PROTEINS: Structure, Function, and Genetics 20:85–97 (1994)

Intrinsic pK_as of Ionizable Residues in Proteins: An Explicit Solvent Calculation for Lysozyme

Gabriela S. Del Buono, Francisco E. Figueirido, and Ronald M. Levy
Department of Chemistry, Rutgers, The State University of New Jersey, New Brunswick, New Jersey 08903

ABSTRACT **Molecular dynamics simulations of triclinic hen egg white lysozyme in aqueous solution were performed to calculate the intrinsic pK_as of 14 ionizable residues. An all-atom model was used for both solvent and solute, and a single 180 ps simulation in conjunction with a Gaussian fluctuation analysis method was used. An advantage of the Gaussian fluctuation method is that it only requires a single simulation of the system in a reference state to calculate all the pK_as in the protein, in contrast to multiple simulations for the free energy perturbation method. pK_{int} shifts with respect to reference titratable residues were evaluated and compared to results obtained using a finite difference Poisson–Boltzmann (FDPB) method with a continuum solvent model; overall agreement with the direction of the shifts was generally observed, though the magnitude of the shifts was typically larger with the explicit solvent model. The contribution of the first solvation shell to the total charging free energies of the titratable groups was explicitly evaluated and found to be significant. Dielectric shielding between pairs of titratable groups was examined and found to be smaller than expected. The effect of the approximations used to treat the long-range interactions on the pK_{int} shifts is discussed.** © 1994 Wiley-Liss, Inc.

Key words: electrostatics, titration curves, solvation, linear response theory, free energy

INTRODUCTION

It is well known that the function and stability of a protein in solution can be greatly affected by the pH; examples include the acid denaturation of proteins, and the variation of activity of many enzymes with the solution pH. For this reason, the accurate determination of the pK_as of charged amino acid residues, and hence an understanding of the electrostatic forces within proteins in solution are a subject of great interest for both biophysical theorists and experimentalists.

The titration curve of a native protein can differ substantially from the sum of the unperturbed titrations of the constituent acidic and basic groups; therefore, one of the main motivations in calculating pK_a values is to predict how the protein environment shifts the pK_a of an amino acid in the protein from that of the isolated amino acid in solution. In early treatments, Kirkwood and Tanford introduced the concept of self-energy terms in the ionized groups of a protein in solution.[1,2] Recently, there has been renewed interest in these terms and their impact in the calculation of titration curves; the current belief is that in addition to pairwise interactions between charged groups, there is a balance between the cost in energy of desolvating the ionizable residues, and their interaction with the remaining neutral, but polar residues. These factors can make significant contributions to the free energy and must be considered in any accurate treatment of pK_as.

In general, the inclusion of solvent effects is known to be essential for a realistic description of structural and dynamic features of biomolecules. The proper description of solvent–solute interactions in biopolymers in water has received increased attention lately, and several approaches have been developed to accurately model the effect of the solvent on the electrostatic potential of polypeptides or proteins.[3–14] One class of methods includes macroscopic models, in which the solvent is approximated by a dielectric continuum.[12,14–17] These methods have received wide acceptance and have led theoreticians to develop this approach further. Several research groups have recently applied the finite-difference Poisson–Boltzmann (FDPB) method using a continuum solvent model to protein pK_a calculations and have obtained reasonably good results.[3,7,12,18] A major drawback encountered in the application of this type of calculation has been the fact that the calculations are based on the protein's crystal-structure conformation which may deviate from the solution conformation.[3,4] More generally, conformational heterogenity and dynamics are not accounted for in a natural way within the Poission–Boltzmann framework.

A second class of methods includes discrete descriptions of solvation, such as the "protein-dipoles Langevin-dipoles" (PDLD) approach,[19] where the

Received March 15, 1994; revision accepted May 20, 1994.
Address reprint requests to Dr. Ronald M. Levy, Department of Chemistry, Rutgers, The State University of New Jersey, New Brunswick, NJ 08903.

solvent is treated by placing point dipoles on a lattice around the protein, and the average orientation of these dipoles is approximated by a Langevin-type function. Another important approach for incorporating the solvent is to use a detailed microscopic model which explicitly includes the water molecules.[5,20,21] In this representation the electrostatic interactions are governed by Coulomb's law; currently used water models involve pairwise potential functions (e.g., SPC[22] or TIP4P[23]) that produce reasonable bulk dielectric constants for water,[24,25] and yield fairly good representations of the solvent hydrogen bonding.[23,26]

Studies of solvation energies employing all-atom solvent models have experienced a considerable boost since the adaptation of free-energy perturbation (FEP) methods to solvation problems.[27,28] The convergence of the calculations has been satisfactory in many cases, and the calculations have yielded reasonable solvation free energies.

The use of all-atom solvent models in conjunction with FEP for the calculation of solvation energies and pK_as becomes far more difficult when applied to proteins in solution because of problems inherent in treating charge changes with explicit solvent models and because of the much larger size of the system. The high computational effort has long prohibited extensive all-atom FEP simulations which would provide a microscopic description of the charging process. Work by Warshel and co-workers on BPTI,[19,29,30] and by Merz on HEWL and HCAII[10] are among the first attempts to include explicit solvent around particular ionizable groups combined with the FEP method to compute free energy differences upon ionization. The problems associated with the evaluation of long-range electrostatic interactions for these types of calculations are examined to some extent by Lee and Warshel.[30] Their work currently does not include any pH dependence, and only a few hundred water molecules surrounding the particular sites are included in the calculation.

In this work, we employ molecular dynamics simulations (MD) using an all-atom solvent model, in conjunction with a Gaussian fluctuation (linear response) analysis method[31] to evaluate the free-energy changes of ionizable groups in a protein due to electrostatics.

An advantage of the Gaussian fluctuation method for calculating pK_as is that only one MD simulation of the system is required to evaluate all the pK_as in a protein, and hence a microscopic treatment is computationally feasible. The method has been tested for smaller systems (e.g., two simple ions, and methylindole) in solution,[31,32] and has yielded very good results when compared with full FEP simulations of the corresponding charging process. Another attractive feature of the Gaussian fluctuation method for calculating pK_as is that it is derived from a linear response theory. Not only has linear response theory been demonstrated to provide an accurate description of many condensed phase electrostatics problems (and indeed the linear PB model is a linear response model),[33,34] but we expect that the linear response predictions will be less sensitive to the exact parameterization of the molecular mechanics force field than an explicit calculation of the ionization free energies. Given the uncertainties in the parameterization of the force fields, one can argue that the use of higher order theory for calculating pK_a shifts is unwarranted, particularly if the focus is on general trends. An MD simulation of the protein in solution should provide a more realistic description of the protein-solvent system than the crystal structure.

To establish the feasibility of the method, we focus here on calculating the intrinsic pK_as of the titratable groups; the intrinsic pK_a of a group involves only the interactions of the charged residue with the solvent and the remaining neutral, polar groups and excludes electrostatic interactions with other charged residues. We compare and contrast our approach to the method developed by Bashford and Karplus[12] in their study of lysozyme in solution. The different contributions to the ionization free-energy difference are analyzed, and the dielectric response between different titratable groups is estimated. We also test the effect of the truncation of long-range electrostatic interactions in our calculations.

