

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/21863334>

AMI-SM2 and PM3-SM3 parametrized SCF solvation models for free energies in aqueous solution

ARTICLE *in* JOURNAL OF COMPUTER-AIDED MOLECULAR DESIGN · JANUARY 1993

Impact Factor: 2.99 · DOI: 10.1007/BF00126219 · Source: PubMed

CITATIONS

278

READS

106

2 AUTHORS:



[Christopher J Cramer](#)

University of Minnesota Twin Cities

532 PUBLICATIONS 23,458 CITATIONS

[SEE PROFILE](#)



[Donald Truhlar](#)

University of Minnesota Twin Cities

1,342 PUBLICATIONS 81,415 CITATIONS

[SEE PROFILE](#)

AM1-SM2 and PM3-SM3 parameterized SCF solvation models for free energies in aqueous solution

Christopher J. Cramer^{a,*} and Donald G. Truhlar^b

^a*U.S. Army Chemical Research Development and Engineering Center, Research Directorate, Physics Division, Chemometric and Biometric Modeling Branch, Aberdeen Proving Ground, MD 21010-5423, U.S.A.*

^b*Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431, U.S.A.*

Received 23 March 1992

Accepted 14 July 1992

Key words: Free energy; Hydration; Molecular modeling; Solubility; Solvent-accessible surface area

SUMMARY

Two new continuum solvation models have been presented recently, and in this paper they are explained and reviewed in detail with further examples. Solvation Model 2 (AM1-SM2) is based on the Austin Model 1 and Solvation Model 3 (PM3-SM3) on the Parameterized Model 3 semiempirical Hamiltonian. In addition to the incorporation of phosphorus parameters, both of these new models address specific deficiencies in the original Solvation Model 1 (AM1-SM1), viz., (1) more accurate account is taken of the hydrophobic effect of hydrocarbons, (2) assignment of heavy-atom surface tensions is based on the presence or absence of bonded hydrogen atoms, and (3) the treatment of specific hydration-shell water molecules is more consistent. The new models offer considerably improved performance compared to AM1-SM1 for neutral molecules and essentially equivalent performance for ions. The solute charges within the Parameterized Model 3 Hamiltonian limit the utility of PM3-SM3 for compounds containing nitrogen and possibly phosphorus. For other systems both AM1-SM2 and PM3-SM3 give realistic results, but AM1-SM2 in general outperforms PM3-SM3. Key features of the models are discussed with respect to alternative approaches.

1. INTRODUCTION

The solvation energy of organic molecules plays a critical role in their reactivity. We have recently presented [1-4] three new general parametrized models and one more specific model for including aqueous solvation effects in the Fock matrix of neglect-of-diatomic-differential-overlap (NDDO) [5,6] molecular orbital theory, treating the solvent as a dielectric continuum. In our first paper [1], we used the Austin Model 1 (AM1) [7] for solute parameters and we developed our own general parameter set, called Solvation Model 1 (SM1), for the solvation terms. The combined pa-

*Present address: Department of Chemistry, University of Minnesota, Minneapolis, MN 55455-0431, U.S.A.

parameter set, AM1-SM1, is quite successful for both free energies of solvation and free energies of reaction for isomeric and acid-base-type equilibria [2].

AM1-SM1 includes cavity, dispersion, solvent structure, and local-field polarization terms. By including polarization effects in the SCF step, we include solvent-induced charge redistribution effects in the solute, i.e., the energy minimization of the solute electronic and geometrical structures converges to the lowest-energy configuration in the presence of solvent rather than to the gas-phase structure. Such effects can be very large for many organic and bio-organic molecules. However, several systematic flaws are apparent in the actual performance of AM1-SM1. The quantitative (positive) free energies of solvation for alkanes are too small, the magnitude of the hydrophobic effect being underestimated. The free energy of solvation for water itself is significantly underestimated, leading to erroneous results when supermolecule approaches involving explicit water are undertaken. Additionally, the SM1 model derives per-atom cavitation and dispersion contributions to the free energy of solvation by consideration of only the atomic number and solvent-accessible surface area for each center. A more specific (i.e., less generally applicable) version of the model, AM1-SM1a [1], was presented that delivers significantly improved quantitative agreement when, *inter alia*, the cavitation and dispersion parameters for a given hydrogen atom depend additionally on the heavy atom to which it is attached. The SM1a approach is proscribed, however, in ambiguous situations, which include essentially all charged systems and also, for example, a neutral transition state for proton transfer between two different kinds of heavy atom or many neutral transition states with changing hybridization at N or O. Originally, neither the SM1 or SM1a parameter set included phosphorus; this limited their application for problems of biological interest.

In two subsequent articles, we described new approaches which address all of the above issues. Moreover, we have not only extended the utility of the AM1 Hamiltonian by development of the general Solvation Model 2 (AM1-SM2) [3], but we have also developed a general Solvation Model 3 (PM3-SM3) [4] based on the widely used semiempirical Parametrized Model 3 (PM3) Hamiltonian of Stewart [8]. In addition to specific improvements in the above-targeted areas, each of the two new models also results in either improved or equivalent quantitative accuracy in comparison to AM1-SM1.

Section 2 reviews the solvation model and its SM1 implementation. Section 3 discusses the theoretical modifications and new parametrizations which lead to AM1-SM2 and PM3-SM3. Section 4 discusses calculational methodology and computer implementation including refinements in the algorithmic implementation that have emerged since our initial paper. Section 5 discusses the performance of the models, paying particular attention to those areas where further improvements would be desirable, and Section 6 offers a brief comparison to previous methods and further discussion. Section 7 provides summarizing comments.

2. THE SOLVATION MODEL AND AM1-SM1

2.1. *Thermodynamics*

The standard-state free energy of a solute in aqueous solution, $G^\circ(\text{aq})$, may be partitioned into two distinct terms:

$$G^\circ(\text{aq}) = G^\circ(\text{g}) + \Delta G_s^\circ \quad (1)$$

where $G^\circ(g)$ is the gas phase standard-state free energy, and ΔG_s° is the free energy of aqueous solvation. For all our work, the standard state for the gas phase is an ideal gas at a concentration of one mole per liter, and the standard state for solution is an ideal solution at a concentration of one mole per liter.

We define a particularly useful partition of ΔG_s° as

$$\Delta G_s^\circ = G_s^\circ - E_{EN}(g) + \Delta G_{vib} + \Delta G_{elec}, \quad (2)$$

where $E_{EN}(g)$ is the gas-phase electronic kinetic and electronic and nuclear coulombic energy. G_s° is the part of the solute aqueous free energy given by

$$G_s^\circ = E_{EN}(aq) + G_p(aq) + G_{CDS}^\circ(aq), \quad (3)$$

$E_{EN}(aq)$ is the sum of the solute electronic kinetic and electronic-nuclear coulombic energies in the presence of solvent, $G_p(aq)$ is the solution polarization free energy, and G_{CDS}° is the sum of the cavitation free energy (which is the energy required to make room in the solvent for the solute), the change in dispersion energy upon dissolution, energetic and entropic effects from structural changes in the solvent, and the sum of the changes in free energy due to changes in PV terms, loss of gas-phase translational and rotational modes, and the creation of libron modes (the three-to-six extra vibrational modes of the solution that replace the gas-phase translation and rotation of the solute). Finally, Δ denotes a change upon dissolution so that

$$\Delta G_{vib} = G_{vib}(aq) - G_{vib}(g) \quad (4)$$

and

$$\Delta G_{elec} = G_{elec}(aq) - G_{elec}(g) \quad (5)$$

where G_{vib} is vibrational free energy, and G_{elec} is electronic free energy. We will later find it convenient to deal with the sum

$$G_{ENP}(aq) = E_{EN}(aq) + G_p(aq). \quad (6)$$

2.2. Molecular modeling

The partition of solvation effects into the G_p term and the G_{CDS}° term in Eq. 3 is the critical physical approximation in all four SM models (SM1, SM1a, SM2, and SM3) reviewed in the present paper. We include electric polarization of the solvent dielectric in G_p , and then we assume that the dominant remaining effects are all approximately proportional to the area of a surface passing through the first hydration shell, but with, however, a different constant of proportionality for different parts of the hydration shell, depending on which part of the solute it is solvating. We next consider the G_p and G_{CDS}° terms in turn.

2.2.1. Electric polarization

The very successful Born model [9] for a monatomic ion immersed in a solvent expresses the po-

larization free energy as

$$G_P = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \frac{q^2}{\alpha} \quad (7)$$

where ϵ is the solvent dielectric constant (relative permittivity), q is the net atomic charge, and α is some effective ionic radius. This expression has been generalized for multicentered situations [10–15] to

$$G_P = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{k,k'} q_k q_{k'} \gamma_{kk'} \quad (8)$$

where k labels an atomic center, and $\gamma_{kk'}$ is a coulomb integral. Still et al. [16] suggested a very useful semiempirical form for the coulomb integrals, given by

$$\gamma_{kk'} = \{r_{kk'}^{-2} + \alpha_k \alpha_{k'} C_{kk'}(r_{kk'})\}^{-\frac{1}{2}} \quad (9)$$

where $r_{kk'}$ is the interatomic distance between atoms k and k' , $C_{kk'}$ is equal to

$$C_{kk'}^{(0)} = \exp(-r_{kk'}^{-2}/d^{(0)}\alpha_k \alpha_{k'}) \quad (10)$$

and $d^{(0)}$ is an empirically optimized constant equal to 4.0. We use a modified $C_{kk'}$ given by Eq. 1

$$C_{kk'} = C_{kk'}^{(0)}(r_{kk'}) + C_{kk'}^{(1)}(r_{kk'}) \quad (11)$$

where $C_{kk'}^{(1)}$ is a localized function given by

$$C_{kk'}^{(1)} = \begin{cases} d_{kk'}^{(1)} \exp(-d_{kk'}^{(2)}/\{1 - [(r_{kk'} - r_{kk'}^{(1)})/r_{kk'}^{(2)}]^2\}), & |r_{kk'} - r_{kk'}^{(1)}| < r_{kk'}^{(2)}, \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

We make $d_{kk'}^{(1)}$ non-zero only for selected atom pairs that exhibit systematic deficiencies for compounds with interatomic distances in a localized range. In all work so far (SM1, SM1a, SM2, SM3), we have included $C_{kk'}^{(1)}$ only for O–O and N–H interactions with parameters such that the localized function affects only bonded and geminal O–O pairs and vicinal N–H interactions. The parameters have not changed since SM1 and are given in our original paper [1].

For a single atom ($k = k' = 1$), it is apparent that Eq. 7 is a special case of Eq. 8 and α_k is then chosen to be equal to an intrinsic coulomb radius, ρ_k , where

$$\rho_k = \rho_k^{(0)} + \rho_k^{(1)} \left[-\frac{1}{\pi} \arctan \frac{q_k + q_k^{(0)}}{q_k^{(1)}} + \frac{1}{2} \right]. \quad (13)$$

The coulomb radius is thus a smooth, sigmoidal function of the atomic charge, q_k , which runs from an empirically derived limit of $\rho_k^{(0)}$ for large positive q_k to $\rho_k^{(0)} + \rho_k^{(1)}$ for large negative q_k .

The parameters $q_k^{(0)}$ and $q_k^{(1)}$ determine the location and the steepness of the inflection point for the function [1]. We have taken $\rho_k^{(0)}$, $\rho_k^{(1)}$, and $q_k^{(0)}$ as empirical parameters fitted to aqueous free energies of solvation. However, $q_k^{(1)}$ has been fixed at 0.1 in all models so far. The reason for this is that the charges found on a given kind of atom in representative neutrals and ions do not span a range uniformly. Instead, they tend to cluster into fairly consistent anionic, neutral, or cationic subranges, so it is not possible to optimize $q_k^{(1)}$ very precisely. We have found $q_k^{(1)} = 0.1$ is a rough optimum for SM1, SM2, and SM3.

In the multicenter case, α_k is determined numerically, again following a procedure introduced by Still et al. [16]. Thus, operating under the assumption that the remainder of the molecule functions to displace the surrounding dielectric medium, concentric spherical shells are expanded about atom k and their contribution to the polarization free energy, based on enclosed dielectric volume, is calculated. The effective radius, α_k , is then chosen so that the G_p derived from Eq. 7, i.e., as if a monatomic case, is equal to the G_p determined via numerical integration. Thus we consider M spherical shells, i , around each atom, k , and calculate

$$\alpha_k^{-1} = \sum_{i=1}^M \frac{A_i(r_i; \{\rho_k\})}{4\pi r_i^2} \left(\frac{1}{r_i - 0.5T_i} - \frac{1}{r_i + 0.5T_i} \right) + \frac{1}{r_{M+1} - 0.5T_{M+1}} \quad (14)$$

where r_i and T_i are defined recursively by

$$r_i = \begin{cases} \rho_k + \frac{1}{2}T_1, & i=1 \\ r_{i-1} + \frac{1}{2}(T_{i-1} + T_i), & i>1 \end{cases} \quad (15)$$

$$T_i = \begin{cases} T_1, & i=1 \\ (1+F)T_{i-1}, & i>1 \end{cases} \quad (16)$$

and $A_i(r_i; \{\rho_k\})$ is an ‘exposed shell area’, i.e., the numerically determined exposed surface area of the sphere of radius r_i , where the exposed area is that area *not* included in any spheres centered around other atoms when those spheres have radii given by the set $\{\rho_k\}$. The summation limit, M , is reached when the sphere of radius $r_i - 0.5T_i$ encompasses the entire molecule (for the monatomic case, $M = 0$; in that case only the term outside the summation is used). (Note that Eqs. 14 and 15 suffered from typographical errors in our original paper [1], but these errors are corrected in an erratum and here.) The expansion factor, F , and the initial shell thickness, T_1 , in Eq. 16 are user-adjustable and are discussed in greater detail in Section 4.

This dielectric screening algorithm of Still et al. allows for arbitrary shaped solutes. Some earlier continuum dielectric treatments required approximating the solute shape by a sphere [17] or ellipsoid [18], and this becomes increasingly unrealistic for larger molecules.

Returning to Eq. 3, by using the SCF formalism [6,12], we may now express the E , N , and P terms as

$$G_{ENP} = \frac{1}{2} \sum_{\mu\nu} P_{\mu\nu} (H_{\mu\nu} + F_{\mu\nu}^{(0)}) + \frac{1}{2} \sum_{k,k' \neq k} \frac{Z_k Z_{k'}}{r_{kk'}} - \frac{1}{2} \left(1 - \frac{1}{\varepsilon} \right) \sum_{k,k'} q_k q_{k'} \gamma_{kk'} \quad (17)$$

where P , H , and $F^{(0)}$ are respectively the density, one-electron, and Fock matrices in the absence of solvent terms, μ and ν run over valence atomic orbitals, and Z_k is the valence nuclear charge of atom k (equal to the nuclear charge minus the number of core electrons). We obtain the partial charges by Mulliken population analysis [19]:

$$q_k = Z_k - \sum_{\mu \in k} P_{\mu\mu}. \quad (18)$$

By defining a complete Fock matrix as $F_{\mu\nu} = \partial G_{\text{ENP}} / \partial P_{\mu\nu}$ with fixed $\gamma_{kk'}$, one obtains [12]

$$G_{\text{ENP}} = \frac{1}{2} \sum_{\mu\nu} P_{\mu\nu} (H_{\mu\nu} + F_{\mu\nu}) + \frac{1}{2} \sum_{k,k'=k} \frac{Z_k Z_{k'}}{r_{kk'}} \quad (19)$$

where

$$F_{\mu\nu} = F_{\mu\nu}^{(0)} + \delta_{\mu\nu} \left(1 - \frac{1}{\varepsilon}\right) \sum_{k'} (Z_{k'} - \sum_{\mu' \in k'} P_{\mu'\mu'}) \gamma_{kk'} \quad \mu \in k \quad (20)$$

and $\delta_{\mu\nu}$ is the Kronecker delta function. As a consequence of the Kronecker delta, only the diagonal elements of the solute Fock matrix are modified. The key point to be emphasized here is that by explicitly including the polarization effects in the Fock operator, the resulting density matrix and converged orbitals are determined self-consistently in the presence of solvent [11,12,14,18,20]. Allowing the molecular charge distribution to relax in the field of the solvent can contribute significantly to the overall G_s° (vide infra).

It may be useful to mention that the Mulliken population analysis embodied in Eqs. 18 and 20 has been widely criticized, in hundreds of papers. But its errors will be systematic when applied within a given semiempirical solute parameterization, such as AM1 or PM3, and thus the semiempirical elements of our general parameterization take these into account, just as they also compensate for errors in the approximate wave function itself (independent of how partial charges are extracted from it) and other deficiencies of the formal underpinning of the theory.

