Langevin dynamics of proteins at constant *p*H

Aleksandra M. Walczak* and Jan M. Antosiewicz†
Department of Biophysics, Warsaw University, 02-089 Warsaw, Poland
(Received 24 May 2002; revised manuscript received 23 July 2002; published 20 November 2002)

An application of the Langevin dynamics algorithm for simulation of protein conformational equilibria at constant *p*H is presented. The algorithm is used to compute average protonation of titratable groups in ovomucoid third domain, as functions of *p*H, resulting in data, basically equivalent to the *p*H dependencies of chemical shifts obtained from multidimensional nuclear magnetic resonance (NMR) spectroscopy, for the protein titratable residues. The pK_a values obtained from the simulation are in reasonable agreement with experimental data. Possible improvements of this methodology, using achievements from other fields of mesoscopic biomolecular simulations, are also discussed.

DOI: 10.1103/PhysRevE.66.051911 PACS number(s): 87.15.Aa

I. INTRODUCTION

One of the important determinants of protein structure and function is related to the ability of side groups of some of its constituting amino acids to exchange protons with the environment [1]. Quantitatively, protonation equilibrium of a given functional group AH is characterized by its pK_a value, defined as the negative decimal logarithm of the acidic dissociation constant K_a for the proton dissociation reaction [2]:

$$AH \rightleftharpoons A^- + H^+. \tag{1}$$

Introducing the well known definition of *p*H,

$$p\mathrm{H} = -\log_{10} a_{H^+}, \tag{2}$$

where a_{H^+} is the activity of protons, we can express the free energy change for the deprotonation process at a given *p*H as in Ref. [3]:

$$\Delta G = 2.303 RT(p\mathrm{H} - pK_a), \tag{3}$$

which indicates that the pK_a has a physical meaning of such a *p*H, at which the probability of protonation of the group is 50%.

Titratable groups in proteins can be either *acidic*, bearing negative elementary charge when deprotonated (Glu, Asp, Tyr, Cys, C terminus), or *basic*, bearing positive elementary charge when protonated (Lys, Arg, His, N terminus). The pK_a values of these groups, in a molecular environment of the protein, usually lie between 2 and 12, which means that at any given *p*H of experimental interest, there are some functional groups that are almost permanently charged. These charged groups are the main source of electrostatic fields inside and outside the protein, and also frequently they are active in protein function.

The pK_a's of titratable groups in biopolymers can be measured, e.g., by multidimensional nuclear magnetic resonance (NMR) spectroscopy [4]. They can also be computed, however, theoretical prediction of the pK_a's of titratable groups in biopolymers is one of the most challenging tasks for theoretical molecular biophysics [5–10].

Among the different approaches developed to predict protonation equilibria in biomolecules, those based on the assumption that the difference in protonation behavior of a given group isolated in solution, for which the ionization constant is assumed to be known, and the protonation behavior in the biopolymer environment is purely electrostatic in origin, have become particularly popular. Calculations of the relevant electrostatic free energies can be based on the Poisson-Boltzmann model of the protein-solvent system and the finite-difference solution to the corresponding Poisson-Boltzmann equation [11].

The first full pK_a calculation using the finite-difference Poisson-Boltzmann (FDPB) method was reported in 1990 [12] in a study of lysozyme. This calculation was done using two different crystal forms of the protein. Both gave different sets of predicted pK_a values indicating that conformation and protonation equilibria are coupled to each other. The use of conformational ensembles generated by molecular dynamics (MD) simulations [10,13–15] or obtained based on NMR experiments [6,16] has improved the accuracy of pK_a predictions, compared with calculations on a single x-ray structure. However, this does not solve the problem of the coupling between the conformational and protonation equilibria because the structures were obtained assuming one fixed protonation state of the titratable residues, and for different protonation patterns of residues in a multisite polymer, different ensembles of structures are expected [17–19].

The first MD simulations at constant *p*H were described by Mertz and Pettitt [20]. Subsequently, two MD simulations at constant *p*H that included titration of functional groups in a protein [21] and in small amines [22] were described. In these methods, however, the protonation state was treated as a parameter that describes a continuous change from protonation to deprotonation. See also a comment on the later work published in Ref. [23]. It seems necessary to include the neutral and charged form of residues explicitly on order to obtain proper ensembles of structures at a given *p*H.

Here we present an approach, where explicit titration is coupled to generation of protein structures consistent with actual *p*H during the titration procedure. In this respect, the presented approach is consistent with the true titration ex-

*Electronic address: alex@exasp.biogeo.uw.edu.pl
†Electronic address: jantosi@biogeo.uw.edu.pl

periment, e.g., by the NMR spectroscopy. After completion of the present work, another explicit protonation method for simulating proteins at constant *p*H was published by Bürgi, Kollman, and van Gunsteren [24]. The algorithm presented by these authors is a combination of MD and Monte Carlo (MC) simulations used to generate a Boltzmann distribution ensemble of protonation states at a given *p*H. The free energies of protonating or deprotonating a residue are calculated through MD simulation, whereas MC steps sample the protonation states of the ionizable residues during an MD simulation [24].

