Free Energy Calculations: Applications to Chemical and Biochemical Phenomena

Peter Kollman

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Received May 5, 1993 (Revised Manuscript Received August 24, 1993)

Contents

I. Abstract

I will review the applications of free energy calculations employing molecular dynamics or Monte Carlo methods to a variety of chemical and biochemical phenomena. The focus is on the applications of such calculations to molecular solvation, molecular association, macromolecular stability, and enzyme catalysis. The molecules discussed range from monovalent ions and small molecules to proteins and nucleic acids.

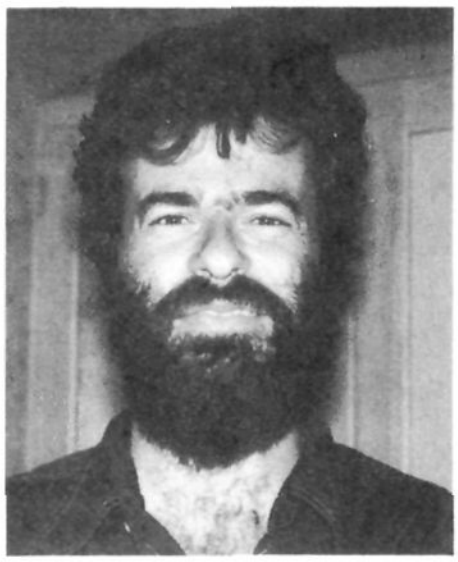
II. Introduction

Free energy is arguably the most important general concept in physical chemistry. The free energies of molecular systems describe their tendencies to associate and react. Thus, being able to predict this quantity using molecular theory in general would be an enormously important advance and is a seductive goal. Progress toward this goal has been made in recent years, and this review attempts to describe this progress as it applies to the use of molecular dynamics and Monte Carlo methods to carry out free energy calculations in the following areas: (1) solvation of small molecules, (2) ligand binding to organic hosts and to proteins and nucleic acids, (3) sequence-dependent stabilities of proteins and nucleic acids, and (4) environmental effects on reactions in solutions and in enzymes.

I will review of the methodologies used in such free energy calculations. After presenting some of the basic equations, I present a detailed discussion of the first application of the methodology to the calculation of the relative solvation free energy of the organic molecules methanol and ethane. The agreement between the calculated and experimental free energy is impressive, as is the inherent statistical error of <1 kcal/mol. I explain why the inherent error is much larger for calculations of the relative solvation enthalpies and entropies.

I then give a more general historical perspective to the development and application of free energy methods to the study of the solvation and association free energies of ionic, polar, and nonpolar molecules. In the practical implementation of free energy methods, there have been two kinds of approaches: single and dual topologies. I will also discuss how well the calculations were able to (a) reproduce experimental free energies, (b) give new mechanistic insights, and (c) be predictive.

0009-2665/93/0793-2395$12.00/0



Peter Kollman received his B.A. in chemistry from Grinnell College in 1966 and his Ph.D. in chemistry from Princeton University in 1970. After a NATO fellowship at Cambridge University in 1970–1971, he joined the faculty of the Department of Pharmaceutical Chemistry, School of Pharmacy, UCSF, where, since 1980, he has been professor of chemistry and pharmaceutical chemistry.

I will discuss the main barriers that hinder the broader application of free energy calculations. These can be succintly summarized as (1) errors in the representation of the energy of the system and (2) limitations in sampling enough of the relevant (low free energy) conformations of the system.

A number of reviews have appeared which have concentrated exclusively or in part on free energy calculations.[1-16] Our review has a number of unique aspects in its emphasis on the practical and greater detail on applications than many of the previous reviews.

III. Methodological Issues

A. Basic Formulation of Free Energy Calculations

The statistical mechanical definition of free energy is in terms of the partition function, a sum of the Bolzmann weights of all the energy levels of the systems. However, only for the simplest model system can this free energy be represented by an analytical function. One can write a classical analog of the quantum mechanical partition function where the energy is viewed as a continuous function, rather than discrete. This is likely to be a good approximation in most systems involving noncovalent interactions near room temperature.

Unfortunately, the free energy represented in this way requires an integration over all $3N$ degrees of freedom, where N = number of atoms in the system. Thus, this is impractical in most cases. However, if one focuses on free energy differences between related systems A and B ($\Delta G = G_B - G_A$) represented by Hamiltonia H_A and H_B, this free energy difference can be represented in eq 1

$$G_B - G_A = \Delta G = -RT \ln \langle e^{-\Delta H/RT} \rangle_A \qquad (1)$$

where $\Delta H = H_B - H_A$ and $\langle\ \rangle_A$ refers to an ensemble average over a system represented by Hamiltonian H_A. Equation 1 is the fundamental equation of free energy perturbation calculations. If systems A and B differ in more than a trivial way, then eq 1 will not lead to a sensible free energy. One can, however, generalize the problem and describe the Hamiltonian $H(\lambda)$ as in eq 2

$$H(\lambda) = \lambda H_B + (1 - \lambda) H_A \qquad (2)$$

where λ can vary from 0 ($H = H_A$) to 1 ($H = H_B$). One can then generalize eq 1 as follows:

$$\Delta G = G_B - G_A = \sum_{\lambda=0}^{1} -RT \ln \langle e^{-\Delta H'/RT} \rangle_\lambda \qquad (3)$$

where $\Delta H' = H_{\lambda+d\lambda} - H_\lambda$. One breaks up the free energy calculation into windows, each one involving a small enough interval in λ to allow the free energy to be calculated accurately.

An alternative to free energy perturbation calculations is *thermodynamic integration*, where the free energy difference between two systems (one characterized by $H = H_A$ or $\lambda = 0$ in eq 2 and the other by $H = H_B$ or $\lambda = 1$ in eq 2

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda \qquad (4)$$

The application of eq 4 requires one to evaluate the ensemble average of the derivative of the hamiltonian with respect to λ, $\langle \partial H/\partial \lambda \rangle_\lambda$ at various values of λ. One can then use numerical integration methods to calculate ΔG by eq 4.

The third commonly used method for free energy calculations is called *slow growth* in which the Hamiltonian is changed an infinitesimal amount over each step of the simulation (eq 5)

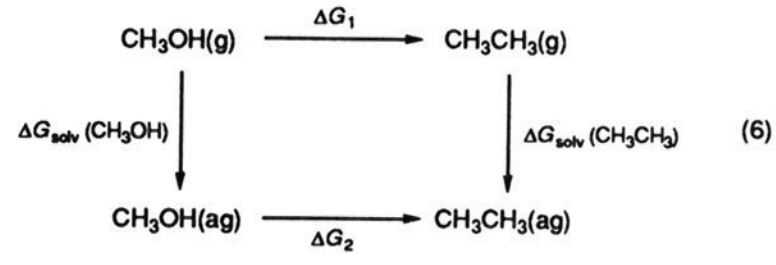
$$\Delta G = \sum_{no.\ steps}^{\lambda=1} \sum_{\lambda=0} (H_{n+1} - H_n) \qquad (5)$$

where H_n is the Hamiltonian for a given λ and H_{n+1} is the Hamiltonian for the next larger λ. This equation can be derived from eqs 1 or 4, using the assumption in eq 1 that ΔG is small and in eq 4, $\partial H/\partial \lambda = \Delta H/\Delta \lambda$.

If evaluated accurately enough, ΔG should be independent of path or simulation protocol, but there are often a number of practical reasons for using one of these three approaches.

As noted above in reference to eq 1, the realism of free energy calculations depends on the realism of the Hamiltonia H_A and H_B. To our knowledge, virtually all applications of this methodology make the assumption that the kinetic energy term in the Hamiltonian can be ignored. How realistic is this assumption? To proceed further on this point, let us focus on the first application of free energy calculations to the solvation of organic molecules: the study of the relative free energy of solvation of methanol and ethane by Jorgensen and Ravimohan[17] (JR). One begins by considering a free energy cycle (6).

B. A Sample Application: The Relative Solvation Free Energy of Methanol and Ethane



(6)

Since free energy is a state function, the difference in free energies of solvation of methanol and ethane,

$\Delta\Delta G = \Delta G_{solv}(CH_3CH_3) - \Delta G_{solv}(CH_3OH) = \Delta G_2 - \Delta G_1$, where ΔG_2 and ΔG_1 are the free energies of mutating methanol into ethane in solution and in the gas phase, respectively. JR used Monte Carlo calculations and eq 3 (free energy perturbation) to calculate ΔG_2. The OPLS solute model used by JR involves a united atom CH_3 group, so CH_3OH is a triatomic molecule and $CH_3 - CH_3$ a diatomic. JR assumed that differences in ΔG due to kinetic energy differences would be identical in calculating ΔG_1 and ΔG_2, so these were not included in either calculation. Van Gunsteren has validated this approximation in calculations of simple systems.[18] Given that, it is reasonable to make the assumption that $H \sim V$. What is a typical potential energy function V? Weiner *et al.*[19,20] use eq 7; the OPLS model uses this form of the equation without the explicit H-bond term.

$$V = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{\substack{i<j \\ nonbonded}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] + \sum_{H\text{-}bonds} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right] \quad (7)$$

Jorgensen and Tirado-Rives[21] have adopted the OPLS model for molecular mechanics and dynamics using the Weiner *et al.* parameters for the first three terms in eq 7: bond stretching, bond bending, and torsional rotation, while employing nonbonded terms they have derived by carrying out Monte Carlo calculations on requisite liquids for the nonbonded part of the potential.

In their Monte Carlo evaluation of ΔG_2 (eq 3 and 6), JR assumed rigid bond lengths and angles; thus no *intramolecular* contribution to the free energy was calculated for the mutation of a "triatomic" molecule (CH_3OH) to a "diatomic" (CH_3—CH_3). As in the case of the kinetic energy term, one is assuming that any such contributions to ΔG_2 are identical to those that would appear in ΔG_1. Recently, Cieplak and Veenstra of UCSF (unpublished) have examined this approximation using molecular dynamics calculations and have found it to be valid for mutations of methane and ethane to methanol and dimethyl ether to propane.

Typically, Monte Carlo calculations on complex molecules assume rigid bond lengths and angles, as did the JR calculation of ΔG_2. As we have noted above, in most practical applications of free energy perturbation, one must use eq 3, in which one creates a number of hybrid systems intermediate between CH_3OH and CH_3-CH_3. For example, the C–O bond length in methanol is 1.43 Å; the C–C bond in ethane is 1.53 Å. A hybrid state ($\lambda = 0.5$) would involve a bond distance between the CH_3 group and the changing atom (O → CH_3) of 1.48 Å. In the OPLS model of methanol, the charge on hydrogen is 0.435, on oxygen, −0.700, and on the methyl group, 0.265. In the $\lambda = 0$ (ethane), the charges are zero. The van der Waals parameters are similarly interpolated between methanol and ethane for the $\lambda = 0.5$ state.

JR began the simulation by inserting the methanol molecule in a box of 125 TIP4P water molecules and carrying out Monte Carlo calculations to equilibrate this system in an isobaric ensemble (constant number of particles, temperature, and pressure). They then evaluated the free energy for mutating methanol to the $\lambda = 0.125$ state, which is 7/8 methanol and 1/8 ethane. They used "double wide sampling", which they found to be a useful test for convergence of the free energy. This involves calculating the free energy difference for both the $\lambda \rightarrow \lambda'$ and $\lambda' \rightarrow \lambda$ intervals. If one applies eq 3 and evaluates the ensemble at state λ and evaluates the free energy to mutate this into λ', this should be the negative of the free energy determined by using the ensemble characteristic of λ' and calculating the free energy to mutate this to state λ.

JR found they needed more (closely spaced) values of λ near the methanol state than the ethane state, because the free energy of interaction with the surrounding waters is changing more rapidly in this range. When they had evaluated ΔG_2 by mutating methanol to ethane, they found a calculated $\Delta G = 6.75 \pm 0.2$ kcal/mol, in excellent agreement with the experimental $\Delta G = 6.93$ kcal/mol. This was a most exciting result, since the molecular mechanical models for water and methanol, derived from reproducing the enthalpies and densities of the respective liquids, could be used without modification in a binary system involving both molecules and the experimental free energies remained excellent. Nowadays, one can achieve such agreement with much simpler models, but at the time this was a most exciting result to this reviewer. It led to the general incorporation of the free energy approach into the simulation program AMBER by Singh.[22]

C. Why Is the Calculation of ΔG More Accurate Than the Calculation of ΔH and ΔS?

To put this result into context, one must appreciate that to directly calculate the difference in *enthalpy* of solvation, ΔH of methanol and ethane would have required separate simulations of the two solutions and taking the difference in the total energy of the two systems (125 waters and 1 solute). The total energy of these systems from Monte Carlo calculations is of the order of -1250 ± 10 kcal/mol. Thus, any directly calculated enthalpy ΔH would have an inherent error of ±10 kcal/mol, not <1 kcal/mol.

The difference between the difficulties in calculating ΔG and the enthalpy ΔH are fundamental and related to the fact that the enthalpy ΔH must be determined by determining the difference between two large numbers, which are dominated by the solvent-solvent interaction energies and ΔG (eqs 1 and 3) can be calculated by determining the ensemble averaged difference between solute–solvent interactions directly. The difference in the Hamiltonia, ΔH, in eqs 1 and 3 includes only the difference in the way methanol and ethane interact with water, since the water–water energy remains the same in the two Hamiltonia. Differences in solvent structure, which certainly occur in response to substituting and OH group with a CH_3 group, are reflected in the ensemble average $\langle\ \rangle$ in applying eqs 1 and 3, not in the ΔH in the exponent.

Fleichman and Brooks[23] have derived a more efficient way to calculate ΔH and ΔS in addition to ΔG and applied it to the relation solvation free energies of methanol, ethane, propane, and butane. They found that the inherent error in the calculated ΔH and ΔS are at least 1 order of magnitude larger than that for ΔG.

D. Historical Perspective of Free Energy Calculations Applied to Chemistry/Biochemistry

Let me now give some historical perspective to the development of free energy calculations for use in systems of interest in organic and biochemistry. The basic equations for free energy perturbation and thermodynamic integration were developed by Zwanzig,[24] Kirkwood,[25] and Valleau and Tomie,[26] but it was in the early 1980s that they were used in analysis and simulation of biophysical systems. Postma *et al.*[27] studied noble gas solvation, Warshel[28] presented "preliminary results" on the solvation free energy contribution to an electron transfer reaction coordinate using two spheres for donor and acceptor and a dipolar model of water, and McCammon showed the usefulness of free energy perturbation calculations on a model system[29] prior to JR's study on the relative solvation free energy of methanol and ethane.[17] The advent of a major enhancement in computational power of the vector CRAY X-MP enabled a major advance in the generality of application of free energy methods by Bash *et al.*[30,31] In a pair of papers in *Science* they studied the relative solvation free energy of a wide variety of amino acid side chains, nucleic acid bases, and other organic molecules, as well as the relative binding free energy of a pair of ligands to the protein thermolysin undergoing experimental studies as well. Those studies clearly demonstrated the power and generality of free energy calculations and the feasibility of studying large mutations such as alanine → tryptophan and methane → 9-CH_3-guanine. The largest mutation attempted prior to Bash *et al.*'s study involved addition or mutation of one or two atoms. Although the calculations were of rather limited duration by today's standards, they clearly showed feasibility of studying a wide variety of chemistry with these approaches and achieving results in reasonable agreement with experiment with modest (±1 kcal/mol) error bars.

In the process of their studies, Bash *et al.* found it to be useful to employ "electrostatic decoupling", i.e. changing only the electrostatic part of the molecular mechanical potential function first and then the remaining. This was motivated by the results of some simulations (e.g. histidine → alanine) where the presence of hydrogens with some remaining charge but very small van der Waals repulsion experienced an artifactually large free energy change upon approach of a solvent water. One could also imagine solving this problem with a different λ dependence in the electrostatic and van der Waals parts of the molecular mechanics potential function.

E. Challenges in Free Energy Calculations on the Solvation of Ionic, Polar, and Nonpolar Molecules

What were the major limitations found by Bash *et al.* in simulating the solvation free energy charges accurately? One can consider three types of molecules: nonpolar, polar, and ionic. The simulation of polar and ionic solvation effects is dominated by the electrostatic energies. These can be simulated reasonably accurately and reproduceably with rather limited simulation lengths. Even the simulation of Ne → Na^+ by Straatsma and Berendsen[32] can be accomplished with convergence to ~1 kcal/mol out of ~100 kcal/mol in 80 ps of molecular dynamics. Polar dominated mutations such as methanol → ethane[17] convergence rapidly also. The main difficulty in the electrostatic dominated perturbations comes when one changes the net charge of the system. For example, the free energy calculated for the Ne → Na^+ mutation will depend strongly on the nonbonded cutoff in the simulation. The solvation free energy estimated using a continuum model for creating a monovalent ion[33] suggests that, using the 8-Å nonbonded cutoff typical of simulations will result in a ~20 kcal/mol underestimate of the absolute value of the solvation free energy. Using the simple Born formula to correct such errors is not rigorously correct when one uses nonspherical boundary conditions, although using the correction is certainly better than not using it.