METHODS

pK_{int} Calculation

The intrinsic pK_a of an ionizable group i, p$K_{int,i}$, is defined as the pK_a the group would have if all other ionizable groups were neutral. p$K_{int,i}$ is given by the following equation:

$$pK_{int,i} = pK^0_{a,i} + \frac{1}{2.303k_BT}[\Delta G_i - \Delta G^0_i] \quad (1)$$

p$K^0_{a,i}$ is the known pK_a of a model compound in solution; ΔG_i and ΔG^0_i are the change in electrostatic energy for charging the ionizable residue in the protein and the model compound, respectively. We chose titratable groups within the protein as our references or model compounds. For each titratable residue type we chose a particular residue from the protein as model compound for that residue class, thereby limiting the calculation of pK_{int} to only those residues that were present more than once in the protein (i.e., Glu, Asp, Lys, and Tyr). Table I lists the amino acid residues used as model compounds and their respective p$K^0_{a,i}$s. The reference residues were chosen based on the criteria that the intrinsic pK_as for these residues, calculated by the FDPB method,[12] were close to the corresponding pK_a for the isolated amino acid.

The free energy change in charging the titratable groups was calculated using the Gaussian fluctua-

TABLE I. Amino Acid Residues in Lysozyme Used as Model Compounds

Residue type	Model compound	$pK^{0}_{a,i}$*	pK_{mod}*
Asp	Asp-18	3.5	4.0
Glu	Glu-7	2.7	4.4
Lys	Lys-97	10.0	10.4
Tyr	Tyr-23	11.2	9.6

*See Ref. 12.

tion formula derived by Levy et al.[31] The formula is derived in the framework of linear response theory and is based on the assumption that the probability distribution function of the fluctuations in the total electrostatic interaction energy between the solute and the solvent (in our case: water + nontitratable groups) is Gaussian. The formula has been applied to the calculation of the free-energy change upon charging two simple ions in aqueous solution[31]; the results agreed well with corresponding free-energy calculations using the thermodynamic integration simulation method.

Under the Gaussian fluctuation approach, the expression obtained for the total free-energy change of going from a reference state to a desired final state is the following:

$$\Delta G = \sum_i \overline{V}_i \Delta q_i - \frac{\beta}{2} \sum_{i,j} \langle \Delta V_i \Delta V_j \rangle \Delta q_i \Delta q_j \,. \qquad (2)$$

In the first term, $\overline{V}_i$ is the average electrostatic potential due to the solvent (in our case, solvent + nontitratable groups) at the ith titratable site, and Δq_i is $+1$ or -1, depending on the type of residue. This term represents the difference in charge distributions between the initial and final state of the solute interacting with the average electrostatic potential of the reference system. In the second term, $\langle \Delta V_i \Delta V_j \rangle$ is the correlation function for the joint fluctuations in the electrostatic potentials at sites i and j averaged over the reference system, where $\Delta V_j = V_j - \overline{V}_j$. This term represents the induction effect caused by the solvent polarization, which in turn is due to the change in solute charge distribution. More details of the derivation are presented in the Appendix.

For the calculation of pK_{int}, only the interactions with the solvent and other nontitratable groups are needed, and therefore the cross-correlation functions involving pairs of different titratable groups, $\langle \Delta V_i \Delta V_j \rangle$ do not contribute. These terms are included in the calculation of the dielectric shielding in a later section. The cross-correlation functions are also important in the estimates of effective pK_as, a subject that will be addressed in future work.

The final expression used to calculate the free-energy difference at each titratable site was the following,

$$\Delta G_i = \overline{V}_i \Delta q_i - \frac{\beta}{2} \langle \Delta V_i^2 \rangle \Delta q_i^2 \qquad (3)$$

where $\overline{V}_i$, and $\langle \Delta V_i^2 \rangle$ can be obtained directly from a molecular dynamics simulation. From knowledge of ΔG_i, pK_{int} is calculated from Eq. (1).

Computer Simulation

Molecular dynamics simulations were carried out on triclinic hen egg white lysozyme in water using the IMPACT computer program.[35] The molecular mechanics parameters used for the protein were those published by Weiner et al.[36] All hydrogens were explicitly included, and the SPC water model was used for the solvent.[22] The equations of motion were integrated using Andersen's RATTLE algorithm with a time step of 1 fs.[35,37]

The initial coordinates of the protein were obtained from the Brookhaven Protein Data Bank.[38-40] The final system consisted of the enzyme (1980 atoms) and 4580 water molecules (~3 layers of solvent) in a 52.0 × 49.6 × 61.8 Å box.

The reference system was chosen with the following properties:

- In order to make a more reliable comparison to the results of Bashford and Karplus the charges on the ionizable groups were distributed as in Ref. 12. Thus the formal charge on the titrating groups was localized on a single atom: the carboxyl carbon atom for Asp, Glu, and the C-terminus, the amino nitrogen atom of Lys and the N-terminus, the N_{ϵ_2} atom for His, and the hydroxyl oxygen atom for Tyr. This approach has shown negligible deviations from the case where the charge is distributed over several atoms.[3,12]
- The arginine side chains were treated as part of the nontitrating background also following Bashford and Karplus. The justification for this treatment is that arginine residues titrate beyond the pH range in which lysozyme is stable.
- Under the Gaussian fluctuation method,[31] the reference system can be either the initial state, an intermediate state, or the final state of interest, and may be chosen for computational convenience. Since counterions would have to be present in solution to balance any net charge on the protein, and this in turn would require a significantly larger system, the charge state of the reference was chosen such that all physiologically charged lysozyme groups were neutral.
- The Gaussian fluctuation formula is derived under the assumption that the distances between ionizable sites are constant (see Appendix). Consequently, the distances between the atoms where the charge is localized in the titratable groups were kept fixed at their crystal structure value using SHAKE.[41] This con-

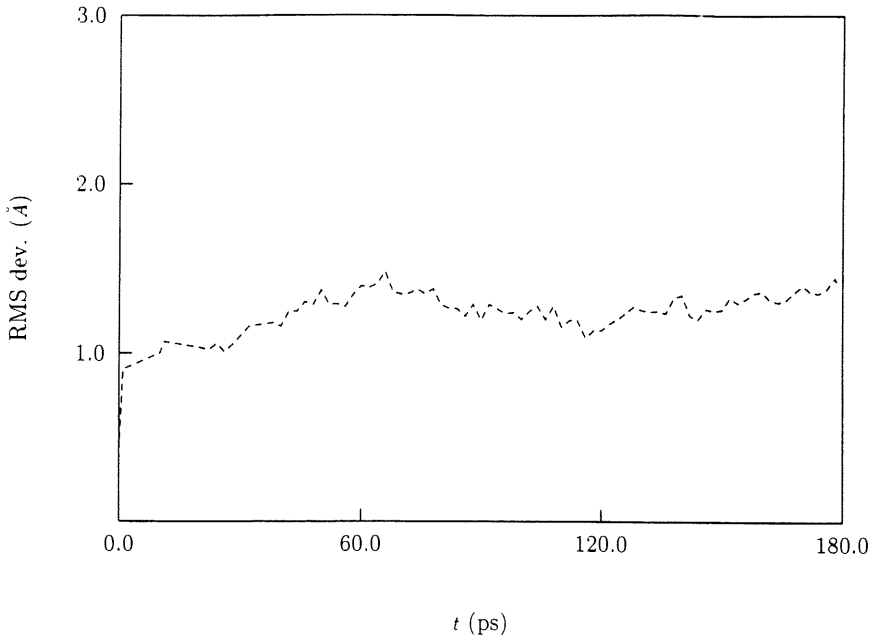

Fig. 1. Rms deviation of the backbone atoms of lysozyme over the simulation time (equilibration + production time).

straint is less restrictive than what is required in the Poisson–Boltzmann framework, where the entire protein is frozen. In our case, this treatment can also be expected to be helpful since by choosing a neutral state of the protein as reference (not the native state of a protein in solution), an unconstrained simulation might lead to unfolding of the protein. This possibility was tested by plotting the rms deviation of the backbone atoms from the crystal structure (see Fig. 1), where we can see that the problem of unfolding does not arise.

