Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution

WILLIAM L. JORGENSEN

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

Received January 3, 1989

Introduction

1988 marked the 25th anniversary of the publication of two classic papers in the evolution of theoretical organic chemistry, Roald Hoffmann's report of extended Hückel theory[1] and the ab initio calculations of Pitzer and Lipscomb on the rotational barrier in ethane.[2] Although one can always point to earlier, more fundamental work, the applications in these papers specifically alerted organic chemists to the potential of molecular quantum mechanics extending beyond π-electron systems. The rapid maturation of this field as embodied in powerful research tools such as Pople's GAUSSIAN/86 program has been remarkable.[3] It is now apparent that the last five years have been another revolutionary period during which theoretical organic chemistry has been extended beyond the gas-phase realm of quantum mechanics to the quantitatively meaningful study of structure and reactivity in solution. This event has been accompanied by parallel developments in theoretical biochemistry which feature the same condensed-phase simulation procedures, Monte Carlo statistical mechanics and molecular dynamics, as the primary computational methods.[4,5] The importance of extending the theoretical studies to fluids is evident from the profound role of solvent effects in organic chemistry and because the liquid state is the predominant venue for organic chemistry as well as living systems.

Methodology

For applications to organic systems such as those described below, Monte Carlo (MC)[6] or molecular dynamics (MD)[5,7] simulations are typically carried out for one or two solutes in a cube or rectangular solid with 200–400 solvent molecules. Periodic boundary conditions are used whereby the central cell is surrounded by images of itself in order to avoid boundary effects and to allow representation of the bulk fluid by a tractable number of molecules. Equilibrium properties are obtained in MC calculations by averaging over millions of instantaneous geometrical configurations of the system that are selected by the Metropolis algorithm,[8] and in MD calculations by solving the Newtonian equations of motion and performing time averages. The simulations often invoke the isothermal–isobaric ensemble to fix the temperature and pressure, e.g., at 25 °C and 1 atm for the results below.

Another element at the heart of the simulations is the selection of intermolecular potential functions that describe the interactions between the components of the system. Typically, interaction sites are located on the nuclei and the nonbonded interactions then consist of Coulombic and Lennard–Jones terms between the sites. The total potential energy is taken as the sum of the individual two-body interactions, so polarization is not included except in an average sense in the parameterization. Bond lengths and bond angles are often fixed; however, torsional motion is normally considered and requires corresponding potential functions that are represented as Fourier series. Potential functions are now available for hydrocarbons, many organic functional groups, and all common protein residues. For the OPLS potentials, the charges and Lennard–Jones parameters have been optimized to yield excellent thermodynamic and structural results for ca. 40 pure organic liquids and water.[9] Derivation of parameters for ions is more problematic; a viable approach is to fit to results of ab initio molecular orbital calculations for ion–solvent molecule complexes.[9,10]

The key thermodynamic quantity that is needed to characterize equilibria is the free energy difference, ΔG, between reactants and products. Fortunately, procedures have emerged for the computation of free energy differences with excellent precision,[11] and they have now been applied successfully to many fundamental problems in organic chemistry. The most straightforward and limited procedure is *direct sampling* of a reaction coordinate (r_c) or other transformation index such as a dihedral angle for a conformational equilibrium. The frequency of occurrence of different values of r_c during the simulation can be accumulated in the distribution function $g(r_c)$, which is simply related to the relative free energy or "potential of mean force" (pmf), $w(r_c)$, by eq 1.

$$w(r_c) = -k_B T \ln g(r_c) + \text{constant} \qquad (1)$$

However, the range of r_c that can be reasonably sampled in one simulation may be limited particularly by

Dr. Jorgensen is a native of New York City and a graduate of Princeton (A.B., 1970) and Harvard (Ph.D., 1975). He has been on the faculty at Purdue University since 1975, where he is currently the H. C. Brown Professor of Chemistry. In December 1988, Dr. Jorgensen enjoyed the hospitality of the Nobel Foundation at Snogeholms Slott, where this manuscript was written.

