Fully Quantum Mechanical Description of Proteins in Solution. Combining Linear Scaling Quantum Mechanical Methodologies with the Poisson−Boltzmann Equation

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY A · JUNE 1999

Impact Factor: 2.69 · DOI: 10.1021/jp990266w

2 AUTHORS, INCLUDING:

Kenneth M Merz
Michigan State University
273 PUBLICATIONS **23,701** CITATIONS

SEE PROFILE

Fully Quantum Mechanical Description of Proteins in Solution. Combining Linear Scaling Quantum Mechanical Methodologies with the Poisson−Boltzmann Equation

Valentin Gogonea and Kenneth M. Merz, Jr.*

Department of Chemistry, 152 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802

Received: January 22, 1999; In Final Form: April 19, 1999

In this paper we report a method for solving the Schrödinger equation for large molecules in solution which involved merging a linear scaling divide and conquer (D&C) semiempirical algorithm with the Poisson−Boltzmann (PB) equation. We then assess the performance of our self-consistent reaction field (SCRF) approach by comparing our D&C-PB calculations for a set of 29 neutral and 36 charged molecules with those obtained by ab initio GVB and DFT (B3LYP) methods, Cramer and Truhlar's semiempirical generalized-Born SM5 model, and with the experimental solvation free energies. Furthermore, we show that our SCRF method can be used to perform fully quantum mechanical calculations of proteins in solution in a reasonable amount of time on a modern workstation. We believe that *all* electrostatic interactions in biological systems require a quantum mechanical description in order to obtain an accurate representation. Thus, our new SCRF method should have an impact on the computational study of physical and chemical phenomena occurring in proteins and nucleic acids, which are, in general, strongly influenced by electrostatic interactions. Moreover, this may lead to novel insights into classic problems like protein folding or drug design.

Introduction

Recent progress in linear scaling algorithms for matrix diagonalization have made it possible for the *first* time to perform *fully* quantum mechanical (semiempirical) calculations on proteins in gas phase.[1−6] However, proteins and other biomolecules function in aqueous solution and, excepting the work of York et al.[7] and Nadig et al.,[8] none have included solvent in any way. In the work of York et al., a linearized version of the continuum based conductor-like screening (COSMO[9]) model was used to estimate solvation effects on biomolecules. Using an explicit solvent model, Nadig et al. studied the solvation of the major cold shock protein (CspA) in explicit water molecules, whose configuration were obtained from a classical MD simulation. From this investigation it was observed that charge transfer interactions play a major role in the solvation of this protein, and, in particular, the carboxylates (from Glu and Asp residues) on the protein surface. While these two efforts represent exciting first steps toward modeling the solvation of *entire* proteins using quantum mechanical methodologies, clearly, much more work needs to be done.

There are a broad range of approaches that have been adopted to include solvation effects within a quantum mechanical methodology. Incorporation of solvation effects via a supermolecule approach, as was done in the work of Nadig et al., is very expensive and realistically cannot be applied routinely to a wide range of problems related to biomolecular structure and function.[8] Continuum solvation models, on the other hand, are much less expensive and do not require extensive statistical sampling of the solvent degrees of freedom because these models are already approximations of the potential of the mean force, thus the statistical weight of different solvent configurations are included.[10] Models based on integral equation formalism (e.g. XSOL) were also proposed, [11] but continuum solvation models remain among the simplest and least expensive computationally, which makes them quite attractive. There are a number of continuum solvation models that can be used in conjunction with quantum mechanical based methodologies. Describing all of them is beyond the scope of this paper, but a review of Tomasi and Persico[12] covers most of the available methodologies. For our purposes we decided to utilize Poisson−Boltzmann (PB) technique because this approach has been widely applied to biological systems using classical charge models.[13,14] For example, PB equation has been used in studies of Cu and Zn superoxide dismutase (SOD),[15] the Klenow fragment of DNA polymerase I[16] and phosphoglycerate mutase,[17,18] and more recently in Brownian dynamic simulations of SOD as a prototype for studying electrostatic steering effects in biomolecular reactions.[19−21] Numerous other examples could be cited, but the interested reader should examine some relatively recent reviews[12,13,22,23] on this subject to get an appreciation of the scope of applications that are possible. Furthermore, the PB approach can readily include salt effects (through the linearized or nonlinear PB equation), which are important in cases where highly charged biomolecules are of interest (e.g., the polyanionic DNA molecule). Moreover, it can be extended to include the fluctuation of ionic concentrations (through the Kirkwood hierarchy of equations) which enables the calculation of higher valence ion distributions around biological molecules.[24] Finally, good quality solvation free energy results using a combination of the PB method and quantum mechanics (the charge distribution was obtained using electrostatic potential (ESP) fitting methodologies[25]) have been reported, [26] giving us confidence that this linkage would work effectively with our linear scaling D&C quantum mechanical methodology.

In order for a combined PB-D&C approach to be useful in investigating biomolecular systems, one needs an accurate charge distribution. ESP methods[25] work reasonably well for small solutes, where few atoms are buried.[27] However, applying

10.1021/jp990266w CCC: $18.00
Published on Web 06/15/1999

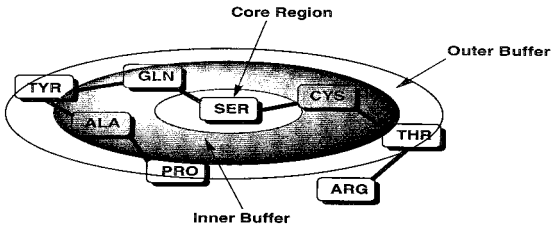

Figure 1. Splitting up a protein into subsystems in the divide and conquer method. The inner and outer buffer regions are introduced to diminish truncation effects (see text).

ESP fitting methods to proteins is not yet possible, which restricts us to obtaining charges directly from the wave function using Mulliken or Coulson charges. However, it is well-known that the calculated dipole moments obtained from Coulson or Mulliken charges are in poor agreement with experiment.[25] Fortunately, this problem has been remedied through the development of so-called class IV charge models,[28,29] which take Coulson and Mulliken charges and scale them to accurately reproduce the dipole moment of small molecules. Without this new charge model, the results outlined below would have been difficult to obtain if the deficiency in the charge model had not been absorbed into the nonpolar part of the solvation free energy.

This paper is divided as follows. First, we review theoretical and technical details relating to the development of the combined D&C-PB methodology. This section is followed by a discussion regarding our calculated solvation free energies for a series of charged and neutral molecules, *N*-acetyl-*N*′-methyl derivatives of the 20 naturally occurring amino acids and DNA bases and nucleotides. We then apply this approach to a few protein systems and one piece of DNA (Dickerson's dodecamer) and then conclude with a summary of our results.

Methodology and Technical Details

Calculation of the Solvation Free Energy Using a Polarizable Continuum−Polarizable Solute Model. *Solvent Polarization.* The solvation free energy (G_{sol}) is the difference between the solute free energy in solution and the gas phase. In the case of a quantum mechanical description of the solute, the solvation free energy can be decomposed into a sum involving the reaction field (electrostatic G_{RF}), the solute wave function distortion (G_{wfd}), and a hydrophobic (G_{np}) term:[30]

$$G_{sol} = G_{RF} + G_{wfd} + G_{np} \tag{1}$$

G_{RF} is calculated from the interaction of the solute charge density with the electrostatic potential generated by the reaction field:[12]

$$G_{RF} = {}^1/_2 \int_V \rho(\mathbf{r})\phi_{RF}(\mathbf{r})\, d\mathbf{r} \tag{2}$$

The use of the PB approach has the advantage of representing the dielectric discontinuity by assigning different dielectric constants to the solute and the continuum solvent.[17] Most of the information necessary to evaluate the solvation free energy is given by this discontinuity in the dielectric.[31] An alternative to the use of the electrostatic potential for the evaluation of the reaction field energy is to use an appropriate virtual surface charge density (σ_{RF}) placed at the dielectric interface which generates the reaction field potential.[32] Consequently, the surface charge density and not the reaction field itself is used to evaluate the reaction field free energy:[33]

$$G_{RF} = \frac{1}{2}\int_V d\mathbf{r}\, \rho(\mathbf{r}) \int_S d\mathbf{r}' \frac{\sigma_{RF}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{3}$$

G_{np} contains contributions from cavity formation and solvent−solute dispersion−repulsion interactions,[12] and for small molecules these two terms are often taken together (and called nonpolar or hydrophobic) and are considered to be proportional to the molecular or solvent-accessible surface area, an idea which was first suggested by Uhlig in 1937,[34] and has been used extensively in modeling solvation phenomena in the last few decades:[35−41]

$$G_{np} = \sum_i \tau_i A_i \tag{4}$$

where A_i is the surface area of one solute atom and τ_i is a surface tension parameter specific for that atom.

Solute Polarization. When the solute is described quantum mechanically, the reaction field potential polarizes its charge distribution (i.e., distorts the gas phase wave function) and this is usually taken into account by perturbing the gas-phase solute Hamiltonian, H^0 with a potential energy operator coming from the interaction of the virtual surface charges with solute electrons and nuclei (i.e., the effective Hamiltonian approach[12]):

$$H = H^0 + \int_S da' \frac{\sigma(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{5}$$

where da' is an area element of the solute surface (S). Thus, the construction of the solute wave function in solution has to be carried out self-consistently with the generation of the reaction field (so-called SCRF method), i.e., the surface charges. Then the solute energy in solution is given as

$$E = \left\langle \Psi \left| H^0 + \int_S da' \frac{\sigma(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \right| \Psi \right\rangle + E_{nucl} + \sum_i^{N_{sc}} \sum_A^{N_{at}} \frac{Z_A \sigma_i}{|\mathbf{r}_i - \mathbf{r}_A|} \tag{6}$$

where σ_i ($= \int_{dS_i} da'\, \sigma(\mathbf{r}')$) is the surface charge obtained by integrating the surface charge density ($\sigma(\mathbf{r}')$) over an element (dS_i) of solute surface area, (N_{Sc}) the number of surface charges, and (N_{at}) the number of solute atoms. The last two terms in eq 6 are the core−core repulsion and core−surface charge interaction, respectively.

The SCRF algorithm can be briefly summarized as follows: a gas-phase calculation (with or without geometry optimization) is performed first. This is necessary in the end to evaluate the energy due to solute polarization. Then one Poisson−Boltzmann calculation is performed for each SCF cycle until the solute wave function is self-consistent with the solvent reaction field. When finite difference (FDM) or finite element (FEM) methods are used to solve the PB equation the electrostatic potential is first determined as the solution of the PB equation and then the surface charges are obtained from the discontinuity in the electric field at the dielectric boundary. On the other hand, in the boundary element method (BEM), the virtual surface charges (VSC) are obtained at the start.[33] Implementations of continuum solvation models into ab initio Hamiltonians have been reported, for example, by Christoffersen (1976),[42] van Duijnen (1980),[43] Tomasi (1981),[44] Rivail (1983), [45] Mikkelsen (1987),[46] Karlstroem (1988),[47] Wiberg (1991),[48] Olivares del Valle (1993),[49] and Honig (1994).[26] Similar implementations into semiempirical methods have been reported, for example, by Rivail (1973−76),[50,51] Tapia (1975),[52] Sakurai (1987),[53] Miertus (1988),[54] Rinaldi (1983),[55] Zerner and Karelson (1986),[56] Cramer and Truhlar (1991),[57,58] Wang and Ford (1992),[59] Luke and Orzoco

TABLE 1: Experimental and Calculated (ab Initio and Semiempirical) Solvation Free Energies of Small Neutral Molecules in Water[a]

compound	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]
water				1.7	−6.3[g]
ab initio/ESP[h]	−11.9	1.7	−1.2 (0.2)		−8.6
AM1/Mul[i]	−2.3	0.1	−0.2		−0.5
AM1/CM1[j]	−7.6	0.3	−0.6		−5.7
AM1/CM2[k]	−7.7	0.3	−0.6		−5.8
PM3/Mul	−1.9	0.1	−0.2		−0.1
PM3/CM1	−7.5	0.3	−0.7		−5.5
PM3/CM2	−7.3	0.3	−0.6		−5.3
methanol				1.9 (1.8[l])	−5.1
ab initio/ESP	−8.6	1.2	−1.1 (−1.1)		−5.5
AM1/Mul	−2.2 (−5.5[l])	0.1	−0.2		−0.2 (−3.7[l])
AM1/CM1	−5.1	0.2	−0.5		−3.0
AM1/CM2	−5.0	0.2	−0.4		−3.0
PM3/Mul	−1.6	0.1	−0.2		0.4
PM3/CM1	−4.9	0.3	−0.5		−2.8
PM3/CM2	−4.7	0.2	−0.5		−2.6
ethanol				2.0 (2.0)	−5.0
ab initio/ESP	−8.2	1.2	−1.0 (−0.7)		−5.0
AM1/Mul	−2.4 (−5.1)	0.1	−0.3		−0.2 (−3.1)
AM1/CM1	−5.4	0.3	−0.5		−3.0
AM1/CM2	−5.2	0.3	−0.5		−2.8
PM3/Mul	−1.7	0.1	−0.2		0.5
PM3/CM1	−5.1	0.3	−0.6		−2.7
PM3/CM2	−4.9	0.3	−0.5		−2.5
acetic acid				2.1 (2.0)	−6.7
ab initio/ESP	−11.0	1.4	−1.4 (2.8)		−7.5
AM1/Mul	−6.4 (−10.1)	0.5	−1.0		−3.8 (−8.1)
AM1/CM1	−8.8	0.5	−1.1		−6.2
AM1/CM2	−9.5	0.6	−1.2		−6.8
PM3/Mul	−5.8	0.6	−1.0		−3.1
PM3/CM1	−8.8	0.6	−1.2		−6.0
PM3/CM2	−9.2	0.7	−1.2		−6.4
acetone				2.1	−3.9
ab initio/ESP	−8.1	1.7	−1.6 (4.0)		−4.2
AM1/Mul	−5.1	0.6	−1.1		−2.3
AM1/CM1	−6.6	0.9	−1.5		−3.6
AM1/CM2	−7.7	1.0	−1.6		−4.6
PM3/Mul	−4.3	0.6	−1.0		−1.5
PM3/CM1	−6.7	1.0	−1.5		−3.6
PM3/CM2	−7.4	1.0	−1.6		−4.2
cis-*N*-methylacetamide				2.2	−10.1
ab initio/ESP	−12.2	2.2	−2.4 (5.3)		−7.8
AM1/Mul	−8.7	1.0	−1.8		−5.4
AM1/CM1	−11.4	1.0	−1.8		−8.1
AM1/CM2	−12.5	1.4	−2.2		−8.8
PM3/Mul	−6.7	1.0	−1.6		−3.4
PM3/CM1	−9.7	1.1	−1.7		−6.3
PM3/CM2	−10.1	1.2	−1.8		−6.6
trans-*N*-methylacetamide				2.2 (2.2)	−10.1
ab initio/ESP	−12.3	2.5	−2.9 (5.5)		−7.6
AM1/Mul	−8.1 (−10.4)	1.0	−1.7		−4.8 (−8.3)
AM1/CM1	−11.3	1.0	−1.7		−8.0
AM1/CM2	−13.3	1.5	−2.4		−9.5
PM3/Mul	−6.4	1.0	−1.5		−3.1
PM3/CM1	−12.4	1.7	−2.6		−8.4
PM3/CM2	−11.6	1.5	−2.2		−7.8
acetamide				2.1 (2.0)	−9.7
ab initio/ESP	−14.4	2.6	−2.7 (4.9)		−9.7
AM1/Mul	−9.0 (−12.5)	1.1	−1.9		−5.8 (−10.5)
AM1/CM1	−12.0	1.0	−1.7		−8.9
AM1/CM2	−14.6	1.6	−2.5		−10.9
PM3/Mul	−7.3	1.2	−1.8		−4.0
PM3/CM1	−13.5	1.7	−2.6		−9.7
PM3/CM2	−12.6	1.5	−2.3		−8.9
2-propanol				2.2	−4.8
ab initio/ESP	−8.3	1.1	−1.0 (−0.4)		−4.9
AM1/Mul	−2.4	0.1	−0.3		−0.1
AM1/CM1	−5.3	0.3	−0.5		−2.8
AM1/CM2	−5.1	0.3	−0.5		−2.6
PM3/Mul	−1.8	0.1	−0.2		0.5
PM3/CM1	−5.1	0.3	−0.6		2.6
PM3/CM2	−4.9	0.3	−0.5		2.4

