Annu. Rev. Biophys. Biophys. Chem. 1989. 18:431–92

FREE ENERGY VIA MOLECULAR SIMULATION: Applications to Chemical and Biomolecular Systems

D. L. Beveridge and F. M. DiCapua

Department of Chemistry and Department of Molecular Biology and Biochemistry, Hall-Atwater Laboratories, Wesleyan University, Middletown, Connecticut 06457

CONTENTS

0883–9182/89/0610–0431$02.00

PERSPECTIVES AND OVERVIEW

The complement of molecular theory and experiment in gaining new knowledge and solving research problems in the biological sciences has been considerably enhanced by the application of molecular simulation to problems in biophysics and biophysical chemistry (142). Molecular simulation can be defined as the numerical determination of the thermodynamic, energetic, structural, and dynamical properties of a mathematical model of a molecular assembly on a digital computer. Computer simulations on molecular systems can be carried out in a probabilistic mode using Monte Carlo (MC) calculations, or they can be performed deterministically, integrating Newton's Laws of Motion, by means of molecular dynamics (MD) (2). Applications of molecular simulation to diverse chemical and biomolecular systems have included studies of liquid water (29), aqueous solution structure (130), proteins (113, 142), nucleic acids (142, 230), and at the outer limit of system size, membrane dynamics (19). In diverse fields, simulations have been helpful in the analysis and interpretation of experimental results in terms of a theoretical model (122, 159).

The Second Law of Thermodynamics establishes that natural systems achieve at equilibrium or else seek out from a nonequilibrium position a state of minimum free energy. All thermodynamic properties can be obtained from a knowledge of free energy and its derivatives. Thus a particular focus (explicit or implicit) on free energy is necessary when comparing theory and experiment. However, free energy is not readily obtained in a conventional molecular simulation. The problem originates in the intrinsic difficulty of calculating the entropy of the system. Conventional MC or MD simulations sample only a restricted part of configuration space, the low-energy region. Relatively good estimates of internal energy and enthalpy are obtained, since the low-energy regions make the dominant contribution to these quantities. Estimates of entropy, a measure of the relative disorder, require more extensive knowledge about the system and thus a broader if not global sampling of configuration space. For systems of $\mathcal{O}(10^2$–$10^3)$ particles at liquid-state densities, this sampling is essentially intractable. However, special methodologies have been introduced to provide a means for the determination of at least relative free energies for molecular systems via computer simulations. Even so, quite

intensive computational procedures are required in all but the simplest of cases. With the advent of supercomputers, dedicated-purpose mini-supercomputers, and multiprocessors, these computational barriers are being overcome, and free energy simulations for important chemical and biomolecular systems have become feasible. Diverse novel and interesting results are currently being obtained.

In this article we review the literature to date and provide additional perspectives on the application of free energy determination via molecular simulation to systems of interest in biophysics and biophysical chemistry. We include applications to selected chemical systems as well, since the methodology for biomolecular systems is developed and validated on relevant chemical prototypes. There are a number of diverse methods for free energy calculations, but currently thermodynamic integration, the perturbation method, and potential-of-mean-force calculations are of the most general interest in molecular biophysics and are emphasized herein. We also focus mainly on techniques generally applicable to systems with solvent included, and in which diffusive modes of motion are prevalent.

A basic familiarity with statistical thermodynamics and molecular simulation theory is presumed. For further theoretical background, the interested reader is referred to a new and comprehensive text on molecular simulation (2), a recent monograph on molecular dynamics of proteins and nucleic acids (142), and other recent reviews on free energy and computer simulation topics (21, 28, 152, 176, 181, 202, 225, 230a). Barker & Henderson (6) have provided extensive background on free energy methods as well as general aspects of liquid state theory in a comprehensive review.

A number of the initial (79) and more recent (230a) studies involving free energy simulations have been collected in symposium proceedings. Our article in Reference 230a (27) describes in considerable elementary detail this subject area and includes complete derivations of all the free energy equations used herein. That article may serve as an introduction and companion piece to the present review. Also, van Gunsteren (227) and Straatsma (210) have dealt with diverse aspects of the theory at a considerably more advanced level.

BACKGROUND

Statistical Thermodynamic Quantities

Important statistical thermodynamic quantities for a discussion of computer simulations and the free energy problem, referred to as the canonical (T,V,N) ensemble, are given in this section. Here T is temperature, V is volume, and N is the number of particles in the system. The operational

elements of statistical mechanics in the determination of (excess) free energy are configurational integrals; we frame our discussion here completely in these terms. The formalism generalized to phase space is described in Reference 227. The quantities are (*a*) the configurational partition function,

$$Z_N = \int \cdots \int \exp[-\beta E(X^N)] dX^N, \qquad 1.$$

where $E(X^N)$ is the configurational energy, $\beta = (kT)^{-1}$, k is Boltzmann's constant, T is the absolute temperature, and the configurational integration extends over all space dX^N of the N-particle system; (*b*) the Boltzmann probability function,

$$P(X^N) = \exp[-\beta E(X^N)]/Z_N, \qquad 2.$$

for a configuration corresponding to a particular energy, $E(X^N)$; (*c*) the average or mean energy expression,

$$U = \int \cdots \int E(X^N) P(X^N) dX^N = \langle E(X^N) \rangle, \qquad 3.$$

equivalent to the excess internal energy of the system (note the bracket notation for an ensemble average); (*d*) the heat capacity,

$$C_V = \frac{\langle E(X^N)^2 \rangle - \langle E(X^N) \rangle^2}{kT^2}, \qquad 4.$$

and (*e*) the excess or configurational free energy,

$$A = -kT \ln[Z/(8\pi^2 V)^N], \qquad 5.$$

which is the essential focus of interest here. The appearance of the normalization factor $(8\pi^2 V)^N$ in the denominator of Equation 5 establishes that A, a Helmholtz free energy, is defined relative to an ideal gas reference state, $E(X^N) = 0$. For an atomic fluid, or for a macromolecule treated as interacting atoms, the orientational factor $8\pi^2$ is replaced by unity. The corresponding Gibbs free energy is

$$G = A + PV = U + PV - TS, \qquad 6.$$

where P is the pressure of the system and $U + PV$ is the enthalpy, H. All quantities and methodology described here can be readily formulated in the constant pressure (T,P,N) ensemble, in which G is the thermodynamic potential, as well as in the canonical (T,V,N) ensemble. The grand canonical ensemble invites interesting alternative approaches to the problem (see

6, 152, 210), which have not yet proven to be of practical utility at densities typical of chemical and biomolecular systems.

For complete background on statistical thermodynamics, see Reference 147.

Protocols of Molecular Simulation

An MC or MD computer simulation on a molecular system begins with an arbitrarily chosen initial configuration and involves generating an ensemble of structures by well-defined prescriptions. In the initial segment of the simulation the energy of the system changes rapidly, but it eventually settles into a reasonable approximation of an oscillation about a mean. At this point the simulation has achieved "equilibration." For a fully ergodic system, in which all regions of phase space are accessible, the region of equilibration will be the same regardless of the choice of initial configuration. When equilibration has been numerically achieved, the previous history of the calculation is discarded, and the "production" phase commences. Properties are calculated from the averages over the ensemble of structures generated during the production phase of a simulation. This phase usually involves millions of configurations in an MC calculation and tens to hundreds of picoseconds in an MD calculation on present-generation machines. The simulation is allowed to proceed until the calculated properties are no longer significantly changing. At this point the simulation is said to be "converged" (but it is actually only stabilized).

Configuration space is by definition of order infinity. Since a given computer simulation is by definition finite, one obtains from it only estimates of the true, infinite-order ensemble averages of the energy and other quantities being determined. Another finite realization, commencing with a different but perfectly acceptable equilibrated structure, would be expected to lead to a slightly different estimate. The set of possible estimates from various realizations in a proper simulation comprises independent random estimates of the true mean, and these should be normally distributed. In an important early article on MC studies of simple liquids, Wood (251) applied standard small-sample statistical theory to the problem. A realization consisting of n total steps is broken up into some small number m of successive sequences (blocks, batches) from which statistical uncertainties, i.e. standard deviations, can be estimated. The size of the blocks is determined to assure they are asymptotically independent of one another. A display of both cumulative mean energy and the batch means for segments along the realization as a function of configurations in the sampling gives a simple statistical tool for assessing the stability of the calculation. Consideration of the statistical uncertainties is vital in the analysis and interpretation of the results of a simulation and in the critical

comparison of calculated and experimentally observed values. Issues of convergence in various other ensemble averages also figure significantly in the free energy problem, and they crop up frequently in the following discussion.

Each property of the system has a characteristic convergence profile. Simple average quantities, such as mean energy, are the simplest to compute and the fastest to converge. Fluctuation properties, such as heat capacity, achieve convergence slower than mean energy. The convergence of each property can be monitored in the course of the simulation, and the method of batch means described above can be used to assign appropriate statistical uncertainties to each calculated quantity. See Reference 213 for further discussion of the statistical uncertainty problem in molecular simulations.

The Free Energy Problem

Considerations on free energy in molecular simulations take a distinctly different form for intramolecular and intermolecular degrees of freedom. For the intramolecular case, the problem involves vibrational and librational modes of motion on the intramolecular energy surface. In the vicinity of well-defined local energy minima, the free energy is readily accessible. It can be computed by constructing a partition function from the vibrational frequencies obtained from a normal mode calculation and treating the problem in the harmonic approximation, essentially an extension of the Einstein oscillator formalism (76, 147, 197). Anharmonic effects can be introduced by performing an MC or MD molecular simulation on the intramolecular energy surface and calculating the entropy in the quasiharmonic approximation (61, 111).

In fluids, including molecular liquids, solutions, and vapors, the particles of the system undergo diffusional motion, and a harmonic or quasiharmonic approximation breaks down. These considerations apply also for flexible molecules, in which conformational transitions are effectively an intramolecular diffusional mode. Conventional MC and MD procedures for diffusional modes, although firmly grounded in Boltzmann statistics, do not ordinarily involve the direct determination of a partition function or obtain a partition function in the calculation. The exponential dependence of the Boltzmann factor on the energy makes the partition function notoriously slow to converge for densities typical of chemical and biomolecular systems (another manifestation of the entropy sampling problem). The Metropolis method in MC calculations is a Markov process designed to avoid the necessity of determining a partition function in the calculation of internal energy and related thermodynamic properties in a simulation. In MD procedures, the physical nature of the calculated

trajectories of the particles serves this purpose equally well. In any case, in the absence of a partition function one is unable to compute the free energy of the system from Equation 5.

To present the problem at a more mathematical level, we cast the free energy into the form of an ensemble average expression that can be evaluated via simulation in a manner analogous to that used for the internal energy. Inserting unity in the form of

$$(8\pi^2 V)^{-N} \int \cdots \int \exp[\beta E(X^N)] \exp[-\beta E(X^N)] dX^N \qquad 7.$$

into the argument of the logarithm in Equation 5 and inverting results in

$$A = kT \ln \int \cdots \int \exp[\beta E(X^N)] P(X^N) dX^N \qquad 8.$$

or, defining a bracket notation for configurational integrals,

$$A = kT \ln \langle \exp[\beta E(X^N)] \rangle. \qquad 9.$$

In principle, Equation 9 provides a means for calculating excess free energy in a single conventional simulation by determining an ensemble of configurations consistent with the probability function $P(X^N)$ and integrating over all space as in the determination of average energy via Equation 3. However, expanding the exponential, one obtains

$$A = \langle E(X^N) \rangle - \frac{\beta}{2!} \langle E(X^N) \rangle^2 + \ldots, \qquad 10.$$

where the first term is the internal energy and the remainder of the expansion is equivalent to $-TS$, where S is the entropy of the system. Thus evaluating the ensemble averages of the exponential is equivalent to determining the ensemble averages of not only the configurational energy, but also the energy squared (as encountered in heat capacity determinations) and other high powers of the energy (fluctuation terms) to an extent depending upon the radius of convergence of the expansion. As a consequence, the free energy computed in a Metropolis MC or MD simulation by Equation 9 is bound to converge even more slowly than the average energy, which is already time consuming to achieve. Therein lies the difficulty.

The free energy problem can be appreciated from another perspective by examining explicitly the configurational integral of Equation 8 (164), where $P(X^N)$ is proportional to $\exp[-\beta E(X^N)]$ as given in Equation 2. Because of the rapid increase of $\exp[\beta E(X^N)]$ with energy, higher-energy regions of configuration space will be important to the configurational

integral. However, MC-Metropolis or MD molecular simulation, sampling on a probability measure proportional to $\exp[-\beta E(X^N)]$, i.e. $P(X^N)$, seeks out preferentially the lower regions of configuration space. For densities typical of molecular liquids, the high- and low-energy regions of configurational space are sufficiently well separated that a conventional simulation of practical length will never adequately sample the high-energy regions that contribute significantly to the ensemble average expression for free energy.

The Coupling Parameter Approach

In molecular systems, the free energy problem can frequently be cast in terms of a free energy difference, ΔA, between two well-defined states, with the calculations focused on the determination of relative free energy. For states denoted 0 and 1, ΔA is formally given in terms of the ratio of the partition functions for the two states, Z_1 and Z_0, as

$$\Delta A = A_1 - A_0 = -kT \ln (Z_1/Z_0). \qquad 11.$$

A straightforward approach to the free energy difference ΔA would require independent determinations of Z_0 and Z_1 based on the energy functions E_0 and E_1 (the dependence on configurational coordinates is implicit), which individually are subject to all the numerical difficulties detailed in the preceding section. The theoretical and methodological approaches to the problem considered herein can all be cast in terms of a very useful construct, the coupling parameter. Let us assume that the potential, E, depends on a continuous parameter λ, such that as λ is varied from 0 to 1, $E(\lambda)$ passes smoothly from E_0 to E_1. This relation defines an analytical continuation,

$$A(\lambda) = -kT \ln Z(\lambda). \qquad 12.$$

ΔA can be calculated by integrating the derivative of $A(\lambda)$ along λ (thermodynamic integration); by designating intermediate states $A(\lambda_i)$ spaced closely enough and using a form of Equation 11 to compute ΔA in a stepwise manner (the perturbation method); or by actually developing $A(\lambda)$ on the [0, 1] interval from simulations where λ is variable and obtaining ΔA as $A(\lambda = 1) - A(\lambda = 0)$. The function $A(\lambda)$ is equivalent to the potential of mean force with respect to the coordinate λ in the statistical mechanics of fluids. Thermodynamic integration, the perturbation method, and the potential-of-mean-force calculations are the main approaches to free energy determinations via molecular simulation currently in wide use in molecular biophysics.

In a modern sense the coupling parameter approach originates in Kirkwood's derivation of an important integral equation in liquid state theory

(118), but the seeds of the idea can be traced back to work on chemical affinity and the degree-of-advancement parameter for a chemical process (56). The coupling parameter, λ, may be considered a generalized extent parameter, defining the progress of the system along a path between the initial and final states 0 and 1. On a physical path, $A(\lambda)$ can be used in certain cases to determine the free energy of activation, $\Delta A^{\ddagger}$, as well as ΔA, and thus leads to estimates of rate constants via transition state theory (73). It is important to note that nonphysical, "fictitious" paths are admissible if the quantity of interest is a state function like ΔA, where value is independent of the path. This turns out to bc a critical advantage in solving actual problems, since a nonphysical pathway may be computationally more convenient. For example, in a mutational process (108), relative free energies for a series of molecules can be computed by using a λ coordinate to change computationally one functional group, subunit, or even entire molecule into another. There is a quite timely correspondence of this idea with the newly developed area of site-specific mutagenesis in molecular biology, where some of the current interesting biophysical applications are to be found. Overall, there is considerable freedom in the choice of the λ coordinate, and decisions on the selection of λ are usually made by combining conceptual and numerical requirements with the nature of the problem at hand.

The transition from the initial to final state in a chemical system frequently involves changes in what may generally be called the molecular topography. The various types of changes on the λ coordinate have been categorized in terms of topographical transition coordinates (27, 28, 152). Mutation is sometimes referred to as "creation/annihilation" and "fade and grow."

THEORY AND METHODOLOGY

Thermodynamic Integration

Given the free energy function, $A(\lambda)$, the free energy difference between an initial state $\lambda = 0$ and a final state $\lambda = 1$ follows simply from the definition of an integral in elementary calculus as

$$\Delta A = \int_0^1 [\partial A(\lambda)/\partial\lambda]\, d\lambda. \qquad 13.$$

Here

$$A(\lambda) = -kT \ln Z(\lambda), \qquad 14.$$

where

$$Z(\lambda) = \int \cdots \int \exp\left[-\beta E(X^N, \lambda)\right] dX^N. \quad 15.$$

It follows after some rearrangement (27) that

$$\partial A(\lambda)/\partial\lambda = \langle \partial E(X^N, \lambda)/\partial\lambda \rangle_\lambda, \quad 16.$$

where the subscript λ indicates an ensemble average over the λ probability function,

$$P(X^N, \lambda) = \frac{\exp\left[-\beta E(X^N, \lambda)\right]}{\int \cdots \int \exp\left[-\beta E(X^N, \lambda)\right]}, \quad 17.$$

that is, the probability of finding the system in configuration X^N at a position λ between the initial and final state of the transformation. Finally,

$$\Delta A = \int_0^1 \langle \partial E(X^N, \lambda)/\partial\lambda \rangle_\lambda \, d\lambda. \quad 18.$$

The direct numerical evaluation of Equation 18 in molecular simulation, integrating between initial and final states using some thermodynamic relationship, is called "thermodynamic integration." Textbook cases in physical chemistry are recovered when λ is identified with a volume change or a temperature change in the system, giving rise to an integral involving the virial expression for pressure or the Gibbs-Helmholtz equation, respectively (6). A particularly important point is that the integration variable, λ, need not be restricted to thermodynamic variables like volume and temperature; it can be used to define an analytical continuation between initial and final states in many conceivable ways, particularly, as mentioned in the previous section, in terms of changes in molecular topography. This idea has broad use in chemical applications of free energy simulations.

For the special but commonly encountered case in which $E(X^N, \lambda)$ is linear in λ, e.g.

$$E(X^N, \lambda) = (1-\lambda)E_0(X^N) + \lambda E_1(X^N) \quad 19.$$

$$= E_0(X^N) + \lambda \Delta E(X^N), \quad 20.$$

where

$$\Delta E(X^N) = E_1(X^N) - E_0(X^N), \quad 21.$$

then

$$\partial E(X^N, \lambda)/\partial\lambda = \Delta E(X^N) \qquad 22.$$

and

$$\Delta A = \int_0^1 \langle \Delta E(X^N) \rangle_\lambda \, d\lambda. \qquad 23.$$

Note that although $\Delta E(X^N)$ is now independent of λ, the ensemble average must still be taken over the λ distribution, $P(X^N, \lambda)$.

The methodology for carrying out a thermodynamic integration in the context of a molecular simulation is thus straightforward. A series of simulations is set up corresponding to a succession of discrete λ values on the interval from 0 to 1, giving discrete values for the various $\langle \Delta E(X^N) \rangle_\lambda$. The auxiliary integration over λ is then carried out numerically. The λ integration and the free energy determination thus require typically 5–10 times as much computational effort as a mean energy calculation via molecular simulation, just to get one (albeit important!) number.