The temperature was kept at 298 K using Berendsen's scaling method with a relaxation time of 0.01 ps.[42] A nonbonded interaction list was regenerated every 5 fs using a 9 Å cutoff. An atom based cutoff was used for the protein–protein interactions, and molecule based for the water–water interactions. For the protein–water interactions, the cutoff procedure was atom based on the protein and molecule based on the water. Although a residue cutoff is normally applied for the protein interactions, we used an atom based cutoff to reduce the computational time; for our particular dynamics simulation this is justified due to the restraining effect resulting from the constraints applied to the protein. Since energetic quantities are sensitive to the cutoff procedure the analysis of the trajectory was performed using an infinite cutoff for the intraprotein interactions.

After 1,000 steps of minimization using a conjugate gradient method, the reference system was equilibrated for 30 ps, and a 150 ps production run was completed. Coordinates were saved for subsequent analysis every 100 fs.

RESULTS AND DISCUSSION

Intrinsic pK_a Shifts

In this section, we evaluate the shift of the intrinsic pK_as of ionizable groups in lysozyme with respect to a reference residue (see Table I), and compare them to results obtained by Bashford and Karplus.[12] Root-mean-square (rms) deviations of the main chain atoms (residues 3 to 126) as a function of time are also examined, and the results are shown in Figure 1. The graph illustrates that there is no large conformational variation of the protein structure with respect to the crystal structure over the length of our simulation.

Using Eq. (3), we calculated the running average of ΔG over the simulation time for the different titratable groups in order to examine the convergence properties of the simulation. For most of the residues the convergence is good; Figure 2 shows results for some of these residues. The intrinsic pK_a shifts of 18 of the 21 titratable residues are shown in Table II. The shifts were displaced such that ΔpK_{int} for the model compounds we used would coincide with Bashford and Karplus' ΔpK_{int}, in order to facilitate comparison[12] (the displacements are -0.4 for Lys, -0.5 for Asp, -1.7 for Glu, and 1.6 for Tyr). Four of those groups are used as model compounds and even though they are shown in the table, they are not included in the evaluation of the method.

From the table, 10 out of 14 groups shift in the

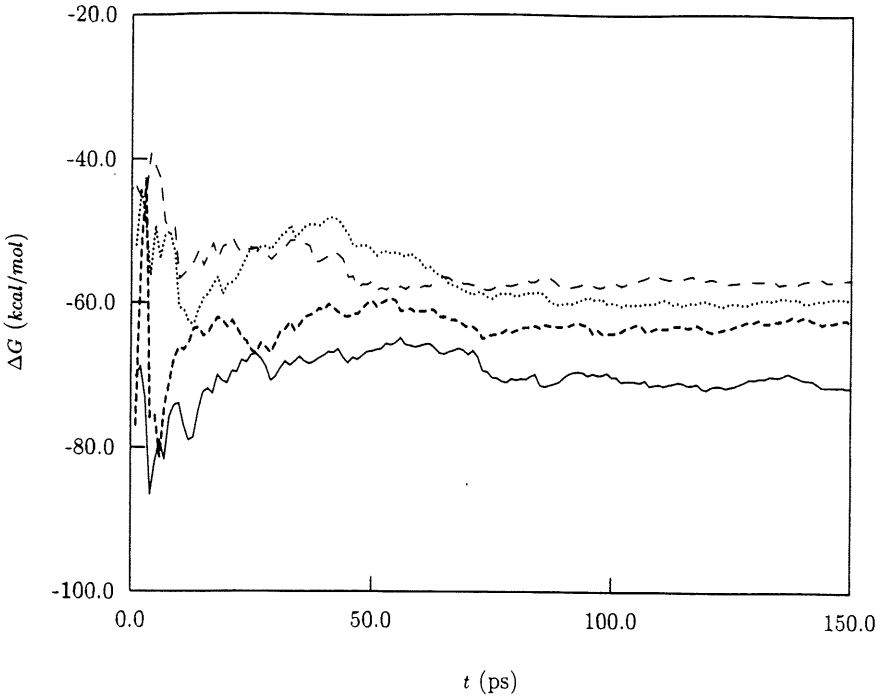

Fig. 2. ΔG running average over the simulation. Solid line, Lys-1; short-dashed line, Asp-119; dotted line, Tyr-53; long-dashed line, Glu-7.

TABLE II. Intrinsic pK_a Shifts: FDPB* vs. Explicit Solvent (ES)

Residue number	ΔpK_{int} (FDPB*)	ΔpK_{int} (9 Å) (ES)	ΔpK_{int} (15 Å) (ES)	ΔpK_{int} (HM†) (ES)
Lys-1	−1.5	−7.4	−6.4	−5.4
Lys-13	−0.8	−14.5	−13.3	−15.7
Lys-33	−1.3	−2.2	−4.2	−6.5
Lys-96	−0.3	−9.9	−9.3	−8.6
Lys-97	−0.4	−0.4	−0.4	−0.4
Lys-116	−0.7	6.9	6.3	0.9
Asp-18	−0.5	−0.5	−0.5	−0.5
Asp-48	−2.9	−6.7	−6.6	−13.2
Asp-52	1.6	−4.2	−8.8	−13.0
Asp-66	−2.3	−3.6	−3.2	−0.4
Asp-87	−1.3	−7.1	−6.8	−11.1
Asp-101	4.3	14.0	9.7	9.4
Asp-119	−0.7	1.0	−1.8	−4.6
Glu-7	−1.7	−1.7	−1.7	−1.7
Glu-35	1.6	−3.2	4.5	9.0
Tyr-20	3.6	21.7	22.8	26.0
Tyr-23	1.6	1.6	1.6	1.6
Tyr-53	2.7	20.6	24.5	29.6

*See Ref. 12.
†Hierarchical method (see "Intrinsic pK_a Shifts").

same direction as the FDPB calculation, even though in most cases the magnitude of the shift is found to be larger than the FDPB result. It has been observed that when the solvent is represented as a dielectric continuum, the magnitude of the pK_a shift is underestimated, and when only a few waters of coordination are included explicitly the shift is overestimated.[4] Moreover, the validity of the quadratic charge dependence of the free energy of a system, predicted by continuum solvent models, is still under scrutiny, especially for systems where the full charging process involving multiple sites is studied.[43] A consistent treatment must include several solvation shells and treat the Coulomb interactions

in a physically consistent way. It is becoming increasingly clear that truncation affects considerably the structure and thermodynamics of the solvent[24,34,44,45] as discussed below.

As a quick test of the effect of the value chosen for the long range cutoff on the results, we recalculated the intrinsic pK_as using a 15 Å cutoff for the protein–water and water–water interactions using the same trajectories obtained with a 9 Å cutoff, and we observed if there were significant changes. Table II shows the results; for several residues there are significant differences between the pK_{int} shifts calculated with the 9 Å cutoff and the 15 Å cutoff. For 10 out of 14 groups the magnitude of the shift is diminished with respect to the calculation based on a 9 Å cutoff. The largest discrepancy is observed for Asp-101 for which ΔpK_{int} is 14 when calculated including all solvent molecules within 9 Å and is reduced to 9.7 when all solvent molecules within 15 Å are included. For the system studied here, using a 9 Å cutoff for the interactions, the number of nonbonded interactions that must be evaluated per step is approximately 2 million. We estimate that if the minimum image approach was employed as our method of truncation (i.e., interactions of the solute with all the solvent molecules in the system are included), the number would increase by approximately a factor of 60.

