PROTEINS: Structure, Function, and Genetics 20:25–33 (1994)

Free Energy Simulations: The Meaning of the Individual Contributions From a Component Analysis

Stefan Boresch, Georgios Archontis, and Martin Karplus
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT **A theoretical analysis is made of the decomposition into contributions from individual interactions of the free energy calculated by thermodynamic integration. It is demonstrated that such a decomposition, often referred to as "component analysis," is meaningful, even though it is a function of the integration path. Moreover, it is shown that the path dependence can be used to determine the relation of the contribution of a given interaction to the state of the system.**

To illustrate these conclusions, a simple transformation (Cl^- to Br^- in aqueous solution) is analyzed by use of the Reference Interaction Site Model–Hypernetted Chain Closure integral equation approach; it avoids the calculational difficulties of macromolecular simulation while retaining their conceptual complexity. The difference in the solvation free energy between chloride and bromide is calculated, and the contributions of the Lennard-Jones and electrostatic terms in the potential function are analyzed by the use of suitably chosen integration paths. The model is also used to examine the path dependence of individual contributions to the double free energy differences ($\Delta\Delta G$ or $\Delta\Delta A$) that are often employed in free energy simulations of biological systems. The alchemical path, as contrasted with the experimental path, is shown to be appropriate for interpreting the effects of mutations on ligand binding and protein stability. The formulation is used to obtain a better understanding of the success of the Poisson-Boltzmann continuum approach for determining the solvation properties of polar and ionic systems.

Key words: hemodynamic integration, RISM theory, alchemical path

INTRODUCTION

Molecular dynamics and Monte Carlo simulations, as well as Poisson-Boltzmann calculations and integral equation methods, are widely used for evaluating free energy differences in solution and in macromolecules.[1–3] Applications to biomolecules have attracted considerable attention, in particular because they provide information concerning contributions of different system components to the overall free energy change. Examples include the effect of mutations on protein stability,[4–6] on ligand binding,[7,8] and on cooperativity.[9] Insights obtained from such simulations concerning the role of different residues and the effect of solvent supplement interpretations are based on structural and thermodynamic data. A recent simulation of substrate binding by Tyrosyl-t RNA synthetase,[8] for example, has provided a deeper understanding of the widely used hydrogen-bond inventory model.[10,11]

In spite of their demonstrated utility and agreement with experiment, the validity of results obtained from free energy simulations of biomolecules has been questioned recently.[12–14] Both calculational and theoretical aspects of the simulations have been criticized. The calculational concerns focus primarily on the possibility of statistical errors in the simulations due to the difficulty of adequate sampling. Although the convergence of the calculated free energy change needs to be assessed in each application, many published simulations have been done carefully and have obtained significant results.[3,8,15] One area in which detailed verification of the statistical error limits has been possible is in the simulation of the effect of sidechain mutations on the stability of small peptides.[16] The present article addresses a theoretical aspect of the published criticisms. Although the overall free energy difference between two end states is independent of the path connecting them, the individual contributions can depend on the path.[8,12,13,17,18] It has been stated[12]: "In a number of studies, a breakdown of a free energy change or difference into components consisting of contributions of groups of atoms, e.g., amino acid residues, or of types of interaction, e.g., van der Waals, electrostatic, bonding interaction, has been presented. Yet such a decomposition cannot be meaningfully defined. . . ." We disagree with this statement. By formulating the thermodynamic integration method so as to make clear the dependence on the system parameters, we show how the decomposition can be used to obtain meaningful in-

Received February 22, 1994; revision accepted May 5, 1994.
Address reprint requests to Martin Karplus, Department of Chemistry, 12 Oxford St., Harvard University, Cambridge, MA 02138.

sights into the nature of the interactions involved. In any such decomposition the integration path must be taken into account in the interpretation of the results. Rather than making the free energy decomposition "meaningless,"[12] the path dependence introduces an additional degree of freedom that can be used to obtain a deeper understanding of the interactions contributing to a free energy change.

We first outline the theory of free energy simulations and explicitly introduce a generalized parameter dependence of the potential energy that makes possible the separation of individual contributions to the free energy change. We use this formulation to show how the contribution of a given interaction depends on the state of the system. We then apply the theory to a very simple example, which contains all the conceptual complexities of macromolecular applications of free energy simulations and yet permits us to obtain results free from statistical errors. This contrasts with two published papers that have claimed to demonstrate a path dependence of the free energy decomposition. In the paper of Shi et al.,[12] cited by the editorial comment in this journal,[14] the mutation Asn → Ser in subtilisin was simulated two times and different overall free energies and different decompositions were obtained. Since the final states found by the two simulations were different, nothing can be concluded about the path dependence. The paper by Simonson and Brünger[17] used a heuristic example to demonstrate that there is a path dependence in the free energy decomposition. However, the specific example was such that the two paths considered yielded identical results; more recently, Hodel et al.[19] briefly described another heuristic example, which could have a path dependence, but no results were given. In the present paper, we use a simple model that explicitly demonstrates the path dependence of the contributions to the overall free energy change and shows how this dependency can be utilized to obtain a better understanding of the underlying behavior of the system. The model is also employed to examine the path dependence of the decomposition of double free energy differences (ΔΔG or ΔΔA). The latter results are of particular interest for the decomposition of the free energy in biomolecular systems. It is shown that the alchemical path, commonly employed in experimental and theoretical analyses of protein mutations, can be used to obtain information that is not directly available from experiment. Finally, the free energy decomposition is used to provide a better understanding of Poisson-Boltzmann calculations of solvation for charged and polar systems.

