association of the probe with Cl^- is ~2.0 × 10^2 (*16*), whereas for $Ru(NH_3)_5^{2+}$ it is only 3 (*17*). The comparison illustrates the electron withdrawing power of η^2-H_2 acting as a π acid.

Dihydrogen-bound η^2-H_2 as a ligand may differ significantly from other ligands in that it may undergo substantial structural changes when the coligands are altered. The values of J_{HD} we have observed cover the range from >20 Hz to a value of <2, suggesting a change in H–D distance, and thus a change in the capacity of HD or H_2 to act as a π acid. If the charge on H_2 remains constant, this tendency is expected to increase as the H–H distance increases, but in remaining attached to the metal, and in the limit of being converted to a dihydride, the antibonding orbitals are filled, so that the value of the distance optimum for backbonding would lie between that of free H_2 and that found in dihydrides.

REFERENCES AND NOTES

1. O. Jardetsky and G. C. K. Roberts, Eds., *NMR in Molecular Biology* (Academic Press, New York, 1981).
2. K. Wüthrich, Ed., *NMR of Proteins and Nucleic Acids*, (Wiley, New York, 1986).
3. Z.-W. Li and H. Taube, *J. Am. Chem. Soc.* **113**, 8946 (1991).
4. See (*1*), chap. 12.
5. All spectra were recorded at 20°C.
6. The equilibrium quotient is defined as [Pr-L]/([Pr-H_2O] [L]), where [L] is the concentration of the free ligand and where [Pr-L]/[Pr-H_2O] is determined from the relative abundance of the two forms of the probe (Pr) as obtained from the ^1H NMR spectrum.
7. G. J. Kubas, *Acc. Chem. Res.* **21**, 120 (1988); R. H. Crabtree, *ibid.* **23**, 95 (1990).
8. L. Stryer, *Biochemistry* (Freeman, New York, ed. 2, 1981).
9. G. N. LaMar, W. D. Horrocks, Jr., R. H. Holm, *NMR of Paramagnetic Molecules* (Academic Press, New York, 1973).
10. A. S. Mildvan, J. Granot, G. M. Smith, M. N. Liebman, *Adv. Inorg. Biochem.* **2**, 211 (1980).
11. M. Pyle and J. K. Barton, *Prog. Inorg. Chem.* **38**, 413 (1990).
12. A. S. Mildvan and R. K. Gupta, *Methods Enzymol.* **49G**, 322 (1978).
13. H. Taube, in *Survey of Progress in Chemistry*, A. F. Scott, Ed. (Academic Press, New York, 1973), vol. 6, 1–46.
14. J. F. Wishart, H. Taube, K. J. Breslauer, S. S. Isied, *Inorg. Chem.* **23**, 2997 (1984).
15. S. S. Isied and H. Taube, *ibid.* **13**, 1545 (1974).
16. Z.-W. Li and H. Taube, unpublished results.
17. H. S. Lim, D. J. Barclay, F. C. Anson, *Inorg. Chem.* 11, 1460 (1972).

18. Support of this work by National Institutes of Health grant GM13638-24 is gratefully acknowledged.

30 October 1991; accepted 4 February 1992

An SCF Solvation Model for the Hydrophobic Effect and Absolute Free Energies of Aqueous Solvation

Christopher J. Cramer and Donald G. Truhlar

A model for absolute free energies of solvation of organic, small inorganic, and biological molecules in aqueous solution is described. This model has the following features: (i) the solute charge distribution is described by distributed monopoles, and solute screening of dielectric polarization is treated with no restrictions on solute shape; (ii) the energetic effects of cavity formation, dispersion interactions, and solute-induced restructuring of water are included by a semiempirical cavity surface tension; and (iii) both of these effects are included in the solute Hamiltonian operator for self-consistent field (SCF) calculations to allow solvent-induced electronic and geometric distortion of the solute. The model is parameterized for solutes composed of H, C, N, O, F, P, S, Cl, Br, and I against experimental data for 150 neutral solutes and 28 ions, with mean absolute errors of 0.7 and 2.6 kilocalories per mole, respectively.

Solubilities of molecules are directly related to the free energy of solvation, and reaction rates in solution are directly related to the differential free energy of solvation of transition states and reactants (*1*). Solvation free energies in aqueous solution are biochemically important for structural and metabolic equilibria and kinetics. For example, in enzyme catalysis a substrate is generally dehydrated, and thus the free energy of desolvation is an important contributor to the free energy of activation (*2*, *3*). Unfavorable solvation energies of associating nonpolar groups may lead to hydrophobic stabilization of their mutual complexes, and this may play an important role in enzyme-substrate or protein-inhibitor binding. Desolvation and hydrophobic stabilization are also important in DNA-binding interactions (*4*). The hydrophobic effect has long been implicated in protein folding, but it is hard to correlate the solvent affinities of residues with solution data for small molecules because nonterminal protein residues in enzymes are less polar than the zwitterionic monomers (*5*). In addition, protein stability may depend on a difficult-to-assess balance of hydrophobic and other noncovalent interactions (*6*), such as electrostatics and exchange repulsion. A theoretical approach can be valuable in allowing the quantitative estimation of the energetics of such solvation effects that are not accessible to direct experimental measurement and also in allowing a self-consistent treatment of hydrophobic effects with other noncovalent and even with covalent interactions.