2.2.2. Hydration-shell effects

We make the assumption that the contributions to the experimental ΔG_s° from ΔG_{vib} and ΔG_{elec} may be neglected in obtaining our parameters. Then the remaining contribution to the free energy of solvation beyond $G_{\text{ENP}}(\text{aq})$ may be calculated from

$$G_{\text{CDS}}^\circ \cong \Delta G_s^\circ - G_{\text{ENP}} + E_{\text{EN}} \quad (21)$$

where ΔG_s° is now the experimental free energy of solvation. A previously employed approximation [16,21,22] is to make all or part of the free energy of solvation proportional to the solute surface area or to its solvent-accessible surface area; the latter choice is more physical because it is proportional, in a continuum model of the solvent, to the number of solvent molecules in the first hydration shell. We employ this kind of strategy by fitting the part of the free energy of solution

given by (21) in just such a way; thus

$$G_{\text{CDS}}^{\circ} = \sum_{k'} \sigma_{k'} A_{k'}(\beta_{k'}, \{\beta_k\}) \quad (22)$$

where $\sigma_{k'}$ is an interfacial atomic surface tension term, and $A_{k'}(\beta_{k'}, \{\beta_k\})$ is the solvent-accessible surface area for atom k' . The latter is defined as the exposed surface area of atom k' , which equals the exposed surface area of the atom-centered sphere with radius

$$\beta_{k'} = R_{k'} + R_s \quad (23)$$

where $R_{k'}$ is the van der Waals radius of atom k' , and R_s is the solvent radius, taken for water as 1.4 Å. The exposed area is defined in this step as the area that is *not* contained in any of the other atomic spheres when they also have the radii given by Eq. 23; this is why $A_{k'}$ depends on the full set of $\{\beta_k\}$. We note that G_{CDS}° was called G_{CD}° in our first paper [1] to emphasize the cavity and dispersion contributions; we have added the S to emphasize that it also includes the effect of changing the structure of the water, in particular those structural changes that go beyond electric polarization. These structural changes may be more profound than the mere creation of a cavity. The G_{CDS}° term also accounts for the effect of dielectric saturation in the nearby solvent molecules.

Further justification of Eq. 22 is discussed, with references, in our first paper [1]. We wish to stress that the solvent-accessible surface area corresponds physically to a boundary that passes through the centers of the solvent molecules in the first hydration shell. However, the effect of the solute on the solvent is not precisely confined to this shell, even for non-polar hydrocarbons [23], and the shell itself is not uniquely defined. A more general term for the significantly affected solvent molecules is the cosphere [24]. Using either the solute radius or the surface tension (i.e., the free energy per unit area associated with the local solvent accessible surface area of the solute), a semiempirical parameter allows the parameterized model to overcome the ambiguity in this concept as much as possible.

By fitting to a data set which encompasses both ionic and neutral solutes, we were able within our SM1 model largely to separate the optimization of those parameters contributing to the G_{ENP}° term ($\rho_k^{(0)}$, $\rho_k^{(1)}$, and $q_k^{(0)}$, and the $C_{kk'}^{(1)}$ terms) from σ_k , which contributes to the G_{CDS}° term. This is because the ENP term heavily dominates the ionic free energies of solvation, as might be expected.

Simultaneous optimization of $\rho_k^{(0)}$, $\rho_k^{(1)}$, $q_k^{(0)}$ and σ_k for $k = \text{H, C, N, O, F, S, Cl, Br, and I}$, and of the $C_{kk'}^{(1)}$ parameters for $k-k' = \text{O-O and N-H}$ to fit a large body of experimental data [25,26] delivered the SM1 model. The parameters are given in Tables 1–4, and selected results obtained with the SM1 model are given in later tables.

3. IMPROVING THE MODEL

Within the experimental data for ions, fairly large errors are implicit since the experimental ΔG_s° values must be calculated by reference to thermodynamic cycles (vide infra). These involve often difficult to measure gas-phase proton affinities or deprotonation enthalpies, assumed entropies, and aqueous pK_a values [26]. Indeed, for many of these cases, it is probable that AM1-SM1 per-

forms accurately to within experimental error. Furthermore, the parametrization in the G_{ENP} term of AM1-SM1 was more flexible than that for G_{CDS}° , with three parameters per atom and two special pair corrections for the former but only one parameter per atom for the latter. We therefore focused our attention on improving the quantitative accuracy of G_{CDS}° in our improved models.

3.1. AM1-SM1a

The first improvement [1] was called SM1a. In this model, the surface tensions for H, N, and O are different for different chemical environments, e.g., the surface tension for H depends on which atom it is bonded to, and the surface tension of O depends on its hybridization. Although we classify this model here as an improvement, we note that there is a considerable price to pay, namely that the model is not completely unambiguous. For example, along a reaction path an H may be partially bonded to two atoms at once, and an O may have intermediate hybridization. Similar ambiguities occur in many molecules with multiple forms, such as partial delocalization of double bonds (e.g., amides), and in some ions and strained neutrals. The parameters are given in Tables 1, 2, and 4, and selected results are given in later tables.

3.2. AM1-SM2

In AM1-SM2 we again attempted to improve G_{CDS}° , but now in a more automatic, unambiguous way. To do this we added two new parameters per atom for C and S and six new parameters each for N and O.

As mentioned in Sections 1 and 3.1, we observed improved accuracy in the specific solvent model SM1a when hydrogenic surface tensions were assigned based on the non-hydrogenic atom to which the proton was attached [1]. As implemented in the computer program, this datum is supplied by the user as part of the input file. However, by deriving the bond-order matrix from our self-consistently determined solute-density matrix, the details of hydrogenic attachment may be calculated from the SCF wave function without the ambiguity of human classification. Fur-

TABLE I
ATOMIC RADII USED IN THE CALCULATION OF SOLVENT-ACCESSIBLE SURFACE AREA

Atom	$\beta_k, \text{\AA}$		
	SM1, SM1a	SM2	SM3
H	2.60	0.00	0.00
C	3.10	3.10	3.10
N	2.95	3.30	3.30
O	2.92	3.20	3.20
F	2.87	2.80	2.80
P	3.20	3.50	3.60
S	3.20	3.20	3.20
Cl	3.15	3.40	3.40
Br	3.25	3.40	3.40
I	3.38	3.40	3.40

TABLE 2
SOLVENT-ACCESSIBLE SURFACE TENSIONS BY ATOM TYPE

Atom	$\sigma_k^{(0)}$, cal mol ⁻¹ Å ⁻²				$\sigma_k^{(1)}$, cal mol ⁻¹ Å ⁻²	
	SM1	SM1a	SM2	SM3	SM2	SM3
H	0.00	4.15 ^a 58.96 ^b -23.39 ^c 3.95 ^d 49.49 ^e	0	0	0	0
C	14.95	4.15	3.36	8.56	3.24	-0.39
N	-73.65	-368.97 ^f -47.38 ^g	-30.70	-42.20	-19.20	-24.25
O	-52.74	-109.70 ^f -25.61 ^g	-25.01	-34.76	-37.28	-22.82
F	18.47	21.17	22.50	20.02	0	0
P	10.67	3.95	3.90	9.79	0	0
S	-18.80	-44.25	-53.25	-53.00	37.03	36.40
Cl	-2.14	-2.84	-2.28	-3.26	0	0
Br	-9.11	-8.93	-8.15	-4.61	0	0
I	-8.21	-13.42	-15.18	-2.43	0	0

^a H-C, constrained to be same as C.

^b H-N.

^c H-O.

^d H-P, constrained to be same as P.

^e H-S.

^f sp³(N, O), amide (N).

^g sp²(N, O), sp (N), aromatic (N).

thermore, we can consider a continuum of bond orders, as occur along a reaction path, rather than just a few discrete bonding prototypes. Thus we define B_{kH} as the sum of the bond orders of atom k to all hydrogen atoms in the solute. Using the definition of bond order used by Armstrong et al. [27], one obtains

$$B_{kH} = \sum_{\mu \in k, v \in H} P_{\mu v}^2 \quad (24)$$

where μ runs over the valence orbitals of atom k, and v runs over all hydrogen 1s orbitals. We note in passing that this definition of bond order works well, with a typical X-H single bond in AM1 having a bond order ranging from about 0.91 for a typical C-H bond to 0.99 for N-H, O-H and S-H. If there is not a 'classically' bonded hydrogen, the associated bond order rarely exceeds 0.05. Of course, fractional bond orders are expected to occur in transition state modeling, and the capability to treat systems with intermediate bond orders is one of the motivations for making the parameters' functions of a continuous bond order measure.

Given such a connection between hydrogen atoms and the non-hydrogenic atoms to which they are attached, we have elected to employ the derived information in a manner similar in spirit to

the united-atom-type approaches [16,28] commonly employed in molecular mechanics. We have thus defined the hydrogen atom to have no solvent-accessible surface area and further, we have defined it not to block the solvent-accessible surface area of the underlying non-hydrogenic atom, i.e., $\beta_H = 0$. Thus methane, for example, for purposes of solvent-accessible surface area *only*, would be regarded as a spherical carbon atom with some additional information known regarding the C-H bond order. We use this information to calculate the heavy atom's overall surface tension; in particular, we have modified Eq. 22 to read

$$G_{\text{CDS}}^{\circ} = \sum_{k'} \{ \sigma_k^{(0)} + \sigma_k^{(1)} [f(B_{kH}) + g(B_{kH})] \} A_k(\beta_{k'}, \{\beta_k\}) \quad (25)$$

where k and k' now run over non-hydrogenic atoms only, and f and g are functions of B_{kH} . Since hydrogen atoms are defined to have zero volume, surface area expansion alone was observed to respond adequately to hydrogenic substitution beyond the first proton in most cases. Thus, a continuously differentiable function, $f(B_{kH})$, which returns zero for $B_{kH} = 0$ and a value close to one for $B_{kH} = 1$ and which rapidly damps for $B_{kH} > 1$, was chosen, i.e.,

$$f(B_{kH}) = \tan^{-1} (\sqrt{3} B_{kH}). \quad (26)$$

Note that f equals 1.05, 1.29, 1.38, and 1.43 for $B_{kH} = 1, 2, 3$ and 4 , respectively. Note also that f is roughly linear for $B_{kH} \lesssim 1$, but it resembles $(B_{kH})^{1/4}$ for $1 \lesssim B_{kH} \lesssim 4$. Table 2 presents the new $\sigma_k^{(0)}$ and $\sigma_k^{(1)}$ values for SM2 and compares them to the older SM1 values. Additionally, since we are using an approach based on the heavy atoms unblocked by hydrogens, we have modified the atom-dependent solvent radii used for the calculation of solvent-accessible surface area, specifically replacing the van der Waals contribution to $R_k + R_S$ with an empirically optimized radius. The new data are presented in Table 1.

While this scheme worked well in most instances, a few special cases were observed. Nitrogen atoms substituted with three (e.g., ammonia and mono-substituted ammonium ions) and four (the ammonium cation) protons gave $-\Delta G_s^{\circ}$ values which were systematically too large. So also did oxygen atoms substituted with two (e.g., water and mono-substituted oxonium ions) and three (the hydronium cation) protons. We have therefore defined

$$g(B_{kH}) = \begin{cases} a_k \exp\{-b_k/[1 - [(B_{kH} - c_k)/w_k]^2]\}, & |B_{kH} - c_k| < w_k \\ 0, & \text{otherwise.} \end{cases} \quad (27)$$

The constants a_k , b_k , c_k and w_k are listed in Table 3. The function $g(B_{kH})$ ameliorates the negative contribution from the very sizable solvent-accessible surface areas of the heteroatoms in the above-listed compounds under our formalism.

We emphasize that calculating the surface tension on the basis of the underlying non-hydrogenic atoms in the surface groups is *not* a calculational simplification designed to reduce the number of degrees of freedom, as united-atom models often are, but rather a critical computational strategy that, as we found empirically, allows us to put chemical group effects into the formalism in a realistic way. The approach recognizes that although solute surfaces are often dominated by

TABLE 3
EMPIRICALLY DERIVED CONSTANTS FOR $g(B_{kH})$

Model	Atom (k)	a	b	c	w
SM2	N	-9.12	1.35	3.69	1.5
	O	-1.94	0.46	2.56	1.0
SM3	N	-4.93	1.00	3.00	0.5
	O	-2.62	0.43	2.66	1.2

hydrogens, these solvation properties are quite variable, depending strongly on the nature of the chemical group to which the hydrogens are attached. Our treatment of the CDS term provides a particularly simple way to include such chemical effects. For example, the AM1-SM2 method treats the effect of water around an -NH₂ group in terms of a single surface tension and solvent-accessible surface area for the whole group rather than artificially partitioning the group into solvent-accessible N and solvent-accessible H. Additional chemical modeling is contained in the dependence of the surface tension on the sum of the N-H bond orders. If, in contrast, we attempted to give a surface tension to H directly, we would have to adopt a scheme to distinguish H atoms in hydrophobic -CH₂- groups from H atoms in hydrophilic -NH₂ groups. Although this was attempted in preliminary calculations, we were more satisfied that we had captured the correct chemistry and physics of the solvation process with the present approach. The quantitative aspects of the hydrophobic solvation free energies of hydrocarbons are particularly well represented, as discussed in Section 5.1.

In principle, since G_{CDS}⁰ now depends on bond orders, it should be included in the Fock operator. However, in practice we have found that it can be added on at the end of the SCF cycles for the cases we have considered and the level of precision we sought. The physical significance is that the variation of the CDS term with respect to possible converged wave functions is relatively small as compared to that of other terms in the Hamiltonian. To avoid misinterpretation, we note that there are two kinds of solvent restructuring in the models under review here; the electric polarization of the solvent (which for water is dominated by dipole-orientation polarization and involves several hydration shells) and the more specific effects that are primarily localized in the first hydration shell and that are treated by the CDS term. The restructuring of the solvent by the presence of the solute is sensitive to the SCF convergence, but the SCF-sensitive part is contained in the electric polarization term.

Finally, although the ENP parameters and the CDS parameters are only weakly coupled, the changes in the SM2 CDS model were significant enough to warrant simultaneous reoptimization of $\rho_k^{(0)}$, $\rho_k^{(1)}$, and $q_k^{(0)}$ as well. (The $C_{kk}^{(1)}$ parameters and $q_k^{(1)}$ were unchanged.) During the least-squares optimization on neutrals we gave octane a weight of 64 and all other compounds were weighted unity. The final parameters are presented in Tables 1-4, and selected results are presented in Tables 5-8. In particular, Tables 5 and 6 contain the solutes used for the parameterization. The experimental values used were selected from previous compilations [25]. The selection process represents an (unsystematic but thoughtful) attempt to produce a 'training set' that includes reasonable coverage of important organic functional groups and a few small molecules.

3.3. PM3-SM3

The utilities and relative utilities of the semiempirical models, AM1 and PM3, have been discussed at length in the literature [8,29–32], and we will not review the arguments here. It suffices to note that each enjoys considerable application in current research. Moreover, AM1 and PM3 differ markedly in their treatment of hydrogen bonding. The former method tends to predict significantly non-linear (sometimes called ‘three-center’ or ‘bifurcated’) hydrogen bonds [33], while the latter favors linear or nearly linear hydrogen bonds [8,30–32]. It is clearly of value to be able to compare the two when explicit water molecules are included in a calculation. Thus, for use with problems when PM3 might be preferred, we optimized a third solvation model based upon PM3. The computational formalism is entirely equivalent to that for AM1-SM2; the parameters themselves and the use of the PM3 gas-phase Hamiltonian are the only differences. Again octane was weighted 64 in the final least-squares step. The final parameters are presented in Tables 1–4, and selected results are presented in Tables 5–8.

3.4. Phosphorus

Both AM1 and PM3 are parametrized for phosphorus. However, there are few data for the aqueous free energies of solvation for phosphorus containing molecules. For neutrals, data are available only for phosphine, PH₃, [34] and three phosphite esters [35]. Pearson has provided estimated data for a handful of other compounds [26], including the phosphide anion and the trimethylphosphonium cation. Given the great importance of phosphorus in many systems of biological interest, we have developed parameter sets for it within our solvation models by optimizing it for the few available data with all other atomic parameters fixed. The parameters are given in Tables 1, 2, and 4. Note that the parameters for SM1 and SM1a are given for the first time in this paper; those for SM2 are from Ref. [3], and those for SM3 from Ref. [4]. The phosphite esters have little effect on the parameters, and, for the same reasons, there is little we can do to fit them better. Our results vary more than the experimental ones with alkyl chain lengthening because the

TABLE 4
OPTIMIZED ENP PARAMETERS BY ATOM TYPE

Atom	$\rho_k^{(0)}, \text{Å}$			$\rho_k^{(1)}, \text{Å}$			$q_k^{(0)}$		
	SM1, SM1a	SM2	SM3	SM1, SM1a	SM2	SM3	SM1, SM1a	SM2	SM3
H	0.57	0.59	0.59	1.303	1.283	1.289	-0.30	-0.30	-0.10
C	1.68	1.68	1.58	0.000	0.000	0.200	a	a	0.20
N	1.40	1.50	1.60	0.620	0.420	0.320	0.40	0.60	0.50
O	1.46	1.46	1.66	-0.250	-0.150	-0.200	0.75	0.75	0.25
F	1.37	1.37	1.37	0.181	0.145	0.149	0.70	0.70	0.70
P	1.30	1.30	1.40	1.000	1.000	0.800	0.00	0.00	-1.50
S	1.30	1.30	1.20	0.800	0.800	0.900	0.70	0.70	0.70
Cl	1.65	1.65	1.65	0.618	0.555	0.559	0.75	0.75	0.75
Br	1.75	1.75	1.75	0.705	0.629	0.610	0.70	0.70	0.70
I	1.88	1.88	1.88	0.932	0.885	0.798	0.60	0.60	0.60

^a Not required.

added methylenes are hydrophobic and shield the ester oxygens in both a dielectric and an accessible-surface area sense.

We note that β_P is larger than β_N , β_O , and β_{Cl} by about the same amount as for the respective van der Waals radii [36], and this is encouraging. Nevertheless, extra caution is advised in applications involving phosphorus because of the limited experimental data that were available to calibrate and test the model. Further comment is provided below.

4. CALCULATIONAL METHODOLOGY

Solvation Models 1 and 1a are contained in AMSOL-version 1.0 [37], based on AMPAC-version 2.1 [38]. Solvation Models 1, 1a, 2 and 3 are all contained in AMSOL-version 3.0.1 [39], which incorporates a greatly portabilized version of AMPAC-version 2.1.