TABLE 1: (Continued)

compound	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]
phenol				2.4	−6.6
ab initio/ESP	−9.8	1.2	−1.4 (−4.0)		−6.3
AM1/Mul	−5.4	0.3	−0.7		−2.7
AM1/CM1	−8.5	0.4	−1.0		−5.7
AM1/CM2	−7.4	0.4	−0.9		−4.6
PM3/Mul	−3.6	0.2	−0.5		−1.0
PM3/CM1	−6.8	0.4	−0.9		−4.0
PM3/CM2	−6.7	0.4	−0.9		−3.9
SM5.2R					−6.5
toluene				2.4 (2.3)	−0.8
ab initio/ESP	−3.5	0.3	−0.5 (−3.1)		−0.8
AM1/Mul	−4.7 (−3.5)	0.2	−0.7		−2.0 (−1.2)
AM1/CM1	−4.7	0.2	−0.7		−2.0
AM1/CM2	−3.5	0.2	−0.5		−0.8
PM3/Mul	−3.0	0.2	−0.5		−0.3
PM3/CM1	−3.0	0.2	−0.5		−0.3
PM3/CM2	−3.1	0.2	−0.5		−0.4
SM5.2R[n]					−0.9
2-pentanone				2.4	−3.5
ab initio/ESP	−7.6	1.7	−1.5 (4.2)		−3.5
AM1/Mul	−4.8	0.6	−1.0		−1.7
AM1/CM1	−6.3	0.8	−1.4		−3.0
AM1/CM2	−7.3	1.0	−1.5		−3.8
PM3/Mul	−4.1	0.6	−0.9		−1.0
PM3/CM1	−6.4	0.9	−1.4		−3.0
PM3/CM2	−7.1	1.1	−1.5		−3.5
ethylbenzene				2.6	−0.8
ab initio/ESP	−3.3	0.3	−0.5 (−2.8)		−0.5
AM1/Mul	−4.8	0.2	−0.7		−2.0
AM1/CM1	−4.8	0.2	−0.7		−2.0
AM1/CM2	−3.6	0.2	−0.5		−0.8
PM3/Mul	−3.1	0.2	−0.5		−0.3
PM3/CM1	−3.1	0.2	−0.5		−0.3
PM3/CM2	−3.2	0.2	−0.5		−0.4
3,5-dimethylpyridine				2.5	−5.5
ab initio/ESP	−8.9	1.2	−1.5 (−2.7)		−4.4
AM1/Mull	−5.9	0.4	−0.9		−2.9
AM1/CM1	−8.3	0.7	−1.9		−5.0
AM1/CM2	−8.3	0.7	−1.5		−5.0
PM3/Mul	−3.5	0.2	−0.6		−0.7
PM3/CM1	−6.3	0.6	−1.2		−3.1
PM3/CM2	−8.2	0.9	−1.6		−4.7
4-methylpyridine				2.4	−4.9
ab initio/ESP	−9.2	1.9	−1.6 (−2.2)		−4.8
AM1/Mul	−5.8	0.3	−0.9		−3.1
AM1/CM1	−8.7	0.7	−2.2		−5.6
AM1/CM2	−8.3	0.7	−1.5		−5.2
PM3/Mul	−3.4	0.2	−0.6		−0.8
PM3/CM1	−6.5	0.6	−1.2		−3.5
PM3/CM2	−8.4	0.9	−1.7		−5.1
4-methyl-2-pentanone				2.5	−3.1
ab initio/ESP	−7.5	1.5	−1.4 (4.1)		−3.4
AM1/Mul	−4.7	0.6	−1.0		−1.5
AM1/CM1	−6.1	0.8	−1.3		−2.7
AM1/CM2	−7.1	1.0	−1.5		−3.5
PM3/Mul	−3.9	0.6	−0.9		−0.7
PM3/CM1	−6.2	0.9	−1.4		−2.7
PM3/CM2	−6.9	1.0	−1.5		−3.3
benzene				2.3	−0.9
ab initio/ESP	−3.3	0.3	−0.5 (−2.3)		−0.7
AM1/Mul	−4.5	0.2	−0.6		−2.0
AM1/CM1	−4.5	0.2	−0.6		−2.0
AM1/CM2	−3.3	0.1	−0.4		−0.9
PM3/Mul	−2.8	0.1	−0.4		−0.4
PM3/CM1	−2.8	0.1	−0.4		−0.4
PM3/CM2	−2.9	0.1	−0.4		−0.5
SM5.2R					−1.0
methanethiol				2.0 (1.9)	−1.2
ab initio/ESP	−4.7	1.0	−0.9 (0.7)		−1.7
AM1/Mul	−0.4 (−3.6)	0.0	0.0		1.6 (−1.6)
AM1/CM1	−1.8	0.1	−0.2		0.3
AM1/CM2	−1.4	0.1	−0.3		0.7
PM3/Mul	−0.3	0.0	0.0		1.7
PM3/CM1	−1.2	0.1	−0.2		0.9
PM3/CM2	−1.2	0.1	−0.2		0.9

TABLE 1: (Continued)

compound	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]
ethanethiol				2.1	−1.2
ab initio/ESP	−4.7	1.0	−0.9 (1.3)		−1.5
AM1/Mul	−0.6	0.0	−0.1		1.6
AM1/CM1	−2.0	0.2	−0.3		0.4
AM1/CM2	−1.6	0.2	−0.3		0.8
PM3/Mul	−0.4	0.0	−0.1		1.8
PM3/CM1	−1.4	0.1	−0.2		0.9
PM3/CM2	−1.4	0.1	−0.3		0.9
SM5.2R	−1.3			0.2[n]	−1.0
dimethyl sulfide				2.1	−1.5
ab initio/ESP	−4.6	0.9	−0.9 (−1.0)		−1.6
AM1/Mul	−0.5	0.0	0.0		1.7
AM1/CM1	−2.0	0.2	−0.3		0.4
AM1/CM2	−1.8	0.1	−0.4		0.5
PM3/Mul	−0.5	0.0	−0.1		1.7
PM3/CM1	−1.9	0.1	−0.3		0.4
PM3/CM2	−1.9	0.1	−0.3		0.4
methyl ethyl sulfide				2.3 (2.2)	−1.4
ab initio/ESP	−4.9	1.0	−0.9 (−0.6)		−1.6
AM1/Mul	−0.7 (−3.1)	0.0	−0.1		1.6 (−0.9)
AM1/CM1	−2.2	0.2	−0.3		0.3
AM1/CM2	−2.0	0.2	−0.4		0.5
PM3/Mul	−0.6	0.0	−0.1		1.7
PM3/CM1	−2.1	0.2	−0.3		0.4
PM3/CM2	−2.1	0.2	−0.4		0.4
diethyl sulfide				2.4 (0.4[n])	−1.4
ab initio/ESP	−4.8	1.1	−0.9 (0.1)		−1.3
AM1/Mul	−1.0	0.0	−0.1		1.5
AM1/CM1	−2.5	0.2	−0.4		0.2
AM1/CM2	−2.3	0.2	−0.4		0.4
PM3/Mul	−0.8	0.0	−0.1		1.7
PM3/CM1	−2.4	0.2	−0.4		0.3
PM3/CM2	−2.5	0.2	−0.5		0.2
SM5.2R	−1.5				−1.1
methylamine				1.9	−4.5
ab initio/ESP	−8.0	1.2	−1.0 (−0.8)		−4.9
AM1/Mul	−1.1	0.0	−0.1		0.8
AM1/CM1	−6.3	0.2	−0.8		−4.2
AM1/CM2	−4.8	0.1	−0.3		−2.8
PM3/Mul	−0.1	0.0	0.0		1.8
PM3/CM1	−3.9	0.1	−0.4		−1.9
PM3/CM2	−3.9	0.1	−0.3		−1.9
dimethylamine				2.1	−4.3
ab initio/ESP	−6.0	0.8	−0.7 (−1.6)		−3.1
AM1/Mul	−1.4	0.0	−0.1		0.7
AM1/CM1	−5.5	0.2	−0.7		−3.2
AM1/CM2	−4.1	0.1	−0.2		−1.9
PM3/Mul	−0.2	0.0	0.0		1.9
PM3/CM1	−3.1	0.1	−0.3		−0.9
PM3/CM2	−3.4	0.1	−0.3		−1.2
trimethylamine				2.2	−3.2
ab initio/ESP	−4.3	0.5	−0.4 (−1.7)		−1.5
AM1/Mul	−1.7	0.1	−0.1		0.7
AM1/CM1	−4.6	0.2	−0.7		−2.1
AM1/CM2	−3.2	0.1	−0.2		−0.8
PM3/Mul	−0.3	0.0	0.0		2.0
PM3/CM1	−1.6	0.1	−0.1		0.8
PM3/CM2	−2.7	0.1	−0.2		−0.3
ethylamine				2.2	−4.5
ab initio/ESP	−5.2	0.7	−1.0 (1.7)		−2.2
AM1/Mul	−1.3	0.0	−0.1		0.8
AM1/CM1	−6.7	0.2	−0.9		−4.4
AM1/CM2	−5.0	0.1	−0.3		−2.8
PM3/Mul	−0.1	0.0	0.0		2.0
PM3/CM1	−4.0	0.2	−0.4		−1.7
PM3/CM2	−4.0	0.2	−0.4		−1.7
n-propylamine				2.4	−4.4
ab initio/ESP	−8.1	1.2	−1.1 (−0.6)		−4.7
AM1/Mul	−1.2	0.0	−0.1		1.1
AM1/CM1	−6.1	0.2	−0.8		−3.6
AM1/CM2	−4.7	0.1	−0.3		−2.3
PM3/Mul	−0.1	0.0	0.0		2.2
PM3/CM1	−3.9	0.2	−0.4		−1.4
PM3/CM2	−3.8	0.1	−0.3		−1.4

TABLE 1: (Continued)

compound	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]
n-butylamine				2.4 (2.3)	−4.4
ab initio/ESP	−8.0	1.2	−1.0 (−0.6)		−4.4
AM1/Mul	−1.3 (−5.2)	0.0	−0.1		1.1 (−2.9)
AM1/CM1	−6.4	0.2	−0.8		−3.8
AM1/CM2	−4.9	0.1	−0.3		−2.4
PM3/Mul	−0.2	0.0	0.0		2.2
PM3/CM1	−4.1	0.2	−0.4		−1.5
PM3/CM2	−4.0	0.2	−0.4		−1.4
SM5.2R					−4.3
diethylamine				2.4	−4.1
ab initio/ESP	−5.2	0.7	−0.7 (−1.2)		−2.0
AM1/Mul	−1.6	0.1	−0.1		0.9
AM1/CM1	−5.2	0.2	−0.7		−2.6
AM1/CM2	−3.7	0.1	−0.3		−1.2
PM3/Mul	−0.3	0.0	0.0		2.1
PM3/CM1	−2.8	0.1	−0.3		−0.3
PM3/CM2	−3.1	0.1	−0.3		−0.6
unsigned max error[m]					
ab initio/ESP					2.5
AM1/Mul					5.8
AM1/CM1					2.1
AM1/CM2					2.9
PM3/Mul					7.0
PM3/CM1					4.0
PM3/CM2					3.5
unsigned av error[m]					
ab initio/ESP					0.7
AM1/Mul					3.6
AM1/CM1					1.1
AM1/CM2					1.4
PM3/Mul					4.4
PM3/CM1					1.9
PM3/CM2					1.8

[a] All entries are in kcal/mol. The solvation energy calculations with the ab initio Hamiltonian were carried on gas-phase geometries optimized at GVB/6-31G** level. Calculations with semiempirical Hamiltonians (AM1 and PM3) used gas-phase geometries optimized at semiempirical level. PB settings: exterior dielectric constant = 80, interior dielectric constant = 1, grid resolution 1.8 grids/Å, van der Waals radii from ref 26, probe radius 1.4 Å, no salts. [b] Reaction field energy. [c] Change in solute self-energy due to polarization of electron density (wave function distortion). [d] Solute polarization free energy $\{G_{RF}(\text{solution}-\text{gas}) + G_{wfd}\}$. The values in parentheses are obtained with Jaguar 3.5 and the other values are taken from ref 26. [e] Cavity plus dispersion−repulsion contribution (nonpolar) calculated as a function of molecular surface. [f] Solvation free energy. [g] Experimental solvation free energies from refs 95 and 96. [h] In ab initio calculations electrostatic potential derived atom charges (ESP) were used to solve the Poisson−Boltzmann equation. The ab initio results were obtained with Jaguar 3.5,[100] see also ref 97; do not include first shell correction. [i] Mulliken charges. [j] CM1 charges.[29] [k] CM2 charges.[28] [l] From ref 7. [m] Difference between the experimental and calculated solvation free energies. [n] From ref 94. The calculations were performed with the AM1 Hamiltonian and class IV charges (CM1[29]).

