Constant-*p*H molecular dynamics study of protonation-structure relationship in a heptapeptide derived from ovomucoid third domain

Maciej Długosz* and Jan M. Antosiewicz†
Department of Biophysics, Warsaw University, 02-089 Warsaw, Poland

Andrew D. Robertson‡
Department of Biochemistry, University of Iowa, Iowa City, Iowa 52240, USA

(Received 18 July 2003; revised manuscript received 14 October 2003; published 27 February 2004)

Molecular dynamics (MD) simulations with implicit solvent and variable protonation states for titratable residues at constant *p*H are performed for a short peptide derived from ovomucoid third domain (OMTKY3), acetyl-Ser-Asp-Asn-Lys-Thr-Tyr-Gly-amide (residues 26–32 of OMTKY3). Nuclear magnetic resonance (NMR) measurements indicate that the pK_a for Asp is 3.6. However, if the charge on Lys is neutralized by acetylation, then the pK_a for Asp is 4.0. These pK_a's, and therefore the Asp-Lys interaction, are insensitive to changes in ionic strength. The constant-*p*H MD simulations for both variants of the heptapeptide yield Asp pK_a values that are 0.6–0.9 *p*H units greater than experimental values, but the difference between the variants that is observed in the NMR experiments is reproduced much better. Moreover, the simulations suggest that Asp-Lys interactions do not dominate the behavior of this heptapeptide, even for normal Lys residue where there is a possibility of forming a salt bridge between negatively charged Asp and positively charged Lys. This is consistent with the experimentally observed independence of Asp pK_a values with respect to ionic strength. Another important result of the simulations with variable protonation states is that they lead to ensembles of the heptapeptide structures that are different from those derived from simulations with fixed protonation states. It should be stressed that these results are for structures generated entirely by computer simulations without any restrictions imposed by experimental data.

DOI: 10.1103/PhysRevE.69.021915 PACS number(s): 87.15.Aa

I. INTRODUCTION

One of the important determinants of protein structure and function is the ability of some functional groups in the amino acids to exchange protons with the environment [1]. There are many indications that a protein's structure and dynamics are sensitive to protonation states of its ionizable groups, and these are manifested primarily in the *p*H dependences for experimental results. However, in molecular simulations the effect of *p*H on protein structure is treated in a simplified way: depending on the *p*H and on the approximate pK_a of a given ionizable group, either the protonated or the unprotonated state of that group is selected and used for the whole simulation. On the other hand, theoretical approaches for prediction of ionization equilibria in proteins including effects of conformational fluctuations developed so far, see, e.g., Refs. [2–4] and references cited in those works, do not take into account the fact that conformational fluctuations and proton exchange phenomena occur simultaneously in solution and that protein conformations usually do not have time to equilibrate with a given protonation pattern before next proton exchange phenomenon occurs. Inclusion of explicit proton exchange in molecular simulations without *a priori* assumptions regarding time scales for conformation and protonation dynamics is necessary for obtaining a physically realistic picture of the molecular mechanisms behind protein dynamics and function, and the development of algorithms of such simulations has already started [5,6]. In addition, constant-*p*H algorithms with implicit proton exchange have been described [7,8]. Ultimately, algorithms for constant-*p*H simulations should become a routine tool in molecular simulations, just as constant temperature and constant pressure algorithms, and they will certainly find numerous useful applications to the study of biologically relevant problems.

One of the main problems to be solved regarding explicit protonation simulations at constant *p*H is the need to include the influence of the aqueous solvent on the protein dynamics. Water molecules participate in the proton exchange phenomena and water molecules and ions of dissolved salts are important sources of electrostatic potential acting on the solute atoms. However, explicit water simulations with explicit protonations are presently too expensive to be used as a routine approach [5]. On the other hand, simple Langevin dynamics of proteins at constant *p*H [6] can be used in more extensive simulations but these do not include electrostatic interactions with the environment. In this regard, simulations with implicit solvent are a reasonable compromise.

Here we present an approach where explicit titration, based on the Poisson-Boltzmann model of the solute-solvent system, is coupled to generation of protein structures by molecular dynamics (MD) simulations with implicit solvent. Our algorithm is applied to a short peptide derived from OMTKY3, acetyl-Ser-Asp-Asn-Lys-Thr-Tyr-Gly-amide (residues 26–32 of OMTKY3, see Fig. 1). There are three titratable residues in this molecule - Asp 2, Lys 4, and Tyr 6.

*Electronic address: mdlugosz@biogeo.uw.edu.pl
†Electronic address: jantosi@biogeo.uw.edu.pl
‡Electronic address: aroberts@blue.weeg.uiowa.edu

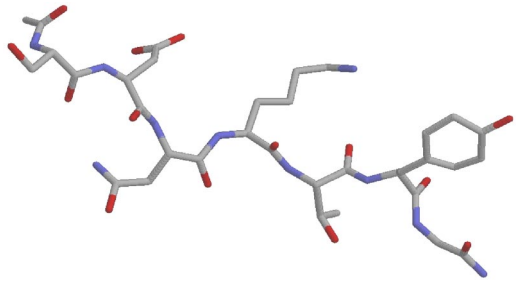

FIG. 1. (Color online) Structure of the heptapeptide, acetyl-Ser-Asp-Asn-Lys-Thr-Tyr-Gly-amide.

Nuclear magnetic resonance (NMR) measurements indicate that the pK_a for Asp is 3.6, which is 0.4 units less than the value for an unperturbed aspartate. This pK_a is insensitive to changes in ionic strength. However, if the charge on Lys is neutralized by acetylation, then the pK_a for Asp increases to 4.0. This system thus shows interesting titration properties and, at the same time, is small enough to allow extensive testing of a variety of simulation protocols.

The main purpose of the present study is to compare fixed and variable protonation state simulations and the extent to which explicit proton exchange phenomena influences the ensemble of polypeptide structures. We would also like to determine if our algorithm predicts the pK_a values and the differences in pK_a values observed experimentally in the two heptapeptides.

II. SIMULATION METHODS

The general outline of our simulation procedure is shown in Fig. 2. The initial heptapeptide is generated by the MD program and assigned a protonation pattern of its titratable residues based on the pH and on pK_a values of the isolated protonable groups [9]. This structure is subject to minimization and a thermalization procedure, as described below. The resulting structure enters the simulation circle shown in Fig. 2. Within each turn of the simulation circle, MD simulation is used to generate a dynamical trajectory of the peptide for a predefined period of time (usually 5 ps), yielding a fraction of the whole trajectory (subtrajectory). The next step is to evaluate the protonation fractions and the most probable protonation patterns for the titratable residues based on the free energies of the states, at a given pH, using a Monte Carlo procedure. This titration step is done for the last structure of the 5 ps subtrajectory. Predicted protonation fractions of the individual residues are stored for further evaluation of trajectory-averaged values. One of the protonation patterns is then chosen based on its Boltzmann law probability and this defines the protonation states for the MD simulation of the next sub-trajectory. This is repeated until the whole trajectory covers the desired length of time. Given that protonation-deprotonation at a given site can be expected to occur very rapidly, 5 ps might appear to be long. The present choice is dictated by the relatively high time cost of the titration step in the simulation circle. Moreover, a previous investigation demonstrated that a constant-pH MD algorithm with implicit proton exchange gave very similar results when protonation changes were made at 0.2, 1, and 5 ps [8]. However proper tuning of the time intervals at which proton exchanges between the molecule and the solvent is clearly an important issue and will be explored in future tests of the procedure. Moreover, this important point illustrates a basic difference between the present algorithm and the very recent algorithms described in Refs. [5,6], on the one hand, and all previous algorithms for pK_a predictions where protein structure is varied, on the other hand. In the older algorithms, simulations with different protonation patterns are done sufficiently long such that conformational equilibrium is reached, and this can have a significant influence on the predicted ionization constants of the residues [10].

