PM3-SM3: A General Parameterization for Including Aqueous Solvation Effects in the PM3 Molecular Orbital Model

2 AUTHORS:

Christopher J Cramer
University of Minnesota Twin Cities
532 PUBLICATIONS **23,445** CITATIONS

SEE PROFILE

Donald Truhlar
University of Minnesota Twin Cities
1,342 PUBLICATIONS **81,315** CITATIONS

SEE PROFILE

PM3-SM3: A General Parameterization for Including Aqueous Solvation Effects in the PM3 Molecular Orbital Model

Christopher J. Cramer*†

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

Donald G. Truhlar*

Department of Chemistry, Supercomputer Institute, and Army High Performance Computing Research Center, University of Minnesota, Minneapolis, Minnesota 55455-0431

Our recently proposed scheme for including aqueous solvation free energies in parameterized NDDO SCF models is extended to the Parameterized Model 3 semiempirical Hamiltonian. The solvation model takes accurate account of the hydrophobic effect for hydrocarbons, as well as electric polarization of the solvent, the free energy of cavitation, and dispersion interactions. Eight heteroatoms are included (along with H and C), and the new model is parameterized accurately for the water molecule itself, which allows meaningful treatments of specifically hydrogen bonded water molecules. The unphysical partial charges on nitrogen atoms predicted by the Parameterized Model 3 Hamiltonian limit the accuracy of the predicted solvation energies for some compounds containing nitrogen, but the model may be very useful for other systems, especially those for which PM3 is preferred over AM1 for the solute properties of the particular system under study. © 1992 by John Wiley & Sons, Inc.

INTRODUCTION

We have recently developed[1-3] a new continuum-dielectric model for including aqueous solvation effects in the Fock matrix of neglect-of-diatomic-differential-overlap (NDDO)[4] molecular orbital theory. We used the Austin Model 1 (AM1)[5] for solute parameters, and we developed first one and then an improved general parameter set, called Solvation Models 1 and 2, for the solvation terms. The second combined parameter set, AM1-SM2,[3] is quite successful for free energies of solvation, with a mean absolute error of 0.7 kcal/mol for 150 neutral solutes and 2.6 kcal/mol for 28 ions.

AM1-SM1 and AM1-SM2 include local-field terms representing solvent electric polarization, cavity creation, dispersion interaction, and change of solvent structure, and these terms are treated self-consistently with a solute electronic Hamiltonian. A particularly important systematic improvement in the performance of AM1-SM2 as compared to AM1-SM1 is for the quantitative free energies of solvation for alkanes, for which the magnitude of the predicted hydrophobic effect is greatly im-

proved. Hydrophobic interactions and other effects associated with restructuring the hydration cosphere are included because the model contains a semiempirical energy term proportional to solvent-accessible surface area. In the present article we present a sister model, PM3-SM3, that performs almost as well as AM1-SM2 for the mean absolute errors mentioned above, but is based on the Parameterized Model 3 (PM3)[6] solute Hamiltonian rather than the AM1 solute parameters. This should be useful for problems for which PM3, for one reason or another, is preferred by an investigator over AM1 for solute modeling.

THEORY

The object of the calculations considered here is either the free energy $G^0(aq)$ in aqueous solution or the free energy of solvation ΔG_S^0. Note that both these quantities depend on the choice of standard state and temperature; we use a one molar ideal solution as the standard state for solution and a one molar ideal gas as the standard state for the undissolved solute, and all parameterization is for 298 K. We partition $G^0(aq)$ as follows

$$G^0(aq) = G_S^0(aq) + G_{vib}(aq) + G_{elec}(aq) \quad (1)$$

---

*Correspondence may be addressed to either author.

†Current address: Department of Chemistry, University of Minnesota, Minneapolis, MN 55455-0431.

where $G_{vib}(aq)$ is the contribution to $G^0(aq)$ of the 3N-6 solute vibrations (3N-5 for linear molecules, 0 for atoms) that also exist in the gas phase, $G_{elec}(aq)$ is the free energy associated with electronic excitation of the solute, and $G_S^0(aq)$ is the rest. In the SMx models (x = 1 or 2 for previous work, 3 for the present paper), $G_S^0(aq)$ is the total energy obtained in an SCF calculation in which the solute electronic-energy-plus-nuclear-repulsion Hamiltonian is augmented by two new terms, called G_P and G_{CDS}^0, discussed in detail in previous work[1,3] and reviewed briefly below. Models SM1 and SM2 differ in the form of G_{CDS}^0 and in the parameters of both G_P and G_{CDS}^0. Having calculated $G_S^0(aq)$, one in principle calculates $G^0(aq)$ from eq. (1) and then ΔG_S^0 from

$$\Delta G_S^0 = G^0(aq) - E_{EN}(g) - G_{vib}(g) - G_{elec}(g) \quad (2)$$

where $E_{EN}(g)$ is the solute electronic energy plus nuclear repulsion obtained in a gas-phase SCF calculation, $G_{vib}(g)$ is the gas-phase vibrational free energy, and $G_{elec}(g)$ is the gas-phase electronic-excitation free energy. However, in all our work so far we have simply assumed that the vibrational and electronic terms in eqs. (1) and (2) cancel each other out. It is important though to note that these terms are sometimes critical, e.g., $G_{vib}(aq)$ − $G_{vib}(g)$ contains the effect of solvation on isotope effects, and the solvation models are designed to be able to address such effects. Another important point to note here is that geometries are optimized separately in the gas phase and solution so that $G_S^0(aq)$ and $E_{EN}(g)$ are calculated at different geometries. Now we turn to G_P and G_{CDS}^0.