(1992),[60] Basilevsky (1992),[61] Rauhut and Clark (1993),[62] and Klamt (1993).[9]

The calculation of solute energy in solution by means of the effective Hamiltonian using any VSC-based approach requires the evaluation of one-electron terms (which contain the contribution from the reaction field).[12,58] A peculiarity of semiempirical methods is that the one-electron integral Fock terms and core−core repulsion terms are evaluated by using two-electron repulsion integrals.[63,64] Consequently, the surface charge−electron interaction terms (the second term in eq 7, vide infra), and the core−surface charge terms (the first term in eq 7) are also evaluated using two-electron repulsion integrals (in the same way that electron−core interactions are calculated).[9,53,54,59−62] Briefly, the virtual surface charges (σ_i) are given a core status; hence, hydrogenoid s-type orbitals are centered on these surface charges in order to calculate the two-electron repulsion integrals necessary to evaluate the one-electron and surface charge−core terms which add the solvent contribution to the solute Hamiltonian. Thus, the SCRF energy is

$$E_{RF} = \sum_{i}^{N_{sc}} \sigma_i \left(\sum_{A}^{N_{at}} \frac{Z_A}{|\mathbf{r}_i - \mathbf{r}_A|} - \left\langle ss \left| \frac{1}{|\mathbf{r}_i - \mathbf{r}|} \right| \Psi \right\rangle \right) \qquad (7)$$

where $\mathbf{r}_i$ is the position of surface charge σ_i and $\mathbf{r}_A$ the position of atom A. Note that the core−surface charge term in eq 7 has a Coulomb type form, but the actual semiempirical implementation uses a two-electron integral based formula over s-type orbitals.

Divide and Conquer (D&C) Algorithm. The cornerstone of our algorithm for performing quantum mechanical calculations on proteins in solution is, of course, the D&C method, which is a linear scaling algorithm for solving Schrödinger equation for large molecules. The D&C method was proposed and applied first to the DFT Hamiltonian by Yang and Lee[1] and shortly afterward was extended to semiempirical Hamiltonians by Dixon and Merz[2,3] and Yang and co-workers.[65,66] Here we review our version of the D&C method briefly. The basic idea in the D&C method is to divide the whole molecule into subsystems (Figure 1) and to replace the diagonalization of the Fock matrix for the whole molecule (which scales as N^3, where N is the number of basis functions) with a series of subsystem diagonalizations. This approach originated in the observation that the basis functions used to expand the molecular orbitals are localized and consequently the Fock matrix for large molecules is sparse.[1,5] The D&C method is equally applicable to both localized and delocalized species, but works more effectively with the former.[67] Following this partitioning, we obtain as many subsystem Fock equations as subsystems:

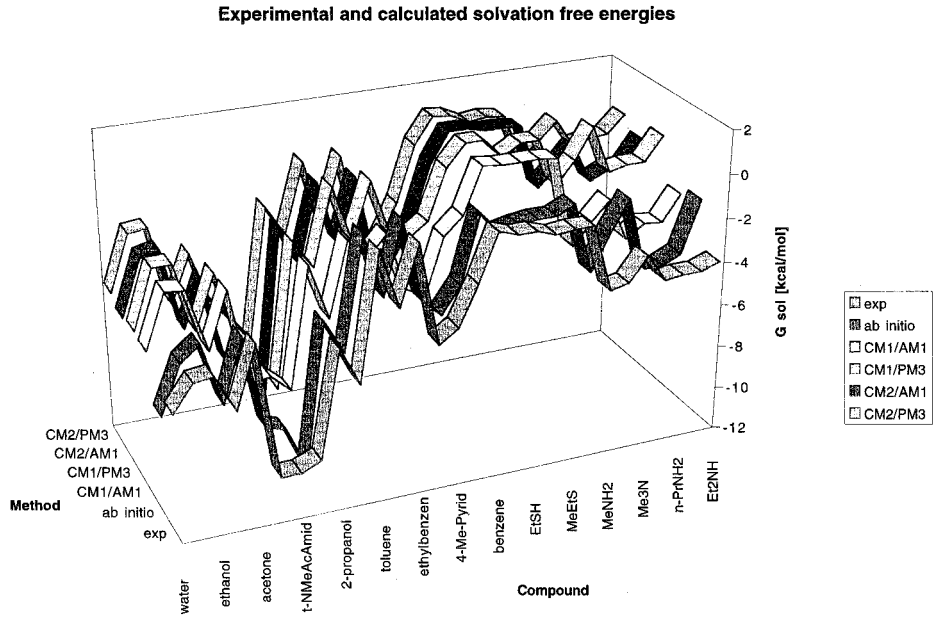

Figure 2. Comparison of experimental solvation free energies for small neutral molecules with those computed by ab initio (GVB) and semiempirical methods (AM1 and PM3 Hamiltonians with CM1 and CM2 charges). All the compounds in Table 1 were used, but on the abscissa only a few of them are shown.

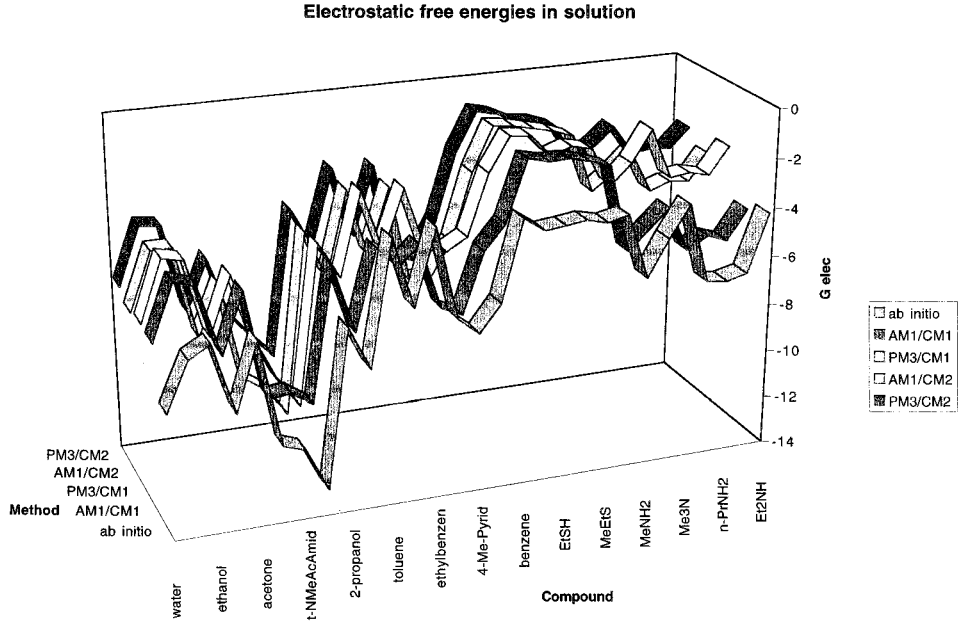

Figure 3. Comparison of ab initio (GVB) reaction field free energies for small neutral molecules (the set of compounds given in Table 1) with those obtained by semiempirical methods (AM1 and PM3).

$$\mathbf{F}^{\alpha}\mathbf{C}^{\alpha} = \mathbf{C}^{\alpha}\mathbf{E}^{\alpha}, \quad \alpha = 1, ..., n_{\text{sub}} \tag{8}$$

As a consequence, we can write the density matrix for the whole molecule $P_{\mu\nu}$ as a sum of the density matrixes of the subsystems $P^{\alpha}_{\mu\nu}$:

$$P_{\mu\nu} = \sum_{\alpha=1}^{n_{\text{sub}}} D^{\alpha}_{\mu\nu} P^{\alpha}_{\mu\nu} \tag{9}$$

The factor $D^{\alpha}_{\mu\nu}$ is assigned a value of 0 or $1/n_{\mu\nu}$ (where n is the number of times basis function μ and ν appear—generally this term is equal to 1) in order to mask out particular density matrix elements.[3] The subsystem density matrix $P^{\alpha}_{\mu\nu}$ in turn is built by introducing an occupation number n^{α}_i:

$$P^{\alpha}_{\mu\nu} = \sum_{i}^{\text{MOs}} n^{\alpha}_i (c^{\alpha}_{\mu i})^* c^{\alpha}_{\nu i} \tag{10}$$

which is calculated according to a recipe proposed by Yang and Lee:[1]

$$n^{\alpha}_i = 2/(1 + \exp[(\epsilon^{\alpha}_i - \epsilon_{\text{F}})/kT]) \tag{11}$$

k is Boltzmann constant and T is the absolute temperature; the actual value of T has little effect on the calculation unless a very large value is chosen. ϵ^{α}_i is the energy eigenvalue of the i "molecular orbital" in subsystem α and ϵ_{F} a Fermi energy level (an adjustable quantity) whose value is determined from the normalization condition on **P** (i.e., the total number of electrons in the system is conserved). Within the semiempirical framework

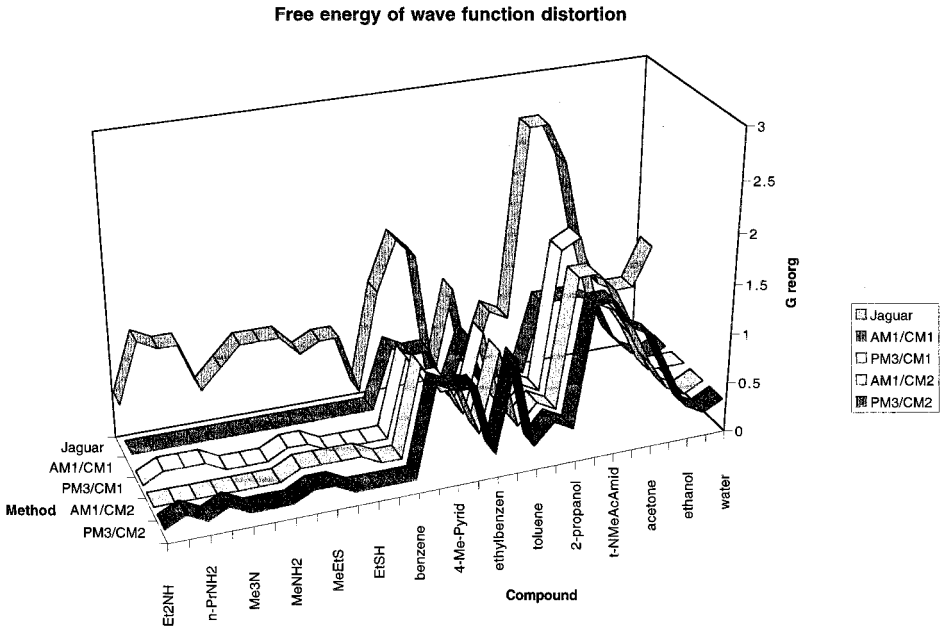

Figure 4. Comparison of ab initio (GVB) calculated free energies due to wave function distortion for small neutral molecules (the set given in Table 1) with the values obtained by semiempirical methods (AM1 and PM3).

this condition involves only the diagonal elements of **P**:

$$n_{\text{elec}} = \sum_{\mu}^{N} \sum_{\alpha}^{n_{\text{sub}}} D_{\mu\mu}^{\alpha} \sum_{i}^{\text{MOs}} n_i^{\alpha} |c_{\mu i}^{\alpha}|^2 \qquad (12)$$

Because $\mathbf{P}^{\alpha}$ depends on ϵ_F, it cannot be assembled until the Roothaan equations have been solved for all subsystems. In practice, we use ϵ_F from the previous SCF cycle to determine the subsystem density matrices.[3] The performance of our D&C method as applied to semiempirical Hamiltonians is fully documented in the literature.[2,3] Below we focus on the coupling between D&C method and the Poisson−Boltzmann equation.

Numerical Solution of the Linear/Nonlinear Poisson−Boltzmann Equation. The PB equation accounts for solvent screening and salt effects and is usually cast in the following two forms (nonlinear/linear):[17]

$$\nabla\cdot[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \kappa(\mathbf{r})^2 \sinh(\phi(\mathbf{r})) = -4\pi\rho(\mathbf{r}) \qquad (13)$$

$$\nabla\cdot[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \kappa(\mathbf{r})^2\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \qquad (14)$$

where $\phi(\mathbf{r})$ is the electrostatic potential at position $\mathbf{r}$, $\epsilon(\mathbf{r})$ is the dielectric constant, $\kappa(\mathbf{r})$ is a modified Debye−Huckel parameter, and $\rho(\mathbf{r})$ is the solute charge density.[21] The PB equation admits analytical solutions for only a few regular shapes, e.g., sphere, cylinder, and only numerical solutions are possible for molecules with complicated shapes.[32] The formidable increase in computer power (both in speed and memory) over the last few decades and the challenging problems raised in life sciences stimulated new efforts toward the development of high-performance algorithms for solving the PB equation for molecules of biological interest. Several mathematical techniques commonly used for solving partial differential equations have been explored: (a) the finite difference method (FDM),[17,21,22,24,68−76] (b) the finite element method (FEM),[31,77−79] and (c) the variational method (VM).[80−83] Another technique, known as the boundary element method (BEM[12,32,33,84−90] does not solve directly the PB equation, but provides the same answer as far as the Poisson equation is concerned (it can be shown that this is equivalent to solving Poisson equation directly for the electrostatic potential, see for example ref 91). In our D&C-PB coupling we have made use of the FDM as implemented in the DelPhi program of Honig and co-workers.[69]

Finite Difference Method. In the FDM a grid with appropriate spacing (h) is constructed around the solute and the dielectric (exterior and interior) constants ϵ, Debye−Hückel parameter κ, and the solute point charges ρ_i are assigned to grid points. Boundary conditions are fulfilled by assigning fixed electric potential values to boundary grid points. These values are calculated either from Coulomb's law or Debye−Hückel theory (when ionic salts are present). Richards' surface[92] is taken as the dielectric interface which separates grid points of different dielectric constants. The ion-accessible surface is the solvent-accessible surface for a probe sphere radius of 2 Å (sodium cation) and separates the grid points for which an ionic salt concentration (Debye−Hückel parameter) is assigned. The above construction leads to a large system of linear equations with the electric potential at the grid points being the unknowns. This system is sparse (tridiagonal with fringes[93]) and can be solved efficiently by iterative methods such as simultaneous overrelaxation (SOR)[68,93] or incomplete Cholesky conjugate gradient (ICCG) methods,[22,93] for example. The finite difference equations give the electrostatic potential at a grid point (ϕ_0) as a function of the electrostatic potentials at neighboring grid points:

$$\phi_0 = \left[\frac{(\sum_{i=1}^{6} \epsilon_i \phi_i) + 4\pi q_0/h}{(\sum_{i=1}^{6} \epsilon_i) + (\kappa_0 h)^2} \right] \qquad (15)$$

where q_0 is the charge on the grid point (if any).

