Microscopic Models for Quantum Mechanical Calculations of Chemical Processes in Solutions: LD/AMPAC and SCAAS/AMPAC Calculations of Solvation Energies

V. Luzhkov† and A. Warshel*

Department of Chemistry, University of Southern California, Los Angeles, California, 90089-1062

Received 28 March 1991; accepted 20 August 1991

Our previously developed approaches for integrating quantum mechanical molecular orbital methods with microscopic solvent models are refined and examined. These approaches consider the nonlinear solute–solvent coupling in a self-consistent way by incorporating the potential from the solvent dipoles in the solute Hamiltonian, while considering the polarization of the solvent by the potential from the solute charges. The solvent models used include the simplified Langevin Dipoles (LD) model and the much more expensive surface constrained All Atom Solvent (SCAAS) model, which is combined with a free energy pertubation (FEP) approach. Both methods are effectively integrated with the quantum mechanical AMPAC package and can be easily combined with other quantum mechanical programs. The advantages of the present approaches and their earlier versions over macroscopic reaction field models and supermolecular approaches are considered. A LD/MNDO study of solvated organic ions demonstrates that this model can yield reliable solvation energies, provided the quantum mechanical charges are scaled to have similar magnitudes to those obtained by high level *ab initio* methods. The incorporation of a field-dependent hydrophobic term in the LD free energy makes the present approach capable of evaluating the free energy of transfer of polar molecules from non polar solvents to aqueous solutions. The reliability of the LD approach is examined not only by evaluating a rather standard set of solvation energies of organic ions and polar molecules, but also by considering the stringent test case of sterically hindered hydrophobic ions. In this case, we compare the LD/MNDO solvation energies to the more rigorous FEP/SCAAS/MNDO solvation energies. Both methods are found to give similar results even in this challenging test case. The FEP/SCAAS/AMPAC method is incorporated into the current version of the program ENZYMIX. This option allows one to study chemical reactions in enzymes and in solutions using the MNDO and AM1 approximations. A special procedure that uses the EVB method as a reference potential for SCF MO calculations should help in improving the reliability of such studies.

INTRODUCTION

Quantum mechanical studies of chemical processes have progressed in an impressive way in recent years due in part to the rapid increase in available computer power. *Ab initio* approaches (e.g., reference 1) allow one to obtain reliable potential surfaces for small molecules in the gas phase, while semiempirical approaches (e.g., reference 2) provide reasonable results for medium size molecules in the gas phase. However, most chemical and biological processes occur in solutions, and gas phase calculations are not expected to reliably describe solution chemistry.[3–6] The effect of the solvent must be taken into account in studies of reaction rates, equilibrium structures, electronic spectra, and in fact any other property, if one intends to consider the solute in its actual environment. Thus it is important to develop practical quantum mechanical methods that can be used in calculating the properties of solvated molecules.

Accounting properly for solvent effects in theoretical models is far from being obvious. Trying to represent the solute and a limited number of solvent molecules by "supermolecule" *ab initio* approach[7] is not expected to provide the relevant energetics, which is associated with many solvent molecules and involves long-range electrostatic forces. One might consider treating the solute and a large number of solvent molecules by a semiempirical quantum mechanical approach. However, the enormous number of degrees of freedom in such a multidimensional system makes it impractical to use energy minimization or other simulation approaches. Thus, it is imperative to somehow limit the explicit quantum mechanical treatment to only a small subspace of the solute–solvent system, while representing the effect of the rest of the system by some type of an effective potential. In particular, it is reasonable to treat the solute quantum mechanically while including the classical electrostatic potential from the sol-

†Permanent address: Institute of Chemical Physics, Chernogolovka, Moscow Region, U.S.S.R.

*Author to whom all correspondence should be addressed.

Journal of Computational Chemistry, Vol. 13, No. 2, 199–213 (1992)

CCC 0192-8651/92/020199-15$04.00

vent molecules in the solute Hamiltonian.[8] The solvent system can then be treated by either macroscopic continuum models or discrete microscopic models. These alternative treatments are briefly considered below.

The incorporation of continuum solvent models in quantum chemical calculations has been explored by several workers.[9] The early work of Klopman[9a] considered the problem in a heuristic way and provided a useful and formally correct expression for the solvation contribution to the energetics of a given solute charge distribution. Subsequent continuum studies[9b-9h] used the Onsager's reaction field (RF) formulations[10a] or related expressions,[10b,10c] considering the solute in a spherical cavity inside a continuum dielectric media. The corresponding potential from the solvent was expressed in terms of the solute charges and incorporated into the solute Hamiltonian. While this approach seems quite rigorous (especially when applied selfconsistently[9d]), it suffers from a major drawback. That is, the results depend on the *third power* of the cavity radius, which is basically an adjustable parameter whose actual value is not given in a reliable way by the available macroscopic prescriptions. Thus the calculations may involve substantial errors and can only be considered as rough estimates (see the Appendix for discussion of this important fact). It is possible that the elegant form of the macroscopic treatments led some to overlook the bottom line, which is the ability to get reasonable solvation energies; a model that cannot give reliable solvation energies should not be used for quantitative studies of processes in solutions regardless of its formal justification.

The above problem may be reduced by replacing the oversimplified cavity model by a discretized continuum treatment of the type used in recent solvation studies (e.g., reference 11), provided that the model is examined and calibrated using experimental, determined solvation energies. Another problem that is hard to treat properly with continuum models is the effect of the solvent electronic polarization. That is, in some cases, and particularly in studies of excited states properties, it is important to treat consistently the solvent induced dipoles in the presence of both the solute charge distribution and the solvent permanent dipoles. Although this issue has been formally addressed,[9h] it is not entirely clear how to treat such a problem in a reliable and practical way while using a continuum model.

Many of the uncertainties associated with the macroscopic models may be eliminated (at least in principle) by considering a large number of solvent molecules explicitly, while using classical analytical potential functions for the solvent–solvent and solute–solvent interactions. In this case the cavity concept can be eliminated altogether and the permanent and induced solvent polarization can be treated explicitly.[8,12] Similarly the consistent incorporation of the solvent potential in the solute Hamiltonian can be accomplished with a clear molecular picture (see the next section). However, using many solvent molecules without properly exploring the phase space of the system, or using uncalibrated solvent parameters, may lead to somewhat meaningless results (similar to those obtained with uncalibrated cavity models).

The first attempts to incorporate classical microscopic solvent models in quantum chemical calculations and to obtain *quantitative* results for chemical processes in solutions and proteins were reported in references 8 and 12. Representing the solvent by simplified yet well calibrated dipolar models, it was possible to obtain reasonable energetics for the fundamental dissociation reaction $H_2O + H_2O \rightarrow H_3O^+ + OH^-$ in aqueous solution, while using the MINDO quantum mechanical scheme for the solute region. Subsequent MO studies that incorporated classical solvent models in the solute Hamiltonian were also reported (e.g., reference 13), including a recent systematic work,[13c] which appears to adopt substantial elements of the formulations of references 8 and 12.

Recent studies seem to indicate that the Valence Bond (VB) formulation provides significant advantages over MO approaches in treating bond breaking chemical reactions in solutions (see discussion in reference 14). Nevertheless, MO methods are widely used and there is now an increasing realization that such calculations should include the effect of the solvent when applied to solvated molecules. Thus we find it useful to refine our early MO approaches and to demonstrate their advantages and general applicability.

In this article we consider two strategies for incorporating solvent effects in MO calculations. The first one involves the Langevin Dipoles (LD) solvent model.[3,8,12b,15] This simplified microscopic model seems superior to the macroscopic reaction field approach since it provides reliable solvation energies and allows one to treat the permanent and induced dipoles of the solvent in a simple and consistent way. The second method considered here involves the Surface Constrained All Atom Solvent (SCAAS) model,[3,16] which can be used in energy minimization and/or in Free Energy Peturbation Molecular Dynamics (FEP/MD) simulations. This model, which is similar in many respects to the one used in our early calculations of reactions in proteins[8] and solutions,[12a] provide more rigorous results than those obtained by the LD model but at the expense of much longer simulation times.

The next section describes our theoretical approach and outlines the incorporation of the different solvent models within SCF/MO quantum chemical approaches taking the AMPAC program[17] and its MNDO option[17] as a specific example. The

third section describes the results of LD/MNDO calculations of the effect of polar solvents on different ions and polar molecules. This section also examines the more rigorous SCAAS/MNDO method calculating the solvation energies of hydrophobic ions. The results are compared to the corresponding LD/MNDO results. General conclusions and comments are given in the final section.