(1) Hoffmann, R. *J. Chem. Phys.* **1963**, *39*, 1397.
(2) Pitzer, R. M.; Lipscomb, W. N. *J. Chem. Phys.* **1963**, *39*, 1995.
(3) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
(4) Beveridge, D. L., Jorgensen, W. L., Eds. *Ann. N.Y. Acad. Sci.* **1986**, *482*.
(5) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University: Cambridge, 1987.
(6) For a review, see: Jorgensen, W. L. *J. Phys. Chem.* **1983**, *87*, 5304.
(7) McCammon, J. A.; Karplus, M. *Acc. Chem. Res.* **1983**, *16*, 187.
(8) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(9) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(10) (a) Gao, J.; Garner, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1986**, *108*, 4784. (b) Jorgensen, W. L.; Briggs, J. M.; Gao, J. *J. Am. Chem. Soc.* **1987**, *109*, 6857. (c) Jorgensen, W. L.; Briggs, J. M. *J. Am. Chem. Soc.*, in press.
(11) For a review, see: Mezei, M.; Beveridge, D. L. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 1.

0001-4842/89/0122-0184$01.50/0

barriers in $w(r_c)$. *Umbrella sampling* can sometimes provide a solution by using an artificial biasing function, $u(r_c)$, which is added to the potential energy.[12] It can constrain the simulation to sample a particular range of r_c or flatten the problematic barriers. The true distribution function can subsequently be recovered from the biased one, $g'(r_c)$, via eq 2 where c is a normalization constant. If an adequate range of r_c is still

$$g(r_c) = cg'(r_c) \exp(u(r_c)/k_BT) \quad (2)$$

not covered, *importance sampling* can be used; i.e., a series of simulations is run with biasing functions that center the sampling in different, overlapping regions of r_c (windows).[13] The resultant individual $g(r_c)$ values are then spliced together to yield the overall $g(r_c)$. This procedure has proven effective in a number of applications including the determination of potentials of mean force for prototypical S_N2 and addition reactions.[14-16] The principal difficulties are the choice of biasing functions and verification of complete sampling for each window.

Over the last five years, an alternative based on statistical perturbation theory (SPT)[17] has become popular. The procedure follows from eq 3, which expresses

$$G_j - G_i = -k_BT \ln \langle \exp[-(H_j - H_i)/k_BT] \rangle_i \quad (3)$$

the free energy difference between systems i and j by an average of a function of their energy difference. The averaging is for sampling based on system i, so system j is treated as a perturbation. If the systems are too different, convergence of eq 3 will be slow. However, multiple simulations may be used to connect i and j through intermediate points using a coupling parameter, λ (eq 4). Features ξ of the systems, including

$$\xi(\lambda) = \xi_i + \lambda(\xi_j - \xi_i) \quad (4)$$

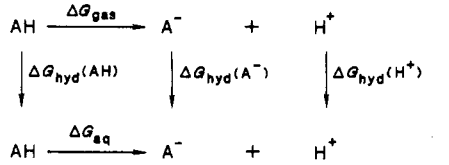
geometry and potential functions, can then be interconverted as λ goes from 0 to 1. Typically, the perturbations are made to mutate one solute to another or to step along a reaction coordinate. The advantages over umbrella sampling are the complete control in choosing the sampled regime via the λ values and the lack of need for biasing functions. However, the choice of λ values for optimal convergence of the incremental ΔG's requires some testing.

Relative Free Energies of Solvation

These procedures allow the computation of the difference in free energies of solvation for two solutes. This was first done for organic solutes (ethane and methanol) in water by using SPT and Monte Carlo simulations, as reported in 1985.[18] Internal degrees of freedom in the solutes were ignored,[19] and since two-body potentials were used, $H_j - H_i$ in eq 3 is then just the difference in total solute–solvent energies. Thus, the two solutes only had to be interconverted in water as λ went from 0 (ethane) to 1 (methanol) and the free energy changes were accumulated to yield $\Delta\Delta G_{hyd}$. The calculated result of 6.8 ± 0.2 kcal/mol agreed well with the experimental value of 6.9 kcal/mol, which supported the utility of the OPLS potential functions[9] including the TIP4P model of water.[20] More importantly, the statistical uncertainty of only ±0.2 kcal/mol, which was estimated by running the mutation in both directions, demonstrated the high precision of the methodology and stimulated much enthusiasm. The procedure was probably not studied in depth previously because three to four separate simulations are needed to span between such a pair of solutes. In 1985, this required the equivalent of four weeks on a Vax 11/780 computer. Now, with hardware that is more than a factor of 10 less expensive, the same calculation can be run in three days.

Once relative free energies of solvation are available, thermodynamic cycles can be set up to obtain free energy changes for many important processes. Some examples are provided in the following.