Our algorithm represents a different approach to the problem of conformational and protonational couplings, and it is much less expensive regarding computational demands. The idea behind our approach is simple. It uses well established simulation methods: Langevin dynamics, finite-difference solution to the Poisson-Boltzmann equation describing a solute-solvent system, and a Monte Carlo sampling (LD/FDPB/MC). The whole simulated protein trajectory is divided into smaller parts. Each subperiod of the simulation is proceeded by the evaluation of protonation probabilities of titratable groups in the starting structure at a chosen *p*H. Having these probabilities, the next step is to decide what is the actual protonation state assumed in the simulation to follow by a Monte Carlo type approach. This step reflects the statistical nature of the proton exchange phenomena and the fact that we do not know, which particular group with a given protonation probability is protonated or deprotonated at the given moment of time. Therefore, we sample a random number from the uniform $\{0-1\}$ distribution. If the sampled number is greater than the probability of protonation for the given site, we accept the unprotonated state of this group for further simulation, otherwise, we consider it protonated. Next, the simulation is done for a defined period of time, the last structure is saved, and protonation probabilities are evaluated again. And the whole procedure is repeated until the simulation is done for the whole assumed period of time. In this way, we account for the coupling between the residue interactions, which induce structural changes and the protonation phenomena. The proposed method is very general and can be applied to different types of protein models. The initial testing of the procedure was carried out on the explicit atom model, which we present below.

This method can be also used with another type of MC sampling. A Monte Carlo program described elsewhere [25] results in a predefined number of protein protonation patterns with the lowest energies found during the Monte Carlo search at a given *p*H. Instead of sampling protonation of individual sites described above, another Monte Carlo sampling can be done, which selects the whole protonation pattern from the list with probability governed by their energies. The simulations using this variant of the MC sampling are currently under way.

II. SIMULATION METHODS

As the first step, an x-ray structure of a protein is taken from the Protein Data Bank (PDB) [26]. All modifications, minimizations, and MD calculations are performed with the CHARMM [27] package using the CHARMM22 [28] force field. Where necessary, we add residues or patches based on parameters of related molecules or groups found in the force field. The left hand side of Fig. 1 presents a ribbon model of the 56-residue ovomucoid third domain available from the PDB under access code 2OVO [29], employed in the present tests.

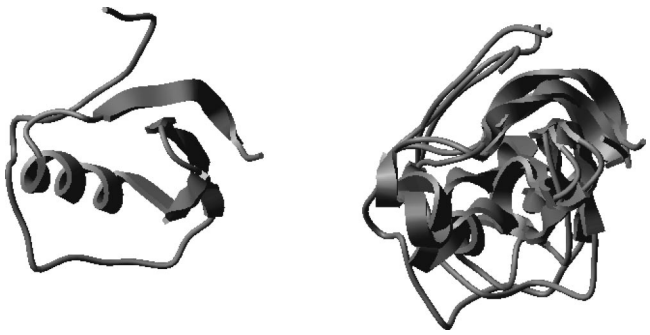

FIG. 1. Ribbon diagram of ovomucoid third domain. Left: original x-ray structure. Right: examples of structures generated for *p*H 1, 2, and 3.

Prediction of protonation states of titratable residues in investigated proteins is done according to procedures described elsewhere [30], using the UHBD [31,32] program for computing necessary electrostatic free energy interaction matrices and the HYBRID program [33] to compute titration curves for the titratable residues from the interaction matrix. We consider all titratable residues, i.e., *N* and *C* termini, Asp, Glu, Lys, Arg, His, and Tyr, as well as Cys, not involved in disulfide bonds.

For each titration run, an ensemble of 100 structures for each *p*H is generated by the Langevin dynamics method (LD) [34] implemented in CHARMM. All LD simulations are performed with a solvent viscosity of 10 ps^{-1}, a time step of 1 fs, and a dielectric constant of 15. A solute-solvent system should rather be represented by two dielectric constants, however the LD algorithm implemented in the CHARMM program, without the use of implicit solvation methods (see the Discussion), does not allow for this. The solute dielectric constant of 15 was used just to decrease effects of electrostatic interactions on the protein's structure, which in true solution are diminished by an aqueous solvent. Electrostatic calculations by the UHBD program also require specification of the solute dielectric constant. We try three values of 4, 8, and 15. Dielectric constant of the solvent was kept 80 in all UHBD simulations. See also the Discussion section.

The initial x-ray structure with assigned protonation states corresponding to $p\mathrm{H}=1$ undergoes minimization (100 steps for the steepest descent, SD, and 40 steps of conjugated gradient, CG, method) and dynamics: 15 ps of heating from 50 K to 400 K, 15 ps of cooling to 293 K, 20 ps (293–600 K), 20 ps (600–293 K), 10 ps (293 K to 450 K), 10 ps (450–293 K), 20 ps of equilibration at 293 K. Prior to titration, each structure is minimized with 50 SD and 50 CG steps. An ensemble of 100 structures is generated for each *p*H. In between these structures, minimization (100 SD and 40 CG steps) and short LD runs (1 ps equilibration) are performed to account for any subtle changes. For every 20 structures, a

