Calculation of the Dipole Moment for Polypeptides Using the Generalized Born-Electronegativity Equalization Method: Results in Vacuum and Continuum-Dielectric Solvent

K. Shimizu,[†] H. Chaimovich,[‡] and J. P. S. Farah[†]

Department of Chemistry, Institute of Chemistry, University of São Paulo, São Paulo, SP, Brazil, and Department of Biochemistry, Institute of Chemistry, University of São Paulo, São Paulo, SP, Brazil

L. G. Dias*

Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290

D. L. Bostick

Department of Physics and Program in Molecular/Cell Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599-3255

Received: November 1, 2003; In Final Form: December 30, 2003

The electronegativity equalization methodology, EEM, is frequently used to calculate the charge distribution and reactivity index (e.g., local softness and hardness, condensed Fukui function) of molecules. However, recent work (Chelli, R. et al., *J. Chem. Phys.* **1999**, *111*, 8569) has shown a serious shortcoming of EEM in the prediction of the polarizability for large molecules. In this paper, our goal is to show that we can obtain a reliable dipole moment for polypeptides in vacuum and continuum-dielectric solvent using the constrained charge approximation and the generalized Born-electronegativity equalization method. Different EEM parameterizations were tested and compared to the expected values of the dipole moment vector operator as calculated at the ab initio B3LYP/6-311G(d,p) level. One EEM parameterization (Bakowies, D., Thiel, W., *J. Comput. Chem.* **1996**, *17*, 87) when used with the constrained charge approximation and the generalized Born-electronegativity equalization method was comparable to the CM1 charge model (Storer et al., *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87) in the prediction of the dipole moment vector in vacuum and continuum-dielectric solvent, but was calculated with a much greater computational efficiency.

I. Introduction

Despite the number of papers dedicated to the inclusion of polarization in molecular mechanics force fields (see for example refs 1−4) using the electronegativity equalization method, EEM, only a few works treat the problem of large molecules.[5−7] A series of three papers[5a−c] detailed a means by which one might construct a polarizable molecular mechanics force field for an arbitrary molecule based on ab initio quantum chemistry. The rationale for a good description of polypeptide charge distribution was obtained using permanent (not fluctuating) charges and polarizable atomic dipoles. Scheraga et al.[6] have shown that atomic monopoles can provide a good description of the dipole moment of a molecule. This work extended the partial equalization of the orbital electronegativity method,[8] taking into account the molecular geometry in the calculation of charge distribution of polypeptides. The work of Cong and Yang[7] has generalized the atom-bond electronegativity equalization method,[9] ABEEM, for the calculation of the charge distribution in polypeptides. More than 200 model molecules with Mulliken charge distributions calculated at the RHF/STO-3G level were used in this parameterization of ABEEM. Although the correlation between ABEEM and RHF/STO-3G charges was very good, the authors did not shown a comparison of molecular observables obtained from their method to experiments or ab initio calculations.

Recently, a serious caveat of EEM that can limit its application for large molecules has been demonstrated.[10] EEM overestimates the polarizability of a molecule as its size grows (it sees molecules as conductors, exaggerating its charge delocalization). This problem may be the rationale for the method of Kaminski et al.[5a−c] In their work, the permanent charge distribution of each amino acid was obtained by an equalization scheme and then was used as a fixed charge model in the following calculations for polypeptides. It may be possible that the method of Scheraga et al.[6] has the same problem.

We are also particularly interested in calculation of the charge distribution of molecules in implicit solvent[11] (continuum-dielectric plus first-solvation-shell semiempirical corrections). These models can be classified by the treatment of the solute charge distribution: quantum mechanically or molecular mechanically. On a historical note, Onsager[12] introduced the first molecular mechanical model of the effect of implicit solvent on a polarizable solute by treating the solute as a polarizable dipole in a spherical cavity. Other molecular mechanic-like methods that treat polarizable solute coupled to the solvent

* Corresponding author: lgdias@email.unc.edu; Department of Chemistry, University of North Carolina−Chapel Hill, Venable Hall CB# 3290, Chapel Hill, NC 27599-3290.

[†] Department of Chemistry.

[‡] Department of Biochemistry.

10.1021/jp037315w CCC: $27.50
Published on Web 03/06/2004

reaction-field have been formulated. As examples, we have the method of Sharp et al.,[13] the method of Nilar,[14] and the generalized Born-electronegativity equalization methodology,[15] GBEEM. The method of Sharp et al.[13] simulates the local polarizabilities by adjusting the value of the local dielectric constant in the portion of the cavity belonging to that atom, while the method of Nilar[14] considers the solute as a collection of polarizable point charges located at the nuclear positions and solves the Poisson and Laplace equations using a boundary element method. The GBEEM is an implementation of the self-consistent reaction-field method based on a generalized electronegativity equalization scheme.[15,16]

In this work, we study the effectiveness of EEM for the calculation of dipole moments in organic molecules and polypeptides. Different parameterizations were tested for calculations in vacuum and continuum-dielectric solvent. We also propose here an approximation for the dipole moment calculation as an attempt to correct the shortcomings of EEM.

II. Theoretical and Computational Methods

Electronegativity Equalization Method. For a molecule with N atomic charges, q_i, the energy, E, is given by

$$E = \sum_{i=1}^{N}\left(E_i^* + \chi_i^* q_i + \frac{\eta_i^*}{2} q_i^2\right) + \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\(j\neq i)}}^{N} \eta_{ij}(r_{ij}) q_i q_j \qquad (1)$$

where χ_i^* and η_i^* are atomic electronegativity and hardness and $\eta_{ij}(r_{ij})$ is expressed as an interpolation between the Coulomb operator and atomic hardness:

$$\eta_{ij}(r_{ij}) = \frac{1}{\sqrt{r_{ij}^2 + \left(\frac{2}{\eta_i^* + \eta_j^*}\right)^2}} \qquad (2)$$

Using the principle of the electronegativity equalization

$$\chi_1 = \chi_2 = \chi_3 = \ldots = \chi_N \qquad (3)$$

with

$$\chi_i = \frac{\partial E}{\partial q_i} = \chi_i^* + \eta_i^* q_i + \sum_{\substack{j=1\\(j\neq i)}}^{N} \eta_{ij}(r_{ij}) q_j$$

plus the constraint for the total molecular charge

$$Q = \sum_{i=1}^{N} q_i$$

we obtain the linear system of equations:

$$\begin{cases} Q = \sum_{j=1}^{N} q_j \\ \chi_1^* - \chi_2^* = \eta_2^* q_2 - \eta_1^* q_1 + \sum_{\substack{j=1\\(j\neq 2)}}^{N} \eta_{2j}(r_{2j}) q_j - \sum_{\substack{k=1\\(k\neq 1)}}^{N} \eta_{1k}(r_{1k}) q_k \\ \vdots \qquad\qquad \vdots \\ \chi_1^* - \chi_N^* = \eta_N^* q_N - \eta_1^* q_1 + \sum_{\substack{j=1\\(j\neq N)}}^{N} \eta_{Nj}(r_{Nj}) q_j - \sum_{\substack{k=1\\(k\neq 1)}}^{N} \eta_{1k}(r_{1k}) q_k \end{cases} \qquad (4)$$

which must be solved to obtain the charges q_i, for each atomic site, i.