THEORY

In the regime of classical statistical mechanics, the Helmholtz free energy, A_λ, for a system described by a potential energy function $V(\mathbf{r}^N,\lambda)$, where $\mathbf{r}^N$ represents the particle coordinates and λ is a parameter, can be written in the form[1]

$$A_\lambda = C(N,V,T) - (1/\beta)\ln Z_\lambda^c. \tag{1}$$

Here C(N,T,V) is obtained from the kinetic energy portion of the partition function and Z_λ^c is the classical configuration integral,

$$Z_\lambda^c = \int_V \exp[-\beta V(\mathbf{r}^N,\lambda)]d\mathbf{r}^N \tag{2}$$

with $\beta = 1/k_BT$ and k_B the Boltzmann constant; the Gibbs free energy can be written correspondingly except that C(N,P,T) replaces C(N,V,T). Differentiating A_λ from Eq. 1 with respect to λ, we have

$$\frac{\partial A\lambda}{\partial\lambda} = (Z_\lambda^c)^{-1}\int_V\left(\frac{\partial V(\mathbf{r}^N;\lambda)}{\partial\lambda}\right)\exp[-\beta V(\mathbf{r}^N,\lambda)]\,d\mathbf{r}^N. \tag{3}$$

Since the right-hand side of Eq. 3 is the expectation value of the quantity $(\partial dV(\mathbf{r}^N;\lambda)/\partial\lambda)$ for an ensemble with probability distribution, $P(\mathbf{r}^N,\lambda)$,

$$P(\mathbf{r}^N,\lambda) = \frac{\exp[-\beta V(\mathbf{r}^N;\lambda)]}{Z_\lambda^c} \tag{4}$$

we can rewrite Eq. 3 in the form

$$\frac{\partial A\lambda}{\partial\lambda} = \left\langle\frac{\partial V(\mathbf{r}^N;\lambda)}{\partial\lambda}\right\rangle A_\lambda \tag{5}$$

where the bracket with subscript λ indicates an ensemble average with the probability distribution (Eq. 4) corresponding to the potential $V(\mathbf{r}^N,\lambda)$.

The parameter λ can be used to change the system from an initial state ($\lambda = 0$) to a final state ($\lambda = 1$). For the solvation free energy difference between ethane and methanol,[20] for example, the potential $V(\mathbf{r}^N,\lambda)$ with $\lambda = 0$ corresponds to that of solvated ethane and that with $\lambda = 1$ to solvated methanol. To calculate the free energy difference, ΔA_{01}, between the initial and final state, we integrate over λ between 0 and 1.[1,21]

$$\Delta A_{01} = A_1 - A_0 = \int_0^1\left(\frac{\partial A_\lambda}{\partial\lambda}\right)d\lambda = \int_0^1\left\langle\frac{\partial V(\mathbf{r}^N;\lambda)}{\partial\lambda}\right\rangle_\lambda d\lambda \tag{6}$$

By taking small enough steps in λ to approximate the integral and evaluating the integrand at each value of λ by a long enough simulation, a reversible change from the initial to the final state is followed on the computer. As in a laboratory experiment, it yields the desired free energy difference between the state corresponding to $\lambda = 0$ and the state corresponding to $\lambda = 1$. From Eq. 6 it is clear that ΔA_{01} is independent of the path λ because the integrand is an exact differential.

If the potential energy function $V(\mathbf{r}^N,\lambda)$ can be expressed as a sum of terms,[22–24]

$$V(\mathbf{r}^N,\lambda) = \sum_m V_m(\mathbf{r}^N,\lambda) \quad (7)$$

we have

$$\left\langle \frac{\partial V(\mathbf{r}^N;\lambda)}{\partial\lambda} \right\rangle_\lambda = \sum_m \left\langle \frac{\partial V_m(\mathbf{r}^N;\lambda)}{\partial\lambda} \right\rangle_\lambda . \quad (8)$$

Substituting into Eq. 6, we obtain

$$\Delta A_{01} = \sum_m (\Delta A_{01})_m \quad (9)$$

with

$$(\Delta A_{01})_m = \int_0^1 \left\langle \frac{\partial V_m(\mathbf{r}^N;\lambda)}{\partial\lambda} \right\rangle_\lambda d\lambda. \quad (10)$$

Eqs. 9 and 10 show that for a system described by a potential function of the form given in Eq. 7, the difference between the initial and final state free energy, ΔA_{01}, can be decomposed into a sum of terms, $(\Delta A_{01})_m$, each of which is associated with the derivative of a specific contribution to the potential function. It is useful to specialize to the case in which the change in the potential functions varies linearly with λ; then

$$V(\mathbf{r}^N,\lambda) = V_0(\mathbf{r}^N) + \Delta V(\mathbf{r}^N)\lambda = V_0(\mathbf{r}^N) + \sum_m \Delta V_m(\mathbf{r}^N)\lambda \quad (11)$$

where the ΔV_m are the terms in the potential energy that are altered between the initial state ($\lambda = 0$) and the final state ($\lambda = 1$). With this form for the potential, the derivative in Eq. 10 becomes

$$(\Delta A_{01})_m = \int_0^1 \langle \Delta V_m(\mathbf{r}^N)\rangle_\lambda d\lambda \quad (12)$$

i.e., the contribution of the m^{th} term in the potential energy function is equal to the expectation value of its difference between the initial and final state, averaged over a distribution corresponding to $V(\mathbf{r}^N,\lambda)$ and integrated from $\lambda = 0$ to $\lambda = 1$.

