JCTC Journal of Chemical Theory and Computation

Atomic Charge Parameters for the Finite Difference Poisson−Boltzmann Method Using Electronegativity Neutralization

Qingyi Yang and Kim A. Sharp*

Johnson Research Foundation and Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received January 6, 2006

Abstract: An optimization of Rappe and Goddard's charge equilibration (QEq) method of assigning atomic partial charges is described. This optimization is designed for fast and accurate calculation of solvation free energies using the finite difference Poisson−Boltzmann (FDPB) method. The optimization is performed against experimental small molecule solvation free energies using the FDPB method and adjusting Rappe and Goddard's atomic electronegativity values. Using a test set of compounds for which experimental solvation energies are available and a rather small number of parameters, very good agreement was obtained with experiment, with a mean unsigned error of about 0.5 kcal/mol. The QEq atomic partial charge assignment method can reflect the effects of the conformational changes and solvent induction on charge distribution in molecules. In the second section of the paper we examined this feature with a study of the alanine dipeptide conformations in water solvent. The different contributions to the energy surface of the dipeptide were examined and compared with the results from fixed CHARMm charge potential, which is widely used for molecular dynamics studies.

Introduction

Accurate representation of the charge distribution within molecules is essential for most atomic resolution modeling techniques. It is crucial for techniques such as finite difference Poisson−Boltzmann (PB) and Generalized Born (GB) methods, since their primary output is electrostatic potentials, energies, or forces. A significant barrier to many modeling studies is that of encountering a 'new' cofactor, ligand, or functional group, for which there are no parameters. Modeling cannot continue until parameters have been obtained. Bond stretch, angle, etc. parameters can usually be transferred from chemically similar groups, but atomic charges are a less local, more context dependent property. There is thus a continual need for rapid determination of molecular charges for use with specific modeling techniques. There are several alternative ways of representing molecular charge distributions. The most accurate uses the nuclear coordinate positions plus the full electron density (electronic orbital) distribution obtained from quantum mechanical (QM) calculations. This representation requires significant memory and computation, and it is not practical for most macromolecular modeling. Thus a reduced representation is required. The most common is point atomic charges, usually centered at the nuclei. For the point atomic charge representation, several approaches have been used to obtain parameter sets suitable for modeling. Again, QM methods are the most fundamental and, in principle, the most accurate. There are, however, two uncertainties encountered with this approach. First, there are different ways to obtain atomic charges from the orbitals, including Mullikan population analysis, apportioning the electron density by locating saddle points in the distribution, and fitting to the Coulomb potential at a set of points surrounding the molecule. The restrained electrostatic potential (RESP) fitting[1] is widely used since it fits to the principal property that the charges will actually be used for: The potential distribution around the molecule. Whatever method is used, however, the point charge representation is an approximation, even if the QM calculation were exact.

* Corresponding author phone: (215)573-3506; fax: (215)898-4217; e-mail: sharpk@mail.med.upenn.edu.

10.1021/ct060009c CCC: $33.50
Published on Web 06/01/2006

The second problem is producing QM based charges that mimic the condensed phase. Inclusion of, for example, a solvent reaction field in the QM calculations helps, but charges usually need to be further scaled or fine-tuned for maximum accuracy in modeling macromolecules in solution. Thus, even atomic charges that start with QM calculations are empirical. Another approach is to just parametrize charges by fitting calculated electrostatic properties to experiment. This can be done completely from scratch, as was the PARSE amino acid set for the finite difference PB (FDPB) method.[2] This worked well for the 20 amino acid side chains and the peptide backbone, with a restricted set of functional group types. It is difficult to extend this approach to more complicated groups/molecules without some starting guesses, however. Even parametrized charge sets usually require some preliminary calculations as a starting point.

Two important criteria for empirical atomic charge parametrization are what theoretical or experimental quantities are to be fit and what method(s) the charges are going to be used with. Atomic charges for molecular dynamics (MD) force fields are usually obtained through a combination of QM calculations and fitting to experiment and other theoretical calculations. In contrast, the OPLS set was fit to solvation data using free energy perturbation methods,[3] and this set would be preferred for solvation energy calculations using the same methods. The original PARSE set was fit to experimental solvation free energies obtained from partition data, using the FDPB method.[2] Several methods have been parametrized to reproduce RESP fit charges from a particular level of QM calculations.[4,5] Different semiempirical QM methods have been assessed by comparison to solvation free energies calculated by free energy perturbation.[6,7] Many of these different charge sets are simultaneously in use for different types of simulations, evidence that different charge sets work better with particular modeling methods, for particular molecules or for particular quantities. This is not surprising given that atomic charges are an abstraction, and all sets are empirical to some degree. Each of the above parametrization strategies has its advantages for particular applications. Given this, we decided that with the wide use of the PB model for macromolecular electrostatics, it would be useful to develop a rapid procedure for generating atomic charges for use specifically with the FDPB method. Our criteria were first that the method be simple and could be applied to diverse cofactors and functional groups with the same small set of parameters. Second, that it be rapid and require little user input. Given the approximate nature of atomic charges, methods that consume a lot of time or that require the user to explore multiple options are not cost-effective. Third, given the empirical nature of atomic charge sets, we wanted to parametrize specifically for a given method, FDPB. As with the original PARSE charge set, we chose solvation free energy data to parametrize against: Solvation free energies can be measured reliably, the electrostatic component can be accurately extracted, and it bears on the important solvation component of macromolecules for which FDPB is often used. However, the original PARSE strategy of ground up parametrization for each chemical subgroup type is difficult to extend to more complicated groups, and it becomes prone to subjectivity. On the other hand, QM based methods are computationally expensive, require several steps and choices of strategy (e.g. level of theory, basis sets, fitting methods), and in the end results in sets that are still partly empirical. We chose a middle course, selecting a charge equalization (CE) through an electronegativity method. Several implementations of this method have been described,[8−13] it is easy to implement, and it is quite flexible. Gilson et al. showed that the equalization method can be made to mimic more rigorous QM calculations well by the careful choice of atomic types and that the method can account for different resonance forms. In this work, however, we focus on reproducing experimental solvation free energies rather than QM calculations. We chose to use the charge equalization method of Rappe and Goddard.[11] They developed a very simple but general scheme to generate charge distributions for use in molecular dynamics simulation based only on molecular geometry and atomic properties. The input parameter data are just atomic ionization energy (IP), electron affinity (EA), available from standard tabulations of element properties, and covalent radius (R), obtained from crystallographic data. The original method was parametrized for gas (vacuum) phase.

In this study we describe an optimization of Rappe and Goddard's charge equilibration method (which they call QEq) for FDPB solvation energies by a combination of adjusting the element data and making a small addition to the algorithm to increase the number of the functional groups it can handle. The method proposed here is thus designed to be a successor to and extension of the PARSE charge set. In the first section of this paper we show the results of the parametrization on small molecule solvation free energies. Following a minimalist strategy we used a rather small set of parameters, a small training set and a large test set, since we believe this is likely to result in a robust set of charges. The QEq method of Rappe and Goddard potentially provides a significant enhancement of the electrostatic treatment in molecular dynamics because it allowed the atomic charge distribution to respond to the molecule's geometry changes, i.e., it is effectively a polarizable potential function, and it is very fast. While polarizable potentials for MD applications are not our primary focus, they are currently under active development and testing.[14−17] In the second section of this paper we applied our parametrized QEq method to a study of the alanine dipeptide conformations. We studied the Ramachandran energy map and compared the results of the QEq method with polarization to the fixed charge CHARMM potential[18] which is widely used for MD. Both sets of maps were calculated with and without a FDPB solvation contribution.

Methods

Calculation of Partial Atomic Charges. Given an atom's ionization potential (IP) and electron affinity (EA) Rappe and Goddard defined an atomic scale chemical potential by taking the derivative of the total electrostatic energy with respect to that atom's charge, Q_A, leading to[11]

$$\chi_A = \chi^0{}_A + J^0{}_{AA}Q_A + \sum_{B\neq A} J_{AB}Q_B \quad (1)$$

where $\chi^0{}_A = (IP+EA)/2$ is the electronegativity, and $J^0{}_{AA} = (IP-EA)$ is the Coulomb repulsion (self-Coulomb integral) between two electrons in A's outer orbital. J_{AB} is the interatomic electrostatic energy between atoms A and B. We followed Rappe and Goddard in representing the electron density of the atoms with Slater type orbitals

$$\phi(r,n) = B_n r^{n-1} e^{-\zeta r} \quad (2)$$

where r is distance from the nucleus, n is the row number of the element in the periodic table, B_n is a normalization constant, and ζ defines the size of the atom (orbital) using the covalent radius of an atom, R_A, through the relation

$$\zeta_A = (2n+1)/(4R_A) \quad (3)$$

For the special case of hydrogen, which has only one electron, the size is charge dependent: $\zeta_H = \zeta^\circ{}_H + Q_H$. J_{AB} is computed from the overlap integral of two Slater type orbitals with the form of eq 2:

$$J_{AB} = \int\int \phi_A(r, n_A)\frac{1}{|\mathbf{r}-\mathbf{s}|}\phi_A(s, n_B)\delta\mathbf{r}\delta\mathbf{s} \quad (4)$$

J_{AB} has the form of the Coulomb potential $1/R$ for large separations. For small separations as the electron density distributions of A and B overlap the interaction is shielded and J_{AB} plateaus to a constant value at $r = 0$. Given eq 1, the atomic charges of a molecule or group with N atoms are obtained by equating all the chemical potentials

$$\chi_i = \chi_N,\ i = 1, N-1 \quad (5)$$

subject to a user specified condition on the total charge of the molecule or group given by

$$\sum Q_i = Q_{net} \quad (6)$$

which leads to a simple matrix equation for the charges

$$C_{ij}Q_j = -D_i \quad (7)$$

where

$$C_{1j} = 1,\ D_1 = Q_{net} \quad (8a)$$

$$C_{ij} = J_{ij} - J_{1j},\ D_i = \chi^0{}_i - \chi^0{}_1,\ \text{for } i \neq 1 \quad (8b)$$

To evaluate the J_{AB} terms required for the elements C_{ij}, eq 4 is integrated numerically by transforming into elliptical coordinates with atoms A and B at the foci and using a Gaussian quadrature routine.[19] Equation 7 is then solved using the linear algebra routines from Numerical Recipes.[19] For each hydrogen atom, initially Q_H is set to zero, and $\zeta_H = \zeta^\circ{}_H$. After solving eq 6, ζ_H is updated using $\zeta_H = \zeta^\circ{}_H + Q_H$ with the current estimate of Q_H. Then eq 7 is resolved. This procedure is repeated to convergence (usually taking about 6−8 cycles). The algorithm was implemented as a Fortran 77 program called QEQUIL. The entire charge determination algorithm typically takes less than 0.1 s on a 2.5 GHz processor for compounds with up to several dozen atoms. Output is in the modified DelPhi PDB format with charge and radii in the occupancy and b-factor columns, so it can be read directly by DelPhi and GRASP.[20,21] The program QEQUIL is available from the authors upon request.