Where dealing with this problem becomes very important is in the calculation of p*K*a's for ionizable groups in proteins, where the presence of numerous charge groups complicates matters and it is important to calculate the solvation free energies of the ionized groups to ±1–2 kcal/mol or better. This large challenge has been undertaken by Lee and Warshel[34] with continuing improved success, including improved models to accurately represent long-range effects. For example, the local reaction field method they propose is significantly more efficient than no-cutoff methods using spherical boundary conditions at a fraction of the computational expense. Warshel recommends a hybrid model, with explicit representation of solvent to a given distance, further waters with the PDLD (Langevin dipole) model, and a continuum electrostatic model beyond that. In free energy calculations, one can also use a hybrid approach, where, for the molecule or fragment which is being mutated, no nonbonded cutoff is used, with an 8-Å cutoff used for the rest of the system. Provided that the system is net neutral or close to it, this approach also offers a significant improvement over the standard 8-Å cutoff at only a modest additional computational expense. There have also been other recent, exciting new approaches to efficiently incorporate long-range electrostatic effects into simulations in general.[35,36]

To accurately simulate small, nonpolar mutations is a particular challenge because the ΔG is very small and can be a small difference between the positive exchange repulsion and the negative dispersion attraction (eq 7). For example, the following relative experimental free energies of solvation[37] in water (in kcal/mol, 1 M standard state) illustrate this (8).

$$\text{nothing} \xrightarrow{+1.94} CH_4 \xrightarrow{0.16} C_2H_6 \xrightarrow{0.21} C_3H_8 \qquad (8)$$

Sun *et al.*[38] have shown how one can simulate methane → ethane and ethane → propane rather accurately, using Spellmeyer's new all atom van der Waals parameters and Pearlman's bond pmf correction[39] and new protocol for the representation of the van der Waals part of $V(\lambda)$ for disappearing groups.[40] Insuring converged free energies requires ~300 ps of simulation in each direction; ethane → propane is calculated to within ~0.1 kcal/mol of experiment; whereas the methane → ethane calculation is more dependent on partial charge model and is overestimated by ~0.1–0.3 kcal/mol.

Why do van der Waals perturbations, where one is disappearing atoms, require long simulations for convergence than the electrostatic dominated ones, whose ΔG are so much larger in magnitude? In our view it is because electrostatic dominated changes require mainly dipolar reorientation of the solvent, whereas the van der Waals charges due to growing and disappearing atoms, involve more slow repacking and translational diffusion of the solvent.

Hermans[41] has shown how one can estimate noise and hysteresis for the slow growth method; these two quantities are related to the relaxation time of the system and the width of an assumed Gaussian distribution in configuration space. Hermans *et al.* have also provided detailed examples of protocols to estimate errors in different types of free energy calculations.[42] Both Pearlman and Kollman[43] and Wood[44] have evaluated the Hamiltonial "lag" in slow growth calculations and the implications of this lag in accurate representation of the calculated free energies.

F. Single and Dual Topologies in Free Energy Calculations

There are two distinct approaches in how to represent the molecular mechanical topologies in free energy calculations. In the approach used by Jorgensen and independently incorporated into the molecular dynamics programs GROMOS and AMBER, one uses a single molecular topology, and the terms in eq 7 change its shape and properties as λ changes. In the molecular mechanics program CHARMM[40] one keeps two independent *topologies*, one for methanol and the other for ethane. For example, in the methanol to ethane perturbation, the methanol OH and ethane CH_3 both exist at the same time in the calculations, but they do not interact with each other. The (only about 0.1 Å apart) interaction of these groups with their environment is calculated using eq 2 or a variant of it, eq 9.

$$H(\lambda) = \lambda^n H_B + (1 - \lambda)^n H_A \qquad (9)$$

The use of different exponents in (9) allows difference pathways and corresponding integrands $\partial H/\partial\lambda$ in eq 4 to be used in determining the free energies. On the other hand, there are a variety of ways within the single topology method to include the λ dependence more directly inside the terms in the potential energy (eq 7).

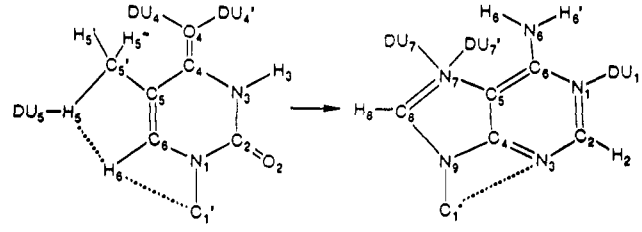
In the single topology method, Pearlman[39] has shown that in molecular dynamics one must determine a "bond pmf" correction when bond length change, in order to determine a rigorously correct free energy for the mutation. Such a correction is not required in the dual topology method, but in the latter method, the best way to overlap or constrain the topologies when mutating, e.g. 1-CH_3 thymine to 1-CH_3 cytosine, is not obvious, and the resulting free energy may be very sensitive to the protocol chosen. At this point, it is fair to say that interesting and useful free energy calculations have been carried out with both approaches.

G. Limitations in the Implementation of Free Energy Methodology in AMBER3 and the Removal of These Limitations in AMBER4

In the Singh *et al.*[46] implementation of free energy approaches within AMBER 3.0, there were three limitations: (1) First, intra "perturbed group" contributions to the energy were calculated in determining the ensemble but were not included in the free energy. This was a deliberate choice because it was reasoned that the large intramolecular energy change involved in, for example, removing an imidizole ring would add too much noise into the free energy and obscure the more important differences in the intermolecular interactions. For example, the inclusion of such terms would require that both ΔG_1 and ΔG_2 be calculated in cycle 6, whereas their neglect would allow the calculation of only ΔG_2. By using a rigid geometry model, that is essentially what was done by JR,[17] with excellent results. (2) Second, because it was not clear how to best carry out the transformation of a 10–12 to a 6–12 nonbonded parameter, only the latter was included in the description of the "perturbed group". As shown by Ferguson *et al.*[47] one can replace 10–12 parameters with appropriate 6–12 parameters so this should not be a large issue; nonetheless, one could not fully correctly implement the Weiner *et al.*[19,20] force field for free energy calculations because of this limitation. Such an issue is not relevant for the OPLS/AMBER force field, which contains only 6–12 parameters. (3) The calculations of Bash *et al.*[30] and Merz *et al.*[48] showed a significant sensitivity to the calculated free energy for disappearing groups depending on whether the bonds were "shrunk" while the group disappeared. For example, if one mutates methane to "nothing" to calculate the absolute free energy of solvation of methane, should one reduce the C–H bond lengths as the molecule disappears and, if so, by how much? Since the free energy must be the same irrespective of the final "C–H" bond length in the dummy methane, it was puzzling that the calculated free energies were so dependent on whether the bonds were "shrunk" in the process.

Pearlman's implementation of the GIBBS module in AMBER 4[40] has removed the three limitations: (1) one now has the option to include or not the intragroup energies, as well as a selected set of them; (2) atom pairs interacting through 10–12 parameters can be mutated to those interacting with 6–12 parameters with fully correct representation of the energy; and (3) "PMF correction" allows the correct free energy to be calculated, irrespective of bond length changes. There are many other implementation improvements included in AMBER 4 GIBBS, including the new combining rules for nonbonded parameters involving disappearing atoms, which were critical in the accurate representation of hydrocarbon solvation free energies by Sun *et al.*[38]

The only remaining limitation in the methodology is a consequence of the use of a single automatically generated topology in free energy calculations. This limitation can be illustrated by considering the mutation of the thymine nucleoside to adenosine (only the C1′ of the sugar is represented below).



In the single topology perturbation, the thymine H_6 is mapped into guanine N_9, H_5 into C_8, C_5, into N_7 DU_5 into H_8, and $H_{5'}$ and $H_{5''}$ become dummy atoms DU_7 and $DU_{7''}$ DU and DU'_4, become the H_6 and $H_{6'}$, O_4 becomes N_6, H_3 become DU_1, and O_2 becomes H_2, etc. Because of the perturbation, bond C1′–N1 disappears, to be replaced by C1′–H6 (N9). A new bond is formed between H6 and H5 (N_9–C_8). These bonds can have a force constant of zero in the state where they do not exist, but their presence leads to an inconsistent implementation of the molecular mechanical model. For example, the "bond" between H5 and H6 means that no van der Waals parameters are evaluated between H5 and H6, even at $\lambda = 0$ (thymine), since the model does not include nonbonded interactions between atoms separated by one or two bonds. The nonbonded interactions between atoms separated by three bonds (1–4 interactions) are also handled different than longer range non-bonded interactions in the Weiner *et al.* force field.[19,20] In the above topology, however, one would treat H6···H5′ and H5″ interactions as 1–4 interactions rather than regular nonbonded interactions.

These limitations may not be critical in semiquantitative studies of sequence dependent perturbations in DNA helices (e.g. AT → TA), since one is always taking the difference between two perturbations and the inconsistent intramolecular energies could largely cancel. For example, Ferguson has examined the molecular mechanical energy dependence of the χ (C1′–N) angle in thymine nucleoside using the perturbation and regular topologies and found reasonable, if not quantitative agreement.[46] Situations like the T → A perturbation are analogous to the neglect of intragroup perturbations required in AMBER3. However, the new combining rules implemented in AMBER 4[40] also allow the "dual topology approach" because atoms that exist at $\lambda = 0$ and not at $\lambda = 1$ (or vice versa) do not experience nonbonded interactions with each other, even in intermediate ($\lambda \neq 0$, 1 states). One would have an extra angle term (N1–C1′–N9), but this could be given a force constant of zero. Whether either or both single or dual topologies will allow the more effective simulation of free energy perturbation within nucleotide double helices remains to be seen, but preliminary results with the single topology method are promising.

Recently, Pearlman (*J. Am. Chem. Soc.*, submitted for publication) has compared the convergence in single and dual topology approaches for an ethane → ethane mutation, where the true answer must be zero and found convergence more rapid for the single topology method.

H. Comparison of Statistical Perturbation Theory, Thermodynamic Integration, and Slow Growth

What are the advantages and disadvantages of the three approaches used in free energy calculations, statistical perturbation (SP), thermodynamic integration (TI), and slow growth (SG)? SP does not require an analytical derivative of H with respect to λ; this derivative is trivial in the dual topology approach, but can be complex in single topology. SP can also give more problems than SG for van der Waals dominated changes as atoms are appearing or disappearing, if $\Delta\lambda$ is too large. It is hard to know in advance what size $\Delta\lambda$ to choose although Pearlman's "dynamically modified windows"[49] has alleviated this problem. There are no fundamental limitations in SP provided computer time is no object and the $\Delta\lambda$ are chosen sufficiently small. The main fundamental weakness in SP is that the total free energy cannot be separated into the sum of the component free energies, because the logarithm of an exponential with different energy components is not equal to the sum of the logarithms of the individual components. Thus free energy component analysis cannot be rigorously carried out with that approach.[50]

Thermodynamic integration, unlike SP, requires a numerical integration of the values of the integrand $\langle \partial H/\partial\lambda \rangle$. This is not a fundamental limitation[51] and, provided that enough λ values are chosen, accurate free energies can be calculated. An advantage of TI (or SP) over slow growth (SG) is that after one has evaluated a number of $\langle \partial H/\partial\lambda \rangle$ and carried out the numerical integration, one concludes that one needs more sampling at already sampled λ's or more λ's in between, this can be easily done without "losing" any of the information previously derived. In SG, one needs to rerun the entire trajectory again, if one decides to double the sampling time. Van Helden and van Gunsteren have shown that, if the system fluctuates over multiple conformations, one is more likely to sample this correctly and efficiently in TI than in SG.[52] SG also suffers from the "Hamiltonian lag", since the Hamiltonian changes at every step.[49]

All in all, TI seems to be the best compromise way to carry out free energy calculations *and* do component analysis;[53,54] however, provided sufficient sampling is done and multiple trajectories examined, all these methods can give useful and insightful results. Far more critical than the choice of which of these methods to use are the two key issues in free energy calculations: (1) accuracy of the Hamiltonian (potential energy) function and (2) the sampling problem.

As noted above, intragroup free energies can be large, and it is not clear how important they are in specific cases. We have examined (P. Cieplak and D. Veenstra, unpublished) their importance in the calculation of the relative free energy of solvation of methanol and ethane or methane and dimethyl ether and propane, finding a <0.3 kcal/mol contribution to the relative free energies of solvation from intragroup free energies. On the other hand, Prevost *et al.*[55] have found that the intragroup energis are ~3 kcal/mol out of a total of ~5 kcal/mol for the mutation of Ile → Ala in native barnase, relative to a model for the denatured protein. This result seems counterintuitive, and further work is required to see how real or general this large magnitude of an intramolecular free energy is.

I. Free Energies Can Be Calculated for Coordinate as Well as Topology Changes

In addition to the free energy due to chemical mutation, free energy calculations have long been applied to calculate the free energy change as a function of a "reaction" coordinate. The earliest work used umbrella sampling, e.g. the McCammon/Karplus study of tyrosine ring rotation in BPTI,[56] but it appears that free energy perturbation methods are more efficient at describing such free energies. A fundamental question related to the hydrophobic effect is how is the gauche → trans conformational equilibrium in butane perturbed in aqueous solution relative to the gas phase or

the neat liquid.[57] Aqueous solvation appears to increase the gauche population by ~10%, a noticeable but not exceptionally large effect. This calculation required the determination of the free energy as a function of torsional angle. Analogously the study of CH_4 dimerization required the determination of the free energy as a function of bond distance, a so-called potential of mean force (PMF) for association.[58] The $(CH_4)_2$, $Na^+\cdots Cl^-$ and $Cl^-\cdots Cl^-$ PMF's have revealed "contact" minima, and, in the case of $(CH_4)_2$ and $Na^+\cdots Cl^-$, also solvent separated minima.[58-60]

Tobias and Brooks[61] have suggested how to implement free energy as a function of coordinates in molecular dynamics and have used this protocol in the simulation of conformational changes in peptides.[62-65] Pearlman has carried out free energy profiles of the χ and γ dihedral angles in nucleosides both in vacuo and in solution.[66] Dang has studied nucleic acid base[67] and K^+/18-crown-6 association in solution using PMF approaches.[68]

Recently Boczko and Brooks[69] suggested that a variant of the multiple histogram method[70] was more efficient than umbrella sampling for calculating free energy as a function of coordinate. Pearlman has also compared methodologies for calculating free energy as a function of coordinate.[71]

Hermans has used a combination of conformational and mutation free energy approaches to characterize the relative helical propensity of various amino acids.[72-74] Not only were these calculations in good agreement with available (and subsequent experiments) but gave nice insight into why certain residues had greater helical propensity than others. For example, the crucial role of greater entropy in the denatured state leads Gly to be less stable in a helix relative to Ala, whereas α-aminobutyric acid (AIB) is more stable than Ala in a helix because its greater rigidity and lowest energy in the helical area of the ψ,ϕ map.

J. Dependence of Calculated Free Energies on Molecular Mechanical Model

There are two issues in assessing the accuracy of potential energy functions such as (7): the accuracy of the parameters in the equation and the inherent ability of the functional form to correctly represent the system. As noted above, the OPLS model[21] has proven to be accurate in its calculation of solvation free energies. This is because the model is inherently well-balanced, having had both solute and solvent parameters derived using Monte Carlo calculations to reproduce the densities and enthalpies of vaporization of liquids. The Bash *et al.*[30] study used mainly the Weiner *et al.* force field,[19,20] which had been derived in a different way, together with the TIP3P water model; however, they modified the charge model to use 6-31G* electrostatic potential derived charges, which are much more "OPLS like" than the Weiner *et al.* charges. Kuyper *et al.*[75] have examined this issue in more detail with the relative solvation energies of methoxy and trimethoxybenzene and benzene and have found that 6-31G* derived electrostatic potential charges led to excellent solvation free energies, STO-3G electrostatic potential charges gave reasonable ones, but 4-31G electrostatic potential derived charges greatly exaggerate the solvation free energy. This can be related to the ability of these charge models to reasonably reproduce the molecular multipole moments. The 6-31G* model inherently overestimates molecular dipole moments by ~10–20% and thus contains some of the "implicit" polarization that the OPLS achieves by fitting to liquids. Often, however, there are cancellation of errors in free energy calculations so rather different charge models can lead to rather similar results. For example, Miyamoto has mutated a 6-31G* charge model of biotin to an STO-3G model both in solution and in the binding site of steptavidin.[76] Both ΔG values were ~15 kcal/mol, reflective of the significantly smaller polarity of the STO-3G model, but their difference was ~1 kcal/mol, small compared to the free energy of binding of ~20 kcal/mol. Deriving charge models that accurately represent intramolecular as well as intermolecular properties is a significant challenge, as is even reproducing the fact that *cis*- and *trans*-*N*-methylacetamide have a nearly identical solvation free energy. The standard OPLS model finds ~2 kcal/mol for the solvation free energy difference,[77] 6-31G* electrostatic potential derived charges for trans NMA finds ~1 kcal/mol for the solvation free energy difference, and using the 6-31G* electrostatic potential charges for cis-NMA to represent cis-NMA and 6-31G* electrostatic potential charges for trans-NMA to represent trans NMA reproduces the nearly identical solvation free energy.[78]

Reynolds *et al.*[79,80] have shown how one can use multiple conformation fitting to improve electrostatic potential derived charge models and Bayly *et al.*[81] Cornell *et al.*,[82] and Cieplak *et al.*[83] have used multiple conformations, multiple molecules, and restrained electrostatic potentials to provide further improvements in the methodology in deriving atomic partial charges for molecular mechanics/free energy calculations.

