

Simulation of Protein Conformational Freedom as a Function of pH: Constant-pH Molecular Dynamics Using Implicit Titration

António M. Baptista, Paulo J. Martel, and Steffen B. Petersen*

MR-Center, SINTEF UNIMED, N-7034 Trondheim, Norway

ABSTRACT Solution pH is a determinant parameter on protein function and stability, and its inclusion in molecular dynamics simulations is attractive for studies at the molecular level. Current molecular dynamics simulations can consider pH only in a very limited way, through a somewhat arbitrary choice of a set of fixed charges on the titrable sites. Conversely, continuum electrostatic methods that explicitly treat pH effects assume a single protein conformation whose choice is not clearly defined. In this paper we describe a general method that combines both titration and conformational freedom. The method is based on a potential of mean force for implicit titration and combines both usual molecular dynamics and pH-dependent calculations based on continuum methods. A simple implementation of the method, using a mean field approximation, is presented and applied to the bovine pancreatic trypsin inhibitor. We believe that this constant-pH molecular dynamics method, by correctly sampling both charges and conformation, can become a valuable help in the understanding of the dependence of protein function and stability on pH. *Proteins* 27:523–544, 1997. © 1997 Wiley-Liss, Inc.

Key words: protonation equilibrium; protein conformation; continuum electrostatics; potential of mean force; force fields; mean field theory

INTRODUCTION

The structural stability and biological function of proteins are usually strongly correlated with pH, and the inclusion of pH effects in computer simulation methods is therefore of great interest. The primary effect of pH on a protein is the control of the protonation equilibrium of its titrable sites, whose most marked consequence is probably the associated change in charges and electrostatic interactions. Thus, pH effects have been traditionally modeled using continuum electrostatic (CE) methods to compute the interaction between sites^{1–3} and, more recently, also the pK_a shift caused by the protein

environment^{4,5}; this methodology will be hereafter referred to as pH-dependent CE (pH-CE). In CE methods the protein is modeled as a polarizable cavity with fixed point charges, immersed in a polarizable solvent and, in some cases, with a counterion atmosphere.^{3,6–11} Recently, CE models have achieved a high level of atomic detail, which should hopefully bring the corresponding calculations into a quantitative stage.^{7,9,10,12} Accordingly, recent pH-CE methods should also provide a route to quantitative predictions. Most emphasis has been put on the prediction of pK_a values,^{4,5,13,14} although studies of the effect of pH on stability^{15,16} and binding¹⁷ have also been done. The predicted pK_a s are generally in reasonable agreement with experimental results, but in some cases a very strong dependence on the protein conformation is observed, which can go up to several pH units. The differences are observed between different conformations from molecular dynamics (MD) simulations,^{13,18} different crystal forms,^{4,14} or even different refinements of the same crystal form.¹⁴ Thus, the choice of a single conformation is problematic. Moreover, with such sensitivity to conformation, it is not clear how fair is the judgment on the predictive capacity of the existing methods, i.e., the failure or success in predicting a particular pK_a value will always have uncertainty associated due to eventual conformation arbitrariness. For example, in previous calculations with bovine pancreatic trypsin inhibitor (BPTI), the question arose of whether or not to use crystal coordinates that disagree with nuclear magnetic resonance (NMR) results.^{5,19} This problem is further complicated by the fact that conformation is also pH dependent, so that experimental mean conformations corresponding to a particular pH value may not be valid over the whole pH range. The conformation dependence is probably the major concern with current pH-CE methods²⁰ and suggests the necessity of developing more appropriate theoretical approaches and methodologies.

Current address of Paulo J. Martel: Instituto de Tecnologia Química e Biológica, R. da Quinta Grande 6, Apartado 127, P-2780 Oeiras, Portugal.

*Correspondence to: Prof. Steffen B. Petersen, MR-Center, SINTEF UNIMED, N-7034 Trondheim, Norway.

Received 21 December 1995; Accepted 30 September 1996

The dependence of pH-CE results on conformation is actually not surprising since it is, in itself, a result of a similar dependence of CE calculations, which is now well known and is the basis of several methods that use CE calculations to include solvation effects in an implicit way.^{21–29} The obvious problem is that the conformational freedom of the protein is not easily included in current CE methods, as opposed to MD and other molecular mechanics (MM) methodologies. Given that the dielectric response of the protein arises from both induced and permanent dipoles, and since the orientation of the latter occurs through conformational changes, it may seem that a judicious choice of the dielectric constant could solve the problem. Along this line, dielectric constants of 4³⁰ or 20³¹ have been suggested as a way of accounting for the protein conformational fluctuations in CE calculations. However, the dielectric response is probably a rather inhomogeneous property of the molecular^{32,33} (e.g., with main chain and side chains contributing quite differently to the overall response³⁴) and thus there seems to be no physical basis for trying to include conformational freedom on a single-dielectric model of the protein. In contrast, there is little doubt about the dielectric constant due only to induced polarization, with both theory and experiment pointing to a (homogeneous) value of 2–4 (see Ref. 7 and references therein). Furthermore, pH-CE methods implicitly assume that the free energy can be decomposed in electrostatic and non-electrostatic terms, and that the latter cancel when differences are taken. However, though this may be a reasonable assumption for a hypothetical non-flexible molecule, protein flexibility is probably strongly coupled with electrostatic interactions,³⁵ rendering such decomposition invalid. In conclusion, it seems that the only consistent view of current CE and pH-CE methods is that they may account for solvent (total) polarization and protein-induced polarization, but not for conformational effects. In other words, they pertain to “stiff” proteins, not to real, flexible ones.

Conversely, usual MM methods are intended to simulate a single protonation state and cannot include titration in a simple way, as pointed out before.³⁶ A possible approach using MM would be to use conventional thermodynamic integration and perturbation methods³⁷ to compute the free energy differences between all protonation states (necessary to describe the protonation equilibrium; see below), but this is not a feasible task with current computer power. An alternative route to compute these free energies using MM is the Gaussian fluctuation method of Levy and co-workers,³⁸ recently applied to the calculation of intrinsic pK_a s in lysozyme.³⁹ A very attractive feature of this method is that a single MD simulation is necessary, and it can in principle be used to compute also the site-site terms of the just mentioned free energy differences. These character-

istics make the Gaussian fluctuation method a serious potential substitute for CE methods in pH-dependent calculations. However, although the method does not assume a totally static structure like pH-CE methods, it assumes that the titrable sites have fixed positions and also (by assuming the potential at the sites to have constant mean and correlation matrix) that the populations of protein conformations are not affected by protonation. Unfortunately, though these approximations may be valid in some cases, the interesting situations in protein applications will often be the ones where titration and conformational changes are correlated, as discussed by Gilson.³⁶ This author has also suggested that a possible way to include pH effects in MM calculations would be to use the most likely protonation state, as computed by a pH-CE method.³⁶ However, since we know that the pH-CE results are conformation dependent, we may ask what happens if the protein conformation undergoes some changes during the simulation. Should we perform periodic updates of the “predominant” protonation state during the MD simulation? Although this may sound like a reasonable procedure, it is not at all clear what its physical meaning is, i.e., it is not obvious what kind of sampling we are performing. Without knowing the nature of the sample it is impossible to relate the simulation to the real system in a proper way, so that, e.g., comparisons with experimental results would be meaningless. The ideal would be to sample both conformations and protonation states that are representative of the pH of interest. However, as the foregoing considerations clearly illustrate, the two aspects are interwoven and we thus need a simulation method that can sample both of them simultaneously.

During the final writing of this article You and Bashford⁴⁰ proposed a method for the explicit inclusion of conformational flexibility on pH-CE calculations. Their approach is based on estimating the conformation-averaged (true) free energies of the different protonation states through a set of conformations whose weight is computed with a combination of MM and CE energy terms, in the spirit of the aforementioned CE-based implicit solvation methods. They present a set of pK_a calculations that again clearly illustrates the sensitivity of usual calculations on conformation and, more importantly, seems to indicate that a general improvement (i.e., predicted values closer to experimental ones) does indeed result from the inclusion of conformational aspects. In spite of these results there is, in our opinion, a major problem with this method: the free energies are computed directly through a sum-over-states, a method known to be extremely inefficient.⁴¹ The only way to mitigate this problem is to do a nearly exhaustive coverage (Boltzmann sampling is not enough) of the conformation space, which is clearly not the case in the work by You and Bash-

ford.⁴⁰ Indeed, the inclusion of even a small number of conformations is computationally so demanding that these authors had to resort to a set of independent, site-based conformational searches; site-site electrostatic effects were treated as being partially uncoupled from conformation. In addition, it seems to us that the absence of single-conformation free energy differences in the formalism makes doubtful the separation in electrostatic and non-electrostatic terms (see comments above). Nevertheless, the method is obviously able to “relieve” unfavorable protonation/conformation pairs (see below), and in this respect it represents a clear improvement over the usual pH-CE calculations per se.

A different approach has been used in the present work, which is based on the assumption that CE calculations provide a satisfactory way of obtaining free energy differences between molecules with the same conformation and different charge distributions. That is, instead of trying to extend the usual CE methods to include flexibility explicitly, as was done by You and Bashford,⁴⁰ we simply eliminate any ambiguity by restricting each calculation to a single protein conformation. This difference in approach may be because our primary aim was not to improve CE methods, but rather to include pH effects on MD simulations, for which the single-conformation calculations prove very useful (see below). The rationale is essentially to restrict CE and MD methods to what they can do best, and combine them in a way that is both theoretically correct and computationally feasible.

In this article we start by presenting how the quantities computed by pH-CE methods should be reinterpreted in terms of constant protein conformation; the assumptions implicit in the usual single-conformation pH-CE calculations are also discussed. We then derive a potential of mean force (PMF) that can treat simultaneously titration and conformational freedom and that can be used as the potential energy in MM calculations. When used in an MD simulation, one obtains a constant-pH MD method, where configurations of the system, including protein conformations, are sampled in a way that properly reflects a given pH value. Due to implementation convenience, a mean field approximation is used for the titration of each conformer. (This approximation can in principle be removed by a simple modification on the calculation of forces in the MD program, without significant computing time costs.) The inclusion of titration is not done in an explicit way, but rather implicitly through the PMF, analogously to, e.g., the way solvent is included in simulations with implicit solvation. An “exact” implementation of the method requires the PMF to be re-evaluated for each conformer, which demands large computation times; we propose instead the use of periodic updating during the simulation, to make the application to average-sized proteins realistic. We have imple-

mented a simple version of the method using this approximations and applied it to BPTI. The results indicate that both the mean field and the periodic updating approximations are valid for the system and methodology used. Also, a general increase of the atomic fluctuations was observed, as should be expected from theoretical considerations (see below).

We finally note that although the present method results from a combination of MM and CE methods, it is *not* a new method for MM with implicit solvation.^{21–29} The PMF, which is central in this work, can in principle be derived from any MM force field, which itself may treat solvation in an implicit or explicit way. Since the PMF requires a periodic updating based on pH-CE calculations, it may actually be particularly easy to implement the method using a CE-based solvation force field, but that is entirely a matter of choice and convenience. The present method may also seem similar to the aforementioned Gaussian fluctuation approach, in the sense that it combines MM and titration.^{38,39} However, while the Gaussian fluctuation method is an MM free-energy method, the scope of the present method is the correct simultaneous sampling of conformations and protonation states of the protein molecule as a function of pH. Although our method relies on the previous existence of MM force fields and pH-CE methods, it can be seen as extending the scope of both, by including titration effects in the former and conformational freedom in the latter.

THEORY

The following treatment is intended for protein ideal solutions (i.e., with non-interacting protein molecules), as are pH-CE methods; thus it suffices to consider a system consisting of a single protein molecule surrounded by a sufficiently large number of solvent molecules. While focusing on a single protein molecule, we can easily relate to solution properties, since standard chemical potential differences are then given by the free energy differences of such single-protein “cells.”

On writing down Eq. (4) (see below) we implicitly assume the degrees of freedom of the protein to be independent of protonation. This may actually be true in terms of the simulation, if one uses a force field with explicit electronic lone pairs on the titrable proton sites. Even as an approximation this is a common practice in free energy calculations and surely a reasonable assumption for such large molecules as proteins.^{37,39}

Also, for simplicity, we will assume the solution to be at constant temperature and volume. The formalism for constant pressure (instead of volume) is analogous. Moreover, for most processes involving protein solutions at constant pressure the volume variations are negligible, so that Gibbs or Helmholtz free energy differences will be identical.