In favorable cases, it is possible to expedite the calculations by allowing λ to vary continuously in the simulation, a procedure called "slow growth" (212). A single extended simulation is used to develop an estimate of free energy via

$$\Delta A = \sum_{i=1}^{N} \langle \partial E(\lambda)/\partial\lambda \rangle_i \Delta\lambda_i, \qquad 24.$$

where i counts the number of steps. Encouraging results have been obtained by this procedure on a picosecond timescale for simple systems. However, strictly speaking, the simulation never properly equilibrates unless λ is infinitesimally small, and the sampling is said to lag behind the Hamiltonian. Furthermore, all structural changes that occur as the system goes from the initial to final state must be achieved on the picosecond time scale. Larger (and indeed, conformational) changes may not have time to occur. Thus in unfavorable cases, slow growth may be less suitable than a discrete method wherein a thorough equilibration may be conducted at each step. An insufficient sampling of configuration space in a simulation is referred to as a "quasi-ergodic" problem in molecular simulations.

Particular care must be taken with the thermodynamic integration calculation (and other methods as well) in the vicinity of $\lambda \to 0$ (and $\lambda \to 1$ for a mutation). Here, when the initial and final states put the atoms at different positions, the simulation tends increasingly to permit the configurations in $P(X^N, \lambda)$ for which the actual ΔE is quite high owing to clashes. The predominance of these configurations leads to a rapid increase in $\langle \Delta E(X^N) \rangle_\lambda$ (possibly a singularity) and a concomitant loss of numerical accuracy, if not a complete blowup of the problem. We refer to this

distortion as the "close encounters" problem. A partial antidote is to employ an energy expression that is nonlinear in λ, such as

$$E(X^N, \lambda^m) = (1-\lambda^m)E_0(X^N)+\lambda^m. \quad 25.$$

Here the energy approaches the asymptotes more sharply than in the linear case, so a knowledge of the coupled particle is built into the system at an earlier point in the integration. The close encounters problem requires considerable practical attention in free energy simulations. Nonlinear methods are discussed further in References 50 and 152.

The Perturbation Method

The perturbation method [based on equations given by Zwanzig (254)] is used to calculate the free energy difference ΔA (or ΔG) for two states by an alternative route to that used in thermodynamic integration. The derivation, however, parallels that used earlier for the ensemble average expression for free energy in Equation 9. We begin with the free energy difference written in the form of Equation 15. By inserting a unity in the form of

$$1 = \exp[\beta E_0(X^N)]\exp[-\beta E_0(X^N)] \quad 26.$$

into the integrand of the partition function Z_1 in the numerator, we obtain

$$\Delta A = -kT\ln\langle\exp[-\beta\Delta E(X^N)]\rangle_0, \quad 27.$$

where the subscript zero indicates configurational averaging over the ensemble of configurations representative of the initial (or reference) state of the system. By symmetry, since the initial and final states are in principle interchangeable,

$$\Delta A = -kT\ln\langle\exp[\beta\Delta E(X^N)]\rangle_1, \quad 28.$$

where the configurational averaging is performed with respect to the probability function representative of the final state of the system. Introduction of virtual particles into the coordinate list allows this treatment to be generalized to any change in the system.

This approach is expected to be numerically accurate only when the energies of the initial and final states of the system differ only slightly, so that one state may be regarded as a perturbation of the other. Thus it is common practice in the chemical literature to refer to this approach as the "perturbation method." However, Equations 27 and 28 are in principle exact, provided there are no configurations for which $E_0(X^N)$ is infinite but $E_1(X^N)$ is finite. This approach is also known as the "direct method," since in principle ΔA is obtained from a single simulation, as opposed to

the series of simulations (or else slow growth) and subsequent λ quadrature involved in thermodynamic integration.

The simplest implementation of the perturbation method involves defining $E_0(X^N)$ and $E_1(X^N)$ and running an MC or MD simulation for the initial state of the system in the usual manner, i.e. taking a sampling based on the probability function $P_0(X^N)$ and forming the ensemble average of $\exp[-\beta\Delta E(X^N)]$. Inserting this average into Equation 27 gives the corresponding free energy difference. An analogous procedure can be formulated with Equation 28. These two calculations are referred to as running forward or backward, respectively, on the λ coordinate. A better procedure, but twice as much work, is to calculate ΔA both ways and average the results. This is called "double-ended" sampling (17). The discrepancy or hysteresis in the two estimates of ΔA is an estimate of the statistical uncertainty in the result due to inaccuracies in the numerical integrations of Equations 27 and 28, and it arises from configurations sampled in $P_0(X^N)$ but not in $P_1(X^N)$ and vice versa. It is essential to note that errors due to insufficient sampling in both directions, i.e. important configurations missed in both $P_0(X^N)$ and $P_1(X^N)$ simulations, are not accounted for in this estimate of error; that is, it is an estimate of precision, but not necessarily of accuracy. Such errors are another example of a quasi-ergodic problem.

The discrepancy between results based on forward and backward integrations is minimal when the initial and final states of the system are very similar. Then configurations important in the ensemble representative of the initial state are also important in the final state and vice versa and are picked up in the sampling for both directions. Successful application of the perturbation method is thus commonly limited to free energy differences of $<2kT$, or ~ 1.5 kcal mol^{-1} (212). However, the free energy difference for many chemical and biomolecular processes is considerably larger than $2kT$. The perturbation method can be used successfully in most cases by defining intermediate states on the λ coordinate that differ successively by no more than $2kT$. The total free energy change can be obtained by summing the ($\leq 2kT$) ΔAs between the intermediate states,

$$\Delta A = \sum_{i=0}^{k-1} \Delta A(\lambda_{i \to i+1}), \qquad 29.$$

where the interval between $\lambda = 0$ and $\lambda = 1$ has been divided up into k subintervals. Obviously, the computation time increases with the number of subintervals involved, since individual simulations are necessary for each intermediate state λ_i. The subinterval perturbation method can be performed most economically using double-wide sampling (17) on each individual leg of double-ended sampling.

The defining equations of thermodynamic integration (Equation 18) and the perturbation method (Equation 27) both clearly show that free energy via molecular simulation is obtained as an ensemble average of a difference ΔE rather than as a difference in ensemble averages, as in the case of ΔU. Furthermore, in ΔE the energy contributions due to any common features between initial and final states cancel. As a consequence of both of these effects, ΔA (or ΔG) can actually be determined with less numerical uncertainty in a simulation than ΔU (or ΔH). In recent advances, Mezei (150) has proposed a hybrid form of thermodynamic integration and the perturbation method and has also been exploring procedures in adaptive importance sampling. Li & Scheraga (125) have also proposed a recursive procedure for evaluation of absolute free energy based on Equation 9.

The Potential of Mean Force

For the special case of two of the N particles of a system fixed in space at a distance R, the expression for Helmholtz free energy takes the form

$$A(R) = -kT \ln \{\exp[-E(X^N|R)kT]\}\, dX^{N-2}, \qquad 30.$$

where $(X^N|R)$ denotes a configuration conditional on a particle separation of R. The radial distribution function for the distance of any two particles has been defined elsehere (14) by the equation

$$g(R) = [N(N-1)/\rho] \exp[-E(X^N|R)/kT]\, dX^{N-2}, \qquad 31.$$

where ρ is the number density. Thus,

$$A(R) = -kT \ln[g(R)] + \text{constant}. \qquad 32.$$

The quantity $A(R)$ is the reversible work involved in the association of the two particles, from infinite separation to the distance R, in solution. The derivative of $A(R)$ with respect to R is the force acting between the two particles, including both direct and solvent-averaged contributions. In the statistical mechanics literature, $A(R)$ is frequently encountered and is known as a potential of mean force, usually denoted as $w(R)$. R can be a simple distance, an internal coordinate representing a torsional degree of freedom, or a generalized structural alteration in the system. A knowledge of $A(R)$ can be useful for the study of free energy for conformational changes, molecular associations, and chemical reactions in condensed phases.

The potential of mean force can in principle be determined from simulation in two ways. The R coordinate can be considered as an additional variable in the simulation, which thus produces a direct estimate of $g(R)$. As an alternative, it is possible to perform separate simulations at different

fixed values of R, compute the forces due to the surroundings, and then integrate the force functions numerically. The latter route is numerically less reliable, and we focus here on approaches involving, in one way or another, a variable R.

In a variable R simulation, $g(R)$ is obtained as the ratio of the probability of sampling the coordinate R and the volume element of the configuration space corresponding to the coordinate R:

$$g(R) = P(R)/V(R). \qquad 33.$$

The volume element can be intepreted as a quantity proportional to the probability of sampling the parameter R with the potential function set to zero. For example, if R is an intermolecular distance, $V(R) = 4\pi R^2$, and if R is a torsion angle, $V(R)$ is constant. Determination of the volume element becomes progressively more complex as the dimensionality of R is increased. Also, subtle consideration of the Jacobian for the transformation (63) may be necessary (see 227); van Gunsteren (227) formulated this in the context of free energy simulation.

In any case, when the R coordinate is considered as simply another degree of freedom in an otherwise conventional mean energy calculation, serious sampling problems arise. In practical applications, $w(R)$ varies by several kT in its range. This variance, however, converts into order of magnitude differences in $g(R)$. The simulation, seeking to describe the equilibrium state dictated by Boltzmann statistics, would end up sampling only the most probable region of the R space. As a result, all values of R would not be represented, and thus an accurate $w(R)$ would not be obtained. The range of sampling can be extended systematically in a special form of importance sampling by running the Metropolis procedure or molecular dynamics with a modified energy function,

$$E'(X^N) = E(X^N) + \omega(X^N). \qquad 34.$$

Here $\omega(X^N)$ is a weighting function that can take various forms depending on the particular application, but typically is of the form

$$\omega_0(X^N) = k_\omega(X^N - X_0^N)^2, \qquad 35.$$

where k_ω is an effective force constant. Values of $\omega(X^N)$ become large for points in configuration space that are far removed from X_0. Thus, a simulation based on the energy function $E'(X^N)$, referred to as a "surrogate potential," will be biased toward the region of X_0 to an extent depending on the value of the force constant k_ω. This technique is called "umbrella sampling," since the resulting distributions are broader than the Boltzmann distribution. The function $\omega_0(X^N)$ is arbitrary, and using $E'(X^N)$ rather than $E(X^N)$ in a simulation and generating umbrella distributions

carries the sampling to a non-Boltzmann regime. Thus, the ensemble averages obtained do not correspond properly to thermodynamic observables.

Valleau & Torrie (225), in research that has come to have broad, important implications in free energy simulations, invented umbrella sampling and devised a procedure for extracting proper Boltzmann-weighted ensemble averages from a non-Boltzmann sampling. For the ensemble average $\langle Q \rangle$ of any property Q,

$$\langle Q \rangle = \frac{\langle Q(X^N) \exp[\beta\omega(X^N)] \rangle_\omega}{\langle \exp[\beta\omega(X^N)] \rangle_\omega}, \qquad 36.$$

where the subscript ω indicates an ensemble average based on the probability $P_\omega(X^N)$ and the energy function $E'(X^N)$. Here

$$P_\omega(X^N) = \frac{\exp[-\beta E'(X^N)]}{\int \cdots \int \exp[-\beta E'(X^N)]\, dX^N}. \qquad 37.$$

Thus, obtaining a proper Boltzmann referenced ensemble average $\langle Q \rangle$ from a simulation that involves non-Boltzmann sampling involves forming the auxiliary ensemble averages indicated in Equation 28 over the probability $P_\omega(X^N)$ and forming the quotient indicated.

One obvious limitation of an umbrella sampling approach to determining $\langle Q \rangle$ is that the more the sampling is to be extended, the larger will be the range of $\omega(X^N)$, making the calculation of $\langle \exp[\beta\omega(X^N)] \rangle_\omega$ more prone to numerical uncertainties. In fact, if the variations are especially large, the computed averages will be dominated by only a very few terms, a clearly undesirable effect. Therefore, a limit must be set on the variations permitted by $\omega_0(X^N)$, which effectively puts a limit on the extension of sampling by this procedure. A succession of simulations using first $\omega_0(X^N)$ and then $\omega_1(X^N)$, $\omega_2(X^N)$, . . . can be devised to overcome this problem (166). The complete $w(R)$ for all R can then be constructed by a matching of individual probability functions for which considerable overlap is observed. Any two successive distributions k and l do not coincide absolutely in the overlap region because the normalization constants for the individual distributions, P_k and P_l, depend in a nontrivial way on the window. However, P_l can be renormalized to obtain a new $P'_l(r)$ by

$$P'_l(r) = P_l(r) N_{lk} \qquad 38.$$

such that k and l are referred to the same normalization factor. A common choice has been simply

$$N_{lk} = P_k(R_\alpha)/P_l(R_\alpha), \quad 39.$$

where R_α is, in principle, any point where P_k and P_l overlap. The statistics for the $P'_k(r)$ get poorer toward the tails of the distributions; thus it is essential to match the distribution at a point R_α where both P_k and P_l have simultaneously reasonable error bars. The final function should not, however, depend sensitively on where the matching is carried out. Ideally, the matching should be based on all overlapping points, with higher weight given to points that were sampled more extensively. A formalism for this matching, applicable to multidimensional cases, has been suggested (153).

Torrie & Valleau (223) showed that umbrella sampling can also be used to alter a sampling distribution so that Equation 9 of the perturbation method can be used more successfully.

The Probability Ratio Method

The idea of extracting ΔA from the end points of a potential-of-mean-force determination is the basis for another free energy determination technique (153). If the free energy function $A(R)$ is evaluated as a function of a coupling parameter λ on the [0, 1] interval, the free energy difference ΔA can be simply obtained as

$$\Delta A = kT \ln [g(R_0)/g(R_1)]. \quad 40.$$

In view of the interpretation of $g(R)$ as a probability per unit volume, the procedure based on Equation 40 is referred to as the probability ratio method (PRM). Also, it can be shown that $V(\lambda)$ is constant for the one-dimensional λ; then

$$\Delta A = kT \ln [P(\lambda_0)/P(\lambda_1)]. \quad 41.$$

In Equation 41 the probability ratio appears explicitly.

Thermodynamic Cycle Integrations

A problem frequently encountered in theoretical studies in biophysics and biophysical chemistry is to account for the relative binding constants of a congeneric series of small molecules interacting with a biological macromolecule. Examples are found in sets of enzyme-substrate complexes, enzyme-inhibitor complexes, and drug-receptor interactions. Furthermore, the newly developed techniques of site-specific mutagenesis permit the question to be inverted: For a given substrate (inhibitor, drug), how does an alteration in the structure of the macromolecule, e.g. an amino acid replacement in a protein or a base-pair mismatch in a DNA, affect the binding constant? The methods of free energy simulation provide a novel theoretical approach for this class of research problems when the

three-dimensional structures of molecules involved are fairly well known from X-ray crystallography or can be reliably assumed. Although thermodynamic integrations as described in the preceding section are well known and thermodynamic processes and in some cases cycles are implicit in nearly every problem, these collective ideas were first applied to the ligand binding process by Tembe & McCammon (220).

Consider, for a solution process, the reaction in Scheme 1 where, for example, A is an enzyme or receptor and B_1 and B_2 are members of a congeneric series of substrates, inhibitors, or drugs. A theoretical account of the relative binding affinity,

$$\Delta\Delta G = \Delta G_2 - \Delta G_1, \qquad 42.$$

is desired.

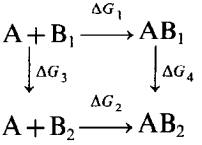
The binding of B_1 and B_2 to A could in principle be studied as independent problems, using for instance umbrella sampling in a series of simulations with the system constrained to successively overlapping regions along a putative reaction path. However, this calculation is complicated by the possible structural changes in A, B_1, B_2, and associated solvent commensurate with binding, so that even if a suitable reaction path could be modeled, a very long and costly sequence of individual simulations would be required to bring each AB complex together from a distant separation and calculate the free energy.

Tembe & McCammon (220) considered the reaction scheme as a thermocycle and suggested that a more practical approach would be to calculate $\Delta\Delta G$ from the equivalent thermodynamic difference

$$\Delta\Delta G = \Delta G_4 - \Delta G_3. \qquad 43.$$

Here ΔG_3 is the relative free energy of B_1 and B_2 in solution and ΔG_4 is the relative free energy of the AB_1 and AB_2 complexes in solution. However, while ΔG_1 and ΔG_2 describe chemical processes that are well defined in principle, ΔG_3 and ΔG_4 describe nonphysical processes, i.e. the free energy changes on mutating B_1 and B_2 in solution and in the AB complex, respectively.

The calculation of ΔG_3 and ΔG_4 using free energy methods turned out to be relatively simple, at least in principle, compared with calculation of



Scheme 1

ΔG_1 and ΔG_2. For ΔG_3, a λ coordinate with $\lambda = 0$ corresponding to B_1 and $\lambda = 1$ corresponding to B_2 is established. Then

$$E(\lambda) = (1-\lambda)E(B_1) - \lambda E(B_2) = E(B_1) - \lambda \Delta E(B_2 - B_1). \quad 44.$$

Equation 44 can be inserted into either a classic thermodynamic integration or the perturbation method to give ΔG_3. Analogous procedures can be applied to the complex to obtain ΔG_4. Tembe & McCammon (220) used the perturbation method for the free energy calculation and called their approach the "thermodynamic cycle perturbation method." The theory is also readily adapted for treating a site-specific mutation within the macromolecular species A; just let A be the ligand and B the protein.

The novelty of the Tembe-McCammon approach emerges from the facts that (*a*) in a cycle alternative routes to the calculation of $\Delta\Delta G$ are available, some physical and some nonphysical, and (*b*) free energy calculations on legs that are nonphysical or fictitious, i.e. legitimate conceptually but impossible in the laboratory, are feasible and in fact readily accomplished by theoretical simulations. The calculation for a nonphysical leg of a thermodynamic process was already implicit in Kirkwood coupling (118) and had in fact been reported several times previously for molecular systems (154, 178). Jorgensen & Ravimohan (108) used the perturbation method to calculate ΔG for converting methanol into ethane. This case is similar to the kind of problem encountered in integrating through a generic series. Tembe & McCammon (220) pointed out that in thermodynamic cycle integrations the computational convenience of treating fictitious legs of a cycle dictates the choice between the alternative routes for calculating $\Delta\Delta G$, which led to Equation 43. Warshel and coworkers (233–241) have made use of thermodynamic cycles in this form for some time in their theoretical calculations, but primarily in the context of Equation 42; the mutational alternative was not explicitly seized until Tembe & McCammon (220) did so. Further perspectives have been given by van Gunsteren & Berendsen (228).

Additional important contributions on thermodynamic cycle integrations via molecular simulations have recently appeared. A quantity of considerable intrinsic interest is the absolute (as opposed to relative) free energy of binding in solution, which is also required to properly define an origin for a potential of mean force determination as described in the preceding section. Cieplak & Kollman (47) described one approach to computing the absolute free energies of binding for nucleic acid base pairs in which, assuming no internal degrees of freedom, a number of independent simulations are involved. Jorgensen et al (103) subsequently pointed out a more efficient procedure that avoids the most computationally demanding steps and all of the gas-phase calculations. Their

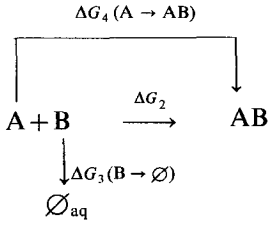
thermocycle (Scheme 2) can be recognized as a special case of Scheme 1 in which $B_1 = \varnothing$, i.e. ΔG_3 corresponds to annihilation of B in solution. Thus the free energy of binding in solution relative to the gas phase can be obtained, and computation of the absolute free energy requires only two independent simulations, annihilation of B in water and mutation of AB to B in the solvated complex.