One might expect that a simple equation such as (7) would break down in treating ionic systems, which would be expected to include significant ionic effects. However, both Urban and Damewood[84] and Aqvist[85] have shown that one can derive "effective" ion parameters that reproduce free energies of solvation even with additive models. Aqvist and more recently Marrone and Merz[86] have derived the parameters by carrying out solvation free energy calculations and adjusting the van der Waals R* and ϵ to reproduce the experimental free energies of solvation and, as well as possible, the radial distribution functions. These models would be expected to be less accurate for small, gas-phase ion clusters; these can be treated with nonadditive effects, as a number of studies have shown.[87-89] Warshel *et al.*[90] have often included explicit polarization effects in their studies, including the calculation of relative pK_a values of protein functional groups.

Cieplak has carried out the first free energy calculation using nonadditive effects on small ion–water clusters, employing Monte Carlo calculations to derive free energies for these clusters.[87] Straatsma and McCammon applied free energy approaches including nonadditive effects in molecular dynamics, studying the free energy of solvation of a small solute and water in water.[91,92] Recently, Sun *et al.* have shown how free energy perturbation including nonadditive effects could improve the calculated Li^+/Na^+ selectivity of an ionophore.[93]

K. The Sampling Issue

Thus, there are clearly a wide variety of chemical phenomena which can be treated quantitatively with free energy methods using molecular mechanical models such as (7) or variants that can include some nonadditive effects. The major roadblock in broader applications of free energy approaches is most often not the accuracy of the potential energy function but the sampling issue. The fact that existing methods have been so successful in reproducing free energies of solvation and binding in simple, well-defined systems supports this. However, for those systems, even implicit solvation models such as GBSA,[94] DELPHI,[95] and AMSOL[96] can often do a rather good job as well; but the full free energy calculations are in any case excellent reference points. Furthermore, new methodologies for more efficient sampling in free energy calculations are continuing to be developed. The multiple histogram method[70] and locally enhanced sampling[97] methodologies are recent, exciting developments.

For small well-defined systems the determination of solvation free energies or binding free energies in solution, adequate sampling is, we feel, no longer an issue. If need be, such systems can be simulated for times approaching 1 ns with periodic boundary conditions, and this appears adequate to describe, with a statistical error of <0.5 kcal/mol and perhaps even better, even van der Waals dominated changes (electrostatic dominated changes generally converge more rapidly). However for systems with rotational degrees of freedom, even for something as simple as 18-crown-6, sampling all the relevant conformations in solution does not occur in ~1 ns at 300 K.[98] For a more complex system, the situation is worse: the time scale for protein folding is milliseconds-seconds. How then can one even apply free energy methods to proteins or nucleic acids? A typical application involves using an X-ray or NMR derived structure for the macromolecule and the assumption that (a) a simulation for 10's or 100's of picoseconds on this structure would retain its essential features and (b) any mutation would involve a small enough structural change that it could occur in the <1 ns of simulation. Obviously one can test (a), which depends on potential function and representation (e.g. boundary condition, long-range electrostatics, etc), but usually one must assume (b). That this is not usually a bad assumption is supported by the many X-ray structures of proteins and their mutants[99] and proteins and their ligands which show small structural differences, but there are exceptions.[100]

With these unprovable uncertainties, are the calculations worth doing, i.e. do they lead to useful insight and are predictive? This is for the reader to decide after reading the applications described below. It seems to this author that in very complex protein and nucleic acids, "more is not necessarily better", given the inaccuracies in the force field/representation. If the system drifts far from the crystal structure, the calculated free energies may well be less accurate/representative of reality than a shorter simulation. However, a particularly sobering example is presented by the calculations by Hirono and Kollman[101] in their mutation of 2′GMP → 2′AMP in RNAse T1 and in solution (see Section IV B).

L. Combining Quantum and Molecular Mechanical Methods

Electronic structure changes cannot be described by equations such as (7), and they are at the heart of chemistry. However, free energy calculations can be combined with quantum mechanical calculations in a number of ways. For example, in the study of tautomeric equilibrium or p*K*a's in solution, one can combine accurate quantum mechanical calculations on gas phase equilibrium and free energy calculations in solution. Warshel's EVB model involves the calibration of a simple valence bond model including noncovalent interactions due to environment described with molecular mechanics to reproduce reaction free energies in solution. Then this same quantum mechanical model is transferred to the enzyme active site and the energetics of the same reaction in the different molecular mechanically represented environment of an enzyme active site are calculated.[16] This enables one to get insight into the nature of enzyme catalysis. Jorgensen has used high level *ab initio* calculations on model organic reactions and then evaluated the solvation free energies at steps along the (gas phase determined) reaction pathway to estimate the solvation free energy contributions to reactions in various solvents.[7] All in all, the future is bright for the continued combination of quantum and statistical mechanical methods to study chemical reactions in solutions and in macromolecules.

In summary, interesting applications of free energy methods in a wide variety of chemical and biochemical systems have been and are continuing to be done. We now review the applications to date, attempting to be exhaustive for the period from 1990 to mid-1993.

IV. Applications

A. Solvation

As noted above, an accurate calculation of the relative solvation free energies of different solutes in a solvent is an essential test for any model which hopes to simulate any chemical or physical process in a condensed phase. Thus, the profound impression that the JR calculations[17] made. Bash's calculations[30] were also a "tour de force" for their time because they showed that one could calculate free energy changes involving ionic, polar, and nonpolar mutations in aqueous solution of impressively large topology change, with rather small statistical error (generally ±1 kcal/mol) and reasonable agreement with available experiments on a variety of models of amino acid side chains. Furthermore, they studied nonadditivity effects in molecular solvation; e.g., *p*-NO_2 phenol has a ΔG_{solv} in water ~0.8 kcal/mol more negative than one would expect from comparing benzene and phenol and benzene and nitrobenzene and adding these two free energy differences. This nonadditivity is a consequence of the resonance between the NO_2 and OH group that makes *p*-nitrophenol a stronger acid than phenol. (Of course, this −0.8 kcal/mol nonadditivity refers to relative solvation of the neutral molecules). Bash *et al.* studied this nonadditive effect in an interesting and direct way that illustrates some of the power of modeling. They created a charge model of *p*-nitrophenol by "merging" the 6-31G* electrostatic

potential charge models of *p*-nitrobenzene and phenol. This model would not include electronic interactions of OH and NO_2 groups. They then determined the 6-31G* electrostatic potential charge model of *p*-nitrophenol directly and mutated one charge model into the other using free energy calculations. The calculated ΔG was in excellent agreement with the experimental nonadditivity. A second set of molecules studied by Bash *et al.*[30] were acetamide, *N*-methylacetamide and *N,N*-dimethylacetamide. Interestingly this is also an example of a nonadditive effect, since the singly methylated *N*-methylacetamide is more soluble in water than either acetamide or *N,N*-dimethylacetamide. The calculations qualitatively reproduced this.

By mutating a molecule to "dummy atoms", one can determine the absolute free energy of solvation. For example, Bash *et al.*[30] mutated all of the N-methylated nucleic acid bases → CH_4 and then, by mutating CH_4 → nothing, were able to predict the absolute free energy of solvation of the bases, none of which had been or have been measured directly. More recently, Ferguson *et al.*[47] have related these calculated free energies to experimental sublimation energy data from which one can reasonably accurately estimate the solvation free energies. The agreement between calculation and experiment was reasonable.

1. Aqueous Solvation

The solvation free energy of a wide variety of molecules has been calculated by Jorgensen and coworkers using the BOSS program[102] and Monte Carlo methodologies. Recent studies include the study of the solvation free energy of benzene and substituted benzenes[103] and the free energy of aromatic···aromatic association in water.[104]

Lee *et al.*[105] have compared simulation approaches to calculate solvation free energies of a wide variety of functional groups, including protein side chains, in water. They showed that PDLD (protein dipoles–langevin dipoles) methodologies provided a useful, inexpensive alternative to full free energy calculations and calculated free energies in impressive agreement with experiment for a wide variety of molecules and phenomena ranging from ionic strength effects on ion pairing, pK_a of protein side chains and molecular association. They make some useful comparisons with the results obtained by purely macroscopic electrostatic models and conclude that the semimicroscopic PDLD/S model is an efficient alternative to fully macroscopic or fully microscopic calculations.

The Warshel[105] approach focuses on the electrostatic energy and uses an empirical correction for hydrophobic effects. They are able to show that, with appropriate treatment of long-range electrostatics,[34] very accurate solvation free energies for polar and ionic molecules can be calculated. On the other hand, a very accurate (<0.3 kcal/mol) representation of aqueous solvation free energies for nonpolar molecules (e.g. CH_4 vs C_2H_6 vs C_3H_8) requires extensive simulation and careful parameterization.[38]

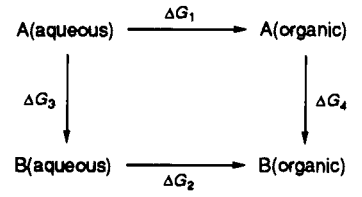
The solvation effect of bringing two water molecules together in water was studied by Mezei and Ben Naim.[106] They noted that there were specific strong "solvent" effects stabilizing the water dimer at particular distances (e.g. ~4 Å) that enable effective bridging interactions. This is analogous to the interpretation of why $Cl^-\cdots Cl^-$ association is calculated to be surprisingly favorable in water.[59]

Free energy solvation calculations have been used on novel nucleic acid bases to aid in the rationalization of relative double helical stabilities of base pairs involving these novel bases.[107]

2. Nonaqueous Solvents and Partition Coefficients

It is of considerable interest to compare the properties of aqueous and nonaqueous solvation using free energy calculations. Thus, Rao and Singh[108,109] have carried out free energy calculations on ionic and nonpolar solvation in hydrazine, methanol, DMSO, CCl_4, and H_2O. The results are intuitively reasonable and the calculated free energies for $Cl^- \rightarrow Br^-$, $Na^+ \rightarrow K^+$, $NH_4^+ \rightarrow N(CH_3)_4^+$, $CH_4 \rightarrow C_2H_6$, and $CH_4 \rightarrow C(CH_3)_4$ are generally in good agreement with experiment. But the most dramatic result in their papers is the difference in the free energy as a function of λ curve for $C(CH_3)_4 \rightarrow CH_4$. In all four nonaqueous solvents, the free energy as one disappears the methyl groups goes monotonically up, reflecting the loss of dispersion attraction. *Only in water* does the free energy first go down, reflecting the reduction in exchange repulsion as the solute decreases in size, and then, as λ decreases below 0.5, does the loss of dispersion attraction dominate and the free energy begin to increase. However, both in the calculation and in experiment, the free energy for the mutation for $C(CH_4)_3 \rightarrow CH_4$ is net negative in water (–0.9 kcal/mol, calculated; –0.5 kcal/mol, experimental), whereas it is 1.9, 1.3, 3.6, and 3.5 kcal/mol in methanol, DMSO, hydrazine, and CCl_4, respectively. This result shows the unique role of exchange repulsion as an important contribution of the "hydrophobic" nature of water.

Of course, the ability to calculate relative solvation free energies in different solvents enables the calculation of partition coefficients. Essex *et al.*[110,111] have shown how one can study partition coefficients by mutating one solute into another using the cycle below:



By mutating A → B in each solvent, one calculates ΔG_4 and ΔG_3. This can be used to determine the difference in the partition coefficient of molecules $\Delta(\log P)$ partition coefficient P as follows:

$$-2.3RT\Delta(\log P) = \Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4$$

Essex *et al.*[110,1111] have carried out such studies both for rigid and more flexible molecules. As noted above, they have found that the use of multiple conformational fit electrostatic potential charges critical in getting relative partition coefficients for propanol and ethanol that agree reasonably with experiment. Jorgensen *et al.*[112] have shown how one can indeed calculate partition coefficients for a variety of organic molecules between H_2O and $CHCl_3$ in reasonable agreement with available experiments.

3. Free Energy as a Function of Conformation

The relative energy of different conformations of molecules in the gas phase often changes significantly in solution and free energy calculations can be used to study this, as described in refs 57 and 63. One of the prototypal systems for study is *n*-butane and how its relative free energy for gauche and trans conformations changes going from the gas phase to solution. Tobias and Brooks[113] have studied the conformational equilibrium of *n*-butane in the gas phase, in water, and in CCl_4 and have shown that only an all-atom and not a united atom model shifts the equilibrium toward the gauche conformation and does so more in H_2O than CCl_4. Other theoretical studies using united atom models[114] found a shift toward favoring a gauche conformation in H_2O; the reason for this discrepancy between the united atom results in refs 113 and 114 is not clear. Pettit and co-workers have shown the usefulness of analytical theories to reproduce conformational dependent free energies in simple peptide systems.[115,116]

Ha *et al.*[117] illustrate the subtle balance between intramolecular electrostatics and intermolecular solvation free energy terms in their study of the $\alpha \rightleftarrows \beta$ equilibrium in D-glucose. The calculated value of −0.3 ± 0.43 kcal/mol (favoring α) is small, consistent with the magnitude of the experimental value but of the wrong sign (the experimental free energy difference is 0.33 kcal/mol). Nonetheless, this is a deceptively difficult system to accurately simulate, given the large number of hydroxyl group conformers and potential for even small force field inaccuracies upsetting the free energy balance.

Sun and Kollman[118] have shown that one can use cartesian coordinate mapping to calculate solvation free energy differences for conformations that differ significantly in many torsional degrees of freedom. One separately evaluates the intramolecular free energy with gas-phase minimization/normal mode analysis. They validated this approach on 18-crown-6, showing the significant solvation stabilization of the D_{3d} conformation relative to other low-energy conformations.

4. Solvent Effects on Tautomerism, Reduction/Oxidation, Acidity/Basicity, Excited States, and Reactions in Solution

In order to study more general phenomena which involve electronic structure changes, one can combine free energy/solvation calculations with quantum mechanics. Cieplak *et al.*[119] showed that high level *ab initio* calculations could reproduce tautomerism of simple aromatic systems (e.g. 2-hydroxypyridine in equilibrium with its keto tautomer) and then, by mutating one tautomer into another with free energy calculations, one could determine the tautomeric equilibrium in solution. There are often dramatic differences in these tautomeric equilibrium in the gas phase and solution, and these could be of relative in DNA base mispairing. Such a combination of quantum mechanics and free energy calculations have been successfully applied to rationalize other tautomeric equilibria as well.[120,121]

As in tautomeric equilibria, basicity/acidity involves proton movement and a large electronic structure change; thus quantum mechanical calculations are necessary to describe such a process. But again, one can combine such calculations with free energy calculations to determine relative basicities/acidities in solution.[122-125] This is particularly useful in estimating difficult to measure pKa values, such as that of ethane.[122] The accurate calculation of pK_a's of enzyme groups is important for interpretation of enzyme mechanisms. Even "negative" results are considerable importance, as in Merz's demonstration[126] that Glu-106 should *not* be considered as a general acid/base in the mechanism of carbonic anhydrase. Along these same lines, Aqvist has shown how the pK_a of H_2O is perturbed by metal ions, which also has implications for enzyme catalysis,[127] discussed further below.

Richards and co-workers[128,129] have nicely combined quantum mechanical calculations for redox processes with solvation free energy contributions to reproduce and make predictions of aqueous redox properties of quinones.

Duffy *et al.*[130] has combined *ab initio* calculations to determine the amide rotational barrier with solvation free energy calculations to estimate the barrier of isomerization in different solvents. Interestingly, there is an increase in $\Delta G^{\ddagger}$ of ~2 kcal/mol in aqueous solution compared to the gas phase, which the calculations reproduce and "clarify". Debolt and Kollman[131] have used a combination of *ab initio* calculations and free energy calculations to simulate the relative solvation free energies of ground and excited states of C=O groups of formaldehyde and acetone in H_2O, CH_3OH, and CCl_4. The calculations are able to rationalize and reproduce the blue shift of the $n \rightarrow \pi^*$ transition in CH_3OH and H_2O, but not the red shift in CCl_4. To reproduce the latter likely requires that the diffuseness of the excited-state charge distribution and its greater dispersion interaction (than the ground state) with the solvent be represented; the model used treats both ground and excited states as simple point charge models, derived from fitting to the respective electrostatic potentials surrounding the molecule. A more general model for including solvation in the study of excited state phenomena has been presented by Luzhkov and Warshel.[132]

Combining quantum mechanical calculations with explicit solvation models has been implemented in the Warshel group both using *ab initio* pseudopotentials[133] and semiempirical quantum mechanical methods.[134] Particularly, the wide range of applicability of the methods of ref 134 is very impressive. It would be interesting to compare that approach on a similar set of molecules, with AMSOL,[96] which uses a more macroscopic solvation model.

5. Protein Solvation

One can consider protein groups as a "solvent"; thus, one has suggested that charges in protein are often stabilized by helix "dipoles". Aqvist *et al.*[135] using free energy calculations, have shown that the stabilization of charges in barnase and sulfate binding protein come mainly from the groups in the first turn of the helix, not from a "macrodipole". They were also able to simulate the actual perturbation of the pK_a of His-18 and sulfate binding constant in sulfate binding protein is impressive agreement with experiment. Earlier, Daggett *et al.* qualitatively simulated the electrostatic

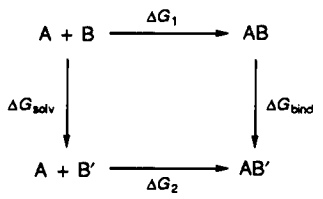
effect of helices on charged groups using simpler models to represent solvent.[136]

The question whether protein cavities provide a suitably attractive environment for water molecules has been studied by Wade *et al.*[137-138] By mutating water → nothing in water ($\Delta A = 6.3$ kcal/mol), the calculations have been able to nicely rationalize the presence of water in one cavity (ΔA for disappearance ~16 kcal/mol) and its absence in another (ΔA for disappearance ~6 kcal/mol).

