanent multipoles, which we will call "direct induction". The second term describes the induction caused by induced dipoles at other sites, that is, "mutual induction". Note that these two types of induction do not have to involve the same sets of atomic sites. Also, it may be desirable to apply different scale factors from 0 to 1 for mutual and direct induction based on bond separation. For example, in the case of water, $\{j\}$ stands for the atomic sites outside the molecule containing i. Meanwhile, $\{j'\}$ includes any atomic sites other than i, which is an intrinsic requirement of Thole's polarizability model as mentioned earlier. However, as the molecule size increases, one might wish to allow permanent multipoles to polarize other parts of the molecule beyond a certain distance. In the same spirit that short-range (in terms of bond separation) intramolecular nonbonded interactions are excluded, direct polarization may be modified at close distances. Definition of the groups $\{j\}$ and $\{j'\}$ coupled with the above formulae specifies an intramolecular polarization model.

Thole's method for computing molecular polarizabilities consists of a modification of the T matrix corresponding to an altered

Table 2. Comparison of NMA Dimer Electrostatic Potentials Computed by Various Models and *Ab Initio* Calculation.

	Fixed charge	Direct DMA	Average DMA	Monomer DMA	Monomer DMA + Intermol. Polarization
Parallel					
RRMS	8.4 (0.89)	6.4 (0.66)	6.0 (0.63)	15.9 (1.65)	5.6 (0.58)
WRRMS	7.5	5.5	5.3	15.5	4.9
Perpendicular					
RRMS	9.7 (0.94)	6.9 (0.65)	7.2 (0.68)	16.6 (1.57)	6.5 (0.61)
WRRMS	8.9	5.7	6.1	16.0	5.5
Stacked					
RRMS	20.0 (1.50)	12.6 (0.91)	18.2 (1.32)	12.7 (0.93)	9.3 (0.68)
WRRMS	19.6	10.8	17.7	11.9	8.5

Relative RMS (RRMS) and weighted relative RMS (WRRMS) deviations are given as percentages. Values in parentheses are absolute RMS deviations in kcal/mol. The fixed charge model taken from Mannfors et al.[31] was derived by simultaneous fit to ESP of multiple H-bonded dimer configurations, and is compared with MP2/6-31++G** results as used in their original work. All other models are compared with MP2/6-311++G(2d,2p) calculations. The average DMA model was obtained by averaging the direct DMA multipoles from the parallel and perpendicular dimers.

charge distribution. Among the several charge distributions suggested by Thole for use with his damping scheme, we have chosen the following exponential form

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

where $u = R_{ij}/(\alpha_i\alpha_j)^{1/6}$ is the effective distance as a function of atomic polarizabilities, and a is a dimensionless damping factor that controls the width of the smeared charge distribution. More details on the form of the T matrix, an iterative solution to yield induced dipoles, the Thole damping scheme, and computation of induction energy and forces are described in a separate work on a water model using the above protocol.[30] The atomic polarizabilities in Å^3 of C, N, O, and H are 1.334, 1.073, 0.837, and 0.496 respectively, as originally derived by Thole. A value of 0.572 was suggested by Thole for the damping factor a after least squares fit to a set of molecular polarizabilities. However, molecular polarizabilities are rather insensitive to a values over a fairly broad range. Interaction energies tend to vary more with changes in the damping factor, and we have chosen a fixed a value of 0.39 as determined by a fit to the binding energies of water clusters through the hexamer.[30]

For a given conformation of a model compound (e.g., alanine dipeptide), one can easily obtain atomic multipoles by means of DMA as

$$M_i = [q_i, \mu_{i,x}, \mu_{i,y}, \mu_{i,z}, Q_{i,xx}, Q_{i,xy}, Q_{i,xz}, \ldots Q_{i,zz}]^T \tag{4}$$

Table 3. Comparison of the Total Dipole Moment and Dipole Components (Debye) of the NMA Dimer as Computed by *Ab Initio* Calculation and Various Force Field Models.

