J. Am. Chem. Soc. 1991, 113, 8305–8311

$$R_i(0\rightarrow 1) = \frac{\hbar}{4c}\left|\frac{\partial\mu_0}{\partial S}\right|^2 \sum_{i,j'}^{N} a_{ij}a_{ij'}\hat{\mu}_j\cdot\mathbf{R}_{j'}^0\times\hat{\mu}_{j'} \qquad \text{(A14)}$$

$$D_i(0\rightarrow 1) = \frac{\hbar}{2\omega_i}\left|\frac{\partial\mu^0}{\partial S}\right|^2 \sum_{j,j'}^{N} a_{ij}a_{ij'}\hat{\mu}_j\cdot\hat{\mu}_{j'} \qquad \text{(A15)}$$

The total dipole strength for the N modes arising from the coupling of the set of oscillators is

$$D_T = \sum_{i=1}^{N} D_i(0\rightarrow 1) = \frac{\hbar N}{2\omega_i}\left|\frac{\partial\mu^0}{\partial S}\right|^2 \qquad \text{(A16)}$$

By substituting eq A16 into eq A14 and rearranging the triple product, we obtain the generalized coupled-oscillator expression for N nonchiral identical oscillators

$$R_i(0\rightarrow 1) = -\pi\nu_i\frac{D_T}{N}\sum_{j>j'}^{N} a_{ij}a_{ij'}\mathbf{R}_{jj'}\cdot\hat{\mathrm{u}}_j\times\hat{\mathbf{u}}_{j'} \qquad \text{(A17)}$$

where $\mathbf{R}_{jj'}$ is the separation vector $(\mathbf{R}_{j'}^0 - \mathbf{R}_j^0)$ between oscillators j and j' and ν_i is the frequency (cm^{-1}) of the ith mode. It is clear from this form of eq A12 that the position of $\mathbf{R}_j^0$ along the bond axis of the oscillator is immaterial and that $\mathbf{R}_j^0$ can be the same point for both atoms in the oscillator, since moving $\mathbf{R}_j^0$ to $\mathbf{R}_j^0 + \lambda\hat{\mathbf{u}}_j$ does not affect the rotational strength. Only pairs of oscillators that are chirally oriented make nonzero contributions in eq A17.

Registry No. 1, 132260-31-2; **2**, 136342-83-1.

General Parameterized SCF Model for Free Energies of Solvation in Aqueous Solution

Christopher J. Cramer[*,†] **and Donald G. Truhlar**[*,‡]

Contribution from the U.S. Army Chemical Research, Development, and Engineering Center, Aberdeen Proving Ground, Maryland 21010-5423, and Department of Chemistry, Supercomputer Institute, and Army High Performance Computing Research Center, University of Minnesota, Minneapolis, Minnesota 55455-0431. *Received March 8, 1991*

Abstract: We present a new general parameterization for aqueous solvation free energies of molecules and ions in aqueous solution. It is obtained by extending a semianalytic treatment of solvation recently proposed for use with molecular mechanics and liquid simulations by Still et al. As extended here, the solvation terms are included in a Fock operator. The model incorporates reaction field polarization effects through the generalized Born functional with charges obtained by AM1 molecular orbital calculations, and it includes cavitation, dispersion, and hydrophobic effects through an empirical function of solvent-accessible surface area. A general parameter set, including parameters for H, C, N, O, F, S, Cl, Br, and I, has been obtained by considering a data set consisting of 141 neutral molecules, 10 cations, and 17 anions. The neutral molecules include alkanes, cycloalkanes, alkenes, arenes, alkynes, ethers, heterocycles, carboxylic acids, esters, nitriles, aldehydes, ketones, alcohols, amines, nitro compounds, sulfides, thiols, halides, and polyfunctional compounds. The general parameterization is called Solvation Model 1, and it is particularly well suited for chemical reaction dynamics and reaction intermediates. We also discuss how the model may be refined for solvation free energies for stable neutral molecules.

1. Introduction

Computational chemistry is continually improving, not only in its ability to correlate experimental trends, but also in the ability to predict qualitative and sometimes quantitative features of structures and reactions not yet observed experimentally. While the advances in ab initio chemistry are dramatic, especially for small molecules,[1] the advances in semiempirical methods, both molecular orbital theory[2] and molecular mechanics,[3] have had an impact on a broader range of chemistry, and these are still the methods of choice for large molecules such as those involved in biochemical processes. Just as the usefulness of ab initio techniques has been closely tied to the availability of well-tested general basis sets[4] and widely applicable computer programs[4] with analytic gradient techniques for stationary point analysis,[5] the revolution in usefulness of semiempirical computational techniques has been closely tied to well-tested general parameterizations,[6–10] such as Austin Model 1 (AM1)[7] and the MM2 force field,[9] and—again—to widely available general computer programs[11–13] with efficient stationary point analyses.[14]

One difficulty that still persists and greatly limits the applicability of computational chemistry techniques is the expense of including solvent effects. Simulations involving large numbers of explicit water molecules[15–18] have proved their usefulness but remain expensive and susceptible to errors in potential energy functions. A complementary approach that should lead to faster progress for some problems is a general parameterization (in the

---

[†] U.S. Army Chemical Research, Development, and Engineering Center.
[‡] University of Minnesota.

(1) (a) *Comparison of Ab Initio Quantum Chemistry with Experiment for Small Molecules*; Bartlett, R. J., Ed.; D. Reidel: Dordrecht, 1985. (b) *Supercomputer Algorithms for Reactivity, Dynamics, and Kinetics of Small Molecules*; Laganà, A., Ed.; Kluwer: Dordrecht, 1989.

(2) (a) Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970. (b) Naray-Szabo, G.; Surjan, P. R.; Angyan, J. G. *Applied Quantum Chemistry*, Reidel: Dordrecht, 1987.

(3) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.

(4) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*, John Wiley & Sons: New York, 1986.

(5) (a) Pulay, P. *Mol.Phys.* **1969**, *17*, 197. (b) Schlegel, H. B.; Wolfe, S.; Bernardi, F. *J. Chem. Phys.* **1975**, *63*, 3632.

(6) (a) Bingham, R. C.; Dewar, M. J. S.; Lo, D. H. *J. Am. Chem. Soc.* **1975**, *97*, 1294. (b) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

(7) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(8) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209, 221.

(9) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.

(10) Allinger, N. L.; Yuh, Y. H.; Li, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(11) Liotard, D. A.; Healy, E. F.; Ruiz, J. M.; Dewar, M. J. S. AMPAC, version 2.1 (Quantum Chemistry Program Exchange program 506). *QCPE Bull.* **1989**, *9*, 123.

(12) (a) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1. (b) Stewart, J. J. P. MOPAC, version 6.0 (Quantum Chemistry Program Exchange program 455). *QCPE Bull.* **1990**, *10*, 86.

(13) Allinger, N. L.; Yuh, Y. H.; Gensmantel, N. P. *MM2* (Quantum Chemistry Program Exchange program 501). *QCPE Bull.* **1985**, *5*, 139.