Table 1. Input Atomic Parameters

element	atomic no.	row	J (eV)	radius[a] (Å)	χ (eV) original	χ (eV) optimized
H	1	1	13.8904	0.371	4.528	4.498
C	6	2	10.126	0.759	5.343	5.723
N	7	2	11.76	0.715	6.899	8.599
O	8	2	13.364	0.669	8.741	8.961
F	9	2	14.498	0.706	10.874	6.374
S	16	3	8.972	0.947	6.928	6.268
Cl	17	3	9.892	0.994	8.564	5.464
Br	35	4	8.85	1.141	7.79	5.790
I	53	5	7.524	1.333	6.822	5.622

[a] Covalent radius used for charge equilibration only.

Solvation Free Energies. Solvation free energies were calculated using the FDPB/SA method as described previously,[2] using the FDPB package DELPHI. Radii were taken from the PARSE set, for H = 1.0 Å, C = 1.9 Å, O = 1.6 Å, N = 1.65 Å, S = 1.9 Å. Radii for the halogens were taken from the AMBER parm99 potential function.[22] Since we aimed to produce a second generation equivalent of the PARSE charges, radii were used as is, and we did not attempt to optimize them also. The covalent radii tabulated by Rappe and Goddard and in Table 1 are used only for the charge determination phase, as described by eqs 2 and 3, not for solvation free energy calculations. Solvent accessible surface (SAS) areas were calculated using the program SURFCV,[23] and the nonpolar contribution to solvation, ΔG^{np}(calc), is obtained by multiplying the SAS by the 'hydrophobic' coefficient $\gamma = 5$ cal/mol/Å^2. The electrostatic component of the solvation free energy, ΔG^{elec}(calc), is computed as the difference in free energy of the molecule between vacuum (exterior dielectric constant ϵ=1) and water (exterior dielectric constant ϵ=80). Solute molecules were assigned a dielectric $\epsilon = 2$, a commonly accepted value that accounts for small solute polarizability. Am1-bcc charges were obtained using the commands "bcctype" and "bcc" within the antechamber module of AMBER V7.0.[24] AMSOL charges were obtained using the AMSOL v 7.1 program,[25] using the class IV CM2 point charge model. AMSOL/GB solvation calculations were performed using the SM5.42R solvation model in the AMSOL v 7.1 program.

Training Set of Solutes. The 14 polar amino acid side-chain analogues and the peptide backbone analogue N, methyl-acetamide, were used as the core of the training set. The omitted amino acid side chains G, A, P, V, L, and I are purely nonpolar and so do not affect parametrization of charges. Experimental vapor-to-water transfer free energies, ΔG^{solv}(expt), were taken from Wolfenden.[26] For the solvation calculations the neutral form of ionizable residues was used, to conform to the experimental measurements.[26] Eight other compounds with a range of functional groups were also used for training (Table 3). Solvation energies for these were taken from Cabani et al.[27] The nonpolar part of the solvation free energy was assumed to be accurately represented by the

Table 2. Atom Pair Shielding Factors

atom pair	S_{AB}	atom pair	S_{AB}
C···N (amide)	0.66	C−O (aldehyde)	1.61
C═N (nitrile)	1.63	C−O (general)	1.03
C–N (general)	1.10	C≡C (-yne)	2.33
N═O (nitro)	2.33		

current SAS method and 'hydrophobic' coefficient γ, so γ was not reparametrized here. Thus to show more clearly the effects of charge parametrization we tabulate the calculated electrostatic contribution to solvation free energy, ΔG^{elec}(calc), and the 'experimental' electrostatic contribution to solvation free energy defined as $\Delta G^{elec}(expt) = \Delta G^{solv}(expt) - \Delta G^{np}(calc)$. Our parametrization was done against data for the vacuum-to-water transfer process for several reasons. First, the most common application of the FDPB method is for calculation of the net hydration free energy, which is the quantity obtained from vapor-water transfer data, so we wanted to optimize for this process. Second such data are of high experimental reliability, there being only one solvent to contend with. We did not aim to produce a general solvation parametrization for use with a range of organic solvents, such as the AMSOL set. This, while of more general applicability, would be less accurate for just water solvent applications.

Input Structures. Small molecules for the training and test sets were built with ISIS/DRAW 2.5 (MDL interactive systems, San Leandro, CA), and their conformations were optimized in InsightII (Accelrys, SanDiego, CA). The lowest energy conformer was used for solvation calculations, except for the study of the alanine dipeptide. For the alanine dipeptide, conformations with different Φ and Ψ values were built in CHARMM,[18] using a torsion angle grid of 10° from −180° to 180°. Experimental solvation free energies were taken from Radzicka and Wolfenden[26] for the side-chain and backbone analogues and from Cabani[27] for the other compounds.

Parametrization Strategy. To minimize the number of adjustable parameters, we initially chose to optimize by adjusting only the atomic electronegativities, χ_i. Electrostatic solvation free energies for the training set were first calculated with the original parameters of Rappe and Goddard[11] (Table 1). To guide the initial direction of parametrization, the resulting data were examined for systematic deviations that could be attributed to specific element electronegativities. For example, compounds that contained sulfur had systematically overestimated electrostatic solvation contributions. The mean squared error in electrostatic solvation free energy, averaged over the training set of 23 compounds in Table 3, was then minimized by adjusting each electronegativity in turn, keeping the others constant. The resulting roughly optimized electronegativities were then used as input into a systematic optimization of the mean squared error in electrostatic solvation free energy by using a grid search over χ_i for the nine elements involved, H, C, O, N, S, F, Cl, Br, and I. The χ_i's were varied over a range of 3.500−11.500 using a step size of 0.015.

Satisfactory optimization could be achieved by adjusting the χ_i's alone. However, since the charge equilibration methods are intrinsically limited in dealing with pi- and delocalized bonds,[28] it is unlikely that the best results can be obtained by optimizing just the individual atomic quantities, χ_i's. Similar considerations imply that adjusting the J_{AA}

Table 3. Training Set Solvation Free Energies (kcal/mol) Using Original and Optimized Parameter Sets

						ΔG^{elec} (calc)	
ID	molecule[a]	SAS (Å^2)	ΔG^{np}(calc)	ΔG^{solv}(expt)	ΔG^{elec}(expt)	original	optimized
1	1-heptyne	278	2.25	0.60	−1.65	−1.91	−2.87
2	fluromethane	133	1.53	−0.22	−1.53	−10.95	−1.77
3	1-chloropropane	208	1.90	−0.27	−2.18	−12.04	−2.16
4	1-bromopropane	224	1.98	−0.56	−2.54	−8.53	−2.66
5	1-iodopropane	233	2.03	−0.59	−2.61	−5.74	−2.62
6	propanenitrile	198	1.85	−3.85	−5.7	−2.89	−5.60
7	1-nitropropane	224	1.98	−3.34	−5.33	−12.02	−5.43
8	pentanal	303	2.38	−3.03	−5.41	−8.69	−5.38
9	N-propylguanidine (arg)	320	2.46	−10.92	−13.38	−5.87	−13.60
10	acetamine (asn)	226	1.99	−9.72	−11.71	−7.05	−11.63
11	acetic acid (asp)	222	1.97	−6.70	−8.67	−8.83	−7.79
12	methylthiol (cys)	202	1.87	−1.24	−3.11	−4.70	−2.91
13	propionamide (gln)	260	2.16	−9.42	−11.58	−7.62	−12.06
14	propionic acid (glu)	256	2.14	−6.47	−8.61	−9.46	−8.45
15	methylimidazole (his)	272	2.22	−10.25	−12.47	−5.81	−12.23
16	N-butylamine (lys)	286	2.29	−4.38	−6.67	−4.46	−8.32
17	methyl ethyl sulfide (met)	272	2.22	−1.49	−3.71	−5.81	−3.99
18	toluene (phe)	306	2.39	−0.76	−3.15	−1.54	−3.09
19	methanol (ser)	180	1.76	−5.08	−6.84	−8.23	−7.28
20	ethanol (thr)	218	1.95	−4.90	−6.85	−8.17	−7.31
21	methylindole (trp)	348	2.60	−5.91	−8.51	−2.84	−6.23
22	*p*-cresole (tyr)	318	2.45	−6.13	−8.58	−7.58	−7.30
23	*N*-methylacetamide (backbone)	266	2.19	−10.08	−12.27	−7.63	−13.15

[a] Amino acid side-chain/backbone analogue in brackets.

values would simply introduce more adjustable parameters without producing qualitatively improved fits, since both contribute in a linear fashion to the effective electronegativity in eq 1. To overcome this limitation, we introduced one additional parameter type, a shielding factor S_{AB} that scales the electrostatic interaction between two atoms, as described by a modified version of eq 1:

$$\chi_A = \chi^0{}_A + J^0{}_{AA}Q_A + \sum_{B \neq A} S_{AB}J_{AB}Q_B \qquad (9)$$

This allows us to incorporate the effect of bond delocalization or atomic-neighbor, bond-dipole, and atomic context type effects in a simple way. Shielding factors were initially set to 1 and then optimized using a grid search. Significantly improved fits could be obtained by introduction of several shielding factors, most involving atom pairs with higher order/resonance bonds (Table 2). The shielding factors for all other atom pairs were 1 and thus could be omitted.

Molecular Mechanics of Alanine Dipeptide. Alanine dipeptide was build using CHARMM, with the CHARMM27 potential.[29] Conformations for the Ramachandran plots were generated using a 10° increment grid of phi and psi angles. For each conformation the dipeptide was minimized subject to constraints on the phi and psi angles using the CHARMM27 steepest descent minimizer by 600 steps, using a dielectric of 1, and a nonbond cutoff of 80.0 Å. The internal CHARMM energy terms were taken from the final minimized structures. Solvation energies were calculated using the FDPB method as described above from the minimized structures using either the CHARMM27 or QEq charges. The geometric center of each peptide bond was used as the origin for the calculation of that peptide's dipole moment as follows

$$p = \Sigma q_i \cdot r_i \qquad (10)$$

where the sum runs over the C, O, N, and H atoms.