The term G_P accounts for electric polarization of the continuum dielectric solvent by a solute distributed-monopole charge distribution.[7] The physical model employed is the generalized[7,8] Born[9] approximation in which various parts of the solute screen other parts from the solvent dielectric. The form of the interatomic Coulomb integrals and the algorithm for solute screening is taken from Still et al.,[10] with localized modifications introduced by us.[1] The treatment has the advantage that it handles arbitrary solute charge distributions and shapes, i.e., it is not necessary to assume that the solute fits in a spherical or elliptical cavity or that its electrostatic potential is calculable from its leading nonzero multipole moment. The functional forms for the G_{ENP} term are the same in both AM1-SM1 and AM1-SM2, and some of the parameters, $d_{kk'}$, $q_k^{(1)}$, $d_{kk'}^{(1)}$, $d_{kk'}^{(2)}$, $r_{kk'}^{(1)}$, and $r_{kk'}^{(2)}$, are also unchanged. We retain these forms and these parameters here as well. Other parameters, $\rho_k^{(0)}$, $\rho_k^{(1)}$, and $q_k^{(0)}$, defined previously,[1] needed to be re-optimized for AM1-SM2 and are also re-optimized here for PM3-SM3.

The term G_{CDS}^0 (which was called G_{CD}^0 in our first article[1]) accounts primarily for the free energy of creating a cavity in the solvent (to make room for the solute), the change in dispersion energy upon dissolution, and solute-induced changes in the structure of water in the first hydration shell. Since G_{CDS}^0 is a semiempirical term, it also accounts for a host of other considerations, e.g., errors in the NDDO partial charges, solute effects on water structure beyond the first hydration shell, dielectric saturation, and so forth. As for the polarization term, the solute is represented as a superposition of atom-centered, overlapping spheres, so again arbitrary solute shapes are accommodated. The form of G_{CDS}^0 changed in passing from AM1-SM1 to AM1-SM2, and in PM3-SM3 we use the latter. We do, however, re-optimize all the parameters, $\sigma_k^{(0)}$, $\sigma_k^{(1)}$, β_k, a_k, b_k, c_k, and d_k, which are all explained—along with the functional form—in a previous article.[3] We simply note here that the critical aspect of the approach is that the CDS terms depend on exposed (solvent-accessible) surface area of nonhydrogenic atoms only; hydrogens have no surface tension, and they have no volume for covering other atoms, but surface tensions of other atoms depend on the sum of their bond orders to hydrogen atoms. The radii used to compute the solvent-accessible surface areas are the semiempirical β_k parameters.

In general, the free energy of solvation should be the difference between the gas-phase free energy weighted over the low-energy gas-phase conformers and the aqueous-phase free energy weighted over the low-energy aqueous-phase conformers. In the present study we approximated this by the free energy of solvation of the lowest-energy conformer.

The full set of equations defining the SM3 model are presented in the Appendix, for the convenience of the reader.

PARAMETERIZATION

In our first article[1] we completely specified a parameterization set of experimental[11-14] free energies of solvation, consisting of data for 141 neutrals, 10 cations, and 17 anions. For the AM1-SM2 parameterization, we deleted one cation (H_3S^+ because the "experimental" solvation energy relies on an estimated pK_a in a thermodynamic cycle since no reliable data are available), and we added 9 neutrals, 1 new cation, and 1 more anion for a total of 150 neutrals, 10 cations, and 18 anions, again with experimental data from literature compilations.[11-15] The full lists of test solutes are completely specified by the previous work,[1,3] and so they are not repeated here.

The parameters in G_P were optimized primarily

Table I. PM3-SM3 parameters.

Element	$\rho_k^{(0)}$ (Å)	$\rho_k^{(1)}$ (Å)	$q_k^{(0)}$	$\sigma_k^{(0)}$ (cal mol^{-1} Å^{-2})	$\sigma_k^{(1)}$ (cal mol^{-1} Å^{-2})	β_k (Å)
H	0.59	1.289	−0.10	0	0	0
C	1.58	0.200	0.20	8.56	−0.39	3.1
N[a]	1.60	0.320	0.50	−42.20	−24.25	3.3
O[b]	1.66	−0.200	0.25	−34.76	−22.82	3.2
F	1.37	0.149	0.70	20.02	0	2.8
P	1.40	0.800	−1.50	9.79	0	3.6
S	1.20	0.900	0.70	−53.00	36.40	3.2
Cl	1.65	0.559	0.75	−3.26	0	3.4
Br	1.75	0.610	0.70	−4.61	0	3.4
I	1.88	0.798	0.60	−2.43	0	3.4

[a] $a_k = -4.93$, $b_k = 1.00$, $c_k = 3.00$, $d_k = 0.5$.
[b] $a_k = -2.62$, $b_k = 0.43$, $c_k = 2.66$, $d_k = 1.2$.

on the ions. The a_k, b_k, c_k, and d_k parameters in G_{CDS}^0 were optimized for ammonia, ammonium cation, and methylammonium cation for N and for water, hydronium ion, and methyoxonium ion for O. The other parameters in G_{CDS}^0 were optimized entirely on the set of neutrals, by minimizing the sum of the squares of the errors. The squared error for octane was weighted 64, and all other weights were unity. The cutoff Gaussians[1] in the ENP terms for N—O and O—O interactions have the same values as in AM1-SM1 and AM1-SM2. The final values of the parameters are given in Table I.