Although Eqs. 10 and 12 are correct, they implicitly involve the choice of an integration path along which all of the terms in the potential vary simultaneously with the parameter λ. To make this explicit, we introduce two parameters, λ_s and λ_m, such that

$$V(\mathbf{r}^N,\lambda_s,\lambda_m) = V_0(\mathbf{r}^N) + \sum_{s\neq m} \Delta V_s(\mathbf{r}^N)\,\lambda_s + \Delta V_m(\mathbf{r}^N)\lambda_m \quad (13)$$

where λ_s includes all terms in the potential function except the m^{th} term. Following a procedure corresponding to that employed in going from Eq. 3 to 6, we find

$$\frac{\partial A_{\lambda_s,\lambda_m}}{\partial\lambda_m} = (Z^c_{\lambda_s,\lambda_m})^{-1} \int_V \Delta V_m(\mathbf{r}^N) \exp[-\beta V(\mathbf{r}^N;\lambda_s,\lambda_m)]\, d\mathbf{r}^N = \left\langle \frac{\partial V(\mathbf{r}^{N,\lambda_s,\lambda_m})}{\partial\lambda_m} \right\rangle_{\lambda_m,\lambda_s} = \langle \Delta V_m(\mathbf{r}^N)\rangle_{\lambda_m,\lambda_s}. \quad (14)$$

To obtain the free energy difference between $\lambda_m = 0$ and $\lambda_m = 1$, we integrate Eq. 14 over λ_m for a fixed value of λ_s,

$$(\Delta A_{01})_{m,\lambda_s} = \int_0^1 \left\langle \frac{\partial V(\mathbf{r}^{N,\lambda_s,\lambda_m})}{\partial\lambda_m} \right\rangle_{\lambda_m,\lambda_s} d\lambda_m = \int_0^1 \langle \Delta V_m(\mathbf{r}^N)\rangle_{\lambda_m,\lambda_s}\, d\lambda_m. \quad (15)$$

Eq. 15 gives the contribution to the free energy change of the term $(\partial V_m(\mathbf{r}^N)/\partial\lambda_m)$ or ΔV_m in going from $\lambda_m = 0$ to $\lambda_m = 1$. It is independent of the integration path for λ_m, but corresponds to a fixed value of λ_s.

RESULTS AND ANALYSIS

To illustrate the formulation outlined above, we consider an example that avoids certain aspects of published work[12] in which the complexity of the system makes it difficult to determine whether there are calculational or theoretical problems in the analysis. We then use the insights obtained from the example to examine the results of simulations for more complex systems of biological interest.

A Simple Example Illustrating the Dependence of the Components on the State of the Rest of the System

The example treats the difference in solvation free energy of two negatively charged monatomic ions that approximate Cl^- and Br^-. The two ions are identical except for their nonbonded (Lennard-Jones) parameters.[25,26] We base the calculations on the Reference Interaction Site Model–Hypernetted Chain Closure (RISM-HNC) integral equation method[27] because it gives exact results within the assumptions of the model, i.e., there are no statistical errors. For the solvation free energy difference, Eqs. 6 and 12 yield

$$\Delta A = \Delta A_{LJ} = \int_0^1 \langle \Delta V\rangle_\lambda\, d\lambda \quad (16)$$

where

$$\Delta V = V(Br^-) - V(Cl^-) = V_{LJ}(Br^-) - V_{LJ}(Cl^-) \quad (17)$$

and V_{LJ} refers to the Lennard-Jones interaction between the ions and the solvent molecules. Since only the solute-solvent Lennard-Jones interaction is being altered, the total free energy change is equal to the solute-solvent Lennard-Jones contribution to the free energy. The other terms in the empirical energy function, which are the electrostatic solute-

TABLE I. Free Energy Differences (kcal/mol)*

	ΔA	ΔΔA
$Cl^- \rightarrow Br^-$	4.65	3.53
$Cl^0 \rightarrow Br^0$	1.12	
$Cl^- \rightarrow Cl^0$	94.09	3.53
$Br^- \rightarrow Br^0$	90.56	

*The parameters used are $\varepsilon = -0.107$ kcal/mol, $\sigma = 4.446$ Å for Cl^- and Cl^0 and $\varepsilon' = -0.090$ kcal/mol, $\sigma' = 4.623$ Å for Br^- and Br^0; a TIP3P water model modified for RISM calculations was used.

solvent interaction terms and the solvent-solvent interaction terms, do not change their values in the transformation of Cl^- into Br^-. They have their full value (Eq. 15 with $\lambda_s = 1$) so that the difference in the Lennard-Jones contribution to the solvation of Br^- versus Cl^- is being evaluated in the presence of the normal system. The numerical results of the RISM-HNC calculation are given in Table I; the Lennard-Jones parameters ε and σ for the two ions are listed in the footnote to the table.