COMPUTATIONAL METHODS

Quantum Chemical Calculations of the Solute Electronic Structure

To clarify the use of our microscopic solvent model we start by considering a rather simple case where the solvent molecules have a fixed charge distribution (neglecting electronic polarizability effects). Such a case can be treated in a *partition* formulation, considering formally the solute–solvent system as a supermolecule and writing the MO functions as an LCAO of the entire system (see also reference 12c).

$$\phi_i = \sum_{\mu} v^S_{\mu i}\chi^S_{\mu} + \sum_{\lambda} v^s_{\lambda i}\chi^s_{\lambda} \tag{1}$$

where S and s designate the solute and solvent, respectively, the v are the MO coefficients and the χ are AO wave functions. We also assume that atomic orbitals χ^S and χ^s are orthogonal. The coefficients $v^S_{\mu i}$ and $v^s_{\lambda i}$ can be obtained by solving the SCF equation for the supersystem.

$$\mathbf{F}\mathbf{v}_i = \varepsilon_i\mathbf{v}_i \tag{2}$$

where the **F** matrix can be separated to blocks, describing the solute–solute, solvent–solvent interactions, and the solute–solvent interactions

$$\begin{vmatrix} \mathbf{F}^S & \mathbf{F}^{Ss} \\ \mathbf{F}^{Ss} & \mathbf{F}^s \end{vmatrix} \tag{3}$$

where the **F** matrix elements are given in reference 12c. The assumption that the solvent and solute AOs are orthogonal to each other, implies, within the CNDO approximation, that $\mathbf{F}^{Ss} = 0$; the assumption that the overlap integral $\langle\chi^S_{\mu}|\chi^s_{\lambda}\rangle$ is zero implies, at this level of approximation, that $\beta^{Ss} = 0$ and therefore $P_{Ss} = 0$ (where β and P designate, respectively, the relevant resonance integral and bond order[12c]). This also implies that the charge transfer interaction between the solute and solvent is neglected. With $\mathbf{F}^{Ss} = 0$, we obtain two sets of solutions to the SCF equation—$\phi^S_i = (\mathbf{v}^S_{\mu i}, \mathbf{0})$ and $\phi^s_j = (\mathbf{0}, \mathbf{v}^s_{\lambda j})$. This means that coefficients $v^s_{\lambda i}$ and $v^S_{\mu j}$ are equal to 0, and allows us to consider MOs of the solute separately from the solvent MOs. Treating the complete **F** matrix within the CNDO-type all valence electron approximation one obtains for the solute part of the **F** matrix

$$\begin{aligned} F^S_{\mu\mu} &= (F^S_{\mu\mu})_0 - \sum_{B\in s} q_B\gamma_{AB} \quad \mu \in A \in S \\ F^S_{\mu\nu} &= (F^S_{\mu\nu})_0 \end{aligned} \tag{4}$$

where $(\mathbf{F}^S)_0$ stands for the SCF matrix of the isolated solute molecule, A and B are, respectively, the solute and solvent atoms and γ_{AB} is the two-electron repulsion integral, which accounts here for the solvent influence on the solute electronic structure. The charge distribution q_B is given by $(Z_B - P_B)$, where Z_B and P_B are, respectively, the core charge and the atomic bond order for the Bth atom (see reference 12c for more details).

Taking eq. (4) and approximating γ_{AB} by e^2/r_{AB}, we obtain for any solvent configuration[8]

$$F^S_{\mu\mu} \cong (F^S_{\mu\mu})_0 - \sum_B e^2q_B/r_{AB} = (F^S_{\mu\mu})_0 - U_A \tag{5}$$

where $\mu \in A$ and U_A designates the total electrostatic potential from the solvent atoms at the site of atom A. This equation can be used in the more general case where the solvent charge distribution is not fixed and can be polarized by the field of the solute (see below). Here, one should estimate the solvent charges and the new value of the corresponding solvent potential U_A at each SCF iteration of the solute electronic structure. Double self-consistency (for the wave function of the solute and for the solute–solvent charge distributions) can be achieved in this manner. Of course, the results of eq. (5) depend on the solvent configurations (that determine the r_{AB}s). These configurations fluctuate around an average polarization which is determined by the solute charges and should be evaluated by the specific solvent model used in the given simulation (see below).

Using the dipole approximation for the solute–solvent interaction we can express the diagonal elements of the solute **F** matrix in the following form

$$F^S_{\mu\mu} \cong (F^S_{\mu\mu})_0 - \sum_k \mathbf{m}_k\mathbf{r}_{kA}/r^3_{kA} \tag{6}$$

where $\mathbf{m}_k$ is the dipole of the kth solvent molecule and $\mathbf{r}_{kA} = \mathbf{r}_A - \mathbf{r}_k$. With this **F** matrix (and if needed with an additional CI treatment based on the corresponding SCF eigenvectors) we can evaluate the potential energy of our solvated molecule (relative to a reference where the solute is kept in the gas phase), by

$$\begin{aligned} E_{\text{sol}}(\mathbf{r}^S) = [&E(\mathbf{U} = \mathbf{U}^s, \mathbf{r}^S) \\ &- E(\mathbf{U} = 0, \mathbf{r}^S_{\text{gas}})] + E'_{Ss} + \Delta E_{ss} \end{aligned} \tag{7}$$

where $\mathbf{U}^s$ is the vector that represents the Coulombic potential due to the solvent at the site of the solute (the components of this vector are the U_A's of eq. (5)). $E(\mathbf{U})$ is the quantum mechanical energy evaluated by the standard SCF formulation (e.g., see reference 12c) with the **F** matrix of eq. (5) that corresponds to the given **U**. Here, $\mathbf{r}^S_{\text{gas}}$ is the gas phase

equilibrium geometry of the solute molecule and E'_{Ss} represents the nonelectrostatic solute solvent interaction term (e.g., the van der Waals interactions) and the energy invested in polarizing the solvent dipoles towards the solute charges. ΔE_{ss} is the *change* in the solvent–solvent interaction potential upon solvation of the given solute molecule. The actual calculations of the solute–solvent interactions depends of course on the solvent model used. Here, we use two microscopic solvent models that will be considered below.

Calculations of Solvation Energies with the LD Model

Probably the simplest way to incorporate the microscopic solvent effect in the solute Hamiltonian and to obtain reasonable solvation energies is provided by the LD solvent model.[3,5] This model divides the solute–solvent system into three regions. The solute molecule (or sometimes only its reactive part) is referred to as region I and is treated quantum mechanically. This region is surrounded by a grid of point dipoles, referred to as region II, which represents the average polarization of the corresponding solvent molecules and include, in general, permanent, and induced components. The solvent outside this region is represented by a continuum media and is referred to as region III. The average polarization of the solvent permanent dipole, $\mathbf{m}^{\text{perm}}$, is related to the corresponding local field in a self-consistent iterative way using a Langevin type equation[8]

$$\mathbf{m}_i^{\text{perm},n+1} = \mathbf{e}_i m_0 \left(\coth X_k^n - \frac{1}{X_k^n} \right) \quad k = 1, N$$
$$X_k = \frac{m_0 \xi_k^n}{k_B T d(r_k)} \tag{8}$$

where ξ_k^n is the electric field on the kth solvent dipole from the solute charge distribution and all the other solvent permanent dipoles in the nth iteration. $\mathbf{e}_k$ is a unit vector in the direction of ξ_k, N is the total number of solvent dipoles. The starting value of ξ_k in the iteration procedure is taken as the corresponding field from the solute. Here k_B is the Boltzmann constant and $d(r_k)$ is a screening function that represents the attenuation of the solute field on the kth dipole by the other solvent molecules[3] (where r_k is the distance between the solvent and the closest solute atom). The function $d(r)$, that *should not* be confused with the frequently used inconsistent distance-dependent dielectric function (see reference 3 for discussion), is usually taken as unity in iterative treatments.[15] However, it is found here that similar results and faster convergence can be obtained with $d(r) = r - 2$ for $r \geq 3$, which is usually used in noniterative treatments.[3] The temperature T in eq. (8) is taken here as 298 K.

In addition to the polarization of the permanent dipoles of the LD model, we also consider the induced solvent dipoles. These dipoles are evaluated by the classical approximation

$$\mathbf{m}_k^{\text{ind}} = \alpha \xi_k \tag{9}$$

where α is the polarizability of the solvent molecules and ξ_k is the electric field on the kth solvent molecule. The field ξ_k is calculated self-consistently considering the field from the solute charges as well as the permanent and induced solvent dipoles.[8] The total polarization of each solvent dipole is taken as the sum of the corresponding permanent and induced components ($\mathbf{m}^{\text{perm}}$ and $\mathbf{m}^{\text{ind}}$). The SCF equation for the LD/MNDO model is taken now as

$$(\mathbf{F}_{\mu\mu}^S)^{(n)} = (\mathbf{F}_{\mu\mu})_0 - \sum_k (\mathbf{m}_k^{\text{perm}} + \mathbf{m}_k^{\text{ind}})^{(n-1)} \mathbf{r}_{kA} / r_{kA}^3 \tag{10}$$

where the index n indicates that we are dealing with an iterative procedure that considers the self-consistent interaction between the solute charges and the solvent dipoles.