RESULTS AND DISCUSSION

Table II summarizes the mean unsigned errors for various classes of compounds used in the parameterization. For the computation of these errors all solutes were weighted equally. The table shows that in general the errors in the solvation energy

Table II. Mean unsigned errors (kcal/mol).

Class	No. of data in class	Mean unsigned error
Neutrals		
Hydrocarbons	33	0.7
H, C, N compounds	18	2.0
H, C, O compounds	38	1.1
H, C, F compounds	4	0.7
H, C, S compounds	6	0.4
H, C, Cl compounds	14	0.7
H, C, Br compounds	8	0.4
H, C, I compounds	5	0.2
Compounds with 4 or more kinds of atoms	20	1.0
NH$_3$, H$_2$O, PH$_3$, H$_2$S	4	0.025
All	150	0.9
Ions		
Cations	10	3.1
Anions	18	3.8
All	28	3.5

are small compared to the errors[6] in the heats of formation of the solutes.

Results for selected compounds are presented in Table III. This table also shows the polarization and cavity-dispersion-solvent structure contributions separately, labeled ENP and CDS, respectively. The former is labeled ENP because it includes not only the G_P contribution but also the change in the electronic energy and nuclear repulsion in passing from the vapor phase to solution.

The reader will notice several trends in Tables II and III; we will comment explicitly on only a few.

First consider the hydrocarbons. For alkanes, cycloalkanes, and alkenes, the polarization contributions are very small. The new approach to hydrophobic effects,[3] by which only nonhydrogenic atoms have nonzero surface tension or volume in the G_{CDS}^0 calculation, but their surface tensions depend on the sum of their bond orders to hydrogen, appears to work quite well here, and it is about equally successful with PM3 as with AM1. Notice that differences in free energy of solvation as one lengthens an alkane are more accurate than absolute free energies of solvation.

One would expect that in general differences in free energy of solvation among similar compounds would be more accurate than absolute values, and Table III bears this out not only for alkanes but also for halogenated alkanes and for compounds containing oxygen. Since free energy perturbation theory, which is addressed to such differences, has proven very useful for applications,[16] the present model may be useful—and perhaps even more accurate—for this kind of application as well, and it has the advantage that it also delivers absolute free energies of solvation with no further effort.

For the study of complex phenomena in aqueous solution, there are many possible reasons why one may wish to treat one or more water molecules explicitly. For example, one may suspect unusual or specific hydrogen bonding interactions, water may function as a possible reagent, or one may

Table III. Examples of components and total free energies of solvation (kcal/mol).