In MD simultaions of proteins with explicit solvent it is usual to include just three layers of solvent and to truncate the interactions at a relatively short cutoff distance to render the calculation computationally tractable. This approach is dangerous since there is already evidence that truncation may lead to incorrect physical interpretations and alternative methods such as the Ewald sum have proven to be more reliable.[34] Several research groups are currently updating their codes to use more modern approaches,[44,46,47] some of which are based on the multipole approximation[48–53] which does not make use of a cutoff for the long-range forces. Our laboratory is working on the implementation of a hierarchical method, based on the algorithm developed by Barnes and Hut,[49] to calculate the nonbonding interactions. Although this work is near completion we do not yet have all the pieces integrated; however, at this point we were able to use this new code to reanalyze the data obtained from our 9 Å cutoff trajectories recomputing the electrostatic potentials without the use of a cutoff. In this way we can further test the effect of long-range interactions on the pK_{int} shifts of ionizable residues in lysozyme. The results are presented in Table II and changes in the pK_{int} shifts from the 15 Å case are observed. This provides further evidence of the sensitivity of the free energies of ionization to the way long-range interactions are calculated, and points to the need for systematic studies of the effects of system size on charging free energies.

Different Contributions to the Free Energy Difference

In this section we examine the contribution to the charging free energy, ΔG, from the average potential and from induction effects, and analyze the relative significance and importance of each term with respect to the final result. We also compare the contribution from the first solvation shell to the more distant waters and the protein background term.

The contributions to the free-energy difference from the first and second terms of Eq. (3) (see the section pK_{int} Calculation) are presented in Table III. The results show that the induction effects associated with the change in solvent polarization and reflected in the correlation functions $\langle \Delta V_i^2 \rangle$ dominate the free-energy change with respect to the first term in Eq. (3). Thus, the charging free energy is dominated by induction effects. However, ΔpK_{int}s are proportional to differences between ΔGs (see Equation 1), hence the induction effect will not necessarily dominate these shifts.

By looking separately at the average electrostatic potential at the titratable sites due to the solvent and the background (rest of the nontitratable groups in the protein) in Table III, we also observe that in most cases the two terms tend to compensate each other; analogous results have been obtained by other groups.[12] For example, there is a −16.4 kcal/mol attractive contribution to the ionization of Asp-87 due to the average electrostatic potential of the hydrating water which is nearly cancelled by a repulsive contribution of 20.2 kcal/mol from the polar groups in the protein. This effect illustrates the balance between the solvation effect to which the ionizable residues in the protein are subjected and their interaction with the remaining nontitratable residues.

The study of the contributions of the background term and the solvent to the total correlation function $\langle \Delta V_i^2 \rangle$ is less straightforward; $\langle \Delta V_i^2 \rangle$ also involves a cross-correlation function according to the following relation:

$$\frac{\beta}{2}\langle \Delta V_i^2 \rangle = \frac{\beta}{2}[\langle \Delta V_i^2 \rangle_{solv} + \langle \Delta V_i^2 \rangle_{back} + 2\langle \Delta V_{i,solv} \Delta V_{i,back} \rangle]. \quad (4)$$

Table IV shows the different terms involved in the total correlation function. Overall, the correlation function due to the interaction of the ionizable groups with the solvent dominates; the solvent-background cross-correlation functions are negative, and the correlation functions due to the 'background' are close in magnitude to the cross-correlation functions but of opposite sign. The negative correlation can be understood qualitatively by the following argument. Imagine that the protein is replaced, as in the continuum model, by a region of dielectric $\epsilon_i < \epsilon_w$,

TABLE III. Different Terms Contributing to the Free Energy of Ionizable Groups in Lysozyme [see Eq. (3)] (energies in kcal/mol)

Residue number	$\langle V_i \rangle_{solv}$	$\langle V_i \rangle_{back}$	$\langle V_i \rangle$	$\frac{\beta}{2}\langle \Delta V_i^2 \rangle$	ΔG_i
N-term	−0.004	35.6	35.6	80.3	−44.7
Lys-1	−1.25	2.29	1.04	72.6	−71.5
Lys-13	−6.21	10.6	4.41	66.2	−61.8
Lys-33	0.79	−9.00	−8.21	70.4	−78.6
Lys-96	3.08	−2.90	0.18	68.2	−68.0
Lys-97	−4.83	−9.10	−13.9	67.1	−81.0
Lys-116	−3.92	−10.2	−14.1	76.8	−90.9
Asp-18	−15.5	29.1	13.6	51.0	−64.6
Asp-48	−0.21	−1.69	−1.90	74.9	−73.0
Asp-52	0.49	3.20	3.69	65.9	−69.6
Asp-66	−0.20	6.35	6.15	62.6	−68.8
Asp-87	−16.4	20.2	3.80	69.8	−73.6
Asp-101	−5.34	−8.60	−13.9	58.7	−44.8
Asp-119	−19.5	26.8	7.30	55.4	−62.7
Glu-7	−15.9	18.8	2.90	53.7	−56.6
Glu-35	−2.34	3.28	0.94	57.7	−58.6
His-15	−2.21	7.89	5.68	83.4	−77.7
Tyr-20	−6.83	−3.28	−10.1	68.4	−58.3
Tyr-23	−6.24	6.78	0.54	85.2	−85.7
Tyr-53	−3.56	5.96	2.40	57.4	−59.8
C-term	−4.41	10.5	6.09	54.5	−60.6

where ϵ_w is the dielectric constant of bulk water. Place a charge q in this inner region, close to the boundary. If the boundary were just a plane, the field produced by the induced polarization of the solvent could be represented, inside this region, by a charge of opposite sign located within the solvent. At any point within the inner region, then, the potential produced by the charge q and that produced by the solvent have opposite signs, i.e., they are negatively correlated. In the general case of a more complicated geometry it is reasonable to expect that if the charge q is close to the boundary the potentials would show the same negative correlation.

Next, we examined the contribution of the first shell of solvent to the total free-energy difference. We isolated the shell of water molecules using a distance criterion; the distance was chosen based on the distribution function of water around a protein calculated by Komeiji et al.[54] They show two different distributions; one around polar groups, with the first minimum at approximately 3.8 Å (Fig. 3b, Ref. 54), and another around apolar groups, with the first minimum appearing at approximately 5 Å (Fig. 3c, Ref. 54). Our preliminary results using either shell size were not very different; we chose a distance criterion of 5 Å for our solvation shell.

Table V shows average interaction potentials, correlation functions, and cross-correlation functions computed from the first shell of solvent, as well as the values using the entire solvent system. These results show that the first solvation shell contributes heavily to the total solvent-background cross-correlation function, as well as a large induction effect [$\langle \Delta V_i^2 \rangle_{solv}$(1st shell)], suggesting that this shell does not behave like a low dielectric region.