Madura & McCammon (133) extended thermodynamic cycle integration to the calculation of free energies of activation in locating condensed-phase transition states by an interpolation procedure.

Thus thermodynamic cycle integrations draw upon a unique aspect of the theoretical calculations: the ability to model a nonphysical mutation of a chemical system to produce convenient estimates of useful thermodynamic quantities. Additional recent examples of this type of calculation are described in succeeding sections. Like all of the other free energy calculations, this method works best when the systems considered are quite similar, so that the change in the structure of the AB complex for B_1 and B_2 is minimal. If a major conformational change were to take place, it might not be picked up in the customary $\mathcal{O}(10–100)$ psec of sampling, and its absence would introduce a serious quasi-ergodic problem into the calculation and an error in the calculated $\Delta\Delta G$. Also, the interpretation of the results to give a molecular account of the binding process may be difficult. Overall, however, in view of the importance of this class of problems in the biological sciences, the thermodynamic cycle integration approach is timely and worth pursuing (141).

Nonsimulation Approaches to Free Energy

Numerical estimates and quantitative accounts of free energy were an important element of molecular biophysics long before the emergence of molecular simulation, although they were by necessity based on ad hoc rather than ab initio (molecular) models. The oldest, but still informative, approaches of this type are continuum models for solvation in which a dissolved molecule is treated as a charge, dipole, or discrete multipole



Scheme 2

expansion in a cavity embedded in a homogenous, structureless dielectric. The multipoles polarize the environment and induce a potential or field, which acts back on the multipoles to an extent measured by the polarization energy of the system. The reaction potential induced by a single charge is used to estimate ion solvation energy via the Born equation (32),

$$\Delta G = \tfrac{1}{2}[(1-\varepsilon)/\varepsilon]q^2/a, \qquad 45.$$

where q is the charge, a is a cavity radius, and ε is the dielectric constant of the continuum. Analogously, the reaction field of a dipole gives the Onsager equation (163),

$$\Delta G = [(1-\varepsilon)/(2\varepsilon+1)]\mu^2/a^3, \qquad 46.$$

where μ is the dipole moment. The right-hand sides of Equations 45 and 46 are both leading terms in a multipole expansion for the polarization energy of a discrete charge distribution (30).

The Born equation and the Onsager equation have been used extensively to provide estimates of the solvation energies of ions and molecules, although empirical modifications and adjustments of these simple equations have been found necessary to account for experiment. For instance, Latimer et al (123) showed that modification of the cavity radii can lead to a quantitative account for ion hydration energies; Rashin & Honig (184) recently extended this approach. Alternatively, the deviations could be treated by a local dielectric region, which leads to concentric continuum theory (30). The Onsager equation has been used extensively to estimate molecular solvation energies (1, 190, 208, 231); one simple important prediction of the reaction field theory in the area of conformational equilibria is that the isomer with the largest dipole moment will be preferentially stabilized by polarizable solvent. Rashin & Namboodiri (185) described a generalization of the reaction field approach to molecular solvation.

The main limitation of the continuum theories is their inability to account for specific solute-solvent interactions such as the hydrogen bonding prevalent in aqueous solutions. This limitation has led to the development of a sequence of hybrid "supermolecule-continuum" type theories (200, 201), of which the extreme is fully discrete molecular simulation. Warshel (233, 234) and Warshel & Levitt (237) have performed a number of studies on this genre of biological molecules involving free energy estimates. Their method, called "protein dipoles/Langevin dipoles" (PDLD), is a nonsimulation theoretical approach to estimating the polarization energy due to the environment. Continuum models have been used with molecular simulation to extrapolate solvation free energies from a finite number of molecules to infinite system size.

Theoretical studies of proteins along this line include the work of Tanford & Kirkwood (219), who considered a protein as a sphere of low

dielectric surrounded by a high dielectric, polarizable continuum representing solvent water and treated the system via the Poisson-Boltzmann (PB) equation. Refinements incorporating the surface accessibility of protein atoms were subsequently introduced with considerable success (136). Recently, numerical solutions to the PB equation for a protein, taking account of all charges as well as the dielectric boundary of the protein, have been pursued (242) and applied extensively (120). PB treatments of nucleic acids have also been performed (137, 165), most recently including treatment of the dielectric boundary by Jayaram et al (94).

Another approach to estimating the free energy of hydration is the hydration shell model developed by Scheraga's group for use in conformational energy studies of proteins (86). The method involves a summation of atom or group contributions empirically parametrized to provide the solvation free energies. Solvent exposure must be determined, and Eisenberg & McGlachlan (62) have reported a method based on solvent accessibilities that has produced interesting results. Here as well, extensions to nucleic acids have been proposed (85), but the method is not yet as well parametrized as that for proteins.

Extensive characterization studies are required to establish that a method predicts solvation energies reliably. Thus methods that specifically consider molecular structure but do not involve the computational demands of a molecular simulation are of interest. The statistical mechanical theory of molecular solutions provides an alternative based on Ornstein-Zernicke integral equation theories for solvation phenomena (147). In the restricted interaction site model (RISM) theory (42, 204), the integral equations are used to determine distribution functions of the solvent relevant to solute atoms, from which solvation energies and solvent effects on conformational stability can be calculated numerically (253). The extended RISM theory of Rossky and coworkers (45, 82, 84, 173, 174) is a useful formulation for the type of problems under consideration herein, but approximations such as closure are necessary, and extended RISM must be considered an approximation to fully discrete simulation. A number of comparisons between extended RISM and simulation have recently been set forth; reasonable accord but not perfect agreement can be expected. Calculated free energies were found to generally agree with simulation at the level of roughly kT (253).

APPLICATIONS TO CHEMICAL SYSTEMS

Liquid Water

An early attempt to calculate the absolute free energy of water from Monte Carlo simulation was reported by Sarkisov et al (196) in 1974. The free

energy calculation was based on direct application of the ensemble average expression for free energy, Equation 9, with only about 60K configurations of sampling for a system of 64 particles. The result, -5.7 kcal mol^{-1}, appears reasonable, but subsequent critical analysis by Owicki & Scheraga (164) indicated that the sampling must be inadequate in the higher-energy regions of configuration space and that the result must be somewhat fortuitous.

Mruzik (157) used thermodynamic integration to calculate free energies for water clusters of 8 and 64 molecules described by both the empirical ST2 potential (209) and a quantum mechanical potential of Lie & Clementi (126), both pairwise-additive models. Simulations were carried out by Monte Carlo procedures, with the close encounters problem handled by an integral transform. The results, in the range of -10 to -12 kcal mol^{-1}, are low, but they are typical in comparison with subsequent calculations for liquid water based on ST2 (29).

Mezei et al (154) reported a successful application of free energy simulation to liquid water in 1978. The water-water interactions were described by the MCY-CI(2) potential function developed by Matsuoka et al (135) from ab initio quantum mechanical calculations. The simulation procedure was Metropolis MC, and the free energy was obtained by thermodynamic integration using an eight-point Gaussian quadrature, the λ parameter coupling up all molecules in parallel. The reference state for the calculation was ideal gas at the corresponding liquid-water density, and the free energy integration was effectively a mutation of ideal gas into liquid water along an isodensity tie line. The calculated free energy was -4.31 ± 0.07 kcal mol^{-1}, while the observed value was -5.74 kcal mol^{-1}; the 10% discrepancy was attributed to deficiencies in the potential functions and the assumption of pairwise additivity with a concomitant neglect of cooperative effects.

Shortly thereafter, Mezei (149) reported an MC determination of the relative free energy for three pairwise water models, ST2, MCY, and SPC (22), with respect to a soft-sphere reference system. Free energy calculations based on thermodynamic integration, the perturbation method incorporating an umbrella sampling approach, and a procedure called the overlap ratio method were compared and contrasted. The three-point SPC model, involving a semiempirical parametrization, showed the closest accord with experimental free energy, $\sim 7\%$. Recently errors were uncovered in some of the calculations (M. Mezei, private communication). Sussman et al (215) calculated the free energy of liquid water at various temperatures based on the perturbation method with umbrella sampling.

Most recently, Hermans and coworkers (81) studied the free energy of liquid water in isothermal molecular dynamics procedures using an ideal

gas reference state and also coupling all molecules in parallel. Applications were reported for two test systems, soft spheres and liquid argon, and for the water models SPC and SPCE (20) from the Groningen group and TIP3P (105) and TIP4P (107) from Jorgensen's laboratory. Stepwise perturbation and slow growth algorithms were both employed. The results for SPC were reported to be within 0.1 kcal mol^{-1} of experiment at three different temperatures. The TIP4P results as recently redetermined by Jorgensen et al (100) were nearly identical, -5.9 ± 0.4 kcal mol^{-1}; the experimental number for comparison is -6.3 kcal mol^{-1}.

Cavities in Water

A series of studies has focused on the free energy required to form cavities of various dimensions in liquid water; this is a preliminary step in the study of aqueous solvation. Postma et al (178) used a series of six independent MD simulations of 25–30 psec duration each to obtain the free energy of a cavity with thermal radius of 3.17 Å. Using SPC water, they first estimated the free energy of creation of a 1.0-Å cavity from scaled particle theory (175), thereby avoiding the close encounters problem. The perturbation method was then used for increasing radii. The results agreed well overall with those of scaled particle theory. The results were used to estimate a theoretical value for the surface tension of water, 0.067 N m^{-1}, which agrees closely with the experimental value of 0.071 N m^{-1} at 305 K.

Postma (177) then explored the method of slow growth for the same problem. The initial state was thus unperturbed liquid water, and the final state was the liquid water system containing the full-size cavity; intermediate states described the gradual growth of the cavity in water. Postma et al (178) calculated the free energy in two different MD simulations, one of 10 psec and the other of 25 psec. The variation of the λ parameter with time was chosen to make $U(\lambda)$ linear in each case. The results of the 10-psec slow growth deviated significantly from those of the discrete method, and a significant hysteresis of ~ 5 kJ mol^{-1} was also observed. The discrepancies are largest for small λ, as here the perturbation of forming the cavity is relatively larger. The results for the 25 psec run were closer to those of the discrete method and exhibited overall less hysteresis. For larger cavities (≥ 7.0 Å) the results of the shorter and longer runs are very similar. Thus the slow growth method with thermodynamic integration led to reasonable estimates of the free energy of formation of a 3.17-Å cavity in one sixth the time of the discrete method. This demonstrates that as long as the phase space is adequately sampled in the simulation, considerable computational economies can be realized by slow growth.

Hydrophobic Effects

The hydrophobic effect is one of the major factors stabilizing the folded, tertiary structures of biological macromolecules, and an accurate account of this phenomenon via simulation is thus of considerable interest in biophysics and biophysical chemistry (66). Hydrophobic effects include both the dissolution of apolar species in water, i.e. hydrophobic hydration, and the spontaneous association of apolar species in water, i.e. hydrophobic interaction or bonding. The interaction of water with solute is weakest for apolar species; thus hydrophobic systems are also important prototypes in the general theory of aqueous solvation and solvent effects. They are not the simplest of systems, however, since entropic contributions tend to dominate and sampling problems can be anticipated. Ravishanker & Beveridge (186) have reviewed theoretical studies of the hydrophobic effect through 1986, including those involving simulations. Both hydrophobic hydration and hydrophobic interactions have been the subject of free energy studies via molecular simulation.

HYDROPHOBIC HYDRATION Apolar molecules such as hydrocarbons and apolar groups such as the methyl group on alanine and the phenyl on tyrosine exhibit a number of unusual properties, particularly a low and temperature-dependent solubility in water. The enthalpies of interaction are usually small and negative and are primarily due to dispersion effects. The low solubility arises as a consequence of entropy; the waters of the hydrophobic hydration complex are indicated to be in a more ordered state than liquid water. This state appears to involve clathrate-like structures (216, 252), originally called "icebergs" by Frank & Evans (65) in an early theoretical account.

The earliest simulation study on apolar solvation was a series of MC calculations by Dashevsky & Sarkisov (53) on a hard sphere and a methane molecule, each solvated by 63 water molecules. This study involved a sampling of only 10K per particle and is now subject to concerns on convergence (164). Nakanishi's group (162) reported a free energy calculation on the hydrophobic hydration of a single Lennard-Jones solute in 63 ST2 water molecules using a single-step perturbation method with a surrogate potential and umbrella sampling. The simulation was carried out using MC methods for the (T,V,N) ensemble. Sampling was reported for 500K configurations, with averages formed over 250K or 3.9 configurations per particle. The calculated free energy was 0.76 kJ mol^{-1}, with an internal energy and an entropy (actually $-TS$) of -1.7 and -2.5 kJ mol^{-1}, respectively. These results give a reasonably good account of the solubility of a simple apolar solute in water.

It is useful in simulation studies of the absolute free energy of solvation

to decompose the overall process into two steps, (*a*) cavity formation and (*b*) introduction of solute, and to determine the free energies independently or sequentially. The simulations and calculated results for the free energy of cavity formation described previously (177, 178) can be used as a reference state in the calculation of the free energy change on introducing a solute into the cavity. Straatsma et al (212) performed such a calculation for the hydration of the rare gas solutes Ne, Ar, Kr, and Xe using the perturbation method. They used the set of simulations described above for cavities of various sizes together with the various rare gas–water interaction potentials to explore the criteria for successful application of the perturbation method. The introduction of a Ne atom into cavities of various sizes demonstrated that the nature of the reference ensemble makes an extremely large difference in the ΔGs calculated via the perturbation method; the best results were obtained for a reference cavity comparable in size or slightly smaller than the solute. Improved agreement was obtained by a two-step process of first altering the cavity radius and then introducing the solute. Smaller perturbations gave more accurate results as expected. The first-order approximation to free energy gave a relatively poor estimate in this case, since entropy effects are important. The best results were in close accord with experiment. However, given the considerable exploratory calculations involved, the ab initio prediction of free energies of solvation at high accuracy would have been difficult. Particularly, Straatsma et al (212) demonstrated that it would have been tempting to accept results with low statistical uncertainty when the problem was really quasi-ergodic. In any case, they developed valuable methodological perspectives in this study.

Swope & Anderson (217) went a considerable step further in simulating the hydrophobic hydration of rare gas atoms, treating the solubility problem by means of MD calculations and thermodynamic integrations to obtain the partial molar free energy of the solute. They also examined the calculated temperature derivatives of solubilities in water; rare gas atoms in water have solubility profiles that exhibit a more negative first derivative and a more positive second derivative than in nonpolar solvents. The authors dealt with the close encounters problem by using a rounding-off procedure at small solute-water separations. Their simulation reproduced the small values, negative slope, and positive curvature characteristic of experimental temperature-dependent solubilities of rare gases in water. However, the exact behavior was sensitive to parameters that describe the solute. The data for Ar and Kr were fit with reasonable choices, but fits for the data for He and Ne were not considered satisfactory. The theory obviously works better for larger solutes. The authors expressed concern about an additional source of error associated with some processes with

long relaxation times in the system, which is a type of quasi-ergodic problem.

There have been several recent reports of the absolute free energy of hydration of methane in water, obtained via the perturbation method with λ as a creation coordinate in the process $[\varnothing \rightarrow CH_4]_{aq}$, where $\varnothing$ represents annihilation of the solute. The experimental value reported for this free energy was about 2 kcal mol^{-1} at 300 K (15). Bash et al (11) studied a system of solute plus ~400 TIP3P water molecules at 300 K and $P = 1$ atm via (T,P,N)-ensemble MD. An 8-Å nonbonded cutoff was applied to the potentials. Slow growth was effected over a period of 30 psec, and the average of forward and backward results was reported. The calculated free energy of hydration of 2.1 ± 0.94 kcal mol^{-1} compared favorably with the experimental value. Curious trouble in changing van der Waals parameters was reported with windowing methods; equilibration would be expected to take care of this.

Jorgensen et al (100) recently calculated the free energy of hydration for methane in water as 2.5 ± 0.4 kcal mol^{-1} using the perturbation method for six windows of doubled-ended, double-wide sampling via MC simulation, with no apparent numerical problems. An exploratory calculation with fewer windows gave a spurious and misleading agreement with experiment, since some of the individual steps were obviously of the wrong sign. This example documents the kind of problem that may arise from improper computational design in free energy simulations.

Fleischman & Brooks (64) applied similar procedures, the perturbation method for free energy and MD simulation, to determine the differential free energies of hydration of ethane and propane and some hydrophilic species mentioned below. For the free energy difference for $[C_2H_6 \rightarrow C_3H_8]_{aq}$, the calculated result (-0.83 kcal mol^{-1}) and the experimental number (0.13 kcal mol^{-1}) disagree in sign, although both are of small magnitude. Fleischman & Brooks (64) presented finite difference formulae for the determination of enthalpy and entropy, which they found more reliable than ensemble average differences. The calculated $\Delta\Delta U^{hyd}$ has the correct sign but differs by 1.6 kcal from experiment. The calculated $\Delta\Delta S$ also has the correct sign but differs from experiment by a factor of 2. The authors expressed concern about the united atom model and the lack of explicit hydrogen atoms.

HYDROPHOBIC INTERACTIONS The central dogma of the hydrophobic effect (115) holds that apolar moieties associate spontaneously in water as a consequence of the increase in entropy obtained when exposure of the apolar surface (and water ordered as a consequence of hydrophobic hydration) is reduced. This idea leads to prediction of an inverse tem-

perature dependence, now considered diagnostic of the phenomenon. The true nature of the hydrophobic interaction has remained a matter of controversy. Franks (66) suggested that it involves a stable solvent-separated association; this view has been supported by extensive theoretical analysis of the hydrophobic effect and integral equation calculations by Pratt & Chandler (179). The simulation era has brought new perspectives on the problem but has not yet produced an unequivocal and fully satisfactory explanation for the experimental observations.

The first simulation on a hydrophobic interaction was a 1979 study by Geiger and coworkers (68) on the molecular dynamics of two Ne atoms in ST2 water (68). Only a short run was possible then, but the study showed that a Ne pair initially in contact diffused in the course of a few picoseconds to another structural form of at least transient stability, with the two solutes separated by on the average of one solvent molecule. MC calculations of the mean force for apolar, Lennard-Jones particles in ST2 water (165a) and for two methanes in MCY water (215a), while subject to considerable numerical inaccuracy, supported the idea of a contact and solvent-separated regions of stability.

In 1979 Pangali et al (166, 167) reported a successful determination of the potential of mean force for two Lennard-Jones particles in ST2 water and described methodology for matched umbrella distributions, which has found wide application. They found clear evidence for the existence of contact and solvent-separated minima in the calculated $w(R)$, with the latter 0.25–0.4 kcal mol^{-1} higher in free energy and a barrier between them of ~ 1 kcal mol^{-1}. The statistical weight of the solvent-separated complex is however higher than that of the correct form owing to the larger volume of configuration space available. In a related study, Ravishanker et al (187) reported conceptually similar results for two methane molecules in MCY water; the $w(R)$s differed in magnitude owing primarily to intrinsic differences in the potentials. These authors (187) and also Zichi & Rossky (252) analyzed the hydration structure of the contact and solvent-separated forms in some detail. Ravishanker & Beveridge (185a) also reported a potential of mean force for the stacking of phenyl rings in aqueous solution, which gives evidence for a solvent-separated but not intercalated structure.