The solvation energy is now evaluated by eq. (7), which can be approximated by

$$E_{\text{sol}} = E_{\text{elec}} + \Delta\Delta E^S_{\text{gas}\rightarrow\text{sol}} \quad E_{\text{vdw}} \tag{11}$$

where E_{elec} is the LD energy that can be expressed as[3]

$$E_{\text{elec}} = E_{\text{perm}} + E_{\text{ind}} + E_{\text{bulk}} = \frac{1}{2} \sum_A q_A U_A + E_{\text{bulk}} \tag{12}$$

Here, E_{perm} and E_{ind} are the contributions from the permanent and induced solvent dipoles, respectively. E_{bulk} designates the interaction between the solute charges and the bulk solvent outside the sphere of the explicit LD dipoles. This interaction is evaluated[3] by using the Kirkwood equation,[10b,c] which is reduced to the Born's formula for a solvated ion and to the Onsager's equation[10a] for a solvated dipole. q_A is the charge of the Ath solute atom and U_A is the potential from the solvent molecules at the site of this atom. The factor half in the $\frac{1}{2}\Sigma q_A U_A$ term reflects the penalty associated with polarizing the solvent, which is estimated by the linear response approximation to be a half of the energy gained by the interaction between the solvent dipoles and the solute charges. The potential U_A does not include the bulk effect, which is estimated separately by E_{bulk}. This approach neglects the small polarization of the solute by the potential from the bulk region (the calculations include, of course, the solute polarization by the LD grid). The van der Waals term is given by

$$E_{\text{vdw}} = \sum_{ij} \varepsilon \left[\left(\frac{r^*}{r_{ij}} \right)^{12} - 2 \left(\frac{r^*}{r_{ij}} \right)^6 \right] \tag{13}$$

where i and j run over the solute and the solvent atoms, respectively. $\Delta\Delta E^S$ is the change in the solute internal energy as a result of its polarization by the solvent molecules. This term is expressed as

$$\Delta\Delta E^S_{\text{gas}\rightarrow\text{sol}} = E\{\mathbf{U} = \mathbf{0}, \mathbf{v} = \mathbf{v}(\mathbf{U} = \mathbf{U}^s)\} - E\{\mathbf{U} = \mathbf{0}, \mathbf{v} = \mathbf{v}(\mathbf{U} = \mathbf{0})\} \quad (14)$$

where the notation $\mathbf{v}(\mathbf{U})$ designates the SCF MO eigenvectors obtained by using the $\mathbf{F}$ matrix with the given Us. Thus, the first term on the right side of eq. (14) is obtained by freezing the $\mathbf{v}$s vectors at their values in solution, while setting U to zero.

In constructing the LD grid we distribute the dipoles in the first solvation shell by a special procedure (see below), while situating the rest of the solvent dipoles on a regular cubic grid which is extended around the center of the solute region up to a radius, R_g, of typically 16 Å (Fig. 1). The coordinates of the dipoles in the first solvation shell are generated by the following procedure: (1) A dense grid of evenly distributed points is built on the solvation surface of the solute molecule. This is done by constructing a spherical surface of grid points around each atom at a radius $R_S + R_s$ (where R_S and R_s are the van der Waals radii of the given atom and the solvent, respectively) and deleting each point which is within the spherical surface of the other atoms. The resulting surface is similar to the *surface accessible to the solvent* as defined by Lee and Richards[18] (see also reference 19). (2) The dense grid is used to construct a coarse grid by sequentially picking points with a spacing (D_w) that corresponds to the density of the given solvent (3.1 Å for water). The computer program also allows one to place point dipoles of the coarse gride preferentially in specific hydration sites of the solute (e.g., along O—H, N—H bonds and near electron lone pairs). The numbers of dipoles generated in the first solvation shell of different molecules coincide relatively well with estimates from MD simulations.

One of the practical problems that may appear in the LD modeling of the solvent is associated with the symmetry properties of the solute wave function. That is, if performing only a single LD calculation one might find that the solvent polarizations deviate from the symmetry of the isolated solute in the gas phase (in this case the solvent potential $\mathbf{U}^s$ may lack the symmetry properties of the charge distribution of the isolated solute). However, it seems reasonable to use a symmetrical charge distribution that reflects an average over a large number of solvent–solute molecular ensembles. In order to reproduce such a charge distribution with the LD model we take the solvent potential at the solute atoms as the average over several (usually 4–6) different coarse grid configurations.

The parameters of the LD model are given in Table I. The dipole moment, molecular polarizability, and the grid spacing of the solvent dipoles are taken as the corresponding experimental values of water.[20] The van der Waals radii of the solvent dipole, R_s, is taken as 1.5 Å. The van der Waals radii of the solute atoms, R_S, correspond to a set of refined radii that were found to give the best fit between the calculated and observed solvation energies for various representative molecules that contain the given atoms.

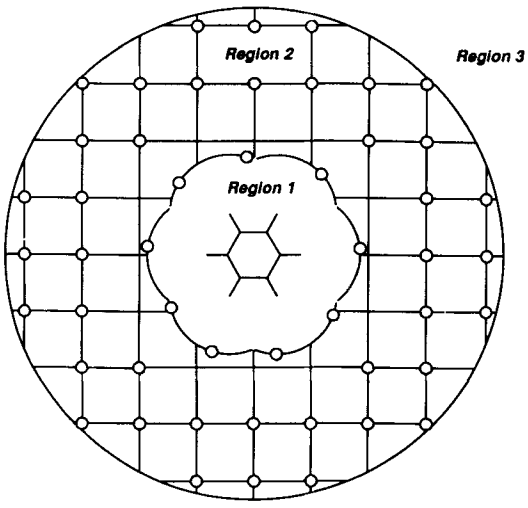


Figure 1. Illustrating the two different regions in the LD model and the structure of the solvation grid. The figure shows the grid points in the plane of a solvated benzene molecule. The inner grid points are generated around the solvation surface of the solute, while the rest of the points are constructed from a regular cubic grid (see text for additional details)

Table I. The parameters of the LD model.[a]

Solvent dipoles parameters	
Parameter	Value
radius, R_s	1.50 Å
grid spacing, D_w	3.10 Å
dipole moment, m_0	1.85 D
polarizability, α	1.48 Å^3
grid radius, R_g	10–19 Å
Dipole–dipole cutoff, R_d	10–38 Å

Solute van der Waals radii	
Atom	R_S(Å)
H(−C)	1.50
H(−C′)	1.20
H(−N, O)	1.10
C	1.90
C′	2.00
N	1.40
O	1.20
N^+, O^-	1.10

[a]C and C′ designate, respectively, aliphatic, and aromatic carbon atoms.

Calculations of Solvation Free Energies with the LD Model

The treatment presented in the previous section uses the LD model to estimate solvation enthalpies. However, the same method can be conveniently used to estimate solvation free energies. Firstly, the solvation free energies of *ions* are usually very similar to the corresponding solvation enthalpies (typically[3,12b] ΔH_{sol} is about 5% larger than ΔG_{sol}). Secondly, electrostatic free energies are *much more* stable with respect to various perturbations (e.g., temperature changes) than electrostatic enthalpies. That is, at low temperatures the solvent dipoles are strongly polarized towards the solute charges, maximizing the enthalpic contribution to the solvation free energy, while at higher temperatures the solvent polarization and the corresponding enthalpic contribution is reduced but the entropic contribution is increased (this minimizes the change in free energy). In fact, electrostatic continuum models, that reflect the solvent average polarization, represent free energies and not enthalpies. Thus, it is reasonable to recalibrate the LD polarization so that it will reproduce solvation free energies at room temperatures.[12b] This can be done by adjusting the solute dipole, m_0, from 1.85 D to 1.63 D using

$$\Delta G_{elec} = E_{elec}(m_0 = 1.63\ \text{D}) \qquad (15)$$

A better approximation can be obtained by a minor recalibration of the van der Waals radii of the solute.

In addition to the above electrostatic term, one has to estimate the free energy associated with the hydrophobic effect. Although this contribution is hard to evaluate in a rigorous way it seems to represent a rather simple physics. That is, a water molecule in the bulk has N_1 orientations with a low energy of about -10 kcal/mol. On the other hand, a water molecule at the surface of a nonpolar molecule has a smaller number (say N_2) of low energy orientations (since some of the orientations that were accessible in the bulk do not allow now for a favorable interactions with the surrounding solvent). Comparing the partition functions of the two water molecules gives an entropy contribution of ca. $-R\ln(N_2/N_1)$. Thus, the overall hydrophobic effect can be estimated by evaluating the above contributions for all the water molecules at the surface of the given solute. This can be done by taking the Langevin dipoles on the solute's surface, rotating them in the six grid directions and evaluating the corresponding partition functions. Instead, however, one may simply count the number of the surface dipoles and scale it by an empirical constant that reproduces the observed free energy of transfer of hydrocarbons from nonpolar solvents to water. Although such a procedure is quite effective,[12b] it does not take into account properly the hydrophobic effect for polar or partially polar solutes. Here again, one may suggest to evaluate the partition functions for each surface dipole. However, considering the strong solute fields, it is unlikely that the corresponding contributions to the partition function will give stable results without a very extensive and expensive averaging procedure. Thus, we introduce here a simplifying procedure *assuming* the dependence of the hydrophobic effect on the local field exerted on the surface solvent molecules. In this way, when the field is very small, we retain the regular hydrophobic effect associated with a nonpolar surface, while at large fields the hydrophobic effect vanishes. Note that the entropic contributions associated with fixing the solvent polarization in strong electric fields are already included in our ΔG_{elec}. The field-dependent hydrophobic energy term is expressed as

$$\Delta G_{hyd} = -T\Delta S_{hyd}$$

$$\Delta S_{hyd} = A + B \sum_{i \in s'} CF(\xi_i^\circ)$$

$$F(\xi_i^\circ) = \begin{cases} 1 & \xi_i^\circ < \xi_{phobic} \\ \left(1 - \dfrac{\xi_i^\circ - \xi_{phobic}}{\xi_{philic} - \xi_{phobic}}\right) & \xi_{phobic} < \xi_i^\circ < \xi_{philic} \\ 0 & \xi_i^\circ > \xi_{philic} \end{cases} \qquad (16)$$

where i runs over the sites of the Langevin dipoles on the surface (s') of the solute. F is a function of the electric field, ξ_i°, on the ith Langevin dipole at the solvation surface. F is 1 when the field is small ($<\xi_{phobic}$) and 0 when the field is large ($>\xi_{philic}$). Here, A, B, C, ξ_{philic} and ξ_{phobic} are scaling constants whose values are respectively, 10.57 cal mol^{-1}K^{-1}, 0.044 cal mol^{-1}K^{-1} Å^{-2}, 10 Å^2, 10^{-2} au Å^{-2}, and 10^{-4} au Å^{-2}. These constants were calibrated by considering the transfer of various molecules from hydrocarbons to water and minimizing the differences between the calculated and observed free energies for these transfer processes.