Effective Dielectric Response

It is common practice when calculating electrostatic energies in proteins using a macroscopic approach to assign to the molecule an internal dielectric constant of the order of 2 to 4; this type of model assumes that the dielectric region within a protein is homogeneous. Various studies concerning charge screening in proteins have also been undertaken by several other research groups.[5,55–57]

Given the quadratic dependence of ΔG on the charges in the Gaussian fluctuation approximation, we can define a collection of "apparent" dielectric constants, ϵ_{ij}, that would include the effect of the surroundings on each pair of residues i and j, by the following expressions,

$$\sum_{i \neq j} q_i q_j \left(\frac{1}{r_{ij}} - \beta \langle \Delta V_i \Delta V_j \rangle \right) = \sum_{i \neq j} \frac{q_i q_j}{\epsilon_{ij} r_{ij}}, \quad (5)$$

$$\epsilon_{ij} = [1 - \beta \langle \Delta V_i \Delta V_j \rangle r_{ij}]^{-1}. \quad (6)$$

Figure 3 shows a plot of the computed ϵ_{ij} as a function of r_{ij}, the distance between two titratable groups. Due to the periodicity of the system the longest physical distance between two residues cannot exceed about 20 Å. In fact, we have taken into account only distances up to 15 Å because of the spurious correlations that appear at longer distances. These correlations arise because water molecules located near pairs of residues that are separated by

TABLE IV. Different Contributions to the Correlation Function Term, $\frac{\beta}{2}\langle \Delta V_i^2 \rangle$ [see Eq. (4)] (Energies in kcal/mol)

Residue number	$\frac{\beta}{2}\langle \Delta V_i^2 \rangle_{solv}$	$\frac{\beta}{2}\langle \Delta V_i^2 \rangle_{back}$	$2\frac{\beta}{2}\langle \Delta V_{i,solv} \Delta V_{i,back} \rangle$	$\frac{\beta}{2}\langle \Delta V_i^2 \rangle$
N-term	93.8	39.3	−52.8	80.3
Lys-1	76.2	15.5	−19.2	72.6
Lys-13	61.9	4.07	0.201	66.2
Lys-33	68.8	13.9	−12.4	70.4
Lys-96	74.4	17.2	−23.4	68.2
Lys-97	70.8	13.2	−16.9	67.1
Lys-116	82.1	11.4	−16.7	76.8
Asp-18	49.3	14.5	−12.8	51.0
Asp-48	76.9	49.1	−51.1	74.9
Asp-52	59.2	32.3	−25.6	65.9
Asp-66	36.0	36.7	−10.1	62.6
Asp-87	66.6	33.6	−30.5	69.8
Asp-101	61.5	9.84	−12.6	58.7
Asp-119	71.1	26.2	−41.8	55.4
Glu-7	72.0	31.8	−50.2	53.7
Glu-35	44.2	24.8	−11.3	57.7
His-15	64.4	52.3	−33.3	83.4
Tyr-20	69.9	18.4	−20.0	68.4
Tyr-23	97.8	49.0	−61.6	85.2
Tyr-53	53.5	28.9	−25.2	57.4
C-term	53.6	4.38	−3.50	54.5

TABLE V. Contribution to the Free Energy Terms From the First Solvation Shell (Energies in kcal/mol)

Residue number	$\langle V_{i,solv} \rangle$ (1st shell)	$\frac{\beta}{2}\langle \Delta V_{i,solv}^2 \rangle$ (1st shell)	$2\frac{\beta}{2}\langle \Delta V_{i,solv} \Delta V_{i,back} \rangle$ (1st shell)	$\langle V_{i,solv} \rangle$ (total)	$\frac{\beta}{2}\langle \Delta V_{i,solv}^2 \rangle$ (total)	$2\frac{\beta}{2}\langle \Delta V_{i,solv} \Delta V_{i,back} \rangle$ (total)
N-term	−1.95	125.9	−50.4	−0.004	93.8	−52.8
Lys-1	−4.32	121.3	−27.4	−1.25	76.2	−19.2
Lys-13	−10.1	99.0	−2.71	−6.21	61.9	0.20
Lys-33	−0.81	108.4	−10.2	0.79	68.8	−12.4
Lys-96	−0.19	111.1	−24.4	3.08	74.4	−23.4
Lys-97	−4.90	117.4	−17.3	−4.83	70.8	−16.9
Lys-116	−1.03	119.6	−13.6	−3.92	82.2	−16.7
Asp-18	−20.6	99.0	−21.7	−15.5	49.3	−12.8
Asp-48	−3.21	113.7	−53.9	−0.21	76.9	−51.1
Asp-52	−2.51	89.5	−26.9	0.49	59.3	−25.6
Asp-66	−3.42	63.4	−21.3	−0.20	36.0	−10.1
Asp-87	−18.1	99.6	−28.6	−16.4	66.6	−30.5
Asp-101	−4.17	102.7	−13.9	−5.34	61.5	−12.6
Asp-119	−20.0	95.9	−35.4	−19.5	71.1	−41.8
Glu-7	−19.7	122.5	−47.5	−15.9	72.0	−50.2
Glu-35	−3.36	68.5	−16.3	−2.34	44.2	−11.3
His-15	−4.43	99.3	−32.6	−2.21	64.4	−33.3
Tyr-20	−6.85	105.2	−25.2	−6.83	69.9	−20.0
Tyr-23	−4.49	133.3	−43.7	−6.24	97.8	−61.6
Tyr-53	−4.93	90.7	−28.0	−3.56	53.6	−25.2
C-term	−5.94	94.1	−4.28	−4.41	53.6	−3.50

half the box (about 20 Å) become effectively closer when periodic boundary conditions are imposed; thus, they could either directly interact with each other (provided the cutoff is much smaller than half the box), or be correlated due to interactions with water molecules located in between them. This effect can be important since the orientational correlation length is a significant fraction of the length of our simulation box.[58]

The shielding observed is much smaller than ex-

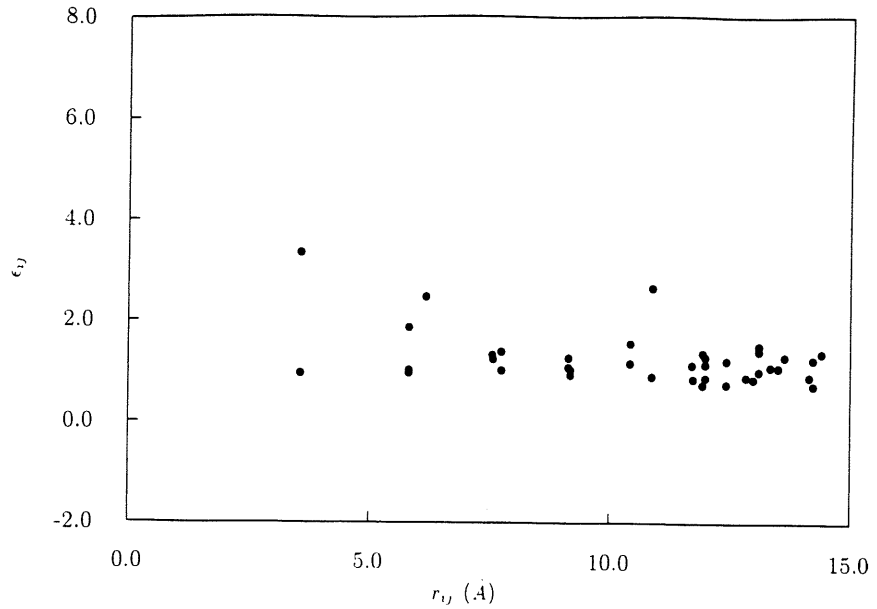

Fig. 3. Distance dependence of point-to-point "apparent" dielectric constant.

pected, and the ϵ_{ij}s appear to fluctuate around a constant value. The trend expected would be for ϵ_{ij}s to asymptotically approach a larger value at larger r_{ij}s due to the promixity of the residues to the surface of the protein, hence to the solvent, and to reach a smaller value at smaller r_{ij}s due to dielectric saturation. Correspondingly, we would expect larger dispersion in the values of ϵ_{ij} at smaller distances, due to the different environments each pair of residues can experience (i.e., both residues in the interior of the protein, or else near the surface); as the distances get larger, the pairs of residues would tend to get closer to the surface, reducing the dispersion in the values of ϵ_{ij}s. Our results in Figure 3 agree with the latter expectations.