Solute	PM3-SM3			Expt.[a]
	ENP	CDS	ΔG_S^0	ΔG_S^0
Hydrocarbons				
Ethane	0.0	1.2	1.2	1.8
Propane	0.0	1.4	1.4	2.0
Cyclopropane	−0.2	1.3	1.1	0.8
Neopentane	0.1	1.8	1.8	2.5
Cyclopentane	0.0	1.6	1.6	1.2
Hexane	0.1	2.1	2.1	2.5
Cyclohexane	0.1	1.8	1.8	1.2
Octane	0.1	2.5	2.6	2.9
Ethene	−0.3	1.2	0.9	1.3
Propene	−0.3	1.4	1.1	1.3
2-Methylpropene	−0.4	1.6	1.2	1.2
E-2-Pentene	−0.3	1.8	1.5	1.3
Cyclopentene	−0.6	1.6	1.1	0.6
1,3-Butadiene	−2.4	1.6	−0.8	0.6
Benzene	−2.0	1.7	−0.3	−0.9
Toluene	−2.0	1.9	−0.1	−0.9
1-Hexyne	−2.4	2.1	−0.3	0.3
H, C, N neutrals				
Ethylamine	−0.1	−5.4	−5.5	−4.5
1-Butanamine	−0.0	−5.0	−5.0	−4.3
Aniline	−2.3	−4.4	−6.7	−4.9
Dimethylamine	−0.3	−2.7	−3.0	−4.3
Piperazine	−0.5	−5.7	−6.1	−7.4
Pyridine	−3.0	−0.9	−3.9	−4.7
2-Methylpyrazine	−3.3	−2.8	−6.1	−5.5
Acetonitrile	−2.8	−3.4	−6.2	−3.9
Butanenitrile	−2.2	−2.8	−4.9	−3.7
H, C, O neutrals				
Ethanol	−1.1	−3.5	−4.6	−5.0
1-Propanol	−1.0	−3.3	−4.3	−4.8
Prop-2-en-1-ol	−0.4	−3.2	−3.6	−5.0
Phenol	−2.7	−2.7	−5.4	−6.6
Propanoic acid	−2.2	−5.2	−7.4	−6.5
Butanoic acid	−2.0	−5.0	−7.0	−6.4
Ethyl acetate	−2.2	−1.8	−4.0	−3.1
Methyl butanoate	−1.6	−1.7	−3.3	−2.8
Butanal	−2.9	−1.2	−4.1	−3.2
Acetophenone	−4.7	−0.4	−5.1	−4.6
1,4-Dioxane	−1.5	−2.0	−3.5	−5.1
1-Methoxypropane	−0.7	0.2	−0.5	−1.7
Tetrahydrofuran	−1.2	−0.4	−1.7	−3.5
1,2-Dimethoxyethane	−1.2	−1.2	−2.5	−4.8
Propanal	−3.0	−1.4	−4.5	−3.5
3-Pentanone	−2.8	−0.3	−3.1	−3.4
4-Heptanone	−2.4	0.1	−2.3	−2.9
Halogenated hydrocarbons				
1,1-Difluoroethane	−1.8	2.5	0.6	−0.1
1,2-Dichloroethane	−0.5	−0.2	−0.7	−1.7
1,1,1-Trichloroethane	−0.3	−0.3	−0.6	−0.3
1,1,2-Trichloroethane	−0.4	−0.5	−0.9	−2.0
Ethyl chloride	−0.6	0.5	−0.1	−0.6
Vinyl chloride	−0.5	0.5	0.0	−0.6
Z-1,2-Dichloroethene	−0.7	−0.1	−0.8	−1.2
E-1,2-Dichloroethene	−0.6	−0.3	−0.9	−0.8
Trichloroethene	−1.1	−0.6	−1.7	−0.4
Tetrachloroethene	−1.8	−0.9	−2.7	0.1
Chlorobenzene	−2.4	1.0	−1.3	−1.1
Bromoethane	−1.6	0.4	−1.2	−0.7
1,2-Dibromoethane	−2.5	−0.3	−2.8	−2.1
Bromobenzene	−2.4	0.9	−1.5	−1.5
Iodomethane	−1.0	0.3	−0.7	−0.9
Iodoethane	−1.6	0.6	−0.9	−0.7

Table III. (continued)

Solute	PM3-SM3			Expt.[a]
	ENP	CDS	ΔG_S^0	ΔG_S^0
H, C, S compounds				
Ethanethiol	−0.4	−0.3	−0.7	−1.3
Diethyl sulfide	−1.0	−0.4	−1.4	−1.3
Thiophenol	−2.8	0.2	−2.5	−2.6
Thioanisole	−3.0	−0.5	−3.5	−2.7
Neutrals with 4 or more kinds of atoms				
2-Methoxyethanamine	−0.8	−6.6	−7.4	−6.6
Nitroethane	1.3	−3.9	−2.6	−3.7
Nitrobenzene	−1.1	−3.1	−4.1	−4.1
2,2,2-Trifluoroethanol	−3.7	−1.3	−4.9	−4.3
bis-2-Chloroethyl sulfide	−0.9	−1.9	−2.8	−3.9
p-Bromophenol	−3.1	−3.5	−6.5	−7.1
2-Methyl-1-nitrobenzene	−1.1	−2.5	−3.6	−3.6
Chlorodifluoromethane	−1.7	1.5	−0.2	−0.5
2-Chloro-1,1,1-trifluorethane	−2.0	2.2	0.2	0.1
2-Bromo-1-chloroethane	−1.0	−0.3	−1.4	−2.0
Neutrals without C				
Ammonia	0.0	−4.3	−4.3	−4.3
Water	−1.9	−4.4	−6.3	−6.3
Phosphine	−1.0	1.6	0.6	0.6
Hydrogen sulfide	0.0	−0.8	−0.8	−0.7
Cations				
H_3O^+	−98.9	−3.5	−102.4	−104
NH_4^+	−71.2	−10.5	−81.7	−79
$MeOH_2^+$	−81.5	−3.5	−84.6	−83
$MeSH_2^+$	−76.7	0.1	−76.6	−74
$MeNH_3^+$	−64.7	−2.4	−67.1	−70
Me_2OH^+	−67.5	−1.6	−69.1	−70
$Me_2NH_2^+$	−60.0	−3.1	−63.1	−63
Me_3NH^+	−54.3	−0.2	−54.5	−59
Me_3PH^+	−58.8	1.8	−57.0	−53
$CH_3C(OH)NH_2^+$	−66.6	−9.0	−75.6	−66
Anions				
H^-	−89.0	0.0	−89.0	−89
F^-	−109.0	2.0	−107.0	−107
Cl^-	−76.5	−0.5	−77.0	−77
Br^-	−71.3	−0.7	−72.0	−72
I^-	−62.6	−0.4	−63.0	−63
OH^-	−103.9	−7.5	−111.4	−106
CN^-	−76.2	−3.4	−79.6	−77
O_2^-	−79.5	−5.4	−84.9	−87
HS^-	−74.7	−1.9	−76.7	−76
PH_2^-	−68.4	1.6	−66.8	−67
HC_2^-	−78.0	1.2	−76.8	−73
HO_2^-	−82.1	−7.4	−89.5	−101
N_3^-	−64.7	−7.9	−72.6	−74
NO_2^-	−70.4	−6.5	−76.9	−72
CH_3O^-	−79.0	−2.6	−81.6	−95
NO_3^-	−40.7	−7.1	−47.7	−65
CH_2CN^-	−67.8	−3.5	−71.3	−75
$CH_3CO_2^-$	−71.3	−4.3	−75.5	−77