Results and Discussion

Charge Parametrization. Figure 1 shows the results of QEQUIL on the training set of 23 compounds using the original parameters of Rappe and Goddard (Table 1). The correlation is very poor at $r^2 = 0.01$, with a mean unsigned error of 3.5 kcal/mol. Examination of the direction of error for individual compounds reveals systematic errors: For example N and S containing groups have systematically over- and underestimated free energies, respectively (Figure 2). Systematic parametrization of the electronegativities improved the overall correlation to > 0.85. Inclusion of an atomic pair shielding factor improved results still further, with a final $r^2 = 0.96$, a slope very close to unity (0.99), and a mean unsigned error of 0.5 kcal/mol, giving a good fit between experiment and calculated values (Figures 1 and 2). Optimized parameters in Table 1 show that most of the adjustment of nonhalogen elements occurred in the N and S elemental negativity values, with smaller adjustments in those for H, C, and O. To obtain the best fit seven pair shielding factors were required to be significantly different from the null value of 1 (Table 2).

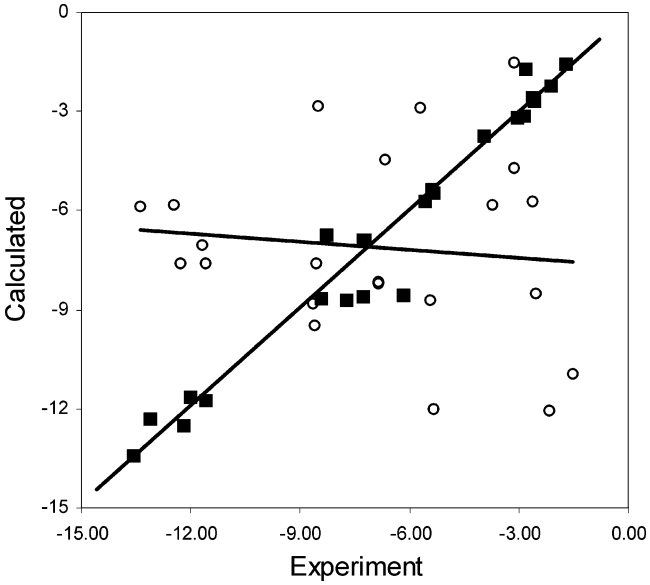

Figure 1. Comparison of experimental and calculated solvation free energies for the training set in Table 3. Energies are in kcal/mol. (o): Original parameter set. (■): Optimized parameter set. Lines are the least-squares fits.

The major error in the amino acid analogues was from the Lys, Trp, and Tyr side-chain analogues (Figure 2). The delocalized electron systems of Trp and Tyr are understandably less well represented in any classical method, such as the one used here.[28] An option would be to add further bond shielding factors, although this was not done here since we wanted to keep the number of adjustable parameters to a minimum. Aside from these deviations, the modified parameters show excellent agreement for side-chain analogues with carboxyl, thiol, amide, and OH moieties. Table 3 gives solvation free energy values for the original and optimized parameters, and the additional breakdown into different terms. In total, nine original element parameters were optimized, along with seven shielding parameters, for a total of 16 parameters on a 23 compound training set. This is a very modest number of parameters for charge sets; they often have dozens of fitted parameters.

The FDPB method for solvation is often used in conjunction with the MD force fields such as CHARMM. The MD is typically used either to generate an ensemble of protein structures for postprocessing with FDPB[30] or more recently with FDPB treatment of solvent forces integrated into the MD simulation.[31−33] We wanted to see how accurate unmodified CHARMM charges were with the same FDPB protocol used for QEq charges. If CHARMM charges could be used as is for solvation energy calculations, it would make integrated MD/FDPB calculations easier and require fewer parameters. Since CHARMM charges were available only for amino acids, results for these charges refer only to the 15 side-chain and backbone analogue compounds of the training set in Table 3. The results were significantly better than unoptimized QEq charges, with a correlation coefficient of 0.85 (Table 6) and a reasonable slope of 0.87. There is, however, a significant mean error of 2.1 kcal/mol so that use of the same set of charges for MD and solvation via FDPB would be less accurate than using two sets.

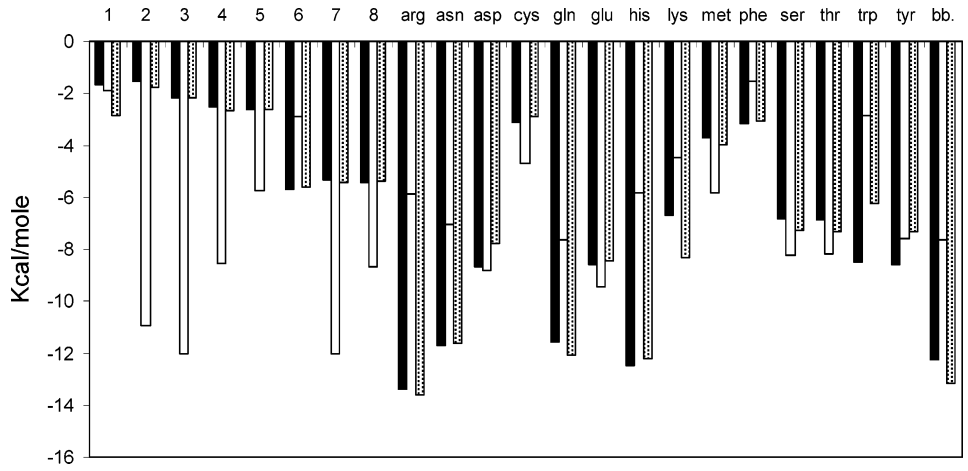

Figure 2. Comparison of experimental and calculated solvation free energies for the training set, including amino acid side-chain analogues (three letter amino acid code) and the backbone analogue NMA (bb). Other compound keys are given in Table 3. Filled bars: experiment. Unfilled bars: original parameters. Shaded bars: optimized parameters.

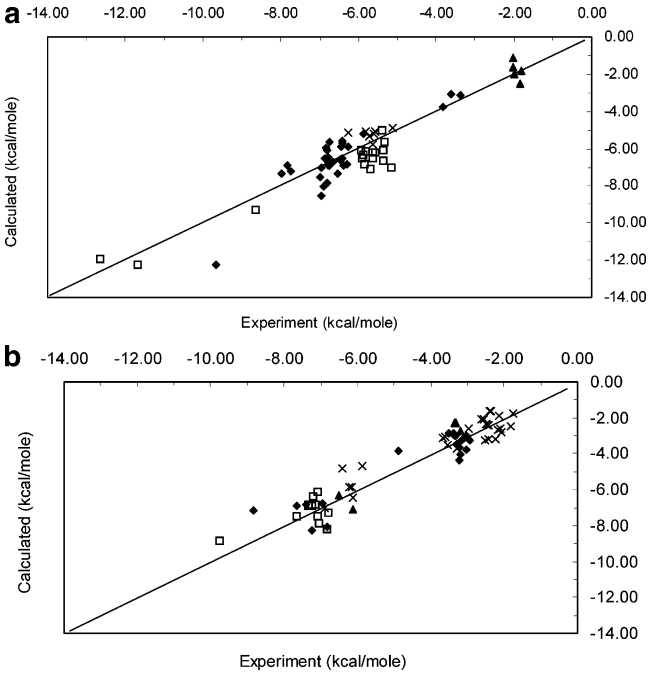

Figure 3. Comparison of experimental and calculated solvation free energies for the test set. X =Y line is shown on each figure. (a) Compounds containing −OH, NH_n, and SH (◆), C═O, COOH, and CHO (□), CN and NO_2 (×), alkynes (▲). (b) Compounds containing aromatic groups (◆), pyridines (□), halogenated alkanes and alcohols (×), halogenated aromatics (▲).

For a test set, we used a different set of 127 compounds with a variety of functional groups that includes those used in the original PARSE parametrization, and for which there are reliable solvation data.[27] Figure 3 shows a comparison of experimental and calculated electrostatic solvation free energies, with compounds grouped into broad categories based on the main type of functional group. Table 4 gives an additional breakdown of the energy terms for each compound. The overall agreement was good, with $R^2 = 0.90$, a calculated vs experimental slope of 0.98, and a mean unsigned error of 0.61 kcal/mol. Overall the method does well considering the small number of parameters and the fact that the test set is five times the size of the training set. Compound classes that contributed more to the mean error include the ketones and pyridines. The error from the latter is understandable as this is a large class of compounds, and the functional group did not occur in the training set. Compounds with alcohol, amide, thiol, sulfide, aldehyde, and benzyl groups were all accurately represented, as were the majority of halogenated compounds.

Comparison with Other Charge Sets. A number of parametrized atomic charged sets have been described in the literature using different approaches and designed for different methods/applications. We compared to two recent and widely used charge sets to see if these could be used with the FDPB method as is: the SM5.42R generated by AMSOL[4,6] and the AM1-bcc set.[5,7] These are both QM derived sets. AM1-bcc was parametrized on 2755 compounds so as to reproduce high level QM charges with a more computationally tractable semiempirical QM method, but they have also been reported to give good solvation energies with the explicit water/free energy perturbation method.[7] The AMSOL set was specifically designed to produce good solvation energies with the generalized Born (GB) solvent model for a wide range of solvents, including water. Our calculations were done using exactly the same FDPB protocol as for the QEq charges, since we wanted to compare charges using the same method. AMBER radii[22] were used with the AM1-bcc set in order to compare these charges on a more equal footing. Both previous charge sets gave significantly larger errors and, in the case of AMSOL charges, a poorer slope (Table 6). Figure 4 compares these charge sets graphically for compounds with a representative range of chemical types, to show where systematic differences, in terms of functional group type, occurred. We emphasize that the AM1-bcc and AMSOL charge sets were not parametrized specifically for the FDPB solvation application, and obviously they perform better with the methods they were designed for. However, the results reiterate the point that