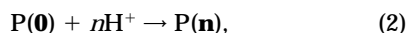
Free Energies From Continuum Electrostatics

A protein with s titrable sites has 2^s possible charge configurations, or *protonation states*, here represented as a vector $\mathbf{n} = (n_1, n_2, \dots, n_s)$, where $n_i = 1$ or 0, depending on whether the site i is protonated or not. The most convenient type of system for studying pH effects is a cell, as described above, that is open to the titrable protons only, i.e., a system in which the individual protonation states of the protein molecule, n_i , can vary but whose amounts of all other chemical species in solution (hereafter simply referred to as “solvent”), eventually including solvated protons, are assumed to be present in fixed amounts, characteristic of their corresponding chemical potentials. (This is equivalent to assuming that exchanges between the bath and the cell do not vary the total contents of the latter; a more realistic system, where the amounts of all solvent species can vary, leads essentially to the same results.) Since we are only interested in sampling configurations of the system, it is not important how the titrating protons enter and leave the system. Such a system corresponds to a semi-grand canonical statistical mechanical ensemble characterized by the chemical potential of the hydrogen ion, temperature, and volume. The probability of finding the system with the protein in protonation state \mathbf{n} can then be obtained as a particular case of the general expression for open systems,⁴² namely

$$p(\mathbf{n}) = \frac{\exp [\beta \mu n - \beta A(\mathbf{n})]}{\sum_{\mathbf{n}'} \exp [\beta \mu n' - \beta A(\mathbf{n}')]} \quad (1)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, T is the absolute temperature, μ is the chemical potential of the hydrogen ion, $n = \sum_i n_i$ is the total protonation, and $A(\mathbf{n})$ is the free energy of the system when the protein is in state \mathbf{n} .

The approach in usual pH-CE methods is to express Eq. (1) in terms of the standard free energies, $\Delta A^\circ(\mathbf{n})$, of the protonation reactions in solution:



where $P(\mathbf{n})$ denotes the protein with protonation state \mathbf{n} [$P(\mathbf{0})$ is the fully deprotonated form]. The difference in the standard chemical potentials of the two protein forms in solution can be expressed as the difference of the free energies of the single-protein system (see above), and thus we can rewrite Eq. (1) as

$$p(\mathbf{n}) = \frac{\exp [-\beta \Delta A^\circ(\mathbf{n}) - n2.3pH]}{\sum_{\mathbf{n}'} \exp [-\beta \Delta A^\circ(\mathbf{n}') - n'2.3pH]} \quad (3)$$

$\Delta A^\circ(\mathbf{n})$ and $p(\mathbf{n})$ are then usually computed using (single-conformation) CE energy differences, in a way that depends on the particular pH-CE method and will not be of concern to us here.^{4,5,36,43–47}

However, as discussed above, the usual CE calculations cannot account for the protein conformational freedom and thus, in a strict sense, the $\Delta A^\circ(\mathbf{n})$ and $p(\mathbf{n})$ computed by pH-CE methods refer to a hypothetical “stiff” protein with a particular conformation c . It is easy to show that the free energy of this “stiff” protein, $A(\mathbf{n}, c)$ (a PMF of the real system), is related to the conditional probability of \mathbf{n} for a given c :

$$\begin{aligned} p(\mathbf{n}|c) &= \frac{\exp [\beta \mu n - \beta A(\mathbf{n}, c)]}{\sum_{\mathbf{n}'} \exp [\beta \mu n' - \beta A(\mathbf{n}', c)]} \\ &= \frac{\exp [-\beta \Delta A^\circ(\mathbf{n}, c) - n2.3pH]}{\sum_{\mathbf{n}'} \exp [-\beta \Delta A^\circ(\mathbf{n}', c) - n'2.3pH]} \end{aligned} \quad (4)$$

where $\Delta A^\circ(\mathbf{n}, c)$ is the standard free energy of a reaction analogous to Eq. (2) but involving the “stiff” protein. It is now evident that $\Delta A^\circ(\mathbf{n}, c)$ is the quantity actually computed by usual pH-CE methods, which can thus only provide conditional mean charges, pK_a s, etc.

It should be noted, however, that in pH-CE methods that use thermodynamic cycles involving the model compound, the latter is also assumed to have a single conformation.^{4,5} (The model compound is usually assumed to have the same conformation as the corresponding amino acid residue in the protein, but the only reason for this is the cancelation of the “lattice singularities” arising in finite-difference techniques; a more consistent choice would be an average conformation of the compound in solution, as is done for the protein.) The present reinterpretation then implicitly assumes that the single-conformer pK_a is equal to the (conformation-averaged) experimental value used by these methods. In fact, the pK_a of a “stiff” model compound with conformation c is related to the conformation-averaged one as

$$pK_{\text{mod}}(c) = pK_{\text{mod}} + \log \frac{p(c|1)}{p(c|0)}, \quad (5)$$

where the $p(c|1)$ and $p(c|0)$ are the conditional probabilities of c when the model compound is protonated and deprotonated, respectively. For fairly extended conformations (as is usually found in proteins) the charged site will be reasonably shielded by surrounding water, and thus the conformation probability should not depend strongly on protonation. Hence, we may expect the experimental pK_{mod} values to be a good estimate of the values required for single-conformer pH-CE calculations.

To see in more detail the assumptions implicit in the usual pH-CE calculations, which substitute $p(\mathbf{n}|c)$ for $p(\mathbf{n})$, we can write the latter in terms of a given conformation c and the conditional probabilities $p(c|\mathbf{n})$:

$$p(\mathbf{n}) = \frac{p(\mathbf{n}|c)/p(c|\mathbf{n})}{\sum_{\mathbf{n}'} p(\mathbf{n}'|c)/p(c|\mathbf{n}')} = \frac{\exp[-\beta\Delta A^\circ(\mathbf{n}, c) - n2.3\text{pH}]/p(c|\mathbf{n})}{\sum_{\mathbf{n}'} \exp[-\beta\Delta A^\circ(\mathbf{n}', c) - n'2.3\text{pH}]/p(c|\mathbf{n}')} \quad (6)$$

Comparison with Eq. (4) shows that pH-CE methods implicitly assume that $p(c|\mathbf{n}) \approx 1$ for all \mathbf{n} and the chosen conformation c . If \mathbf{n} is the predominant protonation state at the same pH conditions in which c was experimentally observed as the predominant or mean conformation, $[\mathbf{n}, c]$ may be virtually the only occurring state of the protein, and the approximation $p(c|\mathbf{n}) \approx 1$ will be a reasonable one. [This is clearly an oversimplification, since conformation space is continuous. However, this simple treatment is equivalent to consider a small region of conformations around c with nearly the same $\Delta A^\circ(\mathbf{n}, c)$.] However, this conformation may rarely occur with other protonation states, as the following simple examples illustrate.

Consider a protein with a titrable site that is facing the protein interior in the experimentally determined conformation, c , but that is solvent exposed in an alternative, less populated conformation c' . When pH-CE calculations are done neglecting the “correcting factors” $p(c|\mathbf{n})$, the term in the summation corresponding to the charged form will be small since in a buried conformation the neutral form will be favored. However, when the site is charged the protein will tend to adopt the solvent-exposed conformation c' and not c , i.e., $p(c|\mathbf{n}) < 1$, and the corresponding term in the summation will increase.

As another simple example, not involving a change in solvent-exposure, consider a protein whose observed conformation c has two close titrable sites. If the two sites bear charges of the same sign, the probability that both are charged may be very low when computed using Eq. (4). However, eventual alternative conformations where the sites are more distant will then be preferred, which may considerably increase that probability through the factors $p(c|\mathbf{n})$ in Eq. (6). Conversely, if the two sites bear charges of opposite signs, Eq. (4) will strongly favor the simultaneously charged forms. But, when at least one of the sites is neutral, the conformation c may lose its energy minimum status, and other

energetically available conformations become also populated. Thus, the mobility of these sites can become drastically increased, $p(c|\mathbf{n}) < 1$, and the probability of the neutral forms will be increased.

Although the differences arising from the corrections noted in the previous examples may simply vanish in the summation with higher terms (especially when many titrable sites exist), they will always show up in the numerator, when computing the corresponding probabilities. One may argue that the occurrence of such a protonation state will probably be low, anyway, since the conformation that favors it is also unlikely, so that its total neglect will not be important. However, one cannot rule out the biological relevance of a given conformation simply because it is low populated. An example is the transient “back door” of acetylcholinesterase observed in a MD simulation, which is apparently involved in the release of reaction products.⁴⁸ Thus, a conformational transient state may be the rate-limiting step in a biologically important process, and an error in the calculation of its (low) probability may correspond to an error of several orders of magnitude in the rate of that process. One should also note that in the foregoing examples we have implicitly assumed the same pH value at which c was the predominant or mean conformation. If a different pH value is considered the situation is even worse, since in such a case c may be inappropriate even for the predominant protonation state.

Examples analogous to those above have been discussed by Gilson,³⁶ although from the opposite point of view, i.e., as a critique of the usual (i.e., constant- \mathbf{n}) MD from the perspective of pH-CE methods. Gilson³⁶ concluded that the inclusion of protonation by a pH-CE approach partially “relieves” interactions that are unfavorable if a fixed \mathbf{n} is assumed. Conversely, we see that the inclusion of conformational freedom partially “relieves” interactions that are unfavorable if a fixed c is assumed. These considerations show that the energetics of the system depend on both conformation and protonation state, and that the respective probabilities are irrevocably linked. We now show how MM and pH-CE methods can be combined to treat both aspects simultaneously, providing a simulation method for constant-pH MD.

Potential of Mean Force for Constant pH

Although PMFs are more common as a way of removing the explicit reference to some configurational degrees of freedom of the system, the concept can be extended to eliminate the explicit reference to a variable number of particles.^{42,49} We can thus define a PMF $W(q)$ for the “cell” protein system (see

beginning of Theory section) as

$$\exp [-\beta W(q)] = D \sum_{\mathbf{n}} \exp [\beta \mu_{\mathbf{n}} - \beta U(\mathbf{n}, q)], \quad (7)$$

where D is some constant, q stands for all configurational coordinates (protein conformation included), and $U(\mathbf{n}, q)$ is the potential energy. We can regard $W(q)$ as the potential energy of a non-titrable *reduced system* (in analogy with the *extended system* formalism)⁵⁰ otherwise identical to the original system with a titrable protein. Definition (7) ensures that the distribution of conformations in the reduced system will be the same as the one of the original system at chemical potential μ .⁴² Therefore, a canonical MD simulation of the reduced system will also reflect a given pH value, i.e., it will be a constant-pH (and constant-temperature) MD simulation of the original titrable system. However, to perform such a simulation we have to express W in terms of quantities that we know how to compute. The PMF defined in Eq. (7) can also be written as⁴²

$$\nabla_k W(q) = \sum_{\mathbf{n}} \nabla_k U(\mathbf{n}, q) p(\mathbf{n}|q), \quad (8)$$

where ∇_k denotes the gradient with respect to the position of an atom k . Except for simulations that include the bare protein (either treating the solvent implicitly or not), $p(\mathbf{n}|q)$ is *not* $p(\mathbf{n}|c)$, the quantity computed in pH-CE methods (see above). Although, if we make the reasonable assumption that the dependence of $p(\mathbf{n}|q)$ on q is mainly due to the protein conformation, i.e., $p(\mathbf{n}|q) \approx p(\mathbf{n}|c)$, we get

$$\nabla_k W(q) = \sum_{\mathbf{n}} \nabla_k U(\mathbf{n}, q) p(\mathbf{n}|c), \quad (9)$$

where c should be understood as the conformation contained in q , i.e., as $c(q)$. All quantities in this equation can in principle be computed, so that a constant-pH MD simulation can be done. The energy $U(\mathbf{n}, q)$ can be obtained from the force field being used (see below), and $p(\mathbf{n}|c)$ can be computed by a pH-CE method, through Eq. (4). The previous approximation, $p(\mathbf{n}|q) \approx p(\mathbf{n}|c)$, probably holds in most situations, since normal solvent configurations are not likely to influence $p(\mathbf{n}|c)$ in a noticeable way. Unusual solvent configurations for which $p(\mathbf{n}|q) \neq p(\mathbf{n}|c)$ (e.g., with water molecules overlapping the titrable site) are not likely to occur during a MD simulation.