To obtain insight into the origin of the free energy difference between the two ions, we consider the solvation of two uncharged atoms (Cl^0, Br^0) with the same set of Lennard-Jones parameters. The change in the potential function, ΔV, is identical to that given by Eq. 17, but the free energy difference, which involves the ensemble average, $\langle \Delta V \rangle_\lambda$, is different. From Table I, it can be seen that the free energy change equals 1.12 kcal/mol, a value significantly smaller than that for the charged system, which is 4.65 kcal/mol. The difference arises from the probability factor $P(\mathbf{r}^N,\lambda)$, which appears in $\langle \Delta V \rangle_\lambda$; see Eq. 4. In the first case, the potential $V(\mathbf{r}^N,\lambda)$ in the exponential of $P(\mathbf{r}^N,\lambda)$ corresponds to that for a negative ion, and in the second case the potential is that for a neutral (Lennard-Jones) sphere. For the present case, we can write[27,28]

$$\Delta A = \rho_{solv} \sum_v \int_0^1 d\lambda_m \int_{Vol} \Delta V(\mathbf{r}^N)\, g_{solu,v}(\mathbf{r}^N,\lambda_m)_{\lambda_s=1} d\mathbf{r}^N \quad (18)$$

where ρ_{solv} is the solvent number density and $g_{solu,v}(\mathbf{r}^N, \lambda_m)_{\lambda_s=1}$ represents the solute-solvent radial distribution functions; the sum over v goes over the hydrogen and oxygen atoms in the solvent molecule. The radial distribution function, which is simply related to the probability distribution $P(\mathbf{r}^N,\lambda)$,[28] depends on the integration parameter λ_m and on λ_s, the parameter determining the other contributions to the potential, i.e., the charged and neutral spheres correspond to different λ_s and the solute-solvent radial distributions functions are different. Figure 1 shows the solute-solvent radial distribution functions for Cl^- and for the neutral sphere with the same van der Waals parameters (Cl^0). The functions are more sharply peaked and closer to the solute for the charged sphere than the neutral sphere. This leads to the larger Lennard-Jones solvation free energy for Cl^- than Cl^0; the interaction between the anion and the water hydrogens is particularly important.

The above results show how the path-dependent decomposition can be used to demonstrate the role of the system properties in the free energy change. In the example, we considered the effect of the particle charge on the Lennard-Jones contribution to the solvation free energy. An exact decomposition was made for both the mononegative ion and the neutral sphere, with fixed values of λ_s in Eq. 15.

Use of the Simple Example to Demonstrate the Essential Path Dependence of the Components

To demonstrate explicitly the path dependence of the free energy decomposition, we consider a thermodynamic cycle (Scheme I) of the type that is often used in such calculations.[3,29] Values for the free energy changes corresponding to the four arrows in Scheme I were calculated with the RISM-HNC model and are listed in the scheme, as well as in Table I; all values are in kcal/mol. Scheme I involves the four species we have considered already. Two of them (Cl^- and Br^-)

$$\begin{array}{ccc} & \Delta A_{LJ}^{-1} = 4.65 & \\ Cl^- & \longrightarrow & Br^- \\ \Delta A_{el} = 94.09 \downarrow & & \downarrow \Delta A'_{el} = 90.56 \\ Cl^0 & \longrightarrow & Br^0 \\ & \Delta A_{LJ}^{0} = 1.12 & \end{array} \quad (I)$$

have a charge of minus one and the other two (Cl^0 and Br^0) are uncharged spheres with the same Lennard-Jones parameters. The transformations described above correspond to the horizontal arrows in Scheme I; they yielded the difference in the solvation free energy for the negative ions, ΔA_{LJ}^{-1}, and for the neutral spheres, ΔA_{LJ}^{0}, due the same change in the Lennard-Jones potential parameters. The vertical arrows, which correspond to discharging the Cl^- and Br^- ions, are discussed in relation to double free energy differences (see below).

By use of Scheme I, we can compare the decompo-

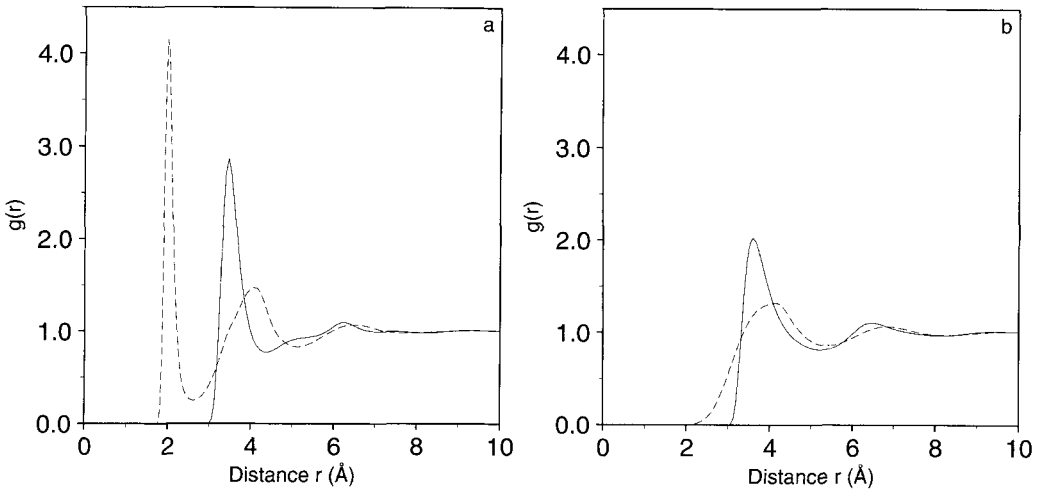


Fig. 1. Radial distribution functions g(r) for monatomic ions and corresponding Lennard-Jones spheres. **a:** Cl^- (—) ion-oxygen, (- - -) ion-hydrogen. **b:** Cl^0 (—) sphere-oxygen, (- - -) sphere hydrogen.