Table 4. Solvation Free Energies (kcal/mol) for Test Set

code	molecule	main functional group(s)[a]	SAS (Å^2)	ΔG^{np}(calc)	ΔG^{solv}(expt)	ΔG^{elec} (expt)	ΔG^{elec} (calc)
1	propanol	OH	250	2.11	−4.83	−6.94	−7.04
2	butanol	OH	282	2.27	−4.72	−6.99	−7.51
3	isopropyl alcohol	OH	246	2.09	−4.76	−6.85	−6.54
4	2-butanol	OH	278	2.25	−4.58	−6.83	−5.99
5	3-methyl-1-butanol	OH	306	2.39	−4.42	−6.81	−7.84
6	ammonia	NH_3	140	1.56	−4.31	−5.87	−5.23
7	methylamine	NH_2	190	1.81	−4.57	−6.38	−6.91
8	ethylamine	NH_2	222	1.97	−4.50	−6.47	−6.62
9	propylamine-N-butylamine	NH_2	254	2.13	−4.39	−6.52	−7.32
10	dimethylamine	NH_2	230	2.01	−4.29	−6.30	−6.84
11	diethylamine	NH_2	298	2.35	−4.07	−6.42	−6.52
12	ethylthiol	SH	240	2.06	−1.30	−3.36	−3.12
13	dimethyl sulfide	CH−S−CH	242	2.07	−1.54	−3.61	−3.06
14	diethyl sulfide	CH−S−CH	308	2.40	−1.43	−3.83	−3.79
15	acetone	>C═O	244	2.08	−3.85	−5.93	−6.07
16	2-butanone	>C═O	276	2.24	−3.64	−5.88	−6.36
17	2-pentanone	>C═O	304	2.38	−3.53	−5.91	−6.50
18	3-pentanone	>C═O	306	2.39	−3.41	−5.80	−6.16
19	3-methyl-2-butanone	>C═O	306	2.39	−3.24	−5.63	−6.20
20	2,4-dimethyl-3-pentanone	>C═O	346	2.59	−2.74	−5.33	−5.66
21	acetaldehyde	HCO	206	1.89	−3.50	−5.39	−4.40
22	propionaldehyde	HCO	240	2.06	−3.44	−5.50	−4.72
23	butyric acid	COOH	284	2.28	−6.36	−8.64	−9.31
24	benzene	arom	258	2.15	−0.87	−3.02	−3.01
25	ethylbenzene	arom	326	2.49	−0.80	−3.29	−3.48
26	pyridine	CH═N−CH═	250	2.11	−4.70	−6.81	−8.10
27	4-methylpyridine	CH═N−CH═	290	2.31	−4.93	−7.24	−8.30
28	2-methylpyridine	CH═N−CH═	290	2.31	−4.63	−6.94	−6.75
29	phenol	OH (arom)	272	2.22	−6.62	−8.84	−7.15
30	pentanol	OH	314	2.43	−4.47	−6.90	−8.06
31	4-methyl-2-pentanol	OH	332	2.52	−3.73	−6.25	−5.89
32	2-pentanol	OH	310	2.41	−4.39	−6.80	−6.65
33	2-methyl-2-butanol	OH	302	2.37	−4.43	−6.80	−6.11
34	2,3-dimethylisobutyl alcohol	OH	324	2.48	−3.92	−6.40	−5.68
35	3-pentanol	OH	310	2.41	−4.35	−6.76	−6.49
36	hexanol	OH	346	2.59	−4.36	−6.95	−8.56
37	3-hexanol	OH	342	2.57	−4.08	−6.65	−6.71
38	2-methyl-3-pentanol	OH	334	2.53	−3.89	−6.42	−5.61
39	4-heptanol	OH	374	2.73	−4.01	−6.74	−6.92
40	2-methylpropanol	OH	276	2.24	−4.52	−6.76	−6.89
41	2-methyl-2-propanol	OH	272	2.22	−4.51	−6.73	−5.64
42	2-methyl-2-pentanol	OH	332	2.52	−3.93	−6.45	−5.93
43	4-methyl-2-pentanone	>C═O	330	2.51	−3.06	−5.57	−6.23
44	4-heptanone	>C═O	366	2.69	−2.93	−5.62	−6.54
45	5-nonnanone	>C═O	430	3.01	−2.67	−5.68	−7.07
46	2-hexanone	>C═O	340	2.56	−3.29	−5.85	−6.84
47	butanal	HCO	272	2.22	−3.18	−5.40	−5.00
48	hexanal	HCO	336	2.54	−2.81	−5.35	−6.08
49	heptanal	HCO	368	2.70	−2.67	−5.37	−6.65
50	octanal	HCO	400	2.86	−2.29	−5.15	−7.03
51	*N*-methylformamide	amide	230	2.01	−10.00	−12.01	−10.79
52	cyclopentanol	OH (cyclic)	278	2.25	−5.49	−7.74	−7.23
53	cyclohexanol	OH (cyclic)	302	2.37	−5.47	−7.84	−6.89
54	cycloheptanol	OH (cyclic)	324	2.48	−5.49	−7.97	−7.37
55	pyrrolidine	5ring NH	260	2.16	−5.48	−7.64	−6.88
56	pipyridine	5ring NH	284	2.28	−5.11	−7.39	−6.85
57	propylbenzene	alkyl arom	358	2.65	−0.53	−3.18	−4.03
58	*o*-xylene	alkyl arom	324	2.48	−0.90	−3.38	−2.85
59	*m*-xylene	alkyl arom	332	2.52	−0.83	−3.35	−2.97

Table 4. (Continued)

code	molecule	main functional group(s)[a]	SAS (Å^2)	ΔG^{np}(calc)	ΔG^{solv}(expt)	ΔG^{elec} (expt)	ΔG^{elec} (calc)
60	*p*-xylene	alkyl arom	332	2.52	−0.81	−3.33	−2.97
61	naphthalene	alkyl arom	328	2.50	−2.39	−4.89	−3.83
62	1,2,4-trimethylbenzene	alkyl arom	360	2.66	−0.86	−3.52	−2.83
63	methylethylbenzene	alkyl arom	352	2.62	−0.30	−2.92	−3.25
64	butylbenzene	alkyl arom	390	2.81	−0.40	−3.21	−4.37
65	methylpropylbenzene	alkyl arom	380	2.76	−0.45	−3.21	−3.68
66	dimethylethylbenzene	alkyl arom	370	2.71	−0.44	−3.15	−3.27
67	dimethylpropylbenzene	alkyl arom	396	2.84	−0.18	−3.02	−3.80
68	3-methylpyridine	−CH═N−CH═	288	2.30	−4.77	−7.07	−7.49
69	2-ethylpyridine	−CH═N−CH═	320	2.46	−4.33	−6.79	−7.27
70	3-ethylpyridine	−CH═N−CH═	320	2.46	−4.60	−7.06	−7.85
71	4-ethylpyridine	−CH═N−CH═	318	2.45	−4.37	−6.82	−8.23
72	2,4-dimethylpyridine	−CH═N−CH═	326	2.49	−4.86	−7.35	−6.93
73	3,4-dimethylpyridine	−CH═N−CH═	316	2.44	−5.22	−7.66	−7.47
74	3,5-dimethylpyridine	−CH═N−CH═	324	2.48	−4.84	−7.32	−6.82
75	2,3-dimethylpyridine	−CH═N−CH═	308	2.40	−4.83	−7.23	−6.92
76	2,5-dimethylpyridine	−CH═N−CH═	326	2.49	−4.72	−7.21	−6.38
77	2,6-dimethylpyridine	−CH═N−CH═	326	2.49	−4.60	−7.09	−6.12
78	dimethylethylpyridine	−CH═N−CH═	364	2.68	−4.46	−7.14	−6.86
79	1,2-ethanediol	OH−OH	226	1.99	−7.66	−9.65	−12.27
80	1-propyne	-yne	173	1.72	−0.31	−2.03	−1.15
81	1-butyne	-yne	200	1.86	−0.16	−2.02	−1.61
82	1-pentyne	-yne	229	2.01	0.01	−1.99	−2.01
83	1-hexyne	-yne	253	2.12	0.29	−1.84	−2.49
84	1-buten-3-yne	-yne	196	1.84	0.04	−1.80	−1.85
85	fluoromethane	F, -alkane	133	1.53	−0.22	−1.75	−1.77
86	chloroethane	Cl, alkane	183	1.78	−0.63	−2.41	−1.60
87	2-chloropropane	Cl, alkane	206	1.89	−0.25	−2.14	−1.87
88	1-chlorobutane	Cl, alkane	234	2.03	−0.14	−2.17	−2.66
89	1-chlororpentane	Cl, alkane	260	2.16	−0.07	−2.23	−3.20
90	2-chloropentane	Cl, alkane	256	2.14	0.07	−2.07	−2.82
91	3-chloropentane	Cl, alkane	254	2.13	0.04	−2.09	−2.63
92	chloroethene	Cl, alkene	180	1.76	−0.59	−2.35	−1.60
93	3-chloror-1-propene	Cl, alkene	208	1.90	−0.57	−2.47	−2.35
94	chlorobenzene	Cl, arom	243	2.07	−1.12	−3.20	−2.74
95	bromoethane	Br, alkane	197	1.85	−0.70	−2.54	−2.06
96	2-bromopropane	Br, alkane	219	1.96	−0.48	−2.43	−2.38
97	1-bromobutane	Br, alkane	249	2.11	−0.41	−2.51	−3.26
98	1-bromo-2methylpropane	Br, alkane	243	2.08	−0.03	−2.10	−2.63
99	bromobenzene	Br, arom	257	2.15	−1.46	−3.61	−3.04
100	1-bromo-4-methylbenzene	Br, arom	282	2.27	−1.39	−3.66	−3.12
101	1-bromo-2-ethylbenzene	Br, arom	296	2.34	−1.19	−3.53	−3.53
102	1-bromo-2-(1-methylethyl)benzene	Br, arom	318	2.45	−0.85	−3.30	−3.70
103	iodoethane	I, alkane	207	1.89	−0.72	−2.62	−2.06
104	2-iodopropane	I, alkane	227	1.99	−0.46	−2.46	−2.39
105	1-iodobutane	I, alkane	259	2.15	−0.26	−2.41	−3.22
106	acetonitrile	C≡N	173	1.72	−3.89	−5.61	−5.20
107	butanenitrile	C≡N	225	1.99	−3.65	−5.63	−5.79
108	nitroethane	NO_2	199	1.86	−3.71	−5.57	−5.11
109	2-nitropropane	NO_2	223	1.98	−3.14	−5.12	−4.90
110	nitrobenzene	NO_2, arom	255	2.14	−4.12	−6.26	−5.17
111	1-methyl-2-nitrobenzene	NO_2, arom	273	2.23	−3.59	−5.82	−5.10
112	1-methyl-3-nitrobenzene	NO_2, arom	281	2.27	−3.45	−5.72	−5.31
113	1,1-difluroethane	halo-alkane	171	1.71	−0.11	−1.82	−2.46
114	1,4-dimethylpiperazine	cyclic amine	262	2.17	−7.57	−9.74	−8.84
115	1,1-dichlorobutane	halo-alkane	280	2.26	−0.70	−2.96	−2.59
116	1,3-dichlorobenzene	halo-arom	294	2.33	−0.98	−3.31	−2.30
117	1,4-dichlorobenzene	halo-arom	294	2.33	−1.01	−3.34	−2.28
118	3-hydroxybenzaldehye	arom, HCO	261	2.17	−9.51	−11.68	−12.23