0002-7863/91/1513-8305$02.50/0

spirit of AM1) based on a continuum model of solvent effects. Such a treatment is more economical and will be especially useful for cases where simulations have unacceptably large errors due to deficiencies in the potential function or the fact that many-body electronic polarization effects are neglected. Just as parameterized molecular orbital theory based on single-configuration wave functions[6–8] is in some sense a provisional or temporary practical solution as compared to the eventually preferable approach of large-scale ab initio configuration interaction or many-body perturbation theory calculations,[1,4] such a parameterized treatment of solvent effects is also a "temporary" practical approach. However, for many systems of interest such practical approaches may still be the only feasible approaches well into the 21st century, and they may be the economical choice for many studies even when large-scale calculations are technically possible; thus, this kind of approach appears to us to be one of the most promising areas for current research.

In the present paper, we present two general parameterizations of aqueous solvation free energies based on a continuum treatment of solvation and semiempirical molecular orbital theory for solute atomic charges. We chose to model the solute with AM1 theory,[7] and our treatment of solute–solvent interactions is an extension of the method of Still and co-workers.[19] Their treatment combines surface tension terms,[20,21] based on solvent-accessible surface area,[20–28] for dispersion and cavity effects with the generalized[29–33] Born[34] model for polarization effects. The surface tension terms dominate in the case of hydrophobic interactions, and the polarization effects dominate for solvation free energies of polar and charged solutes.

Still and co-workers applied their solvation model with molecular mechanics charges,[35] by which we mean charges that are constant for a given atom type and chemical environment, e.g., alcoholic oxygens, and with approximate accessible surface areas. One advantage of their approach is that it allows for rapid evaluation of gradients[26] and hence for convenient stationary point analyses. Our generalization to include molecular orbital charges is based on a variational self-consistent-field (SCF) approach to solvation,[31,32] which is closely related to the solvaton model, self-consistent reaction field theory, and other similar methods.[31,36–41] The approach we have chosen has several advantages. For example, it allows the prediction of substituent effects, the treatment of reactive intermediates and transition states, and the treatment of solvent-induced charge reorganization. A critical difference from most previous work in which solvent effects were incorporated in the Fock operator for SCF calculations[31,32,36–38,40,41] is that our treatment includes not only mutual solute–solvent polarization effects of the solute-induced reaction fields[29–34,36,37,40,42] but also semiempirical approximations to surface area dependent cavity and dispersion terms. Since the latter contain parameters adjusted to experiment, they include not only the free energy of cavity formation, which is positive, but also the free energy of solvent reorganization[43] and dispersion energy when the solute is placed in the cavity; the sum of the latter terms is negative. In the parameterization presented here, these terms do not depend on the solute charge distribution so they affect the final energy and optimized geometry but not the SCF orbitals. Alternative approaches, involving the inclusion of dispersion in the SCF equations, have been presented by Thole and Van Duijnen[40d] and Rinaldi et al.,[44,45] and earlier work including van der Waals interactions and solvent-accessible surface area in semiempirical quantum mechanical solvation models was presented by Warshel.[39]

---

(14) Komornicki, A.; Flanigan, M. C.; McIver, J. W., Jr. In *Semiempirical Methods of Electronic Structure Calculation, Part B: Applications*; Segal, G. A., Ed.; Plenum: New York, 1977; p 1.

(15) (a) Watts, R. O.; Clementi, E.; Fromm, J. *J. Chem. Phys.* **1974**, *61*, 2550. (b) Clementi, E.; Ranghino, G.; Scordamaglia, R. *Chem. Phys. Lett.* **1977**, *49*, 218. (c) Clementi, E.; Corongiu, G. *Biopolymers* **1981**, *20*, 551. (d) Clementi, E.; Corongiu, G. *J. Biol. Phys.* **1983**, *11*, 33. (e) Kim, K. S.; Clementi, E. *J. Phys. Chem.* **1985**, *89*, 3655.

(16) (a) Owicki, J. C.: Scheraga, H. A. *J. Am. Chem. Soc.* **1977**, *99*, 7413. (b) Swaminathan, S.; Harrison, S. W.; Beveridge, D. L. *J. Am. Chem. Soc.* **1978**, *100*, 5705. (c) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1979**, *101*, 1913. (d) Okazaki, S.; Nakamishi, K.; Touhara, H.; Adachi, Y. *J. Chem. Phys.* **1979**, *71*, 2421. (e) Mehrota, P. K.; Beveridge, D. L. *J. Am. Chem. Soc.* **1980**, *102*, 4287. (d) Nêmethy, G.; Peer, W. J.; Scheraga, H. A. *Annu. Rev. Biophys. Bioeng.* **1981**, *10*, 459. (e) Ravishanker, G.; Mezei, M.; Beveridge, D. L. *Faraday Symp. Chem. Soc.* **1982**, *17*, 79. (f) Chandrasekhar, J.; Spellmeyer, D. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 903. (g) Heinzinger, K. *Pure Appl. Chem.* **1985**, *57*, 1031. (h) Alagona, G.; Ghio, C.; Kollman, P. *J. Am. Chem. Soc.* **1986**, *108*, 185.

(17) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.

(18) (a) Jorgensen, W. L. In *Supercomputer Research in Chemistry and Chemical Engineering*; Jensen, K. F., Truhlar, D. G., Eds.; American Chemical Society: Washington, DC, 1987; p 200. (b) Jorgensen, W. L. *Adv. Chem. Phys.* **1988**, *70*, 469.

(19) Still, W. C.; Tempczak, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(20) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.

(21) (a) Amidon, G. L.; Yalkowsky, S. H.; Anik, S. T.; Valvani, S. C. *J. Phys. Chem.* **1975**, *79*, 2239. (b) Rose, G. D.; Geselowitz, A. R.; Lesser, G. J.; Lee, R. H.; Zehfus, M. H. *Science* **1985**, *229*, 834. (c) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199. (d) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086.

(22) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(23) Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973**, *79*, 351.

(24) Hermann, R. B. *J. Phys. Chem.* **1975**, *79*, 163.

(25) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.

(26) (a) Wodak, S. J.; Janin, J. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1736. (b) Richmond, T. J. *J. Mol. Biol.* **1984**, *178*, 63. (c) Hasel, W.; Hendrickson, T. F.; Still, W. C. *Tetrahedron Comput. Methodol.* **1988**, *1*, 103.

(27) Connelly, M. J. *Science* **1983**, *221*, 709.

(28) Jorgensen, W. L.; Gao, J.; Ravimohan, C. *J. Phys. Chem.* **1985**, *89*, 3470.

(29) (a) Hoijtink, G. J.; de Boer, E.; van der Meij, P. H.; Weijland, W. P. *Recl. Trav. Chim. Pays-Bas* **1956**, *75*, 487. (b) Peradejordi, F. *Cah. Phys.* **1963**, *17*, 393.

(30) Jano, I. *C. R. Acad. Sci. (Paris)* **1965**, *261*, 103.

(31) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Veillard, A., Eds.; Reidel: Dordrecht, 1980; Vol. II, p 25.