The computationally demanding step that needed to be added to AMPAC to create AMSOL is the calculation of the exposed surface area. This proceeds as follows. First, the program places a great circle, of radius r_0 , on each atomic sphere and places K_0 uniformly spaced dots on that circle. It then draws $K_0/2$ parallel circles uniformly spaced (same spacing by choice of K_0 as the spacing between points) from pole to pole on the sphere and places K_n points on the smaller circles, where

$$K_n = r_n/r_0. \quad (28)$$

and r_n is the radius of the small circle. This yields the same distribution on the small circles as the large circle.

For the ENP term, we use 50 points on the great circle and 25 parallel circles. This was empirically optimized to give a good combination of speed and consistency. For $K_0 < 50$, the random errors due to finite K_0 begin to affect SCF convergence; for $K_0 > 50$ the computer time, of course, goes up. Since the CDS part of the calculation is much smaller, we can afford a greater point density, and we use 90 points and 45 circles. These choices lead to 816 points for the ENP calculation and 2610 for the CDS calculation.

We next address both the choice of T_1 and F in Eqs. 14–16. The question of the convergence of the numerical integration for $M > 0$ naturally arises, and we carefully examined this point. The numerically determined G_P is largely insensitive to the choice of T_1 , unless unreasonably large values are chosen. The time required for the calculation is similarly little affected. Exactly the opposite behavior is observed for the expansion factor, F . As F increases from 0 (i.e., constant shell size), the numerically calculated G_P decreases in magnitude. For neutral molecules, the differences between converged values and those obtained with an F of 0.5 rarely exceed 0.3 kcal for the entire molecule, and indeed they are usually considerably smaller for all but the most polar molecules. For molecules bearing a unit charge, the same convergence error is on the order of 1–4 kcal. These errors, being regular in terms of underestimation and dependence on molecular size, are largely corrected for by the semiempirical determination of the parameter ρ_k . Thus, finer integrations would tend to yield larger G_P values and the various ρ_k would all increase to return the methodology to internal consistency.

However, more accurate integration requires geometrically increasing computational resources. As a benchmark, reducing the expansion factor, F , from 0.5 to 0.2 decreases the conver-

gence error by roughly one-half but also approximately doubles the required computer time. Finally, we adopted standard values of $T_1 = 0.1 \text{ \AA}$ and $F = 0.5$ for SM1 and SM1a, and $T_1 = 0.02 \text{ \AA}$ and $F = 0.5$ for SM2 and SM3; these values appear to represent a good compromise between speed and accuracy. AMSOL-version 3.0.1 [38] has the option of changing these values for users who wish to assess instances where convergence errors might be nonuniform or to speed up larger calculations. However, we must emphasize that changing these parameters changes the method just as changing any other parameters does. In other words, the final values of the other parameters are chosen to compensate for the systematic differences of the present values of numerical integrals from converged values, and thus these numerical parameters must be used in combination with the choices of numerical parameters used during the parameterization to maintain consistency.

The solvent-accessible surface areas, A_k , depend on geometry, and in SM2 and SM3 some of them depend on the wave function through the bond orders. The exposed shell areas of Eq. 14 also depend on the geometry, the intrinsic coulomb radii depend on the wave function through the partial charges, and the effective coulomb radii depend on the geometry because of dielectric screening. Thus the SM terms must be updated through the SCF cycles and geometry optimizations. As mentioned above, the current code assumes that the CDS parameters do not affect the SCF cycles and adds these terms at the end, but that is a matter of numerical implementation which is acceptable only because of the weak dependence of the final energy on small changes in the bond order.

For a frozen geometry (AMPAC/AMSOL keyword 1SCF), the generalized Born parameters are calculated at the first iteration and updated at the 4th, 8th, 12th, etc. For a geometry optimization, the SCF is updated at iterations 1, 4, 9, 16, However, if the SCF calculation fails to converge, the Fock matrix is updated twice in a row, but only half as often (i.e., for a 1SCF update at iterations 8, 9, 16, 17, 24, 25, etc.). If the SCF process still fails to converge, we update thrice in a row, a third as often, then four times in a row, a fourth as often, and finally five and a fifth. The program informs the user each time it increases the index multiplier, which is called NSTAR, from 1 to 5.

Another important question in calculating free energies of solvation is the existence of multiple conformational minima of the solute. In principle, the free energy of solvation should be a weighted average over the various conformers [12,40,41]. In our work so far, we have approximated this weighted average by the free energy of solvation of the lowest-energy conformer. This approximation is probably good enough for the general survey and parameterization steps we have carried out, but it should be re-examined for individual future applications.

Before discussing the numbers, we emphasize the dependence of all free energies of solvation on standard state. As mentioned in Section 2, our standard state is 1 M for both gas and solution, and our surface tensions are parameterized to yield free energies of solvation for transfer from an ideal gas at this concentration to an ideal solution at this concentration. From these values one can calculate free energies under other conditions by standard thermodynamic formulas. Another common standard state for gases is 1 atmosphere [26], and free energies of solvation could also be quoted for conditions where the solute is present in the gas phase at its vapor pressure. For a gas-phase standard state of 1 atmosphere, all free energies of solvation would be 1.9 kcal more positive than the values for the standard state used here. So one must be careful in comparing data from different sources, in interpreting the energies physically, or in using the calculated free energies of solvation for applications.

TABLE 5
CALCULATED AND EXPERIMENTAL^a FREE ENERGIES OF SOLVATION (KCAL/MOL) FOR NEUTRAL SOLUTES USED IN THE PARAMETERIZATION OF SM2 AND SM3

Compound	SM1a			SM2			SM3			Expt
	ENP	CDS°	ΔG _s	ENP	CDS°	ΔG _s	ENP	CDS°	ΔG _s	
Hydrocarbons										
ethane	0.0	0.8	0.8	0.0	1.2	1.2	0.0	1.2	1.2	1.8
propane	0.0	1.0	1.0	0.0	1.4	1.4	0.0	1.4	1.4	2.0
cyclopropane	-0.2	0.9	0.7	-0.1	1.3	1.1	-0.2	1.3	1.1	0.8
butane	0.1	1.1	1.2	0.1	1.6	1.7	0.0	1.6	1.7	2.1
2-methylpropane	0.1	1.0	1.1	0.1	1.6	1.6	0.0	1.6	1.7	2.3
neopentane	0.1	1.2	1.3	0.1	1.7	1.9	0.1	1.8	1.8	2.5
cyclopentane	0.0	1.0	1.0	0.1	1.5	1.6	0.0	1.6	1.6	1.2
hexane	0.2	1.3	1.5	0.2	2.0	2.2	0.1	2.1	2.1	2.5
cyclohexane	0.2	1.2	1.4	0.2	1.7	1.9	0.1	1.8	1.8	1.2
heptane	0.2	1.4	1.6	0.2	2.2	2.5	0.1	2.2	2.4	2.6
2,4-dimethylpentane	0.3	1.3	1.6	0.3	2.1	2.4	0.1	2.2	2.3	2.9
methylcyclohexane	0.3	1.2	1.5	0.3	1.9	2.1	0.1	2.0	2.1	1.7
octane	0.3	1.6	1.9	0.3	2.4	2.7	0.1	2.5	2.6	2.9
cis-1,2-dimethylcyclohexane	0.3	1.3	1.6	0.3	2.0	2.4	0.2	2.1	2.3	1.6
ethene	-0.3	0.7	0.4	-0.3	1.1	0.8	-0.3	1.2	0.9	1.3
2-methylpropene	-0.4	1.0	0.6	-0.4	1.5	1.1	-0.4	1.6	1.2	1.2
propene	-0.3	0.8	0.5	-0.3	1.3	1.0	-0.3	1.4	1.1	1.3
E-2-pentene	-0.3	1.2	0.9	-0.3	1.7	1.4	-0.3	1.8	1.5	1.3
cyclopentene	-0.5	1.1	0.6	-0.4	1.5	1.0	-0.6	1.6	1.1	0.6
1,3-butadiene	-0.9	0.8	-0.1	-0.9	1.4	0.6	-2.4	1.6	-0.8	0.6
benzene	-1.9	0.9	-1.0	-2.0	1.4	-0.5	-2.0	1.7	-0.3	-0.9
toluene	-1.9	1.1	-0.8	-1.9	1.7	-0.3	-2.0	1.9	-0.1	-0.9
<i>o</i> -xylene	-1.9	1.2	-0.7	-2.0	1.8	-0.1	-2.0	2.1	0.1	-0.9
<i>m</i> -xylene	-1.9	1.2	-0.7	-1.9	1.9	0.0	-1.9	2.1	0.2	-0.8
<i>p</i> -xylene	-1.8	1.0	-0.8	-1.9	1.9	0.0	-1.9	2.1	0.2	-0.8
naphthalene	-3.5	1.2	-2.3	-3.5	1.8	-1.8	-3.8	2.2	-1.5	-2.4
anthracene	-4.9	1.6	-3.3	-4.8	2.2	-2.5	-5.0	2.8	-2.2	-4.2
ethyne	-1.8	0.6	-1.2	-1.8	1.0	-0.9	-2.3	1.2	-1.2	0.0
propyne	-2.1	1.0	-1.1	-2.1	1.2	-0.9	-2.5	1.4	-1.1	-0.3
1-butyne	-1.9	0.9	-1.0	-1.9	1.4	-0.5	-2.4	1.6	-0.8	-0.2
1-pentyne	-1.8	1.1	-0.7	-1.8	1.6	-0.2	-2.3	1.8	-0.5	0.0
1-hexyne	-1.8	1.1	-0.6	-1.8	1.8	0.0	-2.4	2.1	-0.3	0.3
butenyne	-2.5	0.8	-1.7	-2.5	1.3	-1.2	-3.0	1.6	-1.4	0.0
<i>Mean unsigned error</i>										
H, C, N compounds										
ethylamine	-1.3	-3.2	-4.5	-1.3	-3.9	-5.2	-0.1	-5.4	-5.5	-4.5
1-propylamine	-1.3	-3.7	-5.0	-1.3	-3.7	-5.0	-0.1	-5.2	-5.3	-4.4
1-butylamine	-1.2	-3.2	-4.4	-1.2	-3.5	-4.7	0.0	-5.0	-5.0	-4.3
aniline	-2.5	-2.3	-4.8	-2.6	-3.2	-5.8	-2.3	-4.4	-6.7	-4.9
dimethylamine	-2.5	-1.9	-4.4	-2.6	-1.8	-4.3	-0.3	-2.7	-3.0	-4.3
pyrrolidine	-2.0	-1.8	-3.8	-1.8	-1.7	-3.4	-0.1	-2.5	-2.6	-5.5
piperazine	-3.6	-4.0	-7.6	-3.7	-4.1	-7.8	-0.5	-5.7	-6.1	-7.4
trimethylamine	-3.3	-1.6	-4.9	-3.3	0.7	-2.6	-0.6	0.5	-0.2	-3.2

TABLE 5 (continued)

Compound	SM1a			SM2			SM3			Expt
	ENP	CDS°	ΔG° _s	ENP	CDS°	ΔG° _s	ENP	CDS°	ΔG° _s	
pyridine	-3.9	-0.4	-4.3	-3.9	-0.5	-4.4	-3.0	-0.9	-3.9	-4.7
2-methylpyrazine	-4.8	-1.2	-6.0	-4.8	-1.9	-6.7	-3.3	-2.8	-6.1	-5.5
acetonitrile	-2.1	-2.2	-4.3	-2.0	-2.4	-4.3	-2.8	-3.4	-6.2	-3.9
propanenitrile	-1.8	-2.1	-3.9	-1.8	-2.0	-3.8	-2.4	-3.0	-5.4	-3.9
butanenitrile	-1.6	-1.9	-3.5	-1.6	-1.8	-3.4	-2.2	-2.8	-4.9	-3.7
N-methylpiperazine	-4.4	-2.5	-6.9	-4.4	-1.7	-6.2	-0.8	-2.7	-3.5	-7.8
N,N'-dimethylpiperazine	-5.3	-2.0	-7.3	-5.3	0.6	-4.6	-0.9	0.3	-0.6	-7.6
4-methylpyridine	-4.0	-0.2	-4.2	-3.9	-0.3	-4.2	-3.1	-0.7	-3.7	-4.9
azetidine	-1.5	-3.5	-5.0	-1.6	-2.1	-3.7	-0.4	-3.1	-3.4	-5.6
2,6-dimethylpyridine	-2.8	0.4	-2.4	-2.8	0.5	-2.3	-2.4	0.3	-2.1	-4.6
<i>Mean unsigned error</i>			0.6			1.0			2.0	
H, C, O compounds										
methanol	-1.0	-4.4	-5.4	-1.0	-4.7	-5.8	-1.0	-4.3	-5.3	-5.1
ethanol	-1.0	-3.4	-4.4	-1.0	-3.9	-4.9	-1.1	-3.5	-4.6	-5.0
dimethyl ether	-1.2	-1.5	-2.7	-1.1	-0.4	-1.4	-0.8	-0.6	-1.4	-1.9
acetaldehyde	-3.1	-0.7	-3.8	-3.0	-1.5	-4.5	-3.3	-2.0	-5.2	-3.5
acetic acid	-2.3	-4.8	-7.1	-2.1	-5.6	-7.7	-2.7	-5.8	-8.5	-6.7
methyl formate	-2.4	-2.8	-5.2	-2.4	-2.4	-4.8	-3.3	-3.2	-6.5	-2.8
1,2-ethanediol	-1.6	-7.0	-8.6	-1.6	-8.1	-9.6	-1.7	-7.4	-9.1	-7.7
1-propanol	-0.9	-3.2	-4.1	-0.9	-3.7	-4.6	-1.0	-3.3	-4.3	-4.8
<i>i</i> -propanol	-0.9	-2.9	-3.8	-0.9	-3.2	-4.1	-0.8	-2.9	-3.6	-4.8
prop-2-en-1-ol	-1.3	-3.5	-4.8	-1.2	-3.7	-4.9	-0.4	-3.2	-3.6	-5.0
2-methoxyethanol	-1.4	-5.3	-6.7	-1.4	-4.9	-6.3	-1.5	-4.7	-6.2	-6.8
propanal	-2.6	-0.5	-3.1	-2.6	-1.0	-3.6	-3.0	-1.4	-4.5	-3.5
propanone	-3.3	-0.4	-3.7	-3.2	-1.0	-4.1	-3.3	-1.5	-4.7	-3.9
propanoic acid	-1.6	-4.4	-6.0	-1.6	-5.1	-6.7	-2.2	-5.2	-7.4	-6.5
methyl acetate	-2.3	-2.0	-4.3	-2.3	-1.7	-4.0	-2.3	-2.4	-4.7	-3.3
<i>t</i> -butanol	-0.8	-2.3	-3.1	-0.9	-2.4	-3.3	-0.7	-2.1	-2.8	-4.5
diethyl ether	-0.9	0.2	-0.7	-0.9	0.7	-0.2	-0.7	0.6	-0.2	-1.6
1-methoxypropane	-0.9	-0.8	-1.7	-0.9	0.4	-0.5	-0.7	0.2	-0.5	-1.7
2-methoxypropane	-0.9	-0.6	-1.5	-0.9	0.5	-0.4	-0.7	0.3	-0.4	-2.0
1,2-dimethoxyethane	-1.7	-2.5	-4.2	-1.7	-0.8	-2.5	-1.2	-1.2	-2.5	-4.8
butanal	-2.5	-0.3	-2.8	-2.5	-0.8	-3.3	-2.9	-1.2	-4.1	-3.2
butanone	-2.8	-0.1	-2.9	-2.8	-0.5	-3.2	-3.1	-0.9	-4.0	-3.6
butanoic acid	-1.5	-4.2	-5.7	-1.5	-4.9	-6.3	-2.0	-5.0	-7.0	-6.4
ethyl acetate	-2.2	-1.1	-3.3	-2.2	-1.2	-3.3	-2.2	-1.8	-4.0	-3.1
methyl propanoate	-1.8	-1.5	-3.3	-1.8	-1.2	-3.0	-1.8	-1.9	-3.7	-2.9
2-methoxy-2-methylpropane	-0.8	-0.2	-1.0	-0.8	0.8	0.0	-0.6	0.6	0.0	-2.2
2-pentanone	-2.3	0.2	-2.1	-2.3	0.0	-2.3	-2.6	-0.4	-3.1	-3.5
3-pentanone	-2.4	0.2	-2.2	-2.4	0.0	-2.4	-2.8	-0.3	-3.1	-3.4
methyl butanoate	-1.7	-1.4	-3.1	-1.7	-1.0	-2.7	-1.6	-1.7	-3.3	-2.8
4-heptanone	-2.1	0.5	-1.6	-2.0	0.4	-1.6	-2.4	0.1	-2.3	-2.9
5-nonanone	-2.2	0.7	-1.5	-2.2	0.8	-1.4	-2.5	0.5	-2.0	-2.7
tetrahydrofuran	-1.4	-1.7	-3.1	-1.4	-0.2	-1.5	-1.2	-0.4	-1.7	-3.5
1,4-dioxane	-2.0	-3.9	-5.9	-2.0	-1.4	-3.4	-1.5	-2.0	-3.5	-5.1
phenol	-2.5	-3.3	-5.8	-2.5	-3.3	-5.8	-2.7	-2.7	-5.4	-6.6

TABLE 5 (continued)