[a]From refs. 11–13 and 15 for neutrals, ref. 14 for ions. Experimental errors may be substantial (± 5 kcal/mol for ions, up to a few tenths of a kcal/mol for neutrals).

suspect particular water molecules (e.g., those trapped by protein tertiary structure) of having non-bulk-like properties. Alternatively one may be interested in monitoring the water motions to understand relaxation behavior. The PM3 method may be preferred over AM1 for such studies because the hydrogen bond structures it predicts are closer to traditional models of such structures and hence may be more realistic. For such studies it is useful if the free energy of solvation of a water molecule is predicted correctly. Then the enlargement of the solute into a supersolute consisting of

the original solute plus a hypothetical noninteracting water molecule would not cause a spurious change in solvation free energy, and consequently any change in solvation free energy upon introducing the explicit water may be interpreted physically. (However, such changes are still subject to error due to the approximate nature of the PM3 model itself.) In PM3-SM3, as in AM1-SM2, we have used the dependence of the surface tension of O on its bond order to hydrogens in such a way as to make the solvation free energy of water itself (as a solute) accurate to better than 0.1 kcal. This means that one can add one or more explicit, hydrogen-bonded waters to a solute to study specific bonding effects, and the results should suffer minimal ambiguity due to incorrect solvation of such explicit waters.

For amines, cyano, and nitro compounds, however, PM3-SM3 is not very accurate. This appears to be due to nonphysical partial charges on N in PM3, a problem that has also been noted elsewhere.[17] In general the PM3 nitrogen atom is far too electropositive. Thus, for gas-phase NH_4^+, the AM1 nitrogen partial charge is -0.09, whereas with PM3 it is $+1.00$. This may be compared to the partial (Mulliken) charge from an *ab initio* calculation at the HF/6-311G** level, which yields -0.27. This difference gives rise to anomalously small polarization terms in SM3 for amines—especially tertiary amines. The effect is even more striking for nitro groups, where sizable *positive* values for G_{ENP} are observed for nitroaliphatics. The surface tension terms cannot completely correct for these deficiencies, and Table II shows bigger errors for neutral solutes containing nitrogen than for any other class.

The comparison of theory and experiment for all 178 solutes included in the parameterization set is shown in Figures 1 and 2. The correlation is generally good with many of the worst cases involving nitrogen.

Finally we consider the question of the importance of re-optimizing the electronic wave function and nuclear geometry in the presence of solvent. Table IV shows results for a selection of solutes. The first numerical column, labeled NOPOL, is the solvation energy calculated for the gas-phase electronic wave function at the gas-phase optimized geometry. The next column, labeled 1SCF, is for the geometry frozen at the gas-phase optimum, but with the electronic wave function relaxed. The third column is the fully relaxed value. Differences due to electronic relaxation (NOPOL vs. 1SCF) are in the range 0.1-1 kcal for the neutrals shown and 0.5-3.5 kcal for the ions shown, whereas those due to geometry relaxation (1SCF vs. fully relaxed) range up to 0.1 kcal for neutrals and up to 1.1 kcal for ions, for the
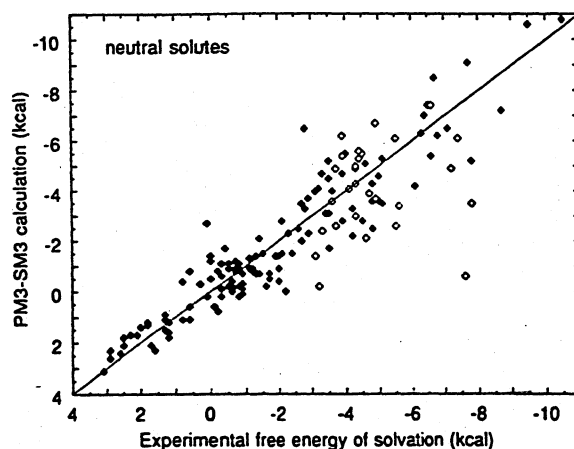

Figure 1. Correlation between calculated free energies of solvation and experimental ones for neutral solutes. Open diamonds represent solutes containing one or more N atoms, and closed diamonds represent other solutes.
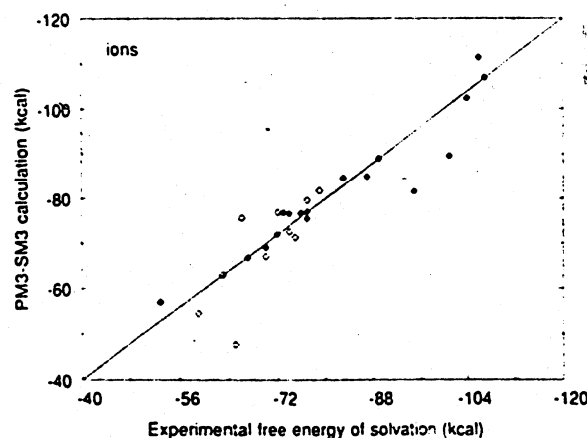

Figure 2. Same as Fig. 1 except for charged solutes.