(32) (a) Fischer-Hjalmars, I.; Hendriksson-Enflo, A.; Hermann, C. *Chem. Phys.* **1977**, *24*, 167. (b) Constanciel, R.; Tapia, O. *Theor. Chim. Acta* **1980**, *55*, 77. (c) Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1. (d) Contreras, R.; Aizman, A. *Int. J. Quantum Chem.* **1985**, *27*, 293. (e) Kozaki, T.; Morihasi, M.; Kikuchi, O. *J. Am. Chem. Soc.* **1989**, *111*, 1547.

(33) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164.

(34) (a) Born, M. *Z. Physik* **1920**, *1*, 45. (b) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588.

(35) (a) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276. (b) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(36) (a) Newton, M. D. *J. Chem. Phys.* **1973**, *58*, 5833. (b) Newton, M. D. *J. Phys. Chem.* **1975**, *79*, 2795. (c) Hylton, J.; Christofferson, R.; Hall, G. G. *Chem. Phys. Lett.* **1974**, *24*, 501. (d) Hylton-McCreery, J.; Christofferson, R. E.; Hall, G. G. *J. Am. Chem. Soc.* **1976**, *98*, 7191.

(37) (a) Tapia, O.; Goscinski, O. *Mol. Phys.* **1975**, *29*, 1563. (b) Tapia, O.; Poulain, E.; Sussman, F. *Chem. Phys. Lett.* **1975**, *33*, 65. (c) Tapia, O.; Sussman, F.; Poulain, E. *J. Theor. Biol.* **1978**, *71*, 49. (d) Tapia, O.; Silvi, B. *J. Phys. Chem.* **1980**, *84*, 2646.

(38) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(39) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640.

(40) (a) Rivail, J.-L.; Rinaldi, D. *Chem. Phys.* **1976**, *18*, 233. (b) Klopman, G.; Andreozzi, P. *Theor. Chim. Acta* **1980**, *55*, 77. (c) Muertius, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117. (d) Thole, T.; van Duijnen, P. T. *Theor. Chim. Acta* **1980**, *55*, 307. (e) Tapia, O. In *Molecular Interactions*; Rataczak, H., Orville-Thomas, W. J., Eds.; John Wiley & Sons: New York, 1982; Vol. 3, p 47. (f) Bonaccorsi, R.; Cimaraglia, R.; Tomasi, J. *J. Comput. Chem.* **1983**, *4*, 567. (g) Mikkelson, K. V.; Dalgaard, E.; Swanstrøm, P. *J. Phys. Chem.* **1987**, *91*, 308. (h) Karlstrøm, G. *J. Phys. Chem.* **1988**, *92*, 1315. (i) Hoshi, H.; Sakurai, M.; Inoue, Y.; Chûjô, R. *J. Mol. Struct.*: THEOCHEM **1988**, *180*, 267. (j) Rivail, J. L. In *New Theoretical Concepts for Understanding Organic Reactions*; Bertrân, J., Csizmadia, I. G., Eds.; Kluwer: Dordrecht, 1989; p 219. (k) Bertrân, J., *Ibid.*; p 231. (l) Karelson, M. M.; Katritzky, A. R.; Szafran, M.; Zerner, M. C. *J. Org. Chem.* **1989**, *54*, 6030. (m) Stienke, T.; Hänsele, E.; Clark, T. *J. Am. Chem. Soc.* **1989**, *111*, 9107. (n) van Duijnen, P. T.; Rullmann, J. A. C. *Int. J. Quantum Chem.* **1990**, *38*, 181. (o) Solâ, M.; Lledôs, A.; Doran, M.; Bertrâm, J.; Abboud, J.-L. M. *J. Am. Chem. Soc.* **1991**, *113*, 2873.

(41) Bash, P. A.; Field, M. J.; Karplus, M. *J. Am. Chem. Soc.* **1987**, *109*, 8092.

(42) (a) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486. (b) Kirkwood, J. E. *J. Chem. Phys.* **1939**, *7*, 911. (c) Huron, M.-J.; Claverie, P. *J. Phys. Chem.* **1974**, *78*, 1853. (d) Beveridge, D. L.; Schnuelle, G. S. *J. Phys. Chem.* **1975**, *79*, 2562.

(43) Pohorille, A.; Pratt, L. R.; Burt, S. K.; MacElroy, R. D. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1257.

(44) (a) Rinaldi, D.; Costa Cabrel, B. J.; Rivail, J. L. *Chem. Phys. Lett.* **1986**, *125*, 495. (b) Rinaldi, D.; Costa Cobral, B. J.; Rivail, J. L. *Chem. Phys. Lett.* **1986**, *156*, 125.

(45) See also: Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616.

To avoid confusion, we note that there are two definitions of solvent-accessible surface area—the original one of Lee and Richards[22] and Hermann[20] based on the surface mapped out by the center of a spherical solvent molecule rolling on the atomic van der Waals spheres of the solute and a later one, also discussed by Richards[25] and favored by Connelly,[27] which is defined in terms of the surfaces of contact and the reentrant surfaces.[46] We used the original definition, and we call it the accessible surface area, although Hermann's nomenclature—the cavity surface area[20,24]—is more closely related to the physical justification for making the hydrophobic or cavitation free energy proportional to it, namely, that this free energy contribution is assumed[47] to be proportional to the number of water molecules *in the first hydration shell*.

The justification for making the dispersion energy proportional to surface area requires the use of Gauss's divergence theorem to convert an integral over the interaction volume outside the solute into a surface integral.[45,48] Since we find the accessible surface tension semiempirically, we don't actually use either of these justifications numerically.

We recognize two somewhat distinct types of usage for a general parameterization of the type proposed here. Firstly, it can be used to predict solvation free energies of reactive intermediates or of generalized transition states along reaction paths for dynamics studies. Secondly, it can be used to predict solvation free energies for various stable molecules in a variety of applications; such free energies can be used to predict solubilities or solvation effects on conformational equilibria and isomerizations. A parameterization of the first type must not include parameters that depend explicitly on hybridization or chemical environment since these may change during the course of a reaction or may be ambiguous for intermediates, and this limitation somewhat increases the errors when such a parameterization is used for type 2 applications. Thus, we present two general parameterizations in this article: Solvation Model 1 (SM1) restricts the parameters (but not the charges) for a given atom type, say oxygen, to be independent of chemical environment, and therefore it can be applied unambiguously to species with exotic bonding or to transition states. Solvation Model 1a (SM1a) allows different parameters for say alcholic and aldehydic oxygens, and it gives a better representation (in a root-mean-square-error sense) of static solvation free energies than the more general SM1 when applied to stable neutral molecules. Since both parameterizations are built on the AM1 semiempirical values for unsolvated solute parameters, the two sets of parameters are denoted below as AM1-SM1 and AM1-SM1a.

The final data set for the AM1-SM1 parameterization includes alkanes, cycloalkanes, alkenes, arenes, alkynes, ethers, heterocycles, carboxylic acids, esters, nitriles, aldehydes, ketones, alcohols, amines, nitro compounds, sulfides, thiols, halides, polyfunctional compounds, cations, and anions—for a total of 168 solutes. The AM1-SM1a parameters are based on the subset of 141 neutrals.