The solvation of the $[Fe_4S_4SCys]^{2-/3-}$ redox couple in four different environments has been simulated by Langen *et al.* and the available experimental data reproduced.[139] Although some of the differences could be rationalized by the presence of amide groups stabilizing these anions, the crucial role of water penetration in rationalizing the relative redox potential of two structurally similar protein systems was noted; this is a beautiful example of simulations pointing to something that is hard to analyze from the experimental structure alone.

B. Molecular Association

Free energy calculation have been applied in many exciting examples to molecular association in solution. The basic equation describing this is given in scheme 10.



(10)

Consider A as a "host" and B and B′ two guests for this host. One can measure the free energy of association of B and B′ to A using experimental methods and "mutate" B into B′ free in solution and when bound to A using theoretical methods. Since free energy is a state function, the difference between the experimentally measured and calculated free energies should be equal (eq 11).

$$\Delta\Delta G = \underset{(experimental)}{\Delta G_2 - \Delta G_1} = \underset{(computational)}{\Delta G_{bind} - \Delta G_{solv}} \quad (11)$$

Molecular association is a balance between solute and solvent interactions; this is what happens in molecular association described by (10) and (11); by bringing molecules together, we replace solvent–solute interactions by solute–solute interactions. We calculate the solvation difference between B and B′ (ΔG_3) and the host interaction free energy difference (in solution) between B and B′ (ΔG_4).

1. "Small" Organic Hosts

The first application of eqs 10 and 11 to a complex molecular association was the study by Lybrand *et al.*[140] on the host SC24/4H+ with the two guests Cl^- and Br^-. The calculations were successful in reproducing the ~3 kcal/mol preference of the host for Cl^-, which was impressive given the large charges involved. The key issue, which was appreciated before free energy calculations, but which could not be calculated in a quantitative way, was the balance between ΔG_{bind} and ΔG_{solv} in determining the free energy of association of guests to hosts. In the case of the Lybrand *et al.* study, ΔG_{bind} was ~7 kcal/mol and ΔG_{solv} was ~4 kcal/mol, so the ΔG_{solv} modulated $\Delta\Delta G$.

A more general picture, in qualitative agreement with available experiments, emerged from the studies of Grootenhuis,[141] who studied both dibenzo-18-crown-6 (DB186) and dibenzo-30-crown-10 (DB3010) with various cations Na^+, K^+, Rb^+, Cs^+. As is the case with most ionophores, binding free energy is lowest at an intermediate point in the alkali ion series, i.e. DB186 binds K^+ most tightly and DB3010 binds Rb^+ most tightly. This comes about because ΔG_{bind} and ΔG_{solv} are both in the order $Li^+ < Na^+ < K^+ < Rb^+ < Cs^+$, because the smaller the alkali ion, the stronger and more favorably it will interact with any electron donor atom. However, $\Delta\Delta G$ (eq 11) depends on the difference of these two free energies, and whereas the ΔG_{solv} is the same for all aqueous ionophores, the shape of ΔG_{bind} depends on the precise shape of the host. The smaller K^+ ion is calculated and found to be preferred for DB186, given its shape and maximum six coordination, whereas DB3010 allows eight coordination and a more open conformation that favors Rb^+.

Grootenhuis used only a cluster model for the solvent (MeOH), but Van Eerden *et al.*[142] and Mazor *et al.*[143] studied the cation selectivity in 18-crown-6 in water and MeOH using full periodic box boundary conditions. Both calculations successfully reproduced the K^+ selectivity of 18-crown-6, whose size is "perfect" for binding this cation.

Grootenhuis[144] has studied the interaction of neutral molecules MeCN, $MeNO_2$, and $CH_2(CN)_2$ with 18-crown-6. Both normal mode analysis/free energy calculations and free energy perturbation methods could be applied because the simulations were carried out without solvent, and despite this, gave excellent agreement with the experimental association free energies measured in a nonpolar solvent. This level of agreement both validated the force fields and the geometry of interaction. It is clear that the force fields for neutral systems like this can usually lead to significantly more accurate calculated free energies than found for ionic systems.

Anisole spherands and calixspherands have also been subjects of studies of alkali ion selectivities. The calculations of Miyamoto[145] have beautifully reproduced the fact that a calixpherand has a K^+ selectivity over Na^+ and Rb^+ and showed the basis for this in calculating the radial distribution function (rdf) for the cation–oxygen distance. Only in the case of K^+ was the maximum in the radial distribution function for $M^+\cdots O$ identical in the host and in water; for Na^+, the rdf was smaller in water than in the host; for Rb^+, the rdf was larger in water than in the host. The solvent water is able to reorganize to form the optimum $M^+\cdots O$ distance, 2.45 Å for Na^+, 2.70 Å for K^+, and 2.90 Å for Rb^+. In the far less compressible and expandable host, the first peak in the rdf is 2.55 Å for Na^+, 2.70 Å for K^+, and 2.80 Å for Rb^+. Thus, Na^+ cannot form the optimal interaction in the host because it cannot get close enough to the oxygens and Rb^+ is forced to experience repulsion because the host does not expand sufficiently. Thus, these analyses do support the concept of optimal

preorganization/minimal strain to achieve selectivity.

The simulations can contribute by quantitating the compromise between optimizing intermolecular interactions and minimizing strain.[146] This is illustrated beautifully in the free energy calculations on the relative binding free energy of pyridine and pyrazine to Rebek's diacid host in $CHCl_3$.[146] Whereas the former forms an ideal hydrogen bond, the latter cannot form two because of geometrical constraints. The calculations are able to reproduce the relative free energies well *and* give structural insight. By also carrying out free energy calculations on two rigid acetic acid moieties fixed in space, Jorgensen *et al.* were also able to estimate the *maximum achievable* binding preference of pyrazine vs pyrimidine. Further insight on binding of substituted benzenes to Diederich's hydrophobic hosts is offered by the calculations of Jorgensen,[147,148] where the correction of an experimental error followed from the calculational result.

Jorgensen's calculations have led to general insights into optimizing host–guest interactions involving hydrogen bonds. Why does diaminoadenine–uracil, with its three hydrogen bonds, associate so much more weakly that guanine–cytosine, with an equal number of hydrogen bonds? Jorgensen has been able to reproduce the free energy difference and suggest that "secondary" interactions play a significant role in hydrogen bond strengths.[149] The optimal hydrogen bonding strength occurs when all the proton donors are on one molecule and all the acceptors are on another. The absolute and relative free energies for association of nucleic acid base pairs in $CHCl_3$ were calculated in very good agreement with experiment, further validating this approach.[150,151]

Jorgensen and co-workers have also studied association of imides and lactams in $CHCl_3$,[152] as well as quantitatively reproducing the free energy of binding of 9-methyladenine to Zimmerman's impressive molecular tweezer.[153] Both aromatic interactions and hydrogen bonding contributed significantly to the latter very strong association in $CHCl_3$.

Lopez and Kollman[154] have used free energy methods to help understand how different functional groups attached to hemes can enable differential binding between CO and O_2. Surprisingly good agreement with experiment was achieved in most cases.

The importance of preorganization (and, concomitantly, a correct representation of rotational isomers in flexible hosts) was further analyzed by Cannon *et al.*[155] In their example, conformation of the flexible host that was most complementary to the guest had too high an internal energy to bind it as effectively as the more rigid host.

Matsui and Jorgensen[156] have studied the association of a Na^+ near a model metal electrode. They found that the ion was predicted to reside about one solvent diameter closer to the electrode than had been commonly accepted.

Marrone and Merz (MM)[157] have studied the association of K^+ and Na^+ to nonactin, which is a macrotetralide that selectively recognizes K^+ over Na^+. This molecule has both furan and ester C=O groups that interact with the ion. MM used free energy perturbation calculations and have been able to reproduce the K^+ selectivity in methanol (calculated 1.0 ± 0.8 kcal/mol; experimental 2.0 ± 0.7 kcal/mol). By carrying out PMF calculations, the absolute binding free energy for K^+ was also in good agreement with experiment (−6.4 kcal/mol calculated; −5.5 ± 0.6 kcal/mol experimental) and also showed interesting details of the association process. Specifically, the first drop in the PMF curve (Figure 6 in ref 157) occurs when two C=O groups from the ionophore replace methanol coordination; then there is a barrier in the PMF curve before two more C=O groups and two ether –O– replace methanols leading to a final octacoordination of K^+ in nonactin, including two remaining CH_3OH molecules.

In a series of studies on ion complexation of valinomycin, Eisenman and co-workers[158,159] have shown the sensitivity of the ability to reproduce the observed ion selectivity on the partial charge of the C=O group, both using Warshel's MOLARIS and van Gunsteren's GROMOS parameter sets. In both cases, reasonable charges were found that reproduced this selectivity, even with the ionophore–ion complex studied *in vacuo*, albeit with, of course, the relative hydration free energies taken into account. The "bracelet" like structure of the complex presumably permits an *in vacuo* model to be reasonable, just like Grootenhuis[144] showed the reasonableness of the use of cluster calculations in studying 18-crown-6 selectivities.

Aqvist *et al.*[160] have extended the above studies by carrying out free energy calculations in methanol, both for the ions and valinomycin, in the latter case as a function of C=O charge. The experimental selectivity was obtained for an oxygen charge of −0.58, somewhat higher than one would expect from an ester C=O. This supports the usefulness of the inclusion of explicit nonadditivies for quantitative complexation calculations; it will be interesting to see if the model developed by Sun *et al.*[93] can reproduce this selectivity with "no" adjustable parameters.

Gramicidin A is the simplest, well-characterized model for an ion channel, and two recent free energy perturbation calculations have analyzed its properties. Aqvist and Warshel[161] have calculated the free energy of a Na^+ ion in water and in the gramicidin A channel and have found that the ion is ~5 kcal/mol less stable in the channel, consistent with its rapid transport through the channel. They also note the important role of the surrounding membrane (represented as point polarizable dipoles) in leading to reasonable free energies, although as one would infer from the results of Aqvist *et al.*,[160] the resulting free energies would likely be very sensitive to the C=O charges used.

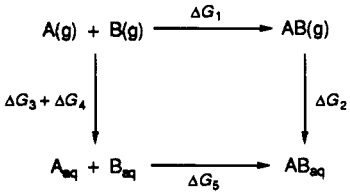
Roux and Karplus[162] have used potential of mean force methods to study the translocation of K^+ and Na^+ along the gramicidin channel. The barrier for translocation between symmetry equivalent positions was determined to be ~5 kcal/mol for Na^+ and ~1 kcal/mol for K^+. They have also calculated the free energy barrier for Na^+ through the entire channel, and found a barrier of ~15 kcal/mol for entering the channel.[163]

The above calculations on host/guest complexes with relatively rigid/small hosts and guests provide critical benchmarks to evaluate the inherent ability of the energy function and representation to lead to quantitative free energies, without the complication of the sampling issue. In that sense, they are complementary

to the simulation of solvation free energies of small solutes described above, where adequate sampling is not an issue either. The generally good quantitative agreement achieved in a wide variety of cases by a large number of labs gives one confidence in the approach and energy functions used.

2. Absolute Free Energies of Association

Cieplak and Kollman[164] showed how one could calculate absolute free energies of association of *N*-methyl nucleic acid bases in solution *both* in the gas phase and solution using cycle (12).



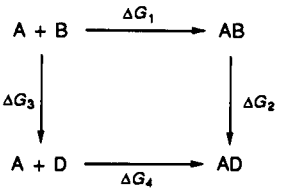
(12)

ΔG_1 is the gas-phase free energy of association, calculated using molecular mechanics energy minimization and normal-mode analysis. The free energies of solvation of A(ΔG_3), B(ΔG_4) and AB(ΔG_2) can be calculated by mutating each of these to nothing in solution, and then one uses eq 13 to calculate ΔG_5, the free energy of association of the bases in aqueous solution.

$$\Delta G_5 = \Delta G_1 + \Delta G_2 - \Delta G_3 - \Delta G_4 \quad (13)$$

Cieplak was able to show that this procedure led to a ΔG_5 for AT (both hydrogen bonded and stacked), GC (both hydrogen bonded and stacked), and AA (stacked) association in water in reasonable agreement with available experiments. Specifically, the stacked forms of AT and GC were calculated to be more stable than the H-bonded, and stacking of AT, GC, and AA bases led to calculated ΔG_{solv} of −1 to −3 kcal/mol, comparable to that found experimentally. Of course, one can calculate ΔG_5 directly with the more time consuming but more accurate potential of mean force approach, as was subsequently done by Dang.[67] He confirmed the Cieplak results for AT association (stacked more stable than H-bonded; ΔG_{assoc} for stacked of −0.8 kcal/mol, compared with the experimental value of −1.2 kcal/mol).

Jorgensen showed that one could calculate the association free energy of two methane molecules in solution more directly[58] using eqs 14 and 15, where all the species are in solution (where A and B are two "real" molecules and D is a "dummy molecule").



(14)

$$\Delta G_1 = \Delta G_2 - \Delta G_3 \quad (15)$$

$\Delta G_4 = 0$ since D is a dummy molecule. Thus, one can calculate the free energy of association ΔG_1 by "disappearing" B when bound to A and, in a separate calculation, when B is free in solution. Reasonable agreement was achieved calculating the ΔG_{assoc} of two methane molecules using eqs 14 and 15 and, secondly, by using PMF methods.

A number of studies have made use of the approach described by eqs 14 and 15. Hermans calculated the absolute free energy of association of Xe gas with myoglobin. In this case, since Xe was not dissolved in solution, only ΔG_3 needed to be calculated.[165] Pranata and Jorgensen used this approach to calculate the absolute free energy of association of nucleic acid bases in nonpolar solvents,[151] Merz calculated the absolute free energy of association of CO_2 to carbonic anhydrase,[166] and Miyamoto calculated the absolute free energy of Rb^+ to a calixpherand.[145] All these calculations led to calculated free energies in quite good agreement with experiment. Lee *et al.*[167] used a related approach to calculate the absolute free energy of association of phosphoryl choline to the antibody MP603, also achieving good agreement between calculated and experimental free energies of association.

PMF approaches have also been used to study $Cl^-\cdots Cl^-$, $Na^+\cdots Cl^-$, and 18-crown-6$\cdots K^+$ [51,60,68,168] association and nonactin/K^+ [157] in aqueous solution. Interestingly, both contact and solvent separated minima have been found in ion–ion associations and, at this point, it appears, but is not definitively established, that the $Cl^-\cdots Cl^-$ association is not purely repulsive, as continuum electrostatic theories would suggest. As noted by Dang *et al.*, the precise properties of the PMF between ions in water is very sensitive to simulation protocol and force field.[59,169] Also, there is not definitive, direct experimental data for comparison. Analogously, the association of two positively charged guanidinium ions is calculated to be surprisingly attractive in water.[170] Further work using nonadditive models is clearly called for, given the likely inaccuracies in effective two body models for systems involving ions.[68,89]

The results on the 18-crown-6/K^+[68] and nonactin K^+[157] association are very interesting since they suggest a minimum displaced from the center of the crown. In each of these cases, encouragingly, the calculated ΔG_{assoc} is in reasonable agreement with available experiments.

3. Protein "Hosts"

There have been a number of applications of cycle 10 to protein–ligand interactions. Here, the calculations are on shakier ground, because the sampling issue is much more problematic for these complexes than for small, rigid organic hosts and guests. The first application of cycle 10 to a protein ligand system was McCammon's perturbation study of substituted benzamidines interacting with trypsin.[171] Mutating a H to a F on the benzene ring has a rather small effect in the theoretically calculated $\Delta\Delta G$, consistent with the experimental result.

Bash *et al.*[31] and subsequently Merz and Kollman[172] used free energy simulations to study thermolysin inhibitors. This was a beautiful system for free energy simulations because both experimental inhibition constants and X-ray data was available for a series of X–PO_2–Y–Z compounds, where X and Z are large, hydrophobic groups, Y is variable, and the PO_2^- group is a ligand for the active site Zn^{2+}. Bash *et al.*[31] were able to reproduce the ~4 kcal/mol tighter binding of Y = NH than Y = O, and Merz[172] studied the Y = CH_2 molecule prior to its experimental study by Bartlett

and co-workers. Contrary to expectations, the Y = CH_2 analog was calculated to bind nearly as tightly as Y = NH, and this prediction was confirmed by the subsequent experiments.[173] These free energy calculations used a +2 charge on the Zn and, to prevent too large movement of the CO_2^- in the binding site, restrained two of the CO_2^- groups.

Merz *et al.*[48] studied the mutation of the tight binding *n*-hexyl sulfonamide inhibitor of carbonic anhydrase, mutating both the hexyl → H and the $SO_2NH^- \rightarrow SO_3^-$. As discussed above, reasonable agreement was achieved for the hexyl → H mutation only if the bond lengths were not "shrunk" as the hexyl group disappeared; it was also surprising that a mutation of $SO_2NH^- \rightarrow SO_3^-$, when corrected for the pK_a difference, led to a quite good agreement with experiment, even though those groups are ligated to the Zn^{2+}. One might expect the very large electrostatic interaction to distort the relative free energies.