We also estimated the error involved in the cross-correlation functions from the standard deviation of $\langle \Delta V_i \Delta V_j \rangle$ obtained for five intervals of the MD simulation, and incorporated this error into the calculation of ϵ. Figure 4 shows separately $\epsilon + \Delta\epsilon$, and $\epsilon - \Delta\epsilon$. Values of ϵ are now more scattered and fluctuate between 1 and 4, closer to the commonly accepted dielectric constant values of the interior of a protein. It is important to emphasize the high sensitivity of ϵ to changes in $\langle \Delta V_i \Delta V_j \rangle$; in order to illustrate this effect we show in Figure 5 the relation between ϵ and the site–site cross-correlation function at a constant distance of 10 Å. The most interesting region is the one with values of ϵ greater than or equal to 1 (i.e., positive cross-correlation function); in this region a change of $\langle \Delta V_i \Delta V_j \rangle$ from 10 to 19 would cause a jump in the "apparent" dielectric constant from 2 to around 30.

Reasons for the relatively small ϵ_{ij} values obtained here can be traced to the particular treatment of long-range interactions; a similar behavior was observed in other studies of systems of two ions in solution.[59] A more systematic study of the effect of the dielectric response on the free energy of ionization of charged particles in aqueous solution is currently underway in our laboratory.

CONCLUSION

Studies of intrinsic pK_a shifts of ionizable groups in hen egg white lysozyme including explicit solvent have been performed using molecular dynamics simulations in conjunction with a Gaussian fluctuation method. An advantage of the Gaussian fluctuation method is that it only requires a single simulation of the system in a reference state to calculate all the pK_as in the protein, in contrast to multiple simulations for the perturbation method (FEP). Overall agreement with the direction of the pK_{int} shifts obtained by Bashford and Karplus using a continuum solvent model was observed,[12] even though the magnitudes of the shifts were found in general to be larger than the Poisson–Boltzmann predictions.

Recalculation of the pK_as using a larger cutoff for water–water and water–protein interactions (protein–protein interactions were already analyzed using infinite cutoff in all cases, see "Computer Simulation") gave preliminary evidence that long-range interactions are an important factor in our calculations, affecting the large magnitudes of the pK_a shifts. This suggests that the implementation of a better treatment of such long-range forces is a necessity, and should be addressed before continuing on to the calculation of effective pK_as.

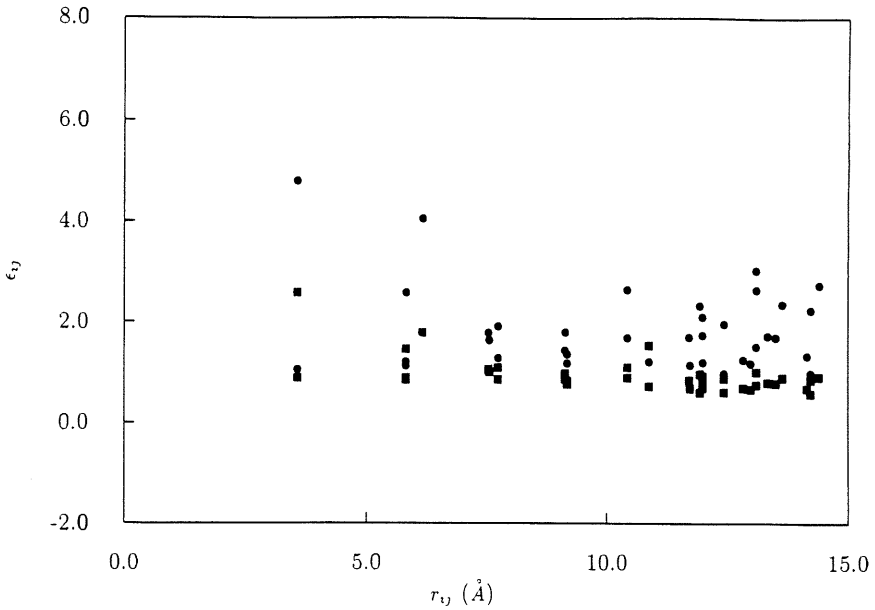

Fig. 4. "Apparent" dielectric constant including errors in cross-correlation functions (see the section Effective Dielectric Response); dots, $\epsilon + \Delta\epsilon$; squares, $\epsilon - \Delta\epsilon$.

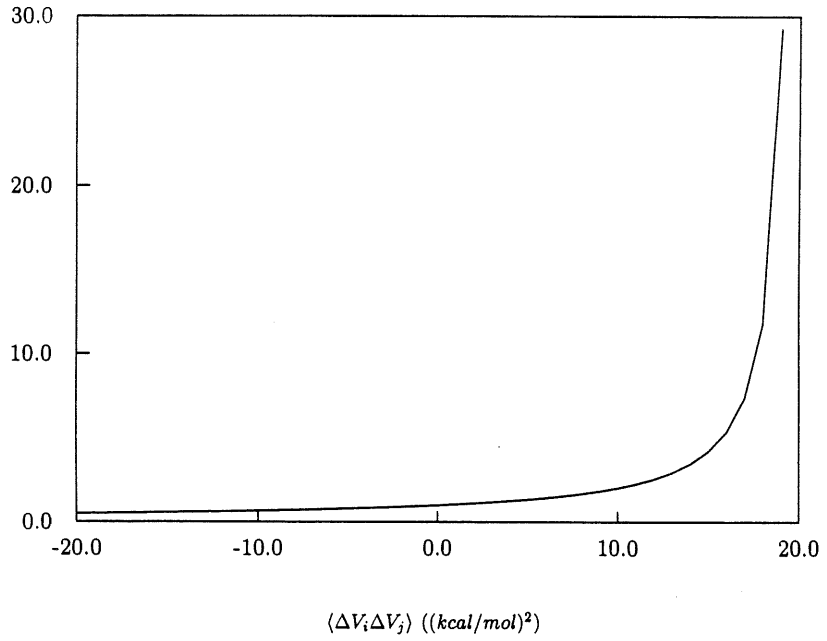

Fig. 5. Dependence of ϵ with respect to the cross-correlation functions at a constant distance between ionizable groups; r_{ij} = 10 Å.

Another factor that might affect the accuracy of results is the magnitude of the Δqs used in the Gaussian fluctuation formula. In previous studies, it was clear that the equation gave optimum results when a reference state was chosen so that Δq was 0.5 charge units or less.[31] We are planning additional studies to optimize the choice of reference state for calculation of protein pK_as from explicit solvent simulations.