Compound	SM1a			SM2			SM3			Expt
	ENP	CDS°	ΔG _s ^o	ENP	CDS°	ΔG _s ^o	ENP	CDS°	ΔG _s ^o	
benzaldehyde	-4.1	-0.3	-4.4	-4.1	-0.7	-4.7	-4.6	-0.9	-5.5	-4.0
acetophenone	-4.3	-0.1	-4.4	-4.3	-0.1	-4.4	-4.7	-0.4	-5.1	-4.6
<i>m</i> -hydroxybenzaldehyde	-4.2	-4.5	-8.7	-4.2	-5.4	-9.5	-5.2	-5.3	-10.6	-9.5
<i>p</i> -hydroxybenzaldehyde	-4.4	-4.6	-9.0	-4.4	-5.3	-9.8	-5.5	-5.3	-10.8	-10.5
<i>Mean unsigned error</i>			0.7			0.8			1.1	
H, C, F compounds										
fluoromethane	-1.4	1.6	0.2	-1.4	1.8	0.4	-0.9	1.7	0.8	-0.2
trifluoromethane	-3.1	3.1	0.0	-3.2	3.2	0.1	-3.3	2.9	-0.4	0.8
tetrafluoromethane	-0.6	4.1	3.5	-0.6	4.0	3.4	-0.4	3.5	3.1	3.1
1,1-difluoroethane	-2.3	2.3	-0.0	-2.4	2.6	0.3	-1.8	2.5	0.6	-0.1
<i>Mean unsigned error</i>			0.4			0.5			0.7	
H, P compound										
phosphine	0.0	0.6	0.6	0.0	0.6	0.6	-1.0	1.6	0.6	0.6
<i>Unsigned error</i>			0.0			0.0			0.0	
H, C, S compounds										
methanethiol	-0.2	-0.8	-1.0	-0.2	-0.7	-0.8	-0.2	-0.7	-0.9	-1.2
ethanethiol	-0.3	-0.3	-0.6	-0.3	-0.3	-0.6	-0.4	-0.3	-0.7	-1.3
dimethyl sulfide	-0.1	-1.4	-1.5	-0.1	-1.5	-1.6	-0.5	-1.6	-2.1	-1.4
diethyl sulfide	-0.1	-0.3	-0.4	-0.1	-0.3	-0.4	-1.0	-0.4	-1.4	-1.3
thiophenol	-3.2	-0.7	-3.9	-3.2	0.0	-3.2	-2.8	0.2	-2.5	-2.6
thioanisole	-3.5	-0.6	-4.1	-3.5	-0.7	-4.2	-3.0	-0.5	-3.5	-2.7
<i>Mean unsigned error</i>			0.8			0.7			0.4	
H, C, Cl compounds										
methyl chloride	-1.0	0.2	-0.8	-1.0	0.3	-0.7	-0.4	0.2	-0.2	-0.6
dichloromethane	-1.0	-0.2	-1.2	-1.0	-0.2	-1.2	-0.3	-0.3	-0.7	-1.4
chloroform	-0.7	-0.5	-1.2	-0.7	-0.5	-1.2	-0.2	-0.7	-0.9	-1.1
ethyl chloride	-1.0	0.2	-0.8	-1.0	0.6	-0.5	-0.6	0.5	-0.1	-0.6
1,2-dichloroethane	-0.9	-0.1	-1.0	-1.0	-0.1	-1.0	-0.5	-0.2	-0.7	-1.7
1,1,1-trichloroethane	-1.0	-0.1	-1.1	-0.9	-0.1	-1.0	-0.3	-0.3	-0.6	-0.3
1,1,2-trichloroethane	-1.0	-0.3	-1.3	-1.1	-0.3	-1.4	-0.4	-0.5	-0.9	-2.0
isopropyl chloride	-1.1	0.6	-0.5	-1.1	0.8	-0.3	-0.6	0.8	0.2	-0.3
vinyl chloride	-0.8	0.3	-0.5	-0.8	0.5	-0.3	-0.5	0.5	0.0	-0.6
Z-1,2-dichloroethene	-1.1	-0.1	-1.1	-1.1	0.0	-1.1	-0.7	-0.1	-0.8	-1.2
E-1,2-dichloroethene	-0.5	-0.2	-0.7	-0.5	-0.2	-0.7	-0.6	-0.3	-0.9	-0.8
trichloroethene	-0.6	-0.4	-1.0	-0.6	-0.4	-1.0	-1.1	-0.6	-1.7	-0.4
tetrachloroethene	-0.3	-0.6	-0.9	-0.3	-0.7	-0.9	-1.8	-0.9	-2.7	0.1
chlorobenzene	-2.0	0.6	-1.4	-2.0	0.9	-1.1	-2.4	1.0	-1.3	-1.1
<i>Mean unsigned error</i>			0.4			0.3			0.7	
H, C, Br compounds										
bromomethane	-0.3	-0.4	-0.7	-0.4	-0.3	-0.7	-1.2	0.1	-1.1	-0.8
bromoethane	-0.4	-0.2	-0.6	-0.4	0.0	-0.4	-1.6	0.4	-1.2	-0.7
2-bromopropane	-0.5	0.1	-0.4	-0.5	0.3	-0.2	-1.8	0.7	-1.1	-0.5
bromobenzene	-2.3	0.1	-2.2	-2.3	0.4	-2.0	-2.4	0.9	-1.5	-1.5
dibromomethane	-0.2	-1.3	-1.5	-0.2	-1.3	-1.5	-1.0	-0.5	-1.5	-2.1
1,2-dibromoethane	-0.6	-0.8	-1.4	-0.6	-0.9	-1.5	-2.5	-0.3	-2.8	-2.1
<i>p</i> -dibromobenzene	-2.4	-0.8	-3.2	-2.4	-0.7	-3.2	-2.5	0.1	-2.3	-2.3

TABLE 5 (continued)

Compound	SM1a			SM2			SM3			Expt	
	ENP	CDS°	ΔG _s ^a	ENP	CDS°	ΔG _s ^a	ENP	CDS°	ΔG _s ^a	ΔG _s ^a	
bromoform	-0.2	-1.9	-2.1	-0.2	-2.0	-2.2	-0.6	-0.9	-1.5	-2.1	
<i>Mean unsigned error</i>			0.4			0.4			0.4	0.4	
H, C, I compounds											
iodomethane	-0.1	-0.9	-1.0	-0.1	-1.0	-1.1	-1.0	0.3	-0.7	-0.9	
iodoethane	0.0	-0.7	-0.7	0.0	-0.7	-0.7	-1.6	0.6	-0.9	-0.7	
1-iodopropane	0.0	-0.5	-0.5	0.0	-0.5	-0.5	-1.2	0.8	-0.4	-0.6	
2-iodopropane	0.0	-0.4	-0.4	0.0	-0.4	-0.4	-1.8	0.9	-0.9	-0.5	
1-iodobutane	0.1	-0.4	-0.3	0.1	-0.3	-0.2	-1.2	1.0	-0.1	-0.3	
<i>Mean unsigned error</i>			0.1			0.1			0.2	0.2	
compounds with 4 or more kinds of atoms											
2-methoxyethanamine	-2.2	-5.3	-7.5	-2.2	-4.9	-7.1	-0.8	-6.6	-7.4	-6.6	
nitroethane	-1.8	-1.9	-3.7	-1.7	-2.6	-4.4	1.3	-3.9	-2.6	-3.7	
1-nitropropane	-1.5	-1.8	-3.3	-1.4	-2.4	-3.8	1.3	-3.7	-2.4	-3.3	
2-nitropropane	-1.9	-1.6	-3.5	-1.9	-2.2	-4.1	2.0	-3.4	-1.4	-3.1	
nitrobenzene	-3.2	-1.6	-4.8	-3.2	-2.1	-5.2	-1.1	-3.1	-4.1	-4.1	
2-methyl-1-nitrobenzene	-3.5	-1.2	-4.7	-3.4	-1.6	-5.0	-1.1	-2.5	-3.6	-3.6	
2,2,2-trifluoroethanol	-3.2	-0.9	-4.1	-3.2	-1.4	-4.5	-3.7	-1.3	-4.9	-4.3	
1,1,1-trifluoropropan-2-ol	-2.3	-1.1	-3.4	-2.4	-0.8	-3.2	-2.5	-0.8	-3.3	-4.2	
chlorofluoromethane	-1.8	1.1	-0.7	-1.8	1.1	-0.7	-1.1	0.9	-0.2	-0.8	
chlorodifluoromethane	-2.1	1.8	-0.3	-2.1	1.7	-0.4	-1.7	1.5	-0.2	-0.5	
2-chloro-1,1,1-trifluoroethane	-2.2	2.6	0.4	-2.2	2.6	0.5	-2.0	2.2	0.2	0.1	
bis-2-chloroethyl sulfide	-1.6	-0.9	-2.5	-1.6	-1.6	-3.2	-0.9	-1.9	-2.8	-3.9	
p-bromophenol	-2.6	-4.3	-6.9	-2.7	-4.4	-7.0	-3.1	-3.5	-6.5	-7.1	
bromotefluoromethane	-1.5	2.1	0.6	-1.4	1.9	0.5	-0.6	1.9	1.3	1.8	
2-bromo-1-chloroethane	-0.7	-0.5	-1.2	-0.7	-0.6	-1.3	-1.0	-0.3	-1.4	-2.0	
3-ethyl-2-methoxypyrazine	-4.6	-0.9	-5.5	-4.3	-1.5	-5.8	-3.2	-2.4	-5.6	-4.4	
morpholine	-2.8	-4.0	-6.8	-2.9	-2.7	-5.6	-1.1	-3.9	-4.9	-7.2	
P(OCH ₃) ₃	-6.0	-3.3	-9.2	-5.9	-0.8	-6.7	-6.3	-0.9	-7.2	-8.7	
P(OC ₂ H ₅) ₃	-5.2	-1.0	-6.2	-5.1	0.7	-4.4	-6.0	0.8	-5.2	-7.8	
P(OC ₃ H ₇) ₃	-4.6	-0.5	-5.1	-4.5	1.3	-3.2	-5.6	1.4	-4.2	-6.1	
<i>Mean unsigned error</i>			0.6			1.1			1.0	1.0	
compounds without C or P											
ammonia	-0.1	-6.1	-6.1	-0.3	-4.0	4.3	0.0	-4.3	-4.3	-4.3	
water	-1.2	-7.4	-8.6	-1.2	-5.1	-6.3	-1.9	-4.4	-6.3	-6.3	
hydrogen sulfide	-0.2	-0.5	-0.7	-0.2	-0.7	-0.9	0.0	-0.8	-0.8	-0.7	
<i>Mean unsigned error</i>			1.0			0.07			0.03	0.03	
Mean unsigned error all compounds			0.6			0.7			0.9	0.9	
RMS error all compounds			0.8			0.9			1.3	1.3	

^a Reference 25.

5. PERFORMANCE OF THE MODELS

In this section we compare AM1-SM2 and PM3-SM3 to one another and, with one exception, to either AM1-SM1a when it is applicable and unambiguous, or AM1-SM1 when AM1-SM1a is ambiguous. The exception is the case of explicit water molecules when AM1-SM1 and AM1-SM1a are not recommended because the newer models are much more accurate for explicit water. Tables 5 and 6 contain the solutes used for the actual parameterization of SM2 and SM3. The remaining Tables 7–12 present other interesting cases for solvation free energies and solvation effects on equilibria.

5.1. *The hydrophobic effect*

Because of the disparate effects of an attached heavy atom, $\sigma_k^{(0)}$ for hydrogen in SM1 was chosen to be zero. Since the vast majority of the solvent-accessible surface area for a saturated hydrocarbon within the SM1 formalism is hydrogenic, this leads to anomalously small, positive free energies of hydration. By the differentiation of heavy atoms to which hydrogen atoms are bound, using Eq. 25, and by going to an approach for the solvent-accessible surface area which exposes carbon atoms that would be otherwise largely covered by non-zero-volume hydrogens, this error has largely been eliminated.

Results for SM1a, SM2, and SM3 for a number of neutral solutes are presented in Table 5. Large improvements are observed for the saturated hydrocarbons using the new models. Moderate or slight improvements are observed for unsaturated hydrocarbons; however, AM1-SM1a is fairly accurate for these molecules. There is little difference in the ΔG_{ENP} term from any of the three methods, except for alkynes and butadiene, for which PM3-SM3 predicts considerably greater solvent polarization. The mean absolute error for 33 hydrocarbons is 0.5 kcal in SM2 and 0.9 kcal in SM3.

5.2. *Amino, hydroxyl, and mercapto groups*

It is evident from the $\sigma_k^{(1)}$ values in Table 1 that G_{CDS}° values for oxygen atoms bearing protons, e.g., alcohols and acids, will be significantly more negative than for the analogous oxygen atoms in carbon-substituted analogs, e.g. ethers and esters. This is entirely in keeping with intuitive expectations, especially since hydrogen bonding to explicit water molecules, which cannot be modeled *per se* by a continuum treatment of the solvent, is accounted for in part by the sizable $\sigma_k^{(1)}$ terms. The same is true for the comparison of primary or secondary amines to their tertiary or aromatic congeners. Conversely, sulfur atoms substituted with protons developed a less negative overall surface tension. It should be recalled, however, that since the proton has zero volume during the calculation of solvent-accessible surface area, the exposed surface of a sulfur atom in a thiol is considerably larger than for one found in a sulfide. Thus, the overall contribution from the sulfur atom to G_{CDS}° in either type of compound is roughly equivalent. Of course, because there are relatively few data with which to parameterize sulfur, any analysis of the quantitative meaning of these values must be tempered with a modicum of skepticism.

Table 5 may be consulted for a comparison of the three solvation models for a number of compounds containing nitrogen, oxygen, and/or sulfur. With SM1, $-\Delta G^{\circ}$ for alcohols and acids tends to be underestimated, while it tends to be overestimated for ketones and esters. By explicit recognition of the hydroxyl proton, SM1a, SM2, and SM3 provide considerable improvement. A

similar situation exists with SM1 for amines in comparison to nitriles. While SM2 offers dramatic improvement, SM3 in this instance does not. This is discussed in more detail below. Mixed results are observed for sulfur, where again, scarcity of data limits interpretation.

5.3. Ions

Table 6 lists predicted free energies of solvation calculated via both the SM2 and SM3 formalisms for 28 different ions for which experimental free energies of solvation are available. The ex-

TABLE 6
CALCULATED AND EXPERIMENTAL FREE ENERGIES OF SOLVATION (KCAL/MOL) FOR IONS USED IN THE PARAMETERIZATION OF SM2 AND SM3

Compound	SM1			SM2			SM3			Expt ^a	
	G _{ENP}	G _{CDS} ⁰	ΔG _s ⁰	G _{ENP}	G _{CDS} ⁰	ΔG _s ⁰	G _{ENP}	G _{CDS} ⁰	ΔG _s ⁰	ΔG _s	
H ⁻	-89.0	0.0	-89.0	-89.0	0.0	-89.0	-89.0	0.0	-89.0	-89	
F ⁻	-106.9	1.9	-105.0	-109.2	2.2	-107.0	-109.0	2.0	-107.0	-107	
Cl ⁻	-74.7	-0.3	-75.0	-76.7	-0.3	-77.0	-76.5	-0.5	-77.0	-77	
Br ⁻	-68.8	-1.2	-70.0	-70.8	-1.2	-72.0	-71.3	-0.7	-72.0	-72	
I ⁻	-59.8	-1.2	-61.0	-60.8	-2.2	-63.0	-62.6	-0.4	-63.0	-63	
OH ⁻	-105.0	-4.2	-109.2	-99.9	-8.2	-108.1	-103.9	-7.5	-111.4	-106	
CN ⁻	-75.9	-3.1	-79.0	-80.8	-2.6	-83.5	-76.2	-3.4	-79.6	-77	
O ₂ ⁻	-84.4	-6.8	-91.2	-84.3	-3.8	-88.1	-79.5	-5.4	-84.9	-87	
HS ⁻	-73.7	-1.9	-75.6	-74.4	-1.9	-76.3	-74.7	-1.9	-76.7	-76	
PH ₂ ⁻	-68.0	0.9	-67.1	-67.6	0.6	-67.0	-68.4	1.6	-66.8	-67	
HC ₂ ⁻	-78.4	1.8	-76.6	-78.5	0.7	-77.8	-78.0	1.2	-76.8	-73	
HO ₂ ⁻	-95.8	-5.5	-101.4	-90.0	-7.0	-97.0	-82.1	-7.4	-89.5	-101	
N ₃ ⁻	-63.6	-11.3	-74.9	-69.5	-5.7	-75.2	-64.7	-7.9	-72.6	-74	
NO ₂ ⁻	-74.8	-8.6	-83.4	-73.0	-4.7	-77.7	-70.4	-6.5	-76.9	-72	
CH ₃ O ⁻	-86.1	-2.8	-88.9	-81.3	-2.2	-83.5	-79.0	-2.6	-81.6	-95	
NO ₃ ⁻	-55.4	-9.1	-64.5	-54.0	-5.0	-59.0	-40.7	-7.1	-47.7	-65	
CH ₂ CN ⁻	-64.9	-3.5	-68.4	-67.3	-2.4	-69.7	-67.8	-3.5	-71.3	-75	
CH ₃ CO ₂ ⁻	-74.5	-5.0	-79.5	-72.9	-3.0	-75.9	-71.3	-4.3	-75.5	-77	
H ₃ O ⁺	-100.6	-1.6	-102.1	-100.0	-3.8	-103.8	-98.9	-3.5	-102.4	-104	
NH ₄ ⁺	-77.8	-0.4	-78.2	-77.3	-1.7	-79.0	-71.2	-10.5	-81.7	-79	
MeOH ₂ ⁺	-79.8	-0.8	-80.5	-79.0	-5.2	-84.2	-81.5	-3.5	-84.6	-83	
MeSH ₂ ⁺	-73.6	-0.8	-74.4	-73.7	0.1	-73.6	-76.7	0.1	-76.6	-74	
MeNH ₃ ⁺	-69.7	0.0	-69.6	-68.6	-3.3	-71.9	-64.7	-2.4	-67.1	-70	
Me ₂ OH ⁺	-61.3	-0.2	-61.5	-60.9	-1.8	-62.8	-67.5	-1.6	-69.1	-70	
Me ₂ NH ₂ ⁺	-61.0	0.3	-60.7	-61.3	-2.0	-63.3	-60.0	-3.1	-63.1	-63	
Me ₃ NH ⁺	-53.2	0.3	-52.9	-53.9	0.2	-53.7	-54.3	-0.2	-54.5	-59	
Me ₃ PH ⁺	-49.6	0.2	-49.4	-49.4	1.5	-47.9	-58.8	1.8	-57.0	-53	
CH ₃ C(OH)NH ₂ ⁺	-60.6	-2.3	-62.9	-60.4	-7.8	-68.2	-66.6	-9.0	-75.6	-66	
<i>Mean unsigned error</i>			2.9			2.6			3.5		
<i>RMS error</i>			4.0			3.9			5.6		

^a From Reference 26. Errors may be substantial (\pm 5 kcal/mol).

perimental numbers, presented for comparison, are derived from [26]

$$-\Delta G_s^\circ(X^-) = PA - 267 - \Delta G_s^\circ(HX) - 1.36 pK_a \quad (29)$$

and

$$-\Delta G_s^\circ(BH^+) = -PA + 267 - \Delta G_s^\circ(B) + 1.36 pK_a \quad (30)$$

where all constants are in kcal. PA is the gas-phase proton affinity, the constant 267 is derived from the absolute potential of the hydrogen electrode and assumed entropic contributions to the gas phase acid dissociation free energies, the constant 1.36 is $RT \ln 10$ at 298 K, and pK_a is the usual aqueous acid equilibrium constant. Given the very large potential uncertainties, especially in gas-phase PA measurements and pK_a measurements for very strong or very weak acids, the observed fit for the two models is probably about as good as can be expected. Neither SM2 nor SM3 performs significantly better than SM1 for ions.