Since the definition of the PMF in Eq. (7) ensures a proper sampling of configurations, the average value of any property $X(q)$ can be done by direct sampling during the MD simulation, as usual. For properties

depending also on the protonation state we have

$$\begin{aligned} \langle X \rangle &= \sum_{\mathbf{n}} \int X(\mathbf{n}, q) p(\mathbf{n}, q) dq \\ &= \sum_{\mathbf{n}} \int X(\mathbf{n}, q) p(\mathbf{n}|q) p(q) dq \\ &= \int \langle X \rangle_q p(q) dq, \end{aligned} \quad (10)$$

where $p(q)$ is the distribution function of the coordinates (i.e., actually a probability *density*). The conditional mean $\langle X \rangle_q$ can be computed at each time step using again the approximation $p(\mathbf{n}|q) \approx p(\mathbf{n}|c)$:

$$\langle X \rangle_q = \sum_{\mathbf{n}} X(\mathbf{n}, q) p(\mathbf{n}|q) \approx \sum_{\mathbf{n}} X(\mathbf{n}, q) p(\mathbf{n}|c). \quad (11)$$

For these protonation-dependent properties the average can thus be computed by direct sampling of $\langle X \rangle_q$ during the simulation (using pH-CE calculations). Since protonation-dependent properties are computed through use of Eq. (10), only the mean values $\langle X \rangle$ can be directly obtained from the simulation. The fluctuations $\sigma(X)$, for instance, must then be computed from mean values using

$$\sigma^2(X) = \langle X^2 \rangle - \langle X \rangle^2. \quad (12)$$

Fluctuations in the constant-pH ensemble will generally be higher than in the usual canonical ensemble. This is a consequence of the additional perturbation introduced by the fluctuations on the proton occupancies of the titrable sites. Formally, this is a consequence of a general relation involving joint random variables, which in this case takes the form⁵¹

$$\sigma^2(X) = \langle \sigma_{\mathbf{n}}^2(X) \rangle + \sigma^2(\langle X \rangle_{\mathbf{n}}), \quad (13)$$

where $\sigma_{\mathbf{n}}^2(X)$ and $\langle X \rangle_{\mathbf{n}}$ are respectively the conditional variance and mean of X at fixed \mathbf{n} . This relation leads in general to a higher fluctuation, unless the comparison is done with an \mathbf{n} state that has a low probability at the considered pH *and* an exceptionally high $\sigma_{\mathbf{n}}^2(X)$. Since the interesting comparisons are generally with reasonably populated \mathbf{n} states [which contribute significantly to the first term in the right hand side of Eq. (13)], we will usually have $\sigma(X) \geq \sigma_{\mathbf{n}}(X)$.

In conclusion, an MD simulation using as potential energy the PMF here defined will produce the habitual structural statistics as well as the protonation-related ones, referring to both the given pH and temperature. However, as happens often with other non-microcanonical MD methods, time has lost its real meaning, and the trajectories cannot be taken literally.⁴¹ Moreover, and contrary to some of those methods, it is not possible to recover the temporal

behavior of the system by adjustment of some parameters (e.g., piston masses), because of the implicit way in which protonation is included through the PMF. As a consequence, true temporal properties (e.g., time correlation functions) cannot be computed with this constant-pH MD method.

METHODS

Implementation of the PMF

Force field

The use of $W(q)$ as the potential energy of the reduced system corresponds in fact to the definition of a new force field based on Eq. (7). Since the conformational sampling in MD simulations is totally determined by the forces, an alternative route is to use Eq. (9) as the starting point for the implementation of the new force field. A simple way of doing this is to change the potential energy expression used in the MD program to a mean potential energy at fixed configuration:

$$V(q) = \langle U \rangle_q = \sum_{\mathbf{n}} U(\mathbf{n}, q) p(\mathbf{n} | c). \quad (14)$$

In a strict mathematical sense $\nabla_k V(q) \neq \nabla_k W(q)$, but if $p(\mathbf{n} | c)$ is simply taken as an additional parameter in the force field, its derivatives will not be considered by the MD algorithm, and the forces will be correctly computed as $\nabla_k W(q)$. Hence, $V(q)$ is the implementation equivalent of the PMF. The use of $p(\mathbf{n} | c)$ is similar to the use of a coupling parameter in conventional free energy calculations, and its inclusion in the original force field has to be done in a similar fashion.³⁷ Although such implementation may at first seem unrealistic due to the 2^s protonation states in the summation in Eq. (9), it turns out that only pairwise terms are necessary. In fact, for the force fields currently being used with biomolecules, most terms in the energy function $U(\mathbf{n}, q)$ will not depend on \mathbf{n} , since most atoms do not belong to titrable sites. The remaining terms will involve one or at most two titrable sites, thus depending on one or two of the n_i values. Dependences on three or more titrable sites will not occur with amino acid residues, even through torsional terms. Thus, if we write the original potential energy function as

$$U(\mathbf{n}, q) = U^{(0)}(q) + \sum_i U_i^{(1)}(n_i, q) + \sum_{i < j} U_{ij}^{(2)}(n_i, n_j, q), \quad (15)$$

we obtain

$$V(q) = U^{(0)}(q) + \sum_i \langle U_i^{(1)} \rangle_q + \sum_{i < j} \langle U_{ij}^{(2)} \rangle_q \quad (16)$$

Therefore, the number of protonation-dependent terms in $V(q)$ actually depends on the square of the number of titrable sites, s^2 , which is rather tractable when compared with the original 2^s dependence.

In the present work we modeled $U(\mathbf{n}, q)$ with the CVFF force field⁵² using the following modifications. Deprotonated titrable sites were treated as having an electronic lone pair, represented by a hydrogen atom without partial charge; this ensures that the protein degrees of freedom are constant (see above), and \mathbf{n} is only a parameter of the potential energy. With this modification, the only changes occurring upon titration besides the atomic partial charges are very small modifications on the bond and angle parameters of the titrable carboxylic groups, even in a detailed all-atom force field like CVFF. We decided to ignore this minor difference and always used the protonated form parameters. In this way the $U_i^{(1)}$ and $U_{ij}^{(2)}$ terms in Eq. (15) become simply a sum of Coulombic interactions (all the other terms being included in $U^{(0)}$), which are thus the only protonation-dependent terms in $V(q)$:

$$V(q) = U^{(0)}(q) + \sum_{a \notin \text{TS}} \sum_{b \in \text{TS}} \frac{Q_a \langle Q_b \rangle_q}{r_{ab}} + \sum_{a \in \text{TS}} \sum_{b \in \text{TS}} \frac{\langle Q_a Q_b \rangle_q}{r_{ab}}, \quad (17)$$

where Q_a and Q_b are the partial charges of atoms a and b , which may belong or not to titrable sites ($\in \text{TS}$ or $\notin \text{TS}$), and r_{ab} is the distance between them. The force arising from these Coulombic terms is similar to the one previously suggested by Gilson,³⁶ except that he used the formal charges of the titrable sites, instead of the atomic (partial) ones. We note, however, that the reason why only electrostatic (Coulombic in this case) terms occur is that we have neglected the effect of titration on the degrees of freedom of the protein and also on the bond and angle energy terms. Therefore, the statement that “the forces associated with ionizable groups are, in effect, purely electrostatic”³⁶ may be misleading without a clear explanation of what is meant by “electrostatic.”

The implementation of Eq. (17) requires in principle a modification of the pairwise electrostatic terms where both atoms belong to titrable sites; indeed, for these cases it is not enough to use the $\langle Q_i \rangle_q$ s as the actual partial charges, because each pair will have its own $\langle Q_a Q_b \rangle_q$. This modification should be easily accomplished in MD software with user-definable energy terms or accessible code. However, the MD program used in the present work (see below) does not have such facilities, and we had to resort to an additional approximation to make the PMF implementation possible. The approximation is one often used in pH-CE calculations, namely, a

mean field (MF) approximation for the titrable sites.^{43,46} The MF approximation corresponds to consider that, although affecting each other, different sites titrate independently (for a given conformation), such that $\langle n_i n_j \rangle_q = \langle n_i \rangle_q \langle n_j \rangle_q$. This implies that for two atoms belonging to different sites we have $\langle Q_a Q_b \rangle_q = \langle Q_a \rangle_q \langle Q_b \rangle_q$. The atoms of a given site whose partial charge changes on titration are invariably very close, usually bonded together, and their local geometry is basically determined by bond, angle, and torsional terms (non-bonded interactions between 1–4 neighbors are typically eliminated in common force fields). Therefore, and although their charges are fully correlated, the previous equation can also be used in this case for implementation purposes, without much error (or no error at all, if they are in the 1–4 neighbor range, as is usually the case). In this way, the only change required in the usual calculation of forces (i.e., compared with constant- \mathbf{n} MD) is the substitution of a few atomic partial charges by their mean values at the current conformation. These mean values are obtained from the $\langle n_i \rangle_q$ value of the site i of which the atom is part, through the equation:

$$\langle Q_a \rangle_q = \langle n_i \rangle_q Q_a(1) + (1 - \langle n_i \rangle_q) Q_a(0), \quad (18)$$

where $Q_a(1)$ and $Q_a(0)$ are, respectively, the partial charges of atom a in the protonated and deprotonated forms of residue i , whose $\langle n_i \rangle_q$ can be computed with a pH-CE method using the MF approximation.

The MF approximation was shown to yield good results except when the two sites interact strongly and titrate in the same pH region,^{5,46} which is not a common situation in proteins, because like-charged groups tend to be apart.⁵³ However, cases may occur where sites are correlated, and it is always convenient to check if the approximation holds for each particular protein. A complete assurance can only be obtained from the exact calculation, although some efficient criteria have been suggested.⁴⁶ Such checking would, however, represent an additional burden for the simulation, and we instead decided on an a posteriori consistency check, based on the following reasoning. Let us suppose that two sites i and j are close to each other and partially protonated (i.e., $\langle n_i \rangle_q, \langle n_j \rangle_q \neq 0, 1$) during a simulation using the MF approximation for the electrostatic terms. In spite of the incorrect MF assumption, the sites still exert a *reciprocal* electrostatic (MF-averaged) effect upon each other, which, given their closeness, will determine the computed $\langle n_i \rangle_q$ and $\langle n_j \rangle_q$ values. Hence, as the conformation varies along the simulation, $\langle n_i \rangle_q$ and $\langle n_j \rangle_q$ will show a positive correlation. That is, even if we neglect a significant correlation in the pH-CE calculations, it will still show up during the simulation. This provides a simple consistency check on the MF assumption, which, in addition may reveal interesting opposite-charge correlations (see below).

pH-CE calculations

The pH-CE calculations necessary to compute the $\langle n_i \rangle_q$ values can in principle be performed with any of the existing methods^{4,5,36,43–47}; in the present study we used a very simple method, due to computer power limitations, as discussed below. The calculations were done with the program TITRA,^{11,54} which implements the accessibility-modified Tanford-Kirkwood method.⁴⁴ By assuming a spherical shape for the protein and neglecting electrostatic self-energies, this method allows for a computationally fast analytical solution of the electrostatic problem.² The neglect of electrostatic self-energies corresponds to the assumption that only nearby charged sites can affect the protonation equilibrium of another site; the non-electrostatic effect of the protein environment is not taken into account. This is not a drastic assumption for the protein used, BPTI, for which it has been observed that the desolvation penalty of the sites is compensated for by nearby dipolar groups.^{5,55} Dielectric constants of 2 and 78.5 were assigned to the protein and solvent regions, respectively. To increase the sensitivity to conformational changes, a null ionic strength was used, which maximizes the interaction between charges. The protein radius is automatically computed for each conformation as the root mean squared distance of the solvent-exposed atoms from the center of the protein. The charge depth parameter of the model was assigned a value of 0.4 Å.⁴³ To account for the non-spherical shape, the interaction of the titrable sites is corrected with their relative solvent accessibility,⁴⁴ computed with a tessellation-based algorithm. Instead of solving Eq. (4), an MF approximation is used (in conformity with the force field implementation; see above), which leads to a much simpler set of equations that can be quickly solved with an iterative algorithm, where the interdependent pK_a s (at fixed conformation) and $\langle n_i \rangle_q$ of all sites are simultaneously computed until self-consistency.^{43,46}

The uncoupling of individual protonations assumed by the MF approximation (see above) simplifies the calculation of the global n_i fluctuations, since it implies that each n_i is independently distributed (at fixed conformation) as a two-state (Bernoulli) distribution, for which $\langle n_i^2 \rangle_q = \langle n_i \rangle_q$.⁴⁶ By combining this with Eqs. (10) and (12) the fluctuation of n_i simply becomes:

$$\sigma^2(n_i) = \langle n_i \rangle - \langle n_i \rangle^2, \quad (19)$$

i.e., only the mean needs to be computed.

PMF update interval

A crucial point in the implementation of the PMF is its update, i.e., the protonation-dependent terms in $V(q)$ should in principle be known for each protein conformation, so that the MD algorithm may com-

pute the correct forces at all times. Each update requires a pH-CE calculation; therefore, if one wants to use accurate simulation methods (at both the MD and pH-CE levels), an "exact" update (i.e., at all MD time steps) is clearly an impossible task with current computer power. The obvious approximation is to update the PMF only after so many MD steps. This is facilitated by the fact that in the energy function $V(q)$ the generalized coordinates can be taken out of the protonation-dependent parts, as can be seen for the Coulombic terms in Eq. (17). It is easy to see that this can be done for other terms (bond, angle, torsion, Lennard-Jones, etc.) in case they also depend on protonation. Therefore, the protonation-dependent part (depending only on the original force field parameters) can be computed only after several steps, while the coordinates are allowed to change at all MD steps. The use of a sufficiently large update interval should make the use of accurate methods possible. However, such an approximation may introduce errors that will override the intended accuracy (e.g., the CE calculation can lead to abrupt changes in the charges, resulting in sudden artificial forces upon updating) and should thus be considered with some care.