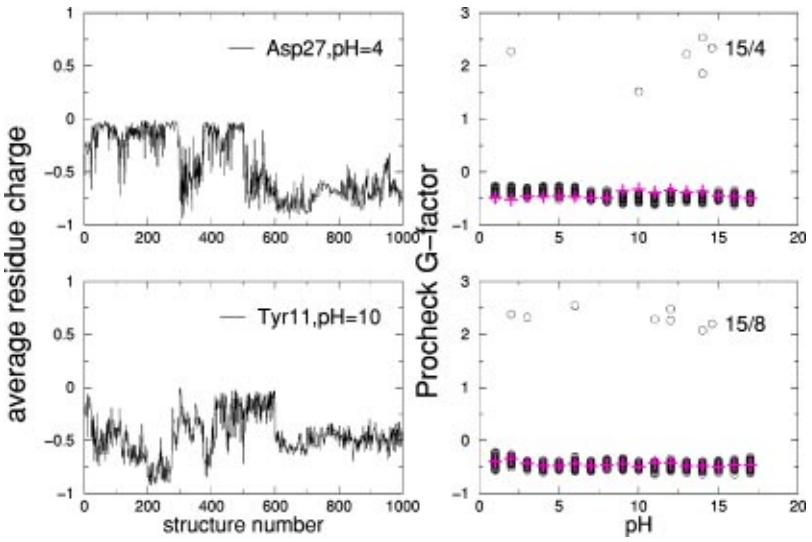

FIG. 2. Left: Average residue charge fluctuations at constant pH values of 4 and 10 for chosen residues of ovomucoid. The structures were generated with a heating procedure for every 100 structures. Right: The results of the PROCHECK test for a pH range from 1–17 for ovomucoid structures from the proposed titration procedure, with different dielectric constants. The crosses indicate the average value for each pH and the circles are values for particular structures.

heating and cooling run (100 steps of SD and 40 steps of CG minimization followed by 20 ps of heating to 400 K and 20 ps cooling to 293 K, 20 ps of equilibration) is added to enable the protein to sample different minima. When switching to a new pH, we allow the protein to undergo longer minimization (100 steps SD and 40 steps CG) and dynamics [20 ps (50–600 K), 20 ps (600–293 K), 15 ps (293 –450 K), 15 ps (450–293 K)], and 20 ps equilibration.

The whole procedure is repeated from $p\mathrm{H}=1$ to $p\mathrm{H}=17-20$. With modifications in the length of simulations and temperatures, it has been applied to tetrapeptides: Gly-Gly-*X*-Ala, where $X=$Tyr, Lys, Glu. As the result of such titration run average protonation fractions and their standard deviations as functions of pH for all titratable residues are stored, and used to determine $p\mathrm{K}_a$'s of the residues by fit of a Henderson-Hasselbalch equation [1] relating protonation fraction p, pH, and $p\mathrm{K}_a$:

$$p=\frac{1}{10^{p\mathrm{H}-p\mathrm{K}_a}+1}. \tag{4}$$

III. RESULTS

A. Generated structures of proteins

Examples of structures of ovomucoid third domain generated for pH 1, 2, and 3 are shown in the right hand side of Fig. 1. All these structures resemble the parent, x-ray structure. This is because of the three disulfide bonds kept intact in all simulations. Ovomucoid third domain is stable over a wide range of solution conditions [35–37], and it is properly folded even at such extreme pH values as 1.5 and 12.5 [38]. Therefore, it was rather justified to keep the disulfide bonds intact in our simulations. For proteins without such bonds, some other means like harmonic restraint potentials can be used in order to preserve the general fold of the protein. Without disulfide bonds or a harmonic restraint potential, the overall structure at low and high pH values would quickly unfold due to strong electrostatic repulsion. However, such restraints should be used with caution as not to prevent large conformational changes that would occur in nature at extreme solution conditions. This point will be discussed further below.

B. Stability and quality of generated structures

The presented Langevin dynamics algorithm at constant pH underwent several tests. First, two 1000-structure test runs at a constant pH value of 7 were done, one with heating every 100 structures, the other without. The predicted protonation factors for both types of simulation are very similar. Comparing the actual structures, we note that all structures are similar to that of the original x-ray structure, all the elements of secondary structure are conserved. However, those generated without prior heating tend to exhibit a drift from the original structure with generation time. The ones from the heated run are stable. The stability of the procedure was also checked by generation of 1000 protein structures at constant pH values of 4 and 10. The left hand side of Fig. 2 shows the sampling of local minima due to the heating procedure. These results show that these changes in structure result in minor protonation fraction, and thus average residue charge changes for pH values close to the $p\mathrm{K}_a$ value of a given residue. It proves the necessity of the heating and cooling runs during generation of an ensemble of structures for a given pH.

Second, structures generated in the above runs were checked with the PROCHECK program [39]. Most of the structures passed the test. The *G* factor, which is the measure of compliance of a structure with experimentally established norms for proteins (i.e., Ramachandran plots, covalent bonds) and ideally should be above -0.5 is presented in the right hand side of Fig. 2. The result is an average taken over all generated structures.