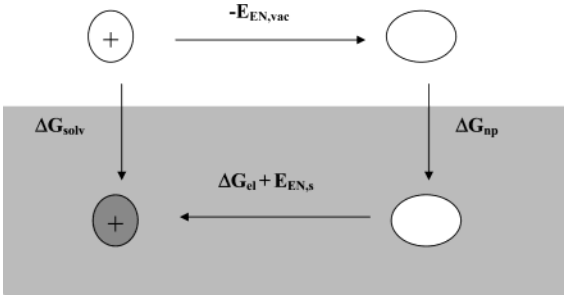

Figure 1. Thermodynamic cycle showing the decomposition of the total solvation free energy of the solute into nonpolar, electrostatic, and internal contributions (see theoretical and computational methods).

Generalized Born-Electronegativity Equalization Method. The GBEEM is a combination of the generalized Born model with the electronegativity equalization method for the treatment of the effect of implicit solvent on the solute charge redistribution.[15] The solvation free-energy minimization with respect to the solute charges produces the renormalized charges due to the coupling to the solvent reaction-field.[15,16] The EEM is a semiempirical version of the density functional theory where the variational principle leads to charges.[1−4]

Using the thermodynamic cycle[11,17] (see Figure 1) and the EEM expressions

$$\begin{aligned} \Delta G_{\text{solv}} &= E_{\text{EN,s}} - E_{\text{EN,vac}} + \Delta G_{\text{np}} + \Delta G_{\text{el}} \\ &= \sum_{i=1}^{N}\left(E_i^* + \chi_i^* q_{i,\text{s}} + \frac{\eta_i^*}{2} q_{i,\text{s}}^2\right) + \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\(j\neq i)}}^{N} \eta_{ij}(r_{ij}) q_{i,\text{s}} q_{j,\text{s}} - \\ &\quad \sum_{i=1}^{N}\left(E_i^* + \chi_i^* q_{i,\text{vac}} + \frac{\eta_i^*}{2} q_{i,\text{vac}}^2\right) - \\ &\quad \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\(j\neq i)}}^{N} \eta_{ij}(r_{ij}) q_{i,\text{vac}} q_{j,\text{vac}} + \Delta G_{\text{np}} + \\ &\quad \left(\frac{1}{\epsilon} - 1\right)\sum_{i=1}^{N}\frac{q_{i,\text{s}}^2}{2\alpha_i^{\text{B}}} + \frac{1}{2}\left(\frac{1}{\epsilon} - 1\right)\sum_{i=1}^{N}\sum_{\substack{j=1\\(j\neq i)}}^{N}\frac{q_{i,\text{s}} q_{j,\text{s}}}{f_{\text{GB}}(r_{ij})} \end{aligned} \qquad (5)$$

We consider the following: (a) ΔG_{np}, the nonpolar contribution[11,17] to the solvation free-energy, does not depend on the charge distribution; (b) $\Delta E_{\text{EN}} = E_{\text{EN,s}} - E_{\text{EN,vac}}$, the change in the total electronic and nuclear energy, can be approximated by the EEM expressions; (c) ΔE_{EN} is *also* the change in the internal free-energy of the solute[11,17] (if zero point and thermal contributions in solution and gas are mutually canceled); (d) the dielectric properties of the solvent are characterized by a constant, ϵ (the solvent is an isotropic and homogeneous continuum).

ΔG_{el} is the electrostatic reversible work as approximated by the generalized Born model,[11,18] GB model. To use the GB model, we need to compute the effective Born radius, α_i^{B}, and the generalized Born function, f_{GB}. The inverse of f_{GB} is similar in form to the η_{ij} term:

$$f_{\text{GB}}(r_{ij})^{-1} = \frac{1}{\sqrt{r_{ij}^2 + \alpha_i^{\text{B}}\,\alpha_j^{\text{B}} e^{\{-r_{ij}^2\}/\{4\alpha_i^{\text{B}}\alpha_j^{\text{B}}\}}}} \qquad (6)$$

Each α_i^{B} was calculated in the Coulomb field approximation[18] but using a surface integral approach.[19] The van der Waals molecular surface (the envelope surface of a set of intersecting

spheres with given atomic radii centered on the atoms) was tiled using the GEPOL[20] algorithm. The coordinates of the center of each *tesserae*, r_j, the normalized vector perpendicular to the surface at this point, $n(r_j)$, and the area of each *tesserae*, Δa_j, were applied in the expression below for obtaining the inverse of α_i^B (bold letters are denoting vectors):

$$\alpha_i^{B\,-1} = \frac{1}{4\pi}\int_S \frac{(\mathbf{r}-\mathbf{R}_i)\,\mathbf{n}(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_i|^4}\,d^2r = \frac{1}{4\pi}\sum_{j=1}^{N_{tesserae}} \frac{(\mathbf{r}_j-\mathbf{R}_i)\,\mathbf{n}(\mathbf{r}_j)}{|\mathbf{r}_j-\mathbf{R}_i|^4}\Delta a_j \quad (7)$$

where $\mathbf{R}_i$ is the position vector for the ith atom. Bondi's radii[21] for the atoms ($H = 0.120$ nm, C = 0.170 nm, N = 0.155 nm, O = 0.152 nm) were used in the construction of the van der Waals molecular surface.