Table 4. (Continued)

code	molecule	main functional group(s)[a]	SAS (Å^2)	ΔG^{np}(calc)	ΔG^{solv}(expt)	ΔG^{elec} (expt)	ΔG^{elec} (calc)
119	4-hydroxybenzaldehye	arom, HCO	262	2.17	−10.47	−12.64	−11.92
120	2-chloropyridine	halo, CH═N−CH	250	2.11	−4.39	−6.51	−6.34
121	3-chloropyridine	halo, CH═N−CH	251	2.11	−4.01	−6.13	−7.13
122	1,1′-thiobis(2-chloroethane)	halo-alkane	325	2.49	−3.92	−6.41	−4.83
123	2,2,2-trifluroethanol	halo, OH	198	1.85	−4.30	−6.15	−5.88
124	1,1,1-trifluropropan-2-ol	halo, OH	220	1.96	−4.16	−6.12	−6.41
125	2,2,3,3-tetrafluropropan-1-ol	halo, OH	230	2.01	−4.88	−6.89	−7.04
126	2,2,3,3,3-pentafluropropan-1-ol	halo, OH	241	2.06	−4.15	−6.22	−5.84
127	1,11,3,3,3-hexafluropropan-2-ol	halo, OH	247	2.10	−3.77	−5.86	−4.71

[a] Arom: aromatic, halo: halogenated.

for the current state of empirical charge sets, it is essential to have a set designed specifically for each application/method.

Although we compared with AMSOL and AM1-bcc charge sets to answer the question of whether it was necessary to parametrize a new charge set for the FDPB method, we also wanted to compare the different charge sets using the solvation methods for which they were designed. When the AMSOL charges are used with the AMSOL radii and method (GB), the results are much better, with a mean unsigned error of only 0.65 kcal/mol and a $R^2 = 0.88$, comparable to the optimized QEq charges on this test set. In making this comparison, it should be noted that the AMSOL charges were parametrized on more than 200 compounds, and so its training set undoubtedly includes some of the common organic compounds in our test set. The true free R^2 for AMSOL for out test set would be somewhat lower, and the unbiased mean unsigned error would be higher, than the values in Table 6.

For the AM1-bcc set, simulations with explicit solvent and the free energy perturbation (FEP) method were used to calculate relative solvation energies of 40 compounds.[7] The mean unsigned error was 0.69 kcal/mol, giving somewhat worse accuracy. In this comparison it should be noted that the set of compounds used was different from those used here, so relative accuracies may vary depending on the mix of compounds used. However, due to the order of magnitude more computation required to do explicit water FEP simulations it is not practical to do the 127 test compounds used here. Also the method gives relative free energies of solvation, i.e., differences between two closely related compounds such as methane and methanol. AMSOL/GB and Qequil/FDPB methods both give absolute solvation free energies and are implicit solvent models with modest computational requirements. These two are easier to compare directly on the same test set.

The test set was chosen to be large relative to the training set (a ratio of about 5:1) in order to obtain a good idea of the robustness of the QEq charge calculation method. However, the variety of compounds one has in a large test set depends on the availability of reliable experimental data. Therefore for experimental reasons certain functional groups are over-represented, such as alcohols, and some under-represented. As the test set is made larger, this imbalance tends to increase. Since any charge set/solvation energy method is likely to perform better on some types of compounds than others, this affects the assessment of accuracy and the comparison between different sets. To counter this we pruned the test set so that each major functional group or group combination occurring in the original set of 127 was represented just twice if possible, but at least once, resulting in 52 test compounds. Table 5 lists this nonredundant test set and the calculated and experimental electrostatic solvation free energies. The nonredundant test set was used to compare the different charge sets, QEq, AMSOL, and AM1-bcc, in a more compound-unbiased way. The results are summarized in Table 6. QEq charges used with FDPB and AMSOL charges used with the GB radii and GB method have about the same accuracy, with mean unsigned errors of 0.50 and 0.57 kcal/mol, respectively, and slopes very close to unity. Again, use of the AMSOL and AM1-bcc charge sets with a method different from the one used for parametrization gives significantly poorer results.

In all these comparisons of calculated and experimental solvation energies, we put more emphasis on the mean unsigned error and the best fit slope and less on the correlation coefficient R^2. First, the mean unsigned error gives what the user is most interested in, an estimate of the method's accuracy in application. Second, in our judgment a large deviation from unity of the best fit slope indicates a poor charge set even if the correlation coefficient is good. With a poor slope there are clearly systematic errors due to over- or under-representation of some factor. This is more than likely to give large errors of *unknown sign* since this factor is altered in an unknown way when the method is used outside the training/test set compounds. From this perspective, the QEq charge set looks robust, with a slope of almost exactly 1. The fact that the QEq method can achieve as good results as AMSOL with far fewer training compounds and parameters is encouraging in terms of the simplicity and robustness of the QEq method.

Alanine Dipeptide Conformational Energy Map Analysis. The Rappe-Goddard algorithm for determining atomic charges is both rapid and takes account of conformation. The conformation dependence comes from the pairwise atomic Slater-Coulomb term J_{AB} in eq 1, which depends inversely on the distance between the A and B atoms. Examination of eq 1 shows that if Q_B is of opposite sign to Q_A, then at constant χ_A the J_{AB} term will tend to increase the magnitude

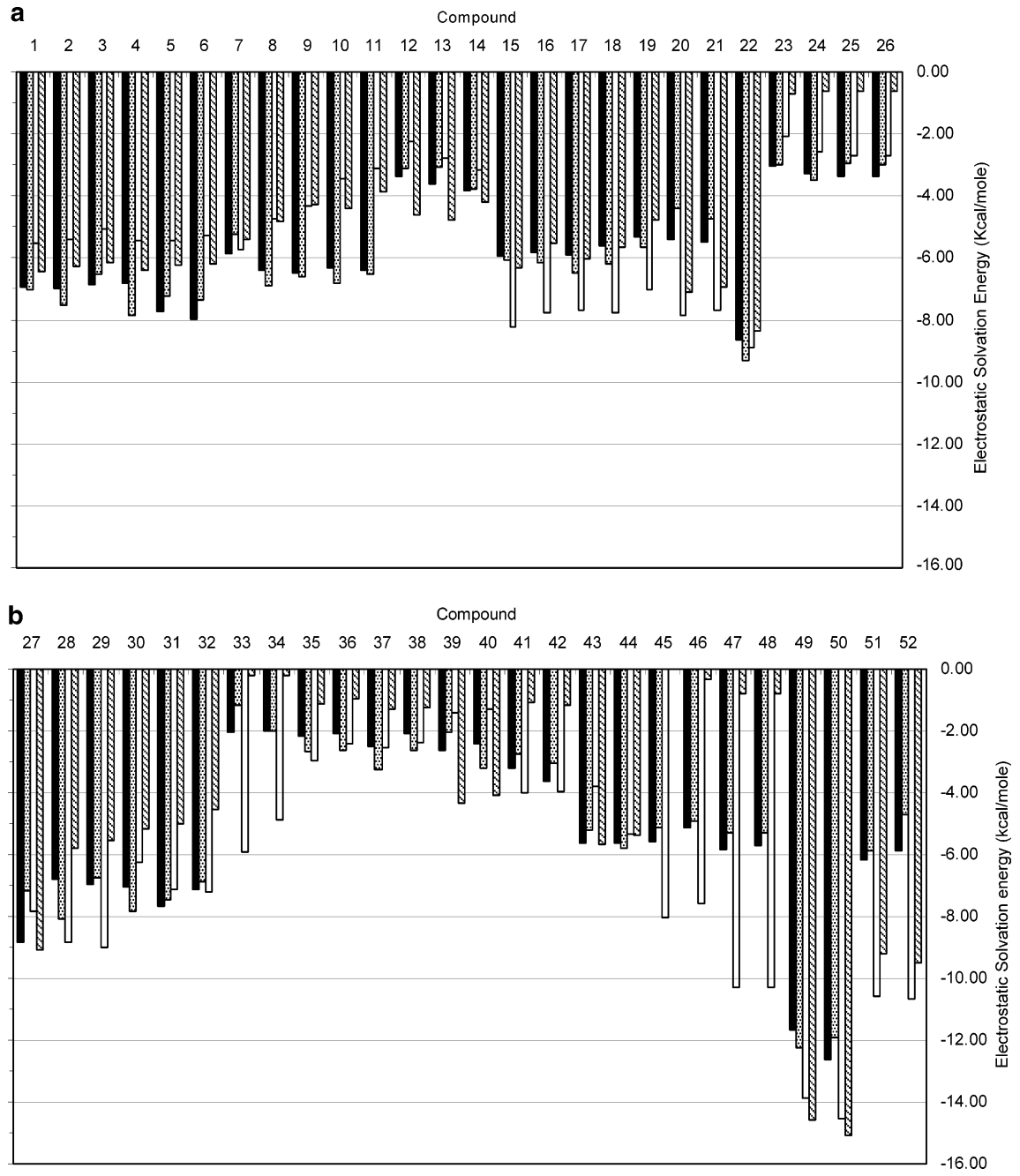

Figure 4. Comparison of experimental and calculated solvation free energies for nonredundant test set. Filled bars: experiment. Shaded bars: QEQUIL charges. Unfilled bars: AMSOL charges. Striped bars: AM1-bcc charges. The FDPB method was used for all three charge sets: (a) compounds 1−26 and (b) compounds 27−52. See Table 5 for compound key.

of Q_A, while if Q_B and Q_A are of the same sign, then the J_{AB} interaction will decrease the magnitude of Q_A. Q_A will have the same effect on Q_B. In other words, favorable intramolecular Coulombic interactions will tend to produce more polarization or larger magnitude partial charges. As pointed out in the original QEq paper, this conformation dependence could be used to implement a polarizable force field in MD simulations. The resulting conformation dependent polarization will also affect the solvation in a conformation dependent way, particularly when using say a FDPB implicit solvent treatment with the MD simulations. To examine the relationship between polarizable charges, conformation, and solvation energy with the QEq charge set, a conformational study of the alanine dipeptide was performed. Alanine dipeptide is the small compound most often used to evaluate most new molecular mechanics methods and for new solvation models designed for use in molecular mechanics of peptides and proteins.[34] The molecule has two planar peptide groups one at the N-terminus, the other at the C-terminus. It also has a side chain and two principle degrees of freedom, the phi and psi torsion angles which determine the relative orientations of the two peptide groups. Analysis of the effect of force field, charge, and solvent is typically displayed in terms of energy surfaces in the dipeptide's phi-psi space, i.e., as Ramachandran type plots. In this work the energies are plotted relative to the lowest point on each energy surface (which is thus set to 0 kcal/mol). This makes it easier to compare relative contributions of each factor at each phi/psi