Table IV. Calculated free energies of solvation (kcal/mol) without and with relaxation.

Solute	NOPOL	1SCF	Relaxed
2-Methylpropene	1.3	1.2	1.2
Benzene	0.1	-0.3	-0.3
Pyridine	-2.9	-3.9	-3.9
Ethanol	-4.4	-4.6	-4.6
3-Pentanone	-2.1	-3.0	-3.1
Bromobenzene	-0.8	-1.4	-1.5
Iodomethane	-0.2	-0.6	-0.7
Thiophenol	-1.8	-2.5	-2.5
p-Bromophenol	-5.8	-6.5	-6.5
Chlorodifluoromethane	0.4	-0.1	-0.2
Water	-5.8	-6.3	-6.3
H_3O^+	-101.8	-102.3	-102.4
$(CH_3)_2NH_2^+$	-60.6	-62.3	-63.1
$CH_3CO_2^-$	-70.9	-74.4	-75.5

cases shown. In some cases not included in our test set, e.g., nucleic acid bases, differences are even larger.

CONCLUDING REMARKS

We have developed a new parameter set, Solvation Model 3 (SM3), for calculating free energies of solvation and hence solute free energies in aqueous solution. The new parameter set is designed for use with the PM3 semiempirical molecular orbital model for the solute. Except for neutral nitrogen-containing compounds we appear to be able to parameterize aqueous solvation free energies about equally well in a PM3 framework as we did with our Solvation Model 2, which was based on AM1. Since many workers have found both the AM1 and PM3 methods useful for a wide variety of problems without solvation, we hope that AM1-SM2 and PM3-SM3 will both be useful for future work on a variety of systems where aqueous solvation is important.

We have developed a portable version, AMSOL v.3.0[18] of a computer package incorporating the SM1, SM1a, SM2, and SM3 solvation models. The resulting code, which is based on AMPAC v.2.1, will be made available through Quantum Chemistry Program Exchange as v.3.0 of program 606. The portable version of AMSOL is called 3.0, and a partially optimized Cray-specific version called 3.0c is also distributed as part of the package that comprises QCPE program 606.

APPENDIX

In this appendix we specify the full set of equations for the SM3 method. The equations are all

$$C_{kk'}^{(1)} = \begin{cases} d_{kk'}^{(1)} \exp\left\{ -\dfrac{d_{kk'}^{(2)}}{1 - [(r_{kk'} - r_{kk'}^{(1)})/r_{kk'}^{(2)}]^2} \right\} & |r_{kk'} - r_{kk'}^{(1)}| < r_{kk'}^{(2)} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

taken from earlier work.[1,3,4,6,8-10,19,20] We calculate the free energy of solvation as

$$\Delta G_S^0 = G_S^0(aq) - E_{EN}(g) \quad (3)$$

where $E_{EN}(g)$ is the gas-phase electronic kinetic and electronic and nuclear coulombic energy, $G_S^0(aq)$ is the part of the solute aqueous free energy given by

$$G_S^0(aq) = G_{ENP}(aq) + G_{CDS}^0(aq) \quad (4)$$

$G_{ENP}(aq)$ is the sum of the solute electronic kinetic and electronic-nuclear coulombic energies in the presence of solvent plus the free energy of electronic polarization, and $G_{CDS}^0(aq)$ is the cavitation-dispersion-solvent-structural free energy.

The first term of (4) is calculated as the final energy of an SCF calculation on the solute with aqueous polarization terms in the Fock matrix $\mathbf{F}$. For a solute treated by the restricted Hartree-Fock, linear-combination-of-atomic-orbitals, all-valence-electrons approach, with a basis of m valence atomic orbitals labeled by μ and ν, this yields

$$G_{ENP}^0(aq) = \frac{1}{2} \sum_{\mu=1}^{m} \sum_{\nu=1}^{m} P_{\mu\nu}(H_{\mu\nu} + F_{\mu\nu}) \quad (5)$$

where $\mathbf{P}$ is the self-consistent density matrix, and $\mathbf{H}$ is the one-electron energy matrix unmodified by the solvent. The Fock matrix at the NDDO level is given by

$$F_{\mu\nu} = F_{\mu\nu}^{(0)} + \delta_{\mu\nu}\left(1 - \frac{1}{\varepsilon}\right) \sum_{k'}$$
$$\times \left(Z_{k'} - \sum_{\mu' \in k'} P_{\mu'\mu'}\right)\gamma_{kk'} \quad \mu \in k \quad (6)$$

where $\mathbf{F}^{(0)}$ is the gas-phase Fock matrix, ε is the dielectric constant of water, Z_k is the valence nuclear charge of atom k (equal to the nuclear charge minus the number of core electrons), and $\gamma_{kk'}$ is a Coulomb integral. The latter is given by

$$\gamma_{kk'} = [r_{kk'}^2 + \alpha_k\alpha_{k'}C_{kk'}(r_{kk'})]^{-1/2} \quad (7)$$

where $r_{kk'}$ is the interatomic distance between atoms k and k', α_k is the effective ionic radius of atom k, and $C_{kk'}$ is given by

$$C_{kk'} = C_{kk'}^{(0)}(r_{kk'}) + C_{kk'}^{(1)}(r_{kk'}) \quad (8)$$

where

$$C_{kk'}^{(0)} = \exp(-r_{kk'}^2/4\alpha_k\alpha_{k'}) \quad (9)$$

and $C_{kk'}^{(1)}$ is given by

We make $d_{kk'}^{(1)}$ nonzero only for O—O and N—H interactions. The parameters in $C_{kk'}^{(1)}$ have not changed since SM1 and are given in our original article.[1] Similar equations hold for the unrestricted Hartree-Fock formalism.