	Parallel	Perpendicular	Stacked	RRMS (%)
Ab initio				
d	8.85	7.88	3.21	
d_x	−8.82	7.73	1.61	
d_y	0.76	0.09	2.62	
d_z	0.02	1.51	−0.93	
Monomer DMA				
d	7.49	6.64	3.34	15.0
d_x	−7.45	6.46	1.57	15.8
d_y	0.75	0.01	2.87	10.2
d_z	0.00	1.52	−0.71	12.5
Monomer DMA + intermolecular polarization				
d	8.85	7.83	3.21	0.4
d_x	−8.81	7.69	1.62	0.4
d_y	0.82	0.03	2.66	3.4
d_z	0.00	1.48	−0.76	9.8

Table 4. Comparisons of ESPs for Five Alanine Dipeptide Conformers as Computed by *Ab Initio* (MP2/6-311G**) Calculation, DMA Multipoles, and Various Models.

	α_L	$C5$	$C7a$	$C7e$	α'
	$\phi = 63.5$	$\phi = -158.4$	$\phi = 73.9$	$\phi = -82.9$	$\phi = -166.1$
	$\psi = 34.8$	$\psi = 161.3$	$\psi = -64.0$	$\psi = 77.9$	$\psi = -37.2$
Direct DMA	3.9 (0.35)	5.4 (0.35)	5.7 (0.38)	5.7 (0.34)	4.0 (0.31)
Avg DMA	23.9 (2.15)	18.1 (1.18)	9.8 (0.65)	8.0 (0.48)	40.2 (3.14)
Atom 1-2	7.4 (0.67)	7.4 (0.48)	9.9 (0.66)	9.5 (0.57)	6.9 (0.53)
Atom 1-3	12.5 (1.12)	14.7 (0.96)	11.0 (0.73)	19.1 (1.14)	6.5 (0.51)
Atom 1-4	15.0 (1.35)	17.6 (1.14)	19.1 (1.27)	29.8 (1.79)	12.1 (0.94)
Group 1-2	6.5 (0.58)	7.6 (0.50)	8.3 (0.55)	9.1 (0.55)	7.3 (0.57)

Both *ab initio* and direct DMA calculations were carried out for each conformer separately. The average DMA model utilizes the average of modified DMA multipoles of the first four conformers, excluding α'. All RRMS deviation values are given as percentages. The RMS deviations listed in parentheses are in kcal/mol.

The resulting multipoles on each atom may be considered as a sum of "permanent" and "induced" moments:

$$M_i = M_i^p + M_i^{\text{ind}} \tag{5}$$

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

$$M_{i,\alpha}^{\text{ind}} = \alpha_i \left(\sum_{\{j\}} T_\alpha^{ij} M_j^p + \sum_{\{j'\}} T_\alpha^{ij'} M_{j'}^{\text{ind}} \right) \tag{6}$$

This is the same relation as in eq. (2) except the induced dipole is replaced by generalized induced moments. Substitution of eq. (5) into the above expression yields

$$M_{i,\alpha}^{\text{ind}} = \alpha_i \left(\sum_{\{j\}} T_\alpha^{ij} (M_j - M_j^{\text{ind}}) + \sum_{\{j'\}} T_\alpha^{ij'} M_{j'}^{\text{ind}} \right) \tag{7}$$

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

$$M_{i,\alpha}^{\text{ind}} = \alpha_i \sum_{\{j\}} T_\alpha^{ij} M_j \tag{8}$$

In the case where $\{j\}$ and $\{j'\}$ are not the same, or when different fractional scaling factors are desired, it is more convenient to write eq. (7) in a more general form:

$$\begin{aligned} M_{i,\alpha}^{\text{ind}} &= \alpha_i \left(\sum_{\text{All Sites}} s_j^p T_\alpha^{ij} (M_j - M_j^{\text{ind}}) + \sum_{\text{All Sites}} s_j^m T_\alpha^{ij} M_j^{\text{ind}} \right) \\ &= \alpha_i \left(\sum_{\text{All Sites}} s_j^p T_\alpha^{ij} M_j + \sum_{\text{All Sites}} (s_j^m - s_j^p) T_\alpha^{ij} M_j^{\text{ind}} \right) \end{aligned} \tag{9}$$

where s_j is a scaling factor between 0 and 1 applied to site j. The superscripts p and m represent permanent and mutual induction, respectively. The above equation can be solved iteratively via the same procedure that is used to compute induce dipoles based on eq. (2). In the present work, the s_j factors are always chosen to be either 0 or 1.