The total solvation free energy is now estimated by

$$\Delta G_{solv} = \Delta G_{elec} + \Delta\Delta E^S + E_{vdw} + \Delta G_{hfb} \qquad (17)$$

FEP Calculations of Solvation Free Energies

Although the LD model provides a very convenient and quite reliable way of obtaining solvent effects on the electronic structure of the solute molecules, it is important to explore more complete solvent models. Thus, we also examine in this work the combination of the Surface Constrained All Atom Solvent (SCAAS) model with the AMPAC package. The SCAAS model has been introduced before[16] and is an extension of our earlier surface constrained dipolar models.[3,12b] This model represents explicitly all the solvent atoms, using a standard force field with

van der Waals and electrostatic terms as well as intramolecular terms.[16] The model involves spherical boundary conditions that constrain the molecules on the surface of the finite simulation system (usually 100 solvent molecules) to have the same polarization as they would have in an infinite solvent system (for more details see reference 16 and for earlier related models see references 12a and 12b).

The SCAAS model has been integrated with the AMPAC package following basically the procedure prescribed in our earlier studies, which incorporated the MINDO approach with a protein force field[8] and with dipolar solvent models.[12b] This procedure involves the incorporation of the electrostatic potential in the Hamiltonian of the solute through eq. (5), while treating the solute–solvent van der Waals interactions classically through eq. (13). The total potential energy of the system is expressed in a similar way to that of eq. (11) as

$$E = E_S\{\mathbf{F}(\mathbf{U})\} + E'_{Ss} + E_{ss} \tag{18}$$

where E_S is the energy obtained quantum mechanically with the **F** matrix that includes the given **U**s. E'_{Ss} is the nonelectrostatic solute–solvent interaction term and E_{ss} is the solvent–solvent classical force field. In studies of very large solute molecules, we sometimes divide the solute region into quantum and classical parts (see also reference 8). The "connection" between the quantum and classical regions is treated by a classical force field (which is included in E'_{Ss}), where the quantum atoms at the boundary are connected to dummy hydrogen like atoms in order to balance the electrons in the quantum system.

In implementing the SCAAS/AMPAC procedure in FEP calculations we use different strategies for calculations of solvation energies and for calculations of potential surfaces of chemical reactions in solutions. In calculations of solvation free energies we consider a two-step cycle. In the first step, we take the gas phase solute and transform it to a hypothetical state where it can only interact with the solvent through the van der Waals interaction term E'_{Ss}. The free energy involved in this step is associated with the formation of the solvent cavity around the solute and is referred to as ΔG_{cav}. In the second step, we turn on the solute–solvent interaction term, obtaining the electrostatic free energies associated with this charging of the solute in the solvent cavity. The total solvation free energy is now given by

$$\Delta G_{\text{sol}} = \Delta G_{\text{elec}} + \Delta G_{\text{cav}} \tag{19}$$

The leading term is, of course, ΔG_{elec}. This term is obtained by a rather simple procedure, replacing the potential surface of the solute–solvent system by a mapping potential of the form (see for example reference 21a)

$$\begin{aligned}\varepsilon_m &= \varepsilon_1(1 - \lambda_m) + \varepsilon_2\lambda_m \\ &= (E - E_{Ss}^{\text{elec}})(1 - \lambda_m) + E\lambda_m\end{aligned} \tag{20}$$

where ε_1 and ε_2 are the potential surfaces that describe, respectively, the solvated solute molecule with zero residual charges and with the charge set obtained from the given quantum mechanical method.

The potential surface ε_2 is taken as the actual SCAAS/AMPAC potential surface, (the E of eq. (18)). The potential surface ε_1 is taken as E minus the solute–solvent electrostatic interaction term E_{Ss}^{elec} (where $E_{Ss}^{\text{elec}} = \Sigma_A q_A U_A$), while the forces (and induced dipoles) associated with ε_1 are obtained by setting the solute charges to zero. Note that with $\lambda_m = 0$, the solvent is not polarized by the solute and $\langle U \rangle = 0$. Using the potential ε_m in a series of molecular dynamics (MD) simulations, where λ_m is varied from zero to one in n steps gives (see for example reference 21)

$$\begin{aligned}\delta G(\lambda_m \rightarrow \lambda_{m'}) &= -(1/\beta)\ \ln[\langle \exp(-(\varepsilon_{m'} - \varepsilon_m)\beta)\rangle_m] \\ \Delta G_{\text{elec}} &= \Delta G(\lambda_n) = \Delta G(\lambda_0 \rightarrow \lambda_n) \\ &= \sum_{m=0}^{n-1} \delta G(\lambda_m \rightarrow \lambda_{m+1})\end{aligned} \tag{21}$$

where $\beta = (k_B T)^{-1}$ and k_B is the Boltzmann constant.

The rather small free energy associated with the cavity formation, ΔG_{cav}, can be evaluated by using the force field of $\varepsilon_1(\mathbf{q} = 0)$ and changing the van der Waals parameters in this force field from zero (which corresponds to the absence of the solute molecule) to their actual value.

In calculating activation free energies of chemical reactions in solutions, one faces a much more serious challenge than in the evaluation of solvation energies. That is, it is not clear what quantum mechanical surfaces should be used in the ε_m that transforms the system from the reactant to the product state. One might suggest to use for ε_1 and ε_2 the SCF surfaces obtained by constraining the bond orders to their average values in the product and reactant state, respectively. Unfortunately, the potential ε_m obtained in this way *does not* correspond to the SCF ground state potential along the reaction coordinate and the corresponding maximum of $\Delta G(\lambda)$ cannot be identified with the activation barrier of a given reactions. One might try to use the bond order as a mapping parameter, but again the resulting mapping potential would not coincide with the proper ground state potential since the unconstrained bond order should change with the solvent fluctuations.[14b] This problem can be resolved by adopting a strategy introduced in our EVB studies,[14b,21a,21b] where a mapping potential based on the diabatic surfaces was

used to obtain the free energy of the quite different adiabatic ground state potential. Here, we exploit this idea by using a combination of two empirical potential functions (ε_1 and ε_2) as a reference potential for the SCF ground state surface. This is done by starting with eq. (20) but using for ε_1 and ε_2 two empirical force fields that correspond, respectively, to the reactant and product states (this is usually done with the EVB potential functions, but the above mentioned potentials obtained with constrained bond orders can also be used). The difference between the mapping potential ε_m and the SCF ground state energy, E, can now be used to transform the mapping free energy $\Delta G(\lambda_m)$ to the ground state free energy using[14b]

$$\exp\{-\Delta g(X_n)\beta\} = \exp\{-\Delta G(\lambda_m)\beta\}\,\langle\exp\{-[E(X_n) - \varepsilon_m(X_n)]\beta\}\rangle_m \quad (22)$$

where X_n is the given value of the reaction coordinate, which is typically taken as the difference between ε_2 and ε_1 and λ_m is the λ that keeps the system near the given X_n.

The Solute Quantum Mechanical Model

The electronic properties of the solute molecules are calculated here with the semiempirical MNDO method using the AMPAC package,[17a] although the AMI method can be used as well. Optimized gas phase geometries are used in all calculations. The electrostatic potential from the solvent permanent and induced dipoles is added to the diagonal elements of the Hamiltonian according to eq. (5). The equilibrium values of the solvent dipoles in the solute field are calculated for each step of the SCF iteration procedure. In this way we achieve self-consistency between the solute charge distribution and the potential from the solvent molecules. The main time consuming stage in this procedure is the iterative recalculation of solvent dipoles at each MNDO SCF iteration. The LD/MNDO calculations presented in this work do not involve geometry optimization of the solvated solute molecule. However, this can be easily accomplished using the MNDO solute gradient vector, while updating the solvent potential every step in the minimization procedure. Such an approach for solvent dependent geometry optimization is provided with the QCFF/SOL program, which is a part of the MOLARIS package.[22]

The atomic charges of isolated molecules obtained by the MNDO method appeared to be very similar to the charges obtained by *ab initio* STO-3G calculations. However, a more realistic calculations with extended basis sets (3-21G, 6-31G*) indicate that the MNDO calculations underestimate significantly the residual atomic charges in different classes of organic compounds.[1] Thus, using the MNDO charges in LD calculations underestimates the corresponding solvation energies. To overcome this drawback we scale the MNDO charges by a factor $f_{sc} = 2.0$, which was empirically determined by comparing calculated MNDO and *ab initio* 6-31G* charges[1] for aliphatic and aromatic hydrocarbons, *n*-alcohols, and amines. The scaling of the charges is performed at the final calculations of the solvation energies (after the end of the quantum chemical SCF iterations). This kind of scaling is performed only for neutral molecules, while for ionic species we use the actual MNDO charges. The above scaling procedure is not fully consistent, since the solvent potential which is included in the solute Hamiltonian reflects the *unscaled* solute charges. However, including the potential due to the scaled solute charges appears to lead to overpolarization of the solute. This problem might be resolved by reparametrizing the solute semiempirical integrals (for example, the solute–solute electron–electron repulsion integrals are underestimated in semiempirical models and therefore the response of such models to external potential might not be properly balanced). However, such a reparametrization procedure is out of the scope of the present study.