Of course, it should be pointed out that minimal basis set calculations on anions are usually much less reliable than for neutral or positively charged molecules. It is unclear, for example, whether the rather large error observed for methoxide arises from a poor description of this ion at the semiempirical level, or whether it is in part an error in the experimental proton affinity.

Two other compounds which are treated very poorly are the nitrite and nitrate anions. In this instance, however, the fault almost certainly rests with the model and not with the experiment, since we have similar difficulties with neutral compounds containing a nitro group, especially using the PM3 Hamiltonian (vide infra). Table 7 presents additional solvation calculations for ions for which there is no experimental data or only questionable data.

Because of the extremely large uncertainties in the aqueous pK_a values for simple carbanions, we have not included any in our parametrization set. It is noteworthy, however, that the two models perform very differently for the simple methyl anion: AM1-SM2 predicts $\Delta G_s^\circ(\text{CH}_3^-)$ to

TABLE 7
CALCULATED AND EXPERIMENTAL FREE ENERGIES OF SOLVATION (KCAL/MOL) FOR ADDITIONAL IONS

Compound	SM1			SM2			SM3			Expt ^a
	G_{ENP}	G_{CDS}°	ΔG_s°	G_{ENP}	G_{CDS}°	ΔG_s°	G_{ENP}	G_{CDS}°	ΔG_s°	
CH_3^-	-68.1	0.7	-67.4	-68.4	0.9	-67.5	-83.5	1.0	-82.5	-82
PH_4^+	-63.5	0.4	-63.1	-61.4	0.6	-60.8	-61.0	1.6	-59.4	-73
MePH_3^+	-59.8	0.3	-59.5	-57.8	0.8	-57.0	-59.6	1.7	-657.9	-66
Me_2PH_2^+	-54.3	0.2	-54.1	-51.1	1.2	-49.9	-61.0	1.8	-59.2	-57
H_2PO_4^-	-120.8	-8.1	-128.9	-95.5	-10.0	-105.5	-75.4	-10.7	-86.1	
HPO_4^{2-}	-223.6	-9.5	-333.1	-291.5	-8.0	-299.5	-256.5	-9.4	-265.9	
PO_4^{3-}	-665.0	-10.6	-675.5	-623.5	-5.7	-629.1	-562.4	-8.1	-570.5	
SH_3^+	-76.3	-1.1	-77.3	-76.4	-0.3	-76.7	-82.1	-0.4	-82.5	-87

^a When given, from Reference 26, but considered uncertain or unreliable (see text). In other cases, unavailable.

be -67.5 kcal, and PM3-SM3 predicts -82.5 kcal. Pearson, using Eq. 29, suggests a value of -82 kcal. Interestingly, AM1 predicts the gas-phase methyl anion to be planar (albeit requiring little energy for pyramidalization), whereas PM3 correctly predicts a pyramidal anion. When solvation effects are included, both models predict a pyramidal anion. It may thus be preferable to use PM3-SM3 in carbanion-like situations.

Another ion of interest which we have not included in our parametrization set is phosphonium, PH_4^+ . Pearson, using Eq. 29 and an estimated pK_a of -3 , predicts that ΔG_s° for this compound is -73 kcal [26]. He points out, however, that the generally quoted value for the pK_a is -14 , which would give $\Delta G_s^\circ = -58$ kcal. Pearson rejects this value because he believes trends in the methylated phosphonium series should mimic those for the ammonium analogues. However, SM2 and SM3 predict ΔG_s° values of -60.8 and -59.4 , respectively, supporting the more negative pK_a estimate. Resolution of this discrepancy will have to await accurate experimental data. For similar reasons, as discussed above, we have not included H_3S^+ in our SM2 and SM3 parametrization sets, although it was included in the development of SM1.

The methylphosphonium and dimethylphosphonium ions are placed in Table 7 because again the only available estimates of ΔG_s° from experiment are obtained [26] from estimated values for the solvation energies of the conjugate bases.

Table 7 also shows H_2PO_4^- , HPO_4^{2-} , and PO_4^{3-} ions, for which we know of no experimental data for free energies of solvation. We can, however, refer to the experimental enthalpies of solvation, which are -76 , -299 , and -637 kcal for H_2PO_4^- , HPO_4^{2-} , and PO_4^{3-} , respectively [42]. We can also check our consistency with a critical pK_a value for metabolic equilibria, in particular the pK_a of H_2PO_4^- . This pK_a of H_2PO_4^- is 7.2 [43]. Thus at physiological pH, the predominant esters are ROPO_3H^- and ROPO_3^{2-} , and they are present in about equal amounts. And, of course, ROPO_3R^- is also typically singly ionized. Table 7 shows a very large difference between AM1-SM2 and PM3-SM3 for phosphate, which is easily traced to the differing gas-phase partial charges on P in AM1 and PM3, namely 2.52 in the former and 1.64 in the latter (vs. 1.32 for a HF/6-311⁺G* calculation). Of course such errors will partly cancel in reactions where the ionization state of PO_4^{3-} does not change, but they are still disturbing. One way to determine which treatment is more realistic is to use the G_s° values themselves (elsewhere in this article we have tested the models in terms of their ΔG_s° values) to calculate pK_a values of H_3PO_4 and H_2PO_4^- . Doing this yields 15.3 and 15.1 for AM1-SM2 and 24.5 and 31.1 for PM3-SM3, vs. 2.1 and 7.2 from standard tables [43]. Clearly AM1-SM2 is somewhat more realistic. It is also clear, as pointed out by Lim et al. [44], that it is very difficult to calculate absolute pK_a values by molecular modeling. As mentioned above, the semi-empirical Hamiltonians are expected to deliver increasingly too positive G° values for the more negatively charged species. This in part explains the anomalously high pK_a values in spite of the ΔG_s° values comparing reasonably well with the experimental solvation enthalpies.

5.4. Neutral molecules

Table 5 presents predicted solvation free energies for all of the 150 neutrals used in the parameterization of SM2 and SM3. These neutral solutes include both mono- and polyfunctional compounds. For the full data set of 150 compounds in Table 5, SM2 and SM3 give mean unsigned errors of 0.69 (root-mean-square error: 0.95) and 0.94 (rms error: 1.29) kcal/mol, respectively. These errors represent considerable improvement over SM1, where for the same data set the mean un-

signed error is 1.17 kcal (rms error: 1.51). As detailed above, the majority of the improvement is found for the classes of compounds specifically addressed above. Other classes, such as the halocarbons, are well treated by all three models. Differences in free energies of solvation for closely related molecules, especially differences due to hydrophobic groups, are predicted even better.

It is interesting to note, however, that AM1 and PM3 treat the halides differently. Thus, PM3 employs gas-phase parameters, that cause fluorine and chlorine to be slightly and moderately less electronegative, respectively, than AM1. In contrast, bromine and iodine atoms are both considerably more electronegative in PM3 than AM1. These differences give rise to the less negative G_{ENP} energies observed for the fluoro- and chlorocarbons and the more negative G_{ENP} energies observed for the bromo- and iodocarbons when comparing SM3 to SM2.

In any large data set there are bound to be interesting special cases that deserve special attention. One interesting solute in Table 5 is methyl formate, which has a very large error in both AM1-SM2 and PM3-SM3. This provides a caution about complications that must be watched for in specific systems. In any event, we have included it in our average error calculations for completeness.

Systematic deficiencies in whole classes of neutrals are primarily confined to the SM3 model, where nitrogen-containing compounds are treated rather poorly. We have discussed this in our previous paper [4], where we explained that it follows from partial charges on N being systematically too positive in PM3, as had been noted previously [30,32]. Thus amines and nitroaliphatics give ΔG_{ENP} terms which are much too small; in some cases the term may even be positive. The surface tension terms cannot completely correct for this deficiency, *inter alia*, because of the large negative G_{ENP} terms found for nitriles with SM3 compared to SM2.

A case worthy of special mention in this regard is N, N'-dimethylpiperazine. Here the effect of methylation is much larger in the models than in the experiments. Examination of surface areas and partial charges leads to the conclusion that this error is more probably due to the inaccuracy of the semi-empirical models for nitrogen partial charges than to the parameterization of surface tensions (since the exposed nitrogen surface area is very small). Although the 3 kcal error for this compound is disconcerting, being 4.5 times larger than the mean unsigned error in AM1-SM2 for neutrals, we can put it in perspective by noting that the molecule has no dipole moment and so the error would be 7.8 kcal in theories based solely on the dipole moment.

Although the nitro group itself does not appear to be particularly well treated even by SM2, where the solvation free energies of nitro compounds are uniformly too negative, amines and nitriles appear to present no significant problems. Tertiary amines continue to deliver $\Delta G_{\text{S}}^{\circ}$ values that are not sufficiently negative.

The only remaining functionality which exhibits systematic deficiencies is the ether class, which does so both with the SM2 and SM3 models. In general, both these models give $\Delta G_{\text{S}}^{\circ}$ values which are not negative enough for ethers. Since a wide variety of other oxygen functionalities all appear to be well treated, there is little room for flexibility within the surface tension terms to absorb this error. While one may imagine analyses of the bond-order matrix which allow for explicit correction of this anomaly, the magnitude of the error is small enough that such an approach is probably not warranted for other than special, particular purposes (*vide infra*). In any event, one should be cautious about making up for systematic deficiencies which may have their origin in G_{ENP} due to poor semiempirical changes by ad hoc adjustments in G_{CDS}° .

Difficulties with unusual atomic charges are not necessarily limited to nitrogen in SM3. Sulfur

and phosphorus atoms may easily take on very large positive charges within both AM1 and PM3, especially in respectively tetra- and pentavalent situations. A paucity of experimental data prevents us from assessing the efficacy of the solvation models in these instances. Of course, when interested primarily in reaction energetics as opposed to absolute free energies of solvation, a fortuitous cancellation of errors may occur.

Table 8 shows selected results for additional compounds, for most of which experimental data are unavailable or unknown to us. To illustrate possible biological applications, Table 8 includes an amino acid in both its uncharged and zwitterionic forms [32], an amide for which an experimental value is available [45], the simplest carbohydrate (a diose), and a simple nucleoside. Table 8 also includes trimethyl phosphine, for which the solvation energy has been estimated by Pearson [26], and phosphoric acid, for which we know of no experimental data. The final six rows of Table 8 will be discussed in detail in the next subsection.

5.5. Illustrative series: pentanols, pentanal and pentanones

Table 8 contains four pentanols, cyclopentanol, pentanal, and cyclopentanone. These are interesting to compare with each other and with the analogous acyclic ketones, 2- and 3-pentanone, for which data are given in Table 5. Within each series, 1-, 2-, and 3-pentanol and pentanal, 2-pentanone, and 3-pentanone, some general trends are apparent. As expected, the solvent-accessible

TABLE 8
CALCULATED AND EXPERIMENTAL FREE ENERGIES OF SOLVATION FOR ADDITIONAL NEUTRAL SOLUTES (KCAL/MOL)

Compound	SM1			SM2			SM3			Expt
	ΔG_{ENP}	G_{DS}	ΔG_{S}	ΔG_{ENP}	G_{DS}	ΔG_{S}	ΔG_{ENP}	G_{DS}	ΔG_{S}	
H ₂ NCH ₂ COOH	-4.1	-5.0	-9.1	-4.0	-9.8	-13.8	-3.0	-11.4	-14.4	
+H ₃ NCH ₂ COO-	-41.3	-4.8	-46.1	-39.5	-6.2	-45.7	-34.5	-6.5	-40.1	
<i>E</i> -N-methylformamide	-8.7	-2.7	-11.4	-8.7	-3.5	-12.2	-7.2	-4.7	-11.8	-10.0^a
glycoaldehyde	-3.5	-4.0	-7.5	-3.4	-5.8	-9.1	-4.1	-5.9	-10.0	
thymine	-10.5	-4.9	-15.4	-10.4	-6.1	-16.5	-11.6	-8.5	-20.1	
uracil	-11.5	-5.2	-16.7	-11.7	-6.6	-18.2	-12.7	-9.0	-21.7	
cytosine	-12.8	-4.8	-17.6	-13.4	-8.4	-21.7	-11.9	-11.3	-23.2	
deoxycytidine	-12.5	-7.3	-19.8	-12.5	-14.2	-26.7	-12.1	-16.3	-28.4	
(CH ₃) ₃ P	-0.5	0.6	0.1	-0.5	1.5	1.0	-0.8	1.9	1.1	-0.9^b
H ₃ PO ₄	-57.2	-6.9	-64.1	-39.7	-12.4	-52.1	-22.8	-12.0	-34.8	
1-pentanol	-0.8	-1.4	-2.2	-0.8	-3.3	-4.1	-1.0	-2.9	-3.9	-4.5^c
2-pentanol	-0.6	-1.1	-1.7	-0.6	-2.4	-2.9	-0.6	-2.0	-2.7	-4.4^c
3-pentanol	-0.5	-0.8	-1.3	-0.5	-2.0	-2.4	-0.6	-1.7	-2.3	-4.4^c
cyclopentanol	-0.8	-1.5	-2.3	-0.7	-3.1	-3.8	-0.7	-2.6	-3.3	-5.5^c
pentanal	-2.5	-2.0	-4.4	-2.5	-0.6	-3.1	-2.9	-1.0	-3.9	-3.0^c
cyclopentanone	-3.0	-2.3	-5.3	-3.0	-0.8	-3.8	-3.1	-1.3	-4.4	

^a Reference 45.

^b Estimate from Reference 26.

^c Reference 25.

surface area of the oxygen atom decreases uniformly as its substitution point is changed from position 1 to 2 to 3. On the other hand, when one ties back the two alkyl groups to make the functionalized cyclopentane, the exposed surface area increases. The observed trends in G_{CDS}° can be almost entirely attributed to this effect; G_{CDS}° for the hydrocarbon portions of the acyclic substrates is effectively constant, whereas forming the ring decreases the hydrophobicity by about 0.3 kcal/mol. The overall effect is greater in the alcohol series than in the ketone series because of the less crowded sp^2 hybridization of the carbon atom bearing oxygen in the latter. The sp^2 hybridization leads to essentially identical exposed oxygen atom surface areas for 2- and 3-pantanone.

The trends in G_{ENP}° are not as obvious, but they remain fairly straightforward. For instance, one expects a trend in the oxygen effective coulomb radius α_O similar to that seen for the solvent-accessible surface area. In particular, as an atom becomes more buried within a molecule, its decreased interaction with the surrounding dielectric causes it to be assigned (Eqs. 14–16) a larger effective coulomb radius, α_k' . In the pentanol series, α_O is 2.2, 2.4, 2.6 and 2.3 Å for 1-, 2-, and 3-pentanol and cyclopentanol, respectively. For the analogous ketones, α_O is 2.0, 2.3, 2.3, and 2.0 Å. From inspection of Eqs. 8 and 9 it is apparent that the magnitude of the G_P term arising from $k = k' = \text{oxygen}$ should inversely follow this trend, i.e., smaller values of α_k' lead to more negative values of G_P (note that $\gamma_{kk} = \alpha_k^{-1}$). In practice, this term accounts for about one-half of the differences in $\Delta G_{\text{ENP}}^{\circ}$ between these molecules, and it is the expected ordering which is observed. The only exception is cyclopentanone, which has a more negative $\Delta G_{\text{ENP}}^{\circ}$ than pentanal. This is due in part to the greater charge separation in the ketone from compression of the sp^2 carbon angle in the five-membered ring: $q_O = -0.38$ compared to $q_O = -0.35$ in pentanal, 2-pantanone, and 3-pantanone.

For the most part, the remaining differences in $\Delta G_{\text{ENP}}^{\circ}$ come from consideration of the oxygen–hydrogen cross terms in Eq. 8. These terms are destabilizing because of the opposite signs for the partial charges found on these two types of atoms. By moving the oxygen atom from one end of the molecule to the center, the majority of the γ_{OH} values decrease, resulting in increased destabilization. This effect is probably the least obvious of the three, but again follows the observed trends in overall $\Delta G_{\text{ENP}}^{\circ}$.