All MD simulations, modifications, and minimizations are performed with the CHARMM package [11]. Electrostatic and Monte Carlo calculations required by the titration step are done using the University of Houston Brownian dynamics (UHBD) [12,13] and the distribution of protonation states (DOPS) [14] programs, respectively.

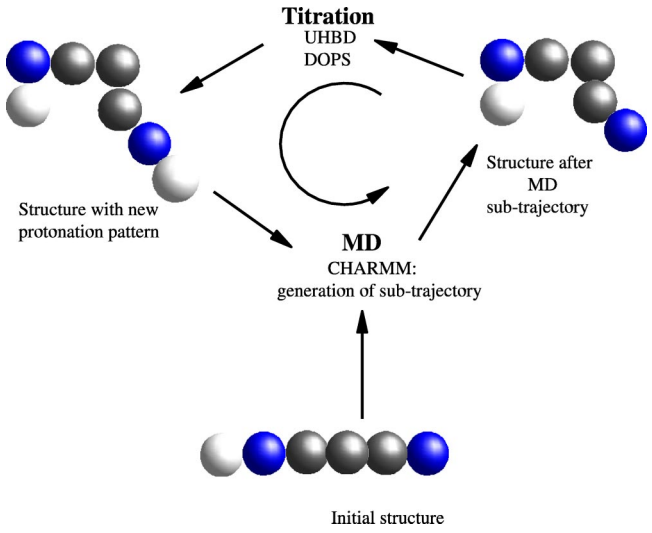

FIG. 2. (Color online) Outline of the simulation procedure.

A. Prediction of protonation states

The prediction of protonation states for titratable residues in the peptides is done according to procedures described elsewhere [15]. The UHBD program is used for computing the electrostatic free-energy interaction matrices. For UHBD calculations, radii and charges from CHARMM all-hydrogen parameter set are used [16]. The dielectric constant of the solute is usually set to 4 and for the solvent a value of 80 is used. Several previous theoretical approaches for prediction of ionization equilibria in proteins used higher solute dielectric constant, therefore, for comparison some of our calculations employ a value of 15 as the solute dielectric constant. All calculations use an ionic strength corresponding to 150 mM of monovalent salt. The DOPS program uses the resulting electrostatic free-energy interaction matrices to compute average protonation fractions at a given pH and to generate a predefined number of peptide protonation patterns with the lowest energies as found by a Monte Carlo algorithm.

Quantitatively, the protonation equilibrium of a given functional group is characterized by its $p\mathrm{K}_a$ value, defined as the negative decimal logarithm of the acidic dissociation constant K_a for the proton dissociation reaction [17]. The average protonation fraction as a function of pH and Eq. (1) is used to compute predicted $p\mathrm{K}_a$ values:

$$p\mathrm{K}_a = p\mathrm{H} - \log_{10}\frac{1-p}{p} \tag{1}$$

where p is the average protonation fraction at a given pH. In principle, with this equation one does not need to simulate the entire titration curve of the peptide. However, if the simulation is done at a pH that is significantly different from the $p\mathrm{K}_a$ value(s) for the residues of interest, then the estimated $p\mathrm{K}_a$ values will be burdened with larger statistical errors.

B. Molecular dynamics—implicit solvent

The MD simulations with implicit solvent use the analytical continuum solvent potential combined with the CHARMM19 potential for proteins [11]. In this method, the electrostatic contribution to the free energy of the solute is calculated using an analytical approximation of the Poisson equation—ACE (analytical continuum electrostatic) [18]. The charges and atomic radii are taken from file param19-1.2.inp—polar hydrogen parameter set of CHARMM. This choice is consistent with the parametrization of ACE potential incorporated in CHARMM.

We modified the original parameter set by adding patches that allow us to change the default protonation state of the simulated structures (i.e., hydrogen connected to OD2 oxygen in ASP, number of hydrogens connected to NZ atom of lysine, ionized state of tyrosine) or to acetylate amino groups or amidate carboxyl groups of the peptides.

We use a dielectric constant of 80 for solvent, which is treated as a continuum, and a dielectric constant of 1 for the space occupied by the atoms of protein. All simulations (also with explicit solvent) are conducted at a constant temperature of 293 K. We use an atom-based switch scheme for nonbonded interactions (cutoff distance 13 Å): during simulations, a heuristic algorithm implemented in CHARMM is used for updating a nonbonded interaction list.

C. Molecular dynamics—explicit solvent

For simulations with explicit solvent we create cubic box (edge 46.65Å) constructed from tip3p water molecules (3324 molecules of water inside the box). In this part of the work, an all-hydrogen parameter set (PARAM27) [16] is used. The dielectric constant is set to 1 everywhere in the box; the cutoff distance for nonbonded interactions (atom-based switch scheme) is set to a slightly smaller distance of 11 Å because explicit solvent simulations are much more time consuming. For fixed periodic boundary conditions we apply the CRYSTAL module of CHARMM.

All simulations (with explicit and implicit solvent) are conducted using the Verlet (Newton's equation of motion) algorithm with a time step 10^{-3} ps.

D. Preparation of initial structures

The initial heptapeptide is assigned a protonation pattern of its titrable residues based on their model $p\mathrm{K}_a$ values [9] and the pH of simulation. Next, starting from fully extended structure, we perform heating from 0 K to three different target temperatures: 600, 800, and 900 K. Finally, each of three structures is cooled (to 293 K) and equilibrated.

III. EXPERIMENTAL METHODS

A. Peptide synthesis

Solid-phase synthesis of acetyl-Ser-Asp-Asn-Lys-Thr-Tyr-Gly-amide was done using Rink amide MBHA resin (Novabiochem) and 9-fluoroenylmethoxycarbonyl (FMOC) amino acid precursors (Novabiochem). The N terminus was acetylated with acetic anhydride. Peptides were cleaved from the resin with 95% trifluoroacetic acid in H2O and purified by C18 reverse-phase chromatography. Peptide identity and purity were verified by analytical 1H-NMR spectroscopy. Acetylation of the side chain amino group of lysine 4 was achieved by exposing the peptide to 10% acetic anhydride in dimethylformamide for 30 min. The derivatized peptide was purified by C18 reverse-phase chromatography and derivatization was confirmed by 1H-NMR spectroscopy.

B. $p\mathrm{K}_a$ determinations

Solutions for $p\mathrm{K}_a$ determinations contained 2 mM peptide, 3 mM potassium citrate, 0.4 mM sodium-3-(trimethylsilyl) propionate-2,2,3,3-d4 (TSP) as a chemical shift reference, 90% H2O and 10% D2O; some samples contained 1 M KCl. COSY and NOESY data were used to assign the 1H resonances for the peptides [19]. For the titration studies, the pH was adjusted with concentrated stock solutions of potassium hydroxide and hydrochloric acid; experiments covered pH 2 to pH 6 in 0.33 pH unit increments. 1D 1H-NMR data were collected at each pH value and the pH dependence of the aspartyl NH resonance was used to determine its $p\mathrm{K}_a$ value, as described previously [20]. All NMR data were collected on a Varian UNITY 500 NMR spectrometer located in the University of Iowa College of Medicine. The estimated experimental uncertainty in $p\mathrm{K}_a$ determinations is 0.1 unit.