RESULTS

LD/MNDO Studies of Solvation Energies of Organic Ions

Previous LD and PDLD calculations of ionic solvation energies have produced reliable estimates of E_{sol} with a fixed charge distribution of the solute and a simplified uniform cubic grid for the water dipoles.[3,12b] Most of these calculations were performed with fixed solute charge distributions, using a large grid radius (R_g) and a large dipole–dipole cutoff distance (R_d). Here, however we are involved with quantum mechanical SCF iterations which require many repeating LD calculations. Thus it is useful to find the smallest possible values for R_g and R_d that still provide reasonable solvation energies. To optimize this parameters we evaluated E_{sol} for the NH_4^+ ion with different values of R_d and R_g. The resulted values of $E_{sol}(NH_4^+)$ as a function of the grid parameters are given in Figure 2 and Table II. The contribution of $E_{perm} + E_{ind}$ to E_{elec} increases and the bulk energy decreases when the dimension of the solvent grid increases. The changes of these two terms start to compensate each other for $R_g \geq 13$ Å, and, accordingly, the results become almost independent of the number of solvent dipoles. Another important feature of the ion solvation is associated with the alignment of the solvent dipoles towards the electric field of the ion. This polarization effect leads to destabilizing electrostatic interactions between the solvent dipoles themselves. Since this dipole–dipole repulsion is a long range effect, the most realistic results

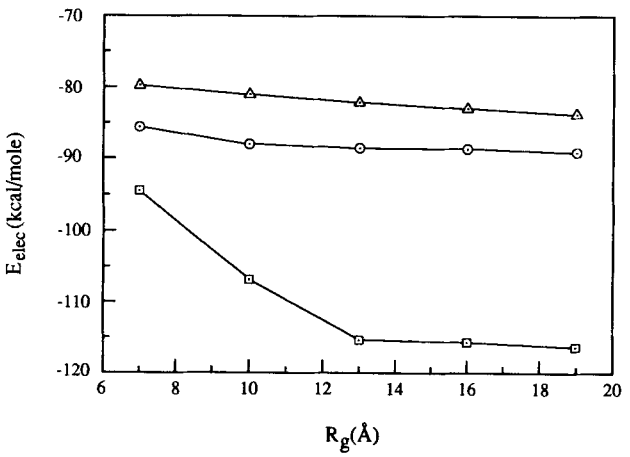


Figure 2. The dependence of the electrostatic solvation energies E_{elec} of an NH_4^+ ion on the radius of the solvation sphere, R_g and the dipole–dipole cutoff radius, R_d. Calculations with $R_d = 2R_g$, $R_d = R_g$, and $R_d = 0.5R_g$ are designated, respectively, by △, ○, and □.

should be obtained with the largest values of R_g and R_d. As can be seen from the plot of E_{sol} for $R_d = 2R_g$ (where the interactions between all the dipoles within the solvation sphere are considered) the absolute value of E_{sol} increases when R_g changes from 7 Å to 16 Å and then remains practically constant. E_{sol} becomes larger as the ratio R_d/R_g decreases and the calculations with $R_d = 0.5R_g$ give unrealistic results for E_{sol} for the almost all R_gs considered. The calculated E_{sol} have the most reasonable values when $R_d = 2R_g$, for $R_g > 13$ Å. Interestingly, calculations with $R_d = R_g$ and $R_g > 10$ Å yield an almost constant value of E_{sol}, which is only ca. 4 kcal/mol lower than that obtained in the most accurate calculations (with large R_g and $R_d = 2R_g$). In view of this results we recommend the use of $R_d = R_g$ as a very useful approximation.

The solvation energies, E_{sol}, for a series of methyl substituted ammonium ions, imidazole, negative carboxy, and metoxy ions are given in Table III. Good agreement is obtained between the calculated results and the corresponding experimental data. For alkylammonium ions the values of E_{sol} decrease upon increase in the number of alkyl groups as a result of the decrease of both E_{perm} and E_{ind}. The

Table II. The dependence of ΔE_{elec} of a solvated NH_4^+ ion on the radius of the solvent grid.[a]

R_g(Å)	$\overline{n}_{dip}$	E_{bulk}	$E_{perm} + E_{ind}$	ΔE_{elec}
7.0	41	−19.2	−66.7	−85.9
10.0	131	−14.2	−74.2	−88.4
13.0	289	−11.3	−77.9	−89.2
16.0	571	−9.3	−80.3	−89.6
19.0	933	−8.0	−81.5	−89.4

[a]Energies in kcal/mol, R_g is the radius of the solvent grid and $\overline{n}_{dip}$ is the average number of the solvent dipoles for the given value of R_g. The value of the dipole–dipole radius, R_d, is taken as the corresponding value of R_g.

Table III. Calculated (LD/MNDO) and observed solvation energies of typical organic ions.[a]

Ion	$\Delta\Delta E^S$	E_{perm}	E_{ind}	ΔE_{sol}	ΔE_{sol}^{obs}
NH_4^+	0.0	−64.9	−15.4	−86.2	−86 ± 5[b]
$N(CH_3)H_3^+$	0.3	−62.7	−10.7	−80.1	−76 ± 5[b]
$N(CH_3)_2H_2^+$	0.3	−58.2	−7.4	−73.4	−70 ± 5[b]
$N(CH_3)_3H^+$	0.1	−53.8	−5.5	−67.7	−63 ± 5[b]
ImH^+	0.1	−56.6	−6.6	−70.1	−65 ± 10[c]
$CHOO^-$	0.8	−66.3	−13.9	−86.6	−83 ± 5[c]
CH_3O^-	1.6	−67.4	−15.2	−94.1	−99 ± 5[c]

[a]Energies in kcal/mol. The calculations were done with the $R_g = R_d = 16$ Å, but the resulted E_{elec} were reduced by 4% to account for the overestimate of the solvation energy by this approximation (Fig. 2). ImH^+ designates protonated imidazole.
[b]The observed value is taken from reference 23.
[c]The observed values are taken from reference 3.

overall difference between the calculated E_{sol} of the two ions with the smallest and largest dimensions in this series (NH_4^+ and $N(CH_3)_3H^+$) is in a very good agreement with the corresponding experimental value[23a] (21.0 kcal/mol and 20.7 kcal/mol for the calculated and observed values, respectively). The solvation energies of imidazole and oxygen containing ions are within the error range of the indirect estimates made from experimental data.[3] It should be noted that in each case the different spatial configurations of the solvent dipoles around the ions provide quite similar solvation energies. The standard deviation of E_{sol} due to averaging over different solvent grids are between 1.1 and 2.2 kcal/mol.

LD/MNDO Calculations of Solvation Energies of Neutral Molecules

The calculated solvation energies of neutral molecules with different polarity are given in Table IV. The calculated E_{elec} are on the average several kcal/mol lower than the corresponding experimental solvation enthalpies. This result is quite reasonable, since for neutral molecules the E_{vdw} term makes a substantial contribution to total solvation enthalpy.

The polarization of the solute by the solvent field manifests itself in the change in the solute dipole moment upon transfer from the gas phase to solution (Table IV). In general the dipole moment of polar molecules appear to increase in solution. On the other hand in nonpolar molecules like benzene or propane we find that although the polarization of the individual C—H bonds increases the total molecular dipole moment remain small, due to the fact that the bond dipoles are pointing in opposite directions.

For methanol our calculations give E_{elec} of −5.90 kcal/mol (a previous estimate based on a continuum approach is −6.0 kcal/mol[11]). One could expect that a water molecule should have a greater electrostatic solvation energy than the −6.25 kcal/mol obtained here. The reason for this underestimate is associated

Table IV. Calculated (LD/MNDO) and observed solvation energies (in kcal/mol) of some neutral molecules.[a]

Molecule	μ_{gas}	μ_{sol}	E_{elec}(OH)	E_{perm}	E_{ind}	E_{elec}	E_{sol}	$\Delta E_{sol}^{obs(b)}$
Water	1.78	1.90	−5.89	−5.57	−0.96	−6.25	−8.80	−9.97
Methanol	1.48	1.69	−6.77	−5.24	−1.02	−5.90	−10.11	−10.25
1-Butanol	1.42	1.78	−6.38	−4.81	−0.96	−5.66	−13.56	−13.95
Phenol	1.22	1.88	−6.22	−7.59	−1.17	−7.97	−13.27	—
Aniline	1.47	1.51	—	−4.41	−0.70	−4.85	−10.45	—
Benzene	0.00	0.03	—	−2.46	−0.33	−2.64	−7.54	−7.08
Propane	0.00	0.00	—	0.00	0.00	0.00	−4.80	−4.83

[a]The calculations were done with $R_g = R_d = 16.0$ Å. E_{elec}(OH) designates the contribution of the O—H group of the given molecule to E_{elec}. μ_{gas} and μ_{sol} are molecular dipole moments (in D) in the gas phase and in solution, respectively.
[b]The observed solvation enthalpies, ΔE_{sol}^{obs}, are taken from reference 24.

with the insufficient accuracy of the atomic charges. That is, the charge separation in the water molecule is underestimated when compared to that in order molecules with a −OH group. The calculations of water solvation with the $f_{sc} = 2.3$ (this value of f_{sc} allows to exactly reproduce the *ab initio* 6-31G* charges for water molecule[1]) give $E_{elec} = -9.2$ kcal/mol, which is closer to the experimental hydration energy.