Menziani *et al.*[174] also studied carbonic anhydrase inhibition with free energy perturbation approaches. By mutating a 4-H in the benzenesulfonamide inhibitor to a Cl both in water and in the enzyme, one was able to calculate a $\Delta\Delta G$ (eq 11) of −1 kcal/mol, in good agreement with the experimental value of −1.5 kcal/mol.

In subsequent work on carbonic anhydrase, Hoops *et al.* developed a distributed charge model using MNDO ESP's,[175] allowing some of the positive charge from the Zn to be delocalized to the histidine ligands. This charge model should be more robust, general and reliable than the +2 model.

The aspartyl proteases have been the subject of a number of free energy perturbation studies. Rao and Singh[176] have shown the sensitivity of the Asp diad protonation state to the geometry of these two acids. By using quantum mechanical calculations, they demonstrated a very low barrier for proton transfer between the aspartic acids. They also found in free energy calculations tha mutating away the OH group of pepstatin to make deshydroxypepstatin was calculated to lead to a $\Delta\Delta G$ of ~5 kcal/mol, in good agreement with experimental estimates.

A number of groups have applied free energy calculations to HIV protease inhibitors. Reddy *et al.*[177] have compared the binding free energy of two peptidic inhibitors, one a heptapeptide and the other a hexapeptide. This rather large mutation involved the disappearance of a valine group and was calculated to lead to a loss in free energy of 3.3 ± 1.1 kcal/mol, in good agreement with the experimental value of 3.8 ± 1.3 kcal/mol. Reddy *et al.*[178] also used free energy calculations in helping to design new HIV protease inhibitors, using the novel compounds studied by Aguron Pharmaceuticals.

Rao *et al.*[179] studied two classes of inhibitors and carried out three mutations. In the one case where the results could be compared to experiment, reasonable agreement was obtained only if the active site was constrained structurally to remain near the experimental structure.

Independently, two groups, using rather different methodologies,[180,181] have used free energy approaches to calculate the R/S selectivity in JG365, as tight binding hydroxyethylene inhibitor of HIV protease. In the course of these calculations, the protonation state of the Asp-25/125 couple had to be determined, and the two groups went about this in rather different ways. Nonetheless, they both came to the same conclusion about which Asp would be first protonated in the presence of JG365. Interestingly, both found binding free energies in reasonable agreement with experiment. Ferguson *et al.*[180] also predicted the binding free energy of the deshydroxy inhibitor of JG365, which has not yet been studied experimentally.

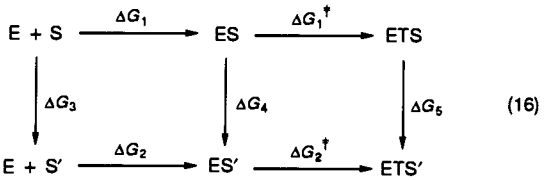
Cieplak *et al.*[182] have carried out a further study of the JG365 inhibitor, considering which amide groups might be mutated to their ethylene or fluoroethylene isostere and lead to more favorable binding. The conceptual basis for this is contained in cycle 10 and eq 11: mutating the polar amide into ethylene or fluoroethylene will likely lead to both ΔG_{bind} and ΔG_{solv} being positive; one seeks locations in the enzyme binding site where the amide group is not interacting particularly strongly with the protein and thus ΔG_{bind} is less positive than ΔG_{solv}. Cieplak was able to find a number of promising candidates for JG365 modifications incorporating either ethylene or fluoroethylene isosters in place of amide bonds.

Aguron Pharmaceuticals has made considerable efforts to design inhibitors of thymidilate synthetase inhibitors and free energy perturbation calculations have played a role in the design process. Reddy *et al.*[183] have calculated the relative binding free energy of TS inhibitors in good agreement with experiment despite the fact that the relative solvation free energies of aldehyde and propargyl inhibitors were surprising and counterintuitive (proparyl better solvated).

Dihydrofolate reductase has been the subject of free energy perturbation calculations by Cummins *et al.*[184] and Brooks *et al.*[185-188] In the former case, the relative free energies of binding of cofactors reduced and oxidized NADP to the enzyme were successfully simulated. Interestingly, the reduced form binds more strongly to the binary complex, and it is calculated to bind much more strongly to the ternary complex involving substrate. A structural interpretation for why this occurs is provided. In the latter case, the relative binding free energies of trimethoprim (with 3-OCH_3 groups) and various substitutions of OCH_3 by CH_2CH_3 were studied, both to the chicken and the *E. coli* enzyme.

In an attempt to design effective inhibitors to the enzyme FDPase, Reddy has applied free energy perturbation approaches to calculate $\Delta\Delta G$ for a number of candidate inhibitors.[189] Reasonable agreement with the available experiments was achieved.

Caldwell[190] has studied the relative binding and transition state stabilities of a number of α-lytic protease substrates using an elaboration of cycle 10, cycle 16, and eq 17 and 18.



(16)

E refers to the enzyme, S and S′ two substrates, ES and ES′ the two Michaelis noncovalent complexes, and ETS and ETS′ the two transition states. One can use

quantum mechanical calculations to create structural models for the transition states. Provided that both substrates do not differ in their mechanism of catalysis or interact differently with the catalytic groups, one can use cycle 16. By mutating S → S′, ES → ES′, and ETS → ETS′ one can relate the calculated free energies ΔG_3, ΔG_4, and ΔG_5 to the experimental binding free energies ΔG_1 and ΔG_2 and catalytic free energies $\Delta G_1^{\ddagger}$ and $\Delta G_2^{\ddagger}$:

$$\Delta\Delta G_{bind} = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \quad (17)$$

$$\Delta\Delta G_{cat} = \Delta G_2^{\ddagger} - \Delta G_1^{\ddagger} = \Delta G_5 - \Delta G_3 \quad (18)$$

By comparing the P_1 substrates Gly, Ala, and Leu, Caldwell *et al.* were able to calculate, in reasonable agreement with experiment, the preference of α-lytic protease for Ala both in terms of binding and transition state stabilization.

Hirono has studied the relative binding free energy of Ribonuclease T_1 inhibitors 2′GMP and 2′AMP and finds, in agreement with experiment, a ~3 kcal/mol preference for 2′GMP.[101] This preference arises from a ΔG_{solv} of ~7 kcal/mol and a $\Delta\Delta G_{bind}$ of ~10 kcal/mol. The fact that ΔG_{solv} is ~7 kcal/mol suggests that a purely hydrophobic site would prefer A over G binding by ~7 kcal/mol; thus, one expects it will be easier for proteins to achieve A specificity rather than G specificity. Given that Glu-46 is likely to play a significant role in the G selectivity, since it forms two H bonds with G (with guanines N1–H and N2–H), Hirono studied the relative free energy of binding of 2′GMP and 2′AMP to the Glu-46-Glu and Gln-46-Ala mutants of the enzyme.[191] The two mutants were calculated to have either a slight A binding preference (Glu-46-Gln) or essentially no G/A preference (Glu-46-Ala). These predictions were qualitatively correct, but this correctness is likely to be fortuitous. That is because Saenger *et al.* showed, by crystallography that 2′AMP in native RNAseT1[100] and both 2′GMP and 2′AMP in the Glu-46-Gln mutant[192] have the base binding in a site removed from the G site in the native protein (presumably the 3′ site for a longer nucleic acid substrate). This illustrates one of the perils of free energy calculations on protein–ligand systems. One must assume a "similar" binding site for the base in 2′GMP and 2′AMP, and Hirono and Kollman were pleased at the partial reversibility of their calculations. However, this assumption of similar binding geometries is not always warranted, as illustrated in this case. The $\Delta\Delta G$ calculated for 2′GMP → 2′AMP was 3 kcal/mol, larger than the experimental value of 2.4 kcal/mol. Thus, the 3 kcal/mol calculated could be "correct", but experimentally, the 2′AMP prefers the site only 2.4 kcal/mol higher in binding free energy than the 2′GMP site. If one knows the structure of one inhibitor protein complex from NMR or crystallography and wishes to predict the binding free energy of another, it is reasonable to expect the calculated $\Delta\Delta G$ to "always" be an upper bound, since the system may not reach the correct geometry of the inhibitor whose structure is unknown.

The strongest known noncovalent binding free energy of a small ligand for a protein is the biotin–avidin interaction. The related biotin–steptavidin interaction, for which the crystal structure was solved, also has a very high binding affinity of $\sim 10^{14}$. Free energy calculations were used both to calculate the relative binding free energy of biotin, thiobiotin, and iminobiotin, in which the ureido O was replaced by S and NH.[76] For the biotin → thiobiotin mutation, ΔG_{solv} was 8 kcal/mol, ΔG_{bind} was ~12 kcal/mol, leading to a $\Delta\Delta G_{calc}$ of ~4 kcal/mol, in good agreement with experiment. For biotin → iminobiotin, ΔG_{bind} was only 1–2 kcal/mol, but ΔG_{solv} = −5 kcal/mol, leading to a $\Delta\Delta G$ = 7 kcal/mol, again in good agreement with experiment. It is interesting that both thiobiotin and iminobiotin lose binding strength relative to biotin in dramatically different ways—thiobiotin interacts much more weakly with the enzyme than it gains in ease of desolvation whereas iminobiotin interacts with the enzyme almost as well as biotin but is much harder to desolvate.

The absolute free energy of binding of biotin for steptavidin was calculated using cycle 14. The partial charges of biotin were mutated to zero, and then the van der Waals parameters were mutated to zero, both in water and when bound to steptavidin. During the mutation in the binding site, the seven hydrogen bonds in the binding site were restrained, in order to prevent biotin from moving too far from its binding location. The restraints were not necessary if the perturbation was carried out relatively rapidly (~100 ps), but one could not get reasonable results running the free energy calculation in the reverse direction, probably because waters are trapped in the binding site. In any case, five independent forward free energy calculations all led to a $\Delta\Delta G$ of ~20–22 kcal/mol, in reasonable agreement with the experimental value of 18.3 kcal/mol. More significantly, the dominant contribution to the $\Delta\Delta G$ was van der Waals, not the electrostatic/polarization that had been suggested by the crystallographers who had solved the structure.[193] The fact that there were three protein hydrogen bonds, all out of plane to the ureido oxygen, led to the suggestion that the enzyme bound tightly to this oxygen because of the resonance structure which placed a formal negative charge on the oxygen. However, the molecular dynamics simulation of biotin in solution found the same number and out of plane character for water O—H···O=C hydrogen bonds. The large van der Waals contribution to the binding comes from the four tryptophans lining the biotin binding site. The calculated results led to the suggestion that van der Waals "preorganization" was likely as important as or more important than electrostatic in optimization of protein–ligand binding free energies.[76]

Other reassuring features of these calculations were (1) the part of the active site which moves most between the apo and biotin bound enzyme is found both experimentally and in the calculations to be residues 46–50 and (2) the apo enzyme crystal structure has 5–6 localized water where biotin bound and a similar number were found in the calculation at the end state when biotin had disappeared. The fact that the biotin binding site was not fully occupied by water could be understood from the free energy calculations of Wade and McCammon,[137,138] who compared the free energy of disappearing water in protein cavities with the free energy of removing a water molecule in liquid water. Unless

the former is larger than the latter (> ~6 kcal/mol), the theory suggests that it is more favorable to have an empty cavity. This is not unreasonable for very hydrophobic cavities.

Given the approximations/constraints required in the biotin/avidin calculations, Miyamoto decided to study a completely different protein–ligand binding interaction to use as a control, *N*-acetyltryptophanamide (NATA) interacting with α-chymotrypsin with an experimental dissociation free energy of +5.2 kcal/mol.[194] The calculation on NATA α-chymotrypsin led to a calculated dissociation free energy of ~9 kcal/mol; as in biotin–avidin, we expect the calculated $\Delta\Delta G$ to be larger than observed. It is interesting that the calculated $\Delta\Delta G$ overshoots experiment similarly to biotin–steptavidin and that this free energy is electrostatic, not van der Waals, dominated; it is also encouraging that the relative ligand association free energies calculated for the two protein systems (11–13 kcal/mol) are reasonably consistent with the observed relative free energies (13 kcal/mol).

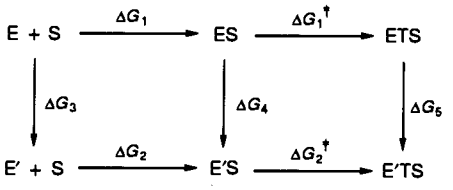
In a rather unusual application of cycle 10, Hansen and Kollman[195] applied it to a case where the protein structure was not known, but must be inferred: the relative binding free energy of two inhibitors of adenosine deaminase. These differ only in the replacement of an OH by an H, but the OH inhibitor binds ~10 kcal/mol more strongly than the H inhibitor. No simple protein active site, using only amino acid side chains, could rationalize this binding free energy difference. It was thus encourating that the subsequently determined crystal structure[196] found an "unexpected" Zn^{2+} in its binding site, which could rationalize that very large selectivity for the OH over the H inhibitor.

In a very interesting model study, Lau and Pettit have suggested an approach to decompose association free energies for interaction into components and have applied this to viral coat protein–inhibitor interactions.[197] Pearlman has developed a related model approach, the use of free energy derivatives; he has applied it to model systems[198] and Cieplak *et al.* to protein–ligand interactions.[199]

To our knowledge, there have been no published papers on relative ligand binding free energies to DNA, presumably because of the lower resolution structural data and greater difficulties in accurately describing the electrostatic properties of such systems. However, Singh *et al.*[200] have recently applied free energy approaches to the relative binding of 2:1 complexes of distamycin (Dis) or its pyrole → imidazole analog (Im) to the DNA minor groove. The calculations were carried out with the NMR structural information, but without knowledge of the free energies. They reproduced the order of binding free energies to DNA and the surprising fact that the mixed complex DNA·Dis·Im was more stable than DNA·Dis·Dis, but DNA·Im·Im less stable. The quantitatively calculated free energies, however, were off by as much as 1.5 kcal/mol.

C. Sequence Dependence on Ligand Binding and Catalysis

One can write a cycle (19) to describe the effect of protein or nucleic acid sequence on binding and catalysis:



(19)

$$\Delta G_{bind} = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \quad (20)$$

$$\Delta G_{cat} = \Delta G_2^{\ddagger} - \Delta G_1^{\ddagger} = \Delta G_5 - \Delta G_4 \quad (21)$$

Rao *et al.*[201] studied the effect of mutating Asn-155 to Ala on the binding and catalysis of a typical substrate by subtilisin to study the role of the oxyanion hole in these processes. The "oxyanion hole" in trypsin is composed of two backbone NH groups, whereas that in subtilisin uses a backbone NH and the side chain of Asn-155. The calculations by Rao *et al.* were done prior to knowledge of the free energy effects of the Asn → Ala mutation, although related mutations had been reported. The results ($\Delta\Delta G_{bind}$ = 0.1 ± 0.8 kcal/mol and $\Delta\Delta G_{cat}$ = 3.4 ± 1.0 kcal/mol) were in encouraging agreement with the experimental values of 0.4 and 3.8 kcal/mol, respectively. Independent calculations by Hwang *et al.*[202] used a more sophisticated approach explicitly considering quantum mechanical effects on the transition state, but calculated similar free energies.

One of the exciting results to emerge from the calculations by Rao *et al.*[201] was the fact that Thr-221 formed an unsuspected interaction with the oxyanion transition state. This led to a set of collaborative experimental/theoretical studies at Genentech and UCSF and showed that the presence of Thr-221 had a ~20-fold contribution to the stabilization of the transition state relative to alanine.[53] Interestingly, the calculations showed that the effects of Asn-155 and Thr-221 were additive, despite the hydrogen bond between the Thr-221 OH and the Asn side chain C=O in the isolated enzyme's crystal structure.

Gago and Richards[203] have used the first part of the above cycle to study the effect on the relative free energy of binding of netropsin to $(ICIC)_2$ vs $(GCGC)_2$ sequences of mutating the G residues, with their 2-NH_2 groups protruding into the minor groove, into I, which lacks the 2-NH_2 group. The preference for I was calculated to be ~4.4 kcal/mol, in excellent agreement with the experimental value of 4.0 kcal/mol.

The sequence specificity of daunomycin for an AT base pair neighboring the intercalation site in daunomycin binding to the DNA fragment $d(CGTACG)_2$ was reproduced by Cieplak *et al.* Interestingly, the preference for CG at the intercalation site was less clear for daunomycin, but was found for 9-NH_2 acridine.[204]

One can use the above cycle to examine the effects of protein mutation on oligomerization where both E and E′ are proteins. Karplus and co-workers have done so for the very important mutation that leads to sickle cell hemoglobin (Glu-6 → Val).[205] Consistent with experiment, the mutation was indeed calculated to stabilize the polymerized state compared to the nonpolymerized.

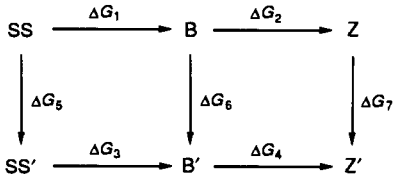
Wade and McCammon[206,207] have examined the relative binding free energies of an antiviral compound WIN53338 to a viral coat protein and its V188L mutant. They found it essential to consider different rotational isomeric states of the side chains of the protein to

calculate free energies in reasonable agreement with experiment. They provide a careful, detailed analysis of the difficulties in correctly dealing with the significant computational complexity of rotational isomerism.