The importance of explicitly including the solvent in studies of biological systems in solution is not to be underestimated; there is evidence that the hydrating waters surrounding a molecule exhibit different behavior from the bulk water[9,60–62]; for our

particular study concerning ionization properties of different groups in a protein, the effect of the water solvating these charged groups is of even greater significance, and it would be advantageous to have these special properties of the surface waters come directly from the model used. Other research groups have also emphasized the importance of the use of explicit solvent when performing pK_a calculations.[4,10,19] The Gaussian fluctuation method is relatively fast and easy to implement; once the accuracy of charging free energy simulations is improved—mostly by improving the treatment of long-range interactions—the method has the potential to become a powerful tool for developing reduced descriptions of complex biological systems. The behavior of the proteins' charged groups plays a critical role in the structure and function of enzymes, and the calculation of pK_a shifts is an important step in understanding the microscopic basis of more complicated processes, such as catalysis in biological reactions. We are currently refining methods for treating electrostatic interactions in solvated protein simulations to increase the accuracy with which pK_a shifts in proteins can be predicted.

ACKNOWLEDGMENTS

This work has been supported by grants from NIH (GM-30580), the Columbia University Center for Biomolecular Simulations (NIH P41RR06892), and a grant of supercomputer time from the Pittsburgh Supercomputer Center.

REFERENCES

1. Tanford, C., Roxby, R. Interpretation of protein titration curves. Application to lysozyme. Biochemistry 11:2192–2198, 1972.
2. Tanford, C., Kirkwood, J.G. Theory of protein titration curves. I. General equations for impenetrable spheres. J. Am. Chem. Soc. 79:5333–5339, 1957.
3. Yang, A.-S., Honig, B. On the pH dependence of protein stability. J. Mol. Biol. 231:459–474, 1993.
4. Yang, A.-S., Gunner, M.R., Sampogna, R., Sharp, K., Honig, B. On the calculation of pK_a's in proteins. Proteins 15:252–265, 1993.
5. Smith, P.E., Brunne, R.M., Mark, A.E., van Gunsteren, W.F. Dielectric properties of trypsin inhibitor and lysozyme calculated from molecular dynamics simulations. J. Phys. Chem. 97:2009–2014, 1993.
6. Tirado-Rives, J., Jorgensen, W.L. Molecular dynamics of proteins with the OPLS potential functions. Simulations of the third domain of silver pheasant ovomucoid in water. J. Am. Chem. Soc. 112:2773–2781, 1990.
7. Bashford, D., Gerwert, K. Electrostatic calculation of the pK_a values of ionizable groups in bacteriorhodopsin. J. Mol. Biol. 224:473–486, 1992.
8. Schreiber, H., Steinhauser, O. Cutoff size does strongly influence molecular dynamics results on solvated polypeptides. Biochemistry 31:5856–5860, 1992.
9. Levitt, M., Sharon, R. Accurate simulation of protein dynamics in solution. Proc. Natl. Acad. Sci. U.S.A. 85:7557–7561, 1988.
10. Merz, K.M. Determination of pK_a's of ionizable groups in proteins: The pK_a of Glu 7 and 35 in hen egg white lysosyme and Glu 106 in human carbonic anhydrase II. J. Am. Chem. Soc. 113:3572–3575, 1991.
11. Niedermeier, C., Schulten, K. Molecular dynamics simulations in heterogeneous dielectric and Debye-Huckel media. Application to the protein bovine pancreatic trypsin inhibitor. Mol. Simulation 8:361, 1992.
12. Bashford, D., Karplus, M. pK_a's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. Biochemistry 29:10219–10225, 1990.
13. Lee, F.S., Chu, Z.T., Warshel, A. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs. J. Comp. Chem. 14:161–185, 1993.
14. Rashin, A.A., Rabinowitz, J.R., Banfelder, J.R. Calculations of pK differences between structurally similar compounds. J. Am. Chem. Soc. 112:4133–4137, 1990.
15. Gilson, M.K., Sharp, K.A., Honig, B.H. Calculating the electrostatic potential of molecules in solution: Method and error assessment. J. Comp. Chem. 9:327–335, 1987.
16. Warwicker, J., Watson, R.S. Calculation of electric potential in the active site cleft due to α-helix dipoles. J. Mol. Biol. 157:671–679, 1982.
17. Davis, M.E., McCammon, J.A. Electrostatics in biomolecular structure and dynamics. Chem. Rev. 90:509–521, 1990.
18. Beroza, P., Fredkin, D.R., Okamura, M.Y., Feher, G. Protonation of interacting residues in a protein by a monte carlo method: Application to lysozyme and the photosynthetic reaction center of rhodobacter sphaeroides. Proc. Natl. Acad. Sci. U.S.A. 88:5804–5808, 1991.
19. Russel, S.T., Warshel, A. Calculations of electrostatic energies in proteins. The energetics of ionized groups in bovine pancreatic trypsin inhibitor. J. Mol. Biol. 185:389–404, 1985.
20. Schreiber, H., Steinhauser, O. Taming cut-off induced artifacts in molecular dynamics studies of solvated polypeptides. The reaction field method. J. Mol. Biol. 228:909–923, 1992.
21. Kitchen, D.B., Reed, L.H., Levy, R.M. Molecular dynamics simulation of solvated protein at high pressure. Biochemistry 31:10083–10093, 1992.
22. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Hermans, J. "Intermolecular Forces." Dordrecht: Reidel, 1981.
23. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79:926–935, 1983.
24. Alper, H.E., Levy, R.M. Computer simulations of the dielectric properties of water: Studies of the simple point charge and transferrable intermolecular potential models. J. Chem. Phys. 91:1242–1251, 1989.
25. Belhadj, M., Alper, H.E., Levy, R.M. Molecular dynamics simulations of water with Ewald summation for the long range electrostatic interactions. Chem. Phys. Lett. 179: 13–20, 1991.
26. Del Buono, G.S., Rossky, P.J., Schnitker, J. Model dependence of quantum isotope effects in liquid water. J. Chem. Phys. 95:3728–3737, 1991.
27. Beveridge, D.L., DiCapua, F.M. Free energy via molecular simulation: Applications to chemical and biomolecular systems. Annu. Rev. Phys. Chem. 18:431–492, 1989.
28. Straatsma, T.P., McCammon, J.A. Computational alchemy. Annu. Rev. Phys. Chem. 43:407–435, 1992.
29. Warshel, A., Sussman, F., King, G. Free energy of charges in solvated proteins: Microscopic calculations using a reversible charging process. Biochemistry 25:8368–8372, 1986.
30. Lee, F.S., Warshel, A. A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. J. Chem. Phys. 97:3100–3107, 1992.
31. Levy, R.M., Belhadj, M., Kitchen, D.B. Gaussian fluctuation formula for electrostatic free-energy changes in solution. J. Chem. Phys. 95:3627–3633, 1991.
32. Levy, R.M., Westbrook, J.D., Kitchen, D.B., Krogh-Jespersen, K. Solvent effects on the adiabatic free-energy difference between the ground and excited states of methylindole in water. J. Phys. Chem. 95:6756–6758, 1991.
33. Marcus, R.A., Sutin, N. Electron transfers in chemistry and biology. Biochem. Biophys. Acta 811:265–322, 1985.
34. Bader, J.S., Chandler, D. Computer simulation study of the mean forces between ferrous and ferric ions in water. J. Phys. Chem. 96:6423–6427, 1992.
35. Kitchen, D.B., Hirata, F., Westbrook, J.D., Levy, R.M.,