This practical necessity of using a large update interval also exists in CE-based implicit solvation methods, where a CE calculation per MD step should also in principle be done.^{24,26} There is obviously a clear difference between the role of CE calculations in these methods and the present one: while the latter uses CE calculations to compute protonation-averaged properties at fixed conformation, the former intend to derive forces that reflect the solvent effects upon conformational changes. Nevertheless, these methods can give us useful information on the effect of a long update interval on CE-computed quantities (in that case solvation forces), and it is thus interesting to consider the arguments and results presented until now. It has been argued that an update interval much lower than the rotational relaxation time of the water molecule (~ 10 ps), e.g., 200 fs, should be a safe choice.^{26,27} However, in our opinion, this argument is somewhat inconsistent with the physical implications of using a CE model. The high relaxation time of the water molecule means that it responds slowly to changes in the electric field caused by the protein faster motions, but then, if one wishes to retain the temporal significance of the trajectories of the implicit solvation method, the use of a CE model as a way of describing the electrostatic effect of the solvent at each time step is actually inappropriate, because in a temporal description the concept of a (non-complex) dielectric constant is only meaningful for processes occurring on time scales much higher than the dielectric relaxation time, which is clearly not the case for the protein motions. Instead, by using a CE model one is implicitly

assuming that the solvent is being modeled in an average way, and therefore all temporal features of the solvent are irrelevant and the resulting implicit solvation method should be seen only as a non-temporal sampling method. We can still characterize the "temporal" behavior of this implicitly solvated protein by analyzing its trajectories during an MD run, but this behavior may be quite different from the original one. By choosing to perform the CE calculations only after several MD steps (because of the above-mentioned computing reasons), we are introducing an approximation on the simulation of this new system, not of the real one. Therefore, the choice of this update time interval should not be based on correlation times of the real system, but rather on correlation times of this new one or, more exactly, on the correlation times of the CE-derived quantities used by the particular solvation method. That is, if the method makes use of the surface charge density²⁴ or reaction field gradient,²⁶ these are the quantities whose correlation times impose an upper limit on the update time interval. The ideal interval in terms of the compromise between exactness and computing time would in general have to be determined by experiment; this is the usual procedure in the multiple time step method,⁴¹ of which this is in fact a particular case, with different time steps being used for the CE-based and non-CE-based interactions in the system. The comparison of the use of different update intervals in a CE-based solvation method has been done by Sharp,²⁴ who has observed that update intervals of 50, 100, and 200 fs lead essentially to the same results, using a CE method with atomic-level detail.

In the present method, the quantity we need from the CE calculations is not the surface charge density, as in Sharp's method, but rather a set of $\langle n_i | q \rangle$ values (or, in general, a set of $p(\mathbf{n} | c)$ values). However, for the same methodology it seems reasonable to expect similar update time intervals to be equally appropriate. Nevertheless, we note that the dynamics of the system will in general depend on both the CE and MD methodologies being used, and also that different systems may show different dynamics. (Sharp has used a dipeptide, not a typical protein molecule.) Hence, the only safe way to determine the appropriate update time interval for a given methodology is to compare the results using different interval sizes. Due to limitations of computer power we had to use a simple methodology in this study: a simple pH-CE method (see above) and a crude solvent representation in the MD simulations (see below). Therefore, this study does not intend to present an accurate application of the proposed method, but rather to illustrate the general aspects of both the implementation and the analysis of the simulation results. On the other hand, the simple methodology allows us to

examine in detail the effect of different update time intervals of the PMF. We have thus done constant-pH MD simulations using several PMF update intervals, including an exact one with updating at each time step (see below). When possible we have also tried to infer how more accurate methods could have influenced the results.

Molecular Dynamics Simulations

To test the present method we have chosen the small protein BPTI, using as a starting structure the entry 4PTI⁵⁶ in the Protein Data Bank.⁵⁷ The energy minimizations and MD calculations were performed with the program *Discover* (Biosym Technologies, San Diego), whose interface with TITRA (referred to above) was done with programs written by the authors. The analysis of the results was partially performed with the program *Insight II* (Biosym Technologies).

Instead of including water molecules in the simulation, the solvation effects were modeled in an approximate way by using a distance-dependent dielectric equal to $4\pi/r$ (where r is the distance between the charges in the Coulombic interaction) in all operations. The explicit presence of water molecules would tend to damp the movement of surface side chains, and thus fast mean charge changes (due to fast conformational change) would be less likely to occur.^{58,59} Hence, the simple solvation method used here should give an overestimate of the differences between the exact and approximate update of the charges, which actually strengthens our conclusion that the use of large intervals is valid (see below). A pH value of 4.0 was chosen for the constant-pH MD simulations since several titrable sites in BPTI are expected to be partially protonated at this pH, the kind of situation in which this new simulation method is of interest.

The mean charges at pH 4.0 were computed for the crystal structure (using TITRA), which was then relaxed for 100 steps of steepest descents of energy minimization, without updating the charges. An initial equilibration of the structure (simulation hereafter referred to as SEQ) was then performed by running an MD simulation for 50 ps using these same charges, which is expected to allow for an overall adjustment of the structure to the force field parameters, of which only some partial charges may differ in the runs to follow. Velocities were initially chosen from a 300 K Maxwell-Boltzmann distribution and directly scaled for 1 ps; after that the bath method of Berendsen et al.⁶⁰ was used to maintain that temperature for the rest of the simulation. The equations of motion were solved by the leapfrog form of the Verlet⁶¹ algorithm, with an integration time step of 1 fs. A group-based cutoff radius of 10.5 Å was applied to non-bonded interactions.

Several constant-pH MD simulations of 50 ps were continued from the SEQ simulation, using different update intervals of the partial charges, namely, 1 fs (exact PMF), 10 fs, 100 fs and 1,000 fs (simulations hereafter referred to as S1, S10, S100, S1000). The simulation conditions were the same as those used for SEQ. Mean protein conformations from these constant-pH MD simulations were obtained by averaging over conformations taken at 1 ps intervals, and submitting the resulting average to 200 steps of steepest descents.

For comparison with the constant-pH MD runs a normal constant-**n** MD simulation was performed using the usual protonation states (i.e., charged Asp, Glu, Arg, Lys, and C- and N-termini). The simulation was started from the crystal structure and run for 100 ps, using the same simulation conditions as above. This simulation will be referred to as SNMD.

RESULTS

Figure 1 displays the temporal evolution of the potential energy of the system at fixed configuration, $\langle U \rangle_q$, given in Eq. (14). Although the potential energy of the system being simulated is the PMF W , $\langle U \rangle_q$ partially characterizes the energetics of the system and can be used as an indication of its equilibration. After an initial small descent from the final value of SEQ, the values from all constant-pH MD simulations seem to have stabilized. A slightly lower curve is obtained for the potential energy U of the control simulation SNMD (not shown), which stabilizes after 40 ps.

Figure 2 shows the root mean squared (RMS) deviation of all α -carbons between instantaneous conformations and the crystal structure. Except for S100, all constant-pH MD simulations have clearly stabilized its conformation (see Discussion). The RMS deviations computed using all heavy atoms give very similar curves, raised by 0.5–1.0 Å (not shown). The RMS deviations for SNMD (not shown) follow the SEQ and S1–1,000 curves closely, showing that the large deviation from the crystal structure is caused by the crude solvation method (distance-dependent dielectric constant) and not by the use of the PMF.

In Table I we show the RMS deviations obtained when superimposing the crystal and the constant-pH and constant-**n** MD mean conformations. Both RMS deviations of α -carbons and heavy atoms indicate that the conformational region sampled by all MD simulations is equally far from the crystal structure, although the constant-pH conformations are much more similar among themselves than with the SNMD one. Indeed, the distance of the S1–1000 structures seems to be nearly the same from the crystal and SNMD structures, specially when all heavy atoms are considered.

Figure 3 shows the fluctuations of the α -carbon positions along the amino acid sequence, for simula-

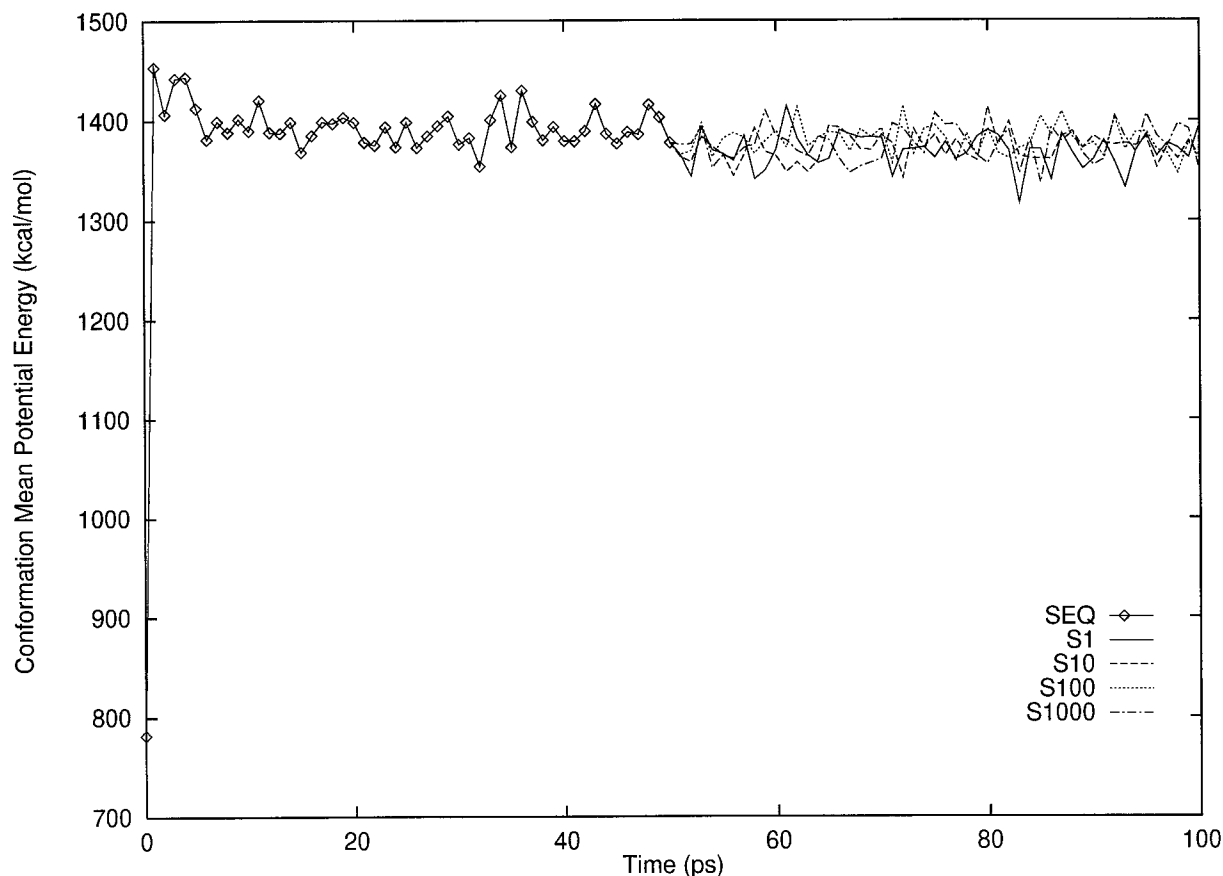


Fig. 1. Temporal evolution of the mean potential energy at fixed configuration, $\langle U_q \rangle$, for simulations SEQ, S1, S10, S100, and S1,000.

tions SNMD and S1. The global fluctuations of the α -carbons is 0.741 Å for SNMD and 1.302 Å for S1. The other constant-pH simulations also give fluctuations consistently higher than the SNMD ones (not shown), with global values of 0.854, 0.995, and 0.909 for S10, S100 and S1,000, respectively. The fluctuations of the heavy atoms in the constant-pH simulations were also consistently higher than the SNMD ones (not shown).

The partially titrated sites at pH 4.0, i.e., those that display non-integer values of $\langle n_i \rangle_q$ during the constant-pH MD runs, are Asp 3, Glu 7, Glu 49, Asp 50, and the C-terminus at residue 58 (C_{tr} 58). These anionic sites have a formal charge $z_i = n_i - 1$. Figures 4–8 show the instantaneous $\langle z_i \rangle_q$ of these titrating sites, for simulations S1–1,000. The same size for the charge range was chosen in all plots for ease of comparison. All points obtained are shown, i.e., as one goes from plots (a) to (d) in each figure, the number of plotted points decreases from 50,000 to 50 by factors of 10.

Table II contains the (non-conditional) mean charges $\langle z_i \rangle$ for the titrating sites, obtained from the constant-pH MD simulations S1–1,000 [see Eq. (10)]. Also shown are the crystal (conditional) mean charges

$\langle z_i \rangle_q$ as obtained from a single pH-CE calculation using the crystal conformation. In addition to the values for the partially titrated sites, the table also shows the total charge of the protein.

In Table III we show the charge fluctuations corresponding to simulation S1. For easier localization of the corresponding range, the mean values from Table II are also shown. Comparison of these individual site fluctuations with the ones obtained in simulations S10–1,000 is superfluous, since in this study they can be directly obtained from the mean values (see section pH-CE calculations above), whose comparison can be done using Table II.