Recently Jorgensen and coworkers (103), following calculations described above for the hydrophobic hydration of methane, reported a redetermination of $w(R)$ for the methane dimer in TIP4P water at high statistical precision using the perturbation method. The result was cast on an absolute basis by means of the thermodynamic cycle integration described in the previous section. The results on the free energy of binding as well as the potential of mean force were consistent with the prior findings from simulation and integral equation theory. Thus the idea of the solvent-

separated hydrophobic interaction is now well established by theoretical calculations. Experimental evidence has not been readily obtained owing to the low solubility of hydrophobic species in water; the idea that solvent-separated structures account for the variation in the partial molar volume of ethanol with concentration and the higher *gauche* population for *n*-butane in water than for the neat liquid remains provocative.

The present view of hydrophobic interactions is further complicated by the results of two independent molecular dynamics simulations, each involving a system of several apolar solutes in water. Rappaport & Scheraga (183) reported a 70-psec MD simulation for four apolar solutes in 339 water molecules at ~300 K. The simulation failed to reveal any tendency of the initially well-separated solutes to associate spontaneously. The authors noted that the duration of the simulation may have been insufficient for observation of the association phenomena. Recently Watanabe & Andersen (243) reported an extensive MD study of Kr-type solutes in the revised central force model for water used in the study of hydrophobic hydration by Swope & Andersen (217) cited above. Results at infinite dilution for various system sizes ranging from 64 to 394 water molecules confirmed that the effective pairwise $w(R)$ for the system had contact and solvent-separated minima. Simulations on systems of multiple solutes in water, run for up to 348 psec, also exhibited no strong tendency of the solutes to associate. In fact, in comparison with randomly distributed particles, the solutes on the average tended to avoid one another! This behavior translates into a positive value for the osmotic second virial coefficient, B_2. Analysis of experimental data on the pressure dependence of several small nonpolar gaseous solutes in water (243) indicates that solution nonideality is dominated by B_2 and appears indeed to exhibit a large positive value. The statistical mechanics and molecular thermodynamic description of the hydrophobic interaction thus remains unsettled.

Hydrophilic Hydration

In hydrophilic and ionic hydration the interactions are stronger than in hydrophobic systems and the perturbations on cavitated water are significantly larger. Direct use of the perturbation method to obtain free energies of solvation is no longer feasible, and a multistep approach is invariably required. Thus investigators in this area turned their attention to calculating the relative free energy of hydration of structurally related solutes.

METHANOL AND ETHANE In 1985, Jorgensen & Ravimohan (108), in what has become a prototypical application of the perturbation method to

relative free energy changes on a mutational coordinate for molecular systems, calculated the differential free energy of hydration for methanol and ethane. The experimentally observed differential free energy change was $+6.93$ kcal mol^{-1} (16), too big a leap for direct use of the perturbation method. The OH of CH_3OH is mutated into CH_3, which involves a complete change from hydrophilic to hydrophobic hydration. Jorgensen & Ravimohan (108) used a stepwise approach with intermediate states between $\lambda = 0$ and $\lambda = 1$ involving simulations on an unphysical "λ-cule," part CH_3OH and part C_2H_6. In the course of the mutation, not only potential functions but also geometrical parameters must be altered as a function of λ; thus the dependence of ΔE on λ is not necessarily linear. In the simulation, the oxygen of methane was maintained coincident with one of the ethane carbons and was mutated into a (united atom) methyl group. The hydroxy hydrogen of methanol has only a charge parameter, which disappeared as a function of increasing λ. The carbon of methanol was progressively mutated into the other carbon of ethane, with λ altering the bond length from 1.43 Å to the ethane value, 1.53 Å. Calculations were performed on the mutational coordinate at $\lambda = 0.0, 0.125, 0.25, 0.50, 0.75$, and 1.0; note that an extra step was inserted as $\lambda \rightarrow 0$ since the solute is altered most drastically in this region. Double-wide and double-ended sampling procedures were used.

The simulations at each λ value were carried out using (T,P,N)-ensemble MC Metropolis procedures at $T = 25°C$ and $P = 1$ atm on a system of one solute and 125 water molecules in a cube under periodic boundary conditions (108). A united atom model was used for CH_3, and the TIP4P model was used for water–water interactions. Each simulation involved 500K steps of equilibration and 1500K steps of production, over which ensemble averages were formed. This processing amounted to 45 psec of dynamics for each simulation. The calculated differential free energy of solvation was 6.75 ± 0.2 kcal mol^{-1}, in close accord with experiment and validating both the perturbation and the potentials. The forward and backward incremental free energy changes along the λ-coordinate were comparable within statistical uncertainties except for the $\lambda \rightarrow 0$ first step. This step involved the largest change in free energy, $\sim 40\%$ of the total. The deviation demonstrates once again how the precision of the method is inversely proportional to the size of the perturbation.

Bash et al (11) compared both stepwise and slow growth techniques for the $[CH_3OH \rightarrow C_2H_6]_{aq}$ mutation. Anticipating applications of the perturbation method to biological problems of large dimension, they explored less extensive sampling protocols. The simulations involved (T,P,N) MD at $T = 300$ K and $P = 1$ atm on a system of one solute and 214 water molecules, with OPLS potentials for solute-water interactions and the

TIP3P model for water. Three determinations were reported; the first consisted of 15 windows, 10 for $\lambda/2$, where the energy change is largest, and 5 for the remainder of the interval. Each window involved 1 psec of equilibration and 1 psec of production. This determination produced a $\Delta\Delta G$ of 6.97 ± 0.1 kcal mol^{-1} based on doubled-ended sampling. Evening out the windows and sampling only one fifth of this amount gave $\Delta\Delta G$ of 6.67 ± 0.3 kcal mol^{-1}. Finally, a $\Delta\Delta G$ of 6.93 kcal mol^{-1} was obtained with only 14 psec of slow growth. Similar results were obtained for the $\Delta\Delta G$ of hydration for $H_3O^+ \rightarrow NH_4^+$. The methanol-to-ethane mutation in water has subsequently become the benchmark system for testing diverse free energy procedures in programs for molecular simulation; it has been repeated in at least six different laboratories to date with substantially similar results.

AMINO ACID RESIDUES AND NUCLEIC ACID BASES Solvent effects on protein and nucleic acid structure and process can ultimately be explored by free energy simulations. Documentation of results for appropriate prototype systems is first appropriate. Wolfenden et al (249) have established a scale of hydration potentials based on partition coefficients for amino acid residues and other biological molecules, which provides an experimental data base for comparison of simulation results with experiment.

Singh et al (206) explored the calculation of $\Delta\Delta G$ for the mutation of $[\text{Ala} \rightarrow \text{Gly}]_{aq}$ and, for a big change, $[\text{Ala} \rightarrow \text{Phe}]_{aq}$. For the former, they considered a system of zwitterionic amino acid and 367 water molecules. Two MD stepwise simulations, a uniform one of 30 psec and a biased one of 45 psec, led to $\Delta\Delta G(\text{Ala} \rightarrow \text{Gly})$ of -1.16 and -1.01 kcal mol^{-1}, respectively. Next a slow growth simulation of 60-psec total length gave $\Delta\Delta G = -0.76$ kcal mol^{-1}. The variation of 0.3–0.4 kcal mol^{-1} was comparable to that found in previous simulations. In comparison, the experimentally observed relative free energy of solvation was -0.2 and -0.3 kcal mol^{-1} for methane/ethane and propane/isobutane, respectively. The discrepancy indicated that the balance of repulsive (cavity formation) and attractive (dispersion) forces might have been a little off because of inaccuracies in either the simulations or the potential functions (206). A serious quasi-ergodic problem would probably not be expected in this case, and the problem is more likely to be in the potential functions.

The $[\text{Ala} \rightarrow \text{Phe}]_{aq}$ problem is somewhat more challenging. Singh et al (206) started the simulation with a phenyl ring of greatly reduced size, i.e. C–C bond lengths of 0.2 Å for the ring and 0.5 Å from C_β to the ring C. As λ goes from initial to final conditions, the phenyl ring grows smoothly to its full dimension. A system of 545 water molecules plus solute was considered in four simulations, two involving windowing and two involv-

ing slow growth. The calculated $\Delta\Delta G$ [Ala → Phe]$_{aq}$ ranged from −2.29 to 3.84 kcal mol^{-1}. The comparable [$CH_4 \rightarrow \phi CH_3$]$_{aq}$ system had an experimentally observed $\Delta\Delta G$ of −2.7 kcal mol^{-1}. The calculated results, while not of the quality obtained for the mutation of methanol to ethane, are at least in the right ballpark, which indicates that calculation of $\Delta\Delta G$ of hydration is feasible when large structural changes are involved. Furthermore, the results from slow growth are comparable to the results from windowing or stepwise procedures.

The Kollman group went on to establish a simulation protocol and ran through calculations of $\Delta\Delta G$ for many amino acid side chains, nucleic acid bases, and small organic model compounds (11). Each MD simulation involved 400 TIP3P rigid water molecules at $T = 300$ K and $P = 1$ atm in a box under periodic boundary conditions with an 8-Å cutoff. The electrostatic charges for the amino acid residues were not those of AMBER 3.0, but fit to electrostatic potentials from 6–31G* ab initio molecular orbital calculations. The charges were linearly scaled if necessary to fit the experimental dipole moment. The calculations were done in two stages. First the charges were altered while the van der Waals parameters were maintained at their initial values. Then the van der Waals radii and well depths were slowly perturbed to their final values while the parameters for bond lengths and angles were left unchanged. Electrically neutral molecules were treated with 20 windows of $\Delta\lambda = 0.05$, with 1-psec equilibration and 1-psec production. For charged residues, 40 windows with $\Delta\lambda = 0.025$ were used with 0.5-psec equilibration followed by 0.5-psec production. Changes in van der Waals parameters were performed with 30 psec of slow growth rather than windowing, owing to large energy fluctuations encountered in the latter. More stable results were obtained with two 30-psec double-ended sampling runs. The overall agreement between calculated and observed $\Delta\Delta G$s was generally within 10%. Free energies due to changes in van der Waals parameters were in general less precise than those due to electrostatics. The results are reasonable, however, and with the results for the additional model compounds acetamide, acetic acid, and phenol support Wolfenden's (249) experimentally determined scale of solvation potentials. Calculations were also reported for the charged amino acids Arg, His, and Lys (11), but uncertainties in reaction-field corrections for a nonspherical case limited comparison with experiment.

Cieplak & Kollman (47) treated the nucleic acid bases A, T, G, and C in a similar manner and obtained results on [$CH_4 \rightarrow A$]$_{aq}$, [$CH_4 \rightarrow T$]$_{aq}$, [$T \rightarrow C$]$_{aq}$, and [$A \rightarrow G$]$_{aq}$. An N_1–CH_3 model was used for pyrimidines and an N_9–CH_3 model for purines. Nucleic acid bases appear to be too water soluble for measuring the free energy of solvation directly, but the calculations indicate that T may be accessible. A and T have almost

equal dipole moments yet considerably different calculated free energies of solvation. The charges show that A has a substantially larger quadrupole moment than T, offering a possible explanation.

Ionic Solutions

As with hydrophobic effects, it is convenient to partition this topic further into ionic hydration and ionic interactions in water.

IONIC HYDRATION The study of ion-water systems via molecular simulation began with Monte Carlo studies of Mruzik et al (158) on alkali metal cations and halide anions with clusters of 2–6 water molecules. These studies were motivated by measurements of the free energy for successive addition of water molecules to the ion-water complexes using high-pressure mass spectroscopy (60). Mruzik et al (158) used MC methods with thermodynamic integration to calculate the differential free energies, with a λ parameter that serves to couple an additional water molecule into the system. The close encounters problem was treated by an integral transform that carried the integration to a nonlinear regime. The intermolecular potential functions were taken from earlier work based on quantum mechanical calculations (119, 126). The magnitudes and general trends in the observed free energies were well supported by the calculations, although the calculated free energies of the cations tended to be more negative than experimental values whereas for the negative ions the comparison was mixed. Discrepancies appeared in the calculated enthalpies as well and were most likely due to inaccuracies in the potential functions because of superposition error in the quantum mechanics and the assumption of pairwise additivity in forming the configurational energy. The experimental free energy for F^- as a function of added water showed an anomalous decrease at $N = 5$, which was probably an artifact of experiment. With this exception, the calculated trends were in close accord with experiment. The Cl^- values were within experimental error and statistical uncertainty.

Lybrand et al (127) determined the relative free energy of transfer of Cl^- and Br^- from gas phase to water as a demonstration of the use of the Tembe-McCammon equation (Equation 43). The problem reduces to the free energy change in mutating Cl^- to Br^- in water. The authors considered an anion with 214 waters at $T = 330$ K, with potentials calibrated to reproduce internal energies of the hydrated complex. MD and the perturbation method were used to calculate the relative free energy change. Since only a small change in ionic radius is involved, the free energy was computed in one step. Close agreement of 3.35 ± 0.15 kcal mol^{-1} calculated versus 3.3 kcal mol^{-1} observed was obtained. However, the internal energy

difference is a leading approximation to the free energy difference and is dominant over entropy in this case (211).

Lybrand et al (128) subsequently incorporated the previous calculations into theoretical binding affinities of the macrotricyclic ionophore SC-24 as a prototype of the ligand-binding problem. The additional calculation involved is the mutation of Cl^- to Br^- in the receptor site. The relative free energy of binding was found to be 4.15 ± 0.35 kcal mol^{-1} versus 4.0 kcal mol^{-1} observed. The selective binding of Cl^- over Br^- was seen to be due to the more favorable interaction of Cl^- with the receptor, which more than compensates for the unfavorable free energy of desolvation. Lybrand et al (128) emphasized that the method is useful only for similar structures; otherwise longer simulations would be required to sample the system.

Brooks (35) subsequently took on the Cl^- to Br^- hydration problem to explore more precise determinations of the energy and entropy of hydration based on temperature derivatives. He claimed savings of 50% for his finite difference formulae (see Reference 36 for corrected versions). Brooks (35) also explored the effect of truncation effects on the potentials by comparison with the calculations of Lybrand et al (127) and showed that 10–25% discrepancies in free energy could be attributed to these effects.

Straatsma & Berendsen (211) are currently exploring the sensitivity of free energy calculations for the aqueous hydration of Na^+, K^+, Ca^{2+}, F^-, Cl^-, and Br^- ions to total simulation time, size of time step, system size, and cutoffs in the potential functions. They have carried out (T,P,N) MD simulations with $T = 298$ K for systems of 216 or 512 SPC water molecules. Free energies were evaluated by thermodynamic integration using a slow growth protocol. The mutation $Ne \rightarrow Na^+$ served as the prototypical case. Following 50 psec of equilibration, five runs of 5–80 psec were carried out, forward and backward. The forward and backward results were invariably smooth, indicating no problematic fluctuations, and for 80 and 40 psec the hysteresis was only ~ 1 kcal mol^{-1}. For shorter runs, the hysteresis became significant, i.e. 18, 28, and 39 kcal mol^{-1} for runs of 20, 10, and 5 psec, respectively; here the relaxation from hydrophobic to ionic hydration did not have sufficient opportunity to equilibrate. The time step of 0.001 psec gives satisfactory results, but when doubled gives a hysteresis of 8.71 kcal mol^{-1} over 40 psec. Increasing the system size from 216 to 512 water molecules led to a change of only -3.36 kJ mol^{-1}.

The studies of cutoff radii indicated that truncation of Coulomb potentials considerably affects the results when charged systems are involved (211). The results also showed a sensitivity to the water–water cutoff values. Partial correction can be obtained with Born's equation (32), but

only if the cutoff is sufficiently large to begin with. The Straatsma & Berendsen calculations on ionic hydration compare well with observed values overall, and the discrepancies generally accord reasonably well with expectations when no semiempirical adjustments are made.

Jorgensen et al (100) recently characterized the $\Delta\Delta G$ of hydration for the mutation of CH_4 into Cl^-, based on MC simulations involving 216 TIP4P solvent molecules. Ten steps on a λ coordinate were involved, two of which showed ΔG of ~ 10 kcal mol^{-1}. Nevertheless, the overall hysteresis was modest, 3–7%. The net $\Delta\Delta G$ was -62.4 kcal mol^{-1}. Adding a correction for forming the methane cavity and a Born equation estimate of the contribution beyond the 8.5-Å cutoff led to an absolute free energy of hydration of -79.1 ± 2 kcal mol^{-1} as compared to -77 kcal mol^{-1} observed. Cutoff corrections are clearly important in cases where charge is created or annihilated. These authors discussed the ability of the calculation to provide the free energy without the inclusion of polarization effects in terms of enthalpy/entropy compensation.

Comparisons with the results of the Born model for ionic hydration (32) have motivated two current free energy studies. Hirata et al (83), using mainly RISM but also MD calculation of free energy by the perturbation method, investigated the effective radius problem and detailed the structural differences in cationic versus anionic water interactions. The former are dominated by ion-dipole interactions and the latter by hydrogen bonding. The authors contend that the cavity radii introduced by Rashin & Honig (184) should be considered a size parameter that effectively takes into account microscopic solvent structure.

Honig and coworkers (91) have performed MC studies on ion hydration that simulated the charging of a cation from 0 to $+3$ and explored the effect of cavity radius and dielectric saturation. They compared the results of a (T,V,N) simulation on a sodium-sized ion in 215 water molecules with the Born model. The reaction potential of the solvent varied linearly with the ionic charge as predicted by the Born model, but two linear regimes were found. The inflection point was found at about $+1.1$, where the effects of electrostriction and orientational alignment were found to plateau and a complete shell of six water molecules with dipoles fully aligned was established.

IONIC INTERACTIONS The interaction of ions and ion pair formation in water have been studied as a potential-of-mean-force problem via molecular simulation. Early papers by McDonald & Rasaiah (146) and Patey & Valleau (168) laid the computational groundwork for this genre of calculation. The general procedure, developed further and applied by Pangali et al (166) to hydrophobic interaction (166), can be applied equally

well to ionic interactions. In 1984 Berkowitz et al (25) calculated a potential of mean force for Na^+ and Cl^- using a 12-window umbrella sampling procedure and molecular dynamics on the ions and 295 TIPS2 water molecules in a rectangular parallelepiped. The simulation revealed a characteristic double-well form for $w(R)$ with a contact minimum at 2.7 Å and a solvent-separated minimum, 1.3 kcal mol^{-1} higher, at $\sim$5 Å. Calculated energies and entropies revealed considerable compensation effects in the final form of $w(R)$. RISM calculations support the same essential features, such as the deeper contact minima. Karim & McCammon (109) followed up the Berkowitz et al study with a dynamical study of rate constants for ion pair formation and dissociation in water.

Dang & Pettitt (52, 52a) calculated the potential of mean force between two chloride anions in water with similar techniques to those used by Berkowitz et al (25); they found a surprising effect, a minimum in $w(R)$ in defiance of ionic repulsion, stabilized by intervening water molecules. This minimum may account for certain anomalous effects observed in NMR relaxation studies.

Jorgensen et al (104), in the course of simulation studies of organic reactions in solution, determined a free energy profile for the separation of tert-butyl cations and chloride anions in dilute aqueous solution and compared it with the results from RISM calculations. A contact ion pair was found, but the solvent separated and free ions were not energetically distinct. The RISM $w(R)$ is similar in shape, but the essential features generally have smaller amplitude.

Conformational Equilibria

The presence of a condensed-phase environment is well known to exert an influence on the conformation of a dissolved molecule. Manifestations of this influence in molecular biophysics range from protein folding and denaturation to the dramatic effect of hydration and salt concentration on the conformation and even helix sense of DNA structures. A reliable theoretical account of solvent effects on conformational stability and equilibria is obviously required of molecular simulation. Progress toward this goal for chemical systems ranging from simple to fairly complex is reviewed in this subsection. Jorgensen (98) reviewed this area earlier with an emphasis on organic systems; his article contains much useful critical perspective, especially on convergence issues.