Semiempirical molecular orbital theory (*7*) provides an increasingly powerful approach to modeling electronic energies of gas-phase molecules and solutes and their interactions with other molecules, including water (*8–10*). However, bulk solvation energies result from the cooperative effect of large numbers of solute molecules. For example, the effects of aqueous solvation on the activation energies of nucleophilic substitution reactions at methyl chlorides converged with respect to increasing the number of water molecules to within only 1 to 2 kcal mol^{-1} when this number was increased from 54 to 66 (*11*). Molecular dynamics and Monte Carlo simulation studies of aqueous solvation and the hydrophobic effect routinely involve 250 or more solvent molecules (*12*), and proper solvation of biomolecules may require many more than 1000 waters (*13*). One way to make the problem more manageable for small charged and polar solutes is to use models that treat the solvent as a continuum dielectric with an electric polarization field. Such effects are typically treated in terms of the solvent dielectric constant, and this approach has a long history for solvent polarization by monatomic ions (*14*), dipoles (*15*), and systems composed of distributed monopoles (*16–20*). Such effects can be incorporated into molecular orbital theory by including local field terms in the Hamiltonian, and the promise of local-field SCF models for biological chemistry has long been recognized (*21*). However, the biological applications of such models are still rather limited, perhaps because such methods are not well suited to treating the hydrophobic effect.

The hydrophobic effect is more subtle than the solvent polarization effects that dominate free energies of solvation for small charged and polar solutes, and it involves additional physical effects not accounted for in dielectric constants. It can, however, be treated by continuum models that assume a proportionality (*2*, *19*, *22*, *23*) between hydrophobicity (or all or part of the free energy of solvation) and the surface area of the solute, or—more physically—the sol-

C. J. Cramer, U.S. Army Chemical Research Development and Engineering Center, Research Directorate, Physics Division, Chemometric and Biometric Modeling Branch, Aberdeen Proving Ground, MD 21010.
D. G. Truhlar, Department of Chemistry, Supercomputer Institute, and Army High Performance Computing Research Center, University of Minnesota, Minneapolis, MN 55455.

vent-accessible surface area, which may be equated to the cavity area or the area of the first hydration shell of a continuum solvent surrounding the solute. This proportionality can be derived under reasonable physical approximations for the free energy of cavitation (*24*) and the interaction energy associated with dispersion forces (*25*), and it provides a natural way to account for loss of entropy of the solvent (*26*) in the first hydration shell due to changes in the water structure [the number of molecules in this shell is approximately proportional to the cavity area (*27*)]. These three effects, cavity formation, dispersion interactions, and structure changing, probably dominate the hydrophobic effect, and the association of the hydrophobic effect with cavity surface area is a dominant part of the shell model of hydration of Scheraga and co-workers (*28*).

Motivated by these kinds of considerations, we developed a semiempirical model of solvation, called AM1-SM1 (*29*), that includes dielectric polarization of the solvent with solute screening effects and a semiempirical term proportional to cavity surface area. Although the accuracy was good, there were three especially troublesome deficiencies from the point of view of application to biological systems, namely: (i) the hydrophobic effect, although included in principle, was not treated accurately; (ii) the solvation of water had a larger than average error; and (iii) P-containing compounds were not included in the parameterization. In this report we present a new general parameterization, called AM1-SM2, that corrects all three deficiencies.

The treatment of the solute Hamiltonian and the electric polarization of the solvent are the same as before, except that most of the parameters are reoptimized. In particular, we continue to treat the solute by the Austin Model 1 (*8*), and our treatment of polarization is based on the generalized Born model (*16–18*) with the dielectric screening algorithm of Still *et al.* (*19*), modified by localized semiempirical corrections (*29*) for bonded and geminal O–O pairs and vicinal N–H pairs. The polarization terms are included with the AM1 solute terms in the SCF Hamiltonian, and the critical Born atomic coulomb integrals were determined semiempirically (*29*). This combination of methods has three very strong features: (i) by including polarization effects in the SCF step, we include the effects of solvent-induced charge redistribution in the solute; (ii) the dielectric screening algorithm of Still *et al.* allows for arbitrary shaped solutes; (iii) by determining the Born coulomb radii semiempirically, we overcame some theoretical limitations of the reaction field theory (*17*) underlying the generalized Born model, such as the assumption that the dielectric properties of the solvent retain their macroscopic description right up to the solute.