Table IV shows the correlation coefficients between the $\langle n_i \rangle_q$ of different sites, for simulation S1. All correlations are very small, specially the positive ones. Identical results (not shown) are observed for simulation S10–1,000.

DISCUSSION

Exact PMF

Although simulation S1 corresponds to the use of the exact PMF, its exactness is only relative to the pH-CE and MD methodologies being used, which in the present case are not expected to be highly

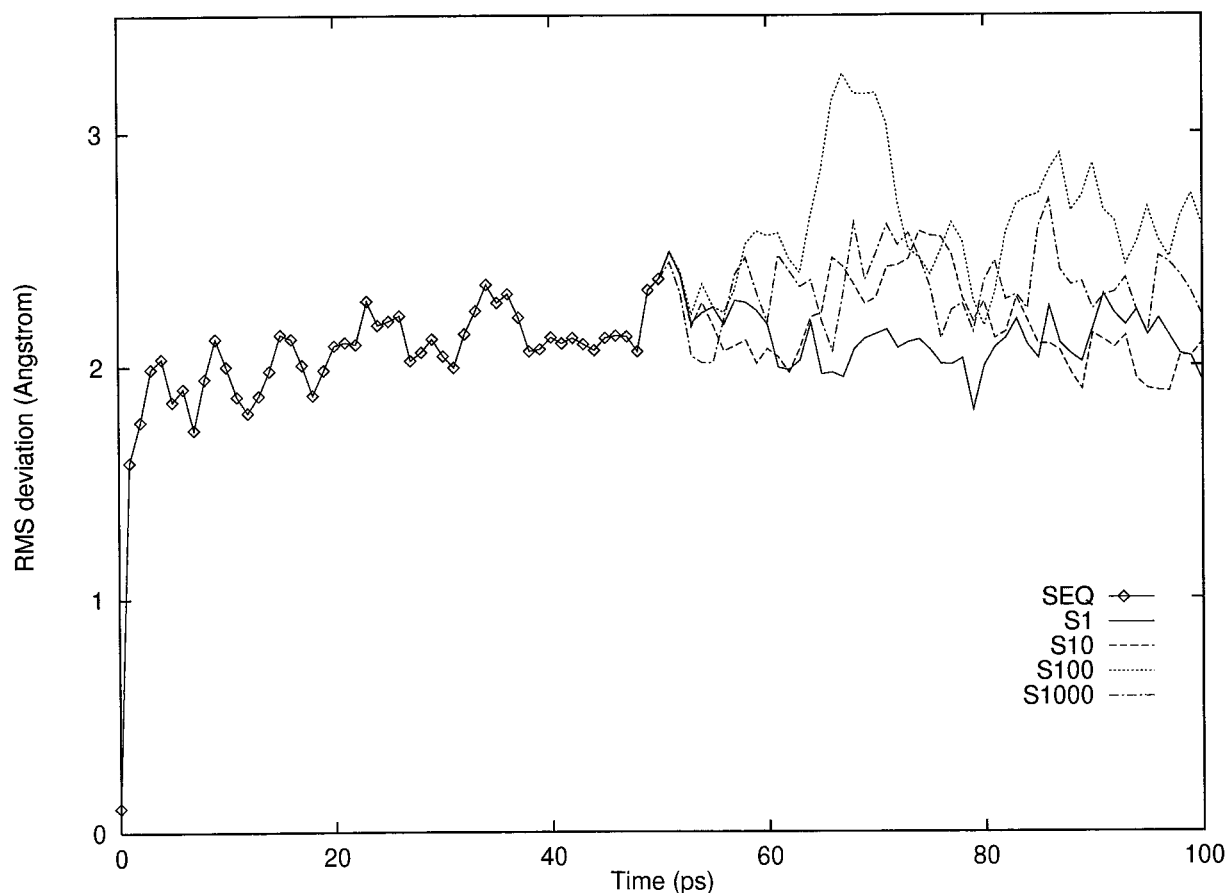


Fig. 2. Temporal evolution of the RMS deviation between α -carbons of instantaneous conformations and the crystal structure, computed after optimal fitting of α -carbons, for simulations SEQ, S1, S10, S100, and S1,000.

TABLE I. RMS Deviations Between Crystal and MD Mean Conformations*

	Crystal	S1	S10	S100	S1,000	SNMD
Crystal		2.534	2.604	2.938	2.686	2.791
S1	2.007		0.610	1.416	0.972	2.506
S10	2.047	0.360		1.505	1.013	2.652
S100	2.461	1.195	1.233		1.167	2.989
S1000	2.177	0.725	0.765	0.897		2.810
SNMD	2.168	1.676	1.805	2.428	2.021	

*Obtained using α -carbons (lower left) and heavy atoms (C, N, O, S) (upper right) for both the optimal fitting and the RMS calculation.

accurate (see section Implementation of the PMF above). However, several general trends can be inferred from the present results, based on the known differences between the methods used here and more accurate ones.

The RMS deviation from the crystal (Fig. 2 and Table I) seems to be caused by the simulation conditions and not by the use of the PMF, since the simulation SNMD shows an identical deviation, which is typical of simulations using a distance-

dependent dielectric. Interestingly, a similar deviation exists between S1 and SNMD, suggesting that the inclusion of pH effects may be as important in MD simulations as the inclusion of the solvent.

The atomic fluctuations are higher in the constant-pH system, as expected from Eq. (13). Some rare exceptions in which the SNMD fluctuation is slightly higher (which occurs for 4 residues in Fig. 3) may be due to statistical error of the sample.

The fluctuations shown by the $\langle z_i \rangle_q$ values in the exact PMF simulation [plot (a) in Figs. 4–8] vary with the site considered. There seems to be a tendency to smaller fluctuations for sites near full deprotonation (Asp 50, Ctr 58). This is to be expected, since the charge of sites near full protonation or deprotonation is less sensitive to alterations in the effective pK_a (due to its approximate sigmoidal dependence), here caused by conformational change. The highest observed fluctuation range is about 0.3 charge units, for Asp 3. The explicit inclusion of water in the simulations would probably increase these fluctuations, since this is known to lead to conformational fluctuations of higher magnitude^{58,59} (although

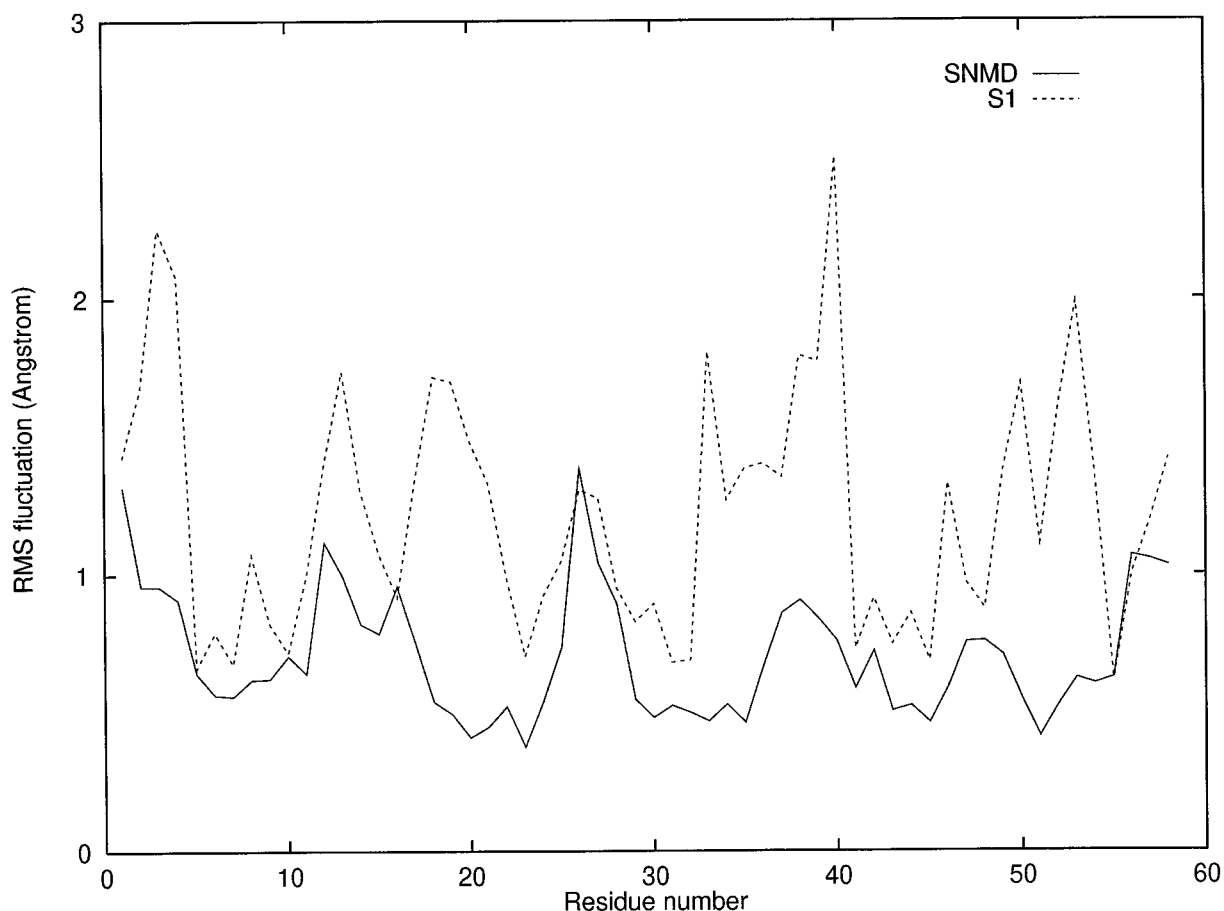


Fig. 3. RMS fluctuation of α -carbon positions along the amino acid sequence, for simulations SNMD and S1.

slower, as noted above). Another possible cause for different fluctuation magnitudes would be the inclusion of electrostatic self-energies when computing $\Delta A^\circ(\mathbf{n}, c)$ [Eq. (4)], as is done by more detailed pH-CE methods.^{4,5} Different self-energy terms will correspond to different conformations; such terms will sum up to the site-site interactions here considered. This should increase the overall $\langle z_{ij} \rangle_q$ fluctuations, unless a reduction of conformational fluctuations also results, which could compensate that effect. However, we expect precisely the opposite, since the burying/exposure of charges would be easier, as discussed above. As a simple comparison of the sensitivity of different pH-CE methods on conformational differences, we used TITRA to compute the apparent pK_a s of sperm whale carbonmonoxy myoglobin, using the same two conformations used by Bashford et al.,¹⁴ and we observed differences much smaller than these workers (not shown). This is probably due to our neglect of self-energy terms and, eventually, to the MF approximation used.

Another interesting aspect visible in plot (a) of Figures 4–8 is that different titrable sites show

different fluctuation patterns of $\langle z_{ij} \rangle_q$. For example, the fluctuations for Asp 3 are generally much faster than the ones for Glu 49, as can be confirmed by comparing their autocorrelation functions (not shown). This probably reflects the particular environment of the site, and a full understanding of this fact would require a more extensive analysis of the conformations.

The mean charges $\langle z_i \rangle$ obtained from the S1 simulation (in Table II) are not very different from the ones directly obtained with the crystal conformation, the larger difference being 0.152 units, for Ctr 58. However, inclusion of the crystal values in the table is simply illustrative, since a quantitative comparison between simulation and experiment would be meaningful only with more accurate simulation methods.