The partition of the solvation energy into atomic contributions makes it possible to analyze the corresponding contributions of different functional groups. Here, the water molecule, *n*-alcohols and phenol share a common hydrophilic O—H group and the solvation energies of the O—H group is similar for these compounds. In addition to the hydration of the O—H group it is important to consider the hydration of the remainder part of the molecule. In case of *n*-alcohols the hydration energy of the hydrocarbon chain is small since the charges on the C and H atoms are rather small. The calculated difference between E_{elec} of benzene and propane (−2.64 kcal/mol) is mainly determined by the electrostatic terms and is close to the experimental difference between the corresponding enthalpies of solvation.[24] The pronounced effect of benzene stabilization is related to the ability of the solvent dipoles to align themselves with respect to the planar C—H bond dipoles, while the structure of the aliphatic hydrocarbon propane does not allow for energetically favourable orientation of the solvent dipoles around the C—H bonds.

To examine the performance of a given model with regard to the relative values of E_{sol} in a series of molecules, it is useful to compare the corresponding partition coefficients, log*P*, between nonpolar and polar media.[25] Here we consider three aromatic compounds (benzene, aniline, and phenol), where the entropy contributions to the total free energy of transfer are expected be similar, since these three molecules are associated with a common hydrophobic aromatic ring. In this case, the trend in log*P* should follow the corresponding difference between E_{sol} in aqueous solutions and in hydrocarbons. Indeed, the order of the calculated value of E_{sol} of this compounds (Table IV), E_{sol}(benzene) > E_{sol}(aniline) > E_{sol}(phenol), corresponds fairly well to the order of the experimental values[25] of log*P* in *n*-heptane/water system: 2.26, −0.03, and −0.92 for benzene, aniline, and phenol, respectively.

A more quantitative agreement should be obtained, however, when one considers the field dependent hydrophobic energy of eq. (16) and the E_{sol} contribution of the nonpolar media. A preliminary examination of the use of eq. (16) in the evaluation of transfer free energies is considered below.

LD/MNDO Calculations of Transfer Free Energies of *n*-Alcohols

It is commonly assumed that the partition coefficient of molecules between nonpolar and aqueous media is closely correlated with their binding affinity to receptor sites.[25,29,30] In fact, one of the primary guides

Table V. Thermodynamic quantities for the transfer of *n*-aliphatic alcohols from hydrocarbon media to water.[a]

$CH_3(—CH_2—)_nOH$	Calculated				Observed		
n	ΔE_{elec}^{w}	ΔE_{tr}	ΔS_{tr}	ΔG_{tr}	ΔE_{tr}	ΔS_{tr}	ΔG_{tr}
0	−6.26	−5.24	−13.2	−1.30	—	—	—
1	−6.19	−5.17	−14.7	−0.78	—	—	—
2	−6.52	−5.43	−16.1	−0.63	—	—	—
3	−6.09	−5.04	−17.7	0.25	−7.51	−25.8	0.19[b]
4	−6.23	−5.18	−19.3	0.58	−7.27	−27.8	1.02[b]
5	−5.77	−4.80	−21.0	1.47	−6.72	−28.1	1.65[b]

[a]ΔE_{elec}^{w}, ΔE_{tr}, ΔS_{tr}, and ΔG_{tr} denote, respectively, the electrostatic solvation energy (kcal/mol) in water, the enthalpy (kcal/mol), the entropy (cal $mol^{-1}K^{-1}$) and the free energy (kcal/mol) of transfer from hydrocarbon to aqueous solution. The calculations are done with $R_g = R_d = 16$ Å.
[b]The observed values are taken from reference 30

in the design of biologically active molecules is the knowledge of their partition coefficient and the corresponding transfer free energy. Thus, there is currently a signicant interest in reliable estimates of transfer free energies. This challenge can be addressed on a semiquantitative level by combining the LD estimate of the electrostatic free energy with the field dependent hydrophobic contributions of eq. (16). Here we considered in a preliminary way the class of alcohols, which contains both hydrophobic and hydrophilic structural groups. The results of our calculations as well as some experimental values of ΔH_{tr}, ΔS_{tr}, and ΔG_{tr} are given in Table V. All the molecules considered were taken in their extended all-trans conformations and the calculated energy contributions were averaged over six sets of solvent configurations in order to provide adequate description of the solvation of the terminal O—H groups.

The calculated values of ΔG_{tr} of the *n*-alcohols series can be correlated with the values of log*P*, obtained from the corresponding observed ΔG_{tr}s, with a slope of 0.80 (Fig. 3). This agrees with the slope of 0.72 expected, at 298K, from the relationship $\log P = \Delta G_{tr}/2.3RT$. More systematic studies are needed, however, in order to examine the validity of the present approach in quantitative evaluation of log*P*.

LD/MNDO and FEP/SCAAS/MNDO Studies of Hydrophobic Ions

In order to examine the relative performance of the LD/MNDO and the FEP/SCAAS/MNDO methods we considered the solvation energies of three organic ions with bulky hydrocarbon groups around their ionic centers. This includes the teraphenylboron (TPB^-) and tetraphenylphosphonium (TPP^+) that belong to the class of "hydrophobic ions" (which are characterized by a good permeability across lipid membranes[26]), and 2,2,6,6-tetramethylpiperidine-1-oxonium ($TEMPOH^+$), which is the protonated form of a stable nitroxyl radical widely used as a membrane probe.[27] Along with these ions the solvation of their neutral analogs—tetraphenylmethyl (TPM) and the stable radical 2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO), was also studied.

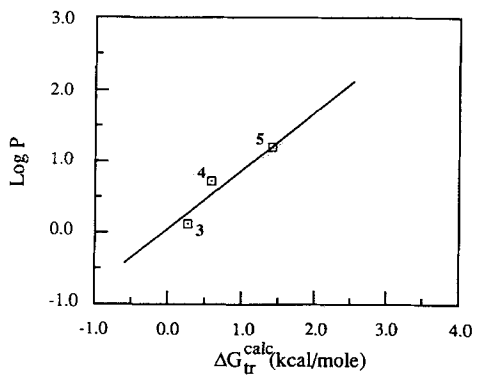
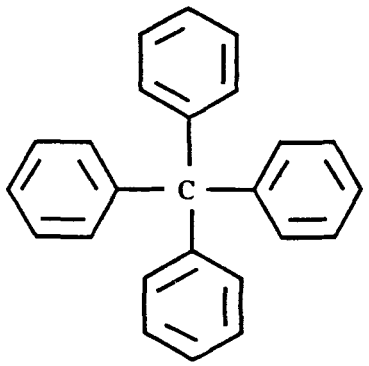


Figure 3. Showing the calculated correlation between log*P* and the free energy of transfer of *n*-alcohols from hydrocarbon media to water (Table V for more details).

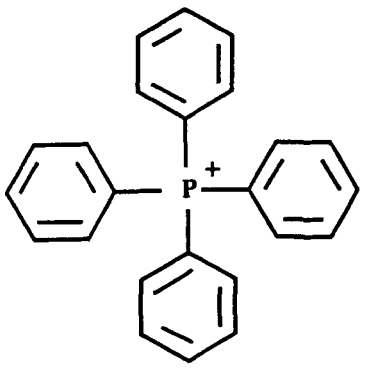
The structural parameters of TEMPO, $TEMPOH^+$, and TPM were determined by MNDO optimization. The geometries of TPB^- and TPP^+ were evaluated by minimizing the corresponding energies with respect to the lengths of the bonds between the phenyl rings and the central atom while leaving all other internal coordinates the same as in TPM. The calculated lengths of the central C—C, C—B, and C—P bonds were found to be 1.575 Å, 1.653 Å, and 1.777 Å respectively. The geometries of TEMPO and $TEMPOH^+$ were minimized at the chair conformation of the piperidine ring. The optimized structural parameters for the isolated molecules were used in the LD/MNDO calculations. The FEP/SCAAS/MNDO simulations used the approach described in the methods section, except that the geometries of the solute molecules were kept fixed at the same structures used in the LD study.

The LD/MNDO atomic charges of TPM, TPB^-, and TPP^+ are given in Table VI. The major differences between the charge distributions of these compounds manifest itself already in the gas phase. In TPB^- the charge of the B atom is only -0.243 e.u., while that of the P atom in TPP^+ is $+0.591$ e.u. and this pattern does not change drastically upon solvation. The electrons of the TPB^- anion are slightly more polarizable than those of the TPP^+ cation and the TPM neutral molecule.