By subtracting induced moments from the M_i obtained from *ab initio* calculation for a single conformer, we are left with the truly "permanent" atomic multipoles with intramolecular polarization removed. In order to test whether these permanent multipoles are indeed transferable among different conformations, one can carry out the procedure on several conformers and compare the modified multipoles. However, such a comparison is usually less than straightforward due to the underdetermined nature of the multipole parameters. Instead, we will compare the electrostatic potential computed using the permanent atomic multipoles plus explicit polarization against *ab initio* results at chosen grid points for various conformations. The root-mean-square deviation of two set electrostatic potentials is computed as

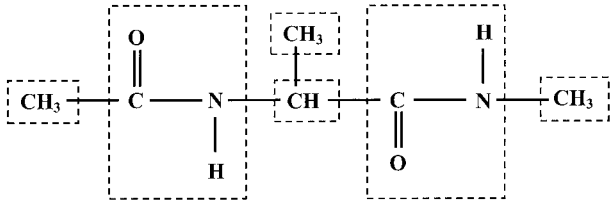

Figure 3. Definition of polarization groups for alanine dipeptide. Dashed regions are outlined such that direct induction is excluded between atoms within a polarization group. Note that groups should generally be chosen such that they have a near-integral net charge and possess at most limited conformational variability. All amide/peptide bonds are trans in the present work; *cis*-bonds require separate parameterization for use in a general force field.

Table 5. Comparison of the Total Molecular Dipole Moment and Dipole Components (Debye) for Alanine Dipeptide Conformers Computed by *Ab Initio* Calculation at the MP2/6-311G** Level and the 1-2 Group-Based Polarization Model.

	α_L	*C*5	*C*7*a*	*C*7*e*	α'	RRMS (%)
Ab initio						
d	5.54	2.99	3.39	2.35	4.77	
d_x	−3.37	2.98	−3.32	−2.38	0.80	
d_y	−1.41	0.16	0.09	0.24	−4.20	
d_z	−4.16	−0.21	−0.70	0.80	2.11	
Group 1-2						
d	5.51	2.91	3.20	2.55	4.90	3.6
d_x	−3.17	2.90	−3.11	−2.45	0.97	5.8
d_y	−1.48	0.17	0.05	0.26	−4.27	2.5
d_z	−4.25	−0.23	−0.76	0.63	2.30	5.8

$$\mathrm{RMS} = \sqrt{\frac{\sum_{i=1}^{N} (v_i^{\mathrm{QM}} - v_i^{\mathrm{MM}})^2}{N}} \tag{10}$$

The relative root-mean-square deviation is defined as

$$\mathrm{RRMS} = \sqrt{\frac{\sum_{i=1}^{N} (v_i^{\mathrm{QM}} - v_i^{\mathrm{MM}})^2}{\sum_{i=1}^{N} v_i^{\mathrm{QM}^2}}} \tag{11}$$

For comparison purposes, we also calculate the weighted relative RMS (WRRMS) quantity used by Mannefors et al.[31] as

$$\mathrm{WRRMS} = \sqrt{\frac{\sum_{i=1}^{N} w_i^2 (v_i^{\mathrm{QM}} - v_i^{\mathrm{MM}})^2}{\sum_{i=1}^{N} w_i^2 v_i^{\mathrm{QM}^2}}} \tag{12}$$

where the weight w_i for each grid point is chosen to be its distance from the closest atom. However, the WRRMS values reported here are systematically somewhat larger for identical structures than those of Mannefors et al., due to our use of grid points slightly closer to the molecular vdW surface.

