8720 *J. Org. Chem.* **1996,** *61,* 8720−8721

A Universal Organic Solvation Model

David J. Giesen, Michael Z. Gu,
Christopher J. Cramer,* and Donald G. Truhlar*

*Department of Chemistry and Supercomputer Institute,
University of Minnesota,
Minneapolis, Minnesota 55455-0431*

Received September 10, 1996

The partitioning of a molecule between two phases has innumerable consequences in chemistry; for example, the partitioning between air and neat liquid determines vapor pressure (Henry's law), the partitioning between aqueous biophases and lipid bilayers is critical to drug delivery, and the differential partitioning of transition states and reactants between two media is the cause of solvent effects on the rates and product ratios of reactions. By the use of free energy cycles, the prediction of all such phenomena can be reduced to the calculation of free energies of solvation for various solutes in various solvents. Here we report a computational model for calculating such quantities for any molecule composed of H, C, N, O, F, S, Cl, Br, and/or I in essentially any organic solvent. The mean unsigned error is under 0.5 kcal.

Historically, most attempts to develop predictive models for solvation or transfer free energies have examined multivariate quantitative structure−property relationships and typically involve assigning additive fragment contributions to different functional groups.[1] An alternative approach is the simulation of a solute surrounded by explicit solvent;[2] because of the size of the system these simulations almost always use classical mechanical force fields. In general, neither of these two approaches explicitly includes either solute polarization or many-body effects. Furthermore, the former method tends to fail when a solute contains fragments not found in the parameterization set or when multiple fragments interact in a nonlinear way, while the latter method is very time-consuming, particularly when a new solvent model must be developed.

Recently, methods for including solvent electrostatic polarization effects in quantum mechanical solute descriptions, in which solute atoms are treated explicitly, the solvent is treated as a continuous fluid, and solute polarization by the solvent is included by a self-consistent reaction field (SCRF) have advanced vigorously.[3−7] Accurate quantitative models must also include nonelectrostatic effects, and we have previously developed such models for both water[8−10] and alkane[11] solvents. In the present paper we describe a parameterized SCRF scheme including both electrostatic and nonelectrostatic effects that encompasses essentially all organic solvents. Our parameterization set involves 1784 free energies for transfer of 205 diverse solutes from the gas phase to 90 diverse solvents, including 13 alkanes, 12 arenes, 14 alcohols, four ketones, two esters, seven ethers, two amines, three pyridines, two nitriles, four nitro compounds, two tertiary amides, 12 haloaliphatics, six haloaromatics, four miscellaneous acidic solvents, and three miscellaneous nonacidic solvents.

Table 1. Performance of SM5A-Organic Model by Solute Functional Group Class

	no. of			mean signed error[b] (kcal)	mean unsigned error[b] (kcal)
solute class	solutes	solvent classes[a]	data		
unbranched alkanes	9	15	76	0.2	0.4
branched alkanes	5	2	7	0.4	0.4
cycloalkanes	4	4	13	1.1	1.1
alkenes	8	3	18	0.4	0.5
alkynes	5	2	9	−0.1	0.3
arenes	9	15	126	−0.1	0.3
alcohols	17	15	369	0.1	0.4
ethers	9	15	71	−0.1	0.5
aldehydes	7	5	32	0.1	0.5
ketones	12	14	191	0.0	0.5
carboxylic acids	5	11	119	0.3	0.7
esters	12	6	227	−0.3	0.5
aliphatic amines	11	8	153	0.0	0.4
aromatic amines	11	10	71	−0.1	0.4
nitriles	4	4	18	−0.1	0.4
nitrohydrocarbons	6	6	32	0.0	0.6
ethanamide	1	3	4	−0.9	0.9
thiols	3	3	10	0.1	0.3
sulfides	4	3	10	0.0	0.4
disulfides	2	2	3	0.0	0.0
nonhalo bifunctional solutes	6	6	26	0.3	0.9
fluorohydrocarbons	5	3	13	0.4	0.5
chlorohydrocarbons	14	4	58	0.0	0.3
bromohydrocarbons	10	4	30	−0.1	0.3
iodohydrocarbons	9	4	20	0.0	0.2
multifunctional halogenated solutes	13	7	44	0.0	0.7
inorganic compds	4	9	34	−0.6	1.4
all solutes	205	15	1784	0.0	0.5

[a] Number of solvent classes for which there are data for this solute class. [b] In molar free energy of solvation at 298 K.