FDM has been extensively developed over the past decade[17,21,22,24,68−76] and extended to deal with the nonlinear form of the PB equation.[71]

Results and Discussion

We assessed the merits of our SCRF methodology (D&C-PB method) by comparing our calculated solvation free energies

TABLE 2: Experimental and Calculated (ab Initio and Semiempirical) Solvation Free Energies of Small Charged Molecules in Water[a]

compound	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]	compound	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]
C_2H^-					−73.0[i]	imidazole (H^+)					−64.0
ab initio/ESP[g]	−74.9	1.5	4.4	2.0[h]	−71.4	ab initio/ESP	−67.2	0.9	−1.3	2.2	−64.2
AM1/Mul[j]	−77.8	0.1	−0.3		−75.8	AM1/Mul	−64.7	0.2	−0.3		−62.4
AM1/CM1[k]	−77.8	0.1	−0.3		−75.8	AM1/CM1	−66.9	0.2	−0.4		−64.5
AM1/CM2[l]	−77.6	0.1	−0.2		−75.6	AM1/CM2	−67.2	0.2	−0.4		−64.9
PM3/Mul	−78.6	0.1	−0.3		−76.6	PM3/Mul	−65.8	0.5	−0.8		−63.2
PM3/CM1	−78.6	0.1	−0.3		−76.6	PM3/CM1	−68.1	0.5	−1.0		−65.5
PM3/CM2	−78.7	0.1	−0.3		−76.7	PM3/CM2	−66.0	0.3	−0.7		−63.6
SM5.2R[m]	−80.5			1.7[n]	−78.8	SM5.2R	−61.4			−0.3	−61.7
$CH_3OH_2^+$					−87.0	pyridine (H^+)				2.3	−58.0
ab initio/ESP	−83.0	1.0	0.1	1.9	−80.1	ab initio/ESP	−61.4	1.0	3.9		−58.1
AM1/Mul	−76.6	0.3	−0.3		−74.5	AM1/Mul	−61.2	0.4	−0.6		−58.5
AM1/CM1	−78.4	0.4	−0.5		−76.1	AM1/CM1	−62.4	0.5	−0.8		−59.6
AM1/CM2	−78.3	0.4	−0.5		−76.0	AM1/CM2	−61.3	0.3	−0.4		−58.7
PM3/Mul	−79.2	0.5	−0.7		−76.8	PM3/Mul	−63.0	1.0	−1.5		−59.7
PM3/CM1	−81.1	0.6	−0.8		−78.6	PM3/CM1	−64.0	1.0	−1.6		−60.7
PM3/CM2	−80.8	0.5	−0.8		−78.4	PM3/CM2	−61.8	0.6	−1.1		−58.9
SM5.2R	−81.4			−1.6	−83.0	SM5.2R	−58.3			−1.1	−59.4
$(CH_3)_2OH^+$				2.1	−70.0	$C_6H_5NH_3^+$					−68.0
ab initio/ESP	−72.0	0.8	2.7		−69.1	ab initio/ESP	−78.2	4.6	8.1	2.4	−71.1
AM1/Mul	−67.0	0.3	−0.1		−64.6	AM1/Mul	−73.1	2.3	−3.7		−68.4
AM1/CM1	−69.1	0.3	−0.3		−66.7	AM1/CM1	−75.6	2.5	−4.2		−70.7
AM1/CM2	−69.0	0.3	−0.3		−66.6	AM1/CM2	−73.5	2.2	−3.5		−68.9
PM3/Mul	−68.4	0.3	−0.4		−66.0	PM3/Mul	−76.6	2.9	−4.6		−71.3
PM3/CM1	−70.3	0.3	−0.5		−67.8	PM3/CM1	−79.0	3.1	−5.1		−73.5
PM3/CM2	−70.2	0.4	−0.5		−67.7	PM3/CM2	−76.0	2.8	−4.6		−70.8
SM5.2R	−69.6			−1.1	−70.7	SM5.2R	−64.1			−3.4	−67.5
$CH_3CH_2OH_2^+$				2.1	−81.0	NH_4^+					−81.0
ab initio/ESP	−80.7	2.2	3.1		−76.4	ab initio/ESP	−92.0	0.1	0.7	1.8	−90.1
AM1/Mul	−73.2	0.7	−1.0		−70.4	AM1/Mul	−90.0	0.2	0.0		−88.1
AM1/CM1	−76.6	0.9	−1.3		−73.6	AM1/CM1	−89.0	0.2	−0.1		−87.1
AM1/CM2	−76.7	0.9	−1.3		−73.7	AM1/CM2	−88.4	0.1	0.1		−86.5
PM3/Mul	−74.7	0.9	−1.4		−71.7	PM3/Mul	−96.0	0.0	−0.2		−94.2
PM3/CM1	−77.9	0.9	−1.6		−74.8	PM3/CM1	−90.8	0.0	0.0		−89.1
PM3/CM2	−77.6	1.0	−1.6		−74.5	PM3/CM2	−90.8	0.0	0.0		−89.1
SM5.2R	−76.4			−1.7	−78.1	SM5.2R	−84.5			−4.1	−88.6
$(CH_3)_2COH^+$				2.1	−64.0	CN^-					−75.0
ab initio/ESP	−77.9	4.5	8.3		−71.2	ab initio/ESP	−69.3	0.1	−1.3	2.0	−67.2
AM1/Mul	−63.2	0.2	−0.1		−60.8	AM1/Mul	−82.8	0.0	−0.4		−81.0
AM1/CM1	−65.6	0.3	−0.3		−63.1	AM1/CM1	−87.1	0.0	−1.7		−85.3
AM1/CM2	−65.8	0.3	−0.3		−63.3	AM1/CM2	−88.6	0.0	−2.1		−86.8
PM3/Mul	−64.2	0.2	−0.3		−61.8	PM3/Mul	−83.0	0.0	−0.5		−81.2
PM3/CM1	−66.6	0.3	−0.5		−64.1	PM3/CM1	−91.2	1.4	−2.3		−88.0
PM3/CM2	−66.3	0.2	−0.4		−63.9	PM3/CM2	−91.5	1.5	−2.4		−88.2
SM5.2R	−66.5			−2.6	−69.1	SM5.2R	−77.5			−0.1	−76.6
H_3O^+					−105.0	CH_2CN^-					−75.0
ab initio/ESP	−101.7	0.5	−1.8	1.7	−99.5	ab initio/ESP	−77.2	3.1	5.0	2.0	−72.2
AM1/Mul	−96.1	0.0	0.3		−94.4	AM1/Mul	−72.3	0.3	−0.4		−70.0
AM1/CM1	−99.3	0.0	0.2		−97.6	AM1/CM1	−76.2	1.2	−1.9		−73.0
AM1/CM2	−99.3	0.0	0.2		−97.6	AM1/CM2	−76.3	1.3	−1.8		−73.0
PM3/Mul	−95.7	0.0	0.4		−94.0	PM3/Mul	−73.6	0.5	−0.8		−71.0
PM3/CM1	−98.4	0.0	0.1		−96.7	PM3/CM1	−80.0	1.9	−2.9		−76.1
PM3/CM2	−98.2	0.0	0.1		−96.5	PM3/CM2	−81.6	2.3	−3.5		−77.3
SM5.2R	−100.1			0.7	−99.4	SM5.2R	−68.2			−1.9	−70.2
CH_3O^-					−98.0	NH_2^-					−95.0
ab initio/ESP	−83.1	1.1	−6.8	2.0	−79.9	ab initio/ESP	−110.7	5.0	9.5	1.7	−104.1
AM1/Mul	−84.3	1.3	−1.7		−81.1	AM1/Mul	−98.5	0.3	−0.2		−96.5
AM1/CM1	−86.1	1.5	−2.0		−82.7	AM1/CM1	−125.0	1.8	−8.0		−121.5
AM1/CM2	−85.6	1.5	−1.8		−82.2	AM1/CM2	−105.6	0.7	−1.3		−103.2
PM3/Mul	−85.8	1.2	−1.9		−82.7	PM3/Mul	−98.1	0.0	−0.1		−96.4
PM3/CM1	−88.3	1.5	−2.3		−84.9	PM3/CM1	−108.7	0.4	−1.6		−106.6
PM3/CM2	−87.9	1.4	−2.2		−84.6	PM3/CM2	−108.6	0.4	−1.5		−106.5
SM5.2R	−80.6			−2.0	−86.2	SM5.2R	−85.6			−5.0	−90.5
CH_3COO^-					−77.0	NO_2^-					−73.0
ab initio/ESP	−84.2	3.5	6.3	2.1	−78.7	ab initio/ESP	−68.3	0.2	−0.9	2.0	−66.0
AM1/Mul	−79.7 (−83.2[o])	1.4	−2.9	1.9[o]	−76.2 (−81.2[o])	AM1/Mul	−78.6	0.1	−0.1		−76.6
AM1/CM1	−81.2	1.6	−3.2		−77.5	AM1/CM1	−77.9	0.0	0.0		−76.0
AM1/CM2	−81.4	1.6	−3.2		−77.8	AM1/CM2	−78.7	0.0	−0.1		−76.8
PM3/Mul	−80.0	1.5	−2.9		−76.5	PM3/Mul	−79.8	0.3	−0.3		−76.6
PM3/CM1	−82.5	1.8	−3.4		−78.7	PM3/CM1	−78.5	0.2	−0.1		−76.4
PM3/CM2	−82.7	1.8	−3.4	−1.2	−78.9	PM3/CM2	−77.9	0.1	0.0		−75.9
SM5.2R	−71.7				−72.8	SM5.2R	−76.1			0.5	−75.6

TABLE 2: (Continued)