Our procedures have all been incorporated in a computer program we call AMSOL, which is an extended version of the widely available AMPAC.[11] We are placing AMSOL in the public domain and making it available through the Quantum Chemistry Program Exchange.[49]

2. Theory

We start with an expression for a well-defined portion G_S of the free energy G of a solute in a specific electronic state. This portion includes the adiabatic electronic and nuclear repulsion energy of the solute plus a dominant part of the free energy of solvation at temperature T. Although the theory is more general, we limit our explicit attention to the ground electronic state and room temperature. We partition the included part of the standard-state free energy of the solution G_S^0 into two terms:[19]

$$G_S^0 = G_{ENP} + G_{CD}^0 \tag{1}$$

where G_{ENP} includes the SCF electronic (E) energy, solute nuclear (N) repulsion, and the solute–solvent polarization (P) terms, both electrostatic and inductive, and G_{CD}^0 includes the standard-state free energy of cavity (C) creation in the solvent plus the solute–solvent dispersion (D) interactions, and an empirical part of the nuclear motion free energy change delineated below.

Since the usual[2,4] SCF formalism for gas-phase molecules includes the E and N terms, we must augment it by the P and CD terms. We do this using the variational approach of Tapia,[31] representing the P terms by the generalized Born formula,[29-33] which is a function of the charges q_k of the atoms, and representing the CD terms by[19-21]

$$G_{CD}^0 = \sum_{k=1}^{N} \sigma_k A_k \tag{2}$$

where N is the number of atoms in the solute, A_k is the accessible surface area[20-26] of atom k, and σ_k is a parameter for atom k that will be called the accessible surface tension. The generalized Born formula is [29-33,50]

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{k=1}^{N}\sum_{k'=1}^{N} q_k q_{k'} \gamma_{kk'} \tag{3}$$

where ϵ is the dielectric constant (relative permitivity of the solvent, and $\gamma_{kk'}$ is a Coulomb integral, and

$$G_{ENP} = E_{EN} + G_P \tag{4}$$

where E_{EN} is the ground-state electronic energy and nuclear repulsion for the solute.

As in the original Born theory[34] for spherical atoms, the diagonal Coulomb integrals are given by

$$\gamma_{kk} = \frac{1}{\alpha_k} \tag{5}$$

where α_k is an effective atomic radius, which depends on chemical environment according to a prescription suggested by Still et al.[19] For monatomic ions, α_k is set equal to an empirical parameter ρ_k called the coulomb radius. For more complicated ions, we determine α_k as the radius that when substituted in eq 3 with $N = 1$, gives the same polarization energy as calculated for that atom in the given chemical environment under the assumption that all the other atoms in the solute have no effect except to displace the dielectric medium from the volume calculated with a radius of ρ_k.[19] Still et al. set ρ_k equal to the sum of the atomic van der Waals radius R_k and an empirically determined (negative) offset, whereas in our model ρ_k is a function of the partial charge as well as the atomic number.

We used a modified form of the semiempirical functional form suggested by Still et al.[19] for $k \neq k'$, in particular

$$\gamma_{kk'} = \frac{1}{\sqrt{r_{kk'}^2 + \alpha_k \alpha_{k'}[\exp(-r_{kk'}^2/d_{kk'}\alpha_k\alpha_{k'}) + C_{kk'}(r_{kk'})]}} \tag{6}$$

where $r_{kk'}$ is the distance from atom k to atom k', $d_{kk'}$ is a constant, and $C_{kk'}$ is an optional, localized function of $r_{kk'}$ explained below. This expression differs from the Ohno–Klopman formula[51] we used previously[33] in that the arithmetic mean of α_k and $\alpha_{k'}$ is replaced by a geometric mean, and the exponential damping factor and localized function are introduced. The first two modifications were introduced by Still et al.;[19] the third is new.

In the restricted Hartree–Fock approximation[2,4] for the valence orbitals of the solute, the portion of the free energy we are considering may be written

$$G_S^0 = \frac{1}{2}\sum_{\mu=1}^{m}\sum_{\nu=1}^{m} P_{\mu\nu}[H_{\mu\nu} + F_{\mu\nu}^{(0)}] + \sum_{k=1}^{N}\sum_{k'=1}^{N} E_{kk'} \tag{7}$$

---

(46) See also: Pascual-Ahuir, J. L.; Silla, E.; Tomasi, J.; Bonaccorsi, R. *J. Comput. Chem.* **1987**, *8*, 778.

(47) Nemethy, G.; Scheraga, H. A. *J. Chem. Phys.* **1962**, *36*, 3401.

(48) Huron, M. J.; Claverie, P. *J. Phys. Chem.* **1972**, *76*, 2123.

(49) Cramer, C. J.; Truhlar, D. G. AMSOL, version 1, program 606, Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN.

(50) All equations are in hartree atomic units.

(51) (a) Ohno, K. *Theor. Chim. Acta* **1964**, *2*, 219. (b) Klopman, G. *J. Am. Chem. Soc.* **1964**, *86*, 4550.

where μ and ν are valence atomic orbital indices, m is the number of valence atomic orbitals, **P**, **H**, and $\mathbf{F}^{(0)}$ are the density, one-electron, and Fock matrices in the absence of solvent, and $E_{kk'}$ contains the nuclear repulsion and solvation terms. In particular

$$E_{kk'} = \frac{1}{2}\left[\bar{Z}_k\bar{Z}_{k'}g_{kk'} - \left(1 - \frac{1}{\epsilon}\right)q_kq_{k'}\gamma_{kk'}\right] + \delta_{kk'}\sigma_kA_k \qquad (8)$$

where $\bar{Z}_k$ is the valence charge of atom k (equal to the nuclear charge minus the number of core electrons), q_k is the partial charge on atom k, and

$$g_{kk'} = \begin{cases} 1/r_{kk'}, & k \neq k' \\ 0, & k = k' \end{cases} \qquad (9)$$

We obtain the partial charges by Mulliken population analysis:[52]

$$q_k = \bar{Z}_k - \sum_{\mu\epsilon k} P_{\mu\mu} \qquad (10)$$

Defining the new Fock matrix by $F_{\mu\nu} = \partial G_S/\partial P_{\mu\nu}$ with fixed $\gamma_{kk'}$ and A_k then yields[31]

$$F_{\mu\nu} = F_{\mu\nu}^{(0)} + \delta_{\mu\nu}\left(1 - \frac{1}{\epsilon}\right)\sum_k\sum_{\mu'\epsilon k}(\bar{Z}_k - P_{\mu'\mu'})\gamma_{k'k} \qquad \mu \in k' \qquad (11)$$

SCF calculations are carried out in the usual way[2,4,11,12] with this Fock matrix, and G_S^0 is evaluated from the converged orbitals by using eq 7.