Relative pK_a Values. Consideration of the thermodynamic cycle below yields eq 5 for the pK_a of AH in water. Since it is easier to compute differences in



$$2.3RT\ pK_a(AH) = \Delta G_{aq} = \Delta G_{gas} + \Delta G_{hyd}(A^-) + \Delta G_{hyd}(H^+) - \Delta G_{hyd}(AH) \quad (5)$$

free energies of solvation than the absolute values (vide infra), an analogous cycle for a second acid BH can be written down, and the difference in pK_a values is expressed in eq 6. Specifically, the difference in pK_a

$$2.3RT[pK_a(BH) - pK_a(AH)] = \Delta\Delta G_{gas}(BH\text{–}AH) + \Delta\Delta G_{hyd}(B^-\text{–}A^-) - \Delta\Delta G_{hyd}(BH\text{–}AH) \quad (6)$$

values is given by the difference in gas-phase acidities plus the difference in free energies of hydration for the anions and acids. The last two quantities are obtainable from fluid simulations in which A^- and AH are gradually mutated to B^- and BH. Furthermore, experimental data or ab initio molecular orbital theory can be used as the source of $\Delta\Delta G_{gas}$.

Such calculations have been carried out for the acidities of acetonitrile and ethane relative to methanethiol and of methylamine and ethane relative to methanol.[10b,c] Monte Carlo simulations were used for the mutations in water, and high-level ab initio calculations provided $\Delta\Delta G_{gas}$ values in good accord with experimental data. The resultant pK_a values were 28 for acetonitrile, 33 for methylamine, and 50 and 54 for ethane.[10b,c] These values are consistent with the ranges for experimental estimates of the pK_a values. In view of the difficulties in obtaining experimental data for such weak acids in water and the associated uncertainties, the theoretical approach is a valuable alter-

(12) Patey, G. N.; Valleau, J. P. *J. Chem. Phys.* **1975**, *63*, 2334.
(13) Pangali, C. S.; Rao, M.; Berne, B. J. *J. Chem. Phys.* **1979**, *71*, 2975.
(14) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 3049. Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154. Chandrasekhar, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 2974.
(15) Madura, J. D.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1986**, *108*, 2517.
(16) Jorgensen, W. L. *Adv. Chem. Phys., Part II* **1988**, *70*, 469.
(17) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.
(18) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.
(19) Including internal degrees of freedom is not problematic; it simply requires simulations for the interconversion both in solution and in the gas phase.
(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

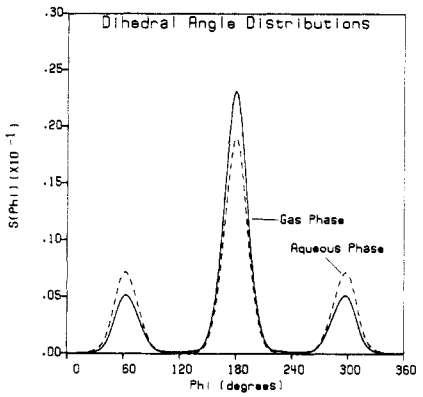


Figure 1. Population distributions for the dihedral angle about the central CC bond of butane in the gas phase and in aqueous solution at 25 °C.

native. Moreover, the simulations provide extensive structural characterization of the hydration of the anions and neutral acids.[10c]

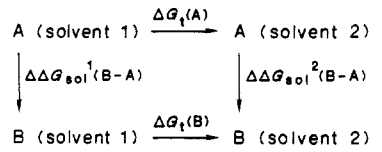
Conformation in Solution. Solvent effects on conformational equilibria can be examined by computing the change in free energy of solvation as one conformer is converted to another.[6] In the simplest case of rotation about one bond, the dihedral angle Φ is the reaction coordinate. The gas-phase population distribution, $s°(\Phi)$, can then be obtained from a Boltzmann distribution for the torsional potential, $V(\Phi)$. It is related to the solution distribution function, $s(\Phi)$, by eq 7 where ΔG_{sol} is the change in free energy of solvation.

$$s(\Phi) = cs°(\Phi) \exp(-\Delta G_{sol}(\Phi)/k_B T) \quad (7)$$

Umbrella sampling, importance sampling, and SPT have all been used to study various systems. For example, the hydrophobic effect on the trans–gauche equilibrium for butane in water has been examined several times.[6,21-24] The results have consistently revealed a 10–20% increase in the gauche population at 25 °C upon transfer from the gas phase to water. The most recent and extensive simulations were of the MC–SPT type and yielded a 12% shift as illustrated in Figure 1.[24] The qualitative result is consistent with more approximate theoretical models[25] and with fundamental notions about hydrophobic effects on biomolecular structure.[26]