While the above trends are simple and intuitively reasonable, the experimental values for ΔG_s° do not necessarily behave in such a fashion. Indeed, 1-, 2-, and 3-pentanol are reported to have almost identical free energies of solvation. While this is disappointing and difficult to rationalize, we note that the difference in ΔG_s° for cyclopentanol compared to 3-pentanol agrees quite closely with experiment, illustrating once more that $\Delta\Delta G_s^{\circ}$ values are often more accurate because of cancellation of systematic deficiencies.

5.6. Equilibria

The models can also be tested against free energy of solvation changes in chemical equilibria. One might expect that the effect of solvation on an equilibrium might often be more accurate than the absolute free energies of solvation because some errors will cancel.

In a study using AM1-SM1, we considered 45 proton transfer equilibria and six isomerization equilibria. We have now recalculated all these with AM1-SM2 and PM3-SM3. The results are given in Table 9. (Note that all results in Ref. 2 for ammonia equilibria suffered from a 2.6 kcal/mol tabulation error. This error is corrected in Table 9.) For the proton transfer equilibria, AM1-SM1 yields a mean unsigned error of 3.6 kcal/mol (rms error of 4.4 kcal). These values have not

TABLE 9
FREE ENERGY OF SOLVATION CHANGES (KCAL/MOL) IN ACID-BASE EQUILIBRIA

Reaction	SM1	SM2	SM3	Expt
$\text{NH}_4^+ + \text{aniline} \rightarrow \text{NH}_3 + \text{aniline}\cdot\text{H}^+$	16.5	16.5	23.1	12.9
$\text{NH}_4^+ + \text{MeNH}_2 \rightarrow \text{NH}_3 + \text{MeNH}_3^+$	8.7	8.9	16.6	9.6
$\text{NH}_4^+ + \text{pyridine} \rightarrow \text{NH}_3 + \text{pyridine}\cdot\text{H}^+$	23.3	21.2	20.2	25.1
$\text{NH}_4^+ + \text{Me}_2\text{NH} \rightarrow \text{NH}_3 + \text{Me}_2\text{NH}_2^+$	18.0	15.7	17.3	17.6
$\text{NH}_4^+ + \text{Me}_3\text{N} \rightarrow \text{NH}_3 + \text{Me}_3\text{NH}^+$	26.1	23.6	23.1	23.7
$\text{NH}_4^+ + \text{PhCO}_2^- \rightarrow \text{NH}_3 + \text{PhCO}_2\text{H}$	143.8	138.7	141.8	137.3
$\text{NH}_4^+ + \text{AcO}^- \rightarrow \text{NH}_3 + \text{AcOH}$	149.0	142.9	144.4	146.9
$\text{NH}_4^+ + \text{PhO}^- \rightarrow \text{NH}_3 + \text{PhOH}$	140.6	134.1	138.3	141.1
$\text{NH}_4^+ + \text{Cp}^- \rightarrow \text{NH}_3 + \text{CpH}^b$	134.8	133.1	139.8	137.6
$\text{aniline}\cdot\text{H}^+ + \text{MeNH}_2 \rightarrow \text{aniline} + \text{MeNH}_3^+$	-7.7	-7.6	-6.5	-3.3
$\text{aniline}\cdot\text{H}^+ + \text{pyridine} \rightarrow \text{aniline} + \text{pyridine}\cdot\text{H}^+$	6.9	4.7	-2.9	12.2
$\text{aniline}\cdot\text{H}^+ + \text{Me}_2\text{NH} \rightarrow \text{aniline} + \text{Me}_2\text{NH}_2^+$	1.6	-0.8	-5.8	4.7
$\text{aniline}\cdot\text{H}^+ + \text{Me}_3\text{N} \rightarrow \text{aniline} + \text{Me}_3\text{NH}^+$	9.8	7.1	0.0	10.8
$\text{aniline}\cdot\text{H}^+ + \text{PhCO}_2^- \rightarrow \text{aniline} + \text{PhCO}_2\text{H}$	127.4	122.2	118.7	124.4
$\text{aniline}\cdot\text{H}^+ + \text{AcO}^- \rightarrow \text{aniline} + \text{AcOH}$	132.6	126.4	121.3	134.0
$\text{aniline}\cdot\text{H}^+ + \text{PhO}^- \rightarrow \text{aniline} + \text{PhOH}$	124.0	117.6	115.2	128.2
$\text{aniline}\cdot\text{H}^+ + \text{Cp}^- \rightarrow \text{aniline} + \text{CpH}$	118.5	116.6	116.7	124.7
$\text{MeNH}_3^+ + \text{pyridine} \rightarrow \text{MeNH}_2 + \text{pyridine}\cdot\text{H}^+$	14.6	12.3	3.6	15.5
$\text{MeNH}_3^+ + \text{Me}_2\text{NH} \rightarrow \text{MeNH}_2 + \text{Me}_2\text{NH}_2^+$	9.4	6.8	0.7	8.0
$\text{MeNH}_3^+ + \text{Me}_3\text{N} \rightarrow \text{MeNH}_2 + \text{Me}_3\text{NH}^+$	17.4	14.7	6.5	14.1
$\text{MeNH}_3^+ + \text{PhCO}_2^- \rightarrow \text{MeNH}_2 + \text{PhCO}_2\text{H}$	135.2	129.8	125.2	127.7
$\text{MeNH}_3^+ + \text{AcO}^- \rightarrow \text{MeNH}_2 + \text{AcOH}$	140.4	134.0	127.8	137.3
$\text{MeNH}_3^+ + \text{PhO}^- + \text{MeNH}_2 + \text{PhOH}$	131.7	125.2	121.7	131.6
$\text{MeNH}_3^+ + \text{Cp}^- \rightarrow \text{MeNH}_2 + \text{CpH}$	126.2	124.2	123.2	128.0
$\text{pyridine}\cdot\text{H}^+ + \text{Me}_2\text{NH} \rightarrow \text{pyridine} + \text{Me}_2\text{NH}_2^+$	-5.3	-5.5	-2.9	-7.6
$\text{pyridine}\cdot\text{H}^+ + \text{Me}_3\text{N} \rightarrow \text{pyridine} + \text{Me}_3\text{NH}^+$	2.9	2.4	2.9	-1.4
$\text{pyridine}\cdot\text{H}^+ + \text{PhCO}_2^- \rightarrow \text{pyridine} + \text{PhCO}_2\text{H}$	120.6	117.5	121.6	112.2
$\text{pyridine}\cdot\text{H}^+ + \text{AcO}^- \rightarrow \text{pyridine} + \text{AcOH}$	125.8	121.7	124.2	121.8
$\text{pyridine}\cdot\text{H}^+ + \text{PhO}^- \rightarrow \text{pyridine} + \text{PhOH}$	117.1	112.9	118.1	116.0
$\text{pyridine}\cdot\text{H}^+ + \text{Cp}^- \rightarrow \text{pyridine} + \text{CpH}$	111.6	111.9	119.6	112.5
$\text{Me}_2\text{NH}_2^+ + \text{Me}_3\text{N} \rightarrow \text{Me}_2\text{NH} + \text{Me}_3\text{NH}^+$	8.2	7.9	5.8	6.2
$\text{Me}_2\text{NH}_2^+ + \text{PhCO}_2^- \rightarrow \text{Me}_2\text{NH} + \text{PhCO}_2\text{H}$	125.8	123.0	124.5	119.7
$\text{Me}_2\text{NH}_2^+ + \text{AcO}^- \rightarrow \text{Me}_2\text{NH} + \text{AcOH}$	131.1	127.2	127.1	129.3
$\text{Me}_2\text{NH}_2^+ + \text{PhO}^- \rightarrow \text{Me}_2\text{NH} + \text{PhOH}$	122.4	118.4	121.0	123.5
$\text{Me}_2\text{NH}_2^+ + \text{Cp}^- \rightarrow \text{Me}_2\text{NH} + \text{CpH}$	116.9	117.4	122.5	121.0
$\text{Me}_3\text{NH}^+ + \text{PhCO}_2^- \rightarrow \text{Me}_3\text{N} + \text{PhCO}_2\text{H}$	117.6	115.1	118.7	113.7
$\text{Me}_3\text{NH}^+ + \text{AcO}^- \rightarrow \text{Me}_3\text{N} + \text{AcOH}$	122.9	119.3	121.3	123.2
$\text{Me}_3\text{NH}^+ + \text{PhO}^- \rightarrow \text{Me}_3\text{N} + \text{PhOH}$	114.3	110.5	115.2	117.4
$\text{Me}_3\text{NH}^+ + \text{Cp}^- \rightarrow \text{Me}_3\text{N} + \text{CpH}$	108.7	109.5	116.7	114.9
$\text{PhCO}_2\text{H} + \text{AcO}^- \rightarrow \text{PhCO}_2^- + \text{AcOH}$	5.3	4.2	2.6	9.4
$\text{PhCO}_2\text{H} + \text{PhO}^- \rightarrow \text{PhCO}_2^- + \text{PhOH}$	-3.5	-4.6	-3.5	3.7
$\text{PhCO}_2\text{H} + \text{Cp}^- \rightarrow \text{PhCO}_2^- + \text{CpH}$	-8.9	-5.6	-2.0	1.3
$\text{AcOH} + \text{PhO}^- \rightarrow \text{AcO}^- + \text{PhOH}$	-8.6	-8.8	-6.1	-5.7
$\text{AcOH} + \text{Cp}^- \rightarrow \text{AcO}^- + \text{CpH}$	-14.2	-9.8	-4.6	-8.1
$\text{PhOH} + \text{Cp}^- \rightarrow \text{PhO}^- + \text{CpH}$	-5.5	-1.0	1.5	-2.4
<i>Mean unsigned error</i>	3.6	3.8	5.4	
<i>RMS error</i>	4.4	4.5	6.6	

TABLE 10
 $\Delta\Delta G_s^\circ$ (KCAL/MOL) FOR TAUTOMERIC EQUILIBRIA

Reaction	SM1	SM2	SM3	Expt
gas-phase geometries and wave function				
I 5-hydroxy-2-pentanone → lactol	3.6	3.0	3.4	
II 4-hydroxypyridine → 4-pyridone	-3.3	-2.0	-2.3	
III 2-hydroxypyridine → 2-pyridone	-1.8	-0.6	-1.5	
IV ethyl acetoacetate → enol	1.1	-1.3	0.5	
V acetylacetone → enol	0.9	-1.3	0.3	
geometry and wave function reoptimized				
I	4.1	3.6	4.4	> 4.0
II	-8.7	-7.6	-9.2	< -8.0
III	-4.4	-2.6	-4.3	-4.3
IV	1.1	-1.3	1.2	1.4
V	0.9	-1.6	0.6	1.0

changed very much with the new methods, becoming 3.8(4.5) kcal/mol for AM1-SM2 and 5.4(6.6) kcal/mol for PM3-SM3. (In each case, the mean unsigned error is followed by the rms error in parentheses.)

The results for keto-lactol, keto-enol, and hydroxypyridine-pyridone equilibria are given in Table 10. Moreover, Table 10 illustrates the importance of solvent-induced reorganization, primarily of the electronic structure but also of the nuclear geometry, on the individual free energies of solvation. Reaction I, the closure of a hydroxyketone to a lactol, is disfavored in the water phase compared to the gas phase (see Ref. 2 for experimental references for these equilibria). Clearly, the separated ketone and alcohol functionalities are better solvated than the single acetal. All three models compare reasonably with experiment, where only a lower bound on $\Delta\Delta G_s^\circ$ has been established. The bulk of the effect from reoptimization of the geometry and wave function is found for the hydroxyketone. This molecule prefers to be internally hydrogen-bonded even in aqueous solution, but the charge distribution relaxes considerably from the vapor-phase, hydrogen-bonded structure upon immersion in the continuum dielectric. The resulting improvement in ΔG_s° accounts for the 0.5–1.0 kcal/mol increase in $\Delta\Delta G_s^\circ$.

A similar situation occurs for reactions II and III; however, the pyridones are much more easily polarized than their hydroxypyridine tautomers, especially 4-pyridone, which develops a very large charge separation. Since this effect is discussed in detail in Section 6, we note here only that in the absence of wave-function relaxation, the fairly close agreement with experiment would be completely lost.

For reactions IV and V, both reactants and products show similar effects from reoptimization of the geometry and wave function, thus leading to small net changes in $\Delta\Delta G_s^\circ$. PM3-SM3, which appears to be more flexible in the degree to which it allows electronic reorganization, gives the largest effects. SM1 and SM3 both compare well with experiment in these cases, while SM2 is somewhat low.

In general, when compared to the gas-phase geometries and wave functions, more than 90% of

the change in ΔG_s° for any given molecule comes from just the electronic relaxation. Logistically speaking, this implies that a 1SCF calculation with one of the Solvation Models at the gas-phase geometry allows very rapid, qualitative analysis of the important solvation effects. Additional changes from geometry relaxation are generally much smaller for stable, ground-state molecules. Of course, for transition states and unusual structures, geometric changes may be more significant.

Table 11 gives results for the conformational equilibrium of N-methylacetamide. Table 11 shows that the change in solvation free energy for rotation about the peptide linkage is in good agreement with experimental observations [46]. In particular, the change is small compared to the magnitude, which is about 10 kcal/mol. This is a prototype of the peptide bond conformational problem. Interestingly, Jorgensen and Gao [47], with 216 explicit water molecules, calculated a $\Delta\Delta G_s^\circ$ value for E → Z (cis peptide linkage → trans peptide linkage) of 2.2 kcal/mol with standard OPLS charges, but 0.1 kcal/mol with parameters adjusted specifically for this reaction, showing the sensitivity of this problem to model parameters. Yu et al. [48] re-examined the problem with interaction-site-model integral equation calculations and calculated $\Delta\Delta G_s^\circ$ values of 1.3, 2.0 and 2.0 kcal/mol for three different approximations within the integral equation context. It is encouraging that the SM2 and SM3 values of 1.0 and 0.9 kcal/mol, respectively, for $\Delta\Delta G_s^\circ$ in Table 11 are in better agreement with experimental observations than most of these more computationally intensive calculations.

The other equilibria in Table 11 are simple trans → gauche internal rotations and exhibit the expected direction of effect, although the magnitude of the solvation effect is smaller for n-butane than might be expected. In particular, our AM1-SM2 value of -0.03 kcal/mol may be compared to -0.07 kcal/mol obtained by Jorgensen [49] with the TIP4P model of water. Perhaps we should emphasize that the two calculations agree within 0.04 kcal/mol. Earlier values in the literature include -0.18 kcal/mol [50] and -0.10 kcal/mol [51] for simulations with the ST2 and TIPS2 water models, respectively.

An interesting question is whether or not the present model is accurate enough to predict pK_a

TABLE 11
 $\Delta\Delta G_s^\circ$ (KCAL/MOL) FOR ROTAMERIC EQUILIBRIA

Reaction	SM1	SM2	SM3	Expt
gas-phase geometries and wave function				
I E-N-methylacetamide → Z	0.9	0.5	0.6	
II <i>trans</i> 1-propanol → <i>gauche</i>	0.4	0.4	0.3	
III <i>trans</i> butane → <i>gauche</i>	-0.01	-0.03	-0.02	
geometry and wave function reoptimized				
I	1.5	1.0	0.0	~0.1 ^a
II	0.4	0.4	0.4	
III	-0.01	-0.03	-0.02	

^a Radzicka et al., Reference 46.

values in aqueous solution. Unfortunately, this is a difficult problem. For example to predict a pK_a value within 0.5 units at 298 K requires an accuracy of 0.7 kcal/mol. Although the solvation energies of the models under review here are often this accurate, the AM1 or PM3 solute energetics are seldom this good.

5.7. Explicit water molecules

While approaches which treat the solvent as a surrounding continuum are appropriate in a variety of instances, there are certainly numerous occasions where one or more explicit water molecules may be crucial to accurate modeling of a given chemical process. Ideally, one would like to pluck one or more water molecules out of the continuum and include it or them explicitly in the first hydration shell, thereby bridging the gap between the all-continuum and the all-explicit-molecule approaches to treating the solvent. However, it is clear that there may be difficulties in this technique if explicit molecular interaction energies of the water molecules are not predicted accurately enough in the enlarged solute.

At first, one might think that adding explicit waters is a *general* way to improve accuracy. Thus, one might add more and more waters until only the second and further hydration shells are treated as a continuum, then continue until only the third and further hydration shells are treated as a continuum, and so forth. However, inclusion of explicit waters should not always be expected to improve the accuracy. This is easily understood when one considers the high accuracy attainable with continuum models and how difficult it is to insure such high accuracy for explicit intermolecular interaction calculations. In addition, systems with many explicit water molecules certainly require statistical averaging over a large number of low-energy structures. Thus, our motivation for improving the accuracy of our parameterization for an explicit water was not because we believe that adding explicit waters is a general way to improve the accuracy (although,

TABLE 12
CALCULATED FREE ENERGIES OF SOLVATION (KCAL/MOL) FOR SOLUTES WITHOUT AND WITH EXPLICIT WATER MOLECULES IN THE FIRST HYDRATION SHELL

Compound	SM2			SM3		
	Interaction ^a	ΔG°	Net ΔE^b	Interaction ^a	ΔG°	Net ΔE^b
$H_2O + H_2O^c$	0.0	-12.6	-12.6	0.0	-12.6	-12.6
$H_2O \cdot H_2O$	-5.0	-9.3	-14.3	-3.5	-10.0	-13.5
acetone + H_2O^c	0.0	-10.4	-10.4	0.0	-11.0	-11.0
acetone- H_2O	-4.5	-6.9	-11.4	-3.6	-7.7	-11.3
piperazine + $2H_2O^c$	0.0	-20.4	-20.4	0.0	-18.7	-18.7
piperazine- $H_2O + H_2O^c$	-3.2	-16.9	-20.1	-2.9	-16.1	-19.0
piperazine- $2H_2O$	-6.5	-13.6	-20.1	-5.9	-13.3	-19.2

^a Gas phase ΔH_f (complex) - ΔH_f (separated educts).