The free energy of solvation depends on standard state and temperature. The present parameterization is based on the standard free energy change for transfer from a 1 M ideal gas to a 1 M ideal aqueous solution at 298 K.[53]

It is important to reiterate that the final result contains the adiabatic electronic energy, the solute nuclear repulsion, and the polarization, cavity, and dispersion parts of the free energy of solvation. Since G_{CD}^0 is empirical and does not fit the experimental data exactly, there is some flexibility in how to interpret it. We interpret it as including not only cavity and dispersion free energies, but also the change in PV terms upon dissolution, the loss of translational and rotational free energy from the gas phase, and the corresponding gain in librational free energy in solution, but not as including the vibrational contributions or that part of the electronic free energy due to electronic excitation. The missing electronic and vibrational contributions could be calculated by sums of states (or for vibration, by the analytic harmonic approximation) from energies or frequencies computed in the presence of solvent. Thus, we approximate the standard-state free energy of solvation as

$$\Delta G_S^0 = G_S^0 - E_{EN}(g) + \Delta G_{vib} + \Delta G_{elec} \qquad (12)$$

where (g) denotes a value in the gas phase, Δ denotes a difference between aqueous solution and the gas phase, $E_{EN}(g)$ is the gas-phase electronic and nuclear repulsion energy of the solute, ΔG_{vib} is the change in vibrational free energy (including both zero-point and thermal effects) in the N_{vib} vibrational modes of the solute ($N_{vib} = 3N - 6$ or $3N - 5$ for molecules and 0 for atoms), and ΔG_{elec} is the change in electronic excitation free energy. *In the present paper*, in comparing to experiment for solvation free energies, we will simply assume that the solution-phase electronic-excitation and vibrational contributions are canceled by the gas-phase electronic-excitation and vibrational contributions. (Note especially that we do not claim that the vibrational zero-point or thermal contributions are absorbed in the empirical parameters. It will often be useful or necessary to explicitly incorporate the vibrational contributions in later work, e.g., to calculate isotope effects[54] or simply for higher accuracy, and we plan to do this. In fact, the present scheme is specifically designed with the intention that one should be able to use it to calculate vibrational frequencies in the presence of solvent.)

Since we parameterize the SCF results using experimental data, they implicitly include electron correlation and configuration mixing effects (just as AM1 does for gas-phase solute energies). Some workers have used similar models in which configuration mixing is explicit.[55] The empirical parameterization also means that hydrophobic effects are included, presumably in G_{CD}^0.

3. Numerical Methods

As mentioned in the Introduction, the calculations were performed by adding solvent terms to AMPAC,[11] resulting in a new computer program called AMSOL.[49]

The slowest step in the calculations is the calculation of the effective Born radii α_k. This calculation is carried out as follows.[19]

$$\alpha_k^{-1} = \sum_{i=1}^{M}\frac{A_i}{4\pi r_i^2}\left(\frac{1}{r_i - 0.5T_i} - \frac{1}{r_i + 0.5T_i}\right) + r_{M+1} \qquad (13)$$

where r_i and T_i are defined by eqs 14 and 15,

$$r_i = \begin{cases} r_k, & i = 1 \\ r_{i-1} + {}^1/_2(T_{i-1} + T_i), & i > 1 \end{cases} \qquad (14)$$

$$T_i = \begin{cases} 0.1\ \text{Å}, & i = 1 \\ 1.5T_{i-1}, & i > 1 \end{cases} \qquad (15)$$

and A_i is the exposed surface area of the atomic sphere with radius r_i (calculated numerically). Thus, the incremental contribution from expanding spherical shells about atom k, beginning with the charge-dependent coulomb radius ρ_k, is summed to yield α_k^{-1}. The summation limit M is reached when the shell with inner radius $r_i - 0.5T_i$ encompasses the entire molecular surface as defined by all the various ρ_k.

The Coulomb integrals α_k are updated on the first, fourth, ninth, etc. iteration, i.e., on square iteration numbers, in the SCF process. The partial charges are updated every iteration.

The atom-specific solvent-accessible surface areas needed for G_{CD} are calculated as the exposed surface area of the atom-centered sphere with radius $R_k + R_S$, where R_k is the van der Waals radius of the atom and R_S is the solvent radius. These surface area terms are calculated exactly, i.e., without the analytic approximations employed by Still et al.

In the present version of AMSOL, all first derivatives of the energy are calculated numerically in Cartesian coordinates.[11] The fact that G_P and G_{CD} are not sums of contributions from atomic pairs, as are the AM1 contributions to the energy, causes no problems in this algorithm, but does change the algorithm. In almost all cases forward differences are sufficient for the SCF calculations, and the two-point central difference algorithm of AMPAC is adequate for the energy derivative with respect to nuclear coordinates, but in those few cases where the method fails and leaves a high gradient norm after optimization, rerunning the calculation with a randomly perturbed geometry usually corrects the problem.

All calculations were carried out on Cray computers at Minnesota Supercomputer Center and the Ballistics Research Laboratory.

4. Results

4.1. AM1–SM1. The parameters were determined as follows. The gas-phase Fock operator was evaluated by using Austin Model 1 (AM1).[7]

The damping parameter $d_{kk'}$ was assumed to be independent of atom type, charge, and chemical environment and was given the value used by Still et al.,[19] namely, $d_{kk'} = 4$. We did double check whether this value is close to optimum, and we found that it is. The solvent radius R_S was set equal to 1.4 Å.[22]

The parameters in G_P were primarily adjusted to data for ions. The coulomb radii were assumed to depend on the charge and were written as

$$\rho_k = \rho_k^{(0)} + \rho_k^{(1)}\left[-\frac{1}{\pi}\arctan\frac{q_k + q_k^{(0)}}{q^{(1)}} + \frac{1}{2}\right] \qquad (16)$$

---

[52] Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

[53] Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.

[54] Jansco, G.; Van Hook, W. A. *Chem. Rev.* **1974**, *74*, 689.

[55] See, e.g.: (a) Beens, H.; Weller, A. *Chem. Phys. Lett.* **1969**, *3*, 666. (b) Yomosa, S. *J. Phys. Soc. Jpn.* (*Nippon Batsuri Gakkai*) **1973**, *35*, 1738; **1974**, *36*, 1655; **1978**, *44*, 602. (c) Tapia, O.; Poulain, E.; Sussman, F. *Theor. Chim. Acta* **1978**, *47*, 171. (d) Mikkelson, K. V.; Ågren, H.; Jensen, H. J. A.; Helgaker, T. *J. Chem. Phys.* **1988**, *89*, 3086.

Table I. Semiempirical Parameters for Determining the Coulomb Radius

k	$\rho_k^{(0)}$, Å	$\rho_k^{(1)}$, Å	$q_k^{(0)}$
H	0.57	1.303	−0.30
C	1.68	0.000	*a*
N	1.40	0.620	0.40
O	1.46	−0.250	0.75
F	1.37	0.181	0.70
S	1.30	0.800	0.70
Cl	1.65	0.618	0.75
Br	1.75	0.705	0.70
I	1.88	0.932	0.60

[a] Not required.