The charge fluctuations in Table III are the true fluctuations in the real system, $\sigma(z_i)$, and should not be confused with the fluctuations seen in Figures 4–8(a), $\sigma(\langle z_{ij} \rangle_q)$ [the two fluctuations are related through an equation analogous to (13), with q replacing \mathbf{n}]. If a more detailed pH-CE method is used $\sigma(\langle z_{ij} \rangle_q)$ is expected to be larger, as noted above,

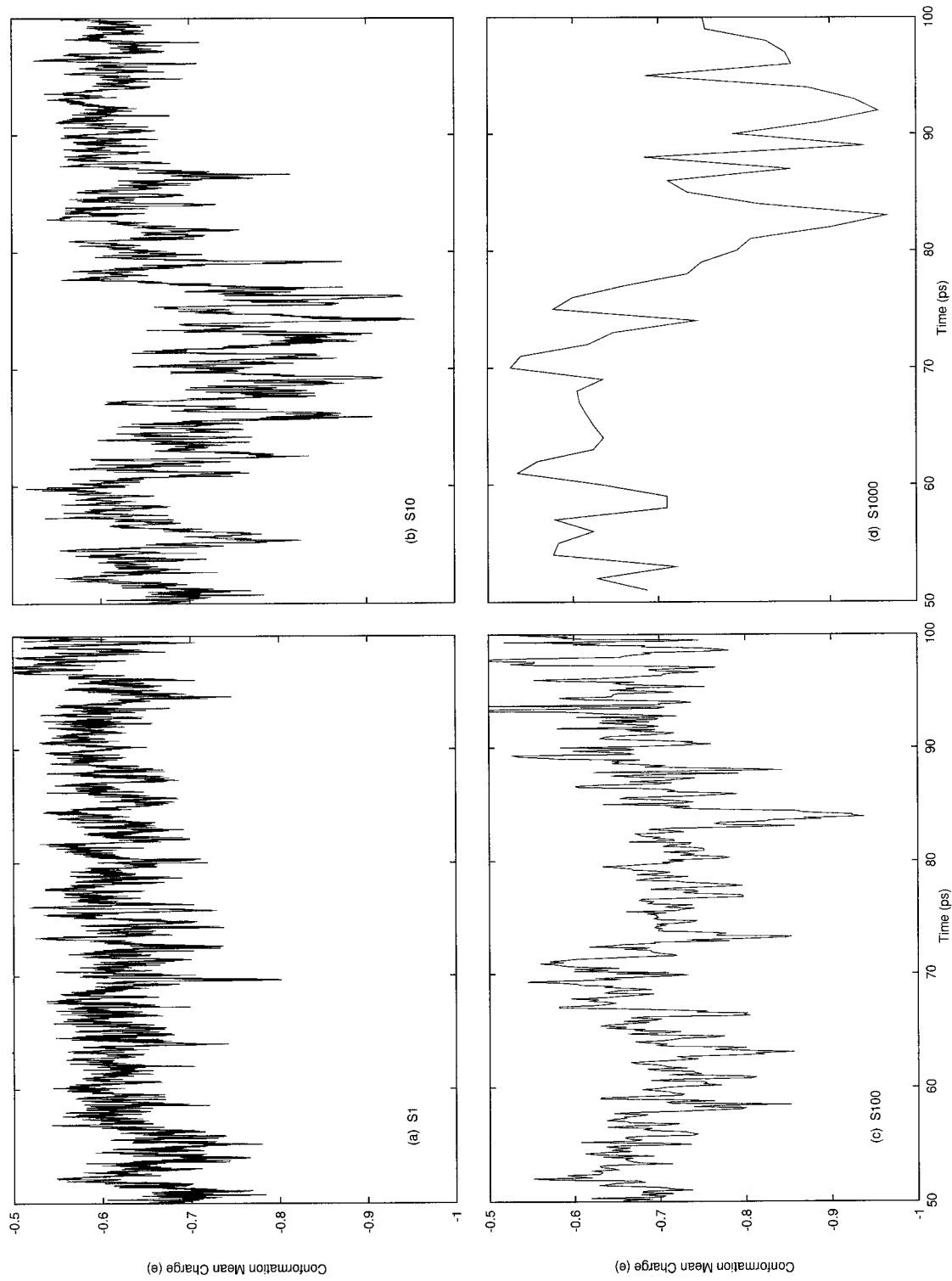


Fig. 4. Temporal evolution of the mean charge at fixed conformation, $\langle Z_{q_i} \rangle$, of site Asp 3 for simulations S1 (a), S10 (b), S100 (c), and S1,000 (d).

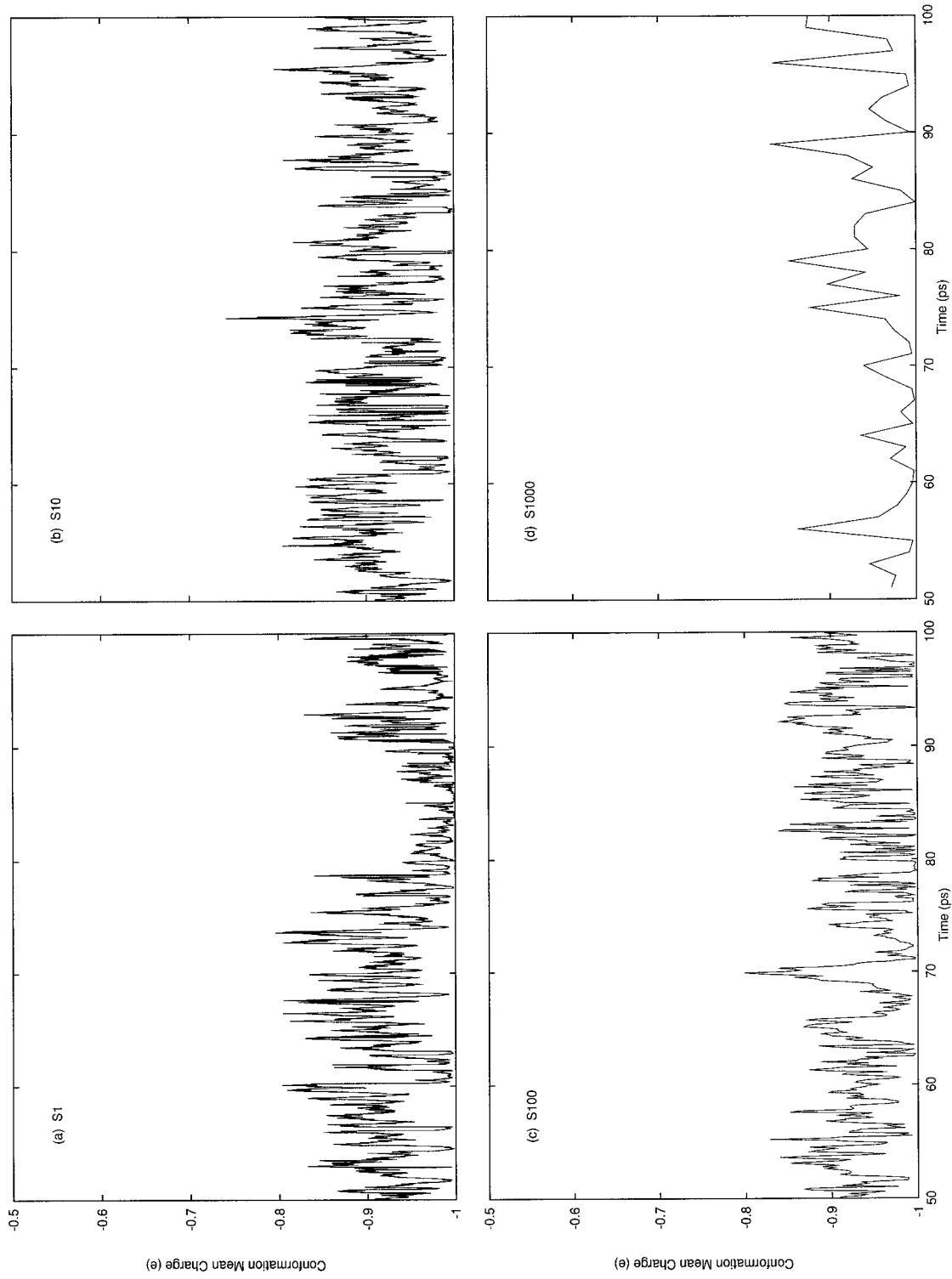


Fig. 5. Temporal evolution of the mean charge at fixed conformation, $\langle z_{\text{Glu 7}} \rangle$ of site Glu 7 for simulations S1 (a), S10 (b), S100 (c), and S1,000 (d).

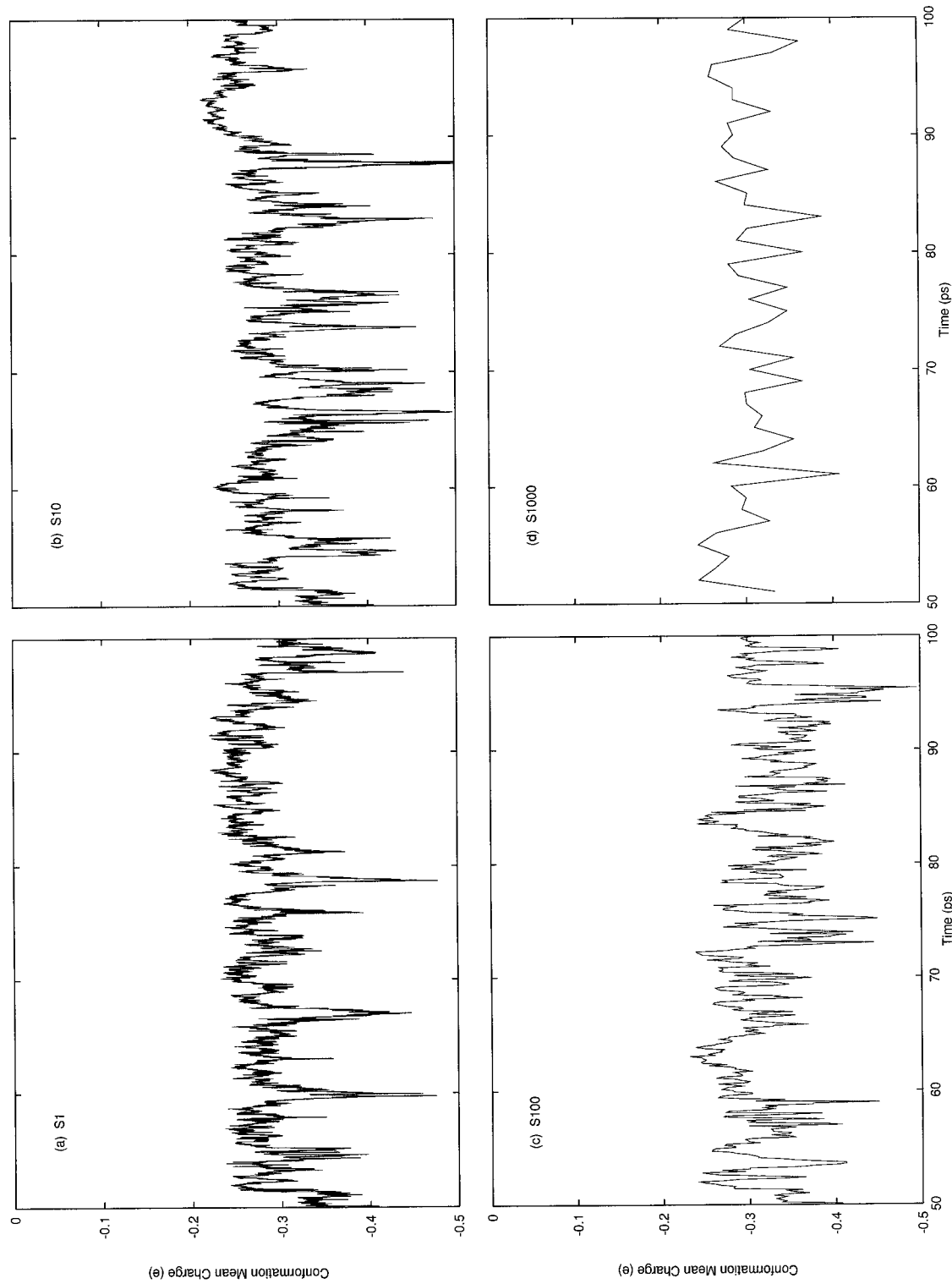


Fig. 6. Temporal evolution of the mean charge at fixed conformation, $\langle Z_{q_i} \rangle$, of site Glu 49 for simulations S1 (a), S10 (b), S100 (c), and S1,000 (d).