Computational Details

The grid points used in ESP calculations were generated using ESTAR.[32] The atomic radii for H, C, N, O were chosen to be roughly their vdW radii of 1.4, 1.9, 1.8, and 1.7 Å, respectively. The inner layer of grid points was 0.5 Å away from the surface defined by the above radii. About 6000 grid points were distributed on a 0.5 Å cubic lattice within a 2 Å shell from 0.5 Å to 2.5 Å above the vdW surface. While selection of grid points will affect the absolute magnitudes in the ESP comparisons, as larger errors usually occur at closer distances, it is important to cover the regions of closest possible contact.

Ab initio calculations were carried out using the GAUSSIAN98 package.[33] Atomic charge, dipole, and quadrupole values were determined by Stone's distributed multipole analysis available through his GDMA program,[34] which takes as its input a GAUSSIAN checkpoint file. ESTAR was used to compute the electrostatic potential at sets of grid points for various potential models.

All force field calculations were performed with the TINKER package,[28] which implements the polarization model described here as part of the TINKER force field.

Results and Discussion

N-methylacetamide (NMA) Monomer and Dimer

Atomic polarizabilities in our model, derived by fitting to experimental molecular polarizabilities, guarantee the correct electric response to an external field or to other molecules. Before discussing methods for intramolecular polarization, it is necessary to verify that the simpler case of intermolecular polarization is modeled adequately. Here we repeat the test introduced by Mannfors et al.[31] In this test, the ESP of an NMA dimer was computed using parameters derived from the monomer with contributions from intermolecular polarization included. This intermolecular test bears some resemblance to our examination below of intramolecular polarization in the alanine dipeptide, a structure that may be viewed as two NMA-like segments joined by a common alpha carbon atom.

As shown in Figure 2, three NMA dimer configurations have been selected for calculation: parallel and perpendicular, which were also used in the earlier study,[31] plus a stacked configuration where the two amide planes are parallel to each other. The NMA monomer was optimized at the MP2/6-311++G(2d2p) level, and atomic multipoles were calculated via DMA at the same level.

These atomic multipoles gave nearly zero error versus the *ab initio* electrostatic potential of the NMA monomer, as observed for many other small molecules.[17] The parallel and perpendicular configurations were then obtained via geometry optimization at the MP2/6-31+G* level with intramolecular degrees of freedom frozen. The stacked configuration, containing no hydrogen bonds, was a minimum on TINKER potential surface. It was chosen to test the transferability of the parameters among different configurations, including those with and without intermolecular hydrogen bonding.

All comparison of electrostatic potentials was made with MP2/6-311++G(2d,2p) results, except for the fixed charge model, which was taken from Mannfors et al.[31] Their model was determined by simultaneously fitting to the MP2/6-31++G** ESP of three hydrogen bonded NMA dimer configurations while constraining charges to be equal for equivalent atom types. The fixed charge model led to a slightly greater than 4% WRRMS deviation from the MP2/6-31++G** ESP for the parallel and perpendicular dimers in that study, but resulted in an almost doubled WRRMS error over the set of grid points adopted in this study, compared with the ESP computed at the same level, as indicated by Table 2. In addition, the fixed charge model behaves rather poorly (19.6% WRRMS deviation) when applied to the stacked configuration, indicative of the problems with transferring a fixed charge model between configurations with different hydrogen bonds.

Using atomic multipoles directly from a distributed multipole analysis of each individual dimer at MP2/6-311++G(2d,2p) level, a WRRMS deviation of less than 6% was achieved for both hydrogen-bonded configurations, and 10.8% for the stacked configuration, as listed in Table 2. These numbers should be considered as the best-case target values when judging the performance of alternative polarization models.

The "average DMA" model was obtained by averaging the DMA multipoles obtained directly from quantum calculations on the parallel and perpendicular dimers. Because the higher order multipoles are defined within specific local coordinate frames, the average was computed with respect to the local frames within each configuration. The resulting averaged multipoles were rotated back into the global frame of each configuration before electrostatic potentials were computed. This averaged model is quite comparable to conformation-specific direct DMA in terms of reproducing the ESP for the two hydrogen-bonded dimers. However, when applied to the stacked configuration it produced as much as an 18.2% RRMS error (17.7% WRRMS) in the ESP.