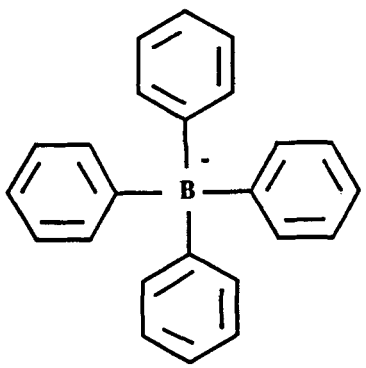
The results of the calculations with the LD/MNDO method and the more rigorous FEP/SCAAS/MNDO method are summarized in Table VII. As seen from the table, both methods give very similar solvation energies. $TEMPOH^+$ has the largest value of ΔE_{sol} among these three ions. This could be attributed to the greater electrostatic stabilization of the central charge, due to the smaller steric hindrance effect. Both methods predict that the TPB^- anion should be more stable than the TPP^+ cation. This could be associated with the much larger charge delocalization in the negative ion, as predicted by the MNDO calculations. The analysis of the contribution to eq. (11) indicates that the stabilization of the peripheral phenyl rings by the solvent dipoles is greater in the anion than in the cation. Apparently, the cation charge is more localized on the central atom than the corresponding anion charge, while the steric effect of the rings prevents the water molecules from approaching the central atom. Thus, the stabilization of the anion by the water molecules is larger than that of the cation. Since the geometries of the TPB^- and TPB^+ ions are similar, the difference in their charge distribution is the crucial factor that determines their relative solvation energies in the LD model. On the other hand, simulations by all-atom solvent models are expected to reflect (at least to



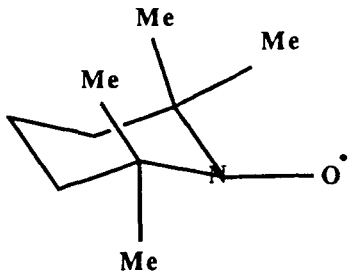
TPM



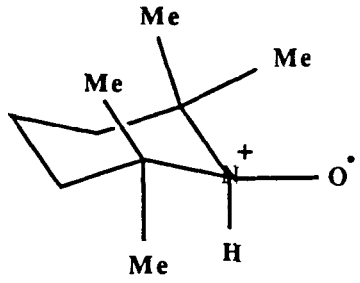
TPP^+



TPB^-



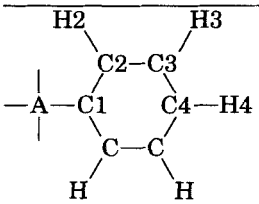
TEMPO



$TEMPOH^+$

some extent) the fact that the water molecules approach the anion center with their hydrogens pointing forward while they approach the cation center with their oxygens pointing forward. These two different cases should involve different steric interactions between the water molecules and the aromatic rings, and therefore different distances between the water molecules and the corresponding charge centers and different electrostatic energies. The existence of this additional steric effect is confirmed by the larger difference between ΔG_{sol} of TPB^- and TPP^+ for the FEP than for the LD/MNDO calculations. Further confirmation of the importance of the steric effect was obtained by FEP simulations of the solvation of TPP^+ and TPB^- where all the charge of each ion was placed at the central atom (the atomic

Table VI. LD/MNDO atomic charges of isolated and solvated hydrophobic ions.[a]



	Charges (e.u.)					
	Isolated			Solvated		
	TPM	TPP^+	TPB^-	TPM	TPP^+	TPB^-
A	0.136	0.583	−0.233	0.136	0.591	−0.243
C1	−0.080	−0.195	0.015	−0.085	−0.198	0.011
C2	−0.041	−0.008	−0.070	−0.051	0.002	−0.092
H2	0.065	0.074	0.054	0.066	0.083	0.049
C3	−0.064	−0.051	−0.079	−0.069	−0.051	−0.087
H3	0.061	0.089	0.032	0.074	0.085	0.052
C4	−0.056	0.002	−0.111	−0.062	−0.013	−0.108
H4	0.059	0.089	0.030	0.074	0.089	0.064

[a]The charges in both the isolated and the solvated states of hydrophobic ions have a symmetrical distribution (within 0.001–0.003 e.u.). The table gives the corresponding averaged values of the atomic charges for the indicated common structural fragment.

Table VII. Calculated solvation energies (in kcal/mol) of hydrophobic ions.[a]

	LD/MNDO					FEP/MD
Ion	$\Delta\Delta E^S$	E_{perm}	E_{ind}	E_{elec}	ΔE_{sol}	ΔG_{sol}
TPB^-	2.68	−38.26	−3.45	−48.37	−43.62	−43.06
TPP^+	0.61	−32.47	−2.54	−43.75	−36.93	−32.43
$TEMPOH^{+b}$	0.07	−43.03	−13.10	−56.06	−48.91	−48.41

[a]The relative solvation energy ΔE_{sol} designates the difference between the E_{sol} of the ion and the corresponding neutral molecule (which is taken in this case as the reference state). The calculations were done with $R_g = R_g = 16$ Å.
[b]The scaling for the reference neutral state of this molecule is taken here as $f_{sc} = 1.0$.

charges for the aromatic rings were taken the same as for TPM). In this case the simulations were done with only four water molecules around the ion, each of them occupying one of the four possible closest positions to the central atom. The corresponding values of ΔG_{sol} for TPB^- and TPP^+ are −10.24 kcal/mol and −5.99 kcal/mol, respectively. These calculations give a clear evidence for the importance of steric effects in the solvation of hydrophobic ions.

It is clear from the above discussion that hydrophobic ions provide an interesting test case which maximizes the relative importance of steric effects. However, even in this extreme case we find (Table VII) that the LD model provides a reasonable approximation.

CONCLUDING REMARKS

This work developed and refined methods for quantum mechanical SCF MO calculations of chemical processes in solutions. These approaches combine microscopic solvent models with semiempirical quantum mechanical methods, incorporating the potential from the solvent into the solute Hamiltonian. The solvent models used include the simplified LD model and the all atom SCAAS model, which is combined with a FEP/MD method. Both methods take into account consistently the permanent and induced polarization of the solvent molecule. These methods (which are closely related to the methods used in our early studies[8,12]) provide practical computational tools for quantum mechanical studies of solvated molecules and are integrated in the program package MOLARIS.[22]

The reliability of the present approach is limited to some extent by the reliability of the quantum mechanical method used. For example, reasonable solvation energies are obtained as long as the solute charges are scaled to have similar magnitude to the charges obtained by *ab initio* method with large basis sets. On the other hand, calculations of charge separation reaction (such as the dissociation of water in aqueous solutions which was examined in our early studies[12a,b]) give less promising results and might require reparameterization of the electron–electron repulsive integrals. In fact, at present we recommend the use of EVB method[14] for such problems.

Regardless of the possible need of reparametrization of the available semiempirical MO models when used in studies of solvated molecules, we have demonstrated here that the LD model gives similar results to those obtained by the FEP approach, even in the stringent test case of hydrophobic ions. In view of this it appears that the LD model provides a simple yet reliable way for quantum mechanical studies of solvated molecules, which is superior to the use of continuum models with an arbitrary cavity radius. Of course, resonable results may be obtained now by descritized continuum models. However, it is instructive to note that such models were introduced in solvation studies much after the LD models, perhaps because of the fact that the fundamental electrostatic concepts were developed much before the emergence of digital computers and were focused on nonnumerical solution of idealized models.

The present approach has been used before in studies of enzymatic reactions[8] and chemical reactions in solutions.[12] Although the effect of the environment on the calculated activation barrier could be reproduced in a qualitative way it was felt that the EVB method[14] should provide more reliable results and could be incorporated with FEP calculations in a more convenient and rigorous way. Nevertheless, the current version of the program ENZYMIX[22] (which is one of the main module of MOLARIS) allows one to use the FEP/AMPAC/MD method in quantum mechanical studies of enzymatic reactions and chemical reactions in solutions. The ENZYMIX program can be conveniently integrated with other quantum mechanical programs, including *ab initio* packages, using the approach of eq. (22). It is important to note, however, that the results of the calculations cannot be more reliable than the particular quantum mechanical method used. Nevertheless, one may exploit the SCF methods in enhancing the reliability of the EVB method. That is, the EVB method is designed to guarantee the correct energetics of the reactant and product states (which might not be predicted in a quantitative way by some SCF methods). However, the absolute energy at the transition state depends on the off-diagonal matrix elements whose calibration is not always straightforward. Although these matrix elements can be

refined by fitting gas phase EVB surface to the corresponding *ab initio* results (see reference 14b) the fitting procedure is not sufficiently general and might not provide sufficient information about the off-diagonal terms along the reaction coordinate in solutions. With the current approach we can obtain the EVB and the corresponding *ab initio* potential energies *in solution* along a well defined reaction coordinate and minimized the difference between these energies by refining the EVB off-diagonal matrix elements (while correcting, when needed, the *ab initio* energies at the reactant and product state). Using such an approach in an iterative way may provide reliable potential surfaces for studies of chemical reactions in solutions and in enzymes.

This work was supported by NIH grant GM-24492 and ONR grant number N00014-91-J-1318.