This class of problem is approached via a special type of potential-of-mean-force study in which the degree of freedom is a torsional coordinate ϕ and the quantity of interest is the potential of mean torsion, $W(\phi)$, defined as

$$W(\phi) = V_0(\phi) + w(\phi), \qquad 47.$$

where $V_0(\phi)$ is the intrinsic intramolecular potential energy with respect to ϕ and $w(\phi)$ describes the solvent-induced effects, literally the solvent-induced reversible work required to rotate the coordinate from a reference state ϕ_0 to ϕ. From $W(\phi)$, the normalized distribution of conformers $s(\phi)$ can be obtained from the Boltzmann equation:

$$s(\phi) = \frac{\exp[-\beta W(\phi)]}{\int_0^{2\pi} \exp[-\beta w(\phi)]\, d\phi}. \qquad 48.$$

The equilibria of interest in most chemical and biomolecular applications are either twofold *cis*/*trans* equilibria, with mole fractions given by

$$x_c/x_t = \frac{\int_{-\pi/2}^{\pi/2} s(\phi)\, d\phi}{\int_{\pi/2}^{3\pi/2} s(\phi)\, d\phi}, \qquad 49.$$

or else threefold *gauche*(−)/*trans*/*gauche*(+) equilibria, with mole fraction ratio

$$x_g/x_t = \frac{\int_{-2\pi/3}^{2\pi/3} s(\phi)\, d\phi}{\int_{2\pi/3}^{4\pi/3} s(\phi)\, d\phi}. \qquad 50.$$

n-BUTANE AND OTHER ALKANES One of the simplest organic molecules with a conformational degree of freedom is *n*-butane, which has been studied extensively by simulation in gas phase, in nonpolar solvents, and in various models of water. The intrinsic rotational potential $V_0(\phi)$ has been well determined in MM2 theory (38, 97) and is considered quite reasonable. The threefold potential is characterized by a *gauche*/*trans* energy difference of ~ 1 kcal mol^{-1} favoring *trans* and by a *gauche* → *trans* barrier of 1.5 kcal mol^{-1}. Thus $s(\phi)$ for the gas phase favors the *trans* isomer by 70%. Bell & Harvey (12) recently compared intramolecular free energy surfaces for united and extended atom models of *n*-butane. The free energy surfaces were found to be similar and dominated by energy. However, substantial differences in the entropy behavior were noted. The entropy of the united atom model exhibited minimal variability, but in the all-atom model entropy was greatest in the *trans* conformer and least in the *cis* barrier. Simulations on biomolecular systems typically assume a

united atom model to keep computational requirements as low as possible; this could be a problem with respect to entropy.

A number of groups have studied the conformational equilibria in *n*-butane as a neat liquid. Ryckaert & Bellemans (194, 195) performed MD simulations at $N = 64$ for temperatures of 200, 274, and 292 K. They reported a 12% increase in *gauche* isomers in going from the gas phase to the neat liquid at 292 K. This result was in qualitative accord with an earlier prediction from integral equations (41, 87, 180). The convergence, however, has been questioned (98) owing to the relatively short simulation times (14 psec). Weber (244, 245) and Weber & Helfand (246) in MD studies found essentially no change, and Bigot & Jorgensen (31) found essentially no shift in an MC calculation. Rebertus et al (189) reported a distinct conformational shift toward the *gauche* isomer. Experimental studies (summarized in 97) do not establish an unequivocal position on this issue.

Bigot & Jorgensen (31) and Jorgensen (95–97) have studied this problem further using MC calculations and umbrella sampling with careful attention to convergence details. They reported no significant change in the conformational equilibrium for liquid vis-à-vis gas *n*-butane; x_g/x_t was 0.49 for the liquid at the boiling point versus 0.50 for the gas. The calculations were carried out for different system sizes and in both (T,V,N) and (T,P,N) ensembles. Recent RISM calculations by Zichi & Rossky (253) gave $x_g/x_t = 1.2$ in liquid, with destabilization of the *trans* isomer. A general trend in extended RISM to overestimate condensed-phase effects by $\sim kT$ was noted, however.

There is general agreement from diverse studies that the *gauche* population of *n*-butane in water is increased over that of the gas phase. Rebertus et al (189) found an increase of 9% in going from gas to a Lennard-Jones solvent. Rosenberg et al (191, 192) reported a 35% increase in the *gauche* population at room temperature in ST2 water, but a slight decrease at 60°C. Jorgensen & Buchner (102) reported an 18% increase using TIPS water models. This conformational change was initially interpreted as a classic hydrophobic effect, an incipient microfolding to bring the terminal methyl groups into closer proximity as a consequence of solvent entropy. Zichi & Rossky (253), however, recently disputed this interpretation, showing that internal energy dominates the results of extended RISM theory. A possible link to the anomalous behavior observed in hydrophobic cluster simulations (183, 243) was noted.

All previous calculations on conformational equilibria in *n*-butane involved umbrella sampling to calculate the conformational preferences. Recently Jorgensen & Buchner (102) redetermined the potential of mean torsion for *n*-butane in TIP4P water using six windows in the stepwise

perturbation method. This methodology was demonstrated to yield results with high precision without the necessity of a surrogate potential, as well as correct averages at discrete values of ϕ rather than averages over distributions. The authors found 12% increase in the population of *gauche* conformers in going from gas phase to water. Recently H. E. Alper & B. J. Berne (private communication) further questioned the convergence issues for *n*-butane in Ne and water; they noted that long-lived fluctuations in the hydrogen bond network of water may affect conformation equilibria.

A related system also treated extensively via molecular simulation is the *n*-butane derivative 1,2-dichloroethane. The calculated results (95, 99) and experimental data (218) both indicated an increase of ~25–60% in *gauche* populations in going from gas phase to neat liquid; the increase was definitively linked to electrostatic effects in the potentials. Jorgensen's group has studied a number of other organic molecules exhibiting *trans/gauche* equilibria (reviewed in 158).

ALANINE DIPEPTIDE The Ala dipeptide, *N*-acetylalanyl-*N*-methyl amide, has been the focus of diverse theoretical studies including recent molecular simulations and free energy calculations. The sterically allowed regions of conformation in (ψ, ϕ) space for the Ala dipeptide are prototypical of the polypeptide backbone in general; thus a knowledge of the conformational stability of this molecule is highly relevant to the structural biochemistry of proteins. The (ψ, ϕ) map of intrinsic energy for the Ala dipeptide shows a large allowed region in the upper left-hand corner in International Union of Pure and Applied Chemistry (IUPAC) convention ($\psi = 200$–$360°$, $\phi = 180$–$60°$). This region contains the equatorial C_7 ($80°, -80°$), equatorial C_5 ($150°, -150°$), and polyproline P_{II} ($149°, -78°$) conformational forms. The conformation found in the α-helix, α_R, is in a well-separated energy minimum at ($-47°, -57°$). The C_7 conformation is well established as the global minimum on the potential energy surface. The C_5 and C_7 structures have been identified by vibrational spectroscopy in CCl_4 solvent, and the crystal structure has revealed two polyproline forms, P_I and P_{II}. Madison & Koppel (131) studied the Ala dipeptide in CCl_4 and H_2O using CD spectra and NMR spin coupling analysis. They confirmed earlier proposals of the presence of C_7 in nonpolar solvents, but also found evidence for contributions from both α_R and P_{II} conformations in water.

The intrinsic energy surface of the Ala dipeptide has been explored extensively based on diverse quantum-mechanical energy calculations and force-field calculations (reviewed in 153); the C_7 conformation, stabilized by a seven-membered intramolecular ring structure closed by a hydrogen bond, is generally indicated to be the lowest energy conformation.

Pettitt & Karplus (169) used normal-mode analysis to explore the influ-

ence of electrostatic interactions on the free energy difference between various conformational minima on the intrinsic energy surface of the Ala dipeptide. Brady & Karplus (33) and Karplus and coworkers (110) calculated the configurational entropy differences between the minima, both in vacuo and with hydration, using quasiharmonic methods. In both studies the entropy contributed 0.5–1.0 kcal mol^{-1} to the stability of C_5 relative to C_7. Ravishanker et al (188) obtained similar results in normal-mode calculations based on quasiharmonic MC calculations.

The Ala peptide has recently served as a model for two independent investigations of the conformational dynamics on the intrinsic energy surface under the influence of constraint and bias. D. S. Tobias & C. L. Brooks III (private communication) investigated the effects of dihedral angle constraints such as those used to define a λ coordinate in a free energy integration. They determined the role of thermal bond and angle fluctuations in the relative vibrational free energies of several low-energy conformations. The results indicated that the contribution of these fluctuations to the vibrational free energy can be significant and is sensitive to the applied constraints. Care is thus advised when constrained models are used to calculate free energy surfaces.

Mezei et al (153) have noted the principal difficulties in applying harmonic umbrella sampling potentials in the calculation of a multidimensional free energy surface: Splicing the separate regions of the multidimensional surface is difficult, and errors propagate in a progressive manner as more windows are spliced in. Bell & Harvey (13) recently proposed a multidimensional sampling method using a single restraint potential consisting of a Fourier series representation of the approximate free energy surface. The free energy can then in principle be determined in a single restrained simulation, which avoids the splicing of data from several independent runs. The Ala-dipeptide was used as a test case for the method in two dimensions. The results were generally encouraging and were in close agreement with previous independent determinations. The determination of the biasing potential was based on a high-temperature MD simulation.

Mezei et al (153) explored the relative free energy of hydration for the C_7, α_R, and P_{II} conformations of the Ala dipeptide in MC simulations. The free energy determinations were carried out using the probability ratio method with correlated conformational coordinates defining tie lines $C_7 \rightarrow \alpha_R$ and $C_7 \rightarrow P_{II}$. $C_7 \rightarrow \alpha_R$ required five windows to obtain suitably overlapping distributions for umbrella sampling, and $C_7 \rightarrow P_{II}$ required three. The simulation system comprised the Ala dipeptide and 201 water molecules at 300 K. The free energy calculations indicate that α_R and P_{II} are preferentially stabilized by hydration, as the solvent water successfully

competes with the intramolecular hydrogen bonding in C_7. The calculated $\Delta\Delta A^{hyd}$, combined with estimates of the intramolecuar energetics, indicates that the coexistence of C_7, α_R, and P_{II} (i.e. conformational flexibility) in water is reasonable.

The extended RISM form of integral equation theory, which requires considerably less computational power than a molecular simulation, has been applied to evaluate the effect of hydration of the complete (ψ, ϕ) map of the Ala dipeptide (170–172). The RISM calculations predict that the solvent generally flattens out the energy surface but that C_7 (equatorial and axial forms) is still lowest in free energy. The P_{II} conformation appears as a well-defined local minimum, as do C_5 and other structures along the edge ($\psi = 180°$ or $-180°$, $\phi = 180°$ or $-180°$) of the (ψ, ϕ) map. The α_R region is stabilized but does not appear as a well-defined energy minimum; this result is sensitive to subtle details in the intramolecular surface as well. The RISM calculations and the MC simulation described above thus show comparable effects of solvent stabilization for α_R and P_{II} relative to C_7.

Preliminary results are available for recent MD simulation studies by Anderson & Hermans (4, 80) in the free energy of the Ala dipeptide in water. A system of one solute and 64 water molecules was treated under periodic boundary conditions with a spherical cutoff of 6 Å. Individual simulations of 50 psec for β, 44 psec for α_R, 50 psec for C_7 (axial), and 89 psec for α_R were carried out. The relative free energies between specific conformations were worked out on correlated conformational coordinates and used to scale the independent probability distributions into a composite form. A β form ($-110°$, $120°$), which was in the vicinity of C_5, was overwhelmingly favored in this calculation.

Thus the simulation studies on the Ala dipeptide have collectively produced somewhat disparate results. In spite of its importance as a structural prototype, this system may not be the best test case for simulations, since the indications are that the Ala dipeptide molecule, like many small peptides, is conformationally flexible in solution and thus may not have a well-defined 2° structure. This problem has plagued NMR analysis of the structure of small polypeptides in water for some time. Recent vibrational spectroscopic studies (124) support conformational flexibility, but also indicate that the corresponding zwitterion may show structure and may be a more suitable candidate for simulation studies.

DIMETHYLPHOSPHATE ANION Phosphodiester torsion angles, (α, ζ) in IUPAC notation, make up an important set of conformational coordinates for the nucleotide backbone of nucleic acids. Three distinct conformations, (g, g), (g, t), and the fully extended (t, t) form, are of interest. Berman & Shieh (26) proposed a classification of nucleotide crystal structures in which the

phosphodiester torsion angles are the principle distinguishing features. While the (g, g) and (g, t) structures are well represented, (t, t) is not commonly found. Jayaram et al (92, 93), following early studies by Frischleder et al (67a), have performed a series of free energy simulations on dimethylphosphate anion in water using quasiharmonic MC simulation for the vibrational free energy (93) and umbrella potentials and the probability ratio method for the hydration component (92). The methodology was generally similar to that described above by Mezei et al (153) for the Ala dipeptide. Both the calculated vibrational and hydration free energies exhibited a preference for the (g, g) form, with (g, t) indicated to be thermally accessible. The (t, t) form was destabilized considerably by hydration (93), which may explain its absence in crystal hydrates. Jayaram et al (93) also compared hydration shell and continuum descriptions of free energy. They observed that the trends in conformational stability were internally consistent.

N-METHYLACETAMIDE The planarity of the peptide bond and the ubiquity of the *trans* form are tenets of structural biochemistry; virtually no *cis* peptide bonds occur in the crystal structures for proteins except in anomalous proline residues. The nature of the preference for the *trans* isomer is thus of considerable fundamental interest, as is the partitioning of the stability between intrinsic and solvent effects. A small molecule prototype for this problem is *N*-methylacetamide. Jorgensen & Gao (106) recently determined the *cis*/*trans* free energy difference for *N*-methylacetamide in water based on (T,P,N) MC simulations and the perturbation method applied to a system of solute and 216 TIP4P water molecules. Jorgensen's OPLS parameter set for the solute led to a $\Delta\Delta G^{hyd}$ of -2.2 ± 0.03, favoring *trans*. Slightly different system configurations and the probability ratio method gave -1.7 kcal mol^{-1} (S. W. Harrison, M. Mezei & D. L. Beveridge, unpublished calculations). Both of these results differ significantly from the experimental estimate, which was near zero (59).

Jorgensen & Gao (106) then explored the charge distribution in *cis* and *trans* *N*-methylacetamide via quantum mechanics. They found that the OPLS functions erroneously favor the N–H$\cdots$H$_2$O hydrogen bond by $\sim$1.3 kcal mol^{-1}. They formulated a revised charge distribution that was consistent with the quantum chemistry. The redetermined $\Delta\Delta G^{hyd}$ was -0.1 ± 0.3 kcal mol^{-1}. These findings emphasize the sensitivity of conformational free energy determinations via molecular simulation to the fine details of intermolecular potentials; here in the absence of the experimental data, the initial OPLS result would not have been questioned. This disclosure is possibly the tip of a sizable iceberg: None of the standard force fields are commonly adjusted for charge redistribution upon con-

formational changes; if the change in net atomic charges with conformation makes a significant difference in conformational equilibria, many simulation studies are compromised.

Chemical Reactions

Only a few chemical reaction systems have been studied in detail via molecular simulation. The point of focus with respect to free energy is the ΔG for a reaction (or $\Delta\Delta G$ for a series) and also $\Delta G^{\ddagger}$, the free energy of activation. The free energy of a reaction is related to the equilibrium constant via

$$K_{eq} = \exp(-\Delta G/RT), \qquad 51.$$

which is usually readily observable in an experiment. The free energy of activation is related to the rate constant of an elementary reaction step via transition state theory (73),

$$k = \kappa(k_B T/h)\exp(-\Delta G^{\ddagger}/RT), \qquad 52.$$

where $k_B T/h$ is the classical frequency of passage through the transition state of a reaction and κ is the transmission coefficient, related to the fraction of trajectories that successfully convert reactants into products. The magnitude of κ reflects the deviation of the rate behavior of the system from classical transition state theory, $\kappa = 1$. The quantity $\exp(-\Delta G^{\ddagger}/RT)$ is the Boltzmann probability of finding the system at the transition state relative to the reactants. Classical reaction dynamics (161) based on earlier gas-phase (5, 116) and solution theory (74, 160) gives the transition state theory estimate of the rate constant as

$$k = \kappa/2\langle|\dot{\zeta}|\rangle_{\zeta=\zeta^{\ddagger}}\rho(\zeta^{\ddagger})\Big/\int\rho(\zeta)\,d\zeta, \qquad 53.$$

where ζ is the reaction coordinate that describes the progress of the reaction from initial to final state through a transition state $\zeta^{\ddagger}$, $\langle\dot{\zeta}\rangle_{\zeta=\zeta^{\ddagger}}$ is the average of the crossing velocity $d\zeta/dt$ evaluated at $\zeta^{\ddagger}$, and $\rho(\zeta)d\zeta$ is the relative frequency of occurrence of configurations with ζ between ζ and $\zeta+\delta\zeta$; the integral is over the range of ζ corresponding to the initial-state valley of the system. The distribution $\rho(\zeta)$ can be determined from molecular simulation by umbrella sampling procedures in a manner analogous to that described previously, and defines a potential of mean force for the reaction

$$W(\xi) = -RT\ln\rho(\zeta) + \text{constant}. \qquad 54.$$

Analogously to the conformational equilibrium case, the $W(\zeta)$ for a reac-

tion may be decomposed into an intrinsic (gas-phase) component $V(\zeta)$ and a solvent-induced contribution $w(\zeta)$:

$$W(\zeta) = V(\zeta) + w(\zeta). \qquad 55.$$

Theoretical study of a reaction thus involves developing or choosing intermolecular potential functions defining the intrinsic reaction surface $V(\zeta)$; determining $w(\zeta)$, $\zeta^{\ddagger}$, and $\Delta G^{\ddagger}$ via molecular simulation; and obtaining $\langle|\dot{\zeta}|\rangle_{\zeta=\zeta^{\ddagger}}$ and κ from trajectory calculations. To obtain κ, the system is configured in the geometry of the activated complex, and unconstrained trajectories are integrated forward and backward in time for $\sim$2 psec in each direction. The quantity κ differs from unity as a consequence of barrier recrossings (18), i.e.

$$\kappa = \frac{\langle|\dot{\zeta}|[1+(-1)^{N+1}]/2N\rangle}{\langle|\dot{\zeta}|\rangle}, \qquad 56.$$

where N is the number of times each trajectory crosses and recrosses the transition-state dividing surface.

The focus of much of the work to date has been on understanding the influence of solvation by calculating energy and free energy profiles along assumed reaction pathways from gas phase and in solution. The molecular system that has produced the most informative studies to date is the prototypical organic S_N2 reaction $Cl^- + CH_3Cl$. Here the reaction coordinate of particular interest is the linear backside attack, and the only energy minima are an ion dipole complex and a symmetric transition state, $[Cl\cdots CH_3\cdots Cl]^-$. The gas-phase reaction potential surface exhibits a double-well form, whereas the free energy profile in solution is generally believed to be unimodal. The role of solvation in S_N2 reactions is a major topic in physical organic chemistry. In anticipation of the future application of molecular simulation to biochemical reaction mechanisms, it is relevant to develop a perspective on a well-defined chemical system before proceeding to more complex cases.