The generalized principle of electronegativity equalization for an N atomic site molecule in a continuum-dielectric solvent is expressed by

$$\langle\chi_1\rangle = \langle\chi_2\rangle = \langle\chi_3\rangle = ... = \langle\chi_N\rangle \quad (8)$$

with

$$\langle\chi_i\rangle = \frac{\partial\Delta G_{sol}}{\partial q_{i,s}} = \frac{\partial(E_{EN,s}+\Delta G_{el})}{\partial q_{i,s}} = \chi_i^* + \eta_i^* q_{i,s} + \sum_{\substack{j=1\\(j\neq i)}}^{N}\eta_{ij}(r_{ij})q_{j,s} + \left(\frac{1}{\epsilon}-1\right)\frac{q_{i,s}}{\alpha_i^B} + \left(\frac{1}{\epsilon}-1\right)\sum_{\substack{j=1\\(j\neq i)}}^{N}\frac{q_{j,s}}{f_{GB}(r_{ij})}$$

thus, we can easily obtain the linear system

$$\begin{cases} Q = \sum_{j=1}^{N} q_{j,s} \\ \chi_1^* - \chi_2^* = \left[\eta_2^* + \left(\frac{1}{\epsilon}-1\right)\frac{1}{\alpha_2^B}\right]q_{2,s} - \left[\eta_1^* + \left(\frac{1}{\epsilon}-1\right)\frac{1}{\alpha_1^B}\right]q_{1,s} + \\ \sum_{\substack{j=1\\(j\neq 2)}}^{N}\left[\eta_{2j}(r_{2j}) + \left(\frac{1}{\epsilon}-1\right)\frac{1}{f_{GB}(r_{2j})}\right]q_{j,s} - \sum_{\substack{k=1\\(k\neq 1)}}^{N}\left[\eta_{1k}(r_{1k}) + \left(\frac{1}{\epsilon}-1\right)\frac{1}{f_{GB}(r_{1k})}\right]q_{k,s} \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \chi_1^* - \chi_N^* = \left[\eta_N^* + \left(\frac{1}{\epsilon}-1\right)\frac{1}{\alpha_N^B}\right]q_{N,s} - \left[\eta_1^* + \left(\frac{1}{\epsilon}-1\right)\frac{1}{\alpha_1^B}\right]q_{1,s} + \\ \sum_{\substack{j=1\\(j\neq i)}}^{N}\left[\eta_{Nj}(r_{Nj}) + \left(\frac{1}{\epsilon}-1\right)\frac{1}{f_{GB}(r_{Nj})}\right]q_{j,s} - \sum_{\substack{k=1\\(k\neq 1)}}^{N}\left[\eta_{1k}(r_{1k}) + \left(\frac{1}{\epsilon}-1\right)\frac{1}{f_{GB}(r_{1k})}\right]q_{k,s} \end{cases} \quad (9)$$

to be solved to obtain the charges.

The comparison between the linear system of equations (4) and (9) shows that the above hardness matrix is a matrix of effective atomic hardnesses. These effective hardnesses contain the dielectric effect of the solvent. Note that for ϵ greater than one, the effective atomic hardnesses are less than the atomic hardnesses in vacuum. This effect was described by Pearson,[22] based on experimental values of the free enthalpy of hydration, and by Lipinsky and Komorowski,[23] based on a virtual charge model for the solvent effect and using a description of the electronic structure at a semiempirical level.

Semiempirical and Ab Initio Quantum Chemistry Calculations. The geometries and CM1[24] charges of the organic molecules were taken from the database used in our recent paper.[25] Also, single point calculations in vacuum were done for the Ala[1]-Ala[2]-Ala[3]-Ala[4]-Ala[5]-Ala[6]-Ala[7] heptapeptide, $(Ala)_7$, Gly[1]-Ala[2]-Ser[3]-Gly[4]-Ala[5]-Ser[6]-Gly[7]-Ala[8]-Ser[9]-Gly[10]-Ala[11]-Ser[12]-Gly[13]-Ala[14]-Ser[15], $(GAS)_5$ and $(Ala)_{15}$ pentadecapeptides, at the semiempirical PM3 level using AMSOL6.6[26] and at ab initio B3LYP/6-31G(d) and B3LYP/6-311G(d,p) levels using GAUSSIAN03.[27] Single point calculations in continuum-dielectric solvent were done for the polypeptides at the PM3 level using the Solvation Model 5.4, SM5.4, as implemented in AMSOL6.6[26] and at the ab initio level using the Integral Equation Formalism for the Polarizable Continuum Model, IEFPCM, as implemented in GAUSSIAN03.[27]

The amino acid building module from the MacSpartan Pro v1.02 program[28] was used to create the polypeptide molecules. These geometries were not optimized.

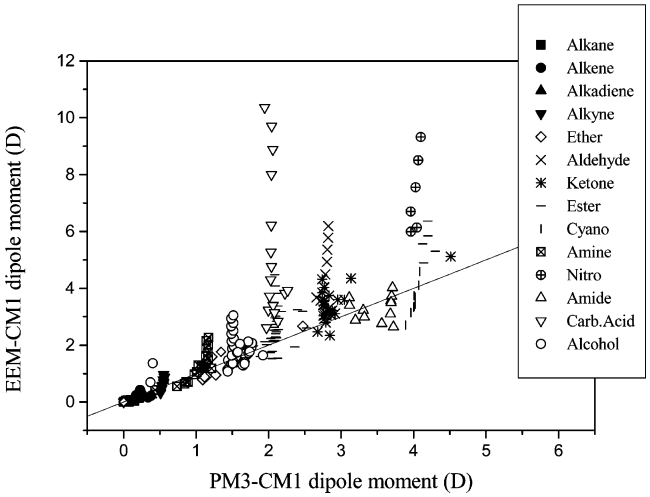

Figure 2. EEM-CM1 dipole moment versus PM3-CM1 dipole moment for molecules classified in different organic functions (see insert). The continuous line is the identity line.

III. Results and Discussion

Organic Molecules in the Vacuum. Figure 2 exemplifies the problem of the EEM as the molecule size grows. The values of the dipole moments calculated using EEM-CM1 were obtained from our recent EEM parameterization[25] with the CM1 charges calculated at the semiempirical PM3 level as targets.

The spiny pattern seen in Figure 2 results from methylene group addition to the molecular structure (e.g., in the aldehyde series, molecules go from propanal to nonanal; in the carboxylic acid series, from propanoic acid to eicosanoic acid; in the nitro series, from nitropropane to nitrohexane). In the aldehyde series, the positive charge center of the molecule (situated on the alkyl chain) is pushed away from the negative charge center (situated on the COH group) as the chain increases.

This feature in the dipole moment calculation is an inherent problem of EEM and does not depend on the parameterization. The interesting aspect is that the EEM charges are well correlated to CM1 charges, but the problem is not evidenced if molecular observables, such as the dipole moments, are not calculated.