sition of the difference in the solvation free energy between Cl^- and Br^- along two paths. There is the direct path, corresponding to the top horizontal arrow, and the path that goes around the cycle ($Cl^- \rightarrow Cl^0 \rightarrow Br^0 \rightarrow Br^-$). Since the two paths for Cl^- to Br^- yield identical free energy differences, we have

$$\Delta A_{LJ}^{-1} = \Delta A_{el} + \Delta A_{LJ}^{0} - \Delta A'_{el} \qquad (19)$$

Eq. 19 is satisfied exactly (see Table I). By the direct path, corresponding to the left-hand side of Eq. 19, a pure Lennard-Jones contribution is obtained. The alternative path, which corresponds to the right-hand side of Eq. 19 and involves the rest of the cycle, yields the difference between the two large electrostatic terms ($\Delta A_{el} - \Delta A'_{el}$) plus a smaller Lennard-Jones contributions, ΔA_{LJ}^{0}. The two different paths project the free energy change on different terms in the potential. This makes clear the importance of choosing the path or paths appropriate for the problem at hand.

Double Free Energy Differences

Macromolecular applications[3,8,29] often focus on the "double" free energy difference, ΔΔA. In Scheme I, the two ΔΔA values are:

$$\Delta\Delta A = \Delta A_{LJ}^{-1} - \Delta A_{LJ}^{0} = \Delta A'_{el} - \Delta A_{el} . \qquad (20)$$

The first contribution, ($\Delta A_{LJ}^{-1} - A_{LJ}^{0}$), corresponding to the horizontal arrows and arises entirely from the change in the Lennard-Jones solute-solvent term in the empirical potential. The second, ($\Delta A'_{el} - \Delta A_{el}$), corresponds to the vertical arrows and arises entirely from the change in the electrostatic solute-solvent term in the empirical potential, i.e., $\Delta V = -V_{el}$, where V_{el} is the Coulomb interaction between the negative (−1) ion and the water molecules. The free energy changes, ΔA_{el} and $\Delta A'_{el}$, correspond to Eq. 15 with λ_s equal to 0 or 1, where $\lambda_s = 0$ now implies that the solute-solvent Lennard-Jones term is fixed at its value for Cl^- and $\lambda_s = 1$ implies that the solute-solvent Lennard-Jones term is fixed at the value for Br^-. A purely electrostatic free energy difference is obtained due to discharging an ion from (−1) to zero for two different Lennard-Jones spheres. The quantities ΔA_{el} and $\Delta A'_{el}$ can be interpreted in terms of a Born solvation model calculation[30,31]; the difference arises primarily from the change in the Born radius associated with the Lennard-Jones parameter σ and σ′.

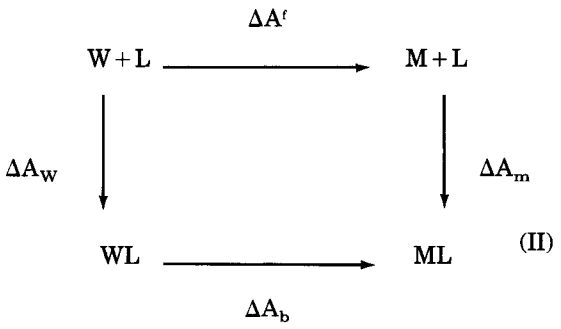
The two ways of calculating ΔΔA yield results that are identical in numerical value (see Table I), as they must. However, the origin of the free energy change is different because the physical changes being compared are different. Each of the four free energy changes in Scheme I corresponds to the alteration of a single contribution to the potential energy: the Lennard-Jones solute-solvent interaction for the horizontal arrows and the electrostatic solute-solvent interaction for the vertical arrows. According to Eq. 15, each "decomposition" is exact and is equal to the entire free energy change in the given process. Nevertheless, for the different paths, the double free energy difference is projected as a different term in the potential. The analysis is neither "meaningless"[12] nor uninteresting. The two ΔΔA decompositions in Eq. 20 provide alternative, valid interpretations for the free energy difference of solvation between the various species examined.

The choice of a specific path in the present example, and in other case, depends on the questions of interest because the path projects the free energy change on certain interactions. Along the horizontal paths in Scheme I one obtains the effect of changing the van der Waals radius for a charged ion or a neu-

tral atom in water. The large difference between the two demonstrates the importance of the electrostatic interaction in determining the Lennard-Jones contribution. Conversely, along the vertical paths, the effect of discharging two ions of different radii in aqueous solution is determined. This provides information on the effect of ion size on the electrostatic interaction with the solvent.