Komeiji *et al.*[208] have studied the relative binding free energy of tryptophan to Trp repressor protein for native and mutant (Ser-88-Cys) Trp repressor. The calculated relative binding free energy of these proteins was 1.6 ± 0.3 kcal/mol, in excellent agreement with experiment (1.7 kcal/mol). Ser-88 forms a hydrogen bond to tryptophan, and one might expect that a Ser-88 → Cys mutation would weaken this hydrogen bond. By carrying out free energy component analysis, Komeiji *et al.* were able to show that electrostatic effects largely cancelled and that it was likely the greater van der Waals repulsion in the Trp–repressor–Trp complex of the mutant when the larger sulfur replaced oxygen that was the cause of the reduced free energy of binding of the mutant. This, interestingly enough, is analogous to the explanation for the reduced stability of a Thr-157 → Val mutant in T4 lyzozyme by Dang *et al.*,[209,210] where the larger CH_3 causes more repulsion in a constricted site in the protein.

D. Sequence Dependent Stabilities

One can use a cycle such as 22 and 25 to describe sequence dependent nucleic acid (22) or protein (25) stabilities, where Z ≡ ZDNA, B ≡ BDNA, SS = single-stranded DNA:



(22)

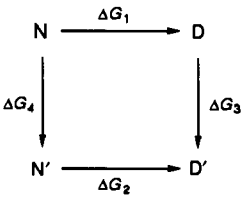
Since one does not know the pathway and it is certain

$$\Delta G(B\text{–}Z) = \Delta G_4 - \Delta G_2 = \Delta G_7 - \Delta G_6 \qquad (23)$$

$$\Delta G(B\text{–}SS) = \Delta G_3 - \Delta G_1 = \Delta G_6 - \Delta G_5 \qquad (24)$$

to be complicated, for the B → Z transition, it is hard to calculate ΔG_2 and ΔG_4 directly. One can measure the sequence dependent tendencies to undergo the B → Z DNA transitions in sequences B (ΔG_2) and B′ (ΔG_4) and compare their difference with the calculated free energies to mutate B → B′ (ΔG_6) and Z → Z′ (ΔG_7). One would also like to understand the relative free energies for nucleic acid sequences to undergo the double strand → single strand transition as the temperature is raised. One can measure the free energy for this transition in sequence B (ΔG_1) and B′(ΔG_3) and simulate it by calculating the free energy to mutate B → B′ in the double-stranded B form (ΔG_6) and in the single-stranded form (ΔG_5).

A similar cycle can be written for studying protein stabilities (see eq 25). One can determine the free energy for denaturation of sequence N (ΔG_1) and N′ (ΔG_2) experimentally and relate this difference to the differences in the free energies for mutating N → N′ in



(25)

the native protein structure (ΔG_4) and in its denatured form (ΔG_3) (eq 26).

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_3 - \Delta G_4 \qquad (26)$$

One must emphasize that the successful theoretical applications of eqs 24 and 26 are particularly difficult because one does not have a well-defined structure for single-stranded DNA or denatured proteins.

Pearlman[211] has studied the B → Z potentiation due to replacing cytosine by 5-Me cytosine using free energy perturbation calculations and cycle 22 both *in vacuo* (with implicit solvation) and in solution (using explicit solvation). Earlier molecular mechanics studies had suggested that the Z potentiation of 5MeC could be rationalized by intranucleic acid interactions, but the crystallographic structure also suggested a hydrophobic contribution, since the 5-Me group was more "buried" in Z than B DNA. Pearlman's calculations were able to beautifully rationalize these results, since the calculated $\Delta\Delta G$ (~0.4 kcal/mol per base pair stabilization of Z by 5MeC compared to C) was in excellent agreement with experiment and, by comparing *in vacuo* and solution simulations, was found to be approximately half due to intramolecular effects and half due to solvation effects.

Dang *et al.*[212] and Singh *et al.*[213] have studied the Z phobicity of an AT base pair compared to a GC by mutating d(CGCGCG)$_2$ into d(CGTGCG)·d(CGCACG) in B and Z DNA. The calculations were able to qualitatively reproduce this tendency, suggest it was mainly intramolecular, and elucidate that these components contained predominantly contributions from the base pair itself and the neighboring bases.

Ross *et al.*[214] have used free energy calculations to nicely rationalize why 5Br substitution stabilizes ZDNA but not ZRNA. In the latter case, of course, the A structure, not the B, was the reference state in a cycle like 22.

Hausheer *et al.*[215] have studied the relative association free energy of 5-Me-cytosine vs cytosine containing single strands to associate with Watson–Crick duplexes and form triple strands. The relative solution pK_a of 5-Me-cytosine and cytosine are well reproduced by the simulation and the calculations suggest a large stabilization of the triple strand by the C → 5MeC substitution.

Ferguson and Kollman[216] and Hausheer *et al.*[217] have applied eq 19 to study the relative stabilities of *R* and *S* methyl phosphonate isomers of DNA "analogs". These compounds could have use in antisense therapy, and the goals of the studies were to simulate and try to understand the different stability of these isomers. Both studies assumed no stability difference in the single stranded forms ($\Delta G_5 \sim 0$), which is probably reasonable given the simplicity of this mutation and the fact that one is studying stereoisomers, albeit they are diastereomers and not enantiomers because of the remaining assymetric centers in the nucleic acid. Both found the

R stereoisomer more stable, independent of base sequence. This is not fully consistent with some experimental data but was rationalized by Ferguson on the basis of "fraying" away from a canonical B structure at the ends leading to a reversal of the calculated/experimental free energies. In addition, Hausheer *et al.*[217] also studied a T → U mutation in a DNA duplex octamer and found that this mutation destabilized the octamer.

Protein stabilities using eq 26 have been the subject of a number of studies. Dang *et al.* have been able to reproduce the observed ~2 kcal/mol destabilization of residue 157 in T4 lyzozyme upon mutation of Thr to either Val[209] or Ala.[210] Component analysis clearly showed that, surprisingly, the Thr → Val destabilization was a van der Waals dominated difference, due to the methyl group being forced to occupy the position of the OH group in the native protein. The Thr → Ala mutation, also surprisingly, seems dominated by the electrostatic energies. This is surprising since the Thr → Val mutation suggested a similar hydrogen bond strength in the native protein and the tripeptide model for the denatured. A dominance of the van der Waals term was also found in the free energy calculations of Yamaotsu *et al.*[218] on G88V and A69T mutants of staph nuclease. Zhang and Hermans[219] have been able to reproduce the added stability of a coiled-coil structure toward denaturation by mutating Ala → Val, Leu at key hydrophobic contact positions.

Tidor *et al.*[220] also studied the relative stability of the native and R96H mutants of T4 lyzozyme. The calculations were consistent with the experimental destabilization of this mutation and highlighted the critical role of the greater solvation stabilization of the denatured histidine cation in this regard. As clear from the studies of Bash *et al.*,[31] the more delocalized the cation (e.g. Arg vs Lys), the easier it is to desolvate and the smaller the free energy to move it to the interior of the protein.

Prevost *et al.* studied the Ile → Ala mutation in barnase[55] and were able to reproduce the considerable destabilization of the protein due to this mutation. There was a surprisingly large intraresidue contribution calculated for the destabilization. Yung-yu *et al.*[221] have carried out protein stability calculations on the mutation in subtilisin. Even the protein stability was reasonably reproduced, but they concluded that free energy calculations could not reliably *predict* protein stability.

Experimental studies have shown a surprising intrinsic stability of isolated peptides in solution; the relative stability of different amino acids in a helix relative to a "random coil" has been the subject of careful and exciting free energy calculations by Hermans and co-workers.[72-74] Relative to Ala as a reference, Aib stabilizes helices, but most other residues (Gly, Val, Pro) destabilize it. Hermans has been able to directly calculate the conformational entropy contribution to helix stability as well as the stabilizing or destabilizing interactions of residues in the helical state. For example, the instability of Gly in the helix is mainly due to its greater backbone conformational free energy (entropy) in the random coil state, whereas the instability of Val comes both from its C_β branding and van der Waals repulsion in the helix state as well as reduction in side-chain entropy. The agreement of calculated relative (to Ala) stabilities with experimental stabilities is, in general, excellent.

Brooks and co-workers have also carried out a number of interesting free energy calculations on small peptide models of protein building blocks. For blocked dipeptides of Ala-Ala and Ala-Pro, they[64] showed that the extended form was much more stable than the reverse term, mainly because of more favorable interactions with solvent for the former. In tetrapeptides, they[63] showed that the α helical structure for Ala is only slightly less stable than the extended structure, but the Val tetrapeptide is significantly less stable, consistent with Hermans calculations[72-74] and experiment. They have used free energy methods to examine the solvation effect on the conformational profile of the alanyl dipeptides.[222] Finally, they[223] have described the surprising stability of a model β sheet structure, where a pair of hydrogen bonds coming together with two peptide fragments in an extended conformation being much more stable than a single hydrogen bonded amide dimerization.[65]

E. Combining Quantum Mechanical Calculations with Free Energy Calculations

It is an exciting and important challenge to be able to calculate the free energy of reactions in solution and in enzyme active sites. This is difficult because *ab initio* quantum calculations require so much computer time for the evaluation of their energies and gradients that it is as yet impractical to evaluate these the 10^4–10^6 times required for fully coupled QM/MM MD/FEP calculations. Promising approaches to do this coupling with AM1 semiempirical MD theory have emerged,[224] but two types of approaches currently dominate the horizon. Jorgensen has studied a number of interesting prototypal organic reactions by carrying out high-level *ab initio* calculations in the gas phase, creating a limited number of geometries along the reaction coordinate and then evaluating the solvation free energy of these snapshots using Monte Carlo simulations. The first study in this series was the simple exchange reaction $CH_3Cl + Cl^{*-} \rightarrow CH_3Cl^* + Cl^-$ where the solvation effect on the reaction is quite dramatic. There is an ion dipole minimum in the gas phase and a barrier of ~10 kcal/mol; in aqueous solution the ion dipole minimum disappears and the barrier increases to ~25 kcal/mol.[225] Calculations and experiments agree very well on the magnitude of this barrier.

Recently, Jorgensen has used the same approach to study the Diels–Alder reaction, which is speeded up substantially (by ~10^3) in aqueous solution.[226] The calculations reproduce this and show it is approximately half hydrophobic and half due to the stronger hydrogen bonding in the transition state. The above reactions have also been studied in other solvents and related to experiments; for example, in organic solvents, the barrier for the $CH_3Cl + Cl^-$ exchange reaction is not as high and there is a residual ion–dipole complex formed.[227] Using an analogous approach,[228] Severance and Jorgensen studied the solvation effect on the Claisen rearrangement.

Bash *et al.*[229] and Karplus *et al.*[230] have used a combined semiempirical AM1/FEP approach to study proton transfer in the enzyme triose phosphate isomerase (TIM). The formation of an imidizolate in

the mechanism was supported in these calculations. Daggett *et al.*[50] have applied free energy component analysis to analyze the effect of the Glu-165 → Asp mutation in this enzyme and have noted the critical role of Lys-13 and His-95 in the reduction of catalytic rate upon E165D mutation. Analogous to Jorgensen's calculations is the study of the reaction $CO_2 + OH^- \rightarrow HCO_3^-$ by Peng and Merz.[231] This reaction is without barrier in the gas phase and has a significant barrier in aqueous solution, due to the better solvation of the reactants. Zheng and Merz have studied the mechanism of carbonic anhydrase, of which the $CO_2 + OH^- \rightarrow HCO_3^-$ reaction is a model.[232] The combination of *ab initio* calculations followed by environmental FEP was useful in distinguishing two mechanisms and leading to a barrier for hydration of CO_2 in reasonable agreement with experiment.

Two other promising and exciting methods for studying chemical reactions in solution should be mentioned at this point. Cramer and Truhlar have combined semiempirical MO theories with aspects from implicit solvation models[96] to create an approach which enables study of chemical reactions in solution.[96,233]

Gao has, in an approach similar to that of Field *et al.*,[224] combined semiempirical MO theories with explicit solvation models, employing Monte Carlo methods to carry out free energy evaluations within such models.[234] The solvation effect on the rotational barrier in dimethylformamide[235] on the relative free energy profile for C–O bond rotation in acetic acid[236] and the PMF for association of tetramethylammonium (TMA) both with benzene and Cl^- [237] have been successfully studied with such an approach. The studies quantitatively reproduce the greater free energy of association of TMA with benzene than Cl^- and show the dramatic effect of solvation in stabilizing the trans conformer of acetic acid.

Many of the studies enzymatic reactions studied by coupled QM/MD free energy calculations have used Warshel's approach, which uses a simple valence bond quantum mechanical approach to represent the reactive part of the system. This enables the QM and MM/MD parts of the reaction to be coupled efficiently; the simplicity of the quantum mechanics is compensated for by calibrating the reaction in solution and then using the same reactive Hamiltonian elements in the enzyme.

Warshel and co-workers have studied the serine proteases, trypsin[238] and subtilisin,[202] staphyloccocal nuclease,[239,240] dehydrogenases,[241,242] electron-transfer reactions,[243,244] p21-catalyzed GTP hydrolysis,[245] and carbonic anhydrase[246–248] using this methodology. Particularly impressive is the ability of the simulations to reproduce the catalytic power of staphyloccocal nuclease and carbonic anhydrase and the role of its divalent ions, Ca^{2+} and Zn^{2+}, in achieving this. These cations must both stabilize anions and allow them to react. The subtle balance between these abilities leads Ca^{2+} and staph nuclease to lower the free energy of activation for phosphate hydrolysis by $\sim 10^{15}$ over the solution reaction and allows carbonic anhydrase to be very efficient, with its reaction diffusion controlled.

V. Summary

The goals of any numerical theoretical approach applied to chemical phenomena are to (a) calculate numerical values that agree with experiment, (b) provide mechanistic insight into the phenomena, and (c) be predictive. Free energy calculations have often reached one or more of these goals, as we now illustrate.

As noted above, there have been many papers in the literature where ΔG values have been calculated that agree "reasonably" with experiment, (within 2 kcal/mol), in good agreement (within 1 kcal/mol) and in excellent agreement (within 0.5 kcal/mol). One of the most impressive recent examples is the work of Jorgensen and Nguyen,[148] where the relative binding free energies of four aromatic molecules to a relatively rigid, hydrophobic host were calculated to within experimental error. This example is one in which the host and guest are relatively rigid, and so one expects the sampling problem not to be too severe. The excellent agreement then quite clearly reflects well calibrated energy functions for this system.

What do we mean by "mechanistic insight"? This refers to a qualitative understanding of a phenomena, not dependent on numerical calculations. If these numerical calculations reproduce the important aspects of the experimental values, one has confidence in deriving mechanistic insight into the system. For example, the surprising biphasic free energies of the octaspherand found by Cram and co-workers[249] has been reproduced by Bayly,[250] and the structures emerging from this suggested a crucial role of water molecules from the mixed $H_2O/CHCl_3$ solvent leading to its dual ion selectivity ($Li^+ < Na^+ < K^+$ and $Na^+ < Rb^+ < Cs^+$). Thomas thus carried out the same free energy calculation in methanol and found that the dual-ion selectivity was *not* present in that solvent.[251] Hopefully experiments will test this prediction.

The fact that the absolute free energy of biotin binding to avidin was calculated[76] in qualitative agreement with experiment suggested that this interaction, among the strongest known noncovalent ligand–macromolecule interactions, was van der Waals dominated, in contrast to the interpretation made on the basis of only examining the structure of the complex. More generally, the subtle balance between solvation and binding free energies in many applications of cycle 10 has led to many qualitative insights into ligand design.

Sometimes the simulations have been genuinely predictive, as in the two organic host/guest systems just noted, in Merz and Kollman's quantitative prediction[172] of the tight binding of a new thermolysin inhibitor, in Singh *et al.*'s[200] prediction of the relative free energy of distamycin analogs binding to DNA, and in a number of other cases reviewed above.

What prevents the free energy approaches from reaching these three goals in more chemical systems? As note above, the two "roadblocks" are (a) inaccuracies in the energy function/representation and (b) limited sampling of the important low-energy conformations.

It is likely that improvements in the energy function and representation will continue, and providing that sampling is not a problem, better representation of long- and short-range electrostatics and nonadditive effects will likely lead to more systems being represented as accurately as that of the Jorgensen/Nguyen[148] example noted above. We should also emphasize the sensitivity of free energy calculations to force field parameters,[75] enabling such calculations to be very useful in the

evaluation of such parameters.

However the major difficulty is that most systems, especially macromolecular ones, have too many local minimum to consider even in nanoseconds of simulation time. In those cases, one relies on X-ray crystallography or NMR to "restrain" the calculations to the relevant areas of conformational space. One must make the assumption that all the ligands bind in approximately the same area of conformation space. An example where this was shown *not* to be the case was the study of the relative binding free energy of 2'GMP and 2'AMP to ribonuclease T1, where only the 2'GMP binding geometry was known when the calculations were done.[101] As noted above, the calculated $\Delta\Delta G$ of $\sim$3 kcal/mol (guanine binding tighter) was in reasonable agreement with experiment, but the binding geometry of the base in 2'AMP was very different from that of 2'GMP[91] and "assumed" in the free energy calculation. Notwithstanding this uncertainty/inaccuracy, it is reasonable to expect that, if one calculates a $\Delta\Delta G$ for two ligands where one has X-ray/NMR data only for the first, the calculated $\Delta\Delta G$ is very likely to be an *upper bound* to the true $\Delta\Delta G$. This is because the "true" conformation for the second ligand will be lower in free energy than the calculated one.