C. Test titrations of tetrapeptides

As a third test, titration of tetrapeptides containing Tyr, Lys, and Glu residues was done. The 12 titration curves obtained using different random number generator seeds for each tetrapeptide is presented in Fig. 3. In such a way, we

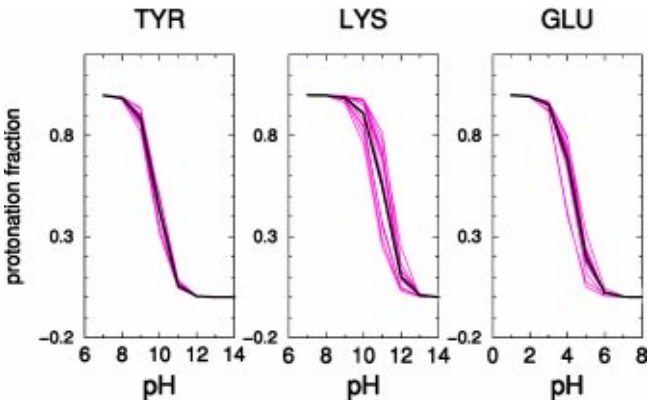

FIG. 3. The titration curves (12 thin gray lines) for tyrosine, lysine, and glutamic acid in the tetrapeptide obtained with $\epsilon_{LD}=15, \epsilon_{tit}=8$. The average (thick black line) results are $p\mathrm{K}_{aTyr}=9.84\pm0.26$, $p\mathrm{K}_{aLys}=11.0\pm0.6$, and $p\mathrm{K}_{aGlu}=4.3\pm0.4$ (at confidence level 0.95); the experimental values reported Ref. [40] are 10.1 ± 0.1 for Tyr, 11.0 ± 0.2 for Lys, and 4.3 ± 0.1 for Glu.

produced a trustworthy statistical ensemble of independent results. In case of Tyr, the curves are all similar and the standard deviation from the average $p\mathrm{K}_a$ value is 0.13, which is of the same order as the experimental error. The average $p\mathrm{K}_a$ result of 9.84 differs by 0.26 pH unit, i.e., twice standard deviation for 12 runs, from the experimentally determined 10.1 ± 0.1 of Ref. [40]. For Lys and Glu, our result of 11.1 ± 0.6 and 4.3 ± 0.4 are such as that of the mentioned experiment: 11.1 ± 0.2 and 4.3 ± 0.1, respectively.

D. Results for ovomucoid

Finally, titration simulations were done for ovomucoid third domain. An example of titration curves for Tyr 31, Glu 43, and Lys 55 are shown in Fig. 4, and all $p\mathrm{K}_a$ values obtained are listed in Table I. Figure 4, besides final titration curves, also visualizes the extent of fluctuations in protonation probability for chosen residues by presenting average protonation fractions for three different titration runs together with their standard deviations. It can be seen that

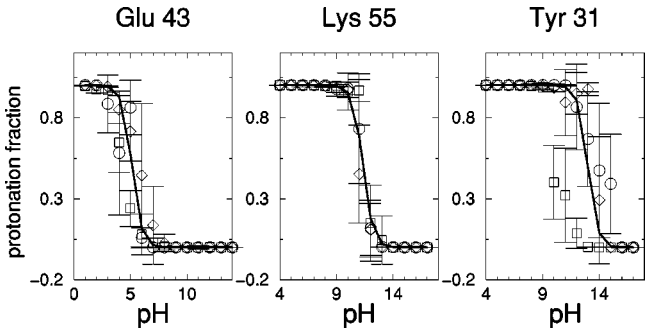

FIG. 4. The protonation fractions as functions of pH, obtained from three separate titration runs (marked with circle, square, and diamond symbols, together with their standard deviations) for tyrosine 31, glutamic acid 43, and lysine 55 in the ovomucoid obtained with $\epsilon_{LD}=15, \epsilon_{tit}=8$. The thick black line illustrates the fit of the Henderson-Hasselbalch equation to the pH dependence of the average protonation fractions from the three runs.

these fluctuations are rather small for Lys 55 and quite substantial for Tyr 31. Table I, besides results for three different dielectric constant pairs used in the LD simulation and titration, also presents results of titration of the original x-ray structure using methods described in Ref. [30] for three solute dielectric constants.

One can see that although there is a huge change in the computational method used for $p\mathrm{K}_a$ prediction in comparison to the traditional way of using FDPB methodology, the overall agreement with the experimental data is satisfactory. Yet our procedure is able to make predictions not available when using the traditional titration methodology. This is explained with the example of tyrosine residues. As it can be seen from Table I, the traditional FDPB methodology does not fit equally well, experimental $p\mathrm{K}_a$'s of tyrosines for one solute dielectric constant. With the solute dielectric constant of 4, the result for Tyr 20 is acceptable, but the results for Tyr 11 and Tyr 31 are by 4.3 and 5.5 pH units too large. When the solute dielectric constant is increased to 8 and 15, there is an improvement regarding Tyr 11 and Tyr 31 but the result for Tyr 20 is by 2 pH units too low. The order of titration of Tyr residues predicted by the traditional methodology seems to be coupled with solvent accessibility and does not agree with that of the experiment. Low solvent accessibility and/or interactions with other titratable groups may act to prevent deprotonation. But these factors can be overcome by conformational change in the protein forced by an increasing trend for a given group to deprotonate when the difference between the pH and $p\mathrm{K}_a$ values of a given group in an isolated state increases. This can cause the given group to deprotonate at a lower pH than predicted by calculations for the x-ray structure or any other rigid structure. This is clearly the result of our LD simulations in the case of the tyrosine residues visible in Table I. Using our titration method based on Langevin dynamics at constant pH, we get a much better agreement with experimental data for these residues. It can be seen that the order of titration of the three Tyr residues in LD simulations is different in each column, and for $\epsilon_{LD}/\epsilon_{tit}=15/8$, it is as observed in the experimental data. For this combination of dielectric constants, the agreement with the experimental data is the best.