The new functional dependence introduced into AM1-SM2 is the following expression for the contribution to the free energy of solvation associated with the cavity area:

$$G^\circ = \sum_{k'} \{\sigma_{k'}^{(0)} + \sigma_{k'}^{(1)}[f(B_{k'H}) + g(B_{k'H})]\}A_{k'}(\beta_{k'},\{\beta_k\}) \quad (1)$$

Here k' labels an atom, $\sigma_{k'}^{(0)}$ and $\sigma_{k'}^{(1)}$ are surface tensions, $f(B_{k'H})$ and $g(B_{k'H})$ are functions specified below, and $A_{k'}(\beta_{k'},\{\beta_k\})$ is the exposed surface area of atom k', which depends on $\beta_{k'}$, the "hydrophobic radius" of atom k', and the full set $\{\beta_k\}$ of all the other hydrophobic radii in the solute. The sum in Eq. 1 runs over all nonhydrogenic atoms. Hydrogen atoms are assumed for the purposes of Eq. 1 to have zero surface tension, zero surface area, and zero volume. The constants $\sigma_{k'}^{(0)}$, $\sigma_{k'}^{(1)}$, and $\beta_{k'}$ are semiempirical parameters for each nonhydrogenic atom type. The exposed surface area on atom k' is defined as the area on a sphere of radius $\beta_{k'}$ centered at its nucleus that is not contained in any sphere of radius β_k on any of the other nonhydrogenic atoms k. The motivation for including bond orders in the surface tension is that exposed H atoms may be hydrophilic or hydrophobic, depending on their environment (*29*). We therefore treat each XH_n unit as a single group with properties that depend on n.

The first function in Eq. 1, the critical one for parameterizing the hydrophobic effect, is given by

$$f(B_{k'H}) = \tan^{-1}(3^{1/2}B_{k'H}) \quad (2)$$

where $B_{k'H}$ is the sum of the bond orders (*30*) from atom k' to all H atoms. This function equals 0, 1.05, 1.29, 1.38, and 1.43 for $B_{k'H}$ = 0, 1, 2, 3, and 4, respectively. Thus, it primarily accounts for the difference between the hydrophobicity of groups containing H and those not containing H. Because H atoms have zero radius in this calculation, a $-CH_3$ group, for example, has more exposed surface area than a $-CH_2$ group, and the combination of this kind of effect with Eq. 2 was found sufficient to account for the major trends in hydrophobicity of hydrocarbons.

The second function in Eq. 1 is included only for k' corresponding to N or O. It corrects small systematic remaining errors in N atoms substituted with three or more H atoms and in O atoms substituted with two or more H atoms (such as NH_3, $H_3CNH_3^+$, H_2O, and H_3O^+). It is given by

$$g(B_{k'H}) = a_{k'} \exp\left\{\frac{-b_{k'}}{1 - [(B_{k'H} - c_{k'})/w_{k'}]^2}\right\} \quad (3)$$

when $|B_{k'H} - c_{k'}| < w_{k'}$, and g is zero elsewhere.

The calculations may be used to predict the absolute free energy in aqueous solution or, by subtraction of the gas-phase value, the free energy of solvation (*29*).

The experimental data (*31–33*) used to optimize the parameters are the same as

Table 1. Parameters.

k	$\rho_k^{(0)}$ (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	1.283	−0.30	0	0	0
C	1.68	0	*	3.36	3.24	3.1
N†	1.50	0.420	0.60	−30.70	−19.20	3.3
O‡	1.46	−0.150	0.75	−25.01	−37.28	3.2
F	1.37	0.145	0.70	22.50	0	2.8
P	1.30	1.000	0.00	3.90	0	3.5
S	1.30	0.800	0.70	−53.25	37.03	3.2
Cl	1.65	0.555	0.75	−2.28	0	3.4
Br	1.75	0.629	0.70	−8.15	0	3.4
I	1.88	0.885	0.60	−15.18	0	3.4

*Not required. †a_k = −9.12, b_k = 1.35, c_k = 3.69, w_k = 1.5. ‡a_k = −1.94, b_k = 0.46, c_k = 2.56, w_k = 1.0.

Table 2. Mean unsigned errors in the free energies of solvation for various classes of solutes in the test set.

Class	Data in class	Error (kcal mol^{-1})
Neutrals		
Hydrocarbons	33	0.5
H, C, N compounds	18	1.0
H, C, O compounds	38	0.8
H, C, F compounds	4	0.5
H, C, S compounds	6	0.7
H, C, Cl compounds	14	0.3
H, C, Br compounds	8	0.4
H, C, I compounds	5	0.1
NH_3, H_2O, PH_3, H_2S	4	0.05
Compounds with four or more kinds of atoms	20	1.1
All	150	0.7
Ions		
Cations	10	2.4
Anions	18	2.7
All	28	2.6

specified completely in (29) augmented by three phosphite triesters (33), heptane, 2,4-dimethylpentane, methylcyclohexane, *cis*-1,2-dimethylcyclohexane, ammonia, phosphine, PH_2^-, $(CH_3)_3PH^+$, and $CH_3C(OH)NH_2^+$ (H_3S^+ was deleted because of uncertainties in the data), for a total of 150 neutrals and 28 ions. The final semiempirical parameters are given in Table 1; the meaning of the first three columns is explained elsewhere (29). Note the magnitudes of the surface tensions. It has been estimated (2) that creating an empty cavity in water costs 104 cal mol^{-1} $Å^{-2}$, so a large part of the surface tension is due to dispersion and structural effects, including hydrogen bonding effects not in the polarization term. In addition, an indeterminate portion of the semiempirical surface tensions may also be making up for part of the systematic error in the generalized Born model of the electric polarization effects.