Chandrasekar et al (43, 44) described in detail a determination of the S_N2 reaction profile in gas phase, in aqueous solution, and in dimethylformamide. They gave a detailed historical background on the problem as well. The gas-phase potential curve for the backside displacement was obtained from ab initio quantum-mechanical calculations with the 6-31G(d) basis set. The typical double-well form was well reproduced, with a computed interaction energy of 10.3 kcal mol^{-1} and an intrinsic barrier of 13.9 kcal mol^{-1}. The solvent-averaged contribution $w(R)$ was obtained via umbrella sampling in an MC simulation involving six overlapping windows. The sampling was extensive, 1–2 million con-

figurations of equilibration followed by 2.6 million configurations of production at each window. The hydration was found to convert the double-well potential curve of the gas phase into an essentially unimodal surface, with the ion-dipole attraction energy offset by the loss in nucleophile solvent interaction. Charge delocalization as the reactants approach causes weakening of the solute-water hydrogen bonds rather than a reduction in number. The calculated $\Delta G^{\ddagger}$ in water was 26.3 ± 0.5 kcal mol^{-1}, in close accord with the experimental estimate of 26.6 kcal mol^{-1}. Results for the less polar solvent, dimethylformamide, were intermediate between those for the gas phase and water and indicated that the reaction mechanism, which is concerted in water, would involve an intermediate in dimethylformamide.

Further theoretical studies of this reaction system were based on MD trajectory calculations (24, 70; reviewed in 248). The energy flow into different modes of the reagents was monitored as the reaction approached the transition state. The calculations showed that at the barrier top recrossings were frequent, and a transmission coefficient of ~ 0.5 was indicated. The local configuration of solvent was found to influence the outcome of trajectories of the reacting atoms considerably. These results indicate a substantive role for solvent fluctuations in simple reaction rates in solution.

Warshel's group (88) modeled the reactant centers in this system via a simplified two-state quantum-mechanical treatment based on valence bond theory. Other degrees of freedom were treated classically, and many-body solvent effects were introduced via an atomic polarizability model. The perturbation method was used to calculate the reaction profile, and the dynamics was treated by propagating trajectories downhill from the transition state. These simulations also suggest that solvent fluctuations have a significant role in the reaction process.

Bash et al (9) developed a hybrid quantum mechanics/molecular mechanics treatment of Cl^- and CH_3Cl based on semiempirical AM1 and MINDO quantum mechanics. Their results were comparable to those of Chandrasekar et al (43, 44) and open up the possibility of general quantum-mechanical treatment of reaction centers in a molecular simulation format (see also 207). Similar hybrid methods had been tried earlier (114, 236), but recent developments in molecular simulation and computational power give the hybridization of molecular quantum mechanics and molecular simulation considerable future potential.

Other reaction systems treated by molecular simulation to date include the addition reaction of OH^- to H_2CO in water; Madura & Jorgensen (132) used methodology analogous to that described above for the S_N2 case. Here once again the intermediates on the gas-phase surface were washed out completely by the water, which indicated that the activation

barrier for the addition reaction was entirely solvent induced. The authors pointed out informative discrepancies with the experimental data on this system, however, which were linked to inadequacies in the description of ion-water polarization effects with two-body potentials.

Cieplak et al (46) recently reported molecular simulation and free energy determination using the perturbation method for the tautomeric equilibrium of 2-oxopyridine, 2-oxopyrimidine, and cytosine in water. The gas-phase component of the problem was treated with ab initio quantum-mechanical calculations. The calculations were found to be sensitive to the geometry assumed for the tautomer, but qualitative agreement with experiment was found for all comparable cases.

Electron transfer and proton transfer reactions, of considerable relevance to biological redox systems and acid dissociation processes, have been the subject of several simulation studies. Warshel (235) has used semiclassical trajectory calculations on a hybrid (quantum mechanics/molecular mechanics) surface based on a two-state valence bond method. Hwang & Warshel (90) recently used the coupling parameter approach with the perturbation method to simulate electron transfer between two benzenelike molecules in polar solvents with variable redox potentials. The solvent contribution to the activation barrier was found to follow Marcus's theory (134), and experimental deviations from the theory were suggested to be due to solute vibronic channels.

Jorgensen et al (101) recently applied thermodynamic cycle integration to the a priori determination of the pK_a of ethane and acetonitrile relative to methane thiol in water. Quantum-mechanical calculation of gas-phase energetics was followed by molecular simulation for the condensed-phase part of the problem. Experimental estimates of the gas-phase quantities could also be used. The predicted pK_a for acetonitrile, 28.6 when quantum results were used and 23.7 with experimental gas phase values, agreed well with the experimental value of 25 ± 1. For ethane, $\Delta\Delta G^{hyd}$ was also well reproduced; ethyl anion was predicted to be better hydrated by 6.4 kcal mol^{-1} than methane thiolate. The pK_a of ethane was predicted to be about 50. The a priori computation of this elusive quantity illustrates nicely the power and versatility of modern simulation methodologies applied to chemical systems (101).

APPLICATIONS TO BIOMOLECULAR SYSTEMS

Bovine Pancreatic Trypsin Inhibitor

The bovine pancreatic trypsin inhibitor (BPTI) molecule is a small (58-residue), extremely stable protein which has frequently served as a prototype system for the development of diverse biophysical methods, both

theoretical and experimental. Studies in the genre of free energy simulations have been reported on side-chain rotational isomerism and on the ionization of acidic groups in BPTI.

The rotational isomerism and dynamics of tyrosine ring rotation in BPTI have been studied by NMR (55, 232); the rate constant was determined to be 1–10 sec^{-1}. A sequence of theoretical studies by McCammon, Karplus, and coworkers (69, 112, 143, 144, 161), focusing particularly on Tyr35, has explored this problem as a general inquiry into the nature of fast motions in the densely packed protein interior. The methods employed were those of dynamical transition state theory outlined earlier; the researchers determined a potential of mean torsion and the free energy of activation for the ring rotation in the protein, and initiated activated dynamical trajectories for the momentum flux and transmission coefficient of the reaction from the transition state. Initial calculations (69) revealed that the observed barriers were due to repulsive nonbonded interactions between the rings and neighboring atoms, and that matrix relaxation has an essential role in permitting the rotations to occur. The activated trajectories showed Tyr35 to be driven over the rotational potential barrier by nearly impulsive collisions with the protein surroundings and appear to be initiated by local packing defects (161). Furthermore, detailed analysis (145) showed how the process is gated by displacement of a section of backbone that covers one face of the ring. The first-order rate constant calculations (161), based on an in vacuo model with charges of the protein atoms set to zero to reflect the dielectric screening properties of water, gave $k \cong 8.9 \times 10^4$ sec^{-1}, considerably too fast. Subsequent studies (71) incorporated a droplet of solvent water explicitly around Tyr35 and obtained improved agreement, 540 sec^{-1}. The transmission coefficient κ was 0.25, indicating considerable deviation from canonical transition state theory for this process. The remaining discrepancy between calculated and observed rates was probably a consequence of the united atom model for Tyr hydrogen atoms. Most recent extensions of this work have looked at solvent viscosity effects (72); a 50% reduction in rate constant was observed, due to relaxation of ring solvent interactions and conformation of the protein matrix.

The free energies of ionization of the acidic Asp3 and Glu7 are quite similar in BPTI in spite of a difference in their local environments. Following earlier PDLD calculations (193), Warshel et al (241) recently calculated the relative free energy of acid dissociation for these groups in BPTI and in aqueous solution and also for aspartic acid in water. The perturbation method in an MD simulation was used, with an equilibration time of 2 psec and production time of 6 psec for each of 12 windows in the charging process. The solvent was included by placing constraining

droplets of 80 water molecules around the reaction site in each case. The $\Delta\Delta G$ results based on an application of Equation 42 at 300 K were all around -70 kcal mol^{-1}, with an error of ~ 1 kcal mol^{-1}. This value is close to the experimental estimate (see 193) and also indicates the sizable contribution of solvation to the process.

Myoglobin

The entrance and exit of dioxygen to and from the heme pocket region of myoglobin has been studied by laser spectroscopy. Considerable rate data have been obtained by following the course of rebinding of ligand after the onset of laser-induced photodissociation. The escape process has a rate constant of 5×10^6 sec^{-1} for CO and 13×10^6 sec^{-1} for dioxygen (78). Egress from the heme pocket probably involves a path roughly parallel to the heme group and passing between the side chains ValE11 and HisE7. MD trajectories for this pathway have been described (39). A gating effect involving the two side chains is indicated, with an energy barrier of 10–15 kcal mol^{-1}.

Case & McCammon (40) subsequently described a potential of mean force and activated dynamics trajectories for this problem and calculated rate constants from dynamical transition state theory. The X-ray structure was used as a point of departure, with only residues <9 Å from the heme center allowed to move. The reaction coordinate was defined as the perpendicular distance from the ligand to the plane bounded by Cβ of ValE11, Cγ of HisE11, and the nitrogen of pyrrole 1 of the heme group. To obtain $W(\zeta)$, equilibration was followed by six simulations of 18 psec each. When the ligand was allowed to equilibrate dynamically in the bottleneck region, the potential barrier disappeared. However, the $W(\zeta)$ showed a barrier of 6.5 kcal mol^{-1}, not due to potential energy but due to loss of translational and rotational entropy in the activated state. The calculated transmission coefficient was near unity, indicating that classical transition-state theory applies. The calculated rate constant, 4.1×10^7 sec^{-1}, compared well with the observed constant, 1.3×10^7 sec^{-1}.

Kottalam & Case (121) recently completed a more extensive MD study of this problem. They treated the protein atoms >9 Å from the active site by Langevin dynamics and considered six different temperatures from 180 to 330 K. The calculated transmission coefficient and rate constants were similar to those described above, but the authors presented additional valuable computational details. They confirmed the entropic nature of the free energy barrier to egress at room temperature and described the temperature dependence. Overall, the dynamical motion of the dioxygen across the transition state in myoglobin was found to be ballistic, as opposed to the diffusive behavior observed for tyrosine ring rotation in

BPTI. The authors emphasized that they have studied only one of the many possible pathways for escape (see also 221).

Trypsin

Several recent free energy simulations have explored the binding of inhibitors and the catalytic mechanism in native and genetically modified forms of the enzyme trypsin, which catalyzes amide hydrolysis. The relative binding affinities of benzamidine and a *p*-fluoro derivative for trypsin have been reported by Mares-Guia et al (134a) as $\Delta\Delta A = 2.1 \pm 1.3$ kJ mol^{-1}. The $\Delta\Delta A$ for a mutant enzyme (Gly216 → Ala216) has been estimated from the Michaelis constants for arginine substrates to be ~ 8.4 kJ mol^{-1} (48). In theoretical calculations, Wong & McCammon (250) took up this problem as a canonical application of the Tembe-McCammon equation to macromolecular binding, with free energies determined by the perturbation method. The calculations were based on (T,V,N) MD simulations, with initial coordinates determined from a 1.7-Å X-ray structure of the inhibited enzyme; the native form was modeled by replacing inhibitor with three water molecules. The entire enzyme-substrate complex together with ~ 5000 water molecules was treated with the benzamidines protonated. Run times for the native and mutant forms were 28.8 and 22.4 psec, respectively, with a cutoff of 8 Å for Lennard-Jones terms and 20 Å for Coulombic interactions. Experiment had shown that benzamidine affinity for trypsin was somewhat larger than that of the *p*-fluoro derivative and affinity for the native was far greater than for the mutant form (134a). These trends were supported by the calculations, which gave $\Delta\Delta A = 3.8 \pm 2.2$ for the former case and $5.6 \pm > 2.2$ kJ mol^{-1} for the latter. The calculations aid here in the interpretation of the experimental results, indicating that the binding preference for benzamidine over *p*-fluoro benzamidine is due to unfavorable desolvation of the derivative, a more dipolar species. The differential free energy for the native versus mutant enzymes was ascribed to steric crowding due to the added methyl group in the mutant, although the breakdown into intrinsic and environmental effects was not reported. Thermodynamic cycle integration was thus indicated to be a feasible route for analysis and prediction of differential affinity for closely related biomolecular systems.

Warshel and coworkers (239, 240) recently reported theoretical studies of the catalytic free energy in native and mutant species of trypsin based on experimental data on genetic modifications of Craik et al (48), specifically the double mutation of Gly to Ala at positions 216 and 226 in rat trypsin. This mutation leads to a relatively small change in the binding constant, less than a factor of 20, and a very large change in the catalytic rate constant, a factor of 2000. The calculations were based on the crystal

structure of bovine trypsin and a lysyl substrate modeled by an Ala-Lys-Gly tripeptide. MD calculations were carried out on atoms within a 16-Å radius of the active site, with the remaining atoms fixed in positions indicated by the crystal structure. A droplet of 100 water molecules in the active site region was also included. A plausible three-state reaction coordinate for the acylation of the substrate was assumed. To describe changes in electronic energy of the active site, a quantum-mechanical treatment of the reaction center based on Warshel's extended valence bond method (235) was interfaced to the otherwise empirical force field for the system. The perturbation method was used to evaluate the relative free energies of the valence bond structures and to estimate a free energy profile for the reaction. The results indicated a $\Delta\Delta G^{\ddagger}$ between the native and mutant form of 6 ± 2 kcal mol^{-1}, leading to a reduction in catalysis rate of a factor of 5×10^{-5} versus 5×10^{-4} observed. Analysis of the results indicated that the calculated effect resided primarily in the solvent contribution to the free energy, with the higher catalytic rate of the native enzyme attributed to better solvent stabilization of the intermediate ionic resonance structures. Warshel (238) proposed that the enzyme active site may act as a supersolvent for the transition state. The calculations from Warshel's group disfavor the charge relay mechanism, since the proton transfer effectively shows a higher calculated activation energy.

Subtilisin

The enzyme subtilisin catalyzes the hydrolysis of amides via a presumed transition state in which the carbonyl oxygen of the reacting moiety is stabilized by hydrogen bonding interactions with two enzyme NH groups in the so-called oxyanion hole (138, 203, 214). In subtilisin, one of the two NH groups comes from the side chain of Asn155, which has been a target for site-specific mutagenesis (37, 247). Mutating Asn155 to Thr, Leu, and Ala is known to have a large effect on the catalytic role, but somewhat less on the binding constant.

This system has recently been the subject of free energy simulations using the perturbation method to obtain $\Delta\Delta G$ and $\Delta\Delta G^{\ddagger}$ (89, 192). Hwang & Warshel (89) simulated on an assumed reaction coordinate and obtained $\Delta\Delta G$ via Equation 44, whereas Rao et al (182) used the mutational coordinate approach, Equation 45. Both studies involved MD calculations on only amino acid residues within a spherical radius of the active site. Hwang & Warshel (89) used their parametrized valence bond treatment in the calculation of energy for the complex; Rao et al (182) used an AMBER force field calculation throughout. In the former study the substrate was a Tyr-Gly dipeptide, whereas the latter authors assumed the tripeptide Ala-Ala-Phe.

Both groups successfully accounted for the magnitudes of the differential free energy changes in the binding and catalytic processes. Hwang & Warshel (89) reported the correct order for mutations of Asn155 to Thr, Leu, and Ala. However, the results were sensitive to the position of advancement along the reaction coordinate, and the position of the transition state was critical in the comparison. Rao et al (182) predicted the $\Delta\Delta G$s of binding and catalysis (0.1 ± 0.8 and 3.4 ± 1.1 kcal mol^{-1}, respectively) for the Ala mutant in advance of experiment (0.41 and 3.67 kcal mol^{-1}, respectively).

Warshel (233) has emphasized that $\Delta\Delta G^{\ddagger}$ can be closely estimated by calculating the solvation energy of the charges of the relevant resonance structures, and he cited this as further evidence for viewing the enzyme as a supersolvent for substrate. According to this theory, catalytic energy is stored not just in the stabilization of the transition-state charge distribution, but also in the prealignment of protein dipoles in the proper direction. Thus the catalytic free energy is stored in the (unfavorable) folding energy needed for prepolarization. This concept is related to the idea of solvent reorganization energy from Marcus's theory (134).

Rao et al (182) analyzed the interactions in the oxyanion hole in some detail in comparing Asn155 with Ala; the former stabilized the complex with a hydrogen bond interaction not present in the mutant. The results did not indicate that direct enzyme–transition state interaction energies are stronger than water–transition state interactions. These authors stated that the most important difference between an enzyme-catalyzed reaction and the corresponding solution reaction is not in electrostatics per se, but in the free energy price paid in synthesis to align and desolvate the catalytic groups.

Thermolysin

Phosphonamidate peptide analogs with $-PO_2^-$–OH replacing the peptide link in a Gly-Leu sequence are potent inhibitors of the zinc endopeptidase thermolysin (7). Their free energies of interaction with the enzyme have been correlated with the (hypothetical) binding energies of the transition states for hydrolysis of the corresponding amide substrates as reflected in turnover ratios (K_m/K_{cat}), which are thought to mimic transition-state configurations of the enzyme-substrate complexes. Bartlett & Marlowe (8) recently studied a series of phosphonate esters, which differ from the amidates in replacement of the NH between the phosphorus and the leucine moiety with an oxygen. The binding energies of five analogs were reduced uniformly by a factor of 840, or ~ 4 kcal mol^{-1}. The interpretation of this quantity via structure determination (224) and free energy calculations (10) has focused on the hypothesis that the oxygen for

N–H change in free energy is due to the loss of a specific phosphonamidate N–H hydrogen bond interaction to an acceptor group on the enzyme.

X-ray diffraction studies (224) have provided support for the above hypothesis as well as for the absence of other specific differences in the way amidates and esters bind. In parallel with binding-constant and crystallographic studies, Bash et al (10) used the perturbation method to calculate $\Delta\Delta G$ values for mutating the –NH– into –O– in the protein. Two calculations, the $\Delta\Delta G$ of hydration for the two inhibitors and $\Delta\Delta G$ for the complexes, were involved. The aqueous hydration of the inhibitor was treated with 782 water molecules, and the complex involved all atoms within 15 Å of the replacement site with a restrained droplet of 168 water molecules. The resulting $\Delta\Delta G$ for binding was calculated to be 4.21 ± 0.54 kcal mol^{-1}; the observed value was 4.1 kcal mol^{-1}. Analysis revealed that the intrinsic difference in the binding energy due to replacement of –NH– by –O– was actually 7.5 kcal mol^{-1}, which includes some effective repulsion between the ester oxygen and the acceptor C=O group of Ala113 on the enzymes. The calculated relative solvation free energy was -3.58 kcal mol^{-1}, favoring the ester. The calculation predicts that the observed free energy is not simply due to the loss of a single quantum of hydrogen bond energy, but is a result of intrinsic attractive and repulsive forces as well as differential solvation.

Dihydrofolate Reductase

Dihydrofolate reductase (DHFR) is an essential enzyme in cell growth. Inhibitors to bacterial DHFR are antibiotics; those to the human enzyme are cytostatic agents. The interactions of the inhibitor trimethoprim (TMP) and a number of congeners thereof with the cofactor-protein complex of NADPH-dihydrofolate reductase are currently the subject of free energy simulations in several laboratories (23, 36, 54, 205, 229). The X-ray structure of the NADPH-DHFR-TMP complex is known (139, 140) and serves as a point of departure for the modeling and simulation studies. The calculations are carried out in the general context of the Tembe-McCammon equation, which requires calculation of relative solvation free energies of the drugs and the difference in free energy for the drugs bound to the presumably common receptor site. van Gunsteren & Berendsen (229) found that the increased possibility for conformational flexibility of the drugs presents quasi-ergodic problems. The integration of TMP to its ethyl analog TEP yielded -55.1 kJ mol^{-1}, whereas the reverse integration of TEP to TMP gave 65.6 kJ mol^{-1}. However, TMP in the latter case was found to differ in a C–O torsion angle from the initial configuration, at a cost of 9.3 kJ mol^{-1}. Still, the difference in forward and backward inte-

grations was considered unacceptable. However, the $\Delta\Delta G$ of the average 4 kJ mol^{-1}, compares with the observed value of 7.1 kJ mol^{-1}, favoring TMP.