For the monatomic ($k = k' = 1$), α_k is set equal to the intrinsic Coulomb radius, ρ_k, which is calculated from

$$\rho_k = \rho_k^{(0)} + \rho_k^{(1)}\left[-\frac{1}{\pi} \arctan \frac{q_k + q_k^{(0)}}{0.1} + \frac{1}{2}\right] \quad (11)$$

where $\rho_k^{(0)}$, $\rho_k^{(1)}$, and $q_k^{(0)}$ are empirical parameters, and q_k is the net charge on atom k given by

$$q_k = Z_k - \sum_{\mu \in k} P_{\mu\mu} \quad (12)$$

In the multicenter case, α_k is determined numerically, following a procedure introduced by Still et al.[10] In this procedure, α_k is chosen so that the polarization energy $G_{B,k}^0(aq)$ calculated for a monatomic ion by the Born equation

$$G_{B,k}^0(aq) = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) q_k^2 \Big/ \alpha_k \qquad (13)$$

equals the polarization energy $G_{P,k}^0(aq)$ of the atom with the same partial charge in the molecular solute, where the other atoms in the solute screen the aqueous dielectric. Thus we consider M spherical shells i around each atom k and calculate

$$G_{P,k}^0(aq) = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) q_k^2 \sum_{i=1}^{M+1} \frac{A_i(r_i\{\rho_{k'}\})}{4\pi r_i^2}$$
$$\times \left(\frac{1}{r_i - 0.5T_i} - \frac{1 - \delta_{i,M+1}}{r_i + 0.5T_i}\right) \quad (14)$$

where $A_i(r_i, \{\rho_{k'}\})$ is the numerically determined exposed surface area of the sphere around atom k of radius r_i, i.e., that area *not* included in any spheres centered around other atoms when those spheres have radii given by the set $\{\rho_{k'}\}$, and r_i and T_i are defined recursively by

$$r_1 = \rho_k + 0.5T_1 \qquad (15)$$
$$r_i = r_{i-1} + 0.5(T_{i-1} + T_i) \qquad i > 1 \quad (16)$$
$$T_1 = 0.02 \text{ Å} \qquad (17)$$
$$T_i = 1.5T_{i-1} \qquad i > 1 \quad (18)$$

For monatomic solutes we use $M = 0$, and for molecular solutes, M is reached for each atom when the sphere of radius $r_M + 0.5T_M$ encloses the entire solute.

To calculate the exposed surface area, 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 $K_0/2$ as the spacing between points) from pole to pole on the sphere and places K_n points on the smaller circle, where

$$K_n = K_0 r_n / r_0 \qquad (19)$$

and r_n is the radius of the small circle. This yields the same distribution on the small circle as the large circle. For A_i in eq. (14), we use $r_0 = r_i$ and $K_0 = 50$. The exposed surface area, A_i, is determined by multiplying the total surface area, $4\pi r_0^2$,

by the fraction of points on the spherical surface not enclosed in spheres centered on other atoms.

The remaining contribution to the free energy of solvation beyond $G_{ENP}(aq)$ is calculated from

$$G_{CDS}^0(aq) = \sum_{k'} \{\sigma_{k'}^{(0)} + \sigma_{k'}^{(1)}[f(B_{k'H}) + g_{k'}(B_{k'H})]\}$$
$$\times A_{k'}(\beta_{k'}, \{\beta_k\}) \quad (20)$$

where $\sigma_{k'}^{(0)}$ and $\sigma_{k'}^{(1)}$ are atomic surface tension parameters, f, $g_{k'}$, and $B_{k'H}$ are functions discussed below, and $A_{k'}(\beta_{k'}, \{\beta_k\})$ is the solvent-accessible surface area for atom k'. Solvent-accessible surface area is calculated in the same way as the exposed surface area described above, except that the atomic radii are the semiempirical parameters β_k instead of ρ_k, and for $A_{k'}$ in eq. (20) we set r_0 equal to $\beta_{k'}$ and K_0 equal to 90. We defined the hydrogen atom to have no solvent-accessible surface area, and, further, we defined it not to block the solvent accessible surface area of the underlying nonhydrogenic atom, i.e., $\beta_H = 0$.