Table 3. Calculated dipole moments (debyes) and calculated and experimental free energies of solvation (kilocalories per mole) and the two components of the calculated free energies for a sample of the solutes in the test set.

Solute	Dipole moment*		Free energy			
			Calculated			Experimental total
	Gas	Solution	ENP	CDS	Total	
Butane	0.0	0.0	0.1	1.6	1.7	2.1
Hexane	0.0	0.0	0.2	2.0	2.2	2.5
Heptane	0.0	0.0	0.2	2.2	2.5	2.6
Octane	0.0	0.0	0.3	2.4	2.7	2.9
2,4-Dimethylpentane	0.0	0.0	0.3	2.1	2.4	2.9
Cyclopropane	0.0	0.0	−0.1	1.3	1.1	0.8
Cyclopentane	0.0	0.0	0.1	1.5	1.6	1.2
Methylcyclohexane	0.0	0.0	0.3	1.9	2.1	1.7
Ethene	0.0	0.0	−0.3	1.1	0.8	1.3
2-Methylpropene	0.4	0.4	−0.4	1.5	1.1	1.2
1,3-Butadiene	0.0	0.1	−0.9	1.4	0.6	0.6
Benzene	0.0	0.0	−2.0	1.4	−0.5	−0.9
Toluene	0.3	0.3	−1.9	1.7	−0.3	−0.9
1-Hexyne	0.4	0.5	−1.8	1.8	0.0	0.3
1-Butanamine	1.6	1.6	−1.2	−3.5	−4.7	−4.3
Aniline	1.5	1.7	−2.6	−3.2	−5.8	−4.9
Dimethylamine	1.2	1.3	−2.6	−1.8	−4.3	−4.3
Piperazine	0.0	0.0	−3.7	−4.1	−7.8	−7.4
Pyridine	2.0	2.5	−3.9	−0.5	−4.4	−4.7
Propanenitrile	2.9	3.6	−1.8	−2.0	−3.8	−3.9
Nitroethane	4.4	5.2	−1.7	−2.6	−4.4	−3.7
1-Nitropropane	4.5	5.3	−1.4	−2.4	−3.8	−3.3
Ethanol	1.6	1.7	−1.0	−3.9	−4.9	−5.0
1-Propanol	1.5	1.7	−0.9	−3.7	−4.6	−4.8
Prop-2-en-1-ol	1.6	1.8	−1.2	−3.7	−4.9	−5.0
Phenol	1.2	1.5	−2.5	−3.3	−5.8	−6.6
Propanoic acid	1.8	2.5	−1.6	−5.1	−6.7	−6.5
Butanoic acid	1.9	2.5	−1.5	−4.9	−6.3	−6.4
Ethyl acetate	1.9	2.6	−2.2	−1.2	−3.3	−3.1
Methyl butanoate	1.7	2.3	−1.7	−1.0	−2.7	−2.8
Propanal	2.6	3.3	−2.6	−1.0	−3.6	−3.5
Butanal	2.6	3.4	−2.5	−0.8	−3.3	−3.2
Acetophenone	3.0	4.2	−4.3	−0.1	−4.4	−4.6
Methanethiol	1.8	1.9	−0.2	−0.7	−0.8	−1.2
Dimethyl sulfide	1.6	1.6	−0.1	−1.5	−1.6	−1.4
2-Methoxyethanol	2.1	2.3	−1.4	−4.9	−6.3	−6.8
m-Hydroxybenzaldehyde	3.1	4.1	−4.2	−5.4	−9.5	−9.5
2-Methoxyethanamine	2.6	2.7	−2.2	−4.9	−7.1	−6.6
Tetrafluoromethane	0.0	0.0	−0.6	4.0	3.4	3.1
1,1-Difluoroethane	2.3	2.7	−2.4	2.6	0.3	−0.1
2,2,2-Trifluoroethanol	3.6	4.3	−3.2	−1.4	−4.5	−4.3
Chlorofluoromethane	1.8	2.2	−1.8	1.1	−0.7	−0.8
Z-1,2-dichloroethene	1.5	1.9	−1.1	0.0	−1.1	−1.2
E-1,2-dichloroethene	0.0	0.0	−0.5	−0.2	−0.7	−0.8
Chloroform	1.2	1.3	−0.7	−0.5	−1.2	−1.1
Bromomethane	1.5	1.7	−0.4	−0.3	−0.7	−0.8
p-Bromophenol	1.6	1.7	−2.7	−4.4	−7.0	−7.1
Iodoethane	1.5	1.5	0.0	−0.7	−0.7	−0.7
2-Iodopropane	1.6	1.7	0.0	−0.4	−0.4	−0.5
H^-	0.0	0.0	−89.0	0.0	−89.0	−89
O_2^-	0.0	0.0	−84.3	−3.8	−88.1	−87
HS^-	1.6	2.0	−74.4	−1.9	−76.3	−76
PH_2^-	2.6	3.1	−67.6	0.6	−67.0	−67
N_3^-	0.0	0.0	−69.5	−5.7	−75.2	−74
$H_3CCO_2^-$	3.8	5.3	−72.9	−3.0	−75.9	−77
H_3O^+	2.3	2.5	−100.0	−3.8	−103.8	−104
NH_4^+	0.0	0.0	−77.3	−1.7	−79.0	−79
$H_3COH_2^+$	2.2	2.8	−79.0	−5.2	−84.2	−83
$H_3CSH_2^+$	1.3	1.5	−73.7	0.1	−73.6	−74
$(CH_3)_3PH^+$	0.3	0.3	−49.4	1.5	−47.9	−53
$CH_3C(OH)NH_2^+$	3.2	4.1	−60.4	−7.8	−68.2	−66

*For ions, origin at center of mass calculated for most abundant isotopes.