IV. RESULTS

Figure 3 shows results of simulations with protonable residues for the peptide with Lys 4 (right) and the peptide in which the amino group of Lys 4 is acetylated (Ac-Lys, left). These are distributions of protonation fractions of the Asp carboxyl group at several pH values. These results are based on 600 structures at each pH sampled from three 1.0 ns MD trajectories, each starting with different initial structure obtained after the heating and equilibration procedure. In all cases, each structure was sampled after 5 ps of a subtrajectory along the complete MD trajectory. For the pH values selected for these calculations, only the Asp residue changes its protonation state; Tyr and Lys always remain protonated.

The distribution of average protonation fractions in Fig. 3

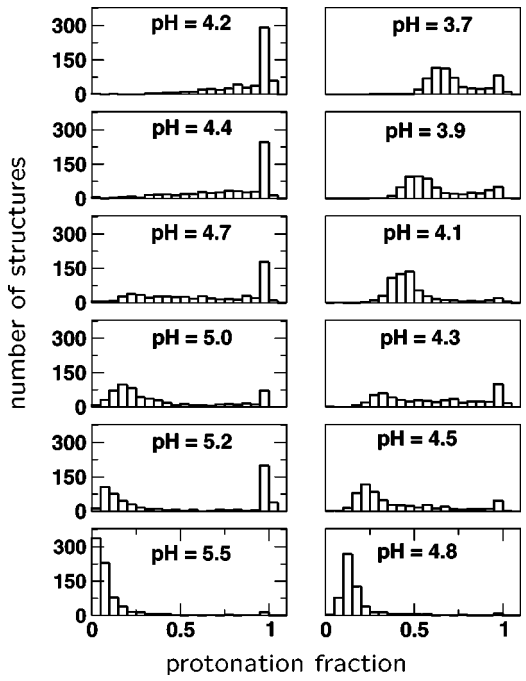

FIG. 3. Distributions of protonation fractions of Asp in structures derived from MD trajectories, with variable protonation states of the residues for solute dielectric constant of 4, at different *p*H values. Right, with normal lysine residue; left, with acetylated lysine residue.

shifts its center of weight from values close to 1 to values close to 0 as the *p*H is increased. For certain *p*H values, e.g., 4.5 for simulations with Lys and 5.0 for Ac-Lys, there are maxima for high and low protonation fractions with less probable values between them, i.e., the distribution is bimodal. For comparison, similar calculations were done for fixed protonation states of the residues: in the MD simulations Asp was either protonated or deprotonated, Lys and Tyr were always protonated, and the average protonation fractions for Asp were estimated based on the titration procedure with UHBD and DOPS. The same number of structures were sampled as in the case of the variable protonation simulations.

Figure 4 includes protonation diagrams obtained from simulations with fixed protonation states. To facilitate comparison, Fig. 4 also includes data from Fig. 3 for variable protonation simulations at the same *p*H. One interesting difference in the calculations for fixed versus variable protonations states is that *p*H dependence for the distribution of average protonation fractions for fixed protonation states is always determined from the same ensemble of structures, whereas the calculated ensemble of structures for variable protonation states is different at each *p*H. The results presented on the right side of Fig. 4 are surprising. The distribution for protonable Asp is not intermediate between the distributions obtained for fixed protonated and deprotonated Asp, respectively, as is the case of the distributions obtained for Ac-Lys, shown on the left-hand side of the figure. We cannot offer a simple explanation of this effect at present, but it seems to be in agreement with results presented below (see comments to Fig. 7 below).

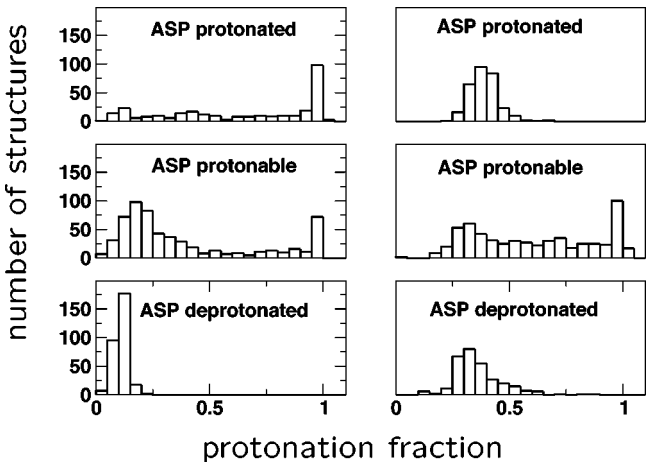

FIG. 4. Examples of distributions of protonation fractions of Asp obtained from MD simulations with fixed protonation states (ASP protonated, ASP deprotonated), compared with corresponding results from simulations at constant *p*H (ASP protonable). Right, normal Lys, *p*H=4.3; left, acetylated Lys, *p*H=5.0.

Tables I and II present the average protonation fractions, minimal and maximal protonation fraction, and aspartyl pK_a values obtained from three different variable protonation state simulations at different *p*H values for peptides containing Lys 4 (Table I) and Ac-Lys 4 (Table II). The estimated pK_a values are all based on application of Eq. (1), so the variability in these values reflects the statistical uncertainty in the results of the simulations. This variability is quite substantial. However, the standard deviations at the 95% confidence level (twice the standard deviation of the mean value) are not large: for the peptide containing Lys 4, the

TABLE I. Protonation fractions of Asp and the corresponding pK_a values obtained from three 1 ns MD simulations with protonable Asp, Lys, and Tyr, with implicit solvent, $\epsilon=4$.

*p*H	Mean	St. dev.	Min.	Max.	pK_a
	0.66	0.08	0.30	0.90	3.99
3.7	0.66	0.08	0.29	0.86	3.98
	0.88	0.14	0.55	1.00	4.55
	0.55	0.09	0.33	0.91	3.99
3.9	0.52	0.08	0.28	0.84	3.93
	0.84	0.15	0.44	1.00	4.61
	0.61	0.22	0.08	1.00	4.29
4.1	0.43	0.07	0.20	0.68	3.98
	0.45	0.10	0.27	0.89	4.02
	0.57	0.27	0.17	1.00	4.43
4.3	0.60	0.28	0.05	1.00	4.47
	0.68	0.25	0.05	1.00	4.62
	0.42	0.26	0.03	1.00	4.37
4.5	0.32	0.21	0.09	1.00	4.18
	0.49	0.27	0.05	1.00	4.49
	0.14	0.05	0.05	0.38	4.01
4.8	0.19	0.18	0.02	1.00	4.17
	0.23	0.19	0.04	1.00	4.28

TABLE II. Protonation fractions of Asp and the corresponding $p\mathrm{K}_a$ values obtained from three 1 ns MD simulations with protonable Asp, and Tyr, with acetylated Lys, and with implicit solvent, $\epsilon=4$.