compound	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]	compound	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]
$CH_3COCH_2^-$					−81.0	NO_3^-					−66.0
ab initio/ESP	−80.7	4.3	5.9	2.1	−74.2	ab initio/ESP	−61.4	0.0	0.4	2.1	−59.2
AM1/Mul	−74.7	1.4	−2.5		−71.1	AM1/Mul	−68.0	0.0	0.0		−66.0
AM1/CM1	−76.1	1.7	−2.8		−72.2	AM1/CM1	−67.7	0.0	0.0		−65.7
AM1/CM2	−75.8	1.7	−2.7		−71.9	AM1/CM2	−69.6	0.0	−0.1		−67.6
PM3/Mul	−74.6	1.4	−2.4		−71.0	PM3/Mul	−68.0	0.1	−0.1		−65.9
PM3/CM1	−76.9	1.8	−2.9		−72.9	PM3/CM1	−67.5	0.1	0.0		−65.4
PM3/CM2	−76.9	1.8	−2.8		72.9	PM3/CM2	−67.3	0.0	0.0		−65.3
SM5.2R	−69.0			−1.6	−70.6	SM5.2R	−68.8			5.2	−63.6
$C_6H_5O^-$					−75.0	N_3^-					−74.0
ab initio/ESP	−72.9	5.5	5.5	2.4	−65.1	ab initio/ESP	−80.5	0.4	10.3	1.8	−78.4
AM1/Mul	−68.6	2.2	−3.3		−64.0	AM1/Mul	−78.3	0.0	0.0		−76.4
AM1/CM1	−69.8	2.5	−3.7		−64.9	AM1/CM1	−92.4	0.4	−3.6		−90.1
AM1/CM2	−69.0	2.5	−3.5		−64.1	AM1/CM2	−78.3	0.0	0.0		−76.4
PM3/Mul	−67.5	2.1	−3.0		−63.0	PM3/Mul	−80.7	0.0	−0.2		−78.8
PM3/CM1	−69.6	2.5	−3.5		−64.7	PM3/CM1	−80.7	0.0	−0.2		−78.8
PM3/CM2	−69.7	2.5	−3.5		−64.8	PM3/CM2	−80.7	0.0	−0.2		−78.8
SM5.2R	−61.7			−3.3	−65.0	SM5.2R	−66.1			−10.1	−76.2
$C_6H_5CH_2^-$					−59.0	$CH_3SH_2^+$					−74.0
ab initio/ESP	−60.6	1.9	−2.4	2.4	−56.2	ab initio/ESP	−76.0	1.0	2.3	2.0	−73.0
AM1/Mul	−63.1	1.2	−2.0		−59.5	AM1/Mul	−77.9	1.1	−1.8		−74.8
AM1/CM1	−63.1	1.2	−2.0		−59.5	AM1/CM1	−72.8	0.4	−0.5		−70.4
AM1/CM2	−63.1	1.0	−1.7		−59.9	AM1/CM2	−74.4	0.7	−1.0		−71.7
PM3/Mul	−61.7	0.8	−1.6		−58.5	PM3/Mul	−75.1	0.7	−0.6		−72.8
PM3/CM1	−61.7	0.8	−1.6		−58.5	PM3/CM1	−72.1	0.3	−0.4		−69.8
PM3/CM2	−61.9	0.9	−1.7		−58.6	PM3/CM2	−73.0	0.3	−0.4		−70.7
SM5.2R	−56.2			−0.2	−56.4	SM5.2R	−74.0			0.8	−73.2
OH^-					−110.0	$(CH_3)_2SH^+$					−61.0
ab initio/ESP	−114.5	3.1	5.0	1.6	−109.7	ab initio/ESP	−67.2	0.7	2.3	2.1	−64.3
AM1/Mul	−101.7	0.3	−0.2		−99.7	AM1/Mul	−73.7	1.5	−2.6		−70.0
AM1/CM1	−105.5	0.6	−0.7		−103.2	AM1/CM1	−63.9	0.3	−0.3		−61.4
AM1/CM2	−105.6	0.6	−0.7		−103.3	AM1/CM2	−68.3	0.9	−1.4		−65.2
PM3/Mul	−102.2	0.3	−0.3		−100.2	PM3/Mul	−67.3	0.4	−0.8		−64.7
PM3/CM1	−106.9	0.6	−0.9		−104.6	PM3/CM1	−63.4	0.2	−0.3		−61.0
PM3/CM2	−106.8	0.6	−0.9		−104.5	PM3/CM2	−64.7	0.2	−0.4		−62.3
SM5.2R	−101.7			−7.4	−109.0	SM5.2R	−68.4			1.3	−67.1
HO_2^-					−101.0	HS^-					−76.0
ab initio/ESP	−103.7	3.5	6.8	1.8	−98.4	ab initio/ESP	−83.2	1.2	5.0	1.8	−80.2
AM1/Mul	−93.4	0.9	−1.2		−90.7	AM1/Mul	−86.3	0.3	−0.6		−84.2
AM1/CM1	−98.7	1.5	−2.1		−95.4	AM1/CM1	−85.3	0.2	−0.3		−83.3
AM1/CM2	−98.0	1.4	−2.2		94.8	AM1/CM2	−87.3	0.3	−0.8		−85.2
PM3/Mul	−92.3	0.7	−1.2		−89.8	PM3/Mul	−84.2	0.0	0.1		−82.4
PM3/CM1	−96.9	1.1	−1.7		−94.0	PM3/CM1	−84.2	0.0	0.1		−82.4
PM3/CM2	−96.2	1.1	−1.7		−93.3	PM3/CM2	−85.5	0.0	−0.1		−83.7
SM5.2R	−87.9			−7.2	−95.1	SM5.2R	−79.9			−2.4	−82.2
$CH_3NH_3^+$					−73.0	CH_3S^-					−76.0
ab initio/ESP	−81.8	1.4	4.4	2.0	−78.4	ab initio/ESP	−85.1	4.3	5.1	2.0	−78.8
AM1/Mul	−77.1	0.5	−0.8		−74.6	AM1/Mul	−84.0	1.4	−2.8		−80.6
AM1/CM1	−77.9	0.6	−1.3		−75.4	AM1/CM1	−82.4	1.2	−2.2		−79.2
AM1/CM2	−77.2	0.6	−0.9		−74.8	AM1/CM2	−88.6	1.9	−3.9		−84.7
PM3/Mul	−83.1	0.6	−1.5		−80.6	PM3/Mul	−81.8	0.7	−1.8		−79.0
PM3/CM1	−82.9	0.8	−1.9		−80.2	PM3/CM1	−81.9	0.7	−1.7		−79.3
PM3/CM2	−80.9	0.7	−1.6		−78.4	PM3/CM2	−85.3	1.0	−2.5		−82.4
SM5.2R	−74.6			−2.3	−77.0	SM5.2R	−75.7			−1.8	−77.5
$HC(OH)NH_2^+$					−78.0	$CH_3CH_2S^-$					−74.0
ab initio/ESP	−79.2	0.8	−4.1	2.0	−76.4	ab initio/ESP	−84.3	5.5	7.9	2.1	−76.6
AM1/Mul	−74.6	0.2	−0.3		−72.4	AM1/Mul	−83.5	2.0	−4.0		−79.4
AM1/CM1	−78.1	0.2	−0.4		−75.9	AM1/CM1	−81.7	1.8	−3.3		−77.8
AM1/CM2	−78.5	0.2	−0.4		−76.3	AM1/CM2	−88.4	2.6	−5.1		−83.7
PM3/Mul	−75.7	0.5	−0.8		−73.2	PM3/Mul	−80.9	1.2	−2.9		−77.6
PM3/CM1	−77.4	0.4	−0.8		−75.0	PM3/CM1	−81.2	1.2	−2.8		−77.9
PM3/CM2	−76.6	0.3	−0.6		−74.3	PM3/CM2	−84.9	1.5	−3.7		−81.3
SM5.2R	−72.7			−7.6	−80.3	SM5.2R	−73.7			−1.8	−75.7
CH_3CNH^+					−69.0	*n*-$C_3H_7S^-$					−76.0
ab initio/ESP	−67.8	0.3	−1.9	2.1	−65.4	ab initio/ESP	−84.5	6.3	8.7	2.3	−75.9
AM1/Mul	−69.3	0.2	−0.2		−67.1	AM1/Mul	−83.0	2.3	−4.8		−78.4
AM1/CM1	−69.6	0.0	0.1		−67.6	AM1/CM1	−81.3	2.2	−4.1		−76.8
AM1/CM2	−69.7	0.1	0.1		−67.6	AM1/CM2	−88.0	2.9	−5.8		−82.8
PM3/Mul	−75.0	1.1	−2.0		−71.9	PM3/Mul	−80.6	1.5	−3.6		−76.8
PM3/CM1	−72.5	0.6	−1.1		−69.9	PM3/CM1	−80.9	1.6	−3.6		−77.0
PM3/CM2	−72.2	0.5	−1.0		−69.7	PM3/CM2	−84.6	1.9	−4.3		−80.4
SM5.2R	−67.1			−1.2	−68.4	SM5.2R	−73.9			−1.7	−75.6

TABLE 2: (Continued)

compound	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]	compound	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]
$CH_3C(OH)NH_2^+$					−70.0	$C_6H_5S^-$					−65.0
ab initio/ESP	−75.5	1.4	−1.4	2.1	−71.6	ab initio/ESP	−76.5	−12.1	−16.6	2.5	−86.1
AM1/Mul	−66.8	0.4	−0.5		−64.2	AM1/Mul	−74.5	4.0	−6.0		−68.0
AM1/CM1	−69.5	0.3	−0.4		−67.0	AM1/CM1	−74.4	4.0	−5.7		−67.9
AM1/CM2	−70.5	0.5	−0.7		−67.8	AM1/CM2	−78.9	5.3	−7.3		−71.1
PM3/Mul	−71.0	1.0	−1.7		−67.8	PM3/Mul	−74.6	3.7	−5.7		−68.4
PM3/CM1	−74.4	1.3	−2.1		−70.9	PM3/CM1	−75.6	3.9	−5.8		−69.2
PM3/CM2	−72.4	1.0	−1.6		−69.2	PM3/CM2	−78.8	4.5	−6.7		−71.7
SM5.2R	−66.2			−8.0	−74.1	SM5.2R	−63.0			−3.4	−66.3
$(CH_3)_2NH_2^+$					−66.0	unsigned max error[n]					
ab initio/ESP	−73.0	1.4	5.3	2.1	−69.6	ab initio/ESP					21.1
AM1/Mul	−68.4	0.5	−0.6		−65.8	AM1/Mul					16.9
AM1/CM1	−70.1	0.6	−1.1		−67.4	AM1/CM1					26.5
AM1/CM2	−69.7	0.6	−0.8		−67.0	AM1/CM2					15.8
PM3/Mul	−72.3	0.6	−1.2		−69.6	PM3/Mul					15.3
PM3/CM1	−74.9	0.9	−2.0		−71.9	PM3/CM1					13.1
PM3/CM2	−72.3	0.7	−1.5		−69.5	PM3/CM2					13.4
SM5.2R	−67.5			0.0	−67.5	SM5.2R					15.4
$(CH_3)_3NH^+$					−59.0	unsigned avg error[p]					
ab initio/ESP	−64.9	0.7	4.0	2.2	−62.0	ab initio/ESP					4.8
AM1/Mul	−60.7	0.3	−0.2		−58.1	AM1/Mul					5.1
AM1/CM1	−61.8	0.4	−0.4		−59.1	AM1/CM1					5.0
AM1/CM2	−61.3	0.3	−0.3		−58.7	AM1/CM2					4.9
PM3/Mul	−63.3	0.3	−0.5		−60.7	PM3/Mul					5.1
PM3/CM1	−64.8	0.5	−1.0		−62.0	PM3/CM1					4.7
PM3/CM2	−63.0	0.4	−0.7		−60.3	PM3/CM2					4.9
SM5.2R	−62.2			2.6	−59.6	SM5.2R					3.8

[a] All entries are in kcal/mol. For geometries used in solvation calculations see footnote *a* in Table 1. PB settings: exterior dielectric constant = 80, interior dielectric constant = 1, grid resolution 1.8 grids/Å, van der Waals radii from ref 26, prove radius 1.4 Å, no salts. [b] Reaction field free energy. [c] Change in solute self-energy due to polarization of electron density. [d] Polarization free energy $\{G_{RF}(\text{solution}-\text{gas}) + G_{wdf}\}$. [e] Cavity plus dispersion−repulsion contribution (nonpolar) calculated as a function of molecular surface. [f] Solvation free energy. [g] The ab initio solvation free energies were calculated with Jaguar 3.5,[100] using electrostatic potential derived atom charges to solve the Poisson−Boltzmann equation. See ref 97. [h] Cavity + dispersion term used for ab initio and AM1/PM3 Hamiltonians (D&C-PB). [i] Experimental solvation free energies.[95,96] [j] Mulliken charges. [k] CM1 charges.[29] [l] CM2 charges.[28] [m] Cramer and Truhlar's solvation model SM5.2R.[94] [n] Cavity + dispersion term used in SM5.2R solvation model.[94] [o] From ref 7. [p] Difference between experimental and calculated solvation free energies.

of a set of 29 neutral organic molecules taken from Tannor et al.[26] (see their Table 2), and 36 ions taken from Cramer and Truhlar (see their Table 10).[94] We then compared our results (see Tables 1 (neutral) and 2 (ions)) against experiment[95,96] and the calculated values of York et al.,[7] Cramer, Truhlar, and co-workers,[94] and Friesner, Honig, Goddard, and co-workers.[26,97] Table 3 gives a comparison of experimental and calculated (semiempirical and ab initio) gas-phase dipole moments and Tables 4 and 5 show semiempirical and ab initio computed solvation free energies for *N*-acetyl-*N*′-methylamide derivatives of the 20 naturally occurring amino acids,[98] and DNA bases and nucleotides. Finally, the solvation free energies in water (and the CPU required to calculate them on a Sillicon Graphics Origin 200 workstation) for a set of proteins ranging from 46 to 275 residues and a piece of DNA (Dickerson's DNA dodecamer[99]) are given in Table 6.

Small Neutral Molecules. Table 1 and Figure 2 shows good agreement between the solvation free energies of small organic solutes calculated by the AM1/PM3 Hamiltonian (using Cramer, Truhlar, and co-workers class IV CM1/CM2 charge models[28,29]), by ab initio methods (GVB: Jaguar 3.5[97,100]) and the experimental values.[95,96] Solvation free energies for a few small neutrals calculated with the SM5.4R model are also included in Table 1 for comparison.[94] Our AM1/CM1 approach gives an average error of only 1.1 kcal/mol, which compares well with the GVB result of 0.6 kcal/mol for the same data set. Tomasi and co-workers [101] reported an average error of 1.6 kcal/mol for a set of 43 small neutral compounds with their HF/PCM model (Pauling radii) and 0.9 kcal/mol for the HF/SCIPCM model. These additional results show that a further increase in the accuracy of estimating solvation free energy is dependent on the selection of the van der Waals radii set and the nonpolar term. Our solvation free energies are also in good agreement with those calculated by Shao et al. [11] (XRISM model), York et al. [7] using an AM1/COSMO approach (Table 1), while a recent SM5 approach, with a more elaborated nonpolar term, gives a solvation free energy error of 0.45 kcal/mol on a set of 248 neutrals.[94] Figure 2 shows that solvation free energies (and reaction field free energies, Figure 3) calculated with the semiempirical-PB method tracks the ab initio results and the experimental values for most compounds in the data set, but notable exceptions are the alchohols, amines, and thiols. The difficulty with these compounds may stem either from the semiempirical Hamiltonian which likely fails to give the proper charge distribution for these systems or the nonpolar term (we used a rather crude approximation) of solvation free energy. Difficulties with amines have already been noted by Friesner and co-workers[26,97] and have been variably ascribed to first shell hydrogen bonding effects, charge transfer effects, etc.

Small Charged Molecules. A comparison between ab initio and semiempirical methods for small charged molecules is given in Table 2. The average error for the PM3/CM1 approach, for example, is 4.7 kcal/mol which is slightly lower than the ab initio result (4.8 kcal/mol). The conclusion from Tables 1 (Figure 2) and 2 is that both AM1 and PM3 Hamiltonians perform quite well with the class IV charges CM1/CM2, while the class II Coulson charges generally give poor agreement with experimental results. For ions, our SCRF method performs as well as the ab initio method, giving us confidence in applying this methodology to larger molecules such as proteins and nucleic acids which are not yet accessible at the ab initio level. Cramer,

TABLE 3: Experimental and Calculated (ab Initio and Semiempirical) Dipole Moments of Small Neutral Molecules in Vaccuum and Water[a]