Table 2 gives mean unsigned errors in the absolute free energies of solvation for various classes of molecules and ions. The "experimental" data for ions are obtained from thermodynamic cycles, and the uncertainties are typically several kilocalories per mole.

Table 3 gives a sample of results for individual solutes. The contribution to the free energy of solvation from Eq. 1 is labeled CDS ("cavity-dispersion-structural"), and the remainder is labeled ENP ("electronic-nuclear-polarization"). Several trends are apparent. For alkanes, cycloalkanes, and alkenes, the ENP term is very small, and the variations are mainly determined by the CDS term. For alkynes and aromatics, the ENP term becomes significant. For systems that contain heteroatoms, it is necessary to treat both terms accurately. Methods that model the ENP term entirely in terms of net charge or dipole would fail for systems with no net dipole, such as benzene, piperazine, and tetrafluoromethane.

The hydrophobic effect consists not only in getting the positive free energies of solvation correct for the nonpolar solutes but also in modeling nonpolar substituent effects, as in 1-propanol versus ethanol or in butanal versus propanal. Notice also the cancellation of effects in passing from 1-butanamine to dimethylamine. A model based only on dipoles or only on surface tensions would be unlikely to explain the similar free energies of hydration for these two solutes.

Another ingredient in being able to treat a wide variety of systems with reasonably uniform accuracy is the inclusion of the solvation terms in the solute SCF calculation. Consider, for example, 4-pyridone. AM1-SM2 predicts that its free energy of solvation is −17.2 kcal mol^{-1}. However, solvating the gas-phase structure yields only −10.6 kcal mol^{-1}. The full calculation predicts that the solute dipole changes from 6.3 to 10.8 D upon dissolution with the partial charge on O

Table 4. Free energies of solvation and differences of free energies of solvation (kilocalories per mole); n.a., not available.

Solute	ΔG_S°		$\Delta G_S^\circ - \Delta G_S^\circ$ (benzene)		
	SM2	Experiment	538 to 612 waters	SM2	Experiment
Benzene	−0.5	−1.0			
Anisole	−2.3	−2.4	−0.9	−1.8	−1.4
1,2-Dimethoxybenzene	−3.2	−3.8	n.a.	−2.7	−2.8
1,2,3-Trimethoxybenzene	−4.0	−5.4	−4.3	−3.5	−4.4

changing from −0.34 to −0.56. This electronic structure is intrinsically less favorable by 11.6 kcal mol^{-1}, but its solvation free energy is −28.8 kcal mol^{-1}, accounting for the net solvation free energy of −17.2 kcal mol^{-1}.

Finally, we compare continuum water models to models that explicitly treat hundreds of waters. One advantage of explicit water models is that they give information about the location and H-bonding patterns of specific inner-shell waters (*34*). In our model we can add one or a few specific waters to the solute to examine such effects because the present model gives the correct solvation free energy of a water molecule, but we cannot easily examine full hydration spheres. However, our model has three significant advantages over most current explicit water simulations. First, it includes electronic distortion of the solute by the solvent. Second, it yields absolute free energies of solvation, whereas most explicit water simulation studies are limited by computational considerations to free energy differences between two solutes. Third, our model requires one to three orders of magnitude less computer time, depending on the system. Yet the accuracy is comparable at worst and better in most cases. Consider, for example, the solvation of benzene, anisole, dimethoxybenzene, and trimethoxybenzene, which have been studied recently in large-scale computer simulations (*35*) involving 538 to 612 explicit waters. This is a difficult test case [Kuyper *et al.* (*35*) found that their results varied by as much as 7.3 kcal mol^{-1} when they varied the basis set model used to obtain solute partial charges]. In addition, ethers often have larger than average errors in our parameterization so these compounds are not especially favorable cases for our model. Our results agree reasonably well with experiment (Table 4) (*31*, *35*), especially when we consider that the uncertainty in the experimental values is about 0.3 to 0.5 kcal mol^{-1}. Another case is the isomerization of 2-hydroxypyridine to 2-pyridone. AM1-SM2 yields a differential solvation effect of −2.6 kcal mol^{-1}, a calculation (*36*) with 600 water molecules gave −3.0 kcal mol^{-1}, and experiment yields (*36*, *37*) −4.1 kcal mol^{-1}. Electronic relaxation in the presence of the solvent increases the magnitude of the net solvation free energies by 10 to 25% in these cases. Thus, models that neglect this effect, as most simulations do, cannot attain this accuracy except perhaps by modeling the effect nonphysically.