pH	Mean	St. dev.	Min.	Max.	$p\mathrm{K}_a$
	0.92	0.16	0.02	1.00	5.29
4.2	0.88	0.15	0.42	1.00	5.08
	0.85	0.19	0.03	1.00	4.96
	0.89	0.22	0.02	1.00	5.33
4.4	0.76	0.24	0.05	1.00	4.91
	0.71	0.24	0.00	1.00	4.78
	0.63	0.32	0.02	1.00	4.93
4.7	0.70	0.28	0.03	1.00	5.06
	0.62	0.32	0.05	1.00	4.91
	0.25	0.18	0.00	0.99	4.52
5.0	0.38	0.30	0.00	1.00	4.78
	0.56	0.34	0.03	1.00	5.10
	0.55	0.41	0.00	1.00	5.29
5.2	0.36	0.37	0.03	1.00	4.96
	0.66	0.41	0.05	1.00	5.49
	0.09	0.07	0.02	0.47	4.50
5.5	0.20	0.25	0.00	1.00	4.89
	0.04	0.03	0.01	0.18	4.15

$\langle p\mathrm{K}_a\rangle_{Asp}=4.24\pm0.12$ while for the peptide with Ac-Lys 4, the $\langle p\mathrm{K}_a\rangle_{Asp}=4.94\pm0.16$. The $p\mathrm{K}_a$ values are 0.6–0.9 pH units greater than the experimental values but the calculated difference between the two peptides, 0.7 ± 0.2 pH unit, agrees well with the experimental results.

Another way of looking at the results of constant pH simulations is to follow changes in protonation states chosen for subsequent subtrajectories of MD simulations. These are shown as a functions of pH in Fig. 5. In simulations with Lys 4 and Ac-Lys 4, there is a transition from more protonations to more deprotonations as pH is increased. At low pH, protonated states last longer and any transition to the deprotonated state is followed by a quick return to the protonated state. For higher pH values, deprotonated states last longer. The pH-induced transitions occur at lower pH for the Lys 4 peptide than they do for the peptide with Ac-Lys 4.

Figure 6 shows representative heptapeptide structures obtained from simulations at the extreme pH values used in this work: 3.7 and 4.8 for the Lys 4 peptide and 4.2 and 5.5 for the Ac-Lys 4 peptide. Some trends are evident in these structures. In all of the structures, the polypeptide chain from Asn 3 to Gly 7 is bent and forms a loop. Asp 2 points towards the central region of the loop in the first and third sets of structures (Lys 4 and Ac-Lys 4 at the lower pH limits) but points in the opposite direction in the two remaining sets of structures (Lys 4 and Ac-Lys 4 at the higher pH limits). The first and third sets of structures are characterized by a high degree of proton occupancy of Asp mainly because the protonated side chain of Asp 2 can form a hydrogen bond with one of carbonyl oxygens within the loop. The two remaining sets of structures are characterized by a low probability of protonated Asp. The Lys side chain does not exhibit any obvious structural regularities. All side chains are relatively well exposed to solvent in all of the structures.

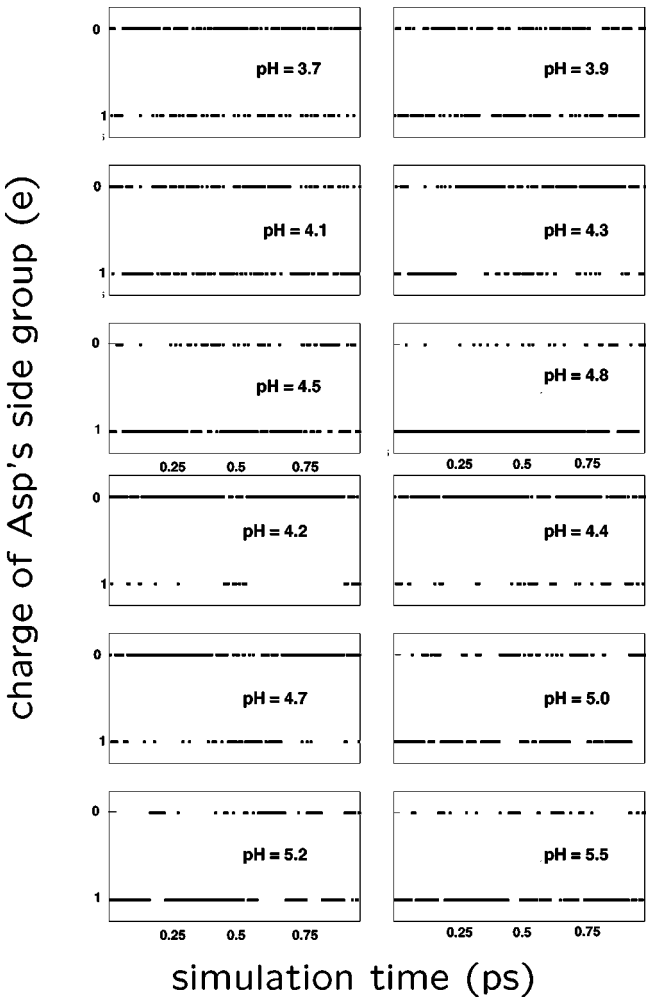

FIG. 5. Changes in the residue charge of Asp along MD trajectories, presented as a function of the simulation time, for constant pH simulations at indicated pH values for ordinary (upper half) and acetylated Lys residue (lower half).

V. DISCUSSION

We have applied an algorithm for simulating polypeptide chains at constant pH to investigate the relationship between protonation states and structure and to predict $p\mathrm{K}_a$ values in a heptapeptide, acetyl-Ser-Asp-Asn-Lys-Thr-Tyr-Gly-amide, derived from OMTKY3. The agreement between predicted and experimental $p\mathrm{K}_a$ values for Asp 2 in peptides with Lys 4 and Ac-Lys 4 is moderately good, especially given that the structures used for prediction of $p\mathrm{K}_a$ values are determined entirely by computer simulations. Nevertheless, it is surprising that, in this simple heptapeptide with all protonable groups well exposed to the solvent, the predicted $p\mathrm{K}_a$ values of Asp 2 are greater than the model compound value of 4.0. One might expect that in the peptide with Lys 4, electrostatic interactions with the positively charged Lys would stabilize the deprotonated form of Asp, i.e., those structures with small Asp-Lys distances, and

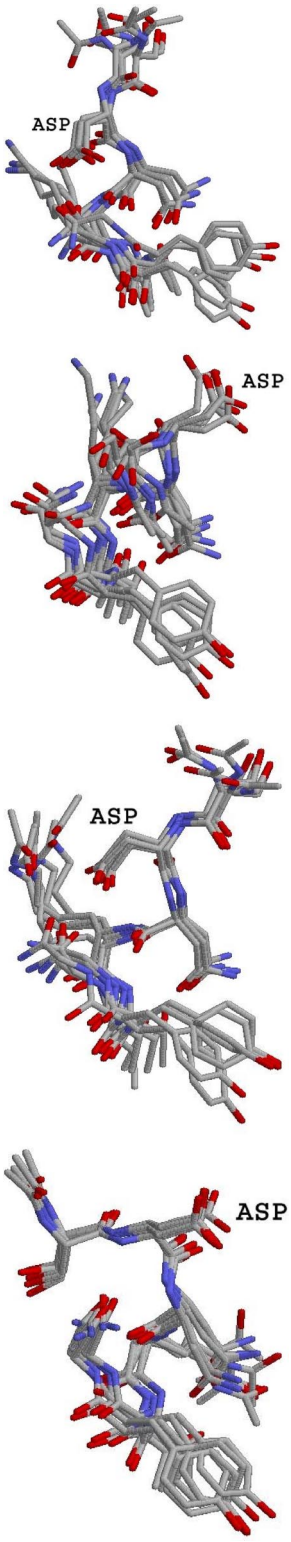

FIG. 6. (Color online) Representative structures obtained from constant pH simulations with ordinary (first and second group from the top) and acetylated (third and fourth groups from the top) lysine residue, at *p*H values given in the text. For structures in the first and third group the average protonation fraction of Asp is above 0.9 and for structures in the second and fourth groups the average protonation fraction is below 0.1.