^b Overall energy change for solvated species relative to the infinitely separated, gas-phase educts.

^c Infinitely separated.

as discussed further below, if one observes a certain insensitivity of the solvation free energy upon adding explicit water molecules, it certainly does increase one's confidence, as it should). Our motivation is different: sometimes one wants to add explicit waters to increase the *understanding* of the behavior of first hydration shell solvent molecules or to check for *unusual* specific bonding effects. In such cases, if water itself were not accurate, then one would be hard pressed to draw conclusions from the results.

We have investigated the quantitative aspects of including explicit water molecules for the water dimer, monohydrated acetone, and both mono- and dihydrated piperazine. The results are presented in Table 12. As observed elsewhere [8,30–32], the AM1 complexes were characterized by long, bifurcated hydrogen bonds, whilst the PM3 complexes evidenced shorter, linear hydrogen bonds. Since we are primarily concerned with the energy partitioning, only moderate effort was put into ensuring that the supermolecule structures examined here are global minima on the hypersurface, so the quantitative values presented should not be regarded as definitive.

The strategy is reasonably successful. For the water dimer, the SM2 net solvation energy is larger than the sum of the solvation energies for two isolated water molecules. However, the majority of the error does not arise from the solvation model, but rather from the overestimation of the gas-phase interaction energy by AM1 (accurate value: –3.6 kcal/mol [52]). Given this precedent, it is reasonable to assume that the acetone-water complexation energy is also overestimated. SM3 delivers more consistent energies in these two instances. On the other hand, piperazine with explicit waters of hydration seems to be well treated within both models.

A noteworthy consideration in adding explicit waters is the way that it changes the dependence of the predicted solvation energy on the various parameters. In general, the most important parameters are those for the atoms close to the surface of the solute. When we add explicit waters, we 'bury' some surface atoms, and we put water oxygens and hydrogens on the surface of the new complex. If the results are reasonably insensitive to this, one cannot help but gain more confidence in the predictions.

5.8. Direct comparison of models

When we compare AM1-SM1 and AM1-SM2, we see a strong systematic improvement in SM2. The situation is similar in some respects to MNDO vs. AM1, where AM1 appears clearly superior. Nevertheless, it is sometimes useful to apply MNDO just because if it agrees well with AM1, confidence is increased slightly. SM1 can probably play a similar role with respect to SM2; in fact, it may be even more useful in this respect because of the entirely different way that hydrogens are treated in the solvent-accessible surface areas in the two theories.

The situation with SM1a is somewhat different. For applications involving only neutrals with typical hybridizations, SM1a gives good results, especially if we exclude alkanes and nitrogen-containing compounds, and it may even be the preferred method for some applications. However, SM1a has a limited applicability, as discussed above.

The comparison of AM1-SM2 to PM3-SM3 is trickier. In general, AM1-SM2 appears to be the better method, and PM3-SM3 shows serious problems for certain classes of nitrogen-containing compounds. Nevertheless, except for these nitrogen-containing compounds, the final decision may be based in many cases on whether AM1 or PM3 provides a better description of the solute, since the average solvation energy errors are similar.

Hydrogen bonding to explicit waters is one area of especially clear difference between AM1 and

PM3. When the final word is in about when each of the two methods is more reliable for hydrogen bonding, it may be a determining factor in choosing a method for many systems.

6. DISCUSSION

6.1. Solvation terms

Our solvation models display certain performance characteristics which are worthy of particular emphasis in comparison to other available approaches. In particular, several versions of the self-consistent reaction field model evaluate polarization energy as the interaction of the surrounding continuum dielectric with a multiple expansion of the solute electron density truncated after the dipole term [e.g., 20e, 20q-s]. With such an approach, symmetrical uncharged molecules characterized by the presence of an inversion center must give zero polarization energies. From inspection of Tables 5 and 8 and such examples as *p*-xylene, 1,4-dioxane, piperazine, *N,N'*-dimethyl-piperazine, tetrafluoromethane, and *E*-1,2-dichloroethene, it is clear that such an approximation is of limited utility for many molecules.

We wish to stress several aspects of the present treatment that we believe are important in a method intended for general applications: (i) The solute charge distribution is represented by distributed monopoles, not a single-center multipole expansion, which can be slowly convergent, or even a single dipole moment, which can represent a drastic oversimplification, especially for large solutes. (ii) The dielectric screening algorithm does not assume a spherical [e.g., 17, 20j, 20l, 20o, 20q-s] or ellipsoidal [e.g., 18] cavity but allows arbitrarily complex shapes. (iii) The electric polarization terms are included in the solute Hamiltonian self-consistently. (iv) Our approach models, simultaneously, electric polarization and solvent-accessible surface area terms. (v) The surface tensions depend on the local nature of the solute at the solute-solvent interface. Although none of these features is unique when considered separately, the methods under review here are apparently unique in including all of these features simultaneously.

Are the parameters of the new solvation models physical? We think so, although interpreting semiempirical parameters physically can be very subjective. The surface tensions probably merit

TABLE 13
AMI-SM2 SOLVENT ACCESSIBLE SURFACE TENSIONS (CAL MOL⁻¹ Å⁻²) FOR COMMONLY OCCURRING GROUPS

X	XH ₀	XH ₁	XH ₂	XH ₃	XH ₄
C	3.36	6.75	7.53	7.83	7.98
N	-30.70	-50.81	-55.46	-25.61	-15.37
O	-25.01	-64.05	-36.09	-35.60	
F	22.50 ^a				
P	3.90 ^a				
S	-53.25	-14.47	-5.49	-2.12	
Cl	-2.28 ^a				
Br	-8.15 ^a				
I	-15.18 ^a				

^a For these elements, the surface tension is independent of bond order to H.

some discussion in this regard. To facilitate this discussion, Table 13 gives AM1-SM2 surface tensions for ideal groups, i.e., groups with integer values of B_{kH} , as computed from

$$\sigma_k(B_{kH}) = \sigma_k^{(0)} + \sigma_k^{(1)}[f(B_{kH}) + g(B_{kH})] \quad (31)$$

along with Eqs. 26 and 27. Let us consider these. As pointed out previously [1], the surface tensions in this kind of model will to some extent make up for systematic deficiencies in the ENP term due to errors in the electronic charge distribution of the solute, e.g., errors due to the restricted flexibility of the minimum basis set. Nevertheless, we find that the major trends in the surface tensions can still be understood in terms of the CDS effects that they are designed to model.

Methyl, methylene, and methine all have similar surface tensions, all close to the universal value $7.2 \text{ cal mol}^{-1} \text{ Å}^{-2}$ used by Still et al. [16]. The value for CH_0 is not very important since quaternary carbon typically has no solvent-accessible surface area.

The values for nitrogen are also reasonable. The large negative values for $-\text{NH}$ and $-\text{NH}_2$ reflect the tendency of these groups to show favorable hydrogen bonding with water. This is a good place to emphasize that the changes in solvent structure included in G_{CDS} include hydrogen bonding of the solvent to the solute (or at least the energetic effect that is not included in the screened electrostatics of the G_p term) as well as changes in the solvent-solvent hydrogen bonds. Both effects are reasonably modeled as proportional to the number of water molecules in the first hydration shell, which in turn is reasonably modeled as proportional to the solvent-accessible surface area. The smaller values of the surface tensions for NH_3 and NH_4^+ , which occur in far fewer molecules and ions, reflect the typically larger solvent-accessible surface areas of these groups.

For oxygen, we find a moderate hydrophilic effect for ethers and carbonyls, a much larger hydrophilic effect for alcohols and other $-\text{OH}$ groups, and smaller surface tensions for H_2O and H_3O^+ , which have solvent accessible surface areas larger than organic oxygen-containing groups.

Fluorine is hydrophobic as expected. The phosphorus value is based on very little data, but is reasonable. The sulfur values are hardest to interpret, but Table 5 does show that net small hydrophilic G_{CDS} terms do improve the accuracy for thiols and sulfides. The trend in the halogens is consistent with increasing dispersion interactions as polarizability increases.

The optimized values of the surface tensions are of course coupled to the optimized atomic radii in Table 1. The AM1-SM1 values are the widely accepted atomic radii of Bondi [53]. The AM1-SM2 values are semiempirical values, and they would be expected to be about the same (or slightly larger because H has no volume in AM1-SM2). In fact, one value is smaller by 2.4%, two are the same, and six are larger by 0.6% to 11.9%. Notice that the radii were only optimized to the nearest 0.1 Å, and we did not add a dependence on hybridization or chemical environment to the radii. The fact that the SM2 group radii, nevertheless, correlate with conventional atomic radii within about 10% indicates that the physical cavity sizes in our model are reasonable, but we caution against interpreting these radii or cavity sizes too closely since the semiempirical parameterization of the solvation models under review here is in terms of macroscopic solvation free energies, not solute sizes. In this spirit, we remind the reader of the difficulty of interpreting AM1 parameters in terms of orbitals since the AM1 model is parameterized to experimental data that include correlation effects. Semiempirical molecular orbital theories parameterized to energies including correlation effects can, in principle, be interpreted [54] in an independent-particle SCF orbital formalism by using projection operators, and, again in principle, projection operator techniques [55]

might be employed to interpret effective solute properties in a few-body model parameterized to include many-body solvation effects. However, it is not clear whether practical insight would result, and, in any event, such an analysis, which would be difficult, has not been carried out.

6.2. Solvent-induced solute charge reorganization

As mentioned above, one critical factor of our semiempirical approach is its incorporation of solvent effects directly into the Fock operator [12], thereby allowing the solute electronic structure to reorganize in response to the solvent. The results of this reorganization may be quite profound, both in terms of overall solvation free energy and of final electronic (and perhaps nuclear) structure. This effect has also been studied with other SCF solvation algorithms [11,12,14,18,20], but we wish to comment on the importance of these effects in the solvation models reviewed here. The present formalism should be especially well suited for treating such effects not only in small, simple solutes but also in large, complicated molecules.

Figure 1 shows the Mulliken-derived atomic charges for 4-pyridone calculated from both the reorganized solute nuclear and electronic structure and from the frozen, gas-phase molecule. We have shown elsewhere that the experimental limit for the hydroxypyridine-pyridone equilibrium free energy of solvation change is well reproduced by AM1-SM1 calculations [2], and similarly good results are obtained with AM1-SM2 and PM3-SM3 (see Reaction II of Table 10). One is immediately struck by the dramatic increase in the dipole moment of this molecule upon electronic relaxation, as judged by the migration of negative charge from the N-H region to the carbonyl. This reorganization costs energy relative to the optimized gas-phase distribution; however, it is offset by sufficient polarization energy to roughly double the net overall solvation energy! Of course, by its nature 4-pyridone is particularly susceptible to such an effect, i.e., it has a large intrinsic dipole and is quite polarizable. However, a number of biologically important molecules are similar in this respect, e.g., purines and pyrimidines. Solvation models which employ gas-phase (typically ab initio) charges and do not allow for solvent-induced relaxation may fail to account for significant portions of the solvation free energies.

SM2 and SM3 predict identical $\Delta G_{\text{ENP}}(\text{aq})$ values for 4-pyridone, -13.7 kcal/mol , but the underlying details are different. $G_p(\text{aq})$ for the gas-phase geometry and wave functions is -7.1 kcal/mol for SM2 and -5.4 kcal/mol for SM3. In SM2, electronic and structural reorganization costs 11.6 kcal/mol , yielding a relaxed-structure $G_p(\text{aq})$ of -25.3 kcal/mol , whereas in SM3 it costs

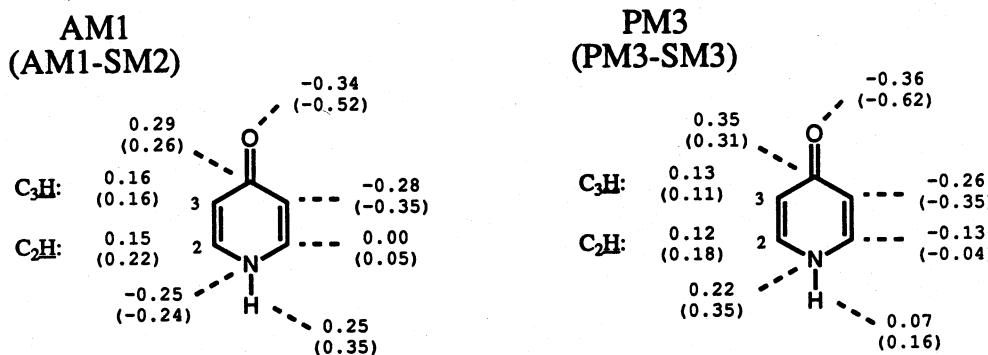


Fig. 1. Effects of solvent-induced electronic reorganization on the atomic charges of 4-pyridone.

20.1 kcal/mol, yielding a final $G_p(aq)$ of -33.8 kcal/mol. Evidently SM3 allows a much greater distortion of the charge distribution. Similar effects of greater solvent-induced distortion in SM3 are seen in butadiene and the alkynes.

Since this effect is so important, it is interesting to analyze a case in greater detail, in particular to sort out geometry changes and changes in CDS terms from changes in the electric polarization. An excellent example of biological interest is deoxycytidine (see Table 8). This nucleoside has an AM1 gas-phase dipole moment of 7.5 D, and using the gas-phase geometry and wave function, AM1-SM2 predicts ΔG_s^o to be -22.6 kcal/mol, of which -14.2 and -8.4 kcal/mol, respectively, derive from G_{CDS}^o and ΔG_{ENP} (the latter term is simply G_p under such circumstances since $\Delta G_N = 0$).

If the geometry of deoxycytidine is kept frozen, but the electronic structure is allowed to relax, G_{CDS}^o remains unchanged. (While at first this might seem a requirement, recall that in SM2 and SM3, the G_{CDS}^o contributions from C, N, O and S depend on hydrogen-bond orders – were these to change as a result of electronic relaxation, G_{CDS}^o could be affected.) Instead, however, the evidence for considerable charge reorganization is found in the increase in the dipole moment to 10.2 D, and in the nearly doubled value of G_p , -16.6 kcal/mol. After accounting for the 4.9 kcal/mol cost of electronic reorganization (i.e., ΔG_E since we are still considering a fixed geometry; ΔG_N continues to be zero), ΔG_{ENP} is -11.8 kcal/mol, giving an AM1-SM2 ΔG_s^o value of -25.9 kcal/mol. Finally, if the geometry is also allowed to relax, the change is insufficient to affect G_{CDS}^o , but rather manifests itself in allowing still greater charge polarization. The very polarizable nature of the nucleoside causes the contribution from nuclear relaxation to be larger than is typically observed for other molecules we have studied. In particular, the dipole moment increases to 10.8 D, and G_p becomes -19.0 kcal/mol; after accounting for 6.5 kcal/mol of reorganization cost (both ΔG_E and ΔG_N now), this gives a ΔG_{ENP} of -12.5 kcal/mol. The final ΔG_s^o value is thus -26.7 kcal/mol. So, the ΔG_{ENP} value increases by 40% from simply electronic relaxation, relative to the gas phase, and by 48% from combined nuclear and electronic relaxation.

Finally, it is noteworthy that in 4-pyridone the increase in ΔG_{ENP} for the analogous process is a more sizable 93%, in spite of the fact that the overall change in dipole, 6.3 to 10.8 D, is fairly similar. Another comparison is provided by the base cytosine (i.e., deoxycytidine without the pentose), which shows a 71% increase in its AM1-SM2 derived ΔG_{ENP} on going from a gas-phase geometry and wave function to a fully relaxed solute. The difference between deoxycytidine and cytosine is attributable to the less polarizable deoxyribose moiety, which displaces a large volume of the dielectric medium, thereby shielding a portion of the pyrimidine from the solvent.

Similar results have been obtained for the other four nucleic acid bases as well, with ΔG_{ENP} calculated by the AM1-SM2 method increasing 54–76% due to solvent-induced geometry and wave function relaxation.

Comparable results are obtained for deoxycytidine and the nucleic acid bases with the PM3-SM3 model, although again greater overall polarization is noted for PM3-SM3 than for AM1-SM2. For example, for deoxycytidine, in going from the gas-phase solute to the frozen-geometry/relaxed-wave function solute, the dipole moment increases from 5.8 to 8.8 D, and it further increases to 9.8 D upon full relaxation. The G_{CDS}^o term remains constant at -16.3 kcal/mol, and G_p goes from -7.1 to -17.0 to -22.2 kcal/mol. Factoring in reorganization costs of 0.0, 6.2, and 10.1 kcal/mol, respectively, this gives ΔG_{ENP} values of -7.1, -10.8, and -12.1 for net ΔG_s^o values of -23.4, -27.1, and -28.4 kcal/mol.