Although the initial parameterization and testing steps have involved equilibria rather than rate processes, we would emphasize that one critical element of the incorporation of solvation effects in a molecular orbital framework, as we have done, is the ability of such a calculation to also treat transition states, for which a molecular mechanics framework is less suitable. The accurate treatment of hydrophobic effects and explicit water in a local-field SCF model by a general parameterization (for H, C, and eight heteroatoms) including both a distributed monopole charge distribution with dielectric screening by arbitrary solute shapes and also semiempirical cavity surface tensions should, we believe, allow for more accurate molecular modeling of a wide variety of organic and biochemical equilibria and rates.

REFERENCES AND NOTES

1. C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry* (VCH, Weinheim, Germany, ed. 2, 1988).
2. A. Fersht, *Enzyme Structure and Mechanism* (Freeman, New York, ed. 2, 1985), pp. 65, 300–301.
3. R. Wolfenden and W. M. Kati, *Acc. Chem. Res.* **24**, 209 (1991).
4. E. C. Long and J. K. Barton, *ibid.* **23**, 271 (1990).
5. R. Wolfenden, *Science* **222**, 1087 (1983).
6. J. T. Kellis, Jr., K. Nyberg, D. Sali, A. R. Fersht, *Nature* **333**, 788 (1988).
7. O. Sinanoglu and K. B. Wiberg, *Sigma Molecular Orbital Theory* (Yale Univ. Press, New Haven, CT, 1970), chaps. 1 to 6; G. A. Segal, *Semiempirical Methods of Electronic Structure Calculation* (Plenum, New York, 1977), parts A and B.
8. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, 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, *ibid.* **187**, 1 (1989); M. J. S. Dewar and Y. C. Yuan, *Inorg. Chem.* **29**, 3881 (1990).
9. J. J. P. Stewart, *J. Comput. Chem.* **10**, 209 (1989); *ibid.*, p. 221; *J. Comput.-Aided Mol. Des.* **4**, 1 (1990).
10. J. J. Dannenberg, *J. Phys. Chem.* **92**, 6869 (1988); I. Juranic, H. S. Rzepa, M. Yi, *J. Chem. Soc. Perkin Trans. 2* **1990**, 877 (1990); H. S. Rzepa and M. Yi, *ibid.*, p. 943; *ibid.* **1991**, 531 (1991).
11. Y. S. Kong and M. S. Jhon, *Theor. Chim. Acta* **70**, 123 (1986).
12. W. L. Jorgensen, J. F. Blake, J. D. Madura, S. D. Wierschke, *ACS Symp. Ser.* **353**, 200 (1987).
13. W. F. Van Gunsteren and H. J. C. Berendsen, in *Molecular Dynamics and Protein Structure*, J. Hermans, Ed. (Polycrystal, Western Springs, IL, 1985), pp. 5–14; W. E. Harte, Jr., *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8864 (1990).
14. M. Born, *Z. Phys.* 1, 45 (1920); A. A. Rashin and B. Honig, *J. Phys. Chem.* **89**, 5588 (1985).
15. J. G. Kirkwood, *J. Chem. Phys.* **2**, 351 (1934); *ibid.* **7**, 911 (1939); L. Onsager, *J. Am. Chem. Soc.* **58**, 1486 (1936).
16. G. J. Hoijtink, E. deBoer, P. H. van der Meij, W. P. Weijland, *Rec. Trav. Chim. Pays-Bas* **75**, 487 (1956); F. Peradejordi, *Cah. Phys.* **17**, 393 (1963); I. Jano, *C. R. Acad. Sci.* **261**, 103 (1965); G. Klopman, *Chem. Phys. Lett.* 1, 200 (1967).
17. O. Tapia, in *Quantum Theory of Chemical Reactions*, R. Daudel, A. Pullman, L. Salem, A. Veillard, Eds. (Reidel, Dordrecht, 1980), vol. 2, pp. 25–72; R. Costanciel, *ibid.*, pp. 73–98; ——— and R. Contreras, *Theor. Chim. Acta* **65**, 1 (1984).
18. R. Contreras and J. S. Gomez-Jeria, *J. Phys. Chem.* **88**, 1905 (1984); T. Kozaki, M. Morihasi, O. Kikuchi, *J. Am. Chem. Soc.* **111**, 1547 (1989); S. C. Tucker and D. G. Truhlar, *Chem. Phys. Lett.* **157**, 164 (1989).
19. W. C. Still, A. Tempczak, R. C. Hawley, T. Hendrickson, *J. Am. Chem. Soc.* **112**, 6127 (1990).
20. A. Jean-Charles *et al.*, *ibid.* **113**, 1454 (1991).
21. O. Tapia, F. Sussman, E. Poulain, *J. Theor. Biol.* **71**, 49 (1978); G. Náray-Szabó and P. R. Surján, in *Theoretical Chemistry of Biological Systems*, G. Náray-Szabó, Ed. (Elsevier, Amsterdam, 1986), p. 1; M. Berndt and J. Kwiatkowski, *ibid.*, p. 349.
22. R. B. Hermann, *J. Phys. Chem.* **76**, 2754 (1972); *ibid.* **79**, 163 (1975).
23. M. J. Harris, T. Higuchi, J. H. Rytting, *ibid.* **77**, 2694 (1973); J. A. Reynolds, D. B. Gilbert, C. Tanford, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2925 (1974); C. Chothia, *Nature* **248**, 338 (1974); G. L. Amidon, S. H. Yalkowski, S. T. Anik, S. C. Valvani, *J. Phys. Chem.* **79**, 2239 (1975); A. Warshel, *ibid.* **83**, 1640 (1979); G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee, M. H. Zehfus, *Science* **229**, 834 (1985); D. Eisenberg and A. D. McLachlan, *Nature* **319**, 199 (1986); T. Ooi, M. Oobatake, G. Nemethy, H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3086 (1987); A. J. Doig and D. H. Williams, *J. Mol. Biol.* **217**, 389 (1991).
24. R. Pierrotti, *Chem. Rev.* **76**, 717 (1976).
25. M. J. Huron and P. Claverie, *J. Phys. Chem.* **76**, 2123 (1972); F. Floris and J. Tomasi, *J. Comput. Chem.* **10**, 616 (1989).
26. H. S. Frank and M. W. Evans, *J. Chem. Phys.* **13**, 507 (1945); W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959).
27. W. L. Jorgensen, J. Gao, C. Ravimohan, *J. Phys. Chem.* **89**, 3470 (1985).
28. K. D. Gibson and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 420 (1967); A. J. Hopfinger, *Macromolecules* **4**, 731 (1971); Z. I. Hodes, G. Nemethy, H. A. Scheraga, *Biopolymers* **18**, 1565 (1979); H. A. Scheraga, *Acc. Chem. Res.* **12**, 7 (1979).
29. C. J. Cramer and D. G. Truhlar, *J. Am. Chem. Soc.* **113**, 8305 (1991); *ibid.*, p. 8552; *ibid.*, p. 9901(E).
30. D. R. Armstrong, P. G. Perkins, J. J. P. Stewart, *J. Chem. Soc. Dalton Trans.* **1973**, 838 (1973).
31. S. Cabani, P. Gianni, V. Mollica, L. Lepori, *J. Solution Chem.* **10**, 563 (1981); A. Ben-Naim and Y. Marcus, *J. Chem. Phys.* **81**, 2016 (1984); R. G. Pearson, *J. Am. Chem. Soc.* **108**, 6109 (1986).
32. D. D. Wagman *et al.*, *J. Phys. Chem. Ref. Data* **11** (Suppl. 2), 2-73 (1982).
33. R. Wolfenden and R. Williams, *J. Am. Chem. Soc.* **105**, 1028 (1983).
34. For example, molecular mechanics calculations [R. C. Wade, *J. Comput.-Aided Mol. Des.* **4**, 199 (1990)] predict especially favorable water binding sites in the substrate pocket of cytochrome P450-cam. Such a water molecule may exhibit cooperative H binding or nonbulk dielectric properties that are not well simulated by a continuum model,