Applications to Biomolecules

With an understanding of the results from the chloride to bromide example, we are now ready to consider some macromolecular problems that have been studied by free energy simulations. The first example compares a ligand (L) binding to a wild type (W) and a mutant (M) protein[8]; studies of binding of two different substrates by the same protein[29,32] can be analyzed by the same approach. The thermodynamic cycle is:

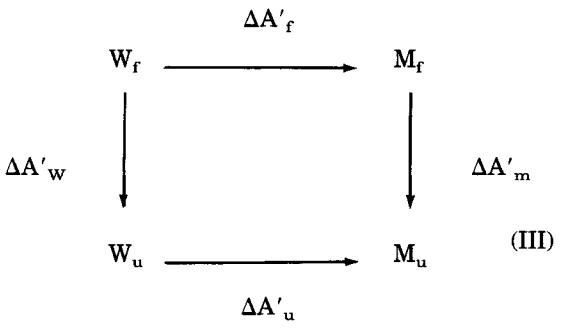


(II)

The ΔA symbols in Scheme II refer to the free energy changes that correspond to the arrows; we have

$$\Delta\Delta A = \Delta A_b - \Delta A_f = \Delta A_m - \Delta A_w. \quad (21)$$

Here f refers to the unliganded protein, b to the liganded protein, w to the wild-type protein, and m to the mutant protein. The second case concerns protein stability[4-6] and can be represented by a similar cycle:



(III)

It compares the difference in the stability of a wild-type (W) and a mutant protein (M). The ΔA symbols in Scheme III refer to the free energy changes corresponding to the arrows; we have

$$\Delta\Delta A' = \Delta A'_u - \Delta A'_f = \Delta A'_m - \Delta A'_w. \quad (22)$$

Here f refers to the folded protein, u to the unfolded protein, w to the wild-type protein, and m to the mutant protein.

Thermodynamic cycles like II and III were introduced by Tembe and McCammon[29] because they realized that it was much simpler to do simulations that follow the horizontal (alchemical) paths, which involve only relatively small changes, and that Hess's law could be used as in Eqs. 21 and 22 to obtain the answer for the vertical (chemical) paths that correspond to the experimental measurements. It is clear from simple physical reasoning, as exemplified by the model system (Scheme I), that for Schemes II and III the free energy contributions along the alchemical and chemical paths are expected to be different, although the values of ΔΔA obtained from the two paths are identical. Because only the alchemical path could be followed easily (although this is not always true, e.g., both paths can be evaluated in Scheme I), alchemical free energy differences have been calculated in most simulations of macromolecules and have been used to determine the contributing interactions.

To make clear the role of the path dependence in the analysis of protein folding, we differentiate between two types of questions: One concerns the effect of an amino acid on protein stability and the other concerns the role of an amino acid along the folding pathway. In most mutant studies, the focus has been on the effect on stability, where the primary interest is in understanding the role of the mutated amino acid in the folded state and in the unfolded state. This is clearly the objective of the experimental work of Matthews et al.[33] on T4 lysozyme, of Fersht[11] on barnase, and of Shortle and Meeker[34] on staphylococcus nuclease. In each case, the analysis of the effect of a given mutation on stability focuses on what happens in the folded state ($\Delta A'_f$) and the unfolded state ($\Delta A'_u$). Attempts are made to infer $\Delta A'_f$ and $\Delta A'_u$ from the experimental data for the wild-type and mutant protein. The earliest mutant analyses, in fact, considered only the folded state, for which it was straightforward to obtain structural information, and neglected effects in the unfolded state.

Free energy simulations along the alchemical path make it possible to examine directly the effect of a mutation in the native state ($\Delta A'_f$) and in unfolded state ($\Delta A'_u$). In an analysis of the mutant of Arg 96 to His in T4 lysozyme,[6] for example, alchemical simulations were made of the transformation of Arg 96 to His in the solvated folded state and in a

solvated extended chain model for the unfolded state. It was shown that a number of contributions that stabilize the wild type or the mutant partially cancel in the overall free energy difference; some of these involved the unfolded state. Comparison of the results with conclusions based on structural and thermodynamic data[35,36] led to new insights into the origin of the stability difference between wild-type and mutant proteins. Of particular interest was the importance of the contributions of more distant residues, solvent water, and the covalent linkage of the mutated amino acid. Also, the analysis of the interactions of Arg/His 96 with the C-terminal end of a helix (residues 82–90) made it clear that the nearby carbonyl groups (Tyr 88 and Asp 89) made the dominant contribution, that the amide groups did not contribute significantly, and that the helix-dipole model was inappropriate for this case.

An essential point that follows from the present treatment is that the decomposition of the free energy change along the alchemical paths is of primary interest for understanding the mutation experiments on ligand binding and protein stability. This has been illustrated here for protein stability and was emphasized for the case of ligand binding (Scheme II) in a simulation study of the effect of the mutation Tyr 169 to Phe in tyrosyl-tRNA synthetase,[8] which provided a microscopic justification of the widely used hydrogen bond inventory model.[10,11]

The component analysis along the chemical paths in Scheme III ($\Delta A'_w$ and $\Delta A'_m$), which correspond to the actual unfolding reaction in the wild-type and mutant protein, respectively, would clearly be very different from that along the alchemical path. The contributions correspond to those expected for the potential of mean force along a reaction path. Examples of reaction path calculations along which the components have been determined include the motion of an ion through a channel[37] and the interaction between two neon atoms as a function of the distance between them.[38] Fersht[39] has used mutants and structure-activity relations to estimate experimentally the contributions of different residues to the free energy along the folding and unfolding paths of barnase. Related simulations of the unfolding of barnase have been reported recently.[40]