Recently, Chelli et al.[10] have suggested the use of the atom−atom charge-transfer model, AACT, to avoid the EEM shortcoming. The AACT model has been introduced as an empirical procedure that avoids some of the instabilities in the fitting procedure of the EEM parameters when the electrostatic potential (ESP) derived charges are used as the target charges.[5c] The AACT adds molecular topology as an important factor in the charge-transfer control.

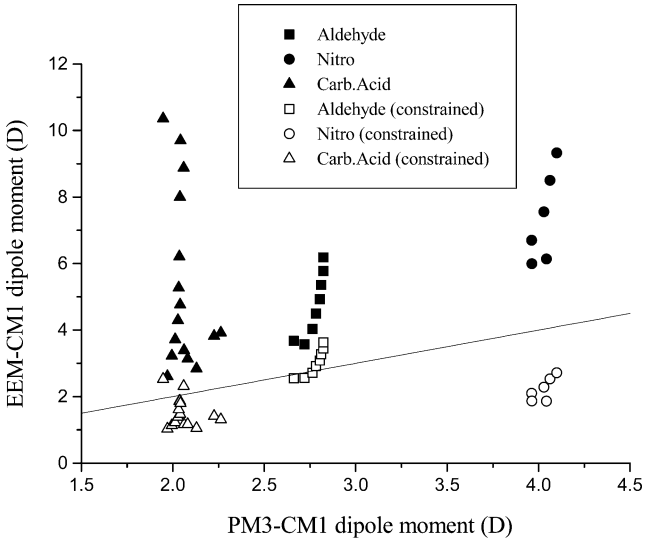

Figure 3. Effect of the constrained charge approximation in the aldehyde, carboxylic acid, and nitro series. The continuous line is the identity line.

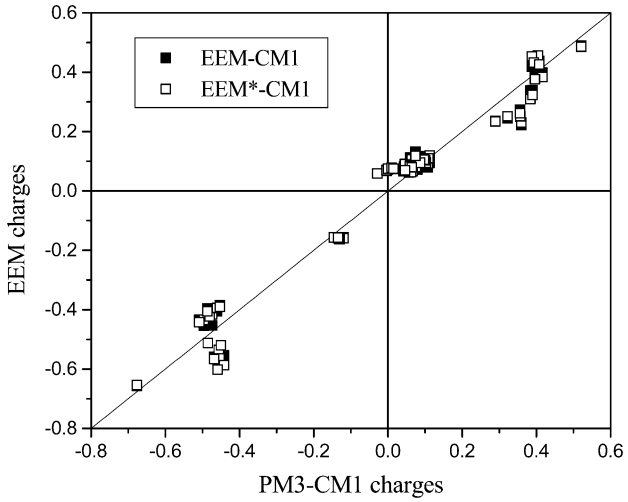

Figure 4. EEM-CM1 ($y = ax + b$, $a = 0.99$, $b = -2.7 \times 10^{-6}$, $R = 0.984$, $SD = 0.054$, $F = 2134$, $N = 73$) and EEM*-CM1 ($y = ax + b$, $a = 0.99$, $b = -2.7 \times 10^{-6}$, $R = 0.982$, $SD = 0.057$, $F = 1883$, $N = 73$) charges versus PM3-CM1 charges for the $(Ala)_7$ polypeptide in the vacuum. The continuous line is the identity line.

TABLE 1: Atomic Electronegativities (χ^*) and Hardnesses (η^*) Used in the EEM Parameterizations[a]

	EEM-M[b]		QEq-PD[c]		QEq-M[c]		EEM-CM1[d]	
atom	χ^*	η^*	χ^*	η^*	χ^*	η^*	χ^*	η^*
H	101.67	317.62	101.98	319.17	89.33	319.17	89.58	424.40
C	131.00	208.71	116.99	232.09	119.38	258.89	107.36	328.93
N	244.43	304.09	178.42	299.08	174.05	325.87	156.45	309.59
O	196.02	255.58	190.92	344.35	205.55	371.15	231.93	477.13

[a] Electronegativities are in kcal mol^{-1} e^{-1} and hardnesses are in kcal mol^{-1} e^{-2}. [b] Ref 4. [c] Ref 29. [d] Ref 25.

Inspired by the work of Chelli et al.,[10] we have tested a simple approximation for the dipole moment calculation: the partial sum of charges pertaining to the defined atomic groups (e.g., chemical function) is fixed and the charge transfer between these groups is not permitted.

Figure 3 shows the effect of the approximation (named as the constrained charge approximation, CCA) on the aldehyde, nitro, and carboxylic acid series. Each molecule has been broken into two parts: the headgroup (−COH, $-NO_2$, or −COOH) and the tail group (alkyl chain). In this case, each headgroup and tail group net charge has been constrained to zero. Note that the approximation of a null charge for these headgroups has been used in all the molecular mechanics force fields and that no EEM reparameterization has been done. As can be seen in Figure 3, the CCA has reduced the overpolarization of the EEM, resulting in reasonable dipole moments.

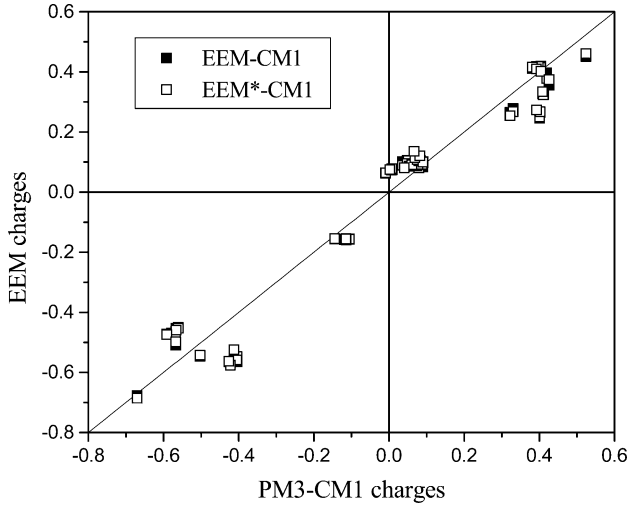

Figure 5. EEM-CM1 ($y = ax + b$, $a = 0.96$, $b = -2.6 \times 10^{-6}$, $R = 0.972$, $SD = 0.072$, $F = 1218$, $N = 73$) and EEM*-CM1 ($y = ax + b$, $a = 0.96$, $b = -2.6 \times 10^{-6}$, $R = 0.973$, $SD = 0.071$, $F = 1254$, $N = 73$) charges versus PM3-CM1 charges for the $(Ala)_7$ polypeptide in the continuum-dielectric solvent ($\epsilon = 78.5$). The continuous line is the identity line.