	ab initio		AM1		PM3				ab initio		AM1		PM3		
compound	gas[c]	water[d]	gas	water	gas	water	exp[b]	compound	gas[c]	water[d]	gas	water	gas	water	exp[b]
water	2.09	2.50	1.09	1.17	0.97	1.05	1.85	4-methyl-2-pentanone	2.69	3.61	2.30	3.12	2.11	2.87	2.70
			2.02	2.14	1.92	2.07					2.69	3.64	2.69	3.66	
			2.00	2.11	1.87	2.00					3.60	4.75	3.50	4.60	
methanol	1.80	2.29	1.14	1.28	0.93	1.05	1.70	methanethiol	1.70	2.35	0.41	0.49	0.55	0.67	1.53
			1.63	1.83	1.59	1.80					1.28	1.50	1.23	1.43	
			1.59	1.78	1.52	1.71					1.25	1.52	1.19	1.41	
ethanol	1.67	2.24	1.12	1.33	0.94	1.11	1.69	ethanethiol	1.71	2.51	0.63	0.76	0.79	0.95	1.58
			1.56	1.84	1.56	1.85					1.37	1.66	1.39	1.65	
			1.51	1.78	1.49	1.75					1.45	1.76	1.34	1.61	
acetic acid	1.56	1.91	2.30	2.86	2.16	2.73	1.74	dimethyl sulfide	1.74	2.49	0.00	0.02	0.36	0.43	1.50
			1.98	2.49	1.96	2.51					1.17	1.41	1.29	1.63	
			2.77	3.37	2.69	3.30					1.06	1.36	1.23	1.49	
acetone	2.70	3.81	2.55	3.28	2.32	3.02	2.88	methyl ethyl sulfide	1.68	2.58	0.23	0.26	0.42	0.52	1.56
			2.95	3.78	2.92	3.77					1.16	1.46	1.30	1.59	
			3.86	4.81	3.72	4.66					1.07	1.43	1.26	1.57	
cis-*N*-methylacetamide	3.96	4.66	3.58	4.62	3.06	4.06		diethyl sulfide	1.74	2.66	0.05	0.00	0.38	0.51	1.50
			3.71	4.66	3.20	4.13					1.09	1.44	1.28	1.63	
			4.44	5.48	4.06	5.02					1.02	1.44	1.25	1.63	
trans-*N*-methylacetamide	3.78	4.55	2.85	3.77	2.63	3.59	3.73	methylamine	1.48	2.01	0.47	0.51	0.10	0.10	1.31
			2.95	3.80	3.72	4.92					1.40	1.61	1.21	1.36	
			4.06	5.09	3.98	5.06					1.17	1.26	1.15	1.27	
acetamide	3.81	4.49	3.21	4.14	2.86	3.83	3.76	dimethylamine	1.09	1.63	0.33	0.37	0.10	0.13	1.03
			3.27	4.09	3.68	4.76					1.09	1.31	0.97	1.14	
			4.30	5.25	4.07	5.06					0.86	0.97	0.95	1.09	
2-propanol	1.45	2.28	1.17	1.41	0.98	1.19	1.66	trimethylamine	0.73	1.25	0.18	0.18	0.05	0.09	0.61
			1.57	1.90	1.58	1.91					0.83	1.03	0.56	0.63	
			1.53	1.89	1.51	1.82					0.63	0.72	0.76	0.88	
phenol	1.36	1.97	0.74	0.98	0.53	0.71	1.45	ethylamine	1.41	2.00	0.47	0.54	0.17	0.20	1.22
			1.27	1.69	1.25	1.68					1.36	1.68	1.14	1.32	
			1.24	1.64	1.18	1.56					1.11	1.27	1.10	1.27	
toluene	0.29	0.42	0.30	0.31	0.25	0.28	0.36	*n*-propylamine	1.50	1.97	0.47	0.53	0.19	0.23	1.17
			0.30	0.31	0.25	0.28					1.35	1.65	1.11	1.29	
			0.29	0.31	0.26	0.28					1.09	1.24	1.06	1.24	
2-pentanone	2.70	3.69	2.39	3.19	2.19	2.95		*n*-butylamine	1.47	2.00	0.47	0.54	0.20	0.24	1.44
			2.79	3.71	2.78	3.72					1.35	1.68	1.11	1.31	
			3.70	4.79	3.58	4.65					1.08	1.26	1.07	1.27	
ethylbenzene	0.20	0.54	0.37	0.39	0.35	0.38	0.59	diethylamine	0.94	1.50	0.27	0.53	0.16	0.21	0.92
			0.37	0.39	0.35	0.38					1.00	1.35	0.89	1.13	
			0.35	0.39	0.35	0.38					0.75	0.95	0.86	1.08	
3,5-dimethylpyridine	2.61	3.71	1.19	1.59	0.88	1.17		unsigned max error	0.39		1.50		1.73		
			1.77	2.70	1.89	2.61					0.84		0.76		
			2.45	3.35	2.59	3.58					1.03		0.95		
4-methylpyridine	2.62	3.89	1.31	1.71	0.97	1.26		unsigned av error	0.12		0.72		0.85		
			1.95	2.94	2.00	2.73					0.24		0.19		
			2.57	3.47	2.68	3.67					0.33		0.27		

[a] All entries are in kcal/mol. For geometries used in solvation calculations see footnote *a* in Table 1. PB settings for solution calculation: exterior dielectric constant = 80, interior dielectric constant = 1, grid resolution 1.8 grids/Å, van der Waals radii from ref 26, probe radius 1.4 Å, no salts. [b] Gas phase values (*CRC Handbook of Chemistry and Physics*, 65th ed., and from *Table of Experimental Dipole Moments*, vol. 1−3, by A. L. McClellan). [c] Gas phase values taken from ref 26. [d] Solution phase values calculated with Jaguar 3.5.[100]

Truhlar, and co-workers SM5.2R model (with class II charges[94]) perform slightly better, due to the appropriate tailoring of the nonpolar term, but it can be seen also that our reaction field energies are closer to the ab initio values than those obtained with the SM5.2R model.[94]

The reason Coulson (Mulliken) charges perform so poorly when it comes to evaluating the solvation free energy can be readily understood from the data of Table 3. It is obvious that Coulson based dipole moments are in poor agreement with experimental values and that CM1 or CM2 based dipole moments are in much better agreement (as they were designed to be[28,29]). Our relatively poor performance with S-containing compounds can be also rationalized by inspecting the results of Table 3. The dipole moments calculated in water are much too low when the semiempirical Hamiltonians were used. For example, dimethyl sulfide has an ab initio calculated gas/solution phase dipole of 1.74/2.58 D, while the corresponding semiempirical values are ∼1.1−1.28/1.4−1.6 D (depending on the Hamiltonian used). The experimental gas-phase value is 1.5 D. As the dipole moment gives the leading term in a multipole expansion of the solvation free energy, it becomes apparent that the gross underestimation of dipole moment introduces large errors in solvation free energy. But, we believe that our difficulties with S-containing compounds (Table 1) are mainly due to a crude model for the nonpolar term rather than the simplicity of the semiempirical Hamiltonian. The nonpolar term includes solvent packing effects (cavity formation) and solvent−solute van der Waals interaction. The assumption that these two terms balance off each other as in the case of hydrocarbons[26] and their sum depends linearly on solvent-accessible surface area is a rather crude one. Better design of the nonpolar term is definitely required for better agreement with the experimental quantities, but our solvation free energy results for amino acids derivatives (vide infra) show that this nonpolar term may be appropriate to use in calculation of proteins. Of the four different methods we examined, the best approach is the AM1/CM1 model followed by AM1/CM2 and the two PM3-based approaches. The differences mainly come into play with the neutral

TABLE 4: Calculated (ab Initio and Semiempirical) Solvation Free Energies in Water of *N*-Methyl-*N′*-acetyl Derivatives of the Naturally Occurring Amino Acids[a]

amino acid	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]	amino acid	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]
glycine						methionine					
ab initio[g]				2.6		ab initio				3.2	
GVB/6-31G**	−19.8	3.5	12.4		−13.7	GVB/6-31G**	*k*				
B3LYP/6-31G**	−22.4	4.4	17.2		−15.4	B3LYP/6-31G**	−22.4	3.8	10.5		−15.4
AM1/CM1[h]	−17.5	1.3	−2.6	2.7	−13.5	AM1/CM1	−18.3	1.1	−2.3	3.2	−13.9
AM1/CM2[i]	−20.9	2.1	−3.5		−16.1	AM1/CM2	−20.9	1.7	−3.0		−15.9
PM3/CM1	−18.6	2.3	−3.5		−13.6	PM3/CM1	−19.7	1.9	−3.2		−14.5
PM3/CM2	−18.13	2.1	−3.2		−13.3	PM3/CM2	−18.6	1.7	−2.9		−13.6
SASA[j]					−1.2	SASA					−1.1
alanine						aspartate					
ab initio				2.7		ab initio				2.9	
GVB/6-31G**	−19.0	3.1	11.2		−13.2	GVB/6-31G**					
B3LYP/6-31G**	−23.1	4.5	15.4		−15.9	B3LYP/6-31G**	−74.5	5.8	16.1		−66.0
AM1/CM1	−17.3	1.4	−2.5	2.8	−13.2	AM1/CM1	−77.4	2.3	−4.7	3.0	−72.1
AM1/CM2	−20.8	2.2	−3.4		−15.9	AM1/CM2	−76.7	2.3	−4.3		−71.4
PM3/CM1	−18.8	2.3	−3.4		−13.8	PM3/CM1	−73.9	2.5	−4.2		−68.5
PM3/CM2	−18.1	2.1	−3.1		−13.3	PM3/CM2	−74.2	2.4	−4.2		−68.8
SASA					−0.6	SASA					−8.0
valine						glutamate					
ab initio:				3.0		ab initio				3.0	
GVB/6-31G**	−17.3	3.0	9.4		−11.3	GVB/6-31G**					
B3LYP/6-31G**	−18.7	2.9	11.4		−12.8	B3LYP/6-31G**	−90.6	10.1	21.1		−77.5
AM1/CM1	−16.5	1.2	−2.2	3.0	−12.3	AM1/CM1	−97.1	5.8	−9.7	3.1	−88.2
AM1/CM2	−19.5	1.7	−2.9		−14.6	AM1/CM2	−98.5	6.2	−10.3		−89.2
PM3/CM1	−17.5	1.8	−3.0		−12.6	PM3/CM1	−102.2	5.0	−8.7		−94.0
PM3/CM2	−16.7	1.7	−2.7		−12.0	PM3/CM2	−100.4	4.8	−8.3		−92.6
SASA					−0.1	SASA					−8.2
leucine						asparagine					
ab initio				3.1		ab initio				3.0	
GVB/6-31G**	−18.6	3.1	9.3		−12.4	GVB/6-31G**	−25.7	3.9	10.1		−18.8
B3LYP/6-31G**	−18.7	2.9	11.5		−12.7	B3LYP/6-31G**	−25.9	4.7	14.0		−18.2
AM1/CM1	−16.6	1.1	−2.1	3.2	−12.3	AM1/CM1	−22.9	1.3	−2.7	3.0	−18.6
AM1/CM2	−19.2	1.6	−2.8		−14.4	AM1/CM2	−26.1	2.0	−3.6		−21.1
PM3/CM1	−17.7	1.8	−3.0		−12.7	PM3/CM1	−24.9	2.3	−3.8		−19.5
PM3/CM2	−16.8	1.6	−2.6		−12.0	PM3/CM2	−23.4	2.0	−3.4		−18.3
SASA					0.2	SASA					−7.9
isoleucine						glutamine					
ab initio				3.1		ab initio				3.1	
GVB/6-31G**	−17.1	2.9	9.2		−11.1	GVB/6-31G**					
B3LYP/6-31G**	−19.2	3.1	11.3		−13.0	B3LYP/6-31G**	−34.2	6.6	22.2		−24.5
AM1/CM1	−17.0	1.2	−2.2	3.1	−12.8	AM1/CM1	−27.6	1.9	−3.6	3.2	−22.5
AM1/CM2	−19.6	1.7	−2.9		−14.8	AM1/CM2	−33.7	3.3	−5.3		−27.2
PM3/CM1	−17.6	1.8	−2.9		−12.6	PM3/CM1	−31.0	3.8	−5.8		−24.0
PM3/CM2	−17.0	1.7	−2.7		−12.2	PM3/CM2	−28.8	3.3	−4.9		−22.3
SASA					0.3	SASA					−7.9
serine						lysine					
ab initio				2.8		ab initio				3.3	
GVB/6-31G**	−18.4	2.5	6.3		−13.1	GVB/6-31G**					
B3LYP/6-31G**	−17.4	2.3	8.1		−12.3	B3LYP/6-31G**	−95.0	7.5	18.7		−84.2
AM1/CM1	−17.0	0.8	−1.7	2.8	−13.4	AM1/CM1	−91.0	3.2	−6.4	3.3	−84.5
AM1/CM2	−17.9	1.1	−2.1		−14.0	AM1/CM2	−93.2	3.7	−6.6		−86.2
PM3/CM1	−17.4	1.3	−2.2		−13.3	PM3/CM1	−96.2	3.9	−7.4		−89.0
PM3/CM2	−16.9	1.2	−2.0		−12.9	PM3/CM2	−93.2	3.6	−7.6		−86.3
SASA					−6.4	SASA					−6.1
threonine						arginine					
ab initio				2.9		ab initio				3.3	
GVB/6-31G**	−18.3	2.4	6.0		−13.0	GVB/6-31G**					
B3LYP/6-31G**	−17.3	2.2	8.2		−12.2	B3LYP/6-31G**	−71.5	5.4	10.3		−62.8
AM1/CM1	−16.3	0.8	−1.7	2.9	−12.6	AM1/CM1	−77.3	1.9	−3.8	3.4	−71.2
AM1/CM2	−17.3	1.1	−2.1		−13.3	AM1/CM2	−79.7	2.6	−4.4		−73.7
PM3/CM1	−16.6	1.3	−2.3		−12.4	PM3/CM1	−80.9	3.4	−5.7		−74.0
PM3/CM2	−15.8	1.2	−2.0		−11.7	PM3/CM2	−77.3	2.9	−4.8		−70.1
SASA					−4.6	SASA					−14.0
cysteine						phenylalanine					
ab initio				2.9		ab initio				3.3	
GVB/6-31G**	−18.3	2.7	8.2		−12.7	GVB/6-31G**					
B3LYP/6-31G**	−18.1	2.9	10.4		−12.3	B3LYP/6-31G**	−24.4	4.5	14.0		−16.6
AM1/CM1	−16.6	1.4	−2.5	2.9	−12.3	AM1/CM1	−20.7	1.7	−3.2	3.3	−15.6
AM1/CM2	−21.2	2.6	−3.8		−15.7	AM1/CM2	−25.3	3.1	−4.7		−18.8
PM3/CM1	−21.1	3.2	−4.4		−14.9	PM3/CM1	−22.1	3.0	−4.4		−15.8
PM3/CM2	−18.8	2.6	−3.6		−13.2	PM3/CM2	−20.3	2.5	−3.6		−14.5
SASA					−2.5	SASA					−2.1

TABLE 4: (Continued)

amino acid	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]	amino acid	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]
tyrosine						tryptophane					
ab initio				3.3		ab initio				3.4	
GVB/6-31G**	−26.4	4.0	5.0		−19.4	GVB/6-31G**					
B3LYP/6-31G**	−27.3	3.4	7.2		−20.6	B3LYP/6-31G**	−26.9	3.5	7.0		−20.0
AM1/CM1	−25.1	1.9	−3.5	3.4	−19.8	AM1/CM1	−28.0	2.2	−4.4	3.5	−22.2
AM1/CM2	−27.4	2.4	−4.1		−21.6	AM1/CM2	−30.2	2.7	−4.8		−24.0
PM3/CM1	−25.8	2.7	−4.4		−19.7	PM3/CM1	−29.4	3.2	−5.4		−22.6
PM3/CM2	−24.8	2.5	−4.0		−18.9	PM3/CM2	−26.9	2.8	−4.7		−20.6
SASA					−9.2	SASA					−5.4
proline						max diff[l]					
ab initio				2.8		GVB/6-31G**					2.7[m]
GVB/6-31G**	−17.6	3.4	14.3		−11.4	AM1/CM1					3.9 (10.7)[n]
B3LYP/6-31G**	−19.3	3.8	16.5		−12.7	AM1/CM2					6.8 (11.7)
AM1/CM1	−15.6	1.1	−2.2	2.9	−11.6	PM3/CM1					5.5 (16.6)
AM1/CM2	−17.8	1.7	−2.8		−13.2	PM3/CM2					5.2 (15.0)
PM3/CM1	−15.0	1.7	−2.6		−10.3	average diff[o]					
PM3/CM2	−14.9	1.6	−2.4		−10.4	GVB/6-31G**					0.8[m]
SASA					−0.7	AM1/CM1					1.3 (6.6)[n]
histidine						AM1/CM2					2.1 (7.6)
ab initio				3.1		PM3/CM1					1.5 (8.8)
GVB/6-31G**						PM3/CM2					1.6 (7.1)
B3LYP/6-31G**	−28.3	5.1	12.7		−20.1						
AM1/CM1	−29.7	2.5	−6.1	3.2	−23.9						
AM1/CM2	−33.2	3.1	−5.5		−26.8						
PM3/CM1	−32.6	3.8	−6.3		−25.6						
PM3/CM2	−32.1	3.6	−6.0		−25.3						
SASA					−6.8						