Kofke, D., Yarmush, M. Conserving energy during molecular dynamics simulations of water, proteins and proteins in water. J. Comp. Chem. 11:1169–1180, 1990.
36. Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A. An all atom force field for simulations of proteins and nucleic acids. J. Comp. Chem. 7:230–252, 1986.
37. Andersen, H.C. RATTLE: a "velocity" version of the SHAKE algorithm for molecular dynamics calculations. J. Comput. Phys. 52:24–34, 1983.
38. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, J.E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
39. Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F., Weng, J. Protein Data Bank. In Crystallographic databases—Information content, software systems, scientific applications. Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, 1987, 107–132.
40. Hodsdon, J.M., Brown, G.M., Sieker, L.C. Refinement of triclinic lysozyme: I. Fourier and least-squares methods. Acta Crystallogr. B 46:54–62, 1990.
41. Ryckaert, J.-P., Ciccotti, G., Berendsen, H. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-Alkanes. J. Comput. Phys. 23:327–341, 1977.
42. Berendsen, H.J.C., Postma, J.P., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684–3690, 1984.
43. Figueirido, F., Del Buono, G.S., Levy, R.M. Molecular mechanics and electrostatic effects. Biophys. Chem. (in press).
44. Schreiber, H., Steinhauser, O. Molecular dynamics studies of solvated polypeptides: Why the cut-off scheme does not work. Chem. Phys. 168:75–89, 1992.
45. Huston, S.E., Rossky, P.J. Free energies of association for the sodium-dimethyl phosphate ion pair in aqueous solution. J. Phys. Chem. 93:7888–7895, 1989.
46. Smith, P.E., Pettitt, B.M. Peptides in ionic solutions: A comparison of the Ewald and switching function techniques. J. Chem. Phys. 95:8430–8441, 1991.
47. Darden, T., York, D., Pedersen, L. Particle mesh Ewald: An $N.\log(N)$ method for Ewald sums in larger systems. J. Chem. Phys. 98:10089–10092, 1993.
48. Appel, A.W. An efficient program for many body simulations. SIAM J. Sci. Statist. Comput. 6:85–103, 1985.
49. Barnes, J., Hut, P. A hierarchical $O(N \log N)$ force-calculation algorithm. Nature (London) 324:446–449, 1986.
50. Greengard, L. "The Rapid Evaluation of Potential Fields in Particle Systems." Cambridge, MA: MIT Press, 1988.
51. Board, J.A., Causey, J.W., Leathrum, J.F.J., Widemuth, A., Schulten, K. Accelerated molecular dynamics simulation with the parallel fast multipole algorithm. Chem. Phys. Lett. 198:89–94, 1992.
52. Ding, H.-Q., Karasawa, N., Goddard, W.A.I. Atomic level simulations on a million particles: The cell multipole method for coulomb and london nonbonded interactions. J. Chem. Phys. 97:4309–4315, 1992.
53. Figueirido, F.E., Levy, R.M. Hierarchical N-body algorithms with both long- and short-range interactions. J. Comp. Phys. (submitted).
54. Komeiji, Y., Uebayasi, M., Someya, J., Yamato, I. A molecular dynamics study of solvent behavior around a protein. Proteins 16:268–277, 1993.
55. Gilson, M.K., Honig, B.H. The dielectric constant of a folded protein. Biopolymers 25:2097–2119, 1986.
56. Simonson, T., Perahia, D., Brünger, A.T. Microscopic theory of the dielectric properties of proteins. Biophys. J. 59: 670–690, 1991.
57. Simonson, T., Perahia, D., Bricogne, G. Intramolecular dielectric screening in proteins. J. Mol. Biol. 218:859–886, 1991.
58. Figueirido, F., Levy, R.M. Unpublished results.
59. Alper, H.E., Levy, R.M. Dielectric and thermodynamic response of a generalized reaction field model for liquid state simulations. J. Chem. Phys. 99:9847–9852, 1993.
60. Teeter, M.M. Water-protein interactions: Theory and experiment. Annu. Rev. Biophys. Biophys. Chem. 20:577–600, 1991.
61. Lounnas, V., Pettitt, B.M. A Connected-Cluster of hydration around myoglobin: Correlation between molecular dynamics simulations and experiment. Proteins 18:133–147, 1994.
62. Lounnas, V., Pettitt, B.M. Distribution function implied dynamics versus residence times and correlations: Solvation shells of myoglobin. Proteins 18:148–160, 1994.

APPENDIX: GAUSSIAN FLUCTUATION FORMULA

In this Appendix we present a short derivation of the formula

$$\Delta G = \sum_i \overline{V}_i \Delta q_i - \frac{\beta}{2} \sum_{i,j} \langle \Delta V_i \Delta V_j \rangle \Delta q_i \Delta q_j \qquad (A1)$$

for the change in free energy in terms of the charge differences. For a more detailed discussion the reader is referred to Ref. 31.

The partition function for a system of charges $\{q_i\}$ at positions $\{\mathbf{x}_i\}$, assumed fixed, can be written as

$$Z(\{q_i, \mathbf{x}_i\}) = e^{-\beta G} = \int d\{\mathbf{X}\} e^{-\beta U(\{\mathbf{x}_i\}, \{\mathbf{X}\})} \qquad (A2)$$

where the interaction potential U has the form

$$U(\{\mathbf{x}_i\}, \{\mathbf{X}\}) = U'(\{\mathbf{x}_i\}, \{\mathbf{X}\}) + U''(\{\mathbf{X}\}) \qquad (A3)$$

where U' is the Coulomb potential of interaction between the charges $\{q_i\}$ and the "solvent" (identified by $\{\mathbf{X}\}$), and U'' is the part of the potential that depends only on the "solvent" coordinates $\{\mathbf{X}\}$. Since U' is just a sum of Coulomb potentials it is a linear function of the charges $\{q_i\}$ and so

$$Z(\{q_i, \mathbf{x}_i\}) = \int d\{\mathbf{X}\}\, e^{-\beta \sum_i q_i V_i - \beta U''(\{\mathbf{X}\})} \qquad (A4)$$

where V_i is the Coulomb potential produced by all charges not in the set $\{q_i\}$ at position $\mathbf{x}_i$. This expression can be rewritten as

$$Z = C \int d\{V_i\}\, p(\{V_i\}) e^{-\beta \sum_i q_i V_i} \qquad (A5)$$

where the constant C is given by the q-independent formula

$$C = \int d\{\mathbf{X}\} e^{-\beta U''(\{\mathbf{X}\})} \qquad (A6)$$

and $p(\{V_i\})$ is the probability distribution function for the random variables $\{V_i\}$. The Gaussian fluctuation formula is derived by making the assumption that p is a Gaussian distribution, i.e., is of the form

$$\ln p(\{V_i\}) = -\frac{1}{2} \sum_{i,j} \Delta V_i C^{ij} \Delta V_j + \text{constant} \qquad (A7)$$

where $\Delta V_i \equiv V_i - \overline{V}_i$ and the matrix C^{ij} can be related to the correlation functions $\langle \Delta V_i \Delta V_j \rangle$. More precisely we have

$$(C^{-1})_{ij} \equiv \sigma_{ij} = \langle \Delta V_i \Delta V_j \rangle_{\{q\}=0} \qquad (A8)$$

where the correlation function is to be computed in

the "reference" state with all charges q_i equal to zero. A simple gaussian integration then yields the desired result.

It should be pointed out that there is another contribution to ΔG which is independent of the "solvent" configuration and is given by the Coulomb interaction between the fixed charges,

$$-\frac{\beta}{2}\sum_{i\neq j}\frac{q_i q_j}{||\mathbf{x}_i-\mathbf{x}_j||} \tag{A9}$$

The above derivation can be extended straightforwardly to include a reference state where some or all the charges are not zero. The details can be found in Ref. 31.