In a more recent application, the MC–SPT method was used to model the interconversion of the trans and cis conformers of *N*-methylacetamide in water.[27] No solvent effect was found in this case. Thus, the extreme rarity of cis peptide bonds in proteins (except adjacent to Pro residues) is an intrinsic, gas-phase effect and not environmentally enhanced. In another study, importance sampling was applied to obtain the effect of hydration on the low-energy conformers of the model dipeptide *N*-acetylalanyl-*N*-methylamide.[28a] As expected, the internally hydrogen bonded C_7 form becomes less favored in water than in the gas phase in comparison to the more extended conformers. In a similar study, hydration has been found to reinforce the intrinsic preference for the gauche–gauche conformer of dimethyl phosphate anion.[28b]

Other Applications. Many additional uses of relative free energies of solvation can be envisaged. A recent application has been to compute differences in partition coefficients.[29] In the thermodynamic cycle below, the partition coefficient for A is just the ratio of the concentrations of A in solvent 2 and solvent 1 at equilibrium (eq 8). The cycle then yields the difference



$$\Delta G_t(A) = -RT \ln ([A]_2/[A]_1) = -2.3RT \log P_A \quad (8)$$

in partition coefficients for A and B in terms of the difference in their free energies of solvation in the two solvents (eq 9). As a simple example, MC–SPT sim-

$$\Delta\Delta G_{sol}^2(B-A) - \Delta\Delta G_{sol}^1(B-A) = \Delta G_t(B) - \Delta G_t(A) = 2.3RT(\log P_A - \log P_B) \quad (9)$$

ulations were used to convert methanol (A) to methylamine (B) in chloroform (solvent 2) and water (solvent 1).[29] The corresponding $\Delta\Delta G_{sol}$ values were 0.30 ± 0.03 and 1.3 ± 0.1 kcal/mol, which translate to a $\Delta \log P$ of −0.7. The experimental result is ca. −0.5,[30] so there is qualitative accord that the methanol to methylamine conversion results in higher affinity for the organic phase. The $\Delta\Delta G_{sol}$ components also clearly show that the preference is dominated by the lower free energy of hydration for methanol than methylamine. Similar calculations for CH_3COOH and CH_3COOCH_3 yielded a $\Delta \log P$ of −4.0, while the experimental result is ca. −3. log *P* data are often a key parameter in structure/activity studies for compounds of pharmacological interest.[31] The perturbation theory approach is well-suited to making predictions on the effects of substituent changes on partitioning and can, therefore, provide an alternative, sophisticated source of log *P* estimates.[31]

In another study, MD–SPT calculations were applied to predict changes in reduction potentials for quinones.[32] The appropriate thermodynamic cycle was constructed to compare the reductions of 1,4- and 1,2-benzoquinone to the hydrobenzoquinones. The intrinsic preference for reduction of 1,2-benzoquinone was largely negated by hydration. Such calculations clearly have potential for electrochemical applications as well as in gauging bioreductive activity.[32]

Free energy methods have also been used to study the effects of hydration on tautomeric equilibria including **1** ⇌ **2**.[33] Consistent with experimental findings, the small gas-phase preference for the hydroxy tautomer

(21) Rosenberg, R. O.; Mikkilineni, R.; Berne, B. J. *J. Am. Chem. Soc.* **1982**, *104*, 7647.
(22) Jorgensen, W. L. *J. Chem. Phys.* **1982**, *77*, 5757.
(23) Jorgensen, W. L.; Gao, J.; Ravimohan, C. *J. Phys. Chem.* **1985**, *89*, 3470.
(24) Jorgensen, W. L.; Buckner, J. K. *J. Phys. Chem.* **1987**, *91*, 6083.
(25) Pratt, L. R.; Chandler, D. *J. Chem. Phys.* **1977**, *67*, 3683.
(26) Ben-Naim, A. *Hydrophobic Interactions*; Wiley: New York, 1980.
(27) Jorgensen, W. L.; Gao, J. *J. Am. Chem. Soc.* **1988**, *110*, 4212.
(28) (a) Mezei, M.; Mehrotra, P. K.; Beveridge, D. L. *J. Am. Chem. Soc.* **1985**, *107*, 2239. (b) Jayaram, B.; Mezei, M.; Beveridge, D. L. *J. Am. Chem. Soc.* **1988**, *110*, 1691.
(29) Jorgensen, W. L.; Briggs, J. M., to be published.
(30) Leo, A. C.; Hansch, C.; Elkins, D. *Chem Rev.* **1971**, *71*, 525.
(31) Dunn, W. J., III, Block, J. H., Pearlman, R. S., Eds. *Partition Coefficient Determination and Estimation*; Pergamon: New York, 1986.
(32) Reynolds, C. A.; King, P. M.; Richards, W. G. *Nature* **1988**, *334*, 80.
(33) Cieplak, P.; Bash, P.; Singh, U. C.; Kollman, P. A. *J. Am. Chem. Soc.* **1987**, *109*, 6283.