Before the monomer DMA multipoles were applied to compute the ESP for the dimers, the atomic multipoles of the three hydrogens in each methyl group were averaged. This modification, which resulted in a 3% deviation from the monomer ESP, is necessary if the multipoles are to be used for general simulation.

Monomer multipoles without intermolecular polarization gave an ESP error of over 15%, more than twice that of the direct DMA baseline value, for the electrostatic potentials of the two hydrogen-bonded dimers. However, as expected, the monomer multipoles are a much more reasonable model for the stacked configuration that has no hydrogen bonds. Once intermolecular polarization is included, the ESP deviation for all configurations drops dramatically, to levels even slightly below those given by the direct DMA values.

The total dipole moment and dipole components for each NMA dimer structure were also computed using monomer multipoles both with and without intermolecular polarization and were compared with *ab initio* results for each dimer. As with the ESP comparisons, the results listed in Table 3 confirm that our intermolecular polarization model is successful in describing the nonadditive variation of electrostatics upon moving from monomer to dimer, as well as among different dimer configurations.

Alanine Dipeptide

Five well-known local minima of alanine dipeptide (α_L, C5, C7a, C7e, and α') were chosen as test configurations. The C5, C7a, and C7e structures each contain an intramolecular hydrogen bond, whereas α_L and α' do not. All structures were obtained from energy optimization at the MP2/6-31G* level of theory. Subsequently, an MP2/6-311G** single-point calculation was carried out for each individual conformer, and the ESPs were computed on a grid of over 6000 points per conformation outside the respective vdW molecular envelopes. Direct DMA was performed at the same level to produce a set of atomic multipoles for each conformer. The ESP computed using these raw conformer-specific DMA multipoles was then compared with *ab initio* results for all five conformers, with results shown in Table 4. In order to demonstrate the nontransferability of these atomic multipoles, an average was taken over the raw DMA multipoles of the α_L, C5, C7a, and C7e structures, and these average values were used to compute an ESP for each of the five dipeptide conformers. The results of the ESP comparison given in Table 4 show errors that vary widely from one conformer to another. In particular, the average DMA model produced a large error (40% RRMS) on the one structure, the α' conformer, which was omitted during the averaging process.

The previously described consistent procedure was employed to remove permanent induction from each set of dipeptide atomic multipoles according to our specific polarization model. Instead of using the "permanent" multipoles from one conformer to compute ESPs for all other conformers, an alternate test of transferability was used. The permanent multipoles of α_L, C5, C7a, and C7e were averaged, and the resulting multipoles were used, with the addition of polarization, to compute the ESP for all five conformers. As before, the α' conformer was left out of the permanent multipole average in order to serve as an unbiased test case.

A natural choice for scaling of the intramolecular polarization is an atom–based scheme such as those used by pairwise force fields for intramolecular nonbonded interactions. Several atom-based models are given in Table 4. The "1-2" model includes permanent polarization everywhere, even between atoms directly bonded (1-2) to each other. The "1-3" model excludes 1-2 polarization but includes permanent polarization between atoms separated by two or more bonds. The "1-4" model allows polarization only between atoms separated by at least three bonds. The comparison in Table 4 of the ESPs computed using these scaling models clearly indicates the 1-2 atom-based model is significantly better than 1-3 and 1-4 models, as it produced an ESP much more comparable to those from DMA performed directly on each of the conformers. This is not particularly surprising because both the 1-3 and 1-4 atom-based models require splitting of contributions from atoms having large counterbalancing charges, such as the carbonyl

carbon and oxygen, leading to unreasonably large induction effects.