The extent of the protein's structural fluctuations, on one hand, and the extent of systematic structural variation with changes in pH are illustrated in Fig. 5. It should be noted that the structures presented in this figure were obtained from continuous simulations going from pH 1, through all subsequent pH values. The end of the polypeptide chain visible on the right hand side is the C terminus. Although the sets of protein structures are not viewed from exactly the same perspective, it seems justified to indicate two following features which can be noticed: the first is the global movement of the C-terminal part of the protein, visible in the two-dimensional picture as an ''opening and closing'' of the ''cleft'' between the C-terminal part and the rest of the chain. The second feature, is the fact that the structure on the left hand side seems to be more stable as the pH is increased.

Structural changes during pH titrations can be also visualized by looking at the potential electrostatic interactions between the groups. These are the potential interactions, as

TABLE I. The $p\mathrm{K}_a$'s of titratable residues in ovomucoid computed for x-ray structure (columns 4–6), and for 100 LD structures (columns 7–9), with different solute dielectric constants, compared with experimental data [38]. Column 2 shows the solvent accessibility of each residue in the x-ray structure.

			X-ray structure			LD structure		
			ϵ			$\epsilon_{LD}/\epsilon_{tit}$		
Residue	% Solvent accessibility	Experimental $p\mathrm{K}_a$	4	8	15	15/4	15/8	15/15
NTER		8.0	8.1	7.9	7.7	7.6	7.9	7.6
ASP 7	58	<2.7	3.6	3.4	3.3	3.9	4.3	3.8
GLU 10	76	4.2	3.7	3.9	4.0	3.9	3.9	4.1
TYR 11	19	10.2	14.5	12.0	9.7	12.6	10.9	10.4
LYS 13	37	9.9	10.0	10.8	11.4	10.2	11.0	11.4
GLU 19	45	3.2	4.3	3.8	3.6	4.5	3.9	4.0
TYR 20	54	11.1	11.5	9.3	9.1	11.2	11.1	9.9
ARG 21			11.8	12.0	12.2	11.6	12.2	12.5
ASP 27	48	<2.3	5.9	4.7	4.1	4.0	3.9	3.7
LYS 29	47	11.1	11.4	11.3	11.2	10.6	11.7	11.0
TYR 31	14	>12.5	18.0	14.3	12.5	11.2	13.0	10.8
LYS 34	34	10.1	10.4	10.8	11.2	10.5	10.4	11.0
GLU 43	84	4.8	4.6	4.5	4.5	5.4	5.1	5.0
HSP 52	33	7.5	7.7	7.3	7.1	6.3	6.3	5.0
LYS 55	50	11.1	10.3	11.4	11.3	11.3	11.3	11.3
CTER		<2.5	3.1	3.2	3.3	3.7	3.0	3.5

their energies contribute to the protein's free energy only when both interacting residues are simultaneously ionized. Figure 6 portrays how the potential interactions of Glu 43 with Lys 29 and Lys 55 vary with the change of pH as a result of accompanying changes in the protein structure, for the three independent titration runs. Results obtained for a particular simulation are displayed on the same level in the figure. It can be seen that the protein went through different regions of the phase space in these three simulations during titration. In the first run, structures at pH around 10 enabled strong interactions between Glu 43 and Lys 29, with negligible interactions between Glu 43 and Lys 55, whereas in the third run, the situation was just the opposite. In the second run, neither of these interactions were significant. However, the sampled structures exhibited another significant interaction of Glu 43 with a basic titratable group, namely His 52 (data not shown). This indicates that satisfactory sampling of protein structures as a function of pH might require more runs. Nevertheless, it seems that these results prove the usefulness of LD simulations at constant pH.

Coming back to the data shown in Table I, we can make some more observations. The results obtained by us are consistent with previous findings of Ref. [41]: Lys 34 does not interact with Glu 19; Arg 21 seems to be responsible for stabilizing interactions with Glu 19 (O—H distance in the appropriate pH range is about 2.4 Å). The distance between the oxygen in Asp 7 and backbone amid, as well as the acid chain hydroxyl of Ser 9 predicts the forming of hydrogen bonds. These results also agree with those of Ref. [41]. It is also worth noting, that in the 15/8 simulation all Tyr, Glu, and Asp residues titrate in the same order as in experiment. Predictions for all residues, but Asp 7, Asp 27, and Lys 13 are consistent with the experiment. This inaccuracy is due to a statistical fluctuation and/or omission of some features in our simulation (see below).