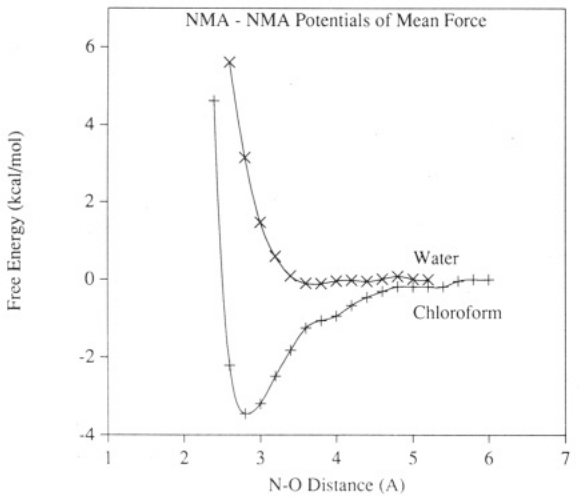


Figure 2. Potentials of mean force for the separation of two NMA molecules in aqueous solution and in chloroform at 25 °C.

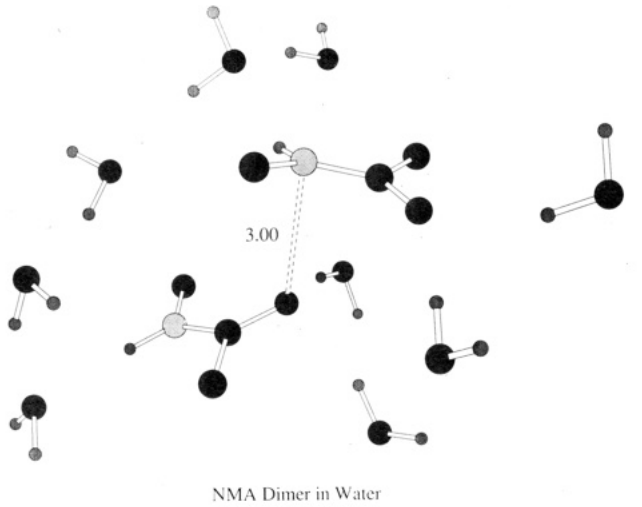


Figure 3. A configuration from the Monte Carlo simulation of two NMA molecules in water at an N–O separation of 3.0 Å. Some of the nearest water molecules are shown.

is computed to be dramatically reversed to a ca. 4 kcal/mol preference for the keto form in water.

1 2

Binding and Molecular Recognition

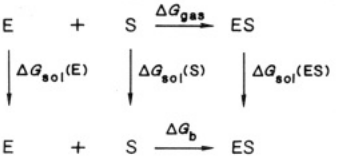
The free energy of host–guest binding is the fundamental thermodynamic index of molecular recognition. The quantity is accessible through simulations along with detailed structural characterization of the complexes. Absolute free energies of binding (ΔG_b) are more time-consuming to compute than relative values ($\Delta\Delta G_b$); however, they have been obtained in several ways. First, a potential of mean force can be constructed as a function of the intersolute separation via importance sampling of SPT. Computation of the association constant K_a then involves integration of the pmf to a cutoff limit c that defines association (eq 10), and ΔG_b is given by $-k_BT \ln K_a$.

$$K_a = 4\pi \int_0^c r^2 \exp(-w(r)/k_BT)\, dr \qquad (10)$$

The importance sampling approach was initially applied to two Lennard–Jones particles[13] and two benzene molecules in water.[34] Contact free energy minima were found, as anticipated from the hydrophobic effect, along with intriguing solvent-separated minima. Subsequently, importance sampling was used to compute pmf's for Na^+Cl^- and Cl^-Cl^- ion pairs in water.[35,36] The first applications of SPT along these lines were for the $(CH_3)_3C^+Cl^-$, $CH_3NH_3^+CH_3COO^-$, $(CH_3)_4N^+Cl^-$, $(CH_3)_4N^+(CH_3)_4N^+$, and Cl^-Cl^- ion pairs in water.[37-39] In general, contact and solvent-separated minima are found for the unlike-charged pairs, while the approach of two $(CH_3)_4N^+$ ions is purely repulsive. The independent studies of Cl^-Cl^- both revealed a controversial minimum near contact owing to the presence of several bridging water molecules.[36,39]