Table 5. Nonredundant Functional Group Test Set

ID	name	functional group(s)	Gexp, ele	GEQ,ele
1	propanol	−OH linear alkane	−6.94	−7.04
2	butanol	−OH linear alkane	−6.99	−7.51
3	isopropyl alcohol	−OH branched alkane	−6.85	−6.54
4	3-methyl-1-butanol	−OH branched alkane	−6.81	−7.84
5	cyclopentanol	−OH cyclic alkane	−7.74	−7.23
6	cycloheptanol	−OH cyclic alkane	−7.97	−7.37
7	ammonia	NH_3	−5.87	−5.23
8	methylamine	$-H_2$ linear alkane	−6.38	−6.91
9	ethylamine	$-H_2$ linear alkane	−6.47	−6.62
10	dimethylamine	$-NH_2$ branched alkane	−6.30	−6.84
11	diethylamine	$-NH_2$ branched alkane	−6.42	−6.52
12	ethylthiol	−SH	−3.36	−3.12
13	dimethyl sulfide	−S−	−3.61	−3.06
14	diethyl sulfide	−S−	−3.83	−3.79
15	acetone	>C═O in middle	−5.93	−6.07
16	3-pentanone	>C═O in middle	−5.80	−6.16
17	2-pentanone	>C═O not in middle	−5.91	−6.50
18	3-methyl-2-butanone	>C═O branched alkane	−5.63	−6.20
19	2,4-dimethyl-3-pentanone	>C═O branched alkane	−5.33	−5.66
20	acetaldehyde	−HCO	−5.39	−4.40
21	propionaldehyde	−HCO	−5.50	−4.72
22	butyric acid	−COOH	−8.64	−9.31
23	benzene	aromatic ring	−3.02	−3.01
24	ethylbenzene	alkylated aromatic	−3.29	−3.48
25	*m*-xylene	bialkylated aromatic	−3.35	−2.97
26	*p*-xylene	bialkylated aromatic	−3.35	−2.97
27	phenol	aromatic, OH	−8.84	−7.15
28	pyridine	−CH═N−CH═	−6.81	−8.10
29	2-methylpyridine	CH═N−CH═ linear alkylation	−6.94	−6.75
30	3-ethylpyridine	CH═N−CH═ linear alkylation	−7.06	−7.85
31	3,4-dimethylpyridine	−CH═N−CH═ multialkylation	−7.66	−7.47
32	dimethylethylpyridine	−CH═N−CH═ multialkylation	−7.14	−6.86
33	1-propyne	−CCH with linear chain	−2.03	−1.15
34	1-pentyne	−CCH with linear chain	−1.99	−2.01
35	1-chlorobutane	halogenated alkane, Cl	−2.17	−2.66
36	3-chloropentane	halogenated alkane, Cl	−2.09	−2.63
37	1-bromobutane	halogenated alkane, Br	−2.51	−3.26
38	1-bromo-2-methylpropane	halogenated alkane, Br	−2.10	−2.63
39	idoethane	halogenated alkane, I	−2.62	−2.06
40	1-iodobutane	halogenated alkane, I	−2.41	−3.22
41	chlorobenzene	halogenated, aromatic	−3.20	−2.74
42	bromobenzene	halogenated, aromatic	−3.61	−3.04
43	acetonitrile	−C≡N	−5.61	−5.20
44	butanenitrile	−C≡N	−5.63	−5.79
45	nitroethane	$-NO_2$	−5.57	−5.11
46	2-nitropropane	$-NO_2$	−5.12	−4.90
47	1-methyl-2-nitrobenzene	$-NO_2$, aromatic	−5.82	−5.31
48	1-methyl-3-nitrobenzene	$-NO_2$, aromatic	−5.72	−5.10
49	3-hydroxybenzaldehye	aromatic −CHO	−11.68	−12.23
50	4-hydroxybenzaldehye	aromatic −CHO	−12.64	−11.92
51	2,2,2-trifluroethanol		−6.15	−5.88
52	1,11,3,3,3-hexafluropropan-2-ol	fluorinated alcohol	−5.86	−4.71

conformation. Figure 5 shows such a plot for the nonelectrostatic contribution to the dipeptide energy using the CHARMM bonded and van der Waals parameters. Approximate centers of commonly defined conformation regions are indicated in the figure. The resulting energy surface recapitulates the original Ramachandran analysis based on steric considerations, with low-energy regions for negative values of phi that span the beta and alpha conformations. Figure 6a shows the electrostatic contribution from fixed CHARMM22 charges. Adding this contribution to the nonelectrostatic contribution yields the in vacuo energy surface, Figure 6b. Figure 6c shows the solvation free energy

Table 6. Summary of Comparisons of Experimental vs Calculated Free Energies

molecule set (no. of compounds)[a]	charge set	method[b]	R^2	slope	mean unsigned error (kcal/mol)
training (23)	QEq (unoptimized)	FDPB	0.01	0.08	3.52
training(15)	CHARMM	FDPB	0.85	0.87	2.1
training (23)	QEq (optimized)	FDPB	0.96	0.99	0.50
test (127)	QEq (optimized)	FDPB	0.90	0.98	0.61
test (127)	AMSOL	FDPB	0.48	0.83	1.60
test (127)	AM1-bcc	FDPB	0.60	1.02	1.77
test (127)	AMSOL	AMSOL/GB	0.88	1.08	0.65
test (40)	AM1-bcc[c]	FEP	n/a	n/a	0.69
nonredundant(52)	Qeq (optimized)	FDPB	0.93	0.98	0.50
nonredundant(52)	AMSOL	FDPB	0.54	0.98	1.71
nonredundant(52)	AM1-bcc	FDPB	0.65	1.17	1.75
nonredundant(52)	AMSOL	AMSOL/GB	0.91	1.10	0.67

[a] Training, test, and nonredundant test molecules are listed in Tables 3−5, respectively. [b] FDPB: finite difference Poisson−Boltzmann, AMSOL/GB: generalized Born with AMSOL parametrization, FEP: free energy perturbation using explicit solvent. [c] Taken from the original AM1-bcc parametrization paper.[7] Mean unsigned error for relative (differences in) solvation free energies for 40 compound pairs listed therein.

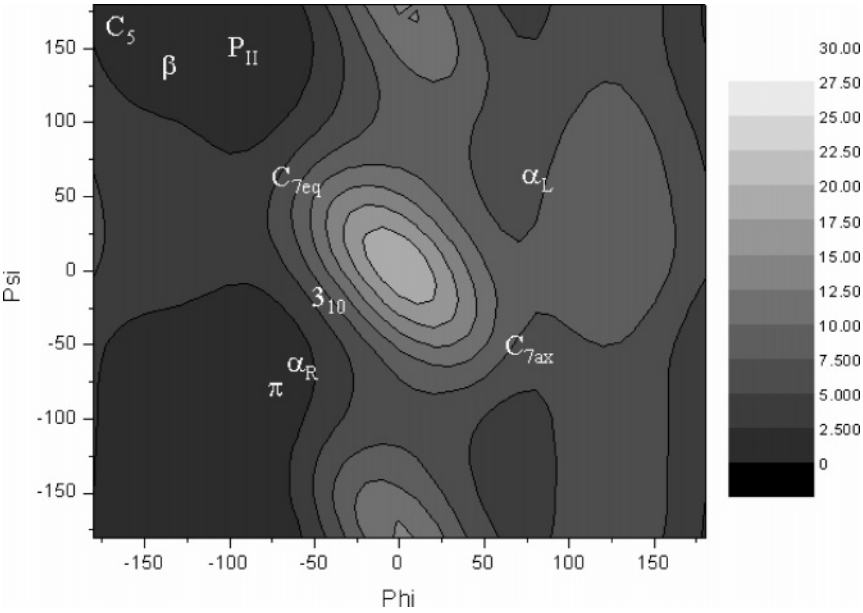

Figure 5. Ramachandran (ϕ, ψ) plot for the alanine dipeptide. Gray scale indicates the nonelectrostatic energy (sum of bond stretch, angle, torsion, and van der Waals energy terms) using the CHARMM force field. Energies are relative to the lowest point on the map (by definition at 0 kcal/mol). Approximate locations of the center of commonly referred to conformations are labeled on the map.

of the dipeptide with these fixed charges calculated using the FDPB solvation model, which, added to the in vacuo energy, yields the total energy in water. The total energy surface is shown in Figure 6d. The resulting low-energy regions are broadly the same as with just the nonelectrostatic terms, but the allowed regions are somewhat more tightly constrained around the canonical alpha and beta regions.

The corresponding energy surfaces for electrostatic and solvation energy terms using the polarizable QEq charges are shown in Figure 7a−d. The nonelectrostatic contribution is the same as with the fixed CHARMM charges (Figure 5). Summing the nonelectrostatic and electrostatic terms gives the in vacuo energy for the QEq charge model (Figure 7b), and adding in the solvation terms gives the total energy surface for the QEq charges (Figure 7d).