IV. DISCUSSION

We have presented an algorithm to simulate proteins at constant pH value and application of this algorithm in prediction of $p\mathrm{K}_a$ values of protein titratable residues. It seems that LD dynamics of proteins, together with the traditional FDPB methodology for titration is a very promising computational tool for prediction of protonation equilibria in biopolymers.

The most important result of the present work is the significant improvement in prediction of $p\mathrm{K}_a$ values of tyrosine residues in comparison to the traditional FDPB methods. If this would be observed for other proteins, remains to be checked. However, there are examples of poor results of application of the traditional FDPB methodology for tyrosine residues in other proteins, e.g., for Tyr 53 in hen egg white lysozyme (HEWL) [9,12], Tyr 25, and Tyr 97 in ribonuclease *A* [30]. In these cases, our method may prove successful. It can be also noted that for the remaining residues, the agreement with experimental data is not worse than that obtained with the traditional FDPB method. And there are numerous ways of improvement. For example, 15/8 simulations predict for Asp 7, $p\mathrm{K}_a=4.3$, whereas experimental value is below 2.7. However, LD simulations with fixed protonation states where Asp 7 was deprotonated resulted in ensemble of structures for which the average $p\mathrm{K}_a$ of Asp 7 was below 2. Similar behavior was observed for Asp 27. Therefore, it seems that the elevated $p\mathrm{K}_a$ values obtained for these residues with free movement in protonation degrees of freedom, together with conformational sampling results from some de-

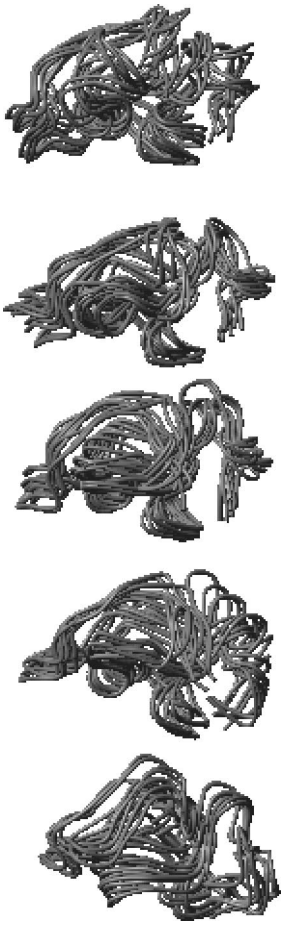

FIG. 5. Ribbon diagrams of ovomucoid third domain structures sampled from simulations at five consecutive pH values, from pH 5 (top) to pH 9 (bottom). Every tenth structure out of one hundred structures sampled at each pH value is shown.

ficiencies in the LD algorithm used for structure generation. One obvious problem is that these simulations were done without realistic aqueous solvent and ionic strength effects. It can be expected that inclusion of these effects should help a lot.

It should be noted that the best results of the present work were obtained for assumed solute dielectric constant, larger than more commonly accepted values of 2–4. As was already written above, the solute dielectric constant used in LD simulations serves, to some extent, as a crude way to decrease the significance of electrostatic interactions. Here, we would rather like to discuss the solute dielectric constant used in FDPB calculations. An extensive account on the history of the problem of solute dielectric constants of solvated proteins was given in a recent review by Schutz and Warshel [42]. As they stated, the protein dielectric constant is simply a parameter that depends on the model used. In an interesting study, introducing a new concept of the reference model compound, used in the FDPB method for $p\mathrm{K}_a$ calculations, Grycuk has indicated [43] that parametrization of this method in terms of the ionization constant, the $p\mathrm{K}_{a,model}$ for the so called model group depends on atomic partial charges and radii of the solute molecule, on the definition of the dielectric boundary between solute and solvent, and on the solute dielectric constant. Having the charges, radii and the boundary defined independently, one is left with the solute dielectric constant and the $p\mathrm{K}_{a,model}$ as the only adjustable parameters in the model being applied for prediction of protonation equilibria by the FDPB method. Having experimental $p\mathrm{K}_a$ values for a number of small, test molecules containing the same model group, one can choose such a solute dielectric constant and corresponding $p\mathrm{K}_{a,model}$, which give the best reproduction of the experimental $p\mathrm{K}_a$'s for all the test molecules. Such a procedure leads to a solute dielectric constant of 6–8, i.e., substantially above values expected on the basis of only electronic polarization contribution. And this is obtained with inclusion of full conformational sampling for molecular structures of test compounds. When this is used for proteins, the optimal solute dielectric constant is even higher, i.e., 12–14 [43,44]. Therefore, the fact that the best results in this study were obtained for solute dielectric constant of 8 in comparison to 4 and 15 is not surprising. However, if 15/8 is the most optimal pair of solute dielectric constants used in LD and FDPB calculations, it remains to be established by studies for other proteins. It should also be