Poisson-Boltzmann Solvation Free Energy Calculations

Eq. 19 is helpful in understanding the success of the linearized Poisson-Boltzmann equation in calculating solvation free energies[41] and free energy differences involved in the evaluation of pK_a values[42] for polar and charged systems. We consider the difference between the solvation free energy of Cl^- and Br^- given in Table I. The Poisson-Boltzmann treatment follows the vertical paths in Scheme I and neglects any direct Lennard-Jones contributions, i.e., it equates the difference in solvation free energy to the electrostatic terms ($\Delta A_{el} - \Delta A'_{el}$). In a more exact calculation, the path on the right-hand side of Eq. 19 would be used. Since the Lennard-Jones contribution to the free energy difference is smaller for an uncharged sphere, neglecting its change from one discharged species to another can often give a satisfactory approximation to the total solvation free energy difference. It should also be noted that the Poisson-Boltzmann results can be decomposed into contributions from individual system components into contributions from individual system components,[41,43] as can the RISM-HNC results.[44,45] The relation of such a decomposition to that obtained from free energy simulations will be considered separately (M. Sommer, M. Schaefer, A. MacKerrell, and M. Karplus, to be published).

CONCLUSIONS

A theoretical formulation of free energy simulations and its application to a model problem have been used to examine the use of thermodynamic integration to determine the contribution of individual terms in the potential energy function to the overall free energy change. If there is more than one potential term related to a given degree of freedom, a path dependence is inherent in the approach. This provides additional flexibility that makes it possible to explore different aspects of the important interactions. However, the path dependence of the results requires care in their interpretation; i.e., the choice of path must be taken into account in the interpretation of the decomposition. It has been shown that the alchemical path used in simulations is appropriate for obtaining physical insights from free energy simulations of macromolecular processes, such as the effect of mutations on ligand binding and protein stability. By contrast, the chemical pathway, which is more difficult to follow, can be used to obtain information about the effect of amino acid changes along the reaction path, such as that involved in protein unfolding.

There are aspects of free energy simulations and their component analysis that need to be considered in addition to the question resolved in this paper. One concerns the meaning of the decomposition when constraints are introduced or special choices are made for the parameter dependence of the potential. An example arises in non-linear coupling schemes, which are widely used for the Lennard-Jones terms when a particle is created or destroyed in an alchemical free energy change.[20] Also, it is likely that the determination of individual contributions along a given path requires longer simulations than evaluation of the total free energy change. Such questions are addressed in a separate paper (S. Boresch and M. Karplus, to be submitted).

The conclusion from the present investigation is that, when used with care and a proper choice of the path, the decomposition of the free energy can give essential insights into the role of the individual contributions that are hidden from direct experimental measurements.[9]

ACKNOWLEDGMENTS

The authors thank A.T. Brünger, J. Hermans, M. Prévost, T. Simonson, T.M. Straatsma, B. Tidor, S. Wodak, and H.-A. Yu for helpful discussions. M.K. thanks W. van Gunsteren for sending a preprint of ref. 12 prior to publication and Escom Publishers for sending preprints of refs. 13 and 18 prior to publication. This work was supported in part by The National Science Foundation and a gift from Molecular Simulations, Inc.