The basic framework of our model is the same as described previously.[8−11] There are three kinds of terms in the solvation free energy: long-range electrostatic contributions (labeled ENP, to denote that they include self-consistent solute electronic and nuclear contributions and solute-solvent electric polarization effects), intermediate-range cavity-structural (CS) contributions, and short-range cavity-dispersion (CD) and cavity-hydrogen-bonding (CH) effects. The model presented here is the same as our SM5.4/A-aqueous model[10] in the way it handles solute electrostatics and the solute-geometry dependence of cavity surface tensions; for these reasons we call the new scheme Solvation Model 5.4/A-organic (SM5.4/A-organic or OSM5.4/A).

The functional forms and parameters of the electrostatic model are identical to those presented elsewhere

---

(1) Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1975**, *40*, 287. Leo, A. J. *Chem. Rev.* **1993**, *93*, 1281. Cramer, C. J.; Famini, G. R.; Lowrey, A. H. *Acc. Chem. Res.* **1993**, *26*, 599.

(2) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395. van Gunsteren, W. F.; Luque, F. J.; Timms, D.; Torda, A. E. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 847. Jorgensen, W. L.; Tirado-Rives, J. *Perspect. Drug Discov. Des.* **1995**, *3*, 123.

(3) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.

(4) Cramer, C. J.; Truhlar, D. G. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1995; Vol. 6; p 1.

(5) Rivail, J.-L.; Rinaldi, D. In *Computational Chemistry, Review of Current Trends*; Leszczynski, J., Ed.; World Scientific: New York, 1996; Vol. 1; p 139.

(6) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán, J., Eds.; Kluwer: Dordrecht, 1996; p 1.

(7) Orozco, M.; Alhambra, C.; Barril, X.; López, J. M.; Busquest, M. A.; Luque, F. J. *J. Mol. Model.* [*Electronic Publication*] **1996**, *2*, 1.

(8) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305. Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213. Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1992**, *13*, 1089.

(9) Storer, J. W.; Giesen, D. J.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G.; Liotard, D. A. In *Structure and Reactivity in Aqueous Solution*; Cramer, C. J., Truhlar, D. G., Eds.; American Chemical Society: Washington, DC, 1994; p 24.

(10) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.

(11) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057. Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1995**, *99*, 7137.

S0022-3263(96)01742-2 CCC: $12.00

Table 2. Contributions to Calculated Molar Free Energy of Solvation from Hydrogen-Bonding, Other Interfacial Effects, and Electrostatics (kcal)

solute	solvent	CH	CD + CS	electrostatic	total[a]	expt
n-hexane	1-octanol	0.2	−3.0	−0.1	−3.0	−3.0
1-pentanol	1-octanol	−0.6	−2.3	−3.3	−6.3	−6.4
1-pentanol	*n*-hexadecane	0.0	−2.3	−1.8	−4.1	−4.2

[a] Calculated; sum of components may differ by 0.1 kcal owing to rounding.

for water,[10] except that the dielectric constant, ϵ, of the organic solvent replaces the dielectric constant of water. The electrostatic treatment involves a three-dimensional integration over the free energy density due to electric polarization of the solvent in the regions of space not occupied by the solute, and therefore, it reflects the solute shape realistically.[6,12] The solute electronic wave functions and solute internal energies are calculated with the semiempirical Austin Model 1[13] (AM1). The competition between solvent polarization and solute distortion is accounted for by placing solvation terms inside the effective one-electron Hamiltonian for the molecular orbitals.[6,14] The electrostatic solvation terms are calculated using class IV partial atomic charges.[9,15] These charges are both highly accurate and very inexpensive to calculate.