and it should probably be included explicitly.
35. L. F. Kuyper, R. N. Hunter, D. Ashton, K. M. Merz, Jr., P. A. Kollman, *J. Phys. Chem.* **95**, 6661 (1991).
36. P. Cieplak, P. Bash, U. C. Singh, P. A. Kollman, *J. Am. Chem. Soc.* **109**, 6283 (1987).
37. P. Beak, F. S. Fry, Jr., J. Lee, F. Steele, *ibid.* **98**, 171 (1976).
38. We are grateful to C. Lim for helpful discussions. This work was supported in part by the National Science Foundation.

11 December 1991; accepted 19 February 1992

Context Dependence of Hydrogen Bond Free Energy Revealed by Substitutions in an RNA Hairpin

John SantaLucia, Jr., Ryszard Kierzek, Douglas H. Turner*

Prediction and modeling of RNA structure requires knowledge of the free energy contributions of various interactions. Many unusual hydrogen bonds were recently proposed in the structure of a GCAA hairpin determined from nuclear magnetic resonance. The contributions of these hydrogen bonds to the folding stability of the hairpin formed by rGGCGCAAGCC have now been investigated through the use of functional group substitutions. These and previous results suggest a strong context dependence for the free energy of hydrogen bond formation. The results also suggest that the phylogenetic preference for GNRA (where N = A, C, G, or U and R = A or G) tetraloops may have a functional rather than thermodynamic basis.

The functional diversity of RNA reflects diversity in three-dimensional (3-D) structure. The 3-D folding of RNA is apparently controlled by the RNA sequence, because proteins are usually not required for proper folding (*1–4*). Prediction of RNA structure from sequence has focused largely on secondary structure (*5*), and predictions have improved as relations between structure and free energy have been determined (*5–7*). In principle, prediction of 3-D folding is amenable to a similar approach and requires a knowledge of the standard free energy (ΔG°) contributions of individual interactions, such as hydrogen bonding, stacking, and hydration in diverse structural contexts. We report ΔG° increments for hydrogen bonds in an RNA hairpin loop, GCAA, for which the structure has been determined with nuclear magnetic resonance (NMR) (*8*). The GCAA sequence is one of a group of sequences, designated "tetraloops," that occur commonly in ribosomal and other RNAs (*9, 10*).