6.3. Usefulness of the models

Some comments should be made regarding the specific objectives of the SM2 and SM3 models. Each has been developed with a view towards generality and balance, and incorporates information into its parameters derived from a wide variety of molecules and functionalities. However, for narrowly focused investigative purposes, it may be desirable to modify the parameters to describe more adequately a specific class of compounds, an approach we call SRP, for Specific Reaction Parameters. An excellent example of when such an approach might be warranted is provided by the ethers, where the SM2 and SM3 models tend to underestimate $-\Delta G^\circ_s$. In a study aimed at modeling the aqueous hydrolysis of these species, for example, one might modify the oxygen σ_k values to more adequately reflect known experimental data. Alternatively, one might introduce a new σ_k which applies only to oxygen atoms that are singly bonded to two unique carbon atoms, as judged by analysis of the bond order matrix. (Of course, depending on the scope of the investigation, another conditional might have to be added to differentiate ether oxygens from ester oxygens, etc.) Alternatively, one could introduce empirical ‘corrections’ to the charges. We have tried to design the solvation models to be sufficiently flexible for such specialized partial reparameterizations to be made with a minimum of effort.

One area of application that we have not discussed yet is transition-state modeling. In fact, the goal of transition-state modeling was one of the primary objectives of our initially taking up this project. We believe that the methods have reached a stage of optimization for stable molecules where the testing and applications for transition states should now receive greater attention, and we hope other researchers will join us in this. We think it is useful to mention two areas of concern in this regard.

First, one should expect somewhat less accuracy for transition states because they are farther from the set of compounds for which both the solute Hamiltonians (AM1 and PM3) and the solvation models were parameterized. For example, transition states may lie in the poorly calibrated inter-subrange regions of charge discussed below Eq. 13.

Second, we note that even for gas-phase reactions, it is often necessary to use specific reaction parameters for quantitative modeling of potential energy surfaces [56]. One difficulty with modifying the solute Hamiltonians to improve the unsolvated energetics is that it may lead to less realistic charges so that the solvation model is spoiled, and users contemplating such approaches should be wary of this.

This brings up a more general point. One possible area for future work would be to incorporate a different solute parameterization that gives more accurate charges. This is an especially promising avenue for improvement since a previous study of AM1 charges has shown some deficiencies. In particular, Besler et al. [57] found that although the AM1 parameterization generally yields better dipole moments than the MNDO [58] model, AM1 charges do not correlate as well as MNDO charges with charges extracted by fitting 6–31G* ab initio electrostatic potentials. Thus the process of optimizing the AM1 parameters to improve the energies may have exacted a toll from the accuracy of the charges. Similarly, we have discussed above how PM3 charges are sometimes significantly less realistic than AM1 charges, presumably for the same reasons. Thus it would be desirable to monitor the charges more carefully in any new general parameterizations that may be planned for solute (i.e., gas-phase) Hamiltonians. In fact, what would probably be advisable would be to (heroically) parameterize to gas-phase energetics and solvation free energies simultaneously.

Although Eq. 1 yields absolute free energies in aqueous solution, the free energy of hydration defined by subtraction, as in Eq. 1, is more useful for calibrating and testing the theory. We note though that for many molecules and ions, especially biomolecules, the gas-phase values are unknown or even unobtainable, and ultimately one will be interested in the absolute free energies. Thus a new parameterization of the solute that yields improved solute *energies* would also be very desirable.

7. CONCLUSION

We believe that the SM2 and SM3 models are reasonable compromises between accuracy, generality, and flexibility. The accuracy for solvation free energies is comparable to that for much more expensive simulations based on an explicit many-body representation of the solvent, even in cases where solvent-induced charge redistribution of the solute is negligible, and the new method – unlike the typical many-body simulation – is also realistic when such charge redistribution is not negligible. Furthermore, the errors in treating the solvation energy with these models are already at least of the order of the errors in the solute energies from the semiempirical electronic structure calculations, and usually considerably less. Hence, more accurate refinements are probably of limited value, except in specialized circumstances where improved accuracy may be required for a small subset of molecules or functionalities. As another consequence, it will sometimes be possible to obtain improved accuracy by combining the present semiempirical techniques for solvation free energies with ab initio calculations of gas-phase solute energy changes, at least in cases where the semiempirical geometries are more accurate than the semiempirical solute energies.

ACKNOWLEDGMENTS

The authors are grateful to Carmay Lim for helpful discussions and references. This work was supported in part by the National Science Foundation through grant no. CHE89-922048. Additional funding was generously provided by the USA-CRDEC ILIR program.

REFERENCES

- 1 Cramer, C.J. and Truhlar, D.G., *J. Amer. Chem. Soc.*, 113 (1991) 8305, 9901 (E).
- 2 Cramer, C.J. and Truhlar, D.G., *J. Amer. Chem. Soc.*, 113 (1991) 8552, 9901 (E).
- 3 Cramer, C.J. and Truhlar, D.G., *Science*, 256 (1992) 213.
- 4 Cramer, C.J. and Truhlar, D.G., *J. Comput. Chem.*, in press.
- 5 Pople, J.A. and Segal, G.A., *J. Chem. Phys.*, 43 (1965) S129.
- 6 Pople, J.A. and Beveridge, D.L., *Approximate Molecular Orbital Theory*, McGraw-Hill, New York, 1970.
- 7 a. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., *J. Amer. Chem. Soc.*, 107 (1985) 3902.
b. Dewar, M.J.S. and Zoebisch, E.G., *J. Mol. Struct. (Theochem)*, 180 (1988) 1.
c. Dewar, M.J.S. and Jie, C., *J. Mol. Struct. (Theochem)*, 187 (1989) 1.
d. Dewar, M.J.S. and Yuan, Y.C., *Inorg. Chem.*, 29 (1990) 3881.
- 8 a. Stewart, J.J.P., *J. Comput. Chem.*, 10 (1989) 209; 221.
b. Stewart, J.J.P., *J. Comput.-Aided Mol. Des.*, 4 (1990) 1.
- 9 a. Born, M., *Physik*, Z., 1 (1920) 45.
b. Rashin, A.A. and Honig, B., *J. Phys. Chem.*, 89 (1985) 5588.

- 10 a. Hoijtink, G.J., de Boer, E., van der Meij, P.H. and Weijland, W.P., *Recl. Trav. Chim. Pays-Bas.* 75 (1956) 487.
b. Peradejordi, F., *Cah. Phys.*, 17 (1963) 393.
- 11 a. Jano, I., *Compt. Rend. Acad. Sci. (Paris)*, 261 (1965) 103.
b. Fischer-Hjalmars, I., Hendriksson-Enflo, A. and Hermann, C., *Chem. Phys.*, 24 (1977) 167.
- 12 Tapia, O., In Daudel, R., Pullman, A., Salem, L. and Veillard, A. (Eds.) *Quantum Theory of Chemical Reactions*. Vol. II, Reidel, Dordrecht, 1980, p. 25.
- 13 Costanciel, R. and Contreras, R., *Theor. Chim. Acta*, 65 (1984) 1.
- 14 Kozaki, T., Morihashi, M. and Kikuchi, O., *J. Amer. Chem. Soc.*, 111 (1989) 1547.
- 15 Tucker, S.C. and Truhlar, D.G., *Chem. Phys. Lett.*, 157 (1989) 164.
- 16 Still, W.C., Tempczak, A., Hawley, R.C. and Hendrickson, T., *J. Amer. Chem. Soc.*, 112 (1990) 6127.
- 17 Onsager, L., *J. Amer. Chem. Soc.*, 58 (1936) 1486.
- 18 See, e.g., Rivail, J.L., In Bertrán, J. and Csizmadia, I.G. (Eds.) *New Theoretical Concepts for Understanding Organic Reactions*, Kluwer, Dordrecht, 1989, p. 219; Rinaldi, D., Ruiz-Lopez, M.F. and Rivail, J.L., *J. Chem. Phys.*, 78 (1983) 834; Rivail, J.L., Terryn, B., Rinaldi, D. and Ruiz-Lopez, M.F., *J. Mol. Struct. (Theochem)*, 120 (1985) 387.
- 19 Mulliken, R.S., *J. Chem. Phys.*, 23 (1955) 1833.
- 20 a. Newton, M.D., *J. Phys. Chem.*, 79 (1975) 2795.
b. Tapia, O. and Goscinski, O., *Mol. Phys.*, 29 (1975) 1653.
c. McCreery, J., Christofferson, R.E. and Hall, G.C., *J. Amer. Chem. Soc.*, 98 (1976) 7191.
d. Rivail, J.-L. and Rinaldi, D., *Chem. Phys.*, 18 (1976) 233.
e. Warshel, A. and Levitt, M., *J. Mol. Biol.*, 103 (1976) 227.
f. Klopman, G. and Andreozzi, P., *Theor. Chim. Acta*, 55 (1980) 77.
g. Miertus, S., Scrocco, E. and Tomasi, J., *J. Chem. Phys.*, 55 (1981) 117.
h. Bonaccorsi, R., Cimaraglia, R. and Tomasi, J., *J. Comput. Chem.*, 4 (1983) 567.
i. Contreras, R. and Gomez-Jeria, J.S., *J. Phys. Chem.*, 88 (1984) 1905.
j. Mikkelsen, K.V., Dalgaard, E. and Swanstrøm, P., *J. Phys. Chem.*, 91 (1987) 3081.
k. Bash, P.A., Field, M.J. and Karplus, M., *J. Amer. Chem. Soc.*, 109 (1987) 8092.
l. Karlstrom, G., *J. Phys. Chem.*, 92 (1988) 1315; 93 (1989) 4952.
m. Hoshi, H., Sakurai, M., Inouye, Y. and Chūjō, R., *J. Mol. Struct. (Theochem)*, 180 (1988) 267.
n. Bertrán, J., In Bertrán, J. and Csizmadia, I.G. (Eds.) *New Theoretical Concepts for Understanding Organic Reactions*, Kluwer, Dordrecht, 1989, p. 231.
o. Stienke, T., Hänseler, E. and Clark, T., *J. Amer. Chem. Soc.*, 111 (1989) 9107.
p. Duijnen van, P.T. and Rullmann, J.A.C., *Int. J. Quantum Chem.*, 38 (1990) 181.
q. Karelson, M.M., Katritzky, A.R., Szafran, M. and Zerner, M.C., *J. Org. Chem.*, 54 (1989) 6030.
r. Wong, M.W., Wiberg, K.B. and Frisch, M., *J. Chem. Phys.*, 95 (1991) 8991.
s. Wong, M.W., Wiberg, K.B. and Frisch, M.J., *J. Amer. Chem. Soc.*, 114 (1992) 523.
- 21 Hermann, R.B., *J. Phys. Chem.*, 76 (1972) 2754.
- 22 a. Harris, M.J., Higuchi, T. and Ryttig, J.H., *J. Phys. Chem.*, 77 (1973) 2694.
b. Reynolds, J.A., Gilbert, D.B. and Tanford, C., *Proc. Natl. Acad. Sci. U.S.A.*, 71 (1974) 2925.
c. Chothia, C., *Nature*, 248 (1974) 338.
d. Amidon, G.L., Yalkowski, S.H., Anik, S.T. and Valvani, S.C., *J. Phys. Chem.*, 79 (1975) 2239.
e. Warshel, A., *J. Phys. Chem.*, 83 (1979) 1640.
f. Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H. and Zehfus, M.H., *Science*, 229 (1985) 834.
g. Eisenberg, D. and McLachlan, A.D., *Nature*, 319 (1986) 199.
h. Ooi, T., Oobatake, M., Nemethy, G. and Scheraga, H.A., *Proc. Natl. Acad. Sci. USA*, 84 (1987) 3086.
- 23 Lee, C.Y., McCammon, J.A. and Rossky, P.J., *J. Chem. Phys.*, 80 (1984) 4448.
- 24 Friedman, H.L. and Krishnan, C.V., *J. Solution Chem.*, 2 (1973) 119.
- 25 a. Hine, J. and Mookerjee, P.K., *J. Org. Chem.*, 40 (1975) 287.
b. Cabani, S., Gianni, P., Mollica, V. and Lepori, L., *J. Solution Chem.*, 10 (1981) 563.
c. Ben-Naim, A. and Marcus, Y., *J. Chem. Phys.*, 81 (1984) 2016.
- 26 Pearson, R.G., *J. Amer. Chem. Soc.*, 108 (1986) 6109.
- 27 Armstrong, D.R., Perkins, P.G. and Stewart, J.J.P., *J. Chem. Soc., Dalton Trans.*, 1973 (1973) 838.
- 28 Jorgensen, W.L., Gao, J. and Ravimohan, C., *J. Phys. Chem.*, 89 (1985) 3470.

- 29 a. Dewar, M.J.S., Healy, E.F., Holder, A.J. and Yan, Y.-C., *J. Comp. Chem.*, 11 (1990) 541.
b. Seeger, D.M., Korzeniewski, C. and Kowalchyk, W., *J. Phys. Chem.*, 95 (1991) 6871.
- 30 Juranic, I., Rzepa, H.S. and Yi, M., *J. Chem. Soc., Perkin Trans.*, 2 (1990) 877.
- 31 Rzepa, H.S. and Yi, M., *J. Chem. Soc., Perkin Trans.*, 2 (1991) 531.
- 32 Rzepa, H.S. and Yi, M., *J. Chem. Soc., Perkin Trans.*, 2 (1990) 943.
- 33 a. Gulera, G., Lluch, J.M., Oliva, A. and Bertrán, J., *J. Mol. Struct. (Theochem)*, 163 (1988) 101.
b. Dannenberg, J.J., *J. Phys. Chem.*, 92 (1988) 6869.
c. Khalil, M., Woods, R.J., Weaver, D.F. and Smith, V.H., Jr., *J. Comput. Chem.*, 12 (1991) 584.
- 34 Wagman, D.D., Evans, W.H., Parker, V.B., Schumm, R.H., Halow, I., Bailey, S.M., Churney, K.L. and Nuttall, R.C., *J. Phys. Chem. Ref. Data*, 11 (1982) Suppl. 2.
- 35 Wolfenden, R. and Williams, R., *J. Amer. Chem. Soc.*, 105 (1983) 1028.
- 36 Pauling, L., *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press, Ithaca, 1960.
- 37 Cramer, C.J. and Truhlar, D.G., AMSOL-version 1.0 (program no. 606 of the Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN), *QCPE Bull.*, 11 (1991) 57.
- 38 a. Dewar Research Group and Stewart, J.J.P., AMPAC-version 1 (program no. 506 of the Quantum Chemistry Program Exchange), *QCPE Bull.*, 6 (1986) 24a.
b. Liotard, D.A., Healy, E.F., Ruiz, J.M. and Dewar, M.J.S., AMPAC-version 2.1 (program no. 506 of the Quantum Chemistry Program Exchange), *QCPE Bull.*, 9 (1989) 123.
- 39 Cramer, C.J., Lynch, G.C. and Truhlar, D.G., AMSOL-version 3.0.1 and 3.0.1c. The new versions are also available from Quantum Chemistry Program Exchange, as version 3.0.1 of program no. 606.
- 40 Pratt, L.R. Chandler, D., *J. Chem. Phys.*, 67 (1977) 3683.
- 41 Wilson, C., Mace, J.E. and Agard, D.A., *J. Mol. Biol.*, 220 (1991) 495.
- 42 George, P., Witonsky, R.J., Trachtman, M., Wu, C., Dorwart, W., Richman, L., Richman, W., Shurayh, F. and Lentz, B., *Biochim. Biophys. Acta*, 223 (1970) 1.
- 43 Weast, R.C., (Ed.) *Handbook of Chemistry and Physics*, 67th ed., CRC Press, Boca Raton, 1986, p. D-163.
- 44 Lim, C., Bashford, D. and Karplus, M., *J. Phys. Chem.*, 95 (1991) 5610.
- 45 a. Wolfenden, R., *J. Amer. Chem. Soc.*, 98 (1976) 1987.
b. Wolfenden, R., *Biochem.*, 17 (1978) 201.
- 46 Radzicka, A., Pederson, L. and Wolfenden, R., *Biochem.*, 27 (1988) 4538.
- 47 Jorgensen, W.L. and Gao, J., *J. Amer. Chem. Soc.*, 110 (1988) 4212.
- 48 Yu, H.-A., Pettit, B.M. and Karplus, M., *J. Amer. Chem. Soc.*, 113 (1991) 2425.
- 49 Jorgensen, W.L., *J. Phys. Chem.*, 87 (1983) 5304.
- 50 Rosenberg, R.O., Mikkeleni, R. and Berne, B.J., *J. Amer. Chem. Soc.*, 104 (1982) 7647.
- 51 Jorgensen, W.L., *J. Chem. Phys.*, 77 (1982) 5757.
- 52 a. Curtiss, L.A., Frurip, D.J. and Blander, M., *J. Chem. Phys.*, 71 (1979) 2703.
b. Szalewicz, K., Cole, S.J., Kolos, W. and Bartlett, R.J., *J. Chem. Phys.*, 89 (1988) 3662.
- 53 Bondi, A., *J. Phys. Chem.*, 68 (1964) 441.
- 54 Freed, K.F., In Segal, G.A. (Ed.) *Semiempirical Methods of Electronic Structure Calculation, Part A: Techniques*, Plenum, New York, 1977, p. 201.
- 55 a. Mori, H., *Progr. Theor. Phys.*, 33 (1965) 423.
b. Grabert, H., *Projection Operator Techniques in Nonequilibrium Statistical Mechanics*, Springer-Verlag, Berlin, 1982.
- 56 a. Gonzalez-Lafont, A., Truong, T.N. and Truhlar, D.G., *J. Phys. Chem.*, 95 (1991) 4618.
b. Viggiano, A.A., Paschkewitz, J., Morris, R.A., Paulson, J.F., Gonzalez-Lafont, A. and Truhlar, D.G., *J. Amer. Chem. Soc.*, 113 (1991) 9404.
- 57 Besler, B.H., Merz, K.M., Jr. and Kollman, P.A., *J. Comput. Chem.*, 11 (1990) 431.
- 58 Dewar, M.J.S. and Thiel, W., *J. Amer. Chem. Soc.*, 99 (1977) 4899; 4907.