[a] All entries are in kcal/mol. For geometries used in solvation calculations see footnote *a* in Table 1. PB settings: exterior dielectric constant = 80, interior dielectric constant = 1, grid resolution 1.8 grids/Å, van der Waals radii from ref 26, probe radius 1.4 Å, no salts. Semiempirical calculations were done with the AM1 and PM3 Hamiltonians. All energy terms are in kcal/mol. [b] Reaction field energy. [c] Change in solute self-energy due to polarization of electron density under the influence of the polarizable solvent. [d] Solute polarization free energy $\{G_{RF}(\text{solution}-\text{gas}) + G_{wfd}\}$. [e] Cavity plus dispersion−repulsion contribution (nonpolar) calculated as a function of molecular surface. [f] Solvation free energy. [g] Ab initio results obtained with Jaguar 3.5[100] at GVB and DFT level of ab initio theory. [h] CM1 charges.[29] [i] CM2 charges.[28] [j] Solvation free energy calculated based on solvent-accessible surface area.[106] [k] SCF did not converge. [l] Maximum difference between solvation free energies calculated with GVB/AM1/PM3 and DFT. [m] The set contains only the neutral amino acids at pH = 7. [n] The set contains only the charged amino acids at pH = 7. [o] Average difference between solvation free energies calculated with GVB/AM1/PM3 and DFT.

molecules (e.g., AM1/CM1 error of 1.1kcal/mol versus PM3/CM1 error of 1.9 kcal/mol), while the error bar for the charged compounds was essentially identical for all models (differences of a few tenths of a kcal/mol).

Amino Acids. Next we addressed the solubility of the amino acids that serve as the building blocks of proteins. Our concern was whether D&C-PB method is able to give accurate solvation free energies for these critically important compounds. Table 4 shows the solvation free energies for the *N*-acetyl-*N*′-methyl derivatives of the 20 naturally occurring amino acids calculated at the semiempirical level and GVB and DFT (B3LYP) levels (single point at the gas-phase GVB/DFT geometry).[100] The ab initio and semiempirical results (AM1/CM1) are in all cases in excellent agreement with each other (even for sulfur- or hydroxyl-containing amino acids). In fact, the ab initio GVB results (Jaguar 3.5) are larger by about 2 kcal/mol than the values shown in Table 4 because they include the first shell solvation correction.[97] This term in some cases is larger; for example, it makes leucine 4 kcal/mol more soluble than alanine. Considering the minor structural differences between the two amino acids, the calculated difference in the solvation free energy seems to be a bit large in this case. As our D&C−PB calculations do not include a first shell correction term we only considered the GVB results without this term. As far as we are aware there is no experimental solvation free energies for these compounds. A free energy perturbation study published by Kollman and co-workers[102] gives a free energy for the mutation of alanine to leucine in solution of 0.65 ± 0.2 kcal/mol. The GVB result was 0.8 kcal/mol (without the first shell correction) and our calculated semiempirical value was 0.9 kcal/mol.

DNA Bases and Nucleotides. Our next series of tests was performed on DNA basis and nucleotides, the basic blocks in constructing DNA/RNA molecules. As in the case of amino acids, we calculated the DNA bases at GVB and DFT (B3LYP) levels of ab initio theory and with semiempirical AM1/PM3 Hamiltonians and CM1/CM2 charge models. These results were compared with those obtained by York et al.[7] using AM1/COSMO approach and Miller and Kollman's FEP results.[103] Table 5 shows that there is good agreement between the semiempirical and DFT results. PM3 Hamiltonian gives good agreement for cytosine, adenine, and guanine, while AM1 agrees well with DFT for thymine and uracil. AM1/CM1 fails badly for cytosine. It appears that the AM1/CM1 charge model is to be blamed for this failure (5.2 kcal/mol off DFT result). For example, both AM1/CM2 and PM3/CM1 reduce the difference to 1.4 kcal/mol. The average difference in solvation free energies (with respect to DFT results) is the same as for small neutral molecules. AM1/CM2 gives an average difference of 1.1 kcal/mol, while AM1/CM1 gives 2.8 kcal/mol because of its difficulty with cytosine. The FEP and York et al. average differences are 1.0 and 0.7 kcal/mol, respectively. The York et al. solvation free energies are closer to both FEP and DFT results, but at the price of reparametrizing the COSMO method. We calculated also the solvation free energies of the nucleotides and compared the DFT and semiempirical results in Table 5. Again cytosine (cytidine phosphate) is handled with difficulty by AM1/CM1 which gives a solvation free energy larger by 13 kcal/mol, while PM3/CM1 gives the same value as DFT. On the other hand, AM1/CM1 performs well for the other four nucleotides. The average difference between our semiempirical

TABLE 5: Calculated (ab Initio, Semiempirical and FEP) Solvation Free Energies in Water of DNA Bases and Nucleotides[a]

amino acid	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]	amino acid	G_{RF}[b]	G_{wfd}[c]	G_{pol}[d]	G_{np}[e]	G_{sol}[f]
1-methylcytosine						3′-Me-cytidine-PO_3−Me					
ab initio[g]						ab initio/B3LYP/6-31G**	−84.9	8.0	10.1	3.6	−73.3
GVB/6-31G**	−26.3	6.0	7.6	2.5	−17.8	AM1/CM1	−94.3	4.5	−8.4	3.8	−86.0
B3LYP/6-31G**	−27.1	5.8	6.8		−18.8	AM1/CM2	−88.4	4.4	−6.7		−80.3
FEP[o]					−18.4	PM3/CM1	−82.0	4.6	−6.3		−73.6
AM1/Mul[h]					−18.8	PM3/CM2	−76.4	4.3	−6.1		−68.3
AM1/CM1[h]	−31.0	4.5	−8.8	2.4	−24.0	3′-Me-Adenosine-PO_3−Me					
AM1/CM2[i]	−26.7	4.1	−5.6	2.5	−20.2	ab initio/B3LYP/6-31G**	−85.8	7.9	−0.6	3.7	−74.2
PM3/CM1	−24.0	4.1	−5.6		−17.4	AM1/CM1	−85.9	3.5	−6.3	3.9	−78.5
PM3/CM2	−24.9	4.3	−5.8		−18.1	AM1/CM2	−82.9	3.6	−6.2		−75.4
9-methyl adenine						PM3/CM1	−76.7	3.9	−6.0		−68.9
ab initio						PM3/CM2	−71.9	4.0	−5.2		−64.4
GVB/6-31G**	*k*					3′-Me-Tymidine-PO_3−Me					
B3LYP/6-31G**	−20.7	3.1	2.0	2.6	−15.0	ab initio/B3LYP/6-31G**	−78.3	3.6	13.6	3.7	−71.0
FEP					−13.6	AM1/CM1	−75.0	2.1	−3.8	3.8	−69.1
AM1/Mul	−23.1	1.8	−4.3	2.6	−15.5	AM1/CM2	−72.7	2.1	−3.9		−66.9
AM1/CM1	−22.5	1.7	−3.6	2.6	−18.7	PM3/CM1	−69.4	2.3	−3.5		−63.2
AM1/CM2	−21.3	2.2	−4.0		−18.1	PM3/CM2	−63.4	2.0	−3.1		−57.6
PM3/CM1	−21.4	2.1	−4.1		−16.4	3′-Me-Guanosine-PO_3−Me					
PM3/CM2					−16.7	ab initio/B3LYP/6-31G**	−103.7	14.3	−4.3	3.7	−85.7
1-methylthymine						AM1/CM1	−93.9	5.7	−8.9	4.0	−84.2
ab initio						AM1/CM2	−95.2	6.8	−9.6		−84.5
GVB/6-31G**	−18.8	3.0	4.6	2.6	−13.2	PM3/CM1	−94.0	7.6	−9.8		−82.4
B3LYP/6-31G**	−18.0	3.0	5.4		−12.4	PM3/CM2	−86.5	6.9	−9.0		−75.5
FEP					−12.4	3′-Me-Uridine-PO_3−Me					
AM1/Mul				2.5	−13.0	ab initio/B3LYP/6-31G**	−81.9	6.1	17.4	3.6	−72.1
AM1/CM1	−15.5	1.3	−2.5	2.6	−11.5	AM1/CM1	−77.2	2.1	−3.9	3.7	−71.4
AM1/CM2	−17.3	1.7	−2.9		−12.9	AM1/CM2	−75.1	2.1	−4.0		−69.2
PM3/CM1	−14.6	1.6	−2.6		−10.4	PM3/CM1	−72.8	2.5	−3.9		−66.4
PM3/CM2	−14.7	1.7	−2.5		−10.4	PM3/CM2	−66.8	2.2	−3.5		−60.7
9-methylguanine						max diff[l]					
ab initio						GVB/6-31G**[m]					0.9
GVB/6-31G**						AM1/Mul[n]					1.6
B3LYP/6-31G**	−34.4	6.6	16.0	2.7	−25.1	FEP/ESP[o]					2.7
FEP					−22.4	AM1/CM1					5.2 (12.7)[p]
AM1/Mul				2.6	−23.4	AM1/CM2					3.2 (6.9)
AM1/CM1	−27.2	2.7	−5.4	2.7	−21.8	PM3/CM1					2.7 (7.6)
AM1/CM2	−31.6	3.7	−5.8		−25.2	PM3/CM2					3.0 (13.3)
PM3/CM1	−30.5	4.3	−6.6		−23.4	average diff[q]					
PM3/CM2	−28.7	3.9	−6.1		−22.1	GVB/6-31G**					0.4
1-methyluracil						AM1/Mul					0.7
ab initio						FEP/ESP					1.0
GVB/6-31G**	−19.7	3.7	14.1	2.5	−13.5	AM1/CM1					2.8 (4.2)
B3LYP/6-31G**	−19.0	2.8	15.1		−13.7	AM1/CM2					1.1 (3.3)
FEP					−14.0	PM3/CM1					1.8 (4.4)
AM1/Mul				2.4	−14.5	PM3/CM2					2.0 (10.0)
AM1/CM1	−16.5	1.4	−2.7	2.5	−12.6						
AM1/CM2	−18.3	1.8	−3.1		−14.0						
PM3/CM1	−15.2	1.7	−2.7		−11.0						
PM3/CM2	−15.4	1.7	−2.7		−11.2						

[a] All entries are in kcal/mol. For geometries used in solvation calculations see footnote *a* in Table 1. PB settings: exterior dielectric constant = 80, interior dielectric constant = 1, grid resolution 1.8 grids/Å, van der Waals radii from ref 26, probe radius 1.4 Å, no salts. Semiempirical calculations were done with the AM1 and PM3 Hamiltonians. All energy terms are in kcal/mol. [b] Reaction field energy. [c] Change in solute self-energy due to polarization of electron density under the influence of the polarizable solvent. [d] Solute polarization free energy {G_{RF}(solution−gas) + G_{wfd}}. [e] Cavity plus dispersion−repulsion contribution (nonpolar) calculated as a function of molecular surface. [f] Solvation free energy. [g] Ab initio results obtained with Jaguar 3.5[100] at GVB and DFT level of ab initio theory. [h] CM1 charges.[29] [i] CM2 charges.[28] [j] Solvation free energy calculated based on solvent-accessible surface area (see ref 106). [k] SCF did not converge. [l] Maximum difference between solvation free energies calculated with GVB/FEP/AM1/PM3 and DFT. [m] The set does not contain the compounds for which the SCF did not converge (see note *k*). [n] Solvation free energies from ref 7. [o] Solvation free energies (FEP) from ref 103. The set contains only the nucleotides. [q] Average difference between solvation free energies calculated with GVB/FEP/AM1/PM3 and DFT.

results and those obtained from DFT calculations is around 4 kcal/mol which compares well with 5.6 kcal/mol (Table 2) obtained for small charged compounds.

Proteins and DNA. We applied the D&C-PB (AM1/CM1) methodology to the calculation of the solvation free energy for few proteins ranging from crambin (642 atoms) to *Subtilisin E* (3854 atoms) and a piece of DNA with 24 nucleotides (Dickerson's dodecamer [99]). For proteins we used (as for the small molecules) a dielectric constant of 1. For crambin we also performed calculations with a dielectric constant of 2, which has been used with nonpolarizable solutes (i.e., with fixed charges) in order to mimic the electronic polarization. The value of −124.9 kcal/mol obtained for the reaction field energy (without SCRF) should be close to the value obtained with a dielectric constant of one with the SCRF (i.e., −316.7 + 23.4). However, the actual calculations show a large difference between these two values, suggesting that the use of a dielectric constant of two with nonpolarizable solutes overestimates the polarization of the electronic charge density. All calculations were carried out on a SGI Origin 200 workstation with 128MB

TABLE 6. Solvation Free Energies of Proteins and DNA in Water Calculated by D&C-PB Methodology[a]

protein[b]	atoms	residues	total charge	G_{RF}[c]	G_{wdf}[d]	G_{pol}[e]	G_{np}[f]	G_{sol}[g]	CPU SCF[h]	CPU PB[i]
Crambin	642	46	0	−316.7	23.4	−37.1	19.7	−273.5	6276	84
				−124.9[j]				−221.4[k]		
BPTI	892	58	+6	−1336.3	69.7	−108.2	26.6	−1239.8	8991	95
								−1021.5		
CspA	1010	69	0	−1175.5	109.3	−158.7	28.6	−1073.5	6256	64
lysozyme	1960	129	+8	−1936.3	129.3	−181.8	45.3	−1761.7	29544	501
								−1575.0		
Subtilisin E[l]	3854	275	−2	−1856.3	166.8	−241.0	74.8	−1614.7	76126	584
Sendai virus	5978	373	−8	−4338.0	343.2	−503.1	138.2	−3856.7	74878	2064
Dickerson's DNA dodecamer	758	24	−22	−6342.9	107.1	−107.3	30.0	−6205.8	9123.1	223.1

[a] Energy entries are in kcal/mol. The geometries were not optimized, see text for a description of geometries preparation. PB settings: exterior dielectric constant = 80, interior dielectric constant = 1, grid resolution 1.2 grids/Å, van der Waals radii from ref 26, probe radius = 1.4 Å, no salts. All energy quantities are given in kcal/mol. [b] X-ray crystal structures (Protein Databank) with protons added. [c] Reaction field energy calculated in the Poisson−Boltzmann module. [d] Wave function distortion energy is obtained following the polarization of electronic density due to solvation (SCRF). [e] Solute polarization free energy $\{G_{RF}(\text{solution}-\text{gas}) + G_{wfd}\}$. [f] Nonpolar contribution to solvation due to cavity and van der Waals interactions; it is calculated as a function of solute molecular surface {Tannor, 1994 #66}. [g] Solvation free energy ($G_{RF} + G_{wfd}, + G_{np}$). [h] CPU required to obtain SCF (in seconds on SGI Origin 200 workstation). [i] CPU time required to solve Poisson−Boltzmann equation. [j] G_{RF} at the beginning of SCRF (interior dielectric constant = 2). This corresponds to a calculation with fixed charges and dielectric constant of 2 to mimic solute electronic polarization. [k] Second row of solvation free energies are from ref 66. [l] Geometry: MD run (AMBER) 352 ps at 300 K (initial structure taken from PDB Databank).

of RAM. The atomic coordinates were taken from the Brookhaven Protein databank and the H's were attached using AMBER 4.1.[104] Previous quantum mechanical calculations on proteins and DNA in solution have been performed by York et al. [66] and a comparison with their results is given in Table 6. We find that the our D&C-PB (SCRF) method gives solvation free energies larger by about 8−15% over the COSMO/AM1 approach. The differences between the two approaches mostly reflect differences in the methodologies used, but other issues regarding the structure, the location of hydrogens, etc. could also be a factor.