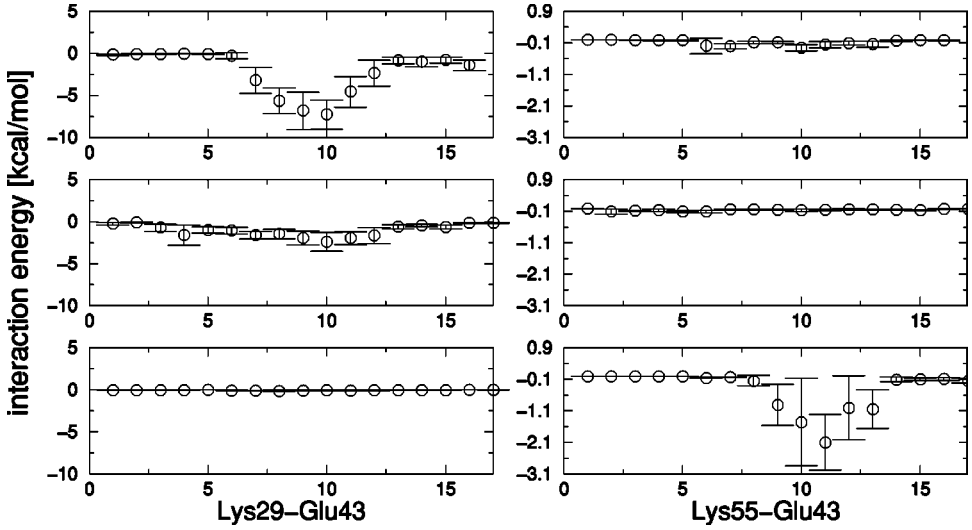

FIG. 6. The change in potential interactions of Glu 43 with Lys 29 and Lys 55, with the change of structure when we change the pH of the environment.

remembered that inclusion of implicit solvation and ionic strength effects in the algorithm certainly will affect 15 as dielectric constant used in LD simulations, as discussed above.

As mentioned in the Introduction, after completion of this work, a paper by Bürgi, Kollman, and van Gunsteren on a similar subject has appeared [24]. Their algorithm combines MD simulations and MC sampling, and as the algorithm described in the present work, it explicitly accounts for dissociated or associated protons. These authors applied their method to HEWL, which has approximately twice the number of residues as ovomucoid third domain, and they performed only three simulations at *p*H values 4, 3, and 2, respectively. Also they considered only titration of carboxylic side groups and that of the C terminus in their study. It seems that their method is much more demanding regarding computer resources than the method presented in this work. With our algorithm, one titration run for ovomucoid third domain from *p*H 1 to *p*H 20 takes 180 h of CPU time on a PC with 1.5-GHz processor working under Linux operating system. Moreover, its accuracy in predicting pK_a values of titratable residues in proteins is not better than that of the algorithm presented in this work. Therefore, it seems that regardless the future of the method by Bürgi, Kollman, and van Gunsteren, the LD/FDPB/MC algorithm presented here is worthy of further work and development.

As it was already mentioned above, the method presented in this work can be extended and/or improved in many ways, using ideas described already by others, although their implementation in the algorithm described above requires a substantial amount of work and is not by any means straightforward and simple. Among these, are inclusions of solvent accessibility and ionic strength in the LD protocol [45]. In CHARMM there is a possibility of adding implicit solvent interactions by means of the generalized Born model [46], however, this elongates the simulation time by a factor of 5. Other methods (effective energy function, [47] and analytical continuum solvent, [48]) are not available with the 22 force field and nonpolar hydrogens. Due to these implementation problems, implicit solvation effect was not employed at the stage of constructing the whole algorithm described in the present paper. Inclusion of implicit solvation effects is related to the problem of stability of the protein during simulations. We expect that structural constraints mentioned above could be significantly reduced or even removed as in the MD simulations described by Bürgi, Kollman, and van Gunsteren [24], when solvent effects are included in a more realistic way. In such a case, the algorithm presented in this work could be used to study structural changes in proteins after *p*H jumps or changes in structural and protonation equilibria in protein ligand association processes and influence of these on the association rates by calculations, similar to computation of rate constants by Brownian dynamics simulations [31]. Finally, it seems worthy to incorporate the new concept of the model group described recently by Grycuk [43] and changes in hydration entropy upon protonation, described by Warwicker [49].

ACKNOWLEDGMENTS

This work was financially supported by Fogarty, USA (Grant No. TW00768), and by the State Committee for Scientific Research, Poland (KBN) (Grant No. 6P04A00121). Molecular structures were visualized using ICMLITE [50].