Some results of recent MC–SPT calculations for a particularly biochemically relevant example are shown in Figures 2 and 3.[40] The *N*-methylacetamide (NMA) dimer was found to form a hydrogen-bonded complex in chloroform; the depth of the free energy well is −3.5 kcal/mol at an N–O separation of 2.8 Å. In contrast, there is no attraction for the amides in water. As illustrated in Figure 3, at short separations the amides do not form a hydrogen bond; rather, they stack with a favorable dipole alignment that also exposes their edges for hydrogen bonding with water. However, the amide–amide interaction at 3.0 Å averages only −3.5 kcal/mol, which is not enough to overcome the more optimal hydration that occurs at larger separations. For comparison, the average NMA–NMA interaction at 3.0 Å in chloroform is −7.5 kcal/mol. The lack of amide self-affinity in water has also been found by spectroscopic and calorimetric measurements; the reported K_a values for NMA in chloroform and water are ca. 3 and 0.005 M^{-1},[41] while the computed values from the pmf's are 4 and 0.04 M^{-1} with a 4-Å cutoff.[40]

Another route to the absolute free energy of binding follows from the cycle below. Knowledge of ΔG_{gas} and the free energies of solvation of the reactants and complex are required. ΔG_{gas} can come from gas-phase



calculations, while the ΔG_{sol} values correspond to the differences in free energies from simulations in which the solutes are grown into the gas phase and into the solution.[42] This procedure was used successfully by

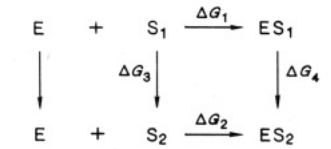
(34) Ravishanker, G.; Mezei, M.; Beveridge, D. L. *Faraday Symp. Chem. Soc.* **1982**, *17*, 79. Ravishanker, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1985**, *107*, 2565.
(35) Berkowitz, M.; Karim, O. A.; McCammon, J. A.; Rossky, P. J. *Chem. Phys. Lett.* **1984**, *105*, 577.
(36) Dang, L. X.; Pettitt, B. M. *J. Am. Chem. Soc.* **1987**, *109*, 5531.
(37) Jorgensen, W. L.; Buckner, J. K.; Huston, S. E.; Rossky, P. J. *J. Am. Chem. Soc.* **1987**, *109*, 1891.
(38) Jorgensen, W. L.; Buckner, J. K.; Gao, J. In *Chemical Reactivity in Liquids*; Moreau, M., Turq, P., Eds.; Plenum: New York, 1988.
(39) Buckner, J. K.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1989**, *111*, 2507.
(40) Jorgensen, W. L. *J. Am. Chem. Soc.*, in press.
(41) Kresheck, G.; Klotz, J. M. *Biochemistry* **1969**, *8*, 8. Krikorian, S. E. *J. Phys. Chem.* **1982**, *86*, 1875.
(42) Jorgensen, W. L.; Blake, J. F.; Buckner, J. K. *Chem. Phys.* **1989**, *129*, 193.

Cieplak and Kollman to obtain ΔG_b for the nucleic acid base pairs, A-T and G-C, in aqueous solution from MD–SPT calculations.[43] A more efficient alternative was subsequently noted[44] based on the following relationships. All that is needed is two simulations in

$$\begin{array}{ll} S \rightarrow 0 & \Delta G_{sol}(S \rightarrow 0) \\ E \rightarrow ES & -\Delta G_{sol}(ES \rightarrow E) \\ \hline E + S \rightarrow ES & \Delta G_b \end{array}$$

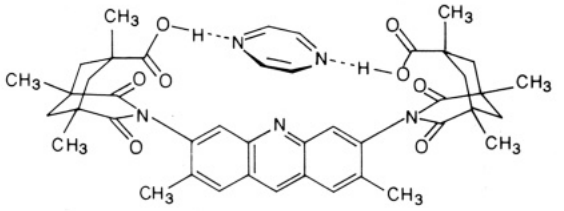
solution, one in which S disappears by itself and one in which it disappears from the complex. However, under any circumstances, simulations in which a solute is made to vanish are arduous since many steps or windows are needed with SPT or importance sampling.[42-44]