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 employed by Armstrong et al.,[20] one obtains

$$B_{kH} = \sum_{\mu \in k, \nu \in H} P_{\mu\nu}^2 \qquad (21)$$

where μ runs over the valence orbitals of atom k and ν runs over all hydrogen 1s orbitals. The functions f and g_k are defined by

$$f(B) = \tan^{-1}(\sqrt{3}B) \qquad (22)$$

and

$$g_k(B) = \begin{cases} a_k \exp\left\{-\dfrac{b_k}{1 - [(B - c_k)/d_k]^2}\right\} & |B - c_k| < d_k \\ 0 & \text{otherwise} \end{cases} \qquad (23)$$

Note added in proof: A more complete set of results for the PM3-SM3 model is provided in a longer, subsequent article[21] with extensive comparison to AM1-SM2 results.

References

1. C.J. Cramer and D.G. Truhlar, *J. Am. Chem. Soc.*, **113**, 8305, 9901(E) (1991).
2. C.J. Cramer and D.G. Truhlar, *J. Am. Chem. Soc.*, **113**, 8552, 9901(E) (1991).
3. C.J. Cramer and D.G. Truhlar, *Science*, **256**, 213 (1992).
4. J.A. Pople, D.P. Santry, and G.A. Segal, *J. Chem. Phys.*, **43**, S129 (1965).
5. M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, and J.J.P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985); M.J.S. Dewar and E.G. Zoebisch, *J. Mol. Struct.* (*THEOCHEM*), **180**, 1 (1988); M.J.S. Dewar and C. Jie, *J. Mol. Struct.* (*THEOCHEM*), **187**, 1 (1989); M.J.S. Dewar and Y.C. Yuan, *Inorg. Chem.*, **29**, 3881 (1990).
6. J.J.P. Stewart, *J. Comp. Chem.*, **10**, 209, 221 (1989); J.J.P. Stewart, *J. Comp.-Aided Mol. Des.*, **4**, 1 (1990).

7. R. Constanciel and O. Tapia, *Theoret. Chim. Acta*, **48**, 75 (1978).

8. G.J. Hoijtink, E. deBoer, P.H. van der Meij, and W.P. Weijland, *Recl. Trav. Chim. Pays-Bas*, **75**, 487 (1956); F. Peradejordi, *Cah. Phys.*, **17**, 393 (1963); I. Jano, *Compt. Rend. Acad. Sci. (Paris)*, **261**, 103 (1965); G. Klopman, *Chem. Phys. Lett.*, **1**, 200 (1967); I. Fischer-Hjalmars, A. Hendriksson-Enflo, and C. Hermann, *Chem. Phys.*, **24**, 167 (1977); O. Tapia, in *Quantum Theory of Chemical Reactions*, R. Daudel, A. Pullman, L. Salem, and A. Veillard, Eds., Reidel, Dordrecht, 1980, pp. 25–72; R. Costanciel and R. Contreras, *Theor. Chim. Acta*, **65**, 1 (1984); R. Contreras and J.S. Gomez-Jeria, *J. Phys. Chem.*, **88**, 1905 (1984); T. Kozaki, M. Morihasi and O. Kikuchi, *J. Am. Chem. Soc.*, **111**, 1547 (1989); S.C. Tucker and D.G. Truhlar, *Chem. Phys. Lett.*, **157**, 164 (1989).

9. M. Born, *Z. Physik,*, **1**, 45 (1920).

10. W.C. Still, A. Tempczyk, R.C. Hawley, and T. Hendrickson, *J. Am. Chem. Soc.*, **112**, 6127 (1990).

11. J. Hine and P.K. Mookerjee, *J. Org. Chem.*, **40**, 292 (1975).

12. S. Cabani, P. Gianni, V. Mollica, and L. Lepori, *J. Solution Chem.*, **10**, 563 (1981).

13. A. Ben-Naim and Y. Marcus, *J. Chem. Phys.*, **81**, 2016 (1984).

14. R.G. Pearson, *J. Am. Chem. Soc.*, **108**, 6109 (1986).

15. D.D. Wagman, W.H. Evans, V.B. Parker, R.H. Schumm, I. Halow, S.M. Bailey, K.L. Churney, and R.C. Nuttall, *J. Phys. Chem. Ref. Data*, **11**, Suppl. 2 (1982).

16. W.L. Jorgensen, *Acc. Chem. Res.*, **22**, 184 (1989); P.A. Kollman and K.M. Merz Jr., *Acc. Chem. Res.*, **23**, 246 (1990).

17. I. Juranic, H.S. Rzepa, and M. Yi, *J. Chem. Soc., Perkin Trans. 2*, **1990**, 887 (1990); H.S. Rzepa and M. Yi, *J. Chem. Soc., Perkin Trans. 2*, **1991**, 513 (1991).

18. C.J. Cramer, G.C. Lynch, and D.G. Truhlar, AMSOL v.3.0, based on AMPAC v.2.1, by D.A. Liotard, E.F. Healy, J.M. Ruiz, and M.J.S. Dewar.

19. B. Lee and F.M. Richards, *J. Mol. Biol.*, **55**, 371 (1971); R.B. Hermann, *J. Phys. Chem.*, **76**, 2754 (1972).

20. D.R. Armstrong, P.G. Perkins, and J.J.P. Stewart, *J. Chem. Soc., Dalton Trans.*, **1973**, 838 (1973).

21. C.J. Cramer and D.G. Truhlar, *J. Comp.-Aided Mol. Des.*, in press.