Brooks (36) considered the inhibitors TMP, TEP, and a paraethyl analog (PET) in protonated as well as neutral forms in a similar free energy study. The free energy differences covered a relatively small range, -1.3 to 1.9 kcal mol^{-1}. However, Brooks's analysis revealed that large energy/entropy compensations were involved that could account for the difference between the calculated $\Delta\Delta A$ of 1.8 kcal mol^{-1} for $TMPH^+$ versus $TETH^+$ and the experimental value of -1.7 kcal mol^{-1}. Experiment (139, 140) indicated the protonated form of the substrates to be the relevant species for binding. The unexpectedly large binding affinity for PET was found to be the result of an additional hydrogen bond. The calculated results indicated that the three drugs bind in somewhat different positions in the binding pocket, which prompts an additional warning about quasi-ergodic problems in drug-binding problems.

Mackay and coworkers (129) have based a current study of ligand binding to DHFR on energy minimization strategies. They have reported a change in ligand entropy on binding or on restriction of ligand motion by the protein. An all-atom model gives a substantially better account of the X-ray structure of the trimethoprim structure than a united atom model.

Singh (205) studied the importance of the salt bridge between Asp27 and the pteridine ring in an *E. coli* DHFR-methotrexate complex using the free energy perturbation method. The calculation suggested that the ion-pair contribution to the binding energy is insignificant, as the enzyme surroundings do not stabilize the salt bridge to the extent of the desolvation of the charged groups. The activation barrier for the proton exchange between the pteridine ring and Asp27 was calculated to be 20.1 kcal mol^{-1}, implying that this may be a channel for the proton exchange from the pteridine ring to the solvent. The $\Delta\Delta G$ of binding between Asn27 and Ser27 mutants was found to be 3.2 kcal mol^{-1} and due mainly to electrostatic interaction.

Cytochrome c

Cutler et al (51) have recently undertaken free energy determinations for site-specific mutations in cytochrome *c*. Arg38 is a residue of cytochrome *c* that hydrogen-bonds to the heme propionate at position 7. Previous work comparing the behavior of cytochrome *c* from several species (156) has suggested that Arg38 lowers the pK_a of this propionate group and thereby accounts for the relative pH independence of the cytochrome reduction potential from pH 5 to 8. The influence of Arg38 on the oxida-

tion-reduction equilibrium of yeast iso-1-cytochrome *c* has now been investigated by electrochemical, NMR, and theoretical analysis of six specifically mutated forms of this protein in which Arg has been replaced by Lys, His, Gln, Asn, Leu, or Ala. As the electron-withdrawing character of the residue at position 38 decreases, the reduction potential of the protein also decreases, with the largest decrease ($\sim$50 mV) observed for the Ala variant. However, the variation in the reduction potentials of the variants as a function of pH was identical to that observed for the wild-type protein. The effects of some of these mutations on the pK_a of His33 and His39 were determined by NMR spectroscopy and found to be minimal. Calculations of the electrostatic free energy for the Leu38 variant predict a decrease in the reduction potential of this mutant that is remarkably close to that observed experimentally. This work suggests that while Arg38 contributes to the relatively high reduction potential of cytochrome *c*, its influence on the pK_a of heme propionate 7 is probably minimal.

Photosynthetic Reaction Centers

The free energy of the primary charge separation process in bacterial photosynthesis was recently studied via molecular simulation by Warshel's laboratory (49). The MD perturbation method was employed, with the coupling parameter used to describe electron transfer as seen in an earlier study (156). The point of departure for the calculation was the crystal structure of reaction centers in *Rhodobacter sphaeroides*. The energies of the charge transfer states in which an electron moves from a special pair of bacteriochlorophylls (P) to a neighboring bacteriochlorophyll (B) or bacteriopheophytin (H) were obtained by combining experimental standard redox potentials with calculated solvation free energies for the electron carriers in solution and in the reaction center. Three different mechanisms were considered for the transfer of an electron from the excited dimer (P*) to H. In one mechanism, an electron first moves from P to B, generating P^+B^- as a transient intermediate state; the dynamics calculated for this pathway agreed well with the kinetics measured experimentally. The P^+H^- complex forms in 3–4 psec, with the concentration of the intermediate state remaining low throughout the reaction. A mechanism in which the electron moves from B to H in the initial step is also possible on the basis of the calculated free energies of the relevant charge transfer states. A mechanism involving a mixture of P*, P^+B^-, and P^+H^- by superexchange gives lower calculated rates but cannot be excluded. The effectiveness of all three mechanisms depends on the ability of the protein to stabilize the charge transfer states while the reorganization energies for the electron transfer reactions remain small.

Protein Conformations

In a series of calculations Hagler and coworkers (76) found that the relative free energies of plausible oligopeptide structures could be dominated by vibrational entropy effects. A formalism for vibrational free energy in the context of proteins has been given by Scheraga (197). The problems involved in determining the relative free energy of protein conformations and free energy pathways for protein folding have been explored in extensive theoretical studies (198). The theoretical description of protein folding has been difficult to obtain as a consequence of the multiple minima problem, which is a quasi-ergodic problem of colossal dimension. A general review of protein folding is outside the scope of this article (see however 67, 117), but free energy calculations on proteins via molecular simulations as described herein by no means overcome the multiple minimum problem; all free energy calculations are promulgated on the basis of an assumed structure in the presumably correct region of configurational space.

Scheraga and coworkers (reviewed in 198) have recently been exploring variants of free energy calculations such as adaptive importance sampling via MC simulation to pursue the protein conformation problem further. Most recently, Meirovitch et al (148) have proposed a Metropolis method that first generates a biased sampling of configuration space and then uses renormalization group theory (199) to develop an unbiased estimate and thus conformational entropy and free energies for protein folding.

Nucleic Acids

Free energy simulations for nucleic acid systems such as DNA and RNA structures have not yet been reported apart from quasiharmonic determinations of the vibrational entropy of B- and Z-DNA (34). Mills et al (155) treated free energy via grand canonical ensemble MC calculations in a study of the thermodynamic coefficients for a primitive model of DNA salt solutions, comparing results with those from a Poisson-Boltzmann calculation. The area of nucleic acids is expected to attract considerable emphasis, especially from the point of view of drug-DNA interactions, in the near future.

SUMMARY AND CONCLUSIONS

The general approaches to free energy determinations via molecular simulation are well formulated theoretically, work well methodologically, and are generally capable, at least in principle, of providing an accurate theoretical account of free energy in chemical and biomolecular systems. The general agreement between calculated and experimentally observed

quantities is quite good at the level of empirical and semiempirical parametrization of intermolecular potentials, where some experimental data are introduced into the parametrization to set the calculations for a class of systems off on the right foot. Ab initio calculation and the pure predictive capabilities of free energy calculations are less well established at present; studies broaching these areas have encountered, as expected, pronounced but informative difficulties. The serious problems here, especially for chemical systems, seem however to be primarily in the potential functions rather than in the free energy methodology. The important but computationally demanding study of free energy for cooperative potentials is yet to be taken up for chemical and biomolecular systems.

A cursory reading of the literature and this review might lead one to conclude prematurely that the agreement between calculated and observed free energies is inversely proportional to system size; the characterization studies on chemical systems point up discrepancies, problems, and challenges, while results on complex biomolecular problems seem invariably to be in close agreement with experiment with no particular problems. Considerable difficulties can be anticipated in biomolecular applications, due to potential functions and particularly the quasi-ergodic problem. There are often several possible ways that a substrate can sit in a binding site or a drug can bind in a receptor area. Some of these will result in similar calculated free energies at the current level of research. There is not yet documentation that simulations can reliably distinguish the alternative possibilities or, on the picosecond time scale, properly sample the possible conformations at a binding site. [Some provocative work has been done on this point with regard to a cold virus problem (M. Pettitt, unpublished studies).] Thus the success of free energy calculations applied to biomolecular systems still depends on looking at relative effects on a series of closely related substances where binding can be reliably modeled based on an X-ray structure. Nevertheless, the results to date are clearly encouraging. However, agreement of calculated and observed free energies does not yet unequivocally validate the calculation or prove the model; one can obtain good agreement between calculated and observed free energies for the wrong reasons, and the multiple minimum problem is still with us. The ultimate utility of molecular simulation in biochemistry and drug design will depend on how well, even with increased computational power, we can deal with the quasi-ergodic problem.

ACKNOWLEDGMENTS

This article was made possible by grants from the National Institutes of Health (GM-37909), the National Science Foundation (CHE-869117), and

the Office of Naval Research (N-0014-87-K-0312). Support from Merck, Sharp & Dohme Research Laboratories and the Bristol-Myers Corporation via a Connecticut Cooperative High Technology Research and Development Grant is also gratefully acknowledged. Participation by DLB in a series of CECAM workshops at the University of Paris at Orsay has contributed significantly to the perspectives set forth in this review. The facilities of Wesleyan University, the Green Mountain Research Center, and the Brattleboro, Vermont Public Library were also essential to this project.

Literature Cited

1. Abraham, R. J., Cavalli, L., Pachler, K. G. R. 1966. *Mol. Phys.* 11: 471
2. Allen, M. P., Tildesly, D. J. 1987. *Computer Simulation of Liquids*. Oxford, UK: Clarenden
3. Deleted in proof
4. Anderson, A. G., Hermans, J. 1988. *Proteins*. In press
5. Anderson, J. B. 1973. *J. Chem. Phys.* 58: 4684
6. Barker, J. A., Henderson, D. 1976. *Rev. Mod. Phys.* 48: 587
7. Bartlett, P. A., Marlowe, C. K. 1983. *Biochemistry* 22: 4618
8. Bartlett, P. A., Marlowe, C. K. 1987. *Science* 235: 569
9. Bash, P. A., Field, M. J., Karplus, M. 1987. *J. Am. Chem. Soc.* 109: 8092
10. Bash, P. A., Singh, U. C., Brown, F. K., Langridge, R., Kollman, P. A. 1987. *Science* 235: 574
11. Bash, P. A., Singh, U. C., Langridge, R., Kollman, P. A. 1987. *Science* 236: 504
12. Bell, C. D., Harvey, S. C. 1986. *J. Phys. Chem.* 90: 6595
13. Bell, C. D., Harvey, S. C. 1988. *J. Phys. Chem.* In press
14. Ben-Naim, A. 1974. *Water and Aqueous Solutions*. New York: Plenum
15. Ben-Naim, A. 1984. *J. Chem. Phys.* 81: 2016
16. Ben-Naim, A., Marcus, Y. 1984. *J. Chem. Phys.* 81: 2016
17. Bennet, C. H. 1976. *J. Comput. Phys.* 22: 245
18. Bennet, C. H. 1977. In *Algorithms for Chemical Computations*, ed. R. E. Christofferson. Washington, DC: Am. Chem. Soc.
19. Berendsen, H. J. C., Egberts, E. 1987. In *Structure, Dynamics and Function of Biomolecules*, ed. A. Ehrenberg et al, p. 275. Berlin: Springer
20. Berendsen, H. J. C., Grigera, J. R., Straatsma, T. P. 1987. *J. Phys. Chem.* 91: 6209
21. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. 1985. See Ref. 79, p. 43
22. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Hermans, J. 1981. In *Intermolecular Forces*, ed. B. Pullman, p. 331. Dordrecht, the Netherlands: Reidel
23. Berendsen, H. J. C., van Gunsteren, W. F., Egberts, E., de Vlcig, J. 1987. *ACS Symp. Ser.* 353: 106
24. Bergsma, J. P., Gertner, B. J., Wilson, K. R. 1987. *J. Chem. Phys.* 86: 1356
25. Berkowitz, M., Karim, O. A., McCammon, J. A., Rossky, P. J. 1984. *Chem. Phys. Lett.* 105: 577
26. Berman, H. M., Shieh, H.-S. 1981. In *Topics in Nucleic Acid Structure*, ed. S. Neidle, p. 17. London: Macmillan
27. Beveridge, D. L., DiCapua, F. M. 1989. See Ref. 230a. In press
28. Beveridge, D. L., Mezei, M. 1985. See Ref. 79, p. 53
29. Beveridge, D. L., Mezei, M., Mehrotra, P. K., Marchese, F. T., Ravishanker, G., et al. 1983. *Adv. Chem. Ser.* 204: 297
30. Beveridge, D. L., Schnuelle, G. W. 1975. *J. Phys. Chem.* 79: 2562
31. Bigot, B., Jorgensen, W. L. 1981. *J. Chem. Phys.* 75: 1944
32. Born, M. 1920. *Z. Phys.* 1: 45
33. Brady, J., Karplus, M. 1985. *J. Am. Chem. Soc.* 107: 6103
34. Brooks, B., Karplus, M. 1988. *Science*. In press
35. Brooks, C. L. III. 1986. *Phys. Chem.* 90: 6680
36. Brooks, C. L. III. 1988. *Int. J. Quantum Chem.* In press
37. Bryan, P., Pantoliano, M. W., Quill,

S. G., Hsiao, H.-Y., Poulos, T. 1986. *Proc. Natl. Acad. Sci. USA* 83: 3743
38. Burkert, U., Allinger, N. L. 1982. *Molecular Mechanics*. Washington, DC: Am. Chem. Soc.
39. Case, D. A., Karplus, M. 1979. *J. Mol. Biol.* 132: 343
40. Case, D. A., McCammon, J. A. 1986. *Ann. NY Acad. Sci.* 482: 222
41. Chandler, D. 1978. *Discuss. Faraday Soc.* 66: 184
42. Chandler, D., Anderson, H. C. 1972. *J. Chem. Phys.* 57: 1930
43. Chandrasekar, J., Smith, S. F., Jorgensen, W. L. 1984. *J. Am. Chem. Soc.* 106: 3049
44. Chandrasekar, J., Smith, S. F., Jorgensen, W. L. 1985. *J. Am. Chem. Soc.* 107: 154
45. Chiles, R. A., Rossky, P. J. 1984. *J. Am. Chem. Soc.* 106: 6867
46. Cieplak, P., Bash, P., Singh, U. C., Kollman, P. A. 1987. *J. Am. Chem. Soc.* 109: 6283
47. Cieplak, P., Kollman, P. A. 1988. *J. Am. Chem. Soc.* 110: 3734
48. Craik, C. S., Langman, C., Fletcher, T., Rocznick, S., Barr, P. J., et al. 1985. *Science* 228: 291
49. Creighton, S., Hwang, J.-K., Warshel, A., Parson, W. W., Norris, J. 1988. *Biochemistry* 27: 774
50. Cross, A. J. 1986. *Ann. NY Acad. Sci.* 482: 89
51. Cutler, R. L., Davies, A., Creighton, S., Warshel, A., Moore, G. R., Smith, M., Grenk Mauk, A. 1989. *Biochemistry*. In press
52. Dang, L. X., Pettitt, B. M. 1987. *J. Chem. Phys.* 86: 6560
52a. Dang, L. X., Pettitt, B. M. 1987. *J. Am. Chem. Soc.* 109: 5531
53. Dashevsky, V. G., Sarkisov, G. N. 1974. *Mol. Phys.* 27: 1271
54. Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., Hagler, A. J. 1988. *Proteins Struct. Funct. Genet.* 4: 31
55. Davis, D. G., Bax, A. 1985. *J. Magn. Reson.* 64: 533
56. de Donder, Th. 1927. *L'Affinite*. Paris: Gauthier-Villars
57. Deleted in proof
58. Deleted in proof
59. Drakenberg, T., Forsen, S. 1971. *J. Chem. Soc. Chem. Commun.* 1971: 1404
60. Dzidic, I., Kebarle, T. 1970. *J. Phys. Chem.* 74: 1466
61. Edholm, O., Berendsen, H. J. C. 1984. *Mol. Phys.* 51: 1011
62. Eisenberg, D., McGlachlan, A. D. 1986. *Nature* 319: 199
63. Fixman, M. 1974. *Proc. Natl. Acad. Sci. USA* 71: 3050
64. Fleischman, S. H., Brooks, C. L. 1987. *J. Chem. Phys.* 87: 3029
65. Frank, H. S., Evans, M. N. 1945. *J. Chem. Phys.* 13: 507
66. Franks, F. 1975. In *Water: A Comprehensive Treatise*, 4: 1. New York: Plenum
67. Frauenfelder, H., Parak, F., Young, R. D. 1988. *Annu. Rev. Biophys. Biophys. Chem.* 17: 451
67a. Frischleder, H., Gleichman, S., Kroll, R. 1977. *Chem. Phys. Lipids* 19: 144
68. Geiger, A., Rahman, A., Stillinger, F. H. 1979. *J. Chem. Phys.* 70: 263
69. Gelin, B. R., Karplus, M. 1975. *Proc. Natl. Acad. Sci. USA* 72: 2002
70. Gerner, B. J., Bergsma, J. P., Wilson, K. R., Lee, S. Y., Hynes, J. T. 1987. *J. Chem. Phys.* 86: 1377
71. Ghosh, I., McCammon, J. A. 1987. *Biophys. J.* 51: 637
72. Ghosh, I., McCammon, J. A. 1987. *J. Phys. Chem.* 91: 4878
73. Glasstone, S., Laidler, K., Eyring, H. 1941. *The Theory of Rate Processes*. New York: McGraw-Hill
74. Grote, R. F., Hynes, J. T. 1980. *J. Chem. Phys.* 73: 2715
75. Deleted in proof
76. Hagler, A. T., Stern, P. S., Sharon, R., Becker, J. M., Naider, F. J. 1979. *J. Am. Chem. Soc.* 101: 6842
77. Deleted in proof
78. Henry, E. A., Sommen, J. H., Hofrictor, J., Eaton, W. A. 1983. *J. Mol. Biol.* 166: 443
79. Hermans, J., ed. 1985. *Molecular Dynamics and Protein Structure*. Western Springs, Ill: Polycrystal
80. Hermans, J. 1986. *Ann. NY Acad. Sci.* 482: 51
81. Hermans, J., Pathiaseril, A., Anderson, A. 1988. *J. Am. Chem. Soc.* In press
82. Hirata, F., Pettitt, B. M., Rossky, P. J. 1982. *J. Chem. Phys.* 77: 509
83. Hirata, F., Redfern, P., Levy, R. M. 1988. *J. Chem. Phys.* In press
84. Hirata, F., Rossky, P. J. 1981. *Chem. Phys. Lett.* 83: 329
85. Ho, F. S., Quigley, G. W., Tilton, R. F., Rich, A. 1988. *J. Phys. Chem.* In press
86. Hodes, Z. I., Nemethy, G., Scheraga, H. A. 1979. *Biopolymers* 18: 1565
87. Hsu, C. S., Pratt, L. R., Chandler, D. 1978. *J. Chem. Phys.* 68: 4213
88. Hwang, J.-K., King, G., Creighton, S., Warshel, A. 1988. *J. Am. Chem. Soc.* In press
89. Hwang, J.-K., Warshel, A. 1987. *Biochemistry* 26: 2669