Relative free energies of binding can be obtained by mutating S_1 to S_2 and ES_1 to ES_2 since $\Delta\Delta G_b = \Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4$. The first computation of this type



was the MD–SPT study by McCammon and co-workers for the binding of Cl^- and Br^- to the macrotricyclic receptor SC24 in water.[45] Subsequently, several very impressive calculations have been carried out for substrates binding to trypsin, thermolysin, and subtilisin in aqueous solution.[46-48] The accord with experimental binding data has been excellent, some predictions have been made and confirmed, and the calculations have provided structural insights and results on the components that make up $\Delta\Delta G_b$. An important theoretical tool has emerged that complements experimental work on site-specific mutagenesis and that has obvious utility in substrate and inhibitor design.

This type of calculation has now been extended to a neutral organic host–guest system.[49] Additional novelty came from use of a nonaqueous solvent and MC–SPT simulations. The difference in free energy of binding for pyrazine and pyridine with Rebek's acridine diacid was investigated in chloroform. Experimentally, pyrazine is observed to be more strongly bound by 1.45 kcal/mol owing to the postulated "two-point binding" illustrated below.[50]



The SPT calculations yielded a preference for pyrazine

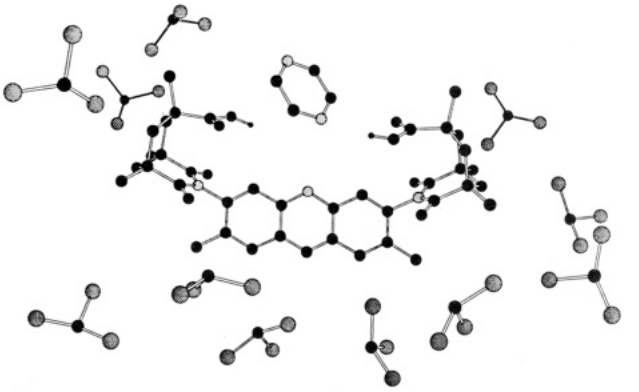


Figure 4. A typical structure for the complex of Rebek's acridine diacid with pyrazine in chloroform. Some of the nearby chloroform molecules are shown with hydrogens implicit.

of 1.2 ± 0.1 kcal/mol with the acid groups syn or anti; however, the computed structures for the complexes did not reveal the "two-point binding". Instead, the complexes are unsymmetrical with one strong hydrogen bond between a pyrazine nitrogen and an acid group, but with only weak interaction for the other nitrogen (Figure 4). The conclusion that the binding cleft is too small to accommodate pyrazine optimally was supported by simulations for a model diacid host which demonstrated "two-point binding" and gave a preference of 3.6 ± 0.1 kcal/mol for pyrazine over pyridine. These studies nicely illustrate the potential of joint experimental and computational investigations of molecular recognition and design.

Reactions in Solution

The greatest technical challenges are in modeling reactions in solution that involve changes in covalency. Ultimately, what is desired thermodynamically is complete characterization of the free energy surface in solution with its implicit mechanistic details. Key problems are that the energy surface must be known in the absence of solvent, the variations in the potential functions for the reactants, e.g. the charges, must be known all over the surface, and multidimensional potentials of mean force must be constructed. Initial progress has been made in our group for some prototypical organic reactions, as reviewed in detail elsewhere.[16] The approach so far has been to (1) determine the minimum energy reaction path (MERP) in the gas phase via ab initio calculations, (2) obtain the solute–solvent potential functions as a function of the reaction coordinate along the MERP by fitting to energetic and structural results of ab initio calculations on complexes of the reacting system with a water molecule in many orientations, and (3) obtain the corresponding one-dimensional pmf in solution from a series of fluid simulations with importance sampling or SPT.

Briefly, the degenerate S_N2 reaction of $Cl^- + CH_3Cl$ was studied first due, in part, to its simplicity and to the profound solvent effects that are observed for such processes.[14] The gas-phase MERP was obtained from ab initio calculations, and the effects of solvation in water and dimethylformamide (DMF) were evaluated by MC simulations with importance sampling. As illustrated in Figure 5, the gas-phase result has the characteristic double-well form, while hydration causes the reaction surface to become almost unimodal and increases the barrier significantly. The reaction profile

(43) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 3734.
(44) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742.
(45) Lybrand, T. P.; McCammon, J. A.; Wipff, G. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 833.
(46) Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 3830.
(47) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1987**, *235*, 574.
(48) Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A. *Nature* **1987**, *328*, 551.
(49) Jorgensen, W. L.; Nguyen, T. B.; Boudon, S. *J. Am. Chem. Soc.* **1989**, *111*, 755.
(50) Rebek, J., Jr.; Askew, B.; Killoran, M.; Nemeth, D.; Lin, F.-T. *J. Am. Chem. Soc.* **1987**, *109*, 2426.