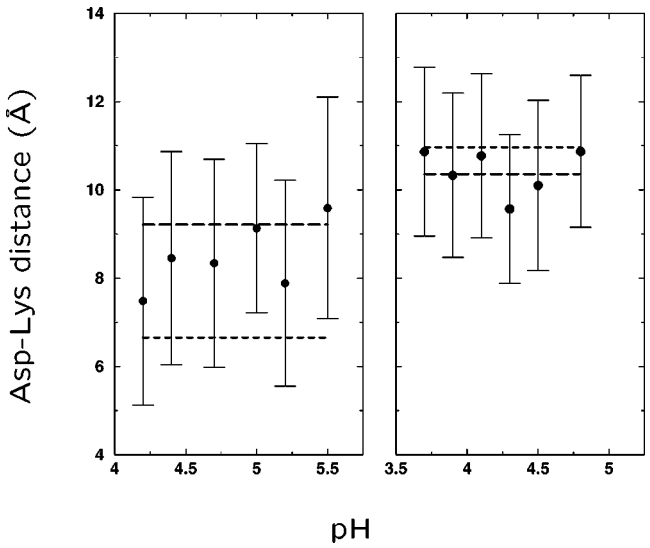

FIG. 7. Distances Asp-Lys as functions of *p*H for normal lysine (right) and acetylated lysine (left) from simulations with implicit solvent. Broken lines are for fixed protonation state simulations; dashed line is for protonated Asp, long-dashed line is for deprotonated Asp, and filled circles with error bars are for constant *p*H simulations. The standard deviations for the distances obtained from simulations with fixed protonation states are of similar magnitude as for the constant *p*H calculations and are not shown for clarity of the picture.

thereby decrease the pK_a value of Asp 2. On the other hand, the pK_a of the solvent-exposed Asp 2 in the peptide containing Ac-Lys 4 peptide should be close to the model compound value because there are no sources of strong negative electrostatic potential (the influence of carbonyl oxygens described in the preceding paragraph has the character of an electric dipole) that would prevent its deprotonation, and there is no extra positive charge that would stabilize its deprotonated form. Therefore, we checked Asp-Lys distances in structures obtained from the variable protonation simulations at different *p*H values and in structures obtained with fixed protonation states. Results of this analysis are shown in Fig. 7.

The Asp-Lys distances and their fluctuations are rather large. Surprisingly, the average distances are slightly larger in the Lys 4 peptide than in the peptide with Ac-Lys 4. In addition, the effect of going from deprotonated to protonated Asp 2 with Ac-Lys 4 is larger and opposite to that seen in simulations for Lys 4. The change in Asp-Lys distances for the Lys 4 peptide in going from protonated to deprotonated Asp 2 shows the expected trend, as the electrostatic attraction should result in the Asp-Lys average distance being smaller than when Asp 2 is neutralized. However, the change is surprisingly small. For simulations with protonable sites, two average values, for *p*H=4.3 and *p*H=4.5, are below the line corresponding to the average distance between permanently deprotonated Asp and protonated Lys residues. This might explain why the distributions of protonation fractions of Asp 2, obtained in simulations with fixed protonation states at *p*H 4.3 and presented in the right-hand side of Fig. 4, are so similar and why the distribution obtained in simulations with

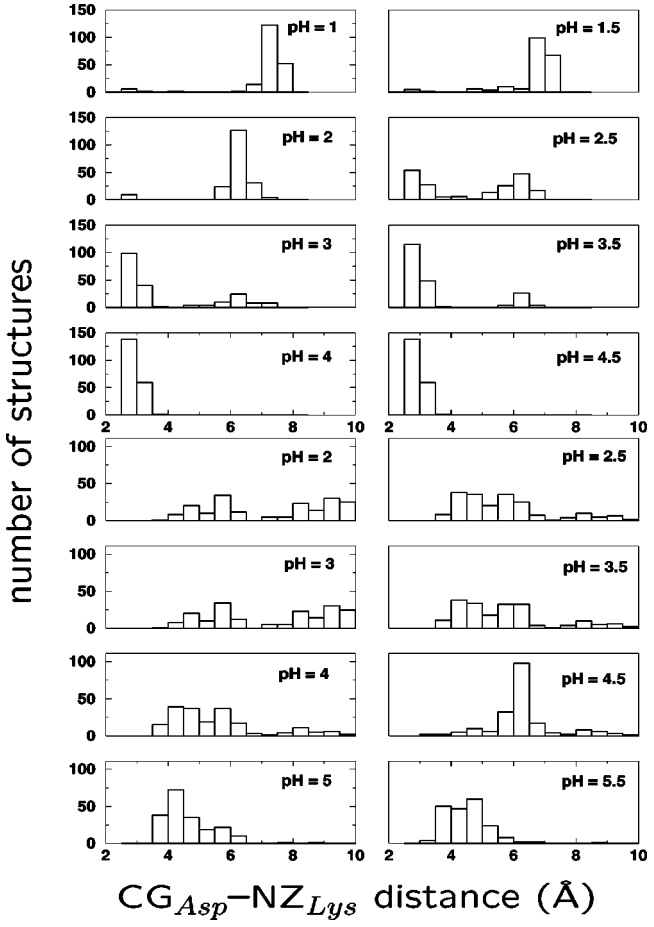

FIG. 8. Distributions of distances between CG_{Asp} and NZ_{Lys} obtained from simulations with protonable residues in vacuum for normal (upper half) and acetylated (lower half) lysine.