TABLE 2: Dipole Moment Vector for the $(Ala)_7$ Polypeptide (α-helical conformation) in Vacuum[a]

	dipole moment (values in Debye)				
model	μ	μ_X	μ_Y	μ_Z	$\sim\%\Delta\mu$[h]
EEM-CM1[b]	11.45	−11.11	0.45	2.77	55
EEM-M[c]	19.51	−18.96	0.53	4.59	23
QEq-M[d]	13.02	−12.65	0.01	3.05	49
QEq-PD[d]	10.19	−9.93	0.01	2.27	60
EEM*-CM1	17.92	−17.34	0.72	4.45	30
EEM*-M	26.50	−24.74	0.37	9.48	4
QEq*-M	23.67	−23.08	0.30	5.25	7
QEq*-PD	14.84	−14.05	0.19	4.77	42
OPLS-AA[e]	22.22	−21.76	0.54	4.50	13
PM3-CM1[f]	27.64	−26.89	0.59	6.36	9
B3LYP/6-31G(d)[g]	25.35	−24.87	0.36	4.86	0.5
B3LYP/6-311G(d,p)[g]	25.47	−24.99	0.33	4.90	0

[a] An asterisk denotes the constrained charge approximation in the electronegativity equalization method (e.g., EEM*-CM1, QEq*-PD, ...). [b] Ref 25. [c] Ref 4. [d] Ref 29. [e] Ref 30. [f] Dipole moment vector calculated using CM1 charges at the semiempirical PM3 level (ref 26). [g] Expected value of the dipole moment vector operator calculated at the ab initio level. [h] $\%\Delta\mu = 100|\mu_{B3LYP/6-311G(d,p)} - \mu_{model}|/\mu_{B3LYP/6-311G(d,p)}$

Polypeptides in a Vacuum and Continuum-Dielectric Solvent. We tested the CCA for polypeptide molecules with the two following aims: (i) we have used different EEM parameterizations in order to determine whether the CCA is dependent upon these parameterizations and (ii) we have considered the solvent effect in the polypeptide charge redistribution.

The following EEM parameterizations have been used: the EEM-CM1,[25] the Bakowies and Thiel[29] parameterization based on Mulliken and ESP charges calculated at the RHF/6-31G* level (QEq-M and QEq-PD, respectively), and the Mortier et al.[4] parameterization based on Mulliken charges calculated at the RHF/STO-3G level (the EEM-M). The CCA consisted of

TABLE 3: Dipole Moment Vector for the $(Ala)_7$ Polypeptide (α-helical conformation) in Continuum-Dielectric Solvent ($\epsilon = 78.5$)[a]

model	dipole moment (values in Debye)			
	μ	μ_X	μ_Y	μ_Z
EEM-CM1[b]	29.97	−29.38	1.11	5.80
EEM-M[c]	92.62	−91.29	2.40	15.43
QEq-M[d]	45.40	−44.66	0.17	8.16
QEq-PD[d]	36.88	−36.30	0.20	6.54
EEM*-CM1	26.94	−25.76	1.38	7.75
EEM*-M	60.30	−55.56	2.25	23.33
QEq*-M	40.79	−39.18	1.10	11.28
QEq*-PD	27.72	−25.81	0.98	10.06
PM3-CM1[e]	38.66	−37.14	1.15	10.69
B3LYP/6-31G(d)[f]	34.44	−33.44	0.96	8.16

[a] An asterisk denotes the constrained charge approximation in the electronegativity equalization method (e.g., EEM*-CM1, QEq*-PD, ...). [b] Ref 25. [c] Ref 4. [d] Ref 29. [e] Dipole moment vector calculated using CM1 charges at the semiempirical PM3 level (ref 26). [f] Expected value of the dipole moment vector operator in continuum-dielectric solvent calculated at the ab initio level using the IEFPCM (ref 27).

TABLE 4: Dipole Moment Vector for the $(GAS)_5$ Polypeptide (α-helical conformation, see Figure 8a) in Vacuum[a]

model	dipole moment (values in Debye)				
	μ	μ_X	μ_Y	μ_Z	$\sim\%\Delta\mu$[g]
EEM-CM1[b]	20.81	20.69	−1.56	−1.64	72
QEq-M[c]	23.61	23.50	−1.64	−1.56	68
QEq-PD[c]	18.83	18.75	−1.32	−1.14	75
EEM*-CM1	49.78	49.72	−0.92	−2.15	33
QEq*-M	69.64	69.59	−0.72	−2.51	6
QEq*-PD	45.52	45.46	−0.38	−2.28	38
OPLS-AA[d]	61.47	61.40	−2.15	−2.04	17
PM3-CM1[e]	74.84	74.77	−1.79	−2.77	1
B3LYP/6-31G(d)[f]	72.72	72.68	−2.05	−1.64	1.5
B3LYP/6-311G(d,p)[f]	73.85	73.81	−2.09	−1.72	0

[a] An asterisk denotes the constrained charge approximation in the electronegativity equalization method (e.g., EEM*-CM1, QEq*-PD, ...). [b] Ref 25. [c] Ref 29. [d] Ref 30. [e] Dipole moment vector calculated using CM1 charges at the semiempirical PM3 level (ref 26). [f] Expected value of the dipole moment vector operator calculated at the ab initio level. [g] $\%\Delta\mu = 100|\mu_{B3LYP/6-311G(d,p)} - \mu_{model}|/\mu_{B3LYP/6-311G(d,p)}$.

TABLE 5: Dipole Moment Vector for the $(GAS)_5$ Polypeptide (see Figure 8b) in Vacuum[a]

model	dipole moment (values in Debye)				
	μ	μ_X	μ_Y	μ_Z	$\sim\%\Delta\mu$[g]
EEM-CM1[b]	15.68	15.51	1.20	−2.00	66
QEq-M[c]	16.40	16.16	1.64	−2.27	64
QEq-PD[c]	12.80	12.58	1.46	−1.85	72
EEM*-CM1	31.67	31.62	1.14	−1.48	31
QEq*-M	42.10	42.04	1.71	−1.59	8
QEq*-PD	25.43	25.39	0.99	−1.04	44
OPLS-AA[d]	38.72	38.54	2.31	−2.91	15
PM3-CM1[e]	46.63	46.45	2.57	−3.22	2
B3LYP/6-31G(d)[f]	45.06	44.86	1.90	−3.80	1
B3LYP/6-311G(d,p)[f]	45.60	45.40	1.98	−3.85	0

[a] An asterisk denotes the constrained charge approximation in the electronegativity equalization method (e.g., EEM*-CM1, QEq*-PD, ...). [b] Ref 25. [c] Ref 29. [d] Ref 30. [e] Dipole moment vector calculated using CM1 charges at the semiempirical PM3 level (ref 26). [f] Expected value of the dipole moment vector operator calculated at the ab initio level. [g] $\%\Delta\mu = 100|\mu_{B3LYP/6-311G(d,p)} - \mu_{model}|/\mu_{B3LYP/6-311G(d,p)}$.

the assumption of null net charge for each amino acid and it was denoted by an asterisk for each EEM parameterization (e.g., EEM*-CM1, EEM*-M ...). The χ_i^* and η_i^* for the EEM parameterization can be seen in Table 1.