Considering first the internal electrostatic energy term, for the fixed CHARMM charges the energy surface shows a broad trough lying along the phi + psi = 0° line. Analyzing this in terms of the N-terminal, or phi-angle dependent peptide group, and the C-terminal, or psi-angle dependent peptide group (each composed of their O→C and N→H dipoles), the trough roughly corresponds to configurations with antiparallel peptide group alignments or, in electrostatic terms, antiparallel alignments of the dipoles associated with these two groups. This results in a favorable interaction between the peptide groups. This trough includes the internally H-bonded C7ax (62, −65) and C7eq (−65, 69) configurations. The solvation map is roughly the inverse of the internal electrostatic map, with a peak along the phi + psi = 0° line, but overall the solvating map is flatter than the internal electrostatic map. Interpreting the map again in terms of two separate dipoles arising from the two peptide groups, the trough along phi + psi = −180° corresponds to configurations with parallel alignments of the dipoles. The

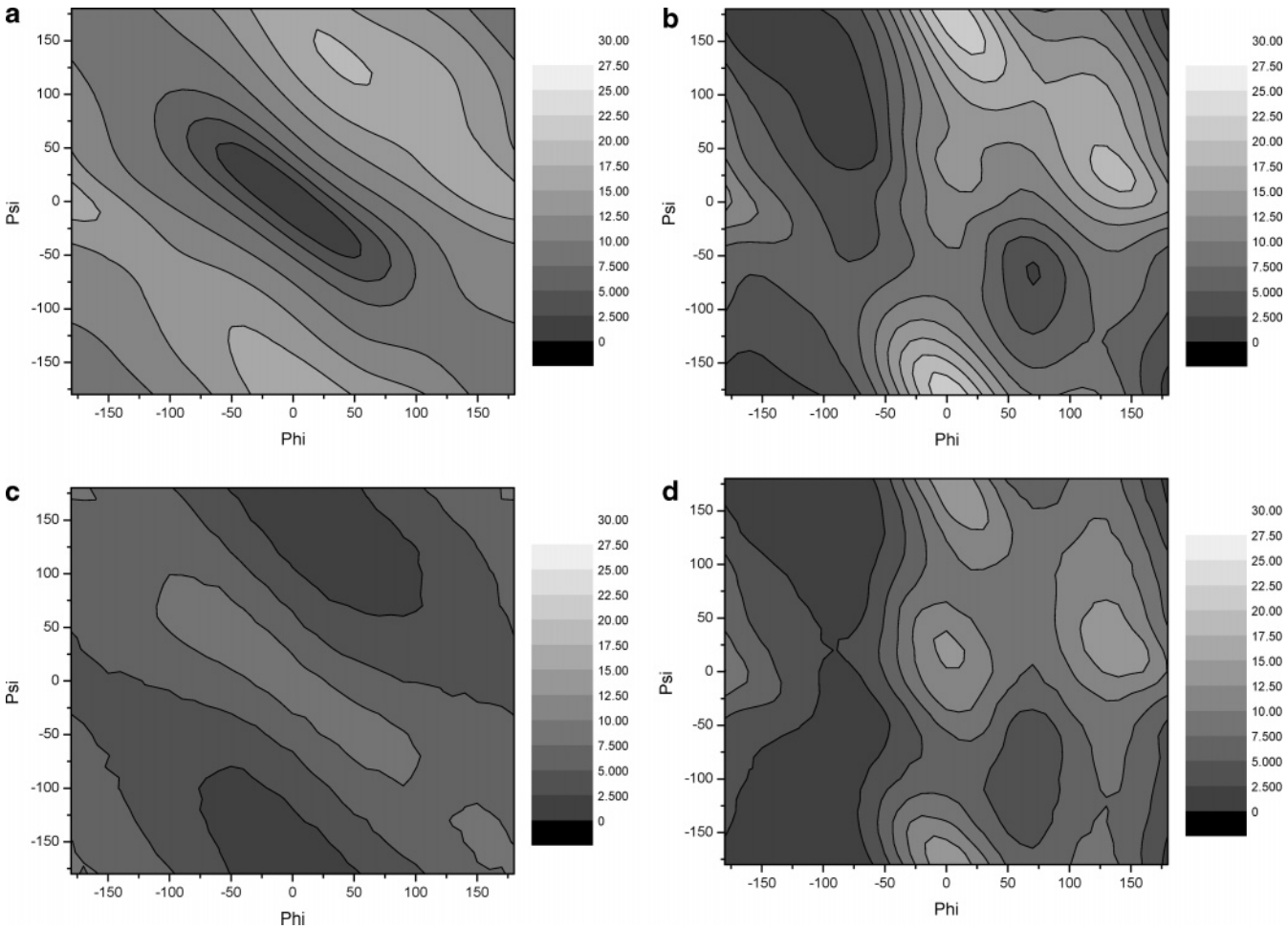

Figure 6. Ramachandran (ϕ, ψ) plot for the alanine dipeptide. Gray scale indicates (a) electrostatic part of internal energy using CHARMM charges, (b) total in vacuo energy (sum of electrostatic and nonelectrostatic energies), (c) solvation energy using the CHARMM charges and the FDPB method, and (d) total energy (sum of nonelectrostatic, electrostatic, and solvation terms). Energies are relative to the lowest point on the map.

two peptide dipoles tend to add with respect to the overall molecular dipole, producing the largest solvent reaction field and hence the most negative solvation free energy. Thus for fixed charges the solvation energy term cancels a lot of the internal electrostatics, giving a total energy surface flatter than, but similar to, the nonelectrostatic internal energy map.

The internal electrostatic energy term in the polarizable QEq model is similar to that of the fixed charged model, but the trough of favorable energy is shorter, less deep, and centered at approximately (−60, 20) rather than at (0, 0). This shrinking and shifting of the minimum region is understandable in terms of the conformation dependent charge polarization, principally the changing magnitude of the phi and psi peptide dipoles (Figure 8 (parts a and b, respectively)). The region where the sum of phi and psi dipole magnitudes is largest lies in the same region as the minimum in the internal electrostatic energy. The shorter trough in the electrostatic energy surface produces an in vacuo energy surface that, while very similar to the CHARMM in vacuo energy surface for phi $< -50°$, is higher (more unfavorable) in the other regions (phi$> -50°$)

The polarizable charges produce a qualitatively different solvation map from the fixed charges (Figure 7c). The map is much flatter. Note that the energy scale in this figure is a factor of 5 smaller. Again this can be interpreted in terms of the magnitude and alignment of the two principle dipolar groups. Conformations with a parallel dipolar alignment have less favorable Coulombic interaction between the two peptide groups and thus significantly less polarized charges. This results in smaller magnitude dipoles, principally not only for the psi dependent group (Figure 8b) but also for the phi dependent group (Figure 8a). This dipole magnitude effect runs in the opposite direction to the dipole orientation effect, the magnitudes being smaller when the dipoles are parallel, and larger when they are 'antiparallel', so the effects on the solvation reaction field tend to cancel, leading to a much flatter map.

Neither the fixed charge solvation energy map nor the polarizable charge solvation energy map can be adequately explained in terms of the total dipole moment of the dipeptide (results not shown), but only in terms of the individual peptide dipole moments. This indicates that representing the molecular charges as a single dipole is a bad approximation for describing solvation when the separation of charged groups is significant compared to the size of cavity formed in solution, as it is in the alanine dipeptide. One must account for the higher order poles, e.g. quadrupoles, etc., or at least describe the distribution as two separate dipole centers.

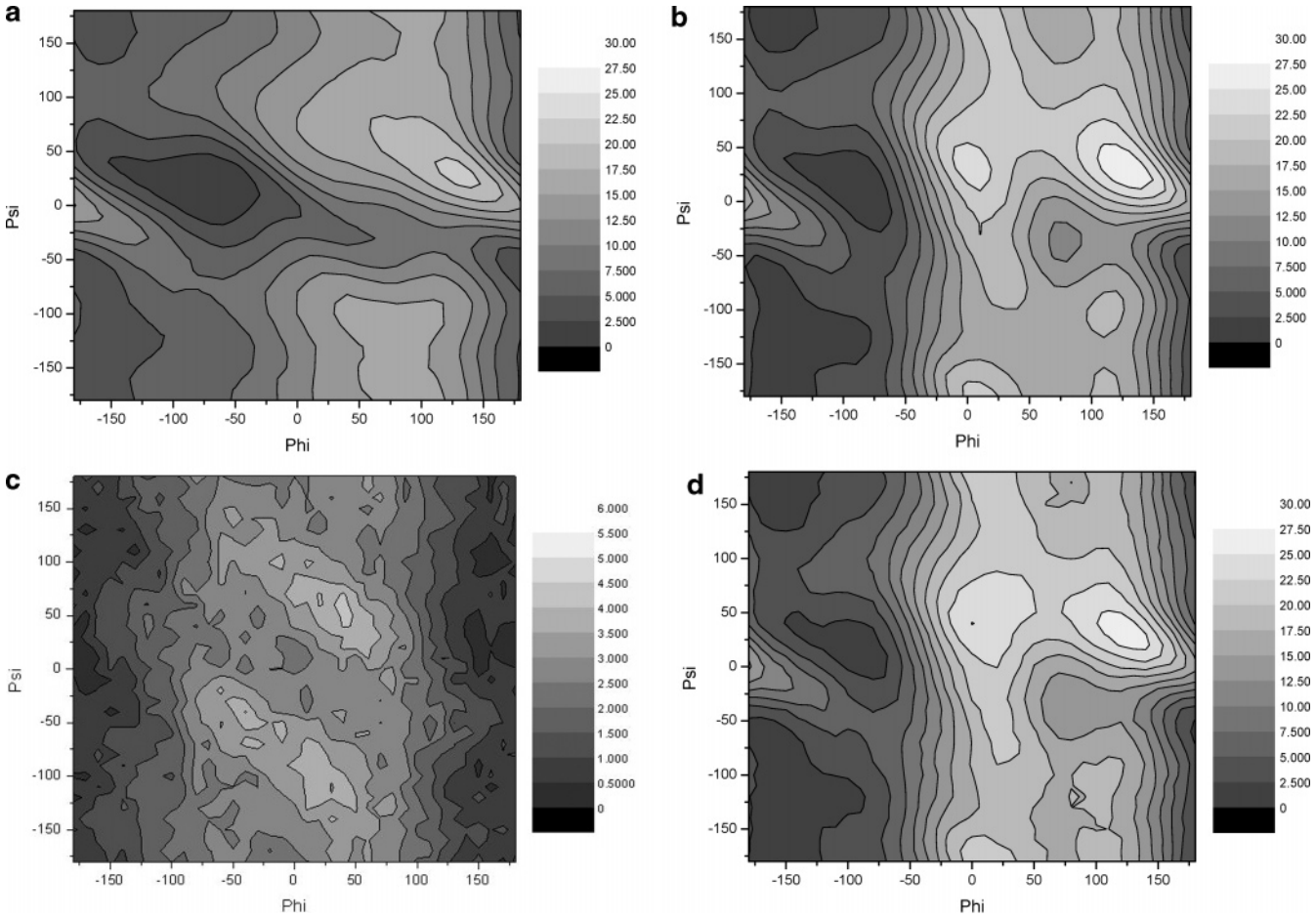

Figure 7. Ramachandran (ϕ, ψ) plot for the alanine dipeptide. Gray scale indicates (a) electrostatic part of internal energy using conformation-dependent QEq charges determined by QEQUIL, (b) total in vacuo energy (sum of electrostatic and nonelectrostatic energies), (c) solvation energy using the QEq charges and the FDPB method, and (d) total energy (sum of nonelectrostatic, electrostatic, and solvation terms). Energies are relative to the lowest point on the map