[1] L. Stryer, *Biochemistry*, 4th ed. (Freeman, New York, 1995).
[2] P.W. Atkins, *Physical Chemistry* (Freeman, New York, 1994).
[3] D. Poland, *Cooperative Equilibria in Physical Biochemistry* (Clarendon Press, Oxford, 1978).
[4] K. Bartik, C. Redfield, and C.M. Dobson, Biophys. J. **66**, 1180 (1994).
[5] A. Warshel, Biochemistry **20**, 3167 (1981).
[6] J. Antosiewicz, J.A. McCammon, and M.K. Gilson, Biochemistry **35**, 7819 (1996).
[7] R.A. Dimitrov and R.R. Crichton, Proteins: Struct., Funct., Genet. **27**, 576 (1997).
[8] Y.Y. Sham, Z.T. Chu, and A. Warshel, J. Phys. Chem. B **101**, 4458 (1997).
[9] H.W.T. van Vlijmen, M. Schaefer, and M. Karplus, Proteins: Struct., Funct., Genet. **33**, 145 (1998).
[10] A.A. Gorfe, P. Ferrara, A. Caflisch, D.N. Marti, H.R. Bosshard, and I. Jelesarov, Proteins: Struct., Funct., Genet. **46**, 41 (2002).
[11] J. Warwicker and H.C. Watson, J. Mol. Biol. **157**, 671 (1982).
[12] D. Bashford and M. Karplus, Biochemistry **29**, 10219 (1990).
[13] D. Bashford and K. Gerwert, J. Mol. Biol. **224**, 473 (1992).
[14] H. Zhou and M. Vijayakumar, J. Mol. Biol. **267**, 1002 (1997).
[15] L. Sandberg and O. Edholm, Biophys. Chem. **65**, 189 (1997).
[16] V. Dillet, H. Dyson, and D. Bashford, Biochemistry **37**, 10298 (1998).
[17] C.H. Pletcher, E.F. Bouhoutsos-Brown, R.G. Bryant, and G.L. Nelsestuen, Biochemistry **20**, 6149 (1981).
[18] D.R. Ripoll, Y.N. Vorobjev, A. Liwo, J.A. Vila, and H.A. Scheraga, J. Mol. Biol. **264**, 770 (1996).
[19] S.T. Wlodek, J. Antosiewicz, and J.A. McCammon, Protein Sci. **6**, 373 (1997).
[20] J.E. Mertz and B.M. Pettitt, Int. J. Supercomput. Appl. **8**, 47 (1994).
[21] A.M. Baptista, P.J. Martel, and S.B. Petersen, Proteins: Struct., Funct., Genet. **27**, 523 (1997).
[22] U. Börjesson and P.H. Hünenberger, J. Chem. Phys. **114**, 9706 (2001).
[23] A.M. Baptista, J. Chem. Phys. **116**, 7766 (2001).
[24] R. Bürgi, P.A. Kollman, and W.F. van Gunsteren, Proteins: Struct., Funct., Genet. **47**, 469 (2002).
[25] J. Antosiewicz, Biophys. J. **69**, 1344 (1995).
[26] F.C. Bernstein, T.F. Koettzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, and M.J. Tasumi, J. Mol. Biol. **123**, 557 (1977).
[27] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[28] A.D. MacKerell, Jr. *et al.*, J. Phys. Chem. B **102**, 3586 (1998).

[29] W. Bode, O. Epp, R. Huber, M. Laskowski, Jr., and W. Ardelt, Eur. J. Biochem. **147**, 387 (1985).
[30] J. Antosiewicz, J.M. Briggs, A.E. Elcock, M.K. Gilson, and J.A. McCammon, J. Comput. Chem. **17**, 1633 (1996).
[31] M.E. Davis, J.D. Madura, B.A. Luty, and J.A. McCammon, Comput. Phys. Commun. **62**, 187 (1991).
[32] J.D. Madura *et al.*, Comput. Phys. Commun. **91**, 57 (1995).
[33] M.K. Gilson, Proteins: Struct., Funct., Genet. **15**, 266 (1993).
[34] G. Widmalm and R.W. Pastor, J. Chem. Soc., Faraday Trans. **88**, 1747 (1992).
[35] W. Ardelt and J.M. Laskowski, J. Mol. Biol. **220**, 1041 (1993).
[36] L. Swint-Kruse and A.D. Robertson, Biochemistry **34**, 4724 (1995).
[37] C.B. Arrington and A.D. Robertson, Biochemistry **36**, 8686 (1997).
[38] W.R. Forsyth, M.K. Gilson, J. Antosiewicz, O.R. Jaren, and A.D. Robertson, Biochemistry **37**, 8643 (1998).
[39] R.A. Laskowski, M.W. MacArthur, D.S. Moss, and J.M. Thornton, J. Appl. Crystallogr. **26**, 283 (1993).
[40] A. Bundi and K. Wüthrich, Biopolymers **18**, 285 (1979).
[41] W.R. Forsyth and A.D. Robertson, Biochemistry **39**, 8067 (2000).
[42] C.N. Schutz and A. Warshel, Proteins: Struct., Funct., Genet. **44**, 400 (2001).
[43] T. Grycuk, J. Phys. Chem. B **106**, 1434 (2002).
[44] M. Wojciechowski, T. Grycuk, J.M. Antosiewicz, and B. Lesyng, Biophys. J. (to be published).
[45] K.A. Sharp, J. Comput. Chem. **12**, 454 (1990).
[46] B.N. Dominy and C.L. Brooks, Jr., J. Phys. Chem. B **103**, 3765 (1999).
[47] T. Lazardis and M. Karplus, Proteins: Struct., Funct., Genet. **35**, 133 (1999).
[48] M. Schaefer, C. Bartels, and M. Karplus, J. Mol. Biol. **284**, 835 (1998).
[49] J. Warwicker, Protein Eng. **10**, 809 (1997).
[50] R.A. Abagyan, M.M. Totrov, and D.N. Kuznetsov, J. Comput. Chem. **15**, 488 (1994).