Table II. COG Parameters

k	k′	$d_{kk'}^{(1)}$	$d_{kk'}^{(2)}$	$r_{kk'}^{(1)}$, Å	$r_{kk'}^{(2)}$, Å
O	O	9.0	1.75	1.7	0.9
N	H	4.0	1.75	1.9	0.5

where q_k is the partial AM1 charge on atom k (in units of the proton charge), and $\rho_k^{(0)}$, $\rho_k^{(1)}$, $q_k^{(0)}$, and $q^{(1)}$ are constants. This expression makes ρ_k a smooth, sigmoidal function of atomic charge, with $q^{(1)}$ determining how rapidly the transition between the limiting $\rho_k^{(0)}$ and $\rho_k^{(0)} + \rho_k^{(1)}$ values occurs. We assumed that $q^{(1)}$ is independent of atom type and roughly optimized its value, obtaining 0.1. The remaining parameters were obtained by fitting the model to selected simple compounds, in which the atom of interest has a variety of charges. For example, the experimental hydration free energies for H^+ and H^- suggest $\rho_k^{(0)}$ and $\rho_k^{(1)}$ values of approximately 0.64 and 1.22 Å. Consideration of additional solutes then permits refinement of these values and optimization of $q_k^{(0)}$ for hydrogen. For Cl and Br, we also used data on aqueous free energies of activation[56] in adjusting $q_k^{(0)}$. The resulting final parameters are collected in Table I by atom type.

For two of the atom type pairs, the optional $C_{kk'}$ function was included in the damping function of eq 6 to reduce systematic errors. The parameters modified this way were γ_{OO} and γ_{NH}. The modifications were carried out by introducing cutoff Gaussians (COGs):

$$C_{kk'} = \begin{cases} d_{kk'}^{(1)} \exp\left\{-d_{kk'}^{(2)} \middle/ 1 - \left[\left(r_{kk'} - r_{kk'}^{(1)}\right) \middle/ r_{kk'}^{(2)}\right]^2\right\}, & \left|r_{kk'} - r_{kk'}^{(1)}\right| < r_{kk'}^{(2)} \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

COGs were originally introduced[57] in an entirely different context (as basis functions for quantum mechanical scattering calculations) because they have the desirable property of being completely localized (i.e., identically zero except for a finite range of their argument) but also have an infinite number of continuous derivatives. This property is equally useful here, where continuous derivatives are important for geometry optimization. (In addition, continuous parameters are required if the computed potential energy functions are to be continuous functions of geometry, which facilitates the use of high-order algorithms in dynamics calculations that involve the potential energy function in differential equations and/or quadratures.) The parameters we adopted for the COGs are given in Table II.

The final parameters are the accessible surface tensions σ_k. The van der Waals radii needed to calculate the accessible surface areas were taken from Bondi.[58] In AM1-SM1, the σ_k depend only on atom type, i.e., atomic number. We set $\sigma_H = 0$ since least-squares adjustment yields different signs for different classes of solutes, e.g., hydroxyl protons vs hydrocarbon protons. The final values of the other σ_k parameters were optimized for a data set involving SCF optimizations for 141 neutral compounds (including the solvent terms in the Hamiltonian) via multiple linear regression over all atom type environments. The data used for fitting was taken from the tables of Cabani et al.,[53] Ben-Naim and Marcus,[59] and Pearson;[60] experimental data from these or additional[61] sources are not completely compatible, but it is beyond the scope of the present study or the accuracy of our parameterization to address such issues. Table III presents the final σ_k parameters for the AM1-SM1 model.

Table III. Accessible Surface Tensions[a]

k	σ_k, cal mol^{-1} Å^{-2}	k	σ_k, cal mol^{-1} Å^{-2}
H	0.00	S	−18.80[b]
C	14.95	Cl	−2.14
N	−73.65	Br	−9.11
O	−52.74	I	−8.21[b]
F	18.47		

[a] AM1-SM1. [b] These values were determined from relatively small data sets.

Table IV. Calculated and Experimental Free Energies for Solvation of Anions[a]

ion	ΔG_{ENP}	G_{CD}^0	ΔG_S^0	ΔG_S^0 (exptl)
H^-	−89.0	0.0	−89.0	−89
F^-	−106.9	1.9	−105.0	−105
Cl^-	−74.7	−0.3	−75.0	−75
Br^-	−68.8	−1.2	−70.0	−70
I^-	−59.8	−1.2	−61.0	−61
OH^-	−105.0	−4.2	−109.2	−106
CN^-	−75.9	−3.1	−79.0	−77
O_2^-	−84.4	−6.8	−91.2	−87
HS^-	−73.7	−1.9	−75.6	−76
C_2H^-	−78.4	1.8	−76.6	−73
HO_2^-	−95.8	−5.5	−101.4	−101
N_3^-	−63.6	−11.3	−74.9	−74
NO_2^-	−74.8	−8.6	−83.4	−72
CH_3O^-	−86.1	−2.8	−88.9	−95
NO_3^-	−55.4	−9.1	−64.5	−65
CH_2CN^-	−64.9	−3.5	−68.4	−75
$CH_3CO_2^-$	−74.5	−5.0	−79.5	−77

[a] AM1-SM1; kcal mol^{-1}.

Table V. Calculated and Experimental Free Energies for Solvation of Cations[a]

ion	ΔG_{ENP}	G_{CD}^0	ΔG_S^0	ΔG_S^0 (exptl)
H_3O^+	−100.6	−1.6	−102.1	−104
H_3S^+	−76.3	−1.1	−77.4	−87
NH_4^+	−77.8	−0.4	−78.2	−79
$CH_3OH_2^+$	−79.8	−0.8	−80.5	−83
$CH_3SH_2^+$	−73.6	−0.8	−74.4	−74
$CH_3NH_3^+$	−69.7	0.0	−69.6	−70
$(CH_3)_2OH^+$	−61.3	−0.2	−61.5	−70
$(CH_3)_2NH_2^+$	−61.0	0.3	−60.7	−63
$(CH_3)_3NH^+$	−53.2	0.3	−52.9	−59
$CH_3CONH_3^+$	−70.3	−2.3	−72.6	−66

[a] AM1-SM1; kcal mol^{-1}.

For all calculations, geometries were separately optimized in the gas phase and in solution to calculate the solvation free energies. Tables IV and V present the model predictions for ΔG_{ENP}, G_{CD}^0, and ΔG_S^0 for the ions in our study, and they compare the ΔG_S^0 values to experiment. The ΔG_{ENP} values are defined as $\Delta G_S^0 - G_{CD}^0$, and they include the change in solute internal energy due to the geometry change upon dissolution. The root-mean-square error in the AM1-SM1 values for the 27 ions in Tables IV and V is 4.4 kcal mol^{-1}. Similar comparisons for selected neutrals are given in Table VI. Solvation free energies for the other neutrals

---

(56) McLennan, D. J. *Aust. J. Chem.* **1978**, *31*, 1897.

(57) Schwenke, D. W.; Truhlar, D. G. In *Computing Methods in Applied Sciences and Engineering*; Glowinski, R., Lichnewsky, A., Eds.; Society for Industrial and Mathematics: Philadelphia, 1990; p 291.

(58) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.

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

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

(61) See, e.g., experimental values cited in Abraham, M. H. *J. Chem. Soc., Perkin Trans. 2* **1973**, 1893. (b) Rao, B. G.; Singh, U. C. *J. Am. Chem. Soc.* **1989**, *111*, 3125.