TABLE 6: Dipole Moment Vector for the $(GAS)_5$ Polypeptide (α-helical conformation, see Figure 8a) in Continuum-Dielectric Solvent ($\epsilon = 78.5$)[a]

model	dipole moment (values in Debye)			
	μ	μ_X	μ_Y	μ_Z
EEM-CM1[b]	66.04	65.85	−3.80	−3.26
QEq-M[c]	101.21	100.97	−5.67	−4.03
QEq-PD[c]	84.30	84.11	−4.42	−3.54
EEM*-CM1	63.75	63.60	−2.15	−3.79
QEq*-M	97.50	97.34	−2.29	−5.19
QEq*-PD	66.95	66.78	−1.51	−4.59
PM3-CM1[d]	90.91	90.77	−2.21	−4.62
B3LYP/6-31G(d)[e]	89.93	89.81	−2.54	−3.75

[a] An asterisk denotes the constrained charge approximation in the electronegativity equalization method (e.g., EEM*-CM1, QEq*-PD, ...). [b] Ref 25. [c] Ref 29. [d] Dipole moment vector calculated using CM1 charges at the semiempirical PM3 level (ref 26). [e] Expected value of the dipole moment vector operator in continuum-dielectric solvent calculated at the ab initio level using the IEFPCM (ref 27).

TABLE 7: Dipole Moment Vector for the $(GAS)_5$ Polypeptide (see Figure 8b) in Continuum-Dielectric Solvent ($\epsilon = 78.5$)[a]

model	dipole moment (values in Debye)			
	μ	μ_X	μ_Y	μ_Z
EEM-CM1[b]	48.87	48.55	2.72	−4.91
QEq-M[c]	70.21	69.63	4.81	−7.67
QEq-PD[c]	57.57	57.11	3.83	−6.15
EEM*-CM1	42.66	42.53	2.18	−2.62
QEq*-M	61.52	61.35	3.39	−3.06
QEq*-PD	40.71	40.61	2.14	−2.05
PM3-CM1[d]	57.26	56.98	4.38	−3.58
B3LYP/6-31G(d)[e]	56.69	56.45	3.35	−4.03

[a] An asterisk denotes the constrained charge approximation in the electronegativity equalization method (e.g., EEM*-CM1, QEq*-PD, ...). [b] Ref 25. [c] Ref 29. [d] Dipole moment vector calculated using CM1 charges at the semiempirical PM3 level (ref 26). [e] Expected value of the dipole moment vector operator in continuum-dielectric solvent calculated at the ab initio level using the IEFPCM (ref 27).

Figures 4 and 5 show the EEM-CM1 and EEM*-CM1 charges for the Alanine heptapeptide in a vacuum and continuum-dielectric solvent compared to the PM3-CM1 charges. The linear regressions are statistically identical but the clustering pattern of points on the third quadrant of the Figures 4 and 5 is not the same, indicating that there are differences between the GBEEM and the SM5.4P in the treatment of the solvent effect on the solute electronic structure.

The dipole moment vectors for $(Ala)_7$ in vacuum and solvent as calculated by the different charge models are presented in the Tables 2 and 3, respectively. Table 2 lists ab initio values for reference. Thus, we can compare the accuracy of the methods. As can be seen in Table 2, the best agreement has been obtained with the EEM*-M, QEq*-M, and PM3-CM1 (in this order). We must note that all the EEM* models (with CCA corrections) were more accurate than the EEM models. Table 3 presents the dipole moment vectors in dielectric solvent. The CCA reduced the dipole moment produced by all EEM parameterizations. The QEq*-M model has a dipole moment near to that of the PM3-CM1 model.

In vacuum, the EEM*-M dipole moments for each amino acid (∼6 D) are greater than the other EEM parameterizations, OPLS-AA and PM3-CM1. In solvent, the dipole moments of the amino acids, as calculated by the EEM*-M, are amplified,

TABLE 8: Dipole Moment Vector for the $(Ala)_{15}$ Polypeptide (α-helical conformation, see Figure 9a) in Vacuum[a]

	dipole moment (values in Debye)				
model	μ	μ_X	μ_Y	μ_Z	$\sim\%\Delta\mu$[g]
EEM-CM1[b]	17.87	17.84	0.13	1.02	73
QEq-M[c]	19.87	19.84	0.11	1.02	70
QEq-PD[c]	15.43	15.41	0.04	0.68	77
EEM*-CM1	46.14	46.11	0.00	1.73	30
QEq*-M	65.23	65.20	−0.14	1.90	2
QEq*-PD	41.84	41.80	−0.01	1.91	37
OPLS-AA[d]	49.58	49.57	−0.49	1.01	25
PM3-CM1[e]	67.51	67.47	−0.65	2.41	2
B3LYP/6-31G(d)[f]	65.63	65.62	−0.34	1.16	1
B3LYP/6-311G(d,p)[f]	66.23	66.22	−0.35	1.22	0

[a] An asterisk denotes the constrained charge approximation in the electronegativity equalization method (e.g., EEM*-CM1, QEq*-PD, ...). [b] Ref 25. [c] Ref 29. [d] Ref 30. [e] Dipole moment vector calculated using CM1 charges at the semiempirical PM3 level (ref 26). [f] Expected value of the dipole moment vector operator calculated at the ab initio level. [g] $\%\Delta\mu = 100|\mu_{B3LYP/6-311G(d,p)} - \mu_{model}|/\mu_{B3LYP/6-311G(d,p)}$.