APPENDIX

Considering the widely held belief that the RF models do provide a reasonable approximation, it is important to clarify the fundamental problems associated with the predictive power of these models. To address this issue in a quantitative way, it is useful to recognize that the solvation energy of an ion pair is given (in kcal/mol) by the quite reliable approximation[3,5]

$$\Delta G_{sol}(R) = \Delta G^{\infty}_{sol} - \sum_{i>j} 332(Q_iQ_j/R_{ij})\left(1 - \frac{1}{\varepsilon}\right) \quad (23)$$

where ΔG^{∞} is the solvation energy of the isolated ions at infinite separation and the last term represents the charge–charge interaction in vacuum with Q in au and R in Å. The solvation energy of the isolated ions can be expressed by a Born-type formula

$$\Delta G^{\infty}_{sol} = -\sum_{i} 166(Q_i^2/\bar{a}_i)(1 - 1/\varepsilon) \quad (24)$$

where $\bar{a}_i$ is the effective radius that reproduces the observed solvation free energy of the given ion. Assuming that $\Delta G_{sol}(R)$ can be expressed by the Onsager's RF formula, one can write

$$\Delta G_{sol}(R) = -166(\mu^2/b^3)(2\varepsilon - 2)/(2\varepsilon + 1) \quad (25)$$

where ΔG is given in kcal/mol, b is the cavity radius (in Å, and μ is given in atomic charge units and Å. Applying eqs. (23), (24), and (25) to ion pairs of equal radius in aqueous solution where $\varepsilon >> 1$, one obtains

$$b^3 \simeq R^2\bar{a}/[2(1 - \bar{a}/R)] \quad (26)$$

This relationship has not been used in previous RF quantum mechanical studies nor does it follow the cavity radius that would be usually deduced from the size of the given molecule. For example, let us consider the amino acids of Table IV of reference 12b. For this system we obtain[12b] solvation energies of -52, -71, and -81 kcal/mol for R values of 3.4, 3.7, and 5.6 Å, respectively. These values would be reproduced by eq. (26) with cavity radii of 3.4, 3.6, and 4.0 Å, which are not directly related to the size of the corresponding systems. Furthermore, similar solvation energies would be obtained for other molecules with the same charge separation distance, and with different overall volume as long as the charges are immersed in water.[3] It is instructive to note in this respect that even the recent impressive study of reference 9j produced solvation energy of about -9 kcal/mol for the most stable tautomer of cytozine, while the experimental estimate[31] and microscopic model give a significantly larger solvation energy ($\Delta G_{sol} < -20$ kcal/mol).

References

1. W.J. Hehre, L. Radom, P.v.R. Schleyer, and J.A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, New York, 1988.
2. a. G.A. Segal, Ed., *Semiempirical Methods of Electronic Structure Calculation*, Plenum Press, New York, 1980. b. M.J.S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99,** 4899 (1977). c. M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, and J.J.P. Stewart, *J. Am. Chem. Soc.*, **107,** 3902 (1985).
3. A. Warshel and S. Russel, *Quart. Rev. Biophys.*, **17,** 283 (1984).
4. L. Salem, *Electrons in Chemical Reactions: First Principles*, John Wiley & Sons, New York, 1982.
5. a. A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, John Wiley & Sons (1991). b. A. Warshel, J. Aqvist, and S. Creighton *Proc. Natl. Acad. Sci. USA*, **86,** 5820 (1989).
6. B.J. Gertner, K.R. Wilson, and J.T. Hynes, *J. Chem. Phys.*, **90,** 353, 3537 (1989).
7. a. A. Pullman in *The New World of Quantum Chemistry, Proceedings of the Second International Congress of Quantum Chemistry*, Pullman B., Parr R., Eds., Reidel, Dodrecht, 1976, p. 149. b. K. Kitaura and K. Morokuma, *Int. J. Quant. Chem.*, **10,** 325 (1976). c. A. Pullman, *Quantum Theory of Chemical Reactions*, vol. 2, R. Daudel, A. Pullman, L. Salem, A. Veillard, Eds., Reidel, Dodrecht, 1981, p. 1.
8. A. Warshel and M. Levitt, *J. Mol. Biol.*, **103,** 227 (1976).
9. a. G. Klopman, *Chem. Phys. Lett.*, **1,** 200 (1967). b. J.H. McCreery, R.E. Cristoffersen, and G.G. Hall, *J. Am. Chem. Soc.*, **98,** 7191, 7198 (1976). c. R. Constanciel and O. Tapia, *Theoret. Chim. Acta*, **48,** 75 (1978). d. J.E. Sanhueza and O. Tapia, *J. Chem. Phys.*, **70,** 3096 (1979). e. O. Tapia and B. Silvi, *J. Phys. Chem.*, **84,** 2646 (1980). f. S. Miertus, E. Scrocco, and J. Tomasi, *Chem. Phys.*, **55,** 117 (1981). g. B.T. Thole and P.T. van Duijnen, *Chemical Physics*, **71,** 211 (1982). h. O. Tapia and O. Goscinski, *Mol. Phys.*, **29,** 1653 (1975). i. M.M. Karelson, A.R. Katritzky, M. Szafran, and M. Zerner, *J. Org. Chem.*, **54,** 6030 (1989). j. A.R. Katritzky and M.M. Karelson, *J. Am. Chem. Soc.*, **113,** 1561 (1991).
10. a. L. Onsager, *J. Am. Chem. Soc.*, **58,** 1486 (1936). b. J.G. Kirkwood, *J. Chem. Phys.*, **2,** 351 (1934). c. J.G. Kirkwood, F.H. Westheimer, *J. Chem. Phys.*, **6,** 506 (1938).

11. A.A. Rashin, *J. Phys. Chem.*, **94,** 1725 (1990).
12. a. A. Warshel, *Chem. Phys. Lett.*, **55,** 454 (1978). b. A. Warshel, *J. Phys. Chem.*, **83,** 1640 (1979). c. A. Warshel and A. Lappicirella, *J. Am. Chem. Soc.*, **103,** 4664 (1981).
13. a. S.J. Weiner, U.C. Singh, and P.A. Kollman, *J. Am. Chem. Soc.*, **107,** 2219, 1985. b. J.A.C. Rullmann, M.N. Bellido, and P.T. van Duijnen, *J. Mol. Biol.*, **206,** 101 (1989). c. M.J. Field, P.A. Bash, and M. Karplus, *J. Comp. Chem.*, **11,** 700 (1990).
14. a. A. Warshel and R.M. Weiss, *J. Am. Chem. Soc.*, **102,** 6218 (1980). b. J.K. Hwang, G. King, S. Creighton, and A. Warshel, *J. Am. Chem. Soc.*, **110,** 5297 (1988). c. A. Warshel, F. Sussman, and J.K. Hwang, *J. Mol. Biol.*, **201,** 139 (1988).
15. S.T. Russell and A. Warshel, *J. Mol. Biol.*, **185,** 389 (1985).
16. a. A. Warshel and G. King, *Chem. Phys. Lett.*, **121,** 124 (1985). b. G. King and A. Warshel, *J. Chem. Phys.*, **91,** 3647 (1989).
17. a. D.L. Liotard, E.F. Healy, J.M. Ruiz, and M.J.S. Dewar, AMPAC (Quantum Chemistry Program Exchange 506), *QCPE Bull.*, **9,** 123 (1989). b. J.J.P. Stewart, MOPAC (Quantum Chemistry Program Exchange 455), *QCPE Bull.*, **10,** 86 (1990).
18. B. Lee and F.M. Richards, *J. Mol. Biol.*, **55,** 379 (1971).
19. J.L. Pascual-Ahuir and E. Silla, *J. Comp. Chem.*, **11,** 1047 (1990).
20. R.C. Weast, Ed., *CRC Handbook of Chemistry and Physics*, The Chemical Rubber Co., Ohio, 1989.
21. a. A. Warshel, *J. Phys. Chem.*, **86,** 2218 (1982). b. A. Warshel, *Pontif. Acad. Sci. Scr. Varia*, **55,** 59 (1984). c. J.P.M. Postma, H.J.C. Berendsen, and J.R. Haak, *Faraday Symp. Chem. Soc.*, **17,** 55 (1982). d. C.F. Wong and J.A. McCammon, *J. Am. Chem. Soc.*, **108,** 3830 (1986). e. S.N. Rao, U.C. Singh, P.A. Bash, and P.A. Kollman, *Nature*, **328,** 551 (1987).
22. A. Warshel and S. Creighton, in *Computer Simulation of Biomolecular Systems*, W.F. van Gunsteren, P.K. Weiner, Eds., p. 120. Leiden: ESCOM, 1989.
23. a. D.H. Aue, H.M. Webb, and M.T. Bowers, *J. Am. Chem. Soc.*, **98,** 318 (1976). b. C.E. Klots, *J. Phys. Chem.*, **85,** 3585 (1981).
24. A. Ben-Naim and Y. Marcus, *J. Chem. Phys.*, **81,** 2016 (1984).
25. A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, **71,** 525, (1971).
26. P. Lauger, R. Benz, G. Stark, E. Bamberg, P.C. Jordan, A. Fahr, and W. Brock, *Quart. Rev. Biophys.* **14,** 513 (1981).
27. L.J. Berliner, Ed. *Spin Labelling. Theory and Applications*, vol. 2, Academic Press, New York, 1976.
28. C. Tanford, *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd edition, John Wiley & Sons, New York, 1980.
29. a. Y.C. Martin, *Quantitative Drug Design*, Marcel Dekker Inc., New York, 1978. b. A.J. Stuper, W.E. Brugger, and P.C. Jurs, *Computer Assisted Studies of Chemical Structure and Biological Function*, John Wiley and Sons, New York, 1989.
30. R. Aveyard and R.W. Mitchell, *Trans. Far. Soc.*, **65,** 2645 (1969).
31. a. M.V. Kilday, *J. Res. Natl. Bur. Std.*, **83,** 539 (1978). b. I.K. Yauson, A.B. Teplitsky, and F. Sukhodub, *Biopolymers*, **18,** 1149 (1979).