Table VI. Calculated and Experimental Free Energies of Solvation for Selected Neutrals (kcal/mol)

	AM1-SM1			exptl	AM1-SM1a
	ΔG_{ENP}	G_S^0	ΔG_S^0	ΔG_S^0	ΔG_S^0
neopentane	0.1	0.4	0.5	2.5	1.3
cyclopropane	−0.2	0.5	0.4	0.8	0.7
ethene	−0.3	0.8	0.5	1.3	0.4
1,3-butadiene	−0.9	1.1	0.3	0.6	0.1
benzene	−1.9	1.2	−0.7	−0.9	−1.0
toluene	−1.9	1.1	−0.7	−0.9	−0.8
p-xylene	−1.8	1.1	−0.8	−0.8	−0.6
naphthalene	−3.5	1.7	−1.7	−2.4	−2.3
1-butyne	−1.9	1.1	−0.8	−0.2	−1.0
dimethyl ether	−1.2	−0.8	−2.0	−1.9	−2.7
1-methoxypropane	−0.9	−0.6	−1.5	−1.7	−1.7
tetrahydrofuran	−1.4	−1.0	−2.4	−3.5	−3.1
1,4-dioxane	−2.0	−2.1	−4.1	−5.1	−5.9
acetic acid	−2.3	−3.8	−6.1	−6.7	−7.1
propanoic acid	−1.6	−3.4	−5.1	−6.5	−6.0
methyl acetate	−2.3	−2.4	−4.6	−3.3	−4.3
ethyl acetate	−2.2	−2.0	−4.3	−3.1	−3.3
methyl propanoate	−1.8	−2.0	−3.8	−2.9	−3.3
methyl butanoate	−1.7	−2.0	−3.7	−2.8	−3.1
acetonitrile	−2.1	−3.7	−5.7	−3.9	−4.3
propanal	−2.6	−2.0	−4.6	−3.5	−3.1
butanal	−2.5	−2.0	−4.5	−3.2	−2.8
benzaldehyde	−4.1	−1.1	−5.2	−4.0	−4.4
butanone	−2.8	−1.7	−4.5	−3.6	−2.9
ethanol	−1.0	−1.5	−2.5	−5.0	−4.4
trimethylamine	−3.3	−0.1	−3.4	−3.2	−4.9
pyridine	−3.9	−0.8	−4.7	−4.7	−4.3
4-methylpyridine	−4.0	−0.8	−4.8	−4.9	−4.2
1,4-dimethylpiperazine	−5.3	−0.5	−5.7	−7.6	−7.3
2-methylpyrazine	−4.8	−2.3	−7.2	−5.5	−6.0
3-ethyl-2-methoxy-pyrazine	−4.6	−1.3	−5.9	−4.4	−5.5
nitrobenzene	−3.2	−3.8	−7.0	−4.1	−4.8
diethyl sulfide	−0.1	−0.3	−0.4	−1.3	−0.4
methanethiol	−0.2	−1.0	−1.2	−1.2	−1.0
ethanethiol	−0.3	−0.9	−1.1	−1.3	−0.6
tetrafluoromethane	−0.6	3.4	2.9	3.1	3.5
1,1-difluoroethane	−2.3	1.8	−0.6	−0.1	−0.0
dichloromethane	−1.0	−0.2	−1.3	−1.4	−1.2
trichloromethane	−0.7	−0.4	−1.1	−1.1	−1.2
chlorofluoromethane	−1.8	0.8	−0.9	−0.8	−0.7
(*Z*)-1,2-dichloroethene	−1.1	0.1	−1.0	−1.2	−1.1
1,1,2-trichloroethane	−1.0	−0.4	−1.4	−2.0	−1.3
bis(2-chloroethyl) sulfide	−1.6	−0.8	−2.4	−3.9	−2.5
bromomethane	−0.3	−0.6	−1.0	−0.8	−0.7
1,2-dibromomethane	−0.6	−1.3	−1.8	−2.1	−1.4
iodoethane	0.0	−0.6	−0.6	−0.7	−0.7

are given in the supplementary material. The root-mean-square error in the solvation free energy for all 141 neutrals is 1.52 kcal mol^{-1}.

4.2. AM1-SM1a. The only difference introduced in Solvation Model 1a is in the σ_k parameters. In AM1-SM1a, σ_k was still assumed to be independent of charge, but now it is allowed to depend on functionality as well as atom type; e.g., alcoholic oxygens have values different from those of aldehydic oxygens. Again the σ_k were optimized by multiple linear regression for all 141 neutrals in the data set. The values obtained are given in Table VII. Solvation free energies for selected compounds are given in the last column of Table VI, and results for the other neutrals are given in the supplementary material. The root-mean-square error for the 141 neutrals is reduced to 0.78 kcal mol^{-1}.

5. Discussion

For ionic species, as easily observed in Tables IV and V, polarization effects far outweigh the accessible surface area terms, so ions are particularly sensitive to the parameters used for the calculation of Coulomb radii. While the rms error for the ions is encouraging, the individual errors for a few of the ions are significant.

Four of the ions in our data set were also included in the study of Still et al.[19] For these four ions, the rms error in their molecular mechanics treatment is 8.6 kcal mol^{-1}, whereas that for the present SCF treatment is 3.3 kcal mol^{-1}. Both methods should be very useful for future work.

Table VII. Accessible Surface Tensions for AM1-SM1a Parameterization

k	environment	σ_k, cal mol^{-1} Å^{-2}
H	–CH	4.15[a]
	–NH	58.96
	–OH	−23.39
	–SH	49.49[b]
C	all	4.15[a]
N	sp^3, amide	−368.97
	sp, sp^2, aromatic	−47.38
O	sp^3	−109.70
	sp^2	−25.61
F	all	21.17
S	all	−44.25[b]
Cl	all	−2.84
Br	all	−8.93
I	all	−13.42[b]

[a] Because of similar σ_k values in preliminary AM1-SM1a fits, in the final regression all carbons and all hydrogens bonded to carbon were treated as one class. [b] These values were determined from relatively small data sets.

The first thing we notice in the σ_k values is the wide range of values they take on. In the parameterization of Still and coworkers[19] with molecular mechanics charges based on the OPLS parameterization,[28] the σ_k values were all assigned as +7.2 cal mol^{-1} Å^{-2}. This difference results in part, of course, from the fact that their value represents an average over all atom types. The fact that their accessible surface tension has a positive value also results in part from the magnitudes of the OPLS partial charges assigned to heteroatoms. These were larger (and presumably less accurate) than the AM1 charges, so that the magnitude of the polarization term was overestimated compared to our value.

Another difference between our methods and those employed previously is that we treat all solute atoms explicitly, whereas Still et al.[19] treated CH$_n$ groups as single units in which hydrogen atoms were not explicitly represented. This too affects the σ_k values.

By the nature of the fitting process, the σ_k values attempt to correct for any systematic inherent error in the semiempirical atomic charges. Indeed this effect may potentially overwhelm the intended function of this term, namely, accounting for cavity and dispersion effects. Fitting the σ_k values only to neutrals alleviated this problem. As a result, most of our σ_k appear physically reasonable, with the possible exception of the value for sp^3 nitrogen atoms in the SM1a parameters.

The inclusion of the COGs corrected for two systematic deficiencies in preliminary fits: (i) The correction to γ_{OO} corrects for (previously) abnormally negative free energies of hydration of esters, acids, nitro groups, and peroxides. (ii) The correction to γ_{NH} corrects for (previously) abnormally positive free energies of hydration for alkylated amines. Thus, the O–O COG is tuned to bonded and geminal O,O pairs, and the N–H COG is tuned to geminal N,H pairs.