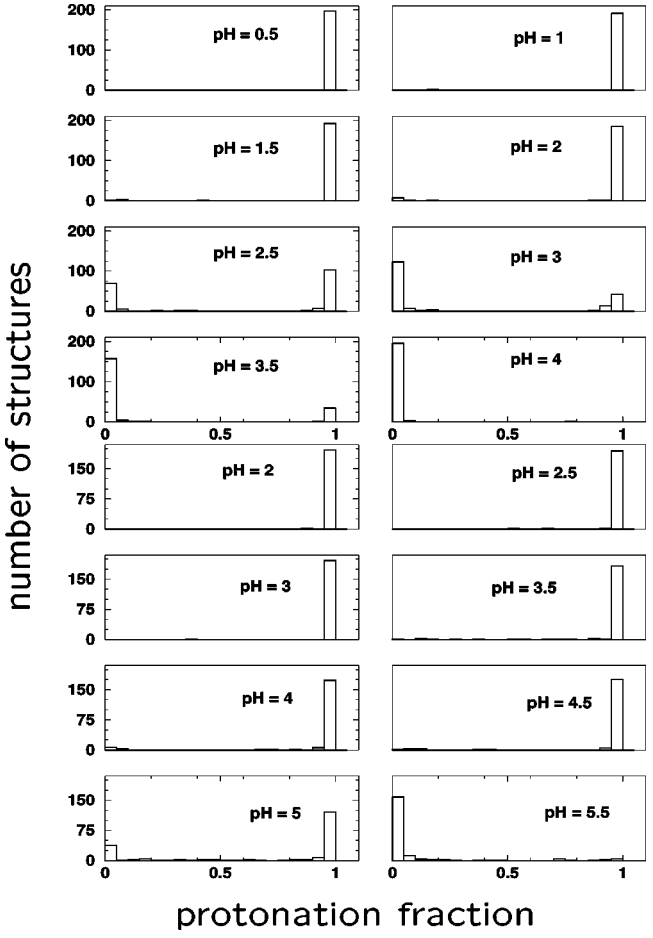

FIG. 9. Distributions of protonation fractions of Asp obtained with protonable residues, at indicated pH values, without solvent, with normal lysine (upper half), and with acetylated lysine (lower half).

protonable residues is not intermediate between the two obtained from fixed states simulations. On the other hand, for Ac-Lys 4 the difference in distances for protonated and deprotonated Asp 2 is much larger, and all average distances for constant pH simulations are between the two limiting dashed lines and change rather smoothly with increasing pH from one case to the other (as illustrated in the left-hand side of Fig. 7). In summary, the results presented in Fig. 7 are rather surprising and not intuitively obvious. Analogous simulations have been performed in vacuum in order to check the extent to which the puzzling results described above are due to inclusion of implicit solvent effects in our simulations. Results of these vacuum simulations are described below.

Figure 8 shows distributions of distances between Asp 2 and Lys 4 for MD simulations without solvent and as a function of pH. These distributions can be explained by electrostatic interactions as the dominant factor. For the Lys 4 peptide, the average distance changes from approximately 7 Å to approximately 3 Å when the pH is changed from 0.5 to 4.0 and the pK_a of Asp 2 is 2.73 ± 0.30 at the 95% confidence level. For the peptide containing Ac-Lys 4, the distribution of Asp-Lys distances is broad at low pH and narrows with increasing pH; the pK_a of Asp 2 is 5.11 ± 0.28. The first pK_a value is thus significantly less than the experimental value of 3.6 and the second is significantly greater than the corresponding experimental value of 4.0. The calculated difference between the pK_a values is almost six times greater than the experimentally determined difference. Both distributions become narrower for higher pH values and, as expected, the Asp-Lys distances in the Lys 4 peptide are shorter than in the Ac-Lys 4 peptide but the overall effect is small.

Figure 9 presents the pH dependence for the distributions of protonation states resulting from these simulations. One striking feature is the negligible contribution of structures with intermediate protonation states; almost all protonations are close to either 1 or 0. Clear bimodal distributions are seen at certain pH values for both the Lys 4 and Ac-Lys 4 peptides and this bimodality is more evident than in the case of Asp-Lys distance distributions (Fig. 8). We therefore conclude that the addition of implicit solvent in the simulation algorithm leads to an increase in the distances between Asp 2 and Lys 4. The fact that inclusion of solvent decreases electrostatic interactions relative to simulations in vacuum is not too surprising but, in the present simulations, the consequences appear to extend beyond simply introducing a uniform dielectric screening parameter for all electrostatic inter-

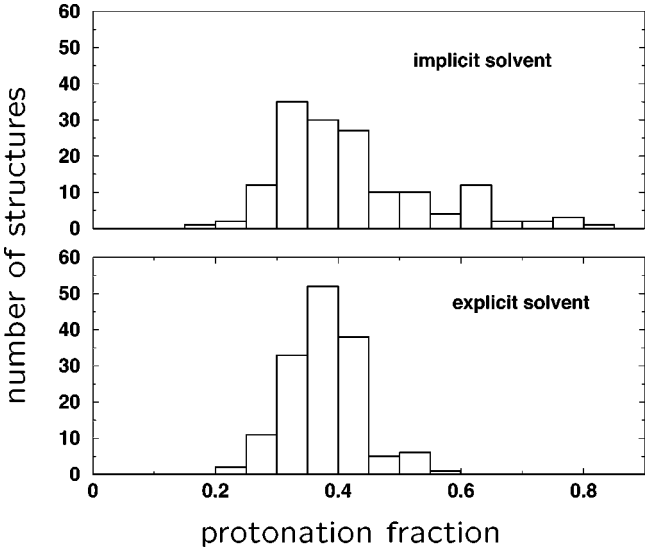

FIG. 10. Example of comparison of distributions of protonation fractions between implicit and explicit solvent simulations with fixed protonation states.

actions, as might be the case with simple Langevin dynamics simulations [6]. The simulations show that the pK_a shift for Asp 2 upon acetylation of Lys 4 cannot be explained solely by the expected changes in direct electrostatic interactions with the side chain amino group of Lys 4. This is qualitatively consistent with the experimentally observed independence of the pK_a on ionic strength.

To further investigate the role of solvent and as an additional check of our procedure, simulations were also performed with explicit solvent. Figure 10 presents a comparison of implicit and explicit solvent simulations, based on 150 structures covering 0.75 ns each. These simulations were done with fixed protonation states, where Asp was deprotonated, Lys and Tyr protonated, and both distributions of protonation fractions refer to pH 4.2. The distributions of protonation fractions obtained in both simulations are similar, although for implicit solvent the distribution is broader than that for the explicit solvent simulations. This results from damping of solute atomic fluctuations by water molecules in explicit solvent simulations and the absence of a corresponding effect in the implicit solvent simulations. The implicit solvent data give a pK_a of Asp 4.06 vs 3.97 derived from the simulation with explicit solvent.

Another factor governing the strength of electrostatic interactions is the solute dielectric constant; for the solvent we use a value appropriate for water at given temperature. Figure 11 presents a comparison of results obtained for solute dielectric constants of 4 and 15. The distribution of protonation fractions of Asp 2 for the Ac-Lys 4 peptide is no longer bimodal with the higher solute dielectric constant. We therefore conclude that the bimodal distributions are exhibited when both conditions are fulfilled: proton exchange phenomena are taken into account and the strength of the electrostatic interactions is sufficiently high. Increasing the solute dielectric constant is probably not an appropriate way of obtaining more reliable results, as the pK_a values predicted for the Lys 4 peptide and the Ac-Lys 4 peptide from