Intermediate- and short-range effects are calculated from solvent-accessible surface areas (SASAs),[16] which are proportional in a continuum sense to the number of solvent molecules involved in the solute−solvent microinterface. The intermediate-range cavity-structural effects are calculated with an effective solvent radius of 3.4 Å for all organic solvents and are included by multiplying the intermediate-range SASA of the solute by an intermediate-range molecular surface tension σ^{CS}. Short-range cavity-dispersion-hydrogen-bonding effects are calculated with an effective solvent radius of 1.7 Å for all solvents and are included by multiplying the SASA for each solute atom i by a short-range atomic surface tension σ_i. The surface tension σ^{CS} depends only on the solvent. In contrast, each σ_i depends on the solvent, on the atomic number of solute atom i, and on the atomic numbers of other solute atoms j and the distances r_{ij} and r_{jk} of such atoms from atom i and from other atoms k.

The new aspect required for this extension is a universal model for the *solvent dependence* of non-bulk-electrostatic effects. We represented the solvent dependence of the short-range surface tensions with components proportional to refractive index n and to Abraham's[17−20] acidity and basicity parameters. The intermediate-range CS surface tensions depend on n and on γ, the macroscopic surface tension. Details are given in the supporting information.

Table 1 provides mean unsigned errors in predicted standard-state free energies of solvation. It shows that the SM5.4/A-organic model does consistently well across *solute* functional group classes. Moreover, the model has a maximum mean unsigned error of 0.6 kcal or less for 13 of the 15 *solvent* classes mentioned above and 0.7 kcal for amine and nitro solvents.

The partition of the free energy of solvation into ENP, CD, CH, and CS contributions is ambiguous because the components are not state functions, and even if they were the partition obtained in the present method would depend on the assumed radii. Nevertheless, the partition provides insight. For hydrogen-bonding groups, most of the exergonic contributions come from the favorable electrostatics. Table 2 shows an example: While the sum of the microinterfacial effects are nearly identical for *n*-hexane and 1-pentanol in an octanol solution, the favorable electrostatic contribution changes by −3.1 kcal, reproducing the experimental trend for these molecules. When the solvent is changed from 1-octanol to *n*-hexadecane, microinterfacial effects associated with hydrogen bonding disappear and favorable electrostatic effects are roughly halved, thereby providing a physical explanation for the 2.2 kcal change in molar transfer free energy.

We have presented a universal model for the solvation energy, in almost any organic solvent, of molecules containing C, H, and the seven most common heteroatoms. The model includes accurate atomic partial charges and quantum mechanical effects of medium-induced solute distortion. It accounts for solute−solvent dispersion interactions and hydrogen-bonding effects through microinterfacial surface tensions. Because the model is quantum mechanical, it can be applied to transition states as well as to stable molecules, and it yields solute wave functions that can be used to calculate expectation values for molecular properties in solution. Because it is based on an economical Hamiltonian (AM1), it can be applied to very large systems, even proteins. Because of the way it is parameterized, it can be used to predict solvation free energies in solvents for which there are *no* measured free energies of solvation. We anticipate interesting applications to many fundamental and applied problems in organic and biological chemistry.

Acknowledgment. D.J.G. gratefully acknowledges a Kodak Fellowship. This work was supported by the National Science Foundation and the U.S. Army Research Office.

Supporting Information Available: Additonal details on the data set, parameterization, final parameters, and model performance, including references (7 pages).

JO9617427

---

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

(13) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902. Dewar, M. J. S.; Zoebisch, E. G. *THEOCHEM* **1988**, *180*, 1. Dewar, M. J. S.; Yate-Ching, Y. *Inorg. Chem.* **1990**, *29*, 3881.

(14) Yomosa, S. *J. Phys. Soc. Jpn.* **1973**, *35*, 1738. Rivail, J.-L.; Rinaldi, D. *Chem. Phys.* **1976**, *18*, 233. Tapia, O.; Goscinski, O. *Mol. Phys.* **1975**, *29*, 1653.

(15) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.

(16) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379. Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.

(17) Abraham, M. H. *Chem. Soc. Rev.* **1993**, 73.

(18) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Duce, P. P.; Morris, J. J.; Taylor, P. *J. Chem. Soc., Perkin Trans. 2* **1989**, 699.

(19) Abraham, M. H. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; p 83.

(20) Abraham, M. H. *J. Phys. Org. Chem.* **1993**, *6*, 660.