One noticeable qualitative deficiency remaining in the current parameterization is the AM1-SM1 underestimate of the hydrophobic effect for unsubstituted alkanes. This is not an inherent deficiency of the method, and the results for alkanes were much better in earlier trial parameterizations in which a larger fraction of the compounds were alkanes. The final parameters are fit to a wide range of data, with the motivation discussed in the Introduction, and we decided not to especially emphasize unsubstituted alkanes.

For the neutral solutes, there is an overlap of 18 molecules between our data set and the results reported by Still et al.[19] For these molecules, the rms errors are 0.75 kcal mol^{-1} for ref 19, 1.74 kcal mol^{-1} for AM1-SM1, and 0.66 kcal mol^{-1} for AM1-SM1a.

For the neutral molecules, the geometry change upon solvation is generally small, although the change in dipole moment may be significant.

6. Concluding Remarks

The overall performance of the new solvation model is quite good. We believe the approach will enable semiempirical molecular orbital theory to be applied to reactions in aqueous solution just as broadly as to gas-phase reactions.

We note that determining the Coulomb radii empirically and using eq 3 has important elements in common with other successful solvation models. For example, in the model reviewed by Warshel and Russell,[62] the vacuum intersection energy of charged groups is scaled by the reciprocal of the dielectric constant and added to the solvation energy of the charged groups at infinite separation. One of the advantages of the present treatment, as compared to formulations based more directly on Onsager's equations (e.g., the method of Karelson et al.[40l]), is that the distributed monopole charge distribution of the generalized Born model does not neglect higher multipoles, whereas Onsager's treatment is based entirely on the dipole moment. In addition, we have replaced the arbitrary cavity radius by an empirical radius, as done earlier, e.g., by Warshel[39] and also by Tucker and one of us.[33] Furthermore, the algorithm of Still et al.[19] provides a very flexible way of treating arbitrarily shaped solutes and buried or partially buried charges. An advantage of the present treatment over molecular-mechanics-based approaches is that, by putting the solvation terms in the Fock operator, it allows for applications to transition states and to solvent-induced charge transfer and solvent-induced charge redistribution.

It is appropriate to comment on possible extensions. One possibility is to use the Gibbs–Helmholtz relation to separate the free energy into enthalpic and entropic parts. This is difficult, though, because we do not know the temperature dependences of the parameters, especially σ_k and ρ_k. It is already well recognized, for example, that the simple Born model with temperature-independent α_k does not yield accurate entropies of hydration for monatomic ionic solutes.[63]

Another possible extension is to improve the performance for the current data set by further optimizing the functional forms, optimizing R_S, letting the σ_k depend on charge, or letting $q^{(1)}$ depend on atom types. Another possibility is to reoptimize the solvation parameters with more accurate charges, e.g., those obtained by ab initio calculations including electron correlation. (Despite the built-in dependence of the σ_k values on AM1 charges, discussed in section 5, it is also possible of course that using more accurate charges could improve the predictions even without reoptimizing parameters; it would be interesting to check this.) One could also consider alternatives to Mulliken population analysis for extracting partial charges from the density matrix. A physically appealing procedure for the present application is to obtain the q_k from the electrostatic potential,[64] but such an algorithm is not as well suited as Mulliken analysis to be added to the Fock operator.

In considering possible improvements to the solvation model, we should keep in mind that AM1–SM1 is already more accurate for solvation free energies than the typical error expected from AM1 or many ab initio techniques for gas-phase barrier heights, so one has already reached a situation of diminishing returns on this kind of improvement for the treatment of reaction dynamics. A more promising approach for quantitative studies of detailed reaction dynamics may be to combine SM1 with neglect-of-differential-overlap calculations with specific reaction parameters (NDDO-SRP).[65]

The present treatment is limited to equilibrium solvation effects, which are expected to be much larger than nonequilibrium effects for most reactions in water. The present model, however, may also be used for the first step of nonequilibrium solvation studies since the equilibrium solvation energy is a critical component of practical theories of nonequilibrium solvation.[66,67]

Although the parameterization presented here is specific to aqueous solution, the general approach could clearly be extended to other solvents, and it would be of interest to obtain parameters for such solvents as well. It seems unlikely that the present parameters for hydrogen-bonding water can simply be transferred to other solvents with only a change in the solvent radius and the dielectric constant. But water is of such unique interest that it certainly merits its own parameter set independent of all transferability considerations.

The solvation models presented here should be applicable not only to simple organic solutes such as considered so far but also to peptides, nucleosides, drug design, and proteins. For biochemical problems, it may be fruitful to use hybrid electronic structure models that treat substrates, reactive parts of large molecules, and active sites of biomolecules by molecular orbital theory and nonreactive but structurally important parts by molecular mechanics[38] or a classical description including the charges and induced dipoles[38,68,69] or effective dielectric constant[70] of the rest of a protein. Salt effects, which may be very important under physiological conditions, can be added by including Debye screening and an effective dielectric constant in eq 3.[71]

The AM1 model and its precursors introduced by Dewar and co-workers have revolutionized the molecular modeling of organic reactions and have clearly demonstrated the value of general parameterizations, and we have recently shown how general parameterizations may serve as useful starting points for even more quantitative dynamical studies by introducing specific reaction parameters. The present study extends the AM1 general parameterization to include aqueous solvation effects by combining it with the SCF local field formalism of Tapia and the solvation model introduced by Still and co-workers. We believe this opens up a wide new class of organic and biochemical reactions to systematic and detailed modeling.

Acknowledgment. This work was supported in part by the National Science Foundation and the USA-CRDEC ILIR program.

Supplementary Material Available: Calculated and experimental free energies of solvation for selected neutrals (3 pages). Ordering information is given on any current masthead page.

---

(62) Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283.

(63) Marcus, Y.; Loewenschuss, A. *Annu. Rep. Prog. Chem., Sect. C* **1984**, *81*, 81.

(64) (a) Smit, P. H.; Derisen, J. L.; van Duijenveldt, F. B. *Mol. Phys.* **1979**, *37*, 521. (b) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304. (c) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(65) Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618.

(66) Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 979.

(67) Lee, S.; Hynes, J. T. *J. Chem. Phys.* **1988**, *88*, 6853. Kim, H. J.; Hynes, J. T. *J. Chem. Phys.* **1990**, *93*, 5194.

(68) (a) Warshel, A. *Biochemistry* **1981**, *20*, 3167. (b) Warshel, A. *Acc. Chem. Res.* **1981**, *14*, 284. (b) Chung, A. K.; Weiss, R. M.; Warshel, A.; Takano, T. *J. Phys. Chem.* **1983**, *87*, 1683.

(69) Rullman, J. A. C.; Bellido, M. N.; van Duijnen, P. T. *J. Mol. Biol.* **1989**, *206*, 101.

(70) Sharp, K.; Honig, B. *Chem. Scr.* **1989**, *29A*, 71.

(71) Jayaram, B.; Sharp, K.; Honig, B. *Biopolymers* **1989**, *28*, 875.