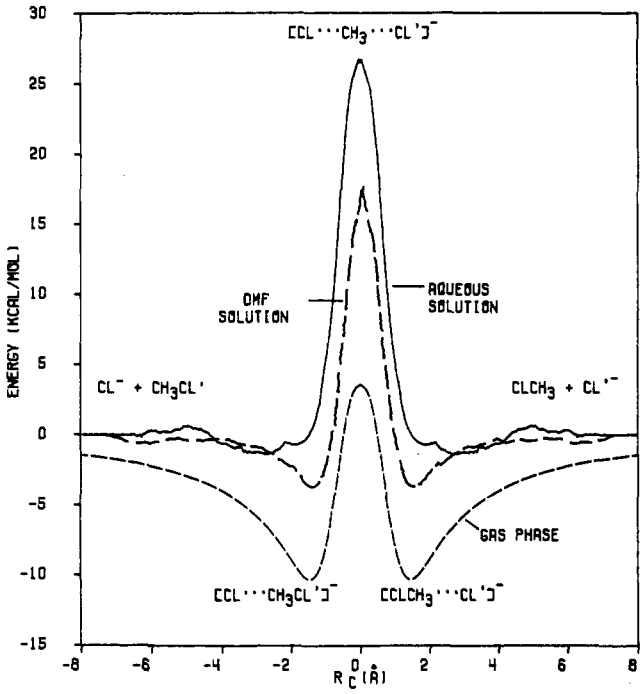


Figure 5. Calculated potential energy in the gas phase and potentials of mean force in water and in DMF for the S_N2 reaction of Cl^- + CH_3Cl. The reaction coordinate is the difference in the two C–Cl distances.

in DMF is intermediate with the ion–dipole complexes still appearing as free energy minima. The latter result requires revision of the traditional notion that S_N2 reactions in solution are concerted processes without intermediates. The computed results were shown to be in good accord with experimental free energies of activation. Moreover, this work has stimulated considerable activity in other research groups;[51-55] and deviations from normal Brønsted behavior in the kinetics for a wide range of S_N2 reactions in DMSO have now been explained by the two-step mechanism.[56]

The same methodology was then applied to the addition reaction of OH^- + $H_2C{=}O$.[15] The reaction proceeds without activation energy into a deep well for the tetrahedral intermediate in the gas phase. Hydration dramatically alters the surface with the introduction of a substantial free energy barrier. For both reactions, the principal source of the solvent-induced barriers in water was found to be weakening of solute–water hydrogen bonds on progression from the charge-localized reactants to the more charge delocalized transition states. Additional studies along these lines are in progress for Diels–Alder reactions and addition of a carbene to an alkene.

Concluding Remarks

The examples reviewed here illustrate the power and potential of molecular dynamics and statistical mechanics simulations for modeling organic chemistry in solution. Recent progress in the field has initiated a new era in which solvent effects can be studied theoretically in a quantitatively meaningful manner. The major contributor to this development has been the maturation of methodologies for free energy calculations and of technical knowledge on their application to complex systems. Many issues need more concerted study, particularly with regard to the adequacy of sampling the configuration space and to time scale. For example, if the initial geometry for a host–guest complex is seriously mistaken, the simulation must be run long enough to permit enough sampling to overcome the error. If there are multiple minima with significant interviewing barriers, the problem may be difficult to detect. The execution of the available computer programs is also not yet as routine as with well-known quantum mechanical packages, and the demands on computer resources are great. Nevertheless, much progress has been made, and new developments and applications will continue to emerge at a rapid pace.

I am grateful to many co-workers and colleagues listed in the references for their assistance and exchange of ideas. Financial support has been provided by the National Science Foundation and National Institutes of Health.

(51) Chiles, R. A.; Rossky, P. J. *J. Am. Chem. Soc.* **1984**, *106*, 6866.
(52) Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T. *J. Chem. Phys.* **1987**, *86*, 1356.
(53) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
(54) Bash, P. A.; Field, M. J.; Karplus, M. *J. Am. Chem. Soc.* **1987**, *109*, 8092.
(55) Hwang, J.-K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297.
(56) Bordwell, F. G.; Hughes, D. L. *J. Am. Chem. Soc.* **1986**, *108*, 7300.