TABLE 9: Dipole Moment Vector for the $(Ala)_{15}$ Polypeptide (see Figure 9b) in Vacuum[a]

	dipole moment (values in Debye)				
model	μ	μ_X	μ_Y	μ_Z	$\sim\%\Delta\mu$[g]
EEM-CM1[b]	9.72	8.65	0.18	4.43	69
QEq-M[c]	10.69	9.49	0.59	4.89	65
QEq-PD[c]	8.24	7.22	0.57	3.92	73
EEM*-CM1	20.66	19.72	1.62	5.96	33
QEq*-M	29.00	28.07	2.30	6.95	6
QEq*-PD	16.88	15.83	1.42	5.70	45
OPLS-AA[d]	22.96	21.57	2.14	7.58	26
PM3-CM1[e]	31.85	30.28	1.33	9.77	3
B3LYP/6-31G(d)[f]	30.55	29.43	1.70	8.02	1
B3LYP/6-311G(d,p)[f]	30.89	29.78	1.83	8.01	0

[a] An asterisk denotes the constrained charge approximation in the electronegativity equalization method (e.g., EEM*-CM1, QEq*-PD, ...). [b] Ref 25. [c] Ref 29. [d] Ref 30. [e] Dipole moment vector calculated using CM1 charges at the semiempirical PM3 level (ref 26). [f] Expected value of the dipole moment vector operator calculated at the ab initio level. [g] $\%\Delta\mu = 100|\mu_{B3LYP/6-311G(d,p)} - \mu_{model}|/\mu_{B3LYP/6-311G(d,p)}$.

TABLE 10: Dipole Moment Vector for the $(Ala)_{15}$ Polypeptide (α-helical conformation, see Figure 9a) in Continuum-Dielectric Solvent ($\epsilon = 78.5$)[a]

	dipole moment (values in Debye)			
model	μ	μ_X	μ_Y	μ_Z
EEM-CM1[b]	61.87	61.84	−0.18	2.18
QEq-M[c]	94.38	94.33	−0.28	3.00
QEq-PD[c]	76.83	76.79	−0.44	2.33
EEM*-CM1	59.67	59.57	0.10	3.44
QEq*-M	91.61	91.47	−0.61	5.06
QEq*-PD	61.66	61.48	−0.60	4.64
PM3-CM1[d]	81.91	81.77	−1.35	4.55
B3LYP/6-31G(d)[e]	79.44	79.39	−0.70	2.80

[a] An asterisk denotes the constrained charge approximation in the electronegativity equalization method (e.g., EEM*-CM1, QEq*-PD, ...). [b] Ref 25. [c] Ref 29. [d] Dipole moment vector calculated using CM1 charges at the semiempirical PM3 level (ref 26). [e] Expected value of the dipole moment vector operator in continuum-dielectric solvent calculated at the ab initio level using the IEFPCM (ref 27).

resulting in exaggerated values (∼10 D). Also, the charges of the N and O atoms are overestimated (not shown). Therefore, the EEM-M parameterization was excluded from the following tests.

Tables 4−11 show the dipole moment vectors as calculated by the charge models and by the ab initio method for the $(GAS)_5$ (Figure 6) and $(Ala)_{15}$ (Figure 7) pentadecapeptides, such that two different conformations are evaluated for each polypeptide. As for the heptapeptide, all the EEM calculations in vacuum using the CCA were more accurate than the EEM calculations without the CCA (using the ab initio values as references). The PM3-CM1 and QEq*-M models have produced similar values and agree well with the ab initio values.

TABLE 11: Dipole Moment Vector for the $(Ala)_{15}$ Polypeptide (see Figure 9b) in Continuum-Dielectric Solvent ($\epsilon = 78.5$)[a]

	dipole moment (values in Debye)			
model	μ	μ_X	μ_Y	μ_Z
EEM-CM1[b]	32.70	31.18	1.64	9.71
QEq-M[c]	50.39	48.30	3.45	13.96
QEq-PD[c]	41.42	39.64	2.76	11.68
EEM*-CM1	28.52	26.49	2.12	10.35
QEq*-M	43.80	41.05	2.73	15.04
QEq*-PD	28.96	26.02	1.47	12.62
PM3-CM1[d]	40.77	37.12	2.15	16.73
B3LYP/6-31G(d)[e]	38.36	35.78	2.45	13.61

[a] An asterisk denotes the constrained charge approximation in the electronegativity equalization method (e.g., EEM*-CM1, QEq*-PD, ...). [b] Ref 25. [c] Ref 29. [d] Dipole moment vector calculated using CM1 charges at the semiempirical PM3 level (ref 26). [e] Expected value of the dipole moment vector operator in continuum-dielectric solvent calculated at the ab initio level using the IEFPCM (ref 27).

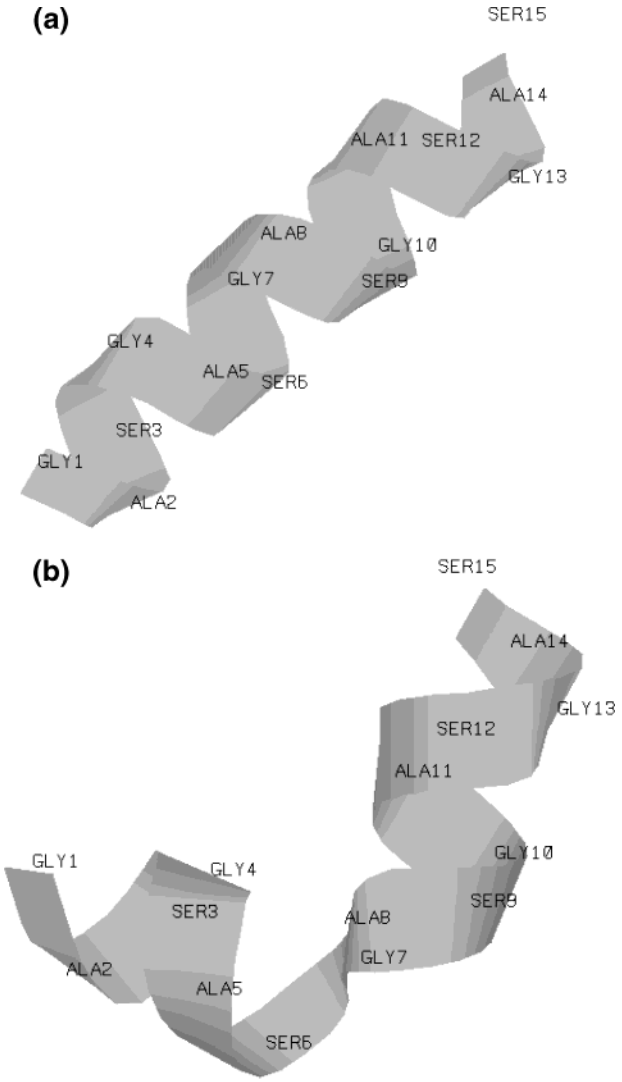

Figure 6. Ribbon representation of the two $(GAS)_5$ conformers: (a) helical conformation; (b) perturbed conformation.