Substitution of hydrogen-bonding groups provides a method for determining the free energy contributions of hydrogen bonds to macromolecular properties (*11–15*). This "atomic mutation" approach is used here. For example, hydrogen bonds involving the amino group of guanosine are studied by substituting guanosine with inosine (I), thus replacing the G amino group with a hydrogen. The ΔG° increment for the hydrogen bonds is then defined as the difference in the ΔG° of folding for the hairpin with and without the amino group. The substitutions made are shown in Fig. 1.

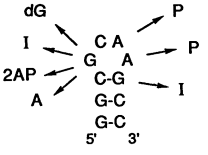

Fig. 1. Summary of individual substitutions made in the GGCGCAAGCC hairpin. P, purine; 2AP, 2-aminopurine; and dG, deoxyguanosine.

The reference sequence is rGGCGCAAGCC, nucleotides 377 to 386 in *Escherichia coli* 16S ribosomal RNA (rRNA). This is a shortened form of the pppGGGCGCAAGCCUUAU sequence used for structure determination with NMR (*8*). Thermodynamic parameters for folding of the hairpins were derived from fits of melting curves to a two-state model (Fig. 2 and Table 1). The ΔG° values at 70°C are more reliable because each requires a small extrapolation from the melting temperature, T_M. Except for rGGCGCAAICC, all of the sequences have concentration-independent T_M's in 0.1 M NaCl, indicating hairpin formation. For rGGCGCAAICC in 0.1 M NaCl, concentration-dependent T_M's were observed, consistent with duplex formation (see footnote to Table 1). In order to quantify the effect of the G to I substitution in the closing base pair, rGGCGCAAGCC and rGGCGCAAICC were melted in 10 mM phosphate buffer without added NaCl. Under these conditions, the T_M's are independent of concentration. The G to I substitution in the closing base pair results in a 16°C drop in T_M, which corresponds to a less favorable (by 1.3 kcal/mol) ΔG° at 70°C (see Table 1). This result is similar to ΔG° decrements observed upon removal of hydrogen-bonding amino groups in GC pairs and GA mismatches (*11, 12*).

Whereas a hydrogen bond in a stem base pair contributes more than 1 kcal/mol to hairpin stability, the results in Table 1 indicate that hydrogen bonds within the hairpin all contribute less than 1 kcal/mol (Fig. 3). The largest effect, 0.7 kcal/mol, is observed for the G to I substitution in the GCAA loop. The NMR structure of the loop (*8*) suggests that the amino group thus

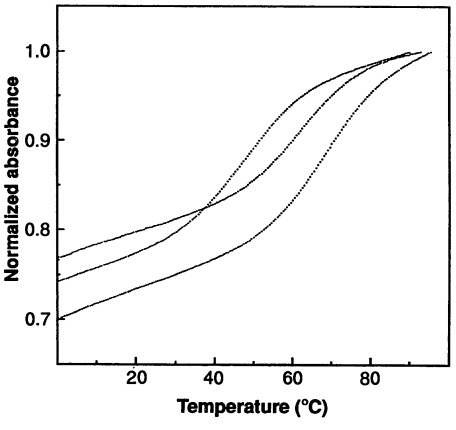

Fig. 2. Typical melting curves. Sequences from highest to lowest melting temperature are GGCGCAAGCC (with 0.1 M NaCl), GGCICAAGCC (with 0.1 M NaCl), and GGCGCAAICC (without NaCl). Oligoribonucleotides were synthesized on solid support with the phosphoramidite method (*12, 23*). After deblocking with ammonia and with acid, the RNA was purified on a Si500F thin-layer chromatography plate (Baker) and eluted for 5 hours with *n*-propanol-ammonia-water (55:35:10 by volume) (*24*). Bands were visualized with an ultraviolet lamp, and the least mobile band was cut out and eluted three times with 1 ml of doubly distilled water. Oligomers were further purified with a Sep-pack C-18 cartridge (Waters). Oligomers for low-salt melts were desalted by continuous-flow dialysis (BRL). Purities were checked with analytical C-8 high-pressure liquid chromatography (Beckman) and were >95%. The buffer for most melting curves was 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM Na_2EDTA at pH 7. For low-salt experiments, 0.1 M NaCl was omitted. Absorbance versus temperature curves were measured at 280 nm with a heating rate of 1°C per minute on a Gilford 250 spectrophotometer. Experiments with a heating rate of 0.5°C per minute gave the same results as those at 1°C per minute, suggesting thermodynamic equilibrium was achieved. Oligomer concentration was varied roughly 100-fold. Absorbance versus temperature curves were fit to a two-state model with sloping baselines through the use of a nonlinear least-squares program.

J. SantaLucia, Jr., and D. H. Turner, Department of Chemistry, University of Rochester, Rochester, NY 14627–0216.
R. Kierzek, Institute of Bioorganic Chemistry, Polish Academy of Sciences, 60-704 Poznan, Noskowskiego 12/14, Poland.
*To whom correspondence should be addressed.