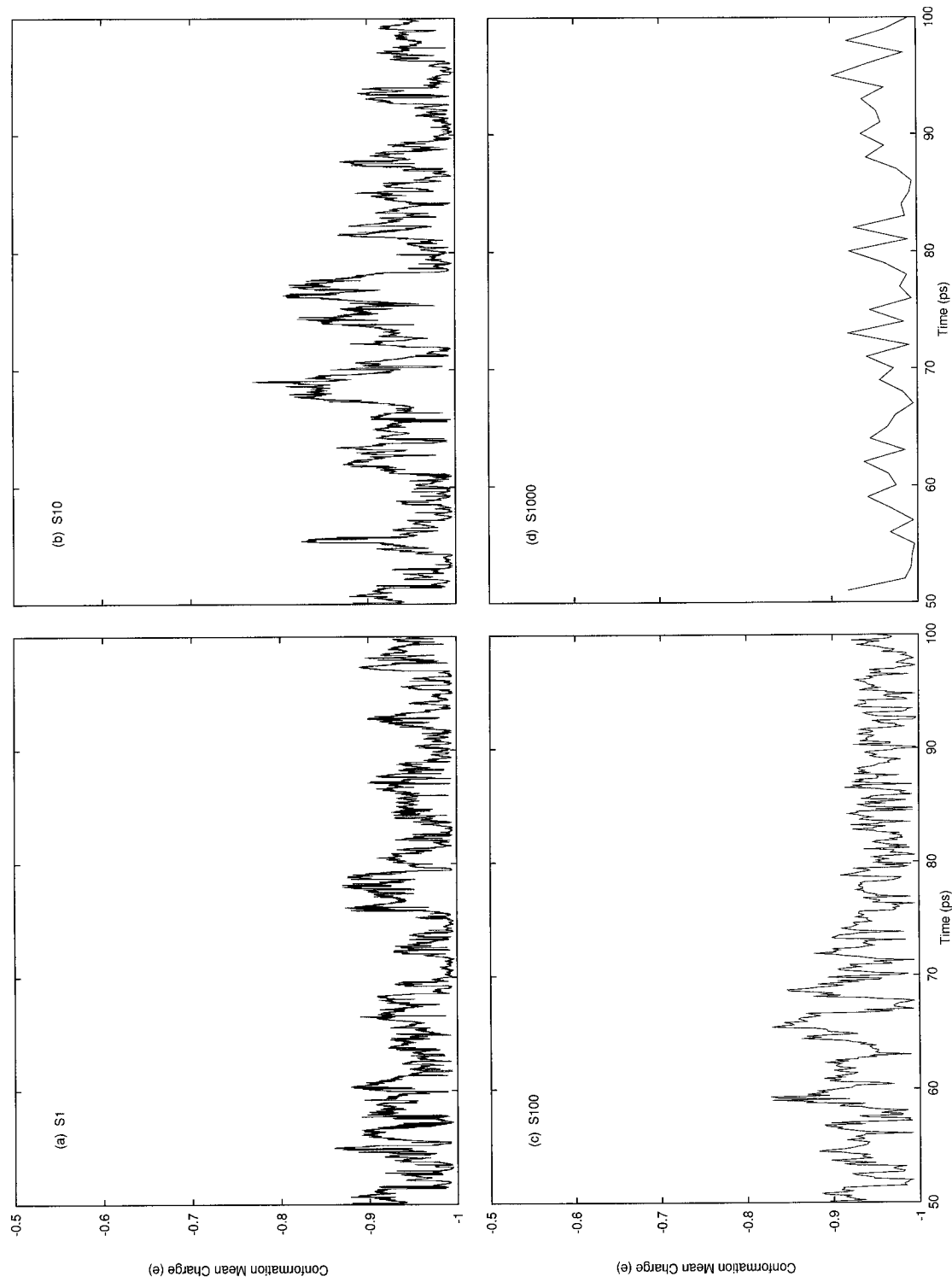


Fig. 7. Temporal evolution of the mean charge at fixed conformation, $\langle Z_{eq} \rangle$, of site Asp 50 for simulations S1 (a), S10 (b), S100 (c), and S1,000 (d).

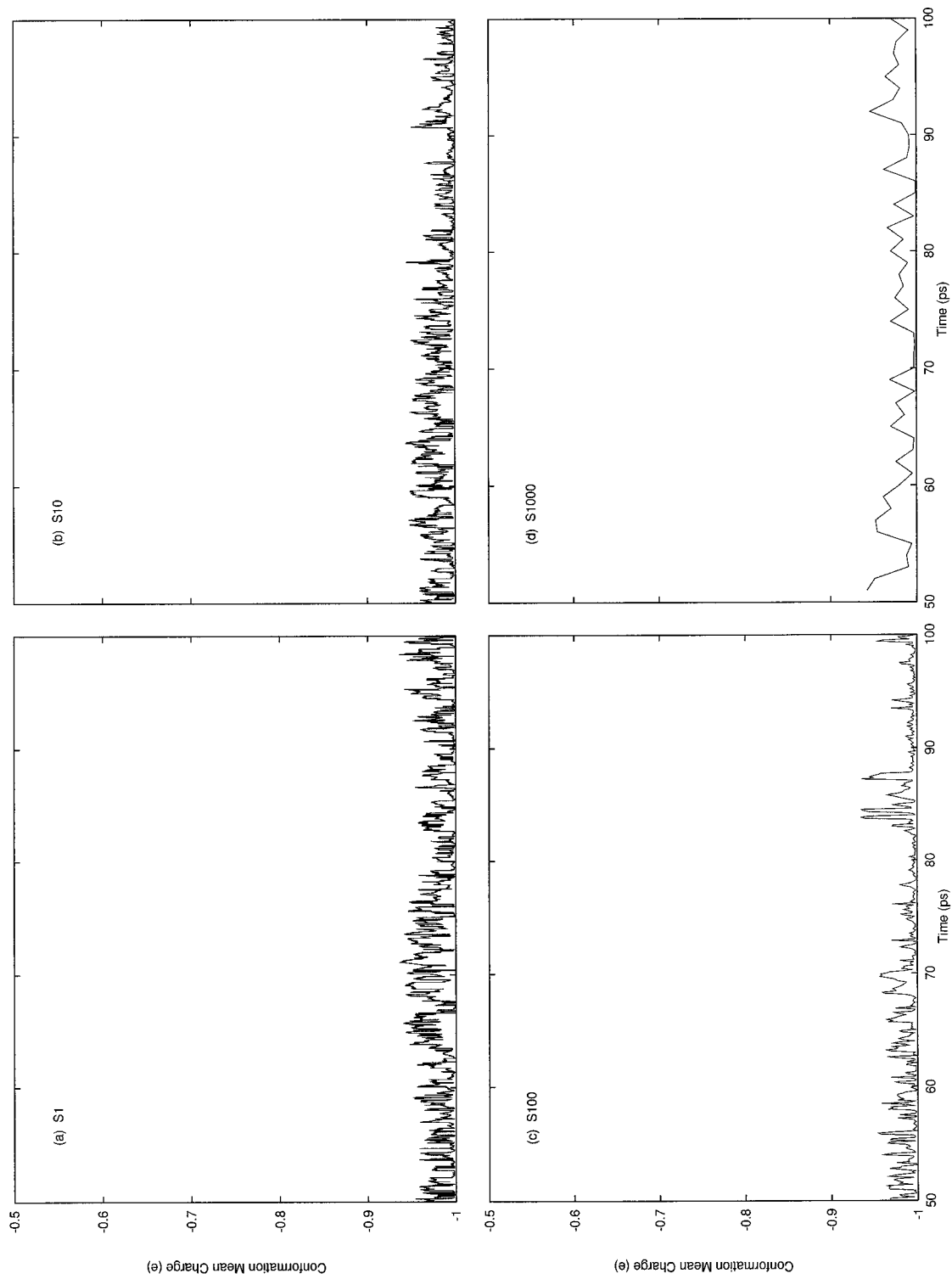


Fig. 8. Temporal evolution of the mean charge at fixed conformation, $\langle z_q \rangle$, of site Ctr 58 for simulations S1 (a), S10 (b), S100 (c), and S1,000 (d).

TABLE II. Site and Total mean Charges at pH 4, Computed From the Crystal Conformation and Constant-pH MD Simulations

Site	Crystal	S1	S10	S100	S1,000
Asp 3	-0.637	-0.626	-0.671	-0.690	-0.713
Glu 7	-0.803	-0.923	-0.917	-0.934	-0.953
Glu 49	-0.353	-0.290	-0.295	-0.321	-0.305
Asp 50	-0.885	-0.948	-0.938	-0.939	-0.965
Ctr 58	-0.826	-0.978	-0.984	-0.988	-0.980
Total	7.496	7.235	7.194	7.128	7.085

TABLE III. Site and Total Mean Charges and Fluctuations at pH 4, Computed From Constant-pH MD Simulation S1

Site	$\langle z \rangle$	$\sigma(z)$
Asp 3	-0.626	0.484
Glu 7	-0.923	0.266
Glu 49	-0.290	0.454
Asp 50	-0.948	0.223
Ctr 58	-0.978	0.148
Total	7.235	0.762

TABLE IV. Correlation Coefficients Between the Conditional Mean Protonation of Different Titrable Sites for Simulation S1

Glu 7	0.008			
Glu 49	0.087	-0.238		
Asp 50	-0.024	0.061	-0.400	
Ctr 58	-0.349	0.168	0.026	-0.065
	Asp3	Glu7	Glu49	Asp50

leading to higher values of $\sigma(z_i)$ [see Eq. (13)]. In this way the fluctuation of the total charge will also be increased, although this effect may be attenuated by negative site-site correlations. We note that the third column in Table III should not be seen as a measure of the error range in the $\langle z \rangle$ values, as in $\langle z \rangle \pm \sigma(z)$, because the distribution of z is not symmetrical. That interpretation does in fact lead to absurd results for all sites (values that are positive or lower than -1).

The results in Table IV indicate a general lack of correlation between the $\langle n_i \rangle_q$ values, with a maximum absolute coefficient of 0.4. The positive correlations are extremely low, as one should expect in case the MF approximation is valid (see section Force field). Even in the case of the highest correlation, Glu 7/Ctr 58, the sites are widely separated (21.1 Å between the carboxylic carbons in the mean structure), showing that some residual correlation may appear without any coupling between the sites. In the case of the two highest negative correlations, Asp 3/Ctr 58 and Glu 49/Asp 50, the distance is also large (11.7 Å and 8.5 Å), but a third residue of opposite charge can be identified in an approximately intermediate position, Arg 1 (and also its Ntr site) and Arg 53,

respectively. The fluctuations of these positively charged sites often bring them closer to one of the negative sites and farther from the other, thus giving rise to the slight negative correlation. In conclusion, the measured correlations indicate that the coupling between the sites titrating in the considered pH region is very small and therefore the MF approximation seems to be valid.

Different PMF Update Intervals

From Figures 1 and 2, the only simulation whose equilibration may be doubtful is S100. It is difficult to judge if the large fluctuation around 70 ps is due to non-equilibration or is rather a fortuitous fluctuation of an equilibrated conformation. In the former case, only times after 80 ps should have been used when computing the mean conformation in Table I. However, the simulation time is too short to draw any valid conclusions. In any case, it does not seem to exist any systematic effect caused by the increasing of the update interval, such as increase or decrease of the RMS deviations in Table I, or of the values or fluctuations in Figures 1 and 2.

The clustering of the constant-pH simulations seen in Table I suggests that the perturbation of the atomic trajectories caused by the larger update intervals is not sufficient to originate any drastic rearrangements of the protein structure (at least with the present methodology). Also, the global α -carbon fluctuations from simulations S10–1,000 are very similar and consistently higher than the SNMD values. Nevertheless, although the increase of update interval does not produce any obvious trend on the α -carbon fluctuations, the values from S1 are generally higher than the S10–1,000 values, which may indicate a more efficient sampling with an exact update. This seems intuitively reasonable, since a less frequent update should “disturb” the trajectory less and produce a constant- n -like local sample (although the fractional charges do not correspond to a true n state), with smaller local fluctuations. Thus, only longer simulations with a large update interval will eventually yield the same fluctuations as the exact one. The interesting point is that this retarding effect, if it exists, seems to be nearly the same with an interval of 10, 100, and 1,000 MD steps.

Plots (b)–(d) in Figures 4–8 show a pattern of fluctuations similar to the exact simulation ones in (a), except for Asp 3. For this site the fluctuations seem to be increased in the three cases, S10–1,000, with values spread over a range of almost half charge unit. For simulations S10 and S100 the curves of *all* sites follow the S1 ones for the first picoseconds, diverging around 5 ps. (In the case of S1,000 it is more difficult to make this comparison, due to the lack of points.) These results show that the perturbation induced by the large update intervals in the trajectory is small enough to be significant

only after 5 ps, even with an update interval 100 times larger than the exact "interval." In our opinion this is a very strong indication that the use of the same mean charges for several time steps is a valid approximation to generate good sampling trajectories. In particular, it seems to rule out the potential occurrence of sudden forces when updating charges, referred to above (see section PMF update interval). This danger may, however, be more serious when including electrostatic self-energies, due to higher charge fluctuations (see above). On the other hand, if solvent is also explicitly included, the fluctuation rate will probably decrease (see above), and it will be less likely that large differences occur at successive updates.

To make a good comparison of the $\langle z_i \rangle$ s from S10–1,000 with the S1 ones (Table II), we would need to know the spread of the latter, i.e., we would need several exact simulations. In any case, the S10–1,000 tend to be closer to the S1 values than to the crystal ones.

In conclusion, the use of a large update interval for the PMF does not seem to produce any consistent trend on the results. The atomic RMS deviations suggest that the explored regions of conformations are the same or very close. Also, the perturbation induced on the "trajectories" of $\langle z_i \rangle_q$ seems to be very small (at least up to an update at each 100 steps), so that their sampling is probably similar. Together, these facts suggest a similar overall sampling, and the use of large update interval seems to be valid, although the sample may be more efficient with an exact update.