Although the 1-2 atom-based model seems successful, it is not preferred because of the large 1-2 direct induction it engenders. Arbitrary separation of electrostatics into underlying permanent and induced components does not affect the electrostatic potential, because the ESP is independent of the permanent versus induced origin of individual atomic moments. However, this distinction, enforced by the polarization model, does have a significant effect on energy computations. As energy is required in order to produce induced dipoles through polarization, unrealistic interaction energies may arise if an unphysically large portion of the local electrostatics is considered to be the result of induction. An ideal model would avoid such unphysical intramolecular polarization, while still accounting, to the greatest extent possible, for through-space polarization as a function of conformational change. A potential solution is to identify small and rigid fragments within the molecule as polarization groups. We propose a 1-2 group-based model wherein direct induction occurs everywhere except within each of the polarization groups. Furthermore, in line with the Thole model, full mutual induction is always allowed between every pair of atoms regardless of whether the atoms are members of the same group. It is also important to group neighboring atoms with large opposite charges together such that each group will have only a small net charge. Following these rules, groups are defined for the alanine dipeptide structure illustrated in Figure 3: (a) the amide groups each containing C(═O), O(═C), N(—H), and H(—N), (b) the alpha C and alpha H atoms, and (c) each of the terminal and side chain CH_3 groups. Net charges on the resulting groups are small, the largest being +0.0155 electrons on the central group. This intramolecular polarization model was tested on the five alanine dipeptide conformers. The results, shown in Table III, indicate the efficacy of the proposed model. The 1-2 group-based polarization scheme does almost as well as the conformation-specific direct DMA multipoles at reproducing the *ab initio* ESPs. Furthermore, this excellent behavior holds for all conformers, including the α' conformer that was not involved in the derivation of the permanent atomic multipoles.

In Table 5, molecular dipole moments computed using the 1-2 group-based model are compared with *ab initio* results for all five dipeptide conformers. The RRMS deviation of the total molecular dipole moments is less than 4%. The molecular dipole moments produced by the direct DMA multipoles are not given as they are essentially identical to those from the corresponding *ab initio* calculations.

As a final note on derivation of atomic multipoles, it is worth mentioning that alternative approaches other than DMA were also examined in the same fashion as that presented above. It was rather difficult to achieve physically meaningful atomic multipoles using an unrestrained fit to the ESP due to the size of the dipeptide molecule. While restrained fits have been used with success to derive atomic partial charge models,[35] we have chosen to consider models with a more obvious dependence on the underlying electron density. The CADPAC[36] program provides a simple scheme for generating distributed multipoles based on a numerical integration of electron density over Voronoi atomic volumes. This method was able to give reasonable atomic multipoles for each individual dipeptide conformer, but fared poorly in the attempt to transfer the resulting "permanent" multipoles among conformers. The indication is that Stone's DMA process results in multipoles that are more "transferable" than others, as long as consistent basis sets are employed throughout. However, because it can be very sensitive to the presence of diffuse functions, the original DMA protocol may not be useful when applied to very large basis set calculations. It will be of interest to explore still other methods, such as the atoms-in-molecules (AIM) approach of Bader[37] and the cumulative atomic multipole moments (CAMM) algorithm,[38,39] in future work. Other workers have found that atomic moments derived from the AIM real-space distribution of electron density perform about as well as DMA moments in simple non-polarizable energy calculations.[40] However, consistency with a polarization scheme such as the one suggested here is a more stringent test of the physical reasonableness of a multipole distribution.

Conclusions

In summary, an intermolecular polarization model has been proposed that is able to account for the nonadditivity in electrostatic potentials, as demonstrated by a series of tests on NMA monomer and dimer configurations. The intermolecular model can be extended via a scheme utilizing interacting polarization groups. Results presented in Tables 4 and 5 show that this extended model is very successful at capturing much of the intramolecular polarization inside the alanine dipeptide molecule. The underlying permanent atomic multipoles are transferable amongst a set of minimum energy conformations. Overall, the described protocol provides a convenient tool for parameterization of molecular mechanics electrostatic terms for flexible model compounds. Thus, a consistent foundation is available for subsequent elaboration of a general polarizable force field for biomolecular and other systems. Future work will present an energetic model compatible with the present polarization methodology.