90. Hwang, J.-K., Warshel, A. 1987. *J. Am. Chem. Soc.* 109: 315
91. Jayaram, B., Fine, R., Sharp, K., Honig, B. 1988. *J. Chem. Phys.* In press
92. Jayaram, B., Mezei, M., Beveridge, D. L. 1988. *J. Am. Chem. Soc.* 110: 1691
93. Jayaram, B., Ravishanker, G., Beveridge, D. L. 1988. *J. Phys. Chem.* 92: 1032
94. Jayaram, B., Sharp, K., Honig, B. 1988. *Biopolymers.* In press
95. Jorgensen, W. L. 1981. *J. Am. Chem. Soc.* 103: 677
96. Jorgensen, W. L. 1981. *J. Am. Chem. Soc.* 103: 4721
97. Jorgensen, W. L. 1982. *J. Chem. Phys.* 77: 5757
98. Jorgensen, W. L. 1983. *J. Phys. Chem.* 87: 5304
99. Jorgensen, W. L., Binning, R. C., Bigot, B. 1981. *J. Am. Chem. Soc.* 103: 4393
100. Jorgensen, W. L., Blake, J. F., Buchner, J. K. 1988. *Chem. Phys.* In press
101. Jorgensen, W. L., Briggs, J. M., Gao, J. 1987. *J. Am. Chem. Soc.* 109: 6857
102. Jorgensen, W. L., Buchner, J. K. 1987. *J. Phys. Chem.* 91: 6083
103. Jorgensen, W. L., Buchner, J. K., Boudon, S., Tirado-Rives, J. 1988. *J. Am. Chem. Soc.* In press
104. Jorgensen, W. L., Buchner, J. K., Huston, S. E., Rossky, P. J. 1987. *J. Am. Chem. Soc.* 109: 1891
105. Jorgensen, W. L., Chandrasekar, J., Madura, J. D., Impey, R. W., Klein, M. L. 1983. *J. Chem. Phys.* 79: 926
106. Jorgensen, W. L., Gao, J. 1988. *J. Am. Chem. Soc.* 110: 4212
107. Jorgensen, W. L., Madura, J. D. 1985. *Mol. Phys.* 56: 1381
108. Jorgensen, W. L., Ravimohan, C. 1985. *J. Am. Chem. Soc.* 83: 3050
109. Karim, O. A., McCammon, J. A. 1988. *J. Chem. Phys.* In press
110. Karplus, M., Brady, J., Brooks, B., Kushick, J., Pettitt, B. M. 1985. See Ref. 79, p. 47
111. Karplus, M., Kushick, J. 1981. *Macromolecules* 14: 324
112. Karplus, M., McCammon, J. A. 1981. *FEBS Lett.* 131: 34
113. Karplus, M., McCammon, J. A. 1983. *Annu. Rev. Biochem.* 52: 263
114. Karplus, M., Porter, R. N., Jaime, R. D. 1965. *J. Chem. Phys.* 43: 3359
115. Kauzmann, W. 1959. *Adv. Protein Chem.* 14: 1
116. Keck, J. C. 1962. *Discuss. Faraday Soc.* 33: 173
117. Kim, P. S., Baldwin, R. L. 1982. *Annu. Rev. Biochem.* 51: 459
118. Kirkwood, J. G. 1968. In *Theory of Liquids*, ed. B. J. Alder. New York: Gordon & Breach
119. Kistenmacher, H., Popkie, H., Clementi, E. 1973. *J. Chem. Phys.* 59: 5842
120. Klapper, I., Hagstrom, R., Fine, R., Sharp, K., Honig, B. 1986. *Proteins* 1: 47
121. Kottalam, J., Case, D. A. 1988. *J. Am. Chem. Soc.* 110: 7690
122. Kuriyan, J., Petsko, G. A., Levy, R. M., Karplus, M. 1986. *J. Mol. Biol.* 190: 227
123. Latimer, W. M., Pitzer, K. S., Slansky, C. M. 1939. *J. Chem. Phys.* 7: 108
124. Lee, O., Roberts, G. M., Diem, M. 1988. *Biopolymers.* In press
125. Li, Z., Scheraga, H. A. 1988. *J. Phys. Chem.* 92: 2633
126. Lie, G. C., Clementi, E. 1975. *J. Chem. Phys.* 62: 2195
127. Lybrand, T. P., Ghosh, I., McCammon, J. A. 1985. *J. Am. Chem. Soc.* 107: 7793
128. Lybrand, T. P., McCammon, J. A., Wipff, G. 1986. *Proc. Natl. Acad. Sci. USA* 83: 833
129. Mackay, D. H. J., Cross, H. J., Hagler, A. J. 1989. In *Prediction of Protein Structure and Principles of Protein Conformation*, ed. G. Fasman. New York: Plenum. In press
130. Madden, P. A., Impey, R. W. 1986. *Ann. NY Acad. Sci.* 482: 91
131. Madison, V., Koppel, K. D. 1980. *J. Am. Chem. Soc.* 102: 4855
132. Madura, J. D., Jorgensen, W. L. 1986. *J. Am. Chem. Soc.* 108: 2517
133. Madura, J., McCammon, J. A. 1988. *Chem. Phys. Lett.* In press
134. Marcus, R. A. 1964. *Annu. Rev. Phys. Chem.* 15: 155
134a. Mares-Guia, M., Nelson, D. L., Rogana, E. 1977. *J. Am. Chem. Soc.* 99: 2336
135. Matsuoka, O., Clementi, E., Yoshimine, M. 1976. *J. Chem. Phys.* 64: 1351
136. Matthew, J. B. 1985. *Annu. Rev. Biophys. Biophys. Chem.* 14: 387
137. Matthew, J. B., Richards, F. M. 1984. *Biopolymers* 23: 2743
138. Matthews, B. W., Sigler, P. B., Henderson, R., Blow, D. M. 1967. *Nature* 214: 652
139. Matthews, D. A., Bolin, J. T., Burridge, J. M., Filman, D. J., Volz, K. W., et al. 1985. *J. Biol. Chem.* 280: 381
140. Matthews, D. A., Bolin, J. T., Burridge, J. M., Filman, D. J., Volz, K. W., Kraut, J. 1985. *J. Biol. Chem.* 280: 392
141. McCammon, J. A. 1987. *Science* 238: 486

142. McCammon, J. A., Harvey, S. C. 1987. *Dynamics of Proteins and Nucleic Acids*. London: Cambridge Univ. Press
143. McCammon, J. A., Karplus, M. 1979. *Proc. Natl. Acad. Sci. USA* 76: 3585
144. McCammon, J. A., Karplus, M. 1980. *Biopolymers* 19: 1375
145. McCammon, J. A., Lee, C. Y., Northrup, S. H. 1983. *J. Am. Chem. Soc.* 105: 2232
146. McDonald, I. R., Rasaiah, J. C. 1975. *Chem. Phys. Lett.* 34: 382
147. McQuarrie, D. A. 1976. *Statistical Mechanics*. New York: Harper & Row
148. Meirovitch, H., Vasquez, M., Scheraga, H. A. 1988. *Biopolymers* 27: 1189
149. Mezei, M. 1982. *Mol. Phys.* 47: 1307
150. Mezei, M. 1988. *J. Chem. Phys.* 86: 7084
151. Deleted in proof
152. Mezei, M., Beveridge, D. L. 1986. *Ann. NY Acad. Sci.* 482: 1
153. Mezei, M., Mehrotra, P. K., Beveridge, D. L. 1985. *J. Am. Chem. Soc.* 107: 2239
154. Mezei, M., Swaminathan, S., Beveridge, D. L. 1978. *J. Am. Chem. Soc.* 100: 3255
155. Mills, P., Anderson, C. F., Record, M. T. Jr. 1988. *J. Phys. Chem.* In press
156. Moore, G. R., Harris, D. E., Leitch, F., Pettigrew, G. W. 1984. *Biochim. Biophys. Acta* 764: 331
157. Mruzik, M. R. 1977. *Chem. Phys. Lett.* 48: 171
158. Mruzik, M. R., Abraham, F. E., Schreiber, D. E., Pound, G. M. 1976. *J. Chem. Phys.* 64: 481
159. Nilsson, L., Clore, G. M., Gronenboun, A. M., Brugner, A. J., Karplus, M. 1986. *J. Mol. Biol.* 188: 455
160. Northrup, S. H., Hynes, J. T. 1980. *J. Chem. Phys.* 73: 2700
161. Northrup, S. H., Pear, M. R., Lee, C. Y., McCammon, J. A., Karplus, M. 1982. *Proc. Natl. Acad. Sci. USA* 79: 4035
162. Okazaki, S., Nakanishi, K., Touhara, H., Adachi, Y. 1979. *J. Chem. Phys.* 71: 2421
163. Onsager, L. 1936. *J. Am. Chem. Soc.* 58: 1486
164. Owicki, J. C., Scheraga, H. A. 1977. *J. Am. Chem. Soc.* 99: 8392
165. Pack, G. R., Klein, B. J. 1984. *Biopolymers* 23: 2801
165a. Pangali, C. S., Rao, M., Berne, B. J. 1978. In *Computer Modeling and Matter*, ed. P. Lykos, p. 32. Washington, DC: Am. Chem. Soc.
166. Pangali, C. S., Rao, M., Berne, B. J. 1979. *J. Chem. Phys.* 71: 2975
167. Pangali, C. S., Rao, M., Berne, B. J. 1979. *J. Chem. Phys.* 71: 2982
168. Patey, G. N., Valleau, J. P. 1975. *J. Chem. Phys.* 63: 2334
169. Pettitt, B. M., Karplus, M. 1985. *J. Am. Chem. Soc.* 107: 1166
170. Pettitt, B. M., Karplus, M. 1985. *J. Chem. Phys.* 83: 781
171. Pettitt, B. M., Karplus, M. 1989. *J. Phys. Chem.* In press
172. Pettitt, B. M., Karplus, M., Rossky, P. J. 1986. *J. Phys. Chem.* 90: 6335
173. Pettitt, B. M., Rossky, P. M. 1982. *J. Chem. Phys.* 77: 1451
174. Pettitt, B. M., Rossky, P. J. 1983. *J. Chem. Phys.* 78: 7296
175. Pierotti, R. A. 1976. *Chem. Rev.* 76: 717
176. Pohorille, A., Pratt, L. R. 1986. *Methods Enzymol.* 127: 48
177. Postma, J. P. M. 1985. PhD thesis. Univ. Groningen, the Netherlands
178. Postma, J. P. M., Berendsen, H. J. C., Haak, J. R. 1982. *Faraday Symp. Chem. Soc.* 17: 55
179. Pratt, L. R., Chandler, D. 1973. *J. Chem. Phys.* 67: 3683
180. Pratt, L. R., Hsu, C. S., Chandler, D. 1978. *J. Chem. Phys.* 68: 4202
181. Quirke, N. 1983. In *The Physics of Superionic Conductors and Electrode Materials*, ed. J. E. Perram. New York: Plenum
182. Rao, S. N., Singh, U. C., Bash, P. A., Kollman, P. A. 1987. *Nature* 328: 551
183. Rappaport, D. C., Scheraga, H. A. 1982. *J. Phys. Chem.* 86: 873
184. Rashin, A. A., Honig, B. 1985. *J. Phys. Chem.* 89: 5588
185. Rashin, A. A., Namboodiri, K. 1987. *J. Phys. Chem.* 91: 6003
185a. Ravishanker, G., Beveridge, D. L. 1985. *J. Am. Chem. Soc.* 107: 2506
186. Ravishanker, G., Beveridge, D. L. 1986. In *Theoretical Chemistry of Biological Systems*, ed. G. Naray-Szabo, p. 423. Amsterdam: Elsevier
187. Ravishanker, G., Mezei, M., Beveridge, D. L. 1982. *J. Chem. Soc. Faraday Trans.* 17: 55
188. Ravishanker, G., Mezei, M., Beveridge, D. L. 1986. *J. Comput. Chem.* 7: 345
189. Rebertus, D. W., Berne, B. J., Chandler, D. J. 1979. *J. Chem. Phys.* 70: 3395
190. Rivall, J. L., Rinaldi, D. 1976. *Chem. Phys.* 18: 233
191. Rosenberg, R. O., Berne, B. J., Chandler, D. 1980. *Chem. Phys. Lett.* 75: 162
192. Rosenberg, R. O., Mikkilineni, R., Berne, B. J. 1982. *J. Am. Chem. Soc.* 104: 7647

193. Russell, S. T., Warshel, A. 1985. *J. Mol. Biol.* 185: 389
194. Ryckaert, J.-P., Bellemans, A. 1975. *Chem. Phys. Lett.* 30: 123
195. Ryckaert, J.-P., Bellemans, A. 1978. *Discuss. Faraday Soc.* 66: 95
196. Sarkisov, G. N., Dashevsky, V. G., Molenkov, G. G. 1974. *Mol. Phys.* 27: 1249
197. Scheraga, H. A. 1971. *Chem. Rev.* 21: 196
198. Scheraga, H. A., Paine, G. H. 1986. *Ann. NY Acad. Sci.* 482: 60
199. Schmitt, K. E. 1983. *Phys. Rev. Lett.* 51: 2175
200. Schnuelle, G. W., Beveridge, D. L. 1975. *J. Phys. Chem.* 79: 2566
201. Schnuelle, G. W., Swaminathan, S., Beveridge, D. L. 1978. *Theor. Chim. Acta* 48: 17
202. Shing, K. S., Gubbins, K. E. 1983. *Adv. Chem. Ser.* 204: 73
203. Shotten, D. M., Watson, H. C. 1970. *Nature* 225: 811
204. Singer, S. J., Chandler, D. 1985. *Mol. Phys.* 55: 621
205. Singh, U. C. 1988. *Proc. Natl. Acad. Sci. USA* 85: 4280
206. Singh, U. C., Brown, F. K., Bash, P. A., Kollman, P. A. 1987. *J. Am. Chem. Soc.* 109: 1607
207. Singh, U. C., Kollman, P. A. 1986. *J. Comput. Chem.* 7: 718
208. Stein, M. L., Claessens, M., Lopez, A., Reisse, J. 1982. *J. Am. Chem. Soc.* 104: 5902
209. Stillinger, F. H., Rahman, A. 1974. *J. Chem. Phys.* 60: 1545
210. Straatsma, T. P. 1987. PhD thesis. Univ. Groningen, the Netherlands
211. Straatsma, T. P., Berendsen, H. J. C. 1989. *J. Chem. Phys.* In press
212. Straatsma, T. P., Berendsen, H. J. C., Postma, J. P. M. 1986. *J. Chem. Phys.* 85: 6720
213. Straatsma, T. P., Berendsen, H. J. C., Storm, A. J. 1986. *Mol. Phys.* 57: 89
214. Stroud, R., Kay, L. M., Dickerson, R. E. 1974. *J. Mol. Biol.* 83: 185
215. Sussman, F. J., Goodfellow, J., Barnes, P., Finney, J. L. 1985. *Chem. Phys. Lett.* 113: 372
215a. Swaminathan, S., Beveridge, D. L. 1979. *J. Am. Chem. Soc.* 101: 5832
216. Swaminathan, S., Harrison, S. W., Beveridge, D. L. 1978. *J. Am. Chem. Soc.* 100: 5705
217. Swope, W. C., Andersen, H. C. 1984. *J. Phys. Chem.* 88: 6548
218. Tanabe, K., Saeki, S. 1975. *J. Mol. Struct.* 27: 79
219. Tanford, C., Kirkwood, J. G. 1957. *J. Am. Chem. Soc.* 79: 5333
220. Tembe, B. L., McCammon, J. A. 1982. *Comput. Chem.* 8: 281
221. Tilton, R. F. Jr., Singh, U. C., Weiner, S. J., Connolly, M. L., Kuntz, I. D., et al. 1986. *J. Mol. Biol.* 192: 443
222. Deleted in proof
223. Torrie, G. M., Valleau, J. P. 1977. *J. Comput. Phys.* 23: 187
224. Tronrud, D. E., Holden, H. M., Matthews, B. W. 1987. *Science* 235: 571
225. Valleau, J. P., Torrie, G. M. 1977. In *Modern Theoretical Chemistry*, ed. B. J. Berne, 5: 169. New York: Plenum
226. Deleted in proof
227. van Gunsteren, W. F. 1989. See Ref. 230a. In press
228. van Gunsteren, W. F., Berendsen, H. J. C. 1987. *J. Comput.-Aided Design* 1: 171
229. van Gunsteren, W. F., Berendsen, H. J. C. 1988. In *Computational Methods in Chemical Design*, ed. J. Stosowski. Oxford, UK: Oxford Univ. Press. In press
230. van Gunsteren, W. F., Berendsen, H. J. C., Guertsen, R. G., Zwinderman, H. R. J. 1986. *Ann. NY Acad. Sci.* 982: 298
230a. van Gunsteren, W. F., Weiner, P., eds. 1989. *Computation of Free Energy for Biomolecular Systems.* Leiden, the Netherlands: Escom. In press
231. Wada, A. 1954. *J. Chem. Phys.* 22: 198
232. Wagner, G. A., DeMarco, A., Wuthrich, K. 1976. *Biophys. Struct. Mech.* 2: 139
233. Warshel, A. 1981. *Acc. Chem. Res.* 14: 284
234. Warshel, A. 1981. *Biochemistry* 20: 3167
235. Warshel, A. 1982. *J. Phys. Chem.* 86: 2218
236. Warshel, A., Karplus, M. 1972. *Chem. Phys. Lett.* 17: 7
237. Warshel, A., Levitt, M. 1976. *J. Mol. Biol.* 103: 227
238. Warshel, A., Russell, S. T. 1984. *Q. Rev. Biophys.* 17: 283
239. Warshel, A., Sussman, F. 1986. *Proc. Natl. Acad. Sci. USA* 83: 3806
240. Warshel, A., Sussman, F., Hwong, J. K. 1988. *J. Mol. Biol.* 201: 139
241. Warshel, A., Sussman, F., King, G. 1986. *Biochemistry* 25: 8368
242. Warwicker, T., Watson, H. C. 1982. *J. Mol. Biol.* 157: 671
243. Watanabe, K., Andersen, H. C. 1986. *J. Phys. Chem.* 90: 795
244. Weber, T. A. 1978. *J. Chem. Phys.* 69: 2347
245. Weber, T. A. 1979. *J. Chem. Phys.* 70: 4272

246. Weber, T. A., Helfand, E. 1980. *J. Chem. Phys.* 72: 4014
247. Wells, J. A., Cunningham, B. C., Craycar, T. P., Estell, D. A. 1986. *Philos. Trans. R. Soc. London Ser. A* 316: 415
248. Wilson, K. R. 1988. In *Chemical Reactivity in Liquids*, ed. M. Moreau, P. Turq. New York: Plenum
249. Wolfenden, R., Anderson, L., Cullis, P. M., Southgate, C. C. B. 1981. *Biochemistry* 20: 849
250. Wong, C. F., McCammon, J. A. 1986. *J. Am. Chem. Soc.* 108: 3830
251. Wood, W. W. 1968. In *Physics of Simple Liquids*, ed. H. N. V. Temperly, J. S. Rowlinson, G. Rushbrooke. p. 115. Amsterdam: North-Holland
252. Zichi, D. A., Rossky, P. J. 1985. *J. Chem. Phys.* 83: 798
253. Zichi, D. A., Rossky, P. J. 1986. *J. Chem. Phys.* 84: 1712
254. Zwanzig, R. W. 1954. *J. Chem. Phys.* 22: 1420

Annual Review of Biophysics and Biophysical Chemistry
Volume 18, 1989

CONTENTS

(*continued*) v