Conclusion

We presented a SCRF-based methodology for calculating the solvation free energies of large molecules using linear-scaling quantum mechanical methods. Our methodology combines the linear scaling D&C algorithm for solving the Schrödinger equation (quantum mechanical representation of the solute) and the widely used Poisson−Boltzmann equation for treating solvation/salts effects (dielectric-continuum representation of the solvent). A critical aspect of this work is the use of the CM1/CM2 class IV charges of Cramer, Truhlar, and co-workers, which allowed us to obtain reasonable agreement between experimental and calculated solvation free energies for both charged and neutral molecules.[28,29] We have shown in this paper that this SCRF method, which uses semiempirical Hamiltonians and CM1/CM2 charges, performs almost as well as the closely related SCRF ab initio GVB [26] and DFT (B3LYP) methods. We used the same nonpolar term as utilized in the ab initio PB model,[26] and the results indicate that taking into account the chemical nature of solute atoms when tailoring this nonpolar term [58] may lead to even better agreement with experiment, as obtained by other groups. We find that the accuracy of this method is comparable to the approach of York et. al.,[7] but the most recent SM5.2R model developed by Cramer, Truhlar, and co-workers[94] which includes a refined nonpolar term performs better even than the ab initio GVB/DFT methods for small neutral molecules, supporting our conclusion that a more accurate evaluation of the nonpolar part of solvation free energy is indeed necessary for better agreement with experimental values. But we found out the nonpolar term that we have used to be good enough for accurate calculation of solvation free energy of amino acid derivatives and nucleotides, and by extension for proteins and DNA. The development of a more elaborate nonpolar term is clearly a way to improve the present results (as well as improving the CM1/CM2 charge models) and work along this lines is underway in our group.

The most significant aspect of this work is the ability to study large biological molecules using a quantum mechanical description of the solute coupled with a continuum representation of the solvent which includes also salt effects through the Debye−Hückel theory (PB equation). Serial CPU times recorded for proteins, as large as 275 residues, and DNA show that it is now practical to carry on calculations on such large systems and even larger. Furthermore, by parallelizing this approach we believe that we will be able to reduce the computational times by up to 25 times for the largest systems studied.[105] We feel that this computational method will become an important tool in investigating thermodynamic and mechanistic aspects of biological phenomena, such as protein folding, enzyme reactions, calculation of pK_a, and binding constants, etc.. In particular, it might be effectively used in drug design, etc., because it allows for quantum mechanical effects like polarization and charge transfer which we found to be quite important in biological systems.[8]

Acknowledgment. We thank the DOE (DE-FGO2-96ER-62270) for supporting this research. We also thank the Pittsburgh Supercomputer Center, the San Diego Supercomputer Center, the National Center for Supercomputer Applications, and the Cornell Theory Center for generous allocations of supercomputer time.

References and Notes

(1) Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995**, *103*, 5674−5678.

(2) Dixon, S. L.; Merz, Jr., K. M. *J. Chem. Phys.* **1996**, *104*, 6643−6649.

(3) Dixon, S. L.; Merz, Jr., K. M. *J. Chem. Phys.* **1997**, *107*, 879−93.

(4) Daniels, A. D.; Millam, J. M.; Scuseria, G. E. *J. Chem. Phys.* **1997**, *107*, 425−31.

(5) Millam, J. M.; Scuseria, G. E. *J. Chem. Phys.* **1997**, *106*, 5569−5577.

(6) Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133−146.

(7) York, D. M.; Lee, T.-S.; Yang, W. *Chem. Phys. Lett.* **1996**, *263*, 297−304.

(8) Nadig, G.; van Zant, L. C.; Dixon, S. L.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1998**, *120*, 5593−5594.
(9) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans 2* **1993**, 799−805.
(10) Roux, B.; Simonson, T. *Biophys. Chem.*, submitted for publication.
(11) Shao, L.; Yu, H.-A.; Gao, J. *J. Phys. Chem.* **1998**, *102*, 10366−10373.
(12) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.
(13) Honig, B.; Nicholls, A. *Science (Wash.)* **1995**, *268*, 1144−1149.
(14) Sharp, K.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, 301.
(15) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins* **1986**, *1*, 47.
(16) Warwicker, J.; Ollis, D.; Richards, F. M.; Steitz, T. A. *J. Mol. Biol.* **1985**, *186*, 645.
(17) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671−679.
(18) Warwicker, J. *J. Theor. Biol.* **1986**, *121*, 199.
(19) Allison, S. A.; Sines, J. J.; A. Wierzbicki *J. Phys. Chem.* **1989**, *93*, 5819−5823.
(20) Davis, M.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509−521.
(21) Holst, M.; Kozack, R. E.; Saied, F.; Subramanian, S. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 231−245.
(22) Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1989**, *10*, 386−391.
(23) Harvey, S. C. *Proteins: Struct. Funct. Genet.* **1989**, *5*, 78−92.
(24) Tomac, S.; Graeslund, A. *J. Comput. Chem.* **1998**, *19*, 893−901.
(25) Besler, B. H.; Merz, K. M. J.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431−439.
(26) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; William A. Goddard, I.; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875−82.
(27) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
(28) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1998**, *102*, 1820−1831.
(29) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87−110.
(30) Honig, B.; Sharp, K. A.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101−1109.
(31) Cortis, C. M.; Friesner, R. A. *J. Comput. Chem.* **1997**, *18*, 1570−1590.
(32) Zauhar, R. J.; Morgan, R. S. *J. Comput. Chem.* **1988**, *9*, 171−187.
(33) Zauhar, R. J.; Morgan, R. S. *J. Mol. Biol.* **1985**, *186*, 115.
(34) Uhlig, H. H. *J. Phys. Chem.* **1937**, *41*, 1215.
(35) Gibson, H. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 420.
(36) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.
(37) Valvani, S. C.; Yalkovsky, S. H.; Amidon, G. L. *J. Phys. Chem.* **1976**, *80*, 829.
(38) Warshel, A. J. *J. Phys. Chem.* **1979**, *83*, 1640.
(39) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(40) Silla, E.; Tunon, I.; Pascual-Ahuir, J. L. *J. Comput. Chem.* **1991**, *12*, 1077.
(41) Cramer, C.; Truhlar, D. G. *J. Comput. Chem.* **1992**, *13*, 1089.
(42) Hylton-McCreery, J.; Christoffersen, R. E.; Hall, G. G. *J. Am. Chem. Soc.* **1976**, *98*, 7191.
(43) Thole, B. T.; van Duijnen, P. T. *Theor. Chim. Acta* **1980**, *55*, 307.
(44) Miertus, S.; Scocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
(45) Rinaldi, D.; Ruiz-Lopez, M. F.; Rivail, J.-L. *J. Chem. Phys.* **1983**, *78*, 834.
(46) Agren, H.; Medina-Llanos, C.; Mikkelsen, K. V. *Chem. Phys.* **1987**, *115*, 43.
(47) Karlstroem, G. *J. Phys. Chem.* **1988**, *92*, 1315.
(48) Wong, M. W.; Frisch, M. J.; Wiberg, K. B. *J. Am. Chem. Soc.* **1991**, *113*, 4776.
(49) Olivares del Valle, F. J.; Aguilar, M. A. *J. Mol. Struct. (THEOCHEM)* **1993**, *280*, 25.
(50) Rinaldi, D.; Rivail, J.-L. *Theor. Chim. Acta* **1973**, *32*, 57.
(51) Rivail, J.-L.; Rinaldi, D. *Chem. Phys.* **1976**, *18*, 233.
(52) Tapia, O.; Goschinski, O. *Mol. Phys.* **1975**, *29*, 1653.
(53) Hoshi, H.; Sakurai, M.; Inoue, Y.; Chujo, R. *J. Chem. Phys.* **1987**, *87*, 1107.
(54) Miertus, S.; Frecer, V.; Majekova, M. *J. Mol. Struct. (THEOCHEM)* **1988**, *189*, 443.
(55) Ruiz-Lopez, M. F.; Rinaldi, D. *J. Mol. Struct. (THEOCHEM)* **1983**, *93*, 277.
(56) Karelson, M. M.; Katritzky, A. R.; Zerner, M. C. *Int. J. Quantum Chem.: Quantum Chem. Symp.* **1986**, *20*, 521.
(57) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305, 9901(E).
(58) Cramer, C.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629.
(59) Wang, B.; Ford, G. P. *J. Chem. Phys*, **1992**, *97*, 4162.
(60) Negre, M. J.; Orozco, M.; Luque, F. J. *Chem. Phys. Lett.* **1992**, *196*, 27.
(61) Ciudinov, G. E.; Napolov, D. V.; Basilevsky, M. V. *Chem. Phys.* **1992**, *160*, 41.
(62) Rauhut, G.; Clark, T.; Steinke, K. *J. Am. Chem. Soc.* **1993**, *115*, 9174.
(63) Pople, J. A.; Beveridge, D. L. *Approximate Molecular Theory*; McGraw-Hill: New York, 1970.
(64) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(65) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *105*, 2744−2750.
(66) York, D. M.; Lee, T.-S.; Yang, W. *J. Am. Chem. Soc.* **1996**, *118*, 10940−10941.
(67) Dixon, S. L.; Merz, K. M., Jr. In *The Divide and Conquer Methodology Applied to Semiempirical SCF MO Calculations*; Dixon, S. L., Merz, K. M., Jr., Eds., in press.
(68) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1987**, *9*, 327−335.
(69) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435−445.
(70) Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1991**, *12*, 909−912.
(71) Luty, B. A.; Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1992**, *13*, 1114−1118.
(72) Holst, M.; Saied, F. *J. Comput. Chem.* **1993**, *14*, 105−113.
(73) Bruccoleri, R. E. *J. Comput. Chem.* **1993**, *14*, 1417−1422.
(74) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(75) Holst, M. J.; Saied, F. *J. Comput. Chem.* **1995**, *16*, 337−364.
(76) Bruccoleri, R. E.; Novotny, J.; Davis, M. E.; Sharp, K. A. *J. Comput. Chem.* **1997**, *18*, 268−276.
(77) Cortis, M. C.; Langlois, J.-M.; Beachy, M. D.; Friesner, R. A. *J. Chem. Phys.* **1996**, *105*, 5472−5484.
(78) Cortis, C. M.; Friesner, R. A. *J. Comput. Chem.* **1997**, *18*, 1591−1608.
(79) Mukherjee, A. In *An Adaptive Finite Element Code for Eliptic Boundary Value Problems in Three Dimensions with Applications in Numerical Relativity*; Mukherjee, A., Ed.; Pennsylvania State University: State College, PA, 1996; p 110.
(80) Reiner, E. S.; Radke, C. J. *J. Chem. Soc., Faraday Trans.* **1990**, *86*, 3901−3912.
(81) Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1990**, *94*, 7684−7692.
(82) Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591−3600.
(83) David, L.; Field, M. J. *J. Comput. Chem.* **1997**, *18*, 343−350.
(84) Rashin, A. A.; Namboodiri, K. *J. Phys. Chem.* **1987**, *91*, 6003−6012.
(85) Yoon, B. J.; Lenhoff, A. M. *J. Comput. Chem.* **1990**, *11*, 1080−1086.
(86) Vorobjev, Y. N.; Grant, J. A.; Scheraga, H. A. *J. Am. Chem. Soc.* **1992**, *114*, 3189−3196.
(87) Purisima, E. O.; Nilar, S. H. *J. Comput. Chem.* **1995**, *16*, 681−689.
(88) Horvath, D.; van Belle, D.; Lippens, G.; Wodak, S. J. *J. Chem. Phys.* **1996**, *104*, 6679−6695.
(89) Zauhar, R. J.; Varnek, A. *J. Comput. Chem.* **1996**, *17*, 864−877.
(90) Vorobjev, Y. N.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 569−583.
(91) Reitz, J. R.; Milford, F. J. *Foundation of Elelctromagnetic Theory*; Addison-Wesley: Reading, MA, 1967.
(92) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
(93) Press: W. H.; Teokolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes. The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: Cambridge, UK, 1992.
(94) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257−3271.
(95) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. *Biochemistry* **1981**, *20*, 849.
(96) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Soln. Chem.* **1981**, *10*, 563.
(97) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775−88.
(98) Vasquez, M.; Nemethy, G.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 1043−1049.
(99) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179.

(100) Jaguar_3.5; Schroedinger Inc.: Portland, OR, 1998.

(101) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210−3221.

(102) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564−568.

(103) Miller, J. L.; Kollman, P. A. *J. Phys. Chem.* **1996**, *100*, 8587.

(104) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 4.1*; University of California: San Francisco, 1993.

(105) Vincent, J. J.; Merz, K. M., Jr. *Theor. Chem. Acc.* **1998**, *99*, 220−223.

(106) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086.