CONCLUSIONS

In the present paper we have presented a novel method that allows for the incorporation of pH effects into MD simulations. This constant-pH MD method does a correct simultaneous sampling of charges and conformations at a given pH value, which has previously not been possible with other simulation methods. The indication that the use of large PMF update intervals is valid (see above) implies that, if simple pH-CE methods are used, the CPU needs are not much different from those of usual MD simulations. This also means that more accurate pH-CE methods and MD simulation methods can be used with this approach, at the expense of a moderate increase of computation time.^{4,5} The present method may be particularly easy to implement with CE-based implicit solvation methods, which already have an interface between the "original" force field and the CE calculations.^{21–29}

The present constant-pH method can be used as a tool in the simulation of protein systems, where it may be a valuable help in the understanding of the effect of pH on function and stability. It may be used to gain insight into the mechanisms responsible for known pH-induced conformational changes, or rather

as a predictive tool guiding the engineering of a protein when the modification of pH-dependent properties (e.g., binding or catalysis) is intended. Its use is not essentially different from the usual MD, the main difference being that the final set of conformations will correctly reflect both temperature and pH. A new type of simulation studies that can be performed is the pH-induced (partial) unfolding of proteins, which should be possible with the current methods and computer power. These unfolding simulations could be an interesting complement to the increasing number of temperature-induced ones,⁶² and also to the recent stability studies based on pH-CE methods.^{15,16}

Besides its use in constant-pH MD, the PMF method here presented can be used in energy minimizations, such as in the geometry optimization of experimental structural data. It should be pointed out that many X-ray diffraction and NMR studies of protein structures have actually been performed at pH conditions deviating markedly from neutrality. Thus, the standard assumption that charge assignment to titrable sites can assume (model) pH neutral conditions is obviously a severe approximation for, e.g., pH 4, which would be a typical value for many NMR studies. By using the present PMF, the pH value can be supplied as one of the parameters to the minimization, which will then properly reflect the experimental conditions.

The present method can also be useful to do conformational searching, i.e., to explore conformational space in a more extensive way, without trying to reflect any particular physical parameters such as pH. The effect of using the PMF as the potential energy is that the charges on titrable groups "adapt" to their particular environment, thereby making the energy surface smoother than it would be using fixed charges (even an all-neutral charge set). The extent of this adaptation is determined by the chosen pH and temperature, which could thus be used as control parameters in the searching. Different pH values could in principle be used to overcome energy barriers involving different titrable groups.

The apparent pK_a s for the titrable sites (i.e., the pH values for which $\langle n_i \rangle = 0.5$) cannot be obtained from a single constant-pH MD simulation, which yields only the "effective" pK_a s that characterize the protonation equilibrium of each group at the pH considered. To determine the apparent pK_a s it is necessary to perform several simulations at different pH values, similarly to what is done with usual pH-CE methods. In this respect the present method can be seen as an improved pH-CE method that takes conformational freedom into account, like the method of You and Bashford.⁴⁰ However, the present method is primarily an MD simulation method, and its use for full pK_a calculations in proteins is still too demanding for current computer power. For such studies the method of You and Bashford⁴⁰ is at

present the only feasible route, since, in spite of some arguable aspects of the conformational sampling (see Introduction), it can obviously "relieve" some protonation/conformation conflicts.

In this study we used a force field conveniently modified to make the implementation of the method as simple as possible. Therefore, since the original force field was parameterized for a different energy function (without electronic lone pairs) and for different simulation conditions (with explicit water molecules), its reliability may have been somewhat impaired by these and hoc modifications. Even though these minor modifications are probably unimportant for qualitative studies, properly parameterized force fields (e.g., where explicit electronic lone pairs are originally considered) should be used when quantitative prediction is intended. Also, the MF approximation assumed for the titration at fixed conformation can be removed, although it appreciably simplifies the implementation of the PMF and it is known to affect the results only when strongly interacting sites titrate at similar pH values⁴⁶; moreover, its validity can be checked with the consistency test used here. It is important to note that the MF approximation can in principle be eliminated by an appropriate change on the calculation of Coulombic terms in the MD program, *without* any significant increase of computing time (see section, Force Field). The $\langle Q_{a|q} \rangle$ and $\langle Q_a Q_{b|q} \rangle$ values necessary for such implementation [see Eq. (17)] can be simultaneously obtained from, e.g., a Monte Carlo sampling of the protonation states.^{5,36,45,47} Finally, we note that the way these quantities are computed is irrelevant for the PMF update mechanism, so that the use of spherical or atomic-level models for the CE calculations is essentially identical in terms of implementation; the major difference will be, of course, the computation time. Work along these lines is in progress.

Although we have focused on proteins, the present method is applicable to any flexible molecule with titrable sites.

ACKNOWLEDGMENTS

A.M.B. and P.J.M. thank the Junta Nacional de Investigação Científica e Tecnológica, Portugal, for grants.

REFERENCES

1. Linderstrøm-Lang, K. On the ionization of proteins. *C.R. Trav. Lab. Carlsberg* 15:1–29, 1924.
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. Matthew, J.B. Electrostatic effects in proteins. *Annu. Rev. Biophys. Biophys. Chem.* 14:387–417, 1985.
4. Bashford, D., Karplus, M. pK_a's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. *Biochemistry* 29:10219–10225, 1990.
5. 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.
6. Rogers, N.K. The modelling of electrostatic interactions in the function of globular proteins. *Prog. Biophys. Mol. Biol.* 48:37–66, 1986.
7. Harvey, S.C. Treatment of electrostatic effects in macromolecular modeling. *Proteins* 5:78–92, 1989.
8. van Gunsteren, W.F., Berendsen, H.J.C. Computer simulations of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Eng.* 29:992–1023, 1990.
9. Sharp, K.A., Honig, B. Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* 19:301–332, 1990.
10. Davis, M.E., McCammon, J.A. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* 90:509–521, 1990.
11. Martel, P.J., Baptista, A., Petersen, S.B. Protein electrostatics. *Biotechnol. Annu. Rev.* 2:315–372, 1996.
12. Rashin, A.A., Bukatin, M.A. A view of thermodynamics of hydration emerging from continuum studies. *Biophys. Chem.* 51:167–192, 1994.
13. Bashford, D., Gerwert, K. Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* 224:473–486, 1992.
14. Bashford, D., Case, D.A., Dalvit, C., Tennant, L., Wright, P.E. Electrostatic calculations of side-chain pK_a values in myoglobin and comparison with NMR data for histidines. *Biochemistry* 32:8045–8056, 1993.
15. Yang, A.-S., Honig, B. On the pH dependence of protein stability. *J. Mol. Biol.* 231:459–474, 1993.
16. Yang, A.-S., Honig, B. Structural origins of pH and ionic strength effects on protein stability. Acid denaturation of sperm whale apomyoglobin. *J. Mol. Biol.* 237:602–614, 1994.
17. MacKerell Jr., A.D., Sommer, M.S., Karplus, M. pH dependence of binding reactions from free energy simulations and macroscopic continuum electrostatic calculations: Application to 2'GMP/3'GMP binding to ribonuclease T₁ and implications for catalysis. *J. Mol. Biol.* 247:774–807, 1995.
18. Wendoloski, J.J., Matthew, J.B. Molecular dynamics effects on protein electrostatics. *Proteins* 5:313–321, 1989.
19. March, K.L., Maskalik, D.G., England, R.D., Friend, S.H., Gurd, F.R.N. Analysis of electrostatic interactions and their relationship to conformation and stability of bovine pancreatic trypsin inhibitor. *Biochemistry* 21:5241–5251, 1982.
20. Honig, B., Nicholls, A. Classical electrostatics in biology and chemistry. *Science* 268:1144–1149, 1995.
21. Davis, M.E., McCammon, J.A. Calculating electrostatic forces from grid-calculated potentials. *J. Comput. Chem.* 11:401–409, 1990.
22. Still, W.C., Tempczyk, A., Hawley, R.C., Hendrickson, T. Semianalytic treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129, 1990.
23. Zauhar, R.J. The incorporation of hydration forces determined by continuum electrostatics into molecular mechanics simulations. *J. Comput. Chem.* 12:575–583, 1991.
24. Sharp, K. Incorporating solvent and ion screening into molecular dynamics using the finite-difference Poisson-Boltzmann method. *J. Comput. Chem.* 12:454–468, 1991.
25. Gilson, M.K., Honig, B. The inclusion of electrostatic hydration energies in molecular mechanics calculations. *J. Comput. Aided Mol. Des.* 5:5–20, 1991.
26. Niedermeier, C., Schulten, K. Molecular dynamics simulations in heterogeneous dielectrica and Debye-Hückel media: Application to the protein bovine pancreatic trypsin inhibitor. *Mol. Simulation* 8:361–387, 1992.
27. Gilson, M.K., Davis, M.E., Luty, B.A., McCammon, J.A. Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. *J. Phys. Chem.* 97:3591–3600, 1993.
28. Smith, K.C., Honig, B. Evaluation of the conformational free energies of loops in proteins. *Proteins* 18:119–132, 1994.
29. Abagyan, R., Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* 235:983–1002, 1994.

30. Gilson, M.K., Honig, B.H. The dielectric constant of a folded protein. *Biopolymers* 25:2097–2119, 1986.
31. Antosiewicz, J., McCammon, J.A., Gilson, M.K. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* 238: 415–436, 1994.
32. Simonson, T., Perahia, D., Bricogne, G. Intramolecular dielectric screening in proteins. *J. Mol. Biol.* 218:859–886, 1991.
33. Simonson, T., Perahia, D., Brünger, A. Microscopic theory of the dielectric properties of proteins. *Biophys. J.* 59:670–690, 1991.
34. 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.
35. Brady, G.P., Sharp, K.A. Decomposition of interaction free energies in proteins and other complex systems. *J. Mol. Biol.* 254:77–85, 1995.
36. Gilson, M.K. Multiple-site titration and molecular modeling: Two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins* 15:266–282, 1993.
37. Beveridge, D.L., DiCapua, F.M. Free energy via molecular simulation: Application to chemical and biomolecular systems. *Annu. Rev. Biophys. Chem.* 18:431–492, 1989.
38. 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.
39. Del Buono, G.S., Figueirido, F.E., Levy, R.M. Intrinsic pK_as of ionizable residues in proteins: An explicit solvent calculation for lysozyme. *Proteins* 20:85–97, 1994.
40. You, T.J., Bashford, D. Conformation and hydrogen ion titration of proteins: A continuum electrostatic model with conformational flexibility. *Biophys. J.* 69:1721–1733, 1995.
41. Allen, M.P., Tildesley, D.J. "Computer Simulation of Liquids." Oxford: Clarendon, 1987.
42. Hill, T.L. "Statistical Mechanics." New York: McGraw-Hill, 1956.
43. Tanford, C., Roxby, R. The interpretation of protein titration curves. Application to lysozyme. *Biochemistry* 11:2192–2198, 1972.
44. Shire, S.J., Hanania, G.I.H., Gurd, F.R.N. Electrostatic effects in myoglobin. Hydrogen ion equilibria in sperm whale ferrimyoglobin. *Biochemistry* 13:2967–2974, 1974.
45. Antosiewicz, J., Porschke, D. The nature of protein dipole moments, experimental and calculated permanent dipole of α -chymotrypsin. *Biochemistry* 28:10072–10078, 1989.
46. Bashford, D., Karplus, M. Multiple-site titration curves of proteins: An analysis of exact and approximate methods for their calculation. *J. Phys. Chem.* 95:9956–9961, 1991.
47. Beroza, P., Fredkin, D.R., Okamura, M.Y., Feher, G. Protein 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. USA* 88:5804–5808, 1991.
48. Gilson, M.K., Straatsma, T.P., McCammon, J.A., Ripoli, D.R., Faerman, C.H., Axelsen, P.H., Silman, I., Sussman, J.L. Open "back door" in a molecular dynamics simulation of acetylcholinesterase. *Science* 263:1276–1278, 1994.
49. Ben-Naim, A. "Statistical Thermodynamics for Chemists and Biochemists." New York: Plenum Press, 1992.
50. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52:255–268, 1984.
51. Mood, A.M., Graybill, F.A., Boes, D.C. "Introduction to the Theory of Statistics." 3rd edit. New York: McGraw-Hill, 1974.
52. Hagler, A.T., Stern, P.S., Sharon, R., Becker, J.M., Naider, F. Computer simulation of the conformational properties of oligopeptides. Comparison of theoretical methods and analysis of experimental results. *J. Am. Chem. Soc.* 101: 6842–6852, 1979.
53. Barlow, D.J., Thornton, J.M. Ion-pairs in proteins. *J. Mol. Biol.* 168:867–885, 1983.
54. Anthonsen, H.W., Baptista, A., Drablos, F., Martel, P., Petersen, S.B. The blind watchmaker and rational protein engineering. *J. Biotechnol.* 36:185–220, 1994.
55. Russell, 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.
56. Marquart, M., Walter, J., Deisenhofer, J., Bode, W., Huber, R. The geometry of the active site and of the peptide groups in trypsin, trypsinogen and complexes with inhibitors. *Acta Crystallogr.* 39:480, 1983.
57. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Jr., E.F.M., 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.
58. McCammon, J.A., Harvey, S.C. "Dynamics of Proteins and Nucleic Acids." Cambridge: Cambridge University Press, 1987.
59. Brooks III, C.L., Karplus, M., Pettitt, B.M. "Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics." New York: John Wiley & Sons, 1988.
60. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690, 1984.
61. Verlet, L. Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* 159:98–103, 1967.
62. Daggett, V., Levitt, M. Protein folding \leftrightarrow unfolding dynamics. *Curr. Opin. Struct. Biol.* 4:291–295, 1994.