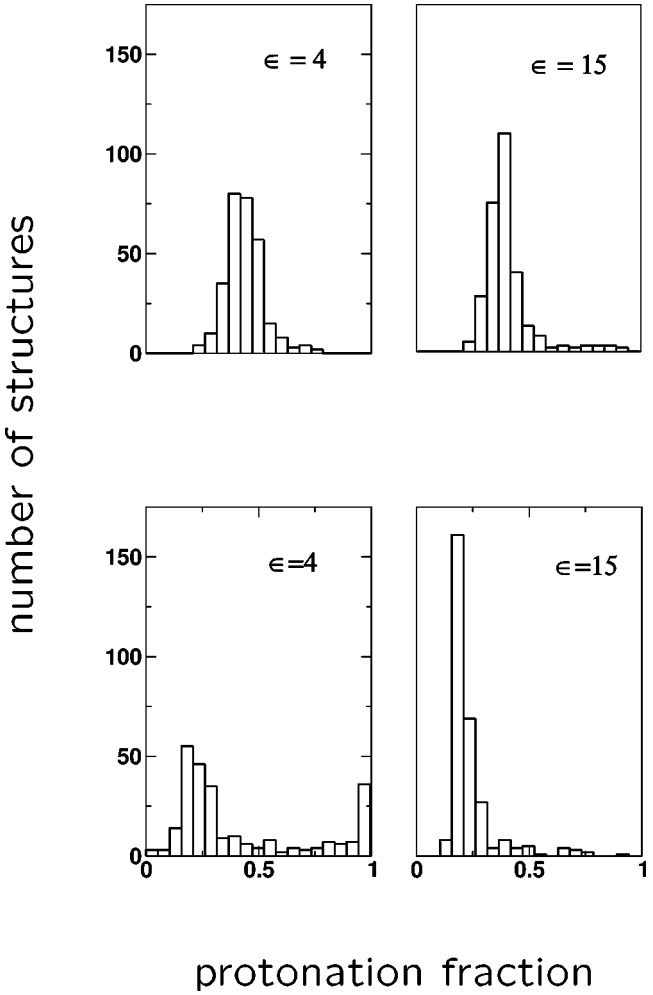

FIG. 11. Influence of the solute dielectric constant used in UHBD calculation on the distributions of protonation fractions in simulations with protonable sites, for normal lysine (top, pH=4.2) and for acetylated lysine (bottom, pH=4.7). For each part there are 300 structures sampled from 1.5 ns of simulations).

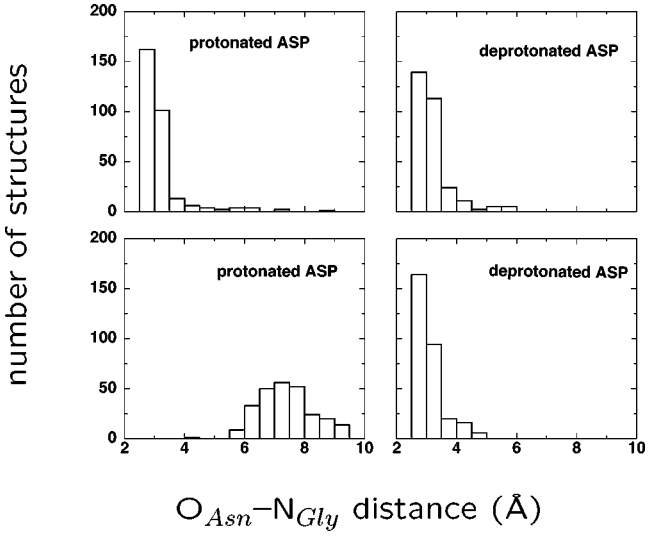

FIG. 12. Distributions of distances O_{Asn}-N_{Gly} for simulations with fixed protonation states and implicit solvent (top, Ac-Lys 4; bottom, Lys 4).

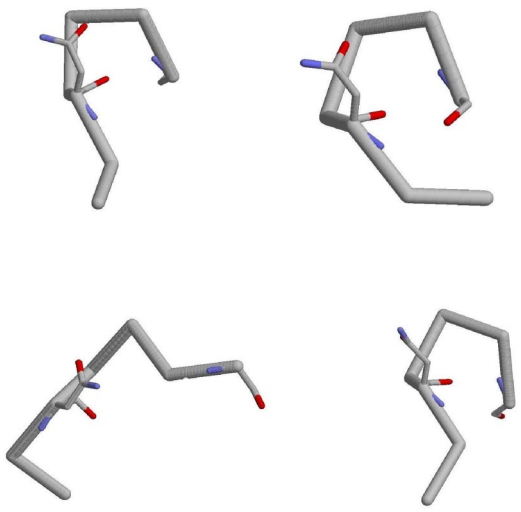

FIG. 13. (Color online) Representative structures for most populated distances O_{Asn}-N_{Gly} shown in Fig. 12.

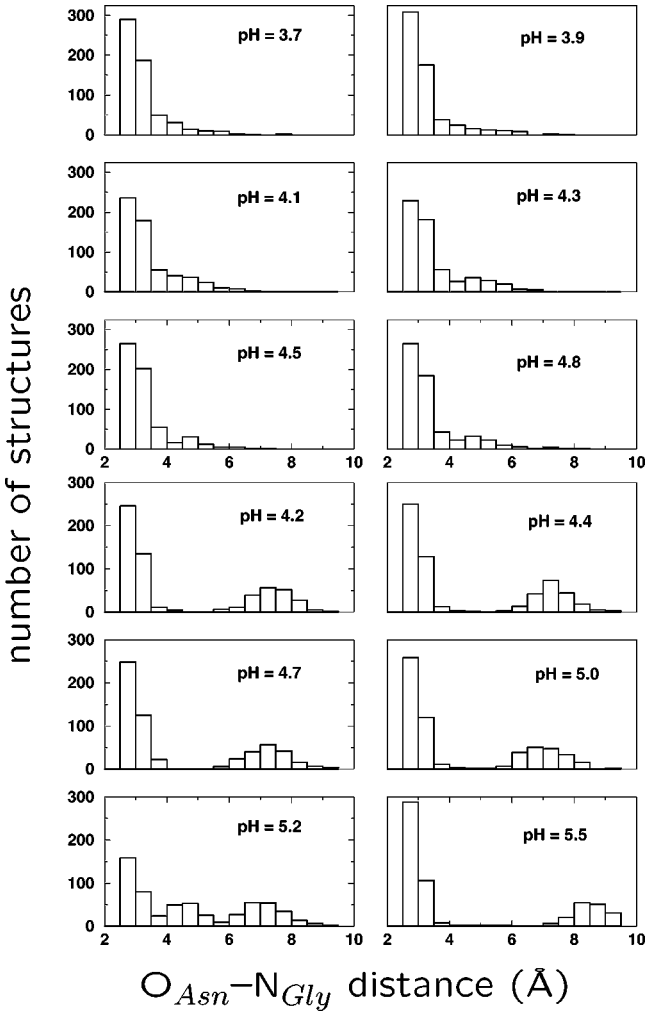

FIG. 14. Distributions of distances O_{Asn}-N_{Gly} at several pH values for normal Lys (upper half) and acetylated Lys (lower half) for simulations with protonable residues and with implicit solvent.