Of course, there are cases where it is currently impossible to even determine the structure to use, such as the denatured protein in cycle 25. Thus, protein stability calculations can usually realistically reach goals (1) and (2) and not be predictive, as is clear from the studies of Dang *et al.*,[209,210] Tidor *et al.*,[220] and Prevost *et al.*,[55] as reemphasized by Yung-yu *et al.*[221] On the other hand, the quantitative propensity of various hydrophobic side chains to form *helices* have been successfully *predicted* by Hermans *et al.*,[73] employing the assumption of a random coil for the nonhelical reference state.

Thus, the prospect is bright for continued interesting applications of free energy calculations to molecular systems in condensed phases. The opportunity to make meaningful contact with experiment is an exciting one that these calculations offer; even when the agreement with experiment may be fortuitous, the model can be continued to be evaluated/tested on other analogs to push it to its limit.

Among the most exciting prospects for free energy calculations have been to make *predictions* in drug design of macromolecular inhibitors. This has been hampered by the fact that only small mutations can be most reliably predicted, and those can often be synthesized faster than the free energy calculated. Ever increasing computer power and such techniques as free energy derivatives,[198,199,252] where one can consider many and more complex structural changes all from the results of one simulation, lead one to hope that free energy calculations will ultimately be of more technological use. But notwithstanding that, it is clear from the above that they have been giving and will continue to give exciting and powerful insights into noncovalent interactions of complex molecules in solution.

Enhancing the power of free energy methods will be the ability to harness massive parallelism in free energy calculations. DeBolt *et al.*[253] have gone a long way to showing how almost perfect parallelism can be achieved in free energy calculations on a message passing NCUBE computer. The problem of different starting geometries for different windows and values of λ was solved using a RAM (Rapidly Accessible Manifold) of starting states. This parallel version of AMBER GIBBS was applied to five problems of increasing complexity and difficulty. The last two problems, a potential of mean force calculation of K^+···salinomycin$^-$ and the relative solubilities of benzene and phenol in water-saturated octanol can only converge with very long simulations of nanoseconds total duration. As the individual processors in such parallel computers become more powerful (e.g. the CRAY T3D apparently will have 500 DEC ALPHA processors, each roughly the speed of a CRAY YMP), more ambitious applications of free energy methods will become possible.

Acknowledgments. We are grateful for research support from the NIH (CA-25644 and GM-29072) and the NSF (CHE-91-13472). We acknowledge the helpful comments of Piotr Cieplak on the manuscript.

Note Added in Proof

Since the completion of this manuscript, a number of new or inadvertently missed articles have been discovered. They are discussed in the order in which they would appear in the review:

Section III.I. Tidor[254] has shown how one can combine molecular dynamics methods to sample Cartesian space and Monte Carlo methods to sample chemical perturbation space on a model system ($Br^- \rightarrow Cl^-$). Such a method may be of use in ligand design.

Section III. J. Ramnarayan *et al.*[255] have shown the dependence on calculated free energies for a wide variety of processes including nonadditive polarization energies. In most cases presented, the calculations with polarization were in worse agreement with experiment, suggesting that the additive part of the force field needed to be reparameterized before polarization was turned on.

Section IV.A. Boudon and Wipff[256] have analyzed the model dependence of the solvation free energies of NH_4^+ in water. Lowis *et al.* have calculated the potential of mean force for an ethyl-substituted imidazole in water.[257] Elcock and Richards have developed a new protocol for mutating purines $\rightarrow$ pyrimidines in water.[258] Reasonable agreement among the purines and among the pyrimidines with the calculations of Bash *et al.*[31] was found, but the relative pyrimidine/purine solvation free energies differ significantly. It is not clear at this point whether force field or simulation protocol is the reason for the difference.

Section IV.B. Gao[259] has studied the PMF for benzene association in superheated water and has shown that the association constant is much smaller at high temperature than in room temperature water. Auffinger and Wipff have studied the relative free energies of M^+ association with C222 in water and methanol.[260]

Section IV.C. The difference in free energy of binding arabinose and fucose by L-arabinose binding protein and its M108L mutant has been calculated in good agreement with experiment by Zacharias *et al.*[261] In this case, the relative free energies for binding the two ligands to the two proteins directly was calculated,

without calculating the relative binding vs solvation of the ligands themselves. Prodhom and Karplus studied the effect of mutation of Asp to Asn in EF hands on ion binding.[262]

Section IV.D. Li and Hermans[263] have studied the effect of mutation on the stabilities of coiled coils and have nicely illustrated the key role of dispersion/hydrophobic effects in determining their stability. Yan *et al.*[264] have studied the free energy of β turn formation in the system CH_3CO-L1-L2-$NHCH_3$, where L1 and L2 are Gly, L-Ala, or D-Ala. Interesting differences in stabilities of both β and inverse β turns was obtained.

Section V. Van Gunsteren has analyzed free energy applications to protein systems.[265] He is quite critical of some applications of free energy calculations to proteins as well as the use of free energy component analysis. I disagree with the harshness of some of his criticism. As noted above, the single largest limitation in the application of free energy calculations to macromolecules is the sampling issue. Probably no protein system meets the criteria of full sampling possible with smaller systems. Given that, are calculations worth doing which do limited sampling or deliberately constrain the system in conformational regions of interest? In my opinion, each application to proteins should be judged on whether it gives mechanistic insight or useful predictions, not on the length of the simulation per se, although the standards for what can/should be done as computers become more powerful will change. We ourselves are guilty (e.g. ref 30) of one of the "sins of inconsistency" mentioned by van Gunsteren: studying the two legs of the thermodynamic cycle (ΔG_3 and ΔG_4 in cycle 16) with different protocols. For the solutes in water (ΔG_3) we used periodic boundary conditions, but for the protein, an 18 Å solvated moveable zone around the ligand (ΔG_4). Given that it was (and still is) rather unpractical to employ full periodic boundary conditions on the protein–ligand system to calculate ΔG_4, the alternative is to use a spherical shell to calculate ΔG_3. We would suggest that except for ligand perturbation with charge changes, the difference between the box and sphere with identical cutoffs would be significantly smaller than other inherent errors or uncertainties (e.g. sampling) being made in the simulations.

The other comment of van Gunsteren that we would take issue with is his assertion that to break the total free energy into components is "rather meaningless". Unlike the total free energy between states, the component free energies are path dependent. However, well-defined paths can be studied and each such study can be judged on its own merits of whether "transferable" mechanistic insight into the process can be derived.

References

(1) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.
(2) Mezei, M.; Beveridge, D. L. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 1.
(3) Van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992.
(4) Van Gunsteren, W. F. *Protein Eng.* **1988**, *2*, 5.
(5) McCammon, J. A. *Curr. Op. Struct. Biol.* **1991**, *1*, 196.
(6) Straastma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407.
(7) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184.
(8) Kollman, P. A.; Merz, K. M. *Acc. Chem. Res.* **1990**, *23*, 246.
(9) Reynolds, C. A.; King, P. M.; Richards, W. G. *Mol. Phys.* **1992**, *76*, 251.
(10) Richards, W. G. *Pure Appl. Chem.* **1993**, *65*, 231–234.
(11) Richards, W. G. *Proc. R. Soc. Edinburgh Sec. B, Biol. Sci.* **1992**, *99*, 105–111.
(12) *Computer Simulation of Biomolecular Systems*; Van Gunsteren, W. F., Weiner, P. K., Eds.; Escom: Leiden, 1989.
(13) Brooks, C. L.; Karplus, M.; Pettit, B. M. *Adv. Chem. Phys.* **1988**, *71*, 1.
(14) Van Gunsteren, W. F.; Mark, A. E. *Eur. J. Biochem.* **1992**, *204*, 947–961.
(15) Warshel, A.; Aqvist, J. *Chem. Scr.* **1989**, *29A*, 75–83.
(16) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley-Interscience: New York, 1991.
(17) Jorgensen, W.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.
(18) Van Gunsteren, W. Private communication.
(19) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(20) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(21) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(22) Singh, U. C.; Weiner, P. K.; Caldwell, J.; Kollman, P. *Amber 3.0*, University of California: San Francisco, 1986.
(23) Fleischman, S. H.; Brooks, C. L. *J. Chem. Phys.* **1987**, *87*, 3029.
(24) Zwanzig, R. *J. Chem. Phys.* **1954**, *22*, 1420.
(25) Kirkwood, J. *J. Chem. Phys.* **1935**, *3*, 300.
(26) Valleau, J. P.; Torrie, G. M. In *Modern Theoretical Chemistry*; Berne, B., Ed.; Plenum Press: New York, 1977; Vol. 5, pp 169–194.
(27) Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. *Faraday Symp. Chem. Soc.* **1982**, *17*, 55.
(28) Warshel, A. *J. Phys. Chem.* **1982**, *86*, 2218.
(29) Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281.
(30) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1987**, *235*, 574.
(31) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564.
(32) Straatsma, T. P.; Berendsen, H. J. C. *J. Chem. Phys.* **1988**, *89*, 5876.
(33) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486.
(34) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100–3107.
(35) Saito, M. *Mol. Sim.* **1992**, *8*, 32.
(36) Darden, T.; York, D.; Pederson, L. *J. Chem. Phys.*, In press.
(37) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.
(38) Sun, Y. X.; Spellmeyer, D.; Pearlman, D. A.; Kollman, P. *J. Am. Chem. Soc.* **1992**, *114*, 6798–6801.
(39) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1991**, *94*, 4532–4545.
(40) Pearlman, D.; Case, D.; Caldwell, J.; Singh, U. C.; Weiner, P. K.; Kollman, P. *Amber 4.0*; University of California: San Francisco, 1992.
(41) Hermans, J. *J. Phys. Chem.* **1991**, *95*, 9029–9032.
(42) Hermans, J.; Yun, R. H.; Anderson, A. G. *J. Comput. Chem.* **1992**, *13*, 429–442.
(43) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1989**, *91*, 7831–7839.
(44) Wood, R. H. *J. Phys. Chem.* **1991**, *95*, 4838.
(45) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(46) Singh, U. C.; Brown, F. K.; Bash, P. K.; Kollman, P. A. *J. Am. Chem. Soc.* **1987**, *109*, 1609.
(47) Ferguson, D. M.; Pearlman, D. A.; Swope, W. C.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 362–370.
(48) Merz, K. M.; Murcko, M. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 4484–4490.
(49) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1989**, *90*, 2460.
(50) Daggett, V.; Brown, F.; Kollman, P. *J. Am. Chem. Soc.* **1989**, *111*, 8247–8256.
(51) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175–1188.
(52) Van Helden, S. Ph.D. Thesis, Leiden University, 1991.
(53) Mizushima, N.; Spellmeyer, D.; Hirono, S.; Pearlman, D. *J. Biol. Chem.* **1991**, *266*, 11801–11809.
(54) Kuczera, K.; Gao, J.; Tidor, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8481–8485.
(55) Prevost, M.; Wodak, S. J.; Tidor, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10880–10884.
(56) Northrup, S. H.; Pear, M. R.; Lee, C.-Y.; McCammon, J. A.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4035–4039.
(57) Tobias, D. J.; Brooks, C. L. *J. Chem. Phys.* **1990**, *92*, 2582–92.
(58) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742.
(59) Dang, L. X.; Pettitt, B. M. *J. Phys. Chem.* **1990**, *94*, 4303–4308.
(60) Dang, L. X.; Rice, J. E.; Kollman, P. A. *J. Chem. Phys.* **1991**, *93*, 7528–9.
(61) Tobias, D. J.; Brooks, C. L. *J. Chem. Phys.* **1988**, *89*, 5115.
(62) Lazaridis, T.; Tobias, D. J.; Brooks, C. L.; Paulaitis, M. E. *J. Chem. Phys.* **1991**, *95*, 7612–7625.
(63) Tobias, D. J.; Brooks, C. L. *Biochemistry* **1991**, *30*, 6059–6070.