References

1. Dykstra, C. E. Chem Rev 1993, 93, 2339.
2. Dunning, T. H. J. J Phys Chem A 2000, 104, 9062.
3. Caldwell, J. W.; Kollman, P. A. J Phys Chem 1995, 99, 6208.
4. Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. J Phys Chem 1994, 98, 4180.
5. Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. J Phys Chem B 1999, 103, 4730.
6. Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. J Phys Chem B 2000, 104, 9746.
7. Brdarski, S.; Astrand, P.-O.; Karlstrom, G. Theor Chem Acc 2000, 105, 7.
8. Burnham, C. J.; Xantheas, S. S. J Chem Phys 2002, 116, 1500.
9. Faerman, C. H.; Price, S. L. J Am Chem Soc 1990, 112, 4915.
10. Price, S. L.; Faerman, C. H.; Murray, C. W. J Comput Chem 1991, 12, 1187.
11. Dinur, U.; Hagler, A. T. J Comput Chem 1995, 16, 154.
12. Koch, U.; Stone, A. J. J Chem Soc Faraday Trans 1996, 92, 1701.
13. Engkvist, O.; Astrand, P.-O.; Karlstrom, G. J Phys Chem 1996, 100, 6950.

14. van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L. I.; Merz, K. M. J. J Phys Chem B 2000, 104, 9554.
15. Roux, B.; Berneche, S. Biophys J 2002, 82, 1681.
16. Thole, B. T. Chem Phys 1981, 59, 341.
17. Williams, D. E. J Comput Chem 1988, 9, 745.
18. Buckingham, A. D.; Fowler, P. W. Can J Chem 1985, 63, 2018.
19. Stone, A. J. Chem Phys Lett 1981, 83, 233.
20. Stone, A. J. Mol Phys 1985, 56, 1047.
21. Cieplak, P.; Caldwell, J.; Kollman, P. J Comput Chem 2001, 22, 1048.
22. Miller, K. J. J Am Chem Soc 1990, 112, 8533.
23. Stout, J. M.; Dykstra, C. E. J Phys Chem A 1998, 102, 1576.
24. Applequist, J.; Carl, J. R.; Fung, K.-K. J Am Chem Soc 1972, 94, 2952.
25. Bode, K. A.; Applequist, J. J Phys Chem 1996, 100, 17820.
26. de Vries, A. H.; van Duijnen, P. T.; Zijlstra, R. W. J.; Swart, M. J Electron Spectrosc 1997, 86, 49.
27. van Duijnen, P. T.; Swart, M. J Phys Chem A 1998, 102, 2399.
28. Ponder, J. W. TINKER: Software Tools for Molecular Design, 3.9; Washington University School of Medicine: Saint Louis, MO, 2001.
29. Applequist, J. J Math Phys 1983, 24, 736.
30. Ren, P.; Ponder, J. W. 2002, in preparation.
31. Mannfors, B.; Mirkin, N. G.; Palmo, K.; Krimm, S. J Comput Chem 2001, 22, 1933.
32. Breuer, M. ESTAR: Electrostatic Properties Research Package, 1.0; Max-Planck-Institut für Biochemie: Martinsried, Germany, 1991.
33. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A. J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98, Revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.
34. Stone, A. J. GDMA; Cambridge University Technical Services: Cambridge, England, 1998.
35. Wang, J.; Cieplak, P.; Kollman, P. A. J Comput Chem 2000, 21, 1049.
36. Amos, R. D. CADPAC: Cambridge Analytic Derivatives Package, 6.5; Cambridge University: Cambridge, England, 2001.
37. Bader, R. F. W. Atoms in Molecules—A Quantum Theory; Oxford University Press: Oxford, 1990.
38. Sokalski, W. A.; Poirier, R. A. Chem Phys Lett 1983, 98, 86.
39. Sokalski, W. A.; Keller, D. A.; Ornstein, R. L.; Rein, R. J Comput Chem 1993, 14, 970.
40. Popelier, P. L. A.; Joubert, L.; Kosov, D. S. J Phys Chem A 2001, 105, 8254.