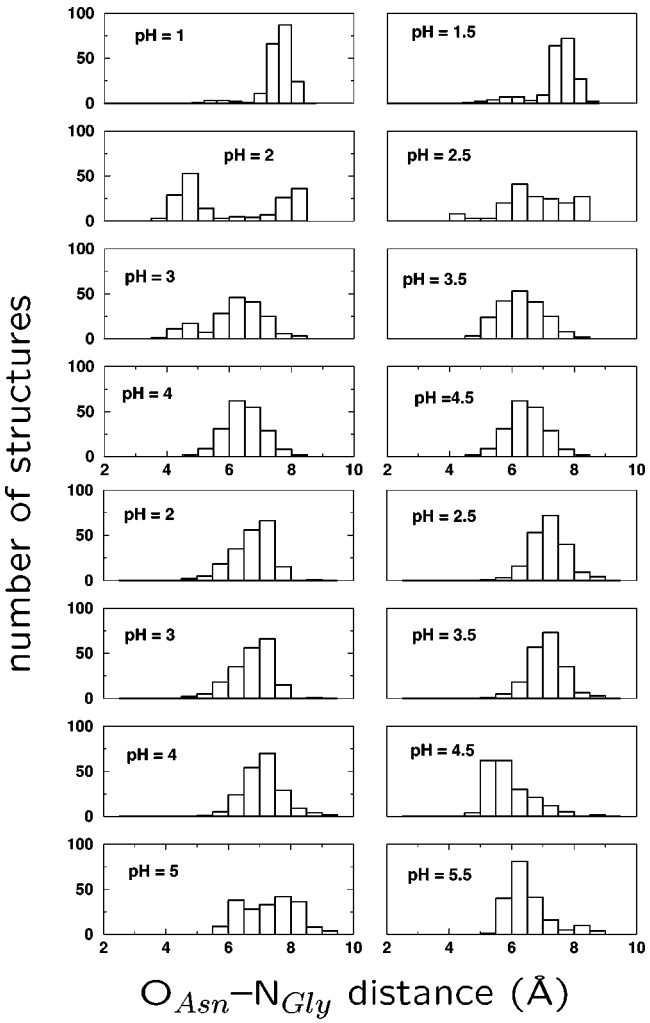

FIG. 15. Distributions of distances between O_{Asn} and N_{Gly} obtained from simulations with protonable residues in vacuum for normal (upper half) and acetylated (lower half) lysine.

the data in Fig. 11 are 4.03 ± 0.06 and 4.19 ± 0.18, respectively, for a solute dielectric constant of 15 vs 4.09 ± 0.04 and 4.68 ± 0.65 for a solute dielectric constant of 4 (all errors are at the 95% confidence). The pK_a difference for the dielectric constant of 15 is too small relative to the experimental values.

In the last part of this section we consider distances between Asn 3 and Gly 7 because they are related to the interesting loop structure mentioned at the end of the Results section. Figure 12 shows distributions of these distances for the four different types of simulations with fixed protonation states and implicit solvent. For the Lys 4 peptide, the average Asn-Gly distance is 7.75 ± 1.26 Å when Asp 2 is protonated and 3.31 ± 0.44 Å when Asp 2 is deprotonated. For the peptide containing Ac-Lys 4, the average Asn-Gly distance is 3.21 ± 0.78 Å when Asp is protonated and 3.20 ± 0.58 Å when Asp is deprotonated. Figure 13 shows representative structures corresponding to these four average distances presented in the same arrangement as parts of Fig. 12.

Figure 14 presents distributions of O_{Asn}-N_{Gly} distances as a function of pH for peptides containing Lys 4 and Ac-Lys 4

obtained from simulations at constant pH and with implicit solvent. The role of solvent in the short distances seen in Fig. 14 can be ascertained by inspection of Fig. 15, which presents distances obtained from constant pH simulations without solvent. For the Lys 4 peptide, all distributions obtained from simulations with implicit solvent are similar: the center of weight of the distribution is about 3 Å and the effect of increasing pH is a slight decrease in the number of structures with short Asn-Gly distances and a slight increase in the number of structures with distances that are 2–3 Å greater. In the Ac-Lys 4 peptide, all distance distributions exhibit a bimodal character with structures in which Asn-Gly distances are centered around 3 and 7–8 Å. For simulations without solvent none of the distributions of Asn-Gly distances in the Ac-Lys 4 peptide is bimodal. A bimodal distribution is visible in vacuum simulation with normal Lys 4 at pH 2. This suggests again that such distribution can be observed when electrostatic interactions are substantial and there is a possibility of changing the protonation state of Asp 2. In Fig. 9 it can be seen that the transition in Asp protonation state in the vacuum simulation occurs between pH 2 and 3 ($p\mathrm{K}_a = 2.73$ as mentioned above). All distances in both sets of distributions shown in Fig. 15 are centered near 7 ± 2 Å, compared to substantial populations centered at approximately half this distance in simulations with implicit solvent (Fig. 14).

VI. CONCLUSIONS

Results presented in this work indicate that proton exchange with the solvent by titratable residues has significant effects on the structure and dynamics of the polypeptide chain. Moreover, the effects of including variable protonation states during MD simulations are complex: calculated $p\mathrm{K}_a$ values and the ensemble of structures are significantly different from those obtained with simulations using fixed protonation states, and the structural basis for the perturbed $p\mathrm{K}_a$ of Asp 2 is not simply a salt-bridge interaction with Lys 4. Overall, the effects observed in this study for a simple peptide are likely to be manifested in proteins as well. Consequently, improving the accuracy of MD simulations for proteins will rely on development of algorithms for simulations of polypeptide chain dynamics at constant pH.

ACKNOWLEDGMENTS

We thank Krzysztof Kuczera for help in running the CHARMM program. This work was supported by the State Committee for Scientific Research, Poland (KBN) (Grant No. 6P04A00121) (M.D. and J.M.A.) and by NIH Grant No. GM46869 (A.D.R.). Partial support of the European Union (Contract No. QLRI-CT-2002-90383) is also acknowledged (M.D. and J.M.A.).

[1] L. Stryer, *Biochemistry*, 4th ed. (Freeman, New York, 1995).
[2] H. Zhou and M. Vijayakumar, J. Mol. Biol. **267**, 1002 (1997).
[3] H.W.T. van Vlijmen, M. Schaefer, and M. Karplus, Proteins: Struct., Funct., Genet. **33**, 145 (1998).
[4] A.A. Gorfe, P. Ferrara, A. Caflisch, D.N. Marti, H.R. Bosshard, and I. Jelesarov, Proteins: Struct., Funct., Genet. **46**, 41 (2002).
[5] R. Bürgi, P.A. Kollman, and W.F. van Gunsteren, Proteins: Struct., Funct., Genet. **47**, 469 (2002).
[6] A.M. Walczak and J.M. Antosiewicz, Phys. Rev. E **66**, 051911 (2002).
[7] U. Börjesson and P.H. Hünenberger, J. Chem. Phys. **114**, 9706 (2001).
[8] A.M. Baptista, V.H. Teixeira, and C.M. Soares, J. Chem. Phys. **117**, 4184 (2002).
[9] C. Tanford and R. Roxby, Biochemistry **11**, 2192 (1972).
[10] S.T. Wlodek, J. Antosiewicz, and J.A. McCammon, Protein Sci. **6**, 373 (1997).
[11] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[12] M.E. Davis, J.D. Madura, B.A. Luty, and J.A. McCammon, Comput. Phys. Commun. **62**, 187 (1991).
[13] J.D. Madura, *et al.*, Comput. Phys. Commun. **91**, 57 (1995).
[14] J. Antosiewicz, Biophys. J. **69**, 1344 (1995).
[15] J. Antosiewicz, J.M. Briggs, A.E. Elcock, M.K. Gilson, and J.A. McCammon, J. Comput. Chem. **17**, 1633 (1996).
[16] A.D. MacKerell, Jr. *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
[17] P. W. Atkins, *Physical Chemistry* (Freeman, New York, 1994).
[18] M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).
[19] K. Wüthrich, *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
[20] W. Schaller and A.D. Robertson, Biochemistry **34**, 4714 (1995).