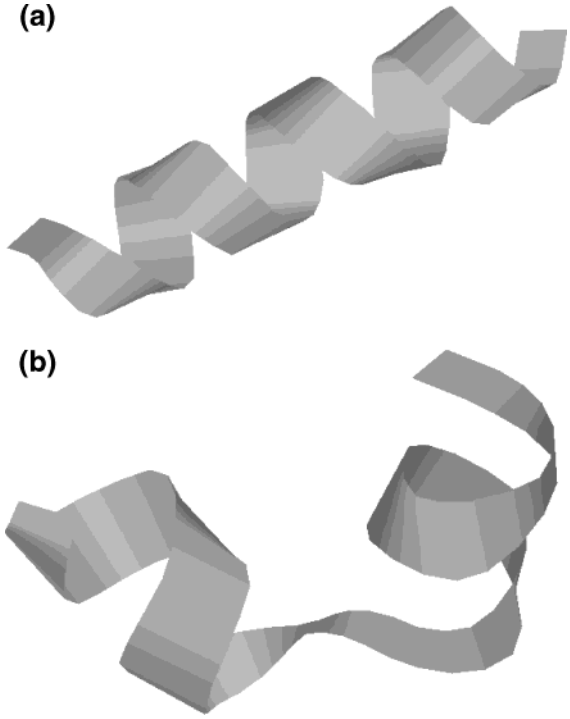

Figure 7. Ribbon representation of the two $(Ala)_{15}$ conformers: (a) helical conformation; (b) perturbed conformation.

In continuum-dielectric solvent, the PM3-CM1 had a very good agreement with the IEFPCM at the ab initio B3LYP/6-31G(d) level (we used Bondi's radii, van der Waals surface, and electrostatic scaling factor equal to 1.0 for the IEFPCM calculations). In general, the QEq*-M had a higher polarization when compared to the PM3-CM1 or IEFPCM. Indeed, the QEq-PD had a better performance than the QEq*-M in the continuum-dielectric calculations for the polypeptides, but this is probably due to a fortuitous compensation of terms since its results in vacuum were not good.

IV. Conclusions and Perspectives

We have shown that the constrained charge approximation can treat the intrinsic problem of EEM in the prediction of molecular observables. The GBEEM has been compared to other models of the continuum-dielectric solvent effect on the renormalization of the dipole moment vector. Namely, we have compared GBEEM to the SM5.4 continuum model at the semiempirical PM3 level and to the IEFPCM continuum model at the ab initio B3LYP/6-31G(d) level. Good or reasonable agreement has been obtained between the QEq*-M (GBEEM using atomic electronegativities and hardnesses from the QEq-M) and these more sophisticated methods.

We believe that the models generated via EEM can replace the fixed charge models (e.g., OPLS-AA). Moreover, these EEM models consider the implicit solvent effect and the effect of conformational changes on the solute charge distribution. We believe that the results presented here show the potential of the EEM model used in combination with the generalized Born solvation model.

Acknowledgment. L.G.D., H.C., and J.P.S.F. are grateful to the FAPESP and CNPq Brazilian agencies for their support. K.S. is a FAPESP graduate fellow (01/05852-8). The ab initio calculations were performed on an IBM P690 Model 681 at the North Carolina Supercomputing Center.

References and Notes

(1) Chelli, R.; Procacci, P. *J. Chem. Phys.* **2002**, *117*, 9175.
(2) Ribeiro, M. C. C.; Almeida, L. C. J. *J. Chem. Phys.* **2000**, *113*, 4722.
(3) York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *104*, 159.
(4) Mortier, W. J.; Ghosh, S. K.; Shankar, S. *J. Am. Chem. Soc.* **1986**, *108*, 4315.
(5) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515. (b) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, *103*, 4730. (c) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741.
(6) Cho, K.-H.; Kang, Y. K.; No, K. T.; Scheraga, H. A. *J. Phys. Chem. B* **2001**, *105*, 3624.
(7) Cong, Y.; Yang, Z.-Z. *Chem. Phys. Lett.* **2000**, *316*, 324.
(8) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
(9) Yang, Z.-Z.; Wang, C. S. *J. Phys. Chem. A* **1997**, *101*, 6315.
(10) Chelli, R.; Procacci, P.; Righini, R.; Califano, S. *J. Chem. Phys.* **1999**, *111*, 8569.
(11) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
(12) Böttcher, C. J. F. *Theory of Electric Polarization*; Elsevier: Amsterdam, 1952.
(13) Sharp, K.; Jean-Charles, A.; Honig, B. H. *J. Phys. Chem.* **1992**, *96*, 3822.
(14) Nilar, S. H. *J. Mol. Struct. (THEOCHEM)* **1996**, *363*, 97.
(15) Dias, L. G.; Shimizu, K.; Farah, J. P. S.; Chaimovich, H. *Chem. Phys.* **2002**, *282*, 237.
(16) Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1996**, *118*, 672.
(17) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
(18) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.
(19) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, 102, 10983.
(20) Pascual-Ahuir, J. L.; Silla, E.; Tuñon, I. *J. Comput. Chem.* **1994**, *15*, 1127.
(21) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
(22) Pearson, R. G. *J. Am. Chem. Soc.* **1986**, *108*, 6109.
(23) Lipinski, J.; Komorowski, L. *Chem. Phys. Lett.* **1996**, *262*, 449.
(24) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.
(25) Menegon, G.; Shimizu, K.; Farah, J. P. S.; Dias, L. G.; Chaimovich, H. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5933.
(26) Hawkins, G. D., Giesen, D. J., Lynch, G. C., Chambers, C. C., Rossi, I., Storer, J. W., Li, J., Winget, P., Rinaldi, D., Liotard, D. A., Cramer, C. J., Truhlar, D. G. *AMSOL-version 6.6*, 1999.
(27) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. A., Montgomery, J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Peterson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Know, J. E., Hratchian, H. P., Cross, J. B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., Pople, J. A. *Gaussian 03, Revision B.04*; Gaussian, Inc.: Pittsburgh, PA, 2003.
(28) Deppmeier, B. J.; Driessen, A. J.; Hehre, W. J.; Jonson, J. A.; Klunzinger, P. E.; Yu, J. *MacSpartan Pro v1.02*, Wavefunction, 1999−2000.
(29) Bakowies, D.; Thiel, W. *J. Comput. Chem.* **1996**, *17*, 87.
(30) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *113*, 11225.