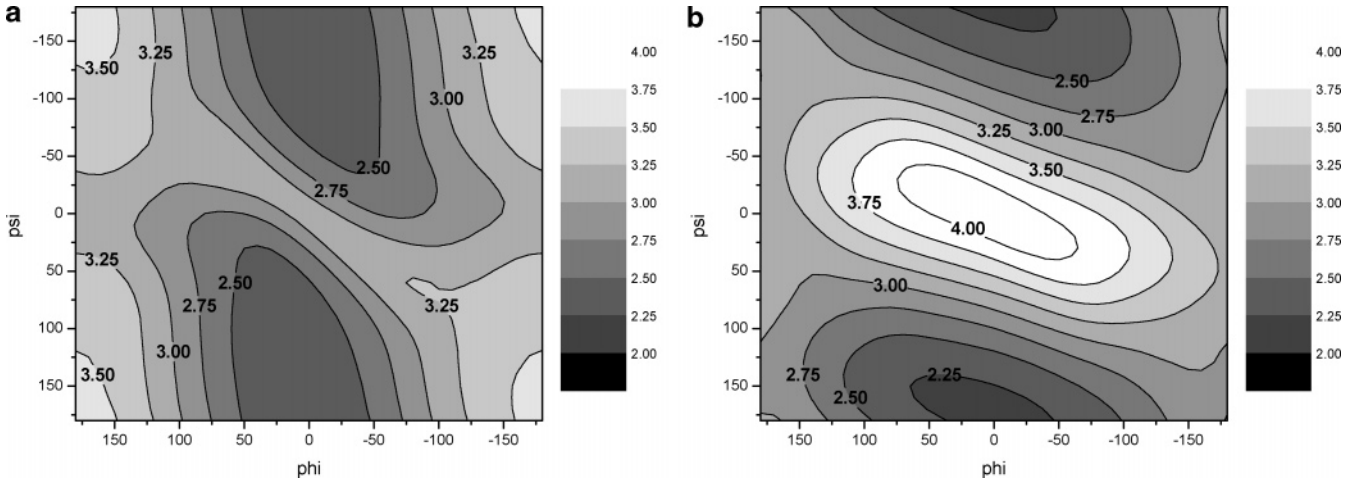

Figure 8. Ramachandran (ϕ, ψ) plot for the alanine dipeptide. Gray scale indicates the magnitude of the peptide moiety dipole moment in eÅ, using conformation-dependent QEq charges: (a) N-terminal (phi) peptide and (b) C-terminal (psi) peptide.

Comparing the total energy surface for the two charge models (Figures 6d and 7d) the QEq model surface has a larger range of energies and a smaller favorable region in the vicinity of the two main secondary structure types, beta sheet (β) and alpha helical (α_R), particularly around the latter. In fact the α_R region lies more on a low saddle in the energy surface than a minimum. The less favorable energy in the alpha region obtained with the polarizable model is larger attributable to the shorter trough in the corresponding internal electrostatic energy surface with QEq charges (Figure 7a vs Figure 6a). The larger difference between high and low regions with the QEq model principally results from higher energies in the region with phi $> -50°$. Again, this is attributable to the shorter trough in the corresponding internal

electrostatic energy surface compared to the fixed charge model. Since the phi > −50° tends to be less populated in folded proteins and in MD simulations (with the exception of Pro and Gly residues) the large difference in energy surfaces in this region between the two models is unlikely to result in as large a difference in conformational preferences. Conformational preference differences between the two models are more sensitive to the more populated beta and alpha regions, and here the differences between the two charge models are much smaller. Whether the QEq charge mode provides an improved description of the peptide backbone energy surface in proteins cannot be determined from examination of the energy surface in phi-psi space alone but must be examined using actual molecular dynamics simulations with say an integrated FDPB solvent treatment. The development of the rapid and solvation-optimized method for charge determination described here should make it easier to perform such simulations in the future. We note, however, that to do accurate molecular dynamics such a charge set would have to be implemented in the MD code in a fully consistent manner including polarization forces for both solvation and internal interactions, a significant undertaking given the complexity of current MD packages.

Conclusions

Partial charges sets are required for most atomic level simulations and calculations. Unfortunately, with the current state of technology no one charge set is adequate for all applications and simulation methods. In this study we describe an optimization and extension of Rappe and Goddard's charge equilibration through electronegativity neutralization method for calculating atomic charges (QEq). The optimized charges are designed to be used specifically with the FDPB method for calculating solvation, and they are in effect a successor to the PARSE charge set. The latter, being effectively parametrized by hand, cannot easily be extended to new functional groups. The method is designed to use a small set of element data in order to handle a wide range of chemical functional groups. The method was optimized using 23 compounds of which 15 were amino acid side-chain and backbone analogues in order to maximize accuracy when applied to peptides and proteins. The method uses 16 adjustable parameters. The mean unsigned error compared to experiment was 0.50 kcal/mol. Testing the method on a larger set of compounds (127) gave a somewhat larger unsigned error of 0.61 kcal/mol and, just as importantly, a best fit slope vs experimental data of close to unity. The fact that we use a large ratio of test to training compounds and that we get a good slope indicates that the method is quite accurate and robust. As with any parametrized charge set, some functional groups are treated better than others. Some functional groups systematically contribute more to the mean error than others, for example ketones, whose magnitudes are systematically overestimated. Pyridines as a class also contribute significantly to the mean error, but are both over- and underestimated, and so contribute little to the overall (functional group independent) systematic error. Systematic error due to particular functional groups can be exacerbated or hidden depending on the composition of the test set. Using a nonredundant functional group set of compounds, the mean unsigned error for the Qequil charges was reduced from 0.61 kcal/mol to 0.50 kcal/mol. This significantly smaller value for the mean unsigned error indicates that larger test sets, with over-representation of the more common functional groups, may not be better tests of charge sets in general. Since the slopes for test sets using the Qequil charges are very close to one, there is little systematic error averaged over the range of functional groups examined here, i.e., there is no systematic under- or overestimation of solvation energy due to some factor common to different functional groups.

Since the QEQUIL program calculates charges in a rapid and conformation dependent manner, we performed a preliminary investigation of the method's potential for implementing a polarizable charge set in MD simulations. We examined the phi-psi energy surface for the alanine dipeptide using the QEQUIL program, calculated the energy surface for different electrostatic components, and compared them to a standard fixed charge set used for MD (CHARMM). The polarizable charges produce a much less conformation dependent solvation reaction field than the fixed charges due to compensating effects from charge polarization. Overall the phi-psi energy surface with polarizable charges has a larger energy difference between high and low regions and somewhat smaller allowed regions around the beta and alpha regions, especially the latter. Testing whether the polarizable charge model can provide an improved description of the peptide energy surface in molecular dynamics simulations of proteins is an obvious future direction for this work. Regardless of its potential for a polarizable force field, the QEq charge determination as implemented in the QEQUIL program does provide an accurate way of generating charges for a range of functional groups, specifically for use with the widely used finite difference Poisson−Boltzmann method of calculating of solvation energies.

Acknowledgment. This work was supported by Grant MCB02-35440 from the NSF. We thank Ninad Prabhu, Ryan Coleman, and John Skinner for helpful discussions.

References

(1) Cornell, W.; Cieplak, P.; Bayly, C.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.

(2) Sitkoff, D.; Sharp, K.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(3) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(4) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.

(5) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132−146.

(6) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257−3271.

(7) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(8) Gasteiger, J.; Marsili, M. *Tetrahedron Lett.* **1978**, *34*, 3181.

(9) Gasteiger, J.; Marsili, M. *Tetrahedron Lett.* **1980**, *36*, 3219−3221.

(10) Tai, K.; Grant, J. A.; Scheraga, H. A. *J. Phys. Chem.* **1990**, *94*, 4732−4739.

(11) Rappe, A.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358−3363.

(12) Bultinck, P.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J. P. *J. Phys. Chem.* **2002**, *106*, 7887−7894.

(13) Gilson, M. K.; Gilson, H. S. R.; Potter, M. J. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 1982−1997.

(14) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141−6151.

(15) Patel, S.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 1−15.

(16) Patel, S.; Alexander, D.; Mackerell, J.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 1504−1514.

(17) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

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

(19) Press: W.; Flannery, B.; Teukolsky, S.; Vetterling, W. *Numerical Recipes*; Cambridge University Press: Cambridge, 1986.

(20) Gilson, M.; Sharp, K. A.; Honig, B. *J. Comput. Chem.* **1988**, *9*, 327−335.

(21) Nicholls, A.; Sharp, K. A.; Honig, H. *Proteins* **1991**, *11*, 281−296.

(22) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(23) Sridharan, S.; Nicholls, A.; Sharp, K. A. *J. Comput. Chem.* **1995**, *16*, 1038−1044.

(24) Case, D.; Pearlman, D.; Caldwell, J. W.; Cheatham, T.; Wang, J.; Ross, C.; Simmerling, T.; Darden, T.; Merz, K.; Stanton, A.; Chenn, J.; Vincent, M.; Crowley, V.; Crowley, V.; Tsui, V.; Gohlke, H.; Radmer, R.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G.; Singh, U.; Weiner, P.; Kollman, P. A. *AMBER 7*; UCSF, 2002.

(25) Hawkins, G. D.; Giesen, D.; Lynch, G.; Chambers, C.; Rossi, I.; Storer, J.; Li, J.; Zhu, T.; Thompson, J.; Winget, P.; Lynch, B.; Rinaldi, D.; Liotard, D.; Cramer, C. J.; Truhlar, D. G. *Amsol v. 7.1*; Regents of the University of Minnesota, 2004.

(26) Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664−1670.

(27) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Soln. Chem.* **1981**, *10*, 563−595.

(28) Moran, A.; Mukamel, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 506−510.

(29) MacKerell, A. D.; Brooks, B.; Brooks, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley& Sons: Chichester, 1998; Vol. 1, pp 271−277.

(30) Schiffer, C.; Caldwell, J.; Kollman, P.; Stroud, R. *Mol. Simul.* **1993**, *10*, 121.

(31) Lu, Q.; Luo, R. *J. Chem. Phys.* **2003**, *119*, 11035−11047.

(32) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244−1253.

(33) Prabhu, N. V.; Zhu, P.-J.; Sharp, K. A. *J. Comput. Chem.* **2004**, *25*, 2049−2064.

(34) Prabhu, N.; Sharp, K. *Chem. Rev.* **2005**, *106*, 1616−1623.

CT060009C