REFERENCES

1. Brooks, C.L. III, Karplus, M., Pettitt, B.M. "Proteins: A Theoretical Perspective of Dynamics, Structure, & Thermodynamics." Adv. Chem. Phys. LXXI. New York: John Wiley & Sons, 1988.
2. McCammon, J.A., Straatsma, T.P. Alchemical free energy simulation. Ann. Rev. Phys. Chem. 43:407, 1992.
3. Kollman, P.A. Free energy calculations: Applications to chemical and biochemical phenomena. Chem. Rev. 93: 2395–2417, 1993.
4. Dang, L.X., Merz, K.M., Kollman, P.A. Free energy calculations on protein stability: Thr-157 → Val-157 mutation of T4 lysozyme. J. Am Chem. Soc. 111:8505–8508, 1989.
5. Prévost, M., Wodak, S.J., Tidor, B., Karplus, M. Contribution of the hydrophobic effect to protein stability: Analysis based on simulations of the Ile 96 → Ala mutation in barnase. Proc. Natl. Acad. Sci. USA, 88:10880–10884, 1991.
6. Tidor, B., Karplus, M. Simulation analysis of the stability mutant R96H of T4 lysozyme. Biochemistry 30:3217–3228, 1991.
7. Komeiji, Y., Uebayasi, M., Someya, J.-I., Yamamoto, I. Free energy perturbation study on a Trp-binding mutant (Ser 88 → Cys) of the trp-repressor. Protein Eng. 5:759–767, 1992.
8. Lau, F.T.K., Karplus, M. Molecular recognition in proteins: Simulation analysis of substrate binding by tryosyltRNA synthetase mutants. J. Mol. Biol. 236:1049–1066, 1994.
9. Gao, J., Kuczera, K., Tidor, B., Karplus, M. Hidden thermodynamics of mutant proteins: A molecular dynamics analysis. Science 244:1069–1072, 1989.
10. Jencks, W.P. "Catalysis & Enzymology." New York: McGraw Hill, 1969.
11. Fersht, A.R. Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. Biochemistry 27: 1577–1580, 1988.
12. Shi, Y.-Y., Mark, A.E., Wang, C., Huang, F., Berendsen, H.J.C., van Gunsteren, W.F. Can the stability of protein mutants be predicted by free energy calculations? Protein Eng. 6:289–295, 1993.
13. van Gunsteren, W.F., Beutler, T.C., Fraternali, F., King, P.M., Mark, A.E., Smith, P.E. Computation of free energy in practice: Choice of approximations and accuracy limiting factors. In: van Gunsteren, W.F., Weiner, P.K., Wilkinson, A.J., eds. "Computer Simulations in Biological Systems." Leiden: Escom, 1994.
14. Hermans, J. An editorial comment: The limits of simulations. Proteins 17:ii, 1993.
15. Karplus, M. Preface. In: van Gunsteren, W.F., Weiner, P.K., Wilkinson, A.J., eds. "Computer Simulations in Biological Systems." Leiden: Escom, 1994.
16. Hermans, J., Yun, R.H., Anderson, A.G. Precision free energies calculated by molecular dynamics simulations of peptides in solution. J. Comp. Chem. 13:427–442, 1992.
17. Simonson, T., Brünger, A.T. Thermodynamics of protein-peptide interactions in the ribonuclease-S system studied by molecular dynamics and free energy calculations. Biochemistry 31:8661–8674, 1992.
18. Straatsma, T.P., Zacharias, M., McCammon, J.A. Free energy difference calculations in biomolecular systems. In: Weiner, P.K., Wilkinson, A.J., eds. "Computer Simulations in Biological Systems." Leiden: Escom, 1994.
19. Hodel, A., Simonson, T., Fox, R.O., Brünger, A.T. Conformational substates and uncertainty in macromolecular free energy calculations. J. Phys. Chem. 97:3409–3417, 1993.
20. Jorgensen, W.L., Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. J. Chem. Phys. 83:3050–3054, 1985.
21. Kirkwood, J.G. Statistical mechanics of fluid mixtures. J. Chem. Phys. 3:300–313, 1935.
22. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comp. Chem. 4:187–217, 1983.
23. Weiner, S.J., Kollman, P.A., Case, D., Singh, U., Ghio, C., Alagona, G., Profeta, S., Jr., Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
24. Jorgensen, W., Swenson, C. Optimized intermolecular potential functions for amides and peptides. Structure and properties of liquid amides. J. Am. Chem. Soc. 107:569–578, 1985.
25. Brooks, C.L. III. Thermodynamics of ionic solutions: Monte Carlo simulations of aqueous chloride and bromide ions. J. Phys. Chem. 90:6684–6687, 1986.
26. Tidor, B. Simulated annealing on free energy surfaces by a combined molecular dynamics and Monte Carlo approach. J. Phys. Chem. 97:1069–1073, 1993.
27. Yu, H., Karplus, M. A thermodynamic analysis of solvation. J. Chem. Phys. 89:2366–2379, 1988.
28. McQuarrie, D.A. "Statistical Mechanics." New York: Harper & Row, 1973.
29. Tembe, B.L., McCammon, J.A. Ligand-receptor interactions. Comput. Chem. 8:281–283, 1984.
30. Hirata, F., Redfern, P., Levy, R.M. Viewing the Born model for ion hydration through a microscope. Int. J. Quant. Chem. 15:179–190, 1988.
31. Roux, B., Yu, H.-A., Karplus, M. Molecular basis for the Born model of ion solvation. J. Phys. Chem. 94:4683–4688, 1990.
32. Miyamoto, S., Kollman, P.A. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. Proteins 16:226–245, 1993.
33. Matthews, B.W. Genetic and structural analysis of the protein stability problem. Biochemistry 26:6885–6888, 1987.
34. Shortle, D., Meeker, A.K. Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. Proteins 1:81–89, 1986.
35. Weaver, L.H., Gray, T.M., Grütter, M.G., Anderson, D.E., Wozniak, J.A., Dahlquist, F.W., Matthews, B.W. High-resolution structure of the temperature-sensitive mutant of phage lysozyme, Arg 96 → His. Biochemistry 28:3793–3797, 1989.
36. Kitamura, S., Sturtevant, J.M. A scanning calorimetric study of the thermal denaturation of the lysozyme of phage T4 and the Arg96 → His mutant form thereof. Biochemistry 28:3788–3792, 1989.
37. Roux, B., Karplus, M. Ion transport in a gramicidin-like channel: Dynamics and mobility. J. Phys. Chem. 95:4856–4868, 1991.
38. Pearlman, D.A. Free energy derivatives: A new method of probing the convergence problem in free energy simulations. J. Comp. Chem. 15:105–123, 1994.
39. Fersht, A.R. Protein folding and stability: The pathway of folding of barnase. FEBS Lett. 325:5–16, 1993.
40. Caflisch, A., Karplus, M. Molecular dynamics simulation

of protein denaturation: Solvation of the hydrophobic cores and secondary structure of barnase. Proc. Natl. Acad. Sci. USA 91:1746–1750, 1994.
41. Gilson, M.K., Honig, B. Energetics of charge-charge interactions in proteins. Proteins 3:32–52, 1988.
42. Bashford, D., Karplus, M. The pk_a's of ionizable groups in proteins: Atomic details from a continuum electrostatic model. Biochemistry 29:10219–10225, 1990.
43. Davis, M.E., McCammon, J.A. Calculating electrostatic forces from grid-calculated potentials. J. Comp. Chem. 10: 386–391, 1991.
44. Yu, H.-A., Roux, B., Karplus, M. Solvation thermodynamics: An approach from analytic temperature derivatives. J. Chem. Phys. 92:5020–5033, 1990.
45. Yu, H.-A., Karplus, M., Pettitt, B.M. Aqueous solvation of N-methyl acetamide conformers: Comparison of simulations and integral equation theories. J. Am. Chem. Soc. 113:2425–2434, 1991.