(64) Tobias, D. J.; Sneddon, S. F.; Brooks, C. L. *J. Mol. Biol.* **1990**, *216*, 783–796.
(65) Sneddon, S. F.; Tobias, D. J.; Brooks, C. L. *J. Mol. Biol.* **1989**, *209*, 817–820.
(66) Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 7167–7177.
(67) Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 503–507.
(68) Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 5716–5720.
(69) Boczko, E.; Brooks, C. L. *J. Phys. Chem.* **1993**, *97*, 4509–4513.
(70) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
(71) Pearlman, D. A. *J. Chem. Phys.* **1993**, *98*, 8946–8957.
(72) Yun, R. H.; Anderson, A.; Hermans, J. *Protein Struct. Funct. Genet.* **1991**, *10*, 219–228.
(73) Hermans, J.; Anderson, A. G.; Yun, R. H. *Biochemistry* **1992**, *31*, 5646–5643.
(74) Yun, R. H.; Hermans, J. *Protein Eng.* **1991**, *4*, 761–766.
(75) Kuyper, L. F.; Hunter, N.; Ashton, D.; Merz, K. M.; Kollman, P. A. *J. Phys. Chem.* **1991**, *95*, 6661–6666.
(76) Miyamoto, S.; Kollman, P. A. *Proteins* **1993**, *16*, 2626–245.
(77) Gao, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1989**, *110*, 4212.
(78) Cieplak, P.; Kollman, P. *J. Comput. Chem.* **1991**, *12*, 1232–1236.
(79) Reynolds, C. A.; Essex, J. W.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 9075–9079.
(80) Reynolds, C. A.; Essex, J. W.; Richards, W. G. *J. Chem. Phys. Lett.* **1992**, *199*, 257–260.
(81) Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* In press.
(82) Cornell, W.; Cieplak, P.; Bayly, C.; Kollman, P. A. *J. Am. Chem. Soc.* In press.
(83) Cieplak, P.; Cornell, W.; Bayly, C.; Kollman, P. A. Manuscript in preparation.
(84) Urban, J. J.; Damewood, J. R. *J. Chem. Soc., Chem. Commun.* **1990**, 1636.
(85) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021–8024.
(86) Marrone, J. J.; Merz, K. M. *J, Phys. Chem.* **1993**, *97*, 6524–6529.
(87) Cieplak, P.; Kollman, P. *J. Chem. Phys.* **1987**, *86*, 6393–6403.
(88) Caldwell, J.; Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 9144–9147.
(89) Dang, L. X.; Rice, J. E.; Caldwell, J.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 2481–2486.
(90) Warshel, A.; Sussman, F.; King, G. *Biochemistry* **1986**, *25*, 8368–8372.
(91) Straatsma, T. P.; McCammon, J. A. *Chem. Phys. Lett.* **1990**, *167*, 252–254.
(92) Straatsma, T. P.; McCammon, J. A. *Chem. Phys. Lett.* **1991**, *177*, 433–440.
(93) Sun, Y.; Caldwell, J.; Kollman, P. A. Manuscript in preparation.
(94) Still, W. C.; Tempcyzk, A.; Hawley, R. C.; Henderson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
(95) Honig, B.; Sharp, K.; Yang, A. S. *J. Phys. Chem.* **1993**, *97*, 1101.
(96) Cramer, C. J.; Truhlar, D. G. *J. Comput. Aided Mol. Design* **1992**, *6*, 629–666.
(97) Verkhivker, G.; Elber, R.; Nowak, W. *J. Chem. Phys.* **1992**, *97*, 7838–7841.
(98) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1989**, *91*, 3631–3637.
(99) Heinz, D. W.; Baase, W. A.; Dahlquist, F. V.; Matthews, B. W. *Nature* **1993**, *361*, 561–564 and the crystal structure of T4 lyzozyme in the Brookhaven Protein Data Bank.
(100) Ding, J. P.; Koellner, G.; Grunert, H. P.; Saenger, W. *J. Biol. Chem.* **1991**, *266*, 15128–15134.
(101) Hirono, S.; Kollman, P. A. *J. Mol. Biol.* **1990**, *212*, 197–209.
(102) Jorgensen, W. L. *BOSS*, version 3.2; Yale University: New Haven, 1992.
(103) Jorgensen, W. L.; Nguyen, T. B. *J. Comput. Chem.* **1993**, *14*, 195–205.
(104) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768–4774.
(105) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161–185.
(106) Mezei, M.; Ben Naim, A. *J. Chem. Phys.* **1990**, *92*, 1359–1361.
(107) Leach, A. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1992**, *114*, 3675–3683.
(108) Rao, B. G.; Singh, U. C. *J. Am. Chem. Soc.* **1991**, *113*, 4381–4389.
(109) Rao, B. G.; Singh, U. C. *J. Am. Chem. Soc.* **1990**, *112*, 3803–3811.
(110) Essex, J. W.; Reynolds, C. A.; Richards, W. G. *J. Chem. Soc., Chem. Commun.* **1989**, *16*, 1152–1154.
(111) Essex, J. W.; Reynolds, C. A.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 3634–3639.
(112) Jorgensen, W. L.; Briggs, J. M.; Contreras, M. L. *J. Phys. Chem.* **1990**, *94*, 1683–1686.
(113) Tobias, D. J.; Brooks, C. L. *J. Chem. Phys.* **1990**, *92*, 2582–2592.
(114) Jorgensen, W.; Buckner, K. *J. Phys. Chem.* **1987**, *91*, 6083.
(115) Yu, Ha.; Pettitt, B. M.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 2425–2434.
(116) Rame, G. L.; Lau, W. F.; Pettitt, B. M. *Int. Peptide Protein Res.* **1990**, *35*, 315–327.
(117) Ha, S. H.; Gao, J. L.; Tidor, B.; Brady, J. W.; Karplus, J. M. *J. Am. Chem. Soc.* **1991**, *113*, 1553–1557.
(118) Sun, Y. X.; Kollman, P. A. *J. Chem. Phys.* **1992**, *97*, 5108–5112.
(119) Cieplak, P.; Bash, P.; Singh, U. C.; Kollman, P. A. *J. Am. Chem. Soc.* **1987**, *109*, 6283.
(120) Worth, G. A.; King, P. M.; Richards, W. G. *Biochim. Biophys. Acta* **1989**, *993*, 134–136.
(121) Reddy, M. R.; Bacquet, R. J.; Varney, M. *J. Chim. Phys. Phys.-Chim. Biol.* **1991**, *88*, 2605–2615.
(122) Jorgensen, W. L.; Briggs, J. M.; Gao, J. *J. Am. Chem. Soc.* **1987**, *109*, 6857.
(123) Jorgensen, W. L.; Briggs, J. M. *J. Am. Chem. Soc.* **1989**, *111*, 4190.
(124) King, P. M.; Reynolds, C. A.; Richards, W. G. *J. Mol. Struct. THEOCHEM.* **1990**, *67*, 205–221.
(125) Grootenhuis, P. D. J.; Kollman, P. A.; Groenen, L. C.; Reinhoudt, D. N.; et al. *J. Am. Chem. Soc.* **1990**, *112*, 4165–4176.
(126) Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 3572–3575.
(127) Aqvist, J. *J. Phys. Chem.* **1991**, *95*, 4587–4590.
(128) Lister, S. G.; Reynolds, C. A.; Richards, W. G. *Int. J. Quantum. Chem.* **1992**, *41*, 293–310.
(129) Reynolds, C. A. *J. Am. Chem. Soc.* **1990**, *112*, 7545–7551.
(130) Duffy, E. M.; Severance, D. L.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1992**, *114*, 7535–7542.
(131) Debolt, S. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 7515–7524.
(132) Luzhkov, V.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4491–4499.
(133) Vaidehi, N.; Wesolowski, T. A.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 4264–4271.
(134) Luzhkov, V.; Warshel, A. *J. Comput. Chem.* **1992**, *13*, 199–213.
(135) Aqvist, J.; Luecke, H.; Quiocho, F. A.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, 88, 2026–2030.
(136) Daggett, V. D.; Kollman, P. A.; Kuntz, I. D. *Chem. Scr.* **1989**, *29A*, 205.
(137) Wade, R. C.; Mazor, M. H.; McCammon, J. A.; Quiocho, F. A. *Biopolymers* **1991**, *31*, 919–931.
(138) Wade, R. C.; Mazor, M. H.; McCammon, J. A.; Quiocho, F. A. *J. Am. Chem. Soc.* **1990**, *112*, 7057–7059.
(139) Langen, R.; Jensen, G. M.; Jacob, U.; Stephens, P. J.; et al. *J. Biol. Chem.* **1992**, *267*, 25625–25627.
(140) Lybrand, T. P.; McCammon, J. A.; Wipff, G. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 833–5.
(141) Grootenhis, P. D. J.; Kollman, P. A. *J. Am. Chem. Soc.* **1989**, *111*, 2152–8.
(142) Van Eerden, J.; Harkema, S.; Feil, D. *J. Phys. Chem.* **1988**, *92*, 5076–5079.
(143) Mazor, M. H.; McCammon, J. A.; Lybrand, T. P. *J. Am. Chem. Soc.* **1990**, *112*, 4411–4419.
(144) Grootenhuis, P. D. J.; Kollman, P. A. *J. Am. Chem. Soc.* **1989**, *111*, 4046–51.
(145) Miyamoto, S.; Kollman, P. A. *J. Am. Chem. Soc.* **1992**, *114*, 3668–3694.
(146) Jorgensen, W. L.; Boudon, S.; Nguyen, T. B. *J. Am. Chem. Soc.* **1989**, *111*, 755–757.
(147) Jorgensen, W. L.; Nguyen, T. B.; Sanford, E. M.; Chao, I.; Houk, K. N.; Diederich, F. *J. Am. Chem. Soc.* **1992**, *114*, 4003–4004.
(148) Jorgensen, W.; Nguyen, T. B. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1194–1200.
(149) Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008–2010.
(150) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.
(151) Pranata, J.; Jorgensen, W. L. *Tetrahedron* **1991**, *47*, 2491–2501.
(152) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 209–216.
(153) Blake, J. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1990**, *112*, 7269–7278.
(154) Lopez, M. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1989**, *111*, 6212–6222.
(155) Cannon, W. R.; Madura, J. D.; Thummel, R. P.; McCammon, J. A. *J. Am. Chem. Soc.* **1993**, *115*, 879–884.
(156) Matsui, T.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1992**, *114*, 3220–3226.
(157) Marrone, T. J.; Merz, K. M. *J. Am. Chem. Soc.* **1992**, *114*, 7542–7549.
(158) Eisenman, G.; Alvarez, O.; Aqvist, J. *J. Inclusion Phenom. Mol. Recognit. Chem.* **1992**, *12*, 23–53.
(159) Eisenman, G.; Aqvist, J.; Alvarez, O. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 2099–2109.
(160) Aqvist, J.; Alvarez, O.; Eisenman, G. *J. Phys. Chem.* **1992**, *96*, 10019–10025.
(161) Aqvist, J.; Warshel, A. *Biophys. J.* **1989**, *56*, 171–182.
(162) Roux, B.; Karplus, M. *Biophys. J.* **1991**, *59*, 961–981.
(163) Roux, B.; Karplus, J. M. *J. Am. Chem. Soc.* **1993**, *115*, 3250–3262.
(164) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 3734–9.
(165) Hermans, J.; Shankar, S. *Isr. J. Chem.* **1986**, *27*, 225–227.
(166) Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 406–411.
(167) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Protein Eng.* **1992**, *5*, 215–228.
(168) Van Eerden, J.; Briels, W. J.; Harkema, S.; Feil, D. *Chem. Phys. Lett.* **1989**, *164*, 370–376.

(169) Dang, L. X.; Pettitt, B. M.; Rossky, P. J. *J. Chem. Phys.* **1992**, *96*, 4046–4047.
(170) Boudon, B.; Wipff, G.; Maigret, B. *J. Phys. Chem.* **1990**, *94*, 6056.
(171) Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 3830–3032.
(172) Merz, K. M.; Kollman, P. A. *J. Am. Chem. Soc.* **1989**, *111*, 5649–5658.
(173) Morgan, B. P.; Scholtz, J. M.; Ballinger, M. D.; Zipkin, I. D.; Bartlett, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 297–307.
(174) Menziani, M. C.; Reynolds, C. A.; Richards, W. G. *J. Chem. Soc., Chem. Commun.* **1989**, 853–855.
(175) Hoops, S. C.; Anderson, K. W.; Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 8262–8270.
(176) Rao, B. G.; Singh, U. C. *J. Am. Chem. Soc.* **1991**, *113*, 6735–6750.
(177) Reddy, M. R.; Viswanadhan, V. N.; Weinstein, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10287–10291.
(178) Reddy, M. R.; Kalish, V.; Palmer, C.; Tatlock, J.; Varney, M.; Wu, B. W.; Appelt, K. Paper presented at Spring, 1993, ACS meeting in Denver.
(179) Rao, B. G.; Tilton, R. F.; Singh, U. C. *J. Am. Chem. Soc.* **1992**, *114*, 4447–4452.
(180) Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. *J. Med. Chem.* **1991**, *34*, 2654–2659.
(181) Tropsha, A.; Hermans, J. *Protein Eng.* **1992**, *5*, 29–33.
(182) Cieplak, P.; Kollman, P. *J. Comput. Assisted Mol. Design* **1993**, *7*, 291–304.
(183) Reddy, M. R.; Bacquet, R. J.; Zichi, D.; Matthews, D. A.; et al. *J. Am. Chem. Soc.* **1992**, *114*, 10117–10122.
(184) Cummins, P. L.; Ramnarayan, K.; Singh, U. C.; Gready, J. E. *J. Am. Chem. Soc.* **1991**, *113*, 8247–8256.
(185) Fleischman, S. H.; Brooks, C. L. *Proteins* **1990**, *7*, 52–61.
(186) Brooks, C. L.; Fleischman, S. H. *J. Am. Chem. Soc.* **1990**, *112*, 3307–3312.
(187) McDonald, J. J.; Brooks, C. L. *J. Am. Chem. Soc.* **1991**, *113*, 2295–2301.
(188) McDonald, J. J.; Brooks, C. L. *J. Am. Chem. Soc.* **1992**, *114*, 2062–2072.
(189) Reddy, M. R. Unpublished calculations at Gensia Pharmaceuticals, San Diego, CA.
(190) Caldwell, J. W.; Agard, D. A.; Kollman, P. A. *Proteins* **1991**, *10*, 140–148.
(191) Hirono, S.; Kollman, P. A. *Protein Eng.* **1991**, *4*, 233–243.
(192) Granzin, J.; Puraslutzke, R.; Landt, O.; Grunert, H. P.; et al. *J. Mol. Biol.* **1992**, *225*, 533–542.
(193) Weber, P. C.; Ohlendorf, D. H.; Wendeloski, J. J.; Salemne, F. R. *Science* **1989**, *243*, 85–88.
(194) Miyamoto, S.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* In press.
(195) Hansen, L. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 994–1002.
(196) Wilson, D. K.; Quiocho, F. *Biochemistry* **1993**, *32*, 1689–1694.
(197) Lau, W. F.; Pettitt, B. M. *J. Med. Chem.* **1989**, *32*, 2542–2547.
(198) Pearlman, D. A. *J. Comput. Chem.* Submitted for publication.
(199) Cieplak, P.; Pearlman, D. A.; Kollman, P. A. Unpublished results.
(200) Singh, S.; Ajay; Kollman, P.; Wemmer, D. *Proc. Natl. Acad. Sci. U.S.A.* Manuscript in preparation.
(201) Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A. *Nature* **1987**, *328*, 551.
(202) Hwang, J. K.; Warshel, A. *Biochemistry* **1987**, *26*, 2669–2673.
(203) Gago, F.; Richards, W. G. *Mol. Pharmacol.* **1990**, *37*, 341–346.
(204) Cieplak, P.; Rao, S. N.; Grootenhuis, P. D. J.; Kollman, P. A. *Biopolymers* **1990**, *29*, 717–727.
(205) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631–639.
(206) Wade, R. C.; McCammon, J. A. *J. Mol. Biol.* **1992**, *225*, 679–696.
(207) Wade, R. C.; McCammon, J. A. *J. Mol. Biol.* **1992**, *225*, 697–712.
(208) Komeiji, Y.; Uebayasi, M.; Someya, J.-I.; Yamato, I. *Protein Eng.* **1993**, *5*, 759–767.
(209) Dang, L. X.; Merz, K. M.; Kollman, P. A. *J. Am. Chem. Soc.* **1989**, *111*, 8505–8508.
(210) Dang, L. X.; Kollman, P. A. *Expanding Frontiers in Polypeptide and Protein Structural Research*; Renogopalakrishnan, Ed.; ESCOM Press: Leiden, 1991.
(211) Pearlman, D. A.; Kollman, P. A. *Biopolymers* **1990**, *29*, 1193–1209.
(212) Dang, L. X.; Pearlman, D. A.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4630–4634.
(213) Singh, S.; Pearlman, D. A.; Kollman, P. A. *J. Biomol. Struct. Dyn.* In press.
(214) Ross, W. S.; Hardin, C. C.; Tinoco, I.; Rao, S. N.; et al. *Biopolymers* **1989**, *28*, 1939–1957.
(215) Hausheer, F. H.; Singh, U. C.; Saxe, J. D.; Flory, J. P.; et al. *J. Am. Chem. Soc.* **1992**, *114*, 5356–5362.
(216) Ferguson, D.; Kollman, P. *Antisense Res. Dev.* **1991**, *1*, 243–254.
(217) Hausheer, F. H.; Rao, B. G.; Saxe, J. D.; Singh, U. C. *J. Am. Chem. Soc.* **1992**, *114*, 3201–3206.
(218) Yamaotsu, N.; Moriguchi, I.; Kollman, P. A.; Hirono, S. *Biochem. Biophys. Acta* **1993**, *1163*, 81–88.
(219) Zhang, L.; Hermans, J. *Proteins* **1993**, *16*, 384–392.
(220) Tidor, B.; Karplus, M. *Biochemistry* **1991**, *30*, 3217–3228.
(221) Yung-yu, S.; Mark, A.; Cun-xian, W.; Fukua, H.; Berendsen, H. J. C.; van Gunsteren, W. *Protein Eng.* **1993**, *6*, 289–295.
(222) Tobias, D. J.; Brooks, C. L. *J. Phys. Chem.* **1992**, *96*, 3864–3870.
(223) Tobias, D. J.; Sneddon, S. F.; Brooks, C. L. *J. Mol. Biol.* **1992**, *227*, 1244–1252.
(224) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
(225) Chandresekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154–163.
(226) Blake, J. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 7430–7432.
(227) Chandesekhar, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 2974–5.
(228) Severance, D. L.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1992**, *114*, 10966–10968.
(229) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; et al. *Biochemistry* **1991**, *30*, 5826–5832.
(230) Karplus, M.; Evanseck, J. D.; Joseph, D.; Bash, P. A.; et al. *Faraday Discuss. Chem. Soc.* **1992**, *93*, 239–248.
(231) Peng, Z.; Merz, K. M. *J. Am. Chem. Soc.* **1992**, *114*, 2733–2734.
(232) Zheng, Y. J.; Merz, K. M. *J. Am. Chem. Soc.* **1992**, *114*, 10498–10507.
(233) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213–217.
(234) Gao, J. L.; Xia, X. F. *Science* **1992**, *258*, 631–635.
(235) Gao, J. *J. Am. Chem. Soc.* **1993**, *115*, 2930–2935.
(236) Gao, J.; Pavelites, J. J. *J. Am. Chem. Soc.* **1993**, *114*, 1912–1914.
(237) Gao, J.; Chou, L.; Auerbach, A. *Biophys. J.* **1993**, *65*, 43–47.
(238) Warshel, A.; Aqvist, J.; Creighton, S. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5820–5824.
(239) Aqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1990**, *112*, 2860–2868.
(240) Aqvist, J.; Warshel, A. *Biochemistry* **1989**, *28*, 4680–4689.
(241) Hwang, J. K.; Chu, Z. T.; Yadav, A.; Warshel, A. *J. Phys. Chem.* **1991**, *95*, 8445–8448.
(242) Yadav, A.; Jackson, R. M.; Holbrook, J. J.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4800–4805.
(243) Warshel, A.; Parson, W. W. *Annu. Rev. Phys. Chem.* **1991**, *42*, 279–309.
(244) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682–8692.
(245) Langen, R.; Schweins, T.; Warshel, A. *Biochemistry* **1992**, *31*, 8691–8696.
(246) Aqvist, J.; Forgill, M.; Warshel, A. *J. Am. Chem. Soc.* **1993**, *115*, 631–635.
(247) Warshel, A.; Hwang, J. K.; Aqvist, J. *Faraday Discuss. Chem. Soc.* **1992**, *93*, 225–238.
(248) Aqvist, J.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 7–14.
(249) Cram, D. J.; Carmack, R. A.; deGrandpre, M. P.; Lein, G. M.; Goldberg, I.; Knobler, C. B.; Maverick, E. F.; Trueblood, K. N. *J. Am. Chem. Soc.* **1987**, *109*, 7068–7073.
(250) Bayly, C.; Kollman, P. *J. Am. Chem. Soc.* Submitted for publication.
(251) Thomas, B. Unpublished results.
(252) Gerber, P.; Mark, A.; van Gunsteren, W. *J. Comput. Assisted Mol. Design* **1993**, *7*, 305–323.
(253) DeBolt, S.; Pearlman, D. A.; Kollman, P. A. *J. Comput. Chem.*, submitted for publication.
(254) Tidor, B. *J. Phys. Chem.* **1993**, *97*, 1069–1073.
(255) Ramnarayan, K.; Rao, B. G.; Singh, U. C. *J. Chem. Phys.* **1990**, *92*, 7057–7067.
(256) Boudon, S.; Wipff, G. *J. Comput. Chem.* **1991**, *12*, 42–51.
(257) Lowis, D. R.; Essex, J. W.; Richards, W. G. *Mol. Simul.* **1993**, *9*, 369–383.
(258) Elcock, A. H.; Richards, W. G. *J. Am. Chem. Soc.* **1993**, *115*, 7930.
(259) Gao, J. L. *J. Am. Chem. Soc.* **1993**, *115*, 6893–6895.
(260) Auffinger, P.; Wipff, G. *J. Chim. Phys.* **1991**, *88*, 2525–2534.
(261) Zacharias, M.; Straatsma, T. P.; McCammon, J. A.; Quiocho, F. A. *Biochemisrty* **1993**, *32*, 7428–7434.
(262) Prodhom, B.; Karplus, M. *Protein Eng.* **1993**, *6*, 585–592.
(263) Li, Z.; Hermans, J. *Proteins* **1993**, *16*, 384–392.
(264) Yan, Y. B.; Tropsha, A.; Hermans, J.; Erikson, B. W. *Proc. Nat. Acad. Sci. U.S.A.* **1993**, *90*, 7898–7902.
(265) van Gunsteren, W. F. *Curr. Opin. Struct. Biol.* **1993**, *3*, 277–281.