The impact of protonation equilibria on protein structure

2005 J. Phys.: Condens. Matter 17 S1607

(http://iopscience.iop.org/0953-8984/17/18/016)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **17** (2005) S1607–S1616 doi:10.1088/0953-8984/17/18/016

The impact of protonation equilibria on protein structure

Maciej Długosz and Jan M Antosiewicz[1]

Department of Biophysics, Warsaw University, Zwirki i Wigury 93, 02-089 Warsaw, Poland

E-mail: jantosi@biogeo.uw.edu.pl

Received 7 October 2004, in final form 21 January 2005
Published 22 April 2005
Online at stacks.iop.org/JPhysCM/17/S1607

Abstract

All proteins contain residues capable of exchanging protons with their environment. The significance of these proton transfer phenomena for protein structure and function has been documented in a number of experimental and theoretical studies. These studies, in principle, were using changes in pH of a protein's environment to induce shifts in the protonation equilibria and explained observed changes in protein's properties by changes in the ionization states (i.e., formal electrical charges) of acidic or basic chemical groups of the molecule, accompanying these shifts.

In this study, we address the problem of significance of protonation equilibria for protein structure at a single value of pH by methods of molecular dynamics at constant pH, applied to a heptapeptide derived from ovomucoid third domain (OMTKY3). We show that the average molecular properties characterizing the equilibrium distribution of the polypeptide conformations obtained from these simulations cannot be reproduced as a linear combination of the properties resulting from the distributions generated by molecular dynamics simulations with fixed protonation states of the titratable residues. This means that consideration of dynamic conformational equilibria of proteins without taking into account explicitly the degrees of freedom resulting from proton exchange phenomena may lead to significant artefacts.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

All proteins contain residues capable of exchanging protons with the environment [1]. These proton transfer phenomena lead to changes in ionization states (i.e., formal electrical charges) of acidic or basic chemical groups and constitute the principal factor determining the charge

[1] Author to whom any correspondence should be addressed.

distribution within a protein molecule. Because of the long-range nature of electrostatic interactions, changes in distribution of molecular charge, resulting from proton transfer phenomena, are expected to have a significant impact on the three-dimensional structure of the polypeptide chain. The most dramatic examples are represented by well known acidic or alkaline denaturation of proteins.

The significance of protonation equilibria for protein structure and function has been documented in a number of experimental and theoretical studies, mainly related to observations of the consequences of changing the solution pH or substituting an unusable residue in protein by a nontitratable one. Some recent interesting examples include an analysis of the pH dependence of protein stability [2–4], triggering of the conformational change of hemagglutinin and membrane fusion by a pH-drop in the endosome during influenza virus infection [5], and other examples of proton-mediated conformational transitions [6–9].

Protonation equilibria are quantitatively characterized by $p\mathrm{K}_a$ values of the titratable groups, which are defined as the negative decimal logarithm of the proton dissociation constant. The $p\mathrm{K}_a$ values for titratable groups can be conveniently measured using well established methods of NMR spectroscopy [10–12]. On the other hand, structure-based theoretical methods for the prediction of $p\mathrm{K}_a$s of titratable groups in proteins have been developed, the most popular being those using the Poisson–Boltzmann model for a solute–solvent system [13–18], and Langevin dipole models [19–21]. These and some others are discussed in several recent reviews [22–25].

The problem of proper inclusion of protonation equilibria in biopolymers is beyond doubt important for their simulations at the microscopic level, using for example molecular dynamics (MD) methods [26]. The effects of pH in MD simulations have generally been treated, until recently, in a simplified way: depending on the selected pH and assumed values of ionization constants, either the protonated or unprotonated state of each titratable group was selected and used for the whole simulation. The most realistic inclusion of solute–solvent proton exchange phenomena into MD requires that titratable groups of the solute molecule and some number of surrounding aqueous solvent molecules are described at the quantum mechanical level. The number of water molecules which need to be included can be extremely large, e.g. if the concentration of the hydrogen ions must be appropriate for the physiological pH values, and this would make corresponding computer simulations impossible to perform at present. Problems of the size can be overcome when calculations of $p\mathrm{K}_a$ values, using quantum mechanical treatments of the solute electronic structure, are combined with a dielectric continuum model for the solvent [27, 28]. Such an approach has also been applied for proteins [29, 30].

During the last few years several constant-pH MD algorithms have been proposed. In some of these algorithms the protonation state was treated as a parameter that describes a continuous change from protonation to deprotonation [31–34], but the majority of the recent work involves discrete protonation states. Bürgi *et al* [35] described an algorithm which combines molecular dynamics with given protonation states of the residues, evaluation of the free energy change accompanying a protonation or deprotonation of a sample residue by thermodynamic integration methods, and Monte Carlo steps for acceptance or rejection of the resulting new states. Baptista *et al* [36] described an explicit solvent molecular dynamics approach combined with the Poisson–Boltzmann model for the evaluation of protonation probabilities. Mongan *et al* [37] proposed a model using a generalized Born (GB) implicit solvation MD algorithm with periodic Monte Carlo sampling of protonation states based on GB-derived energies. Our group initially used constant dielectric Langevin dynamics combined with PB electrostatics [38], and recently switched to implicit solvent MD methods [39, 40].

Current MD algorithms for simulations at constant pH are still in their developmental state and there are many conceptual and/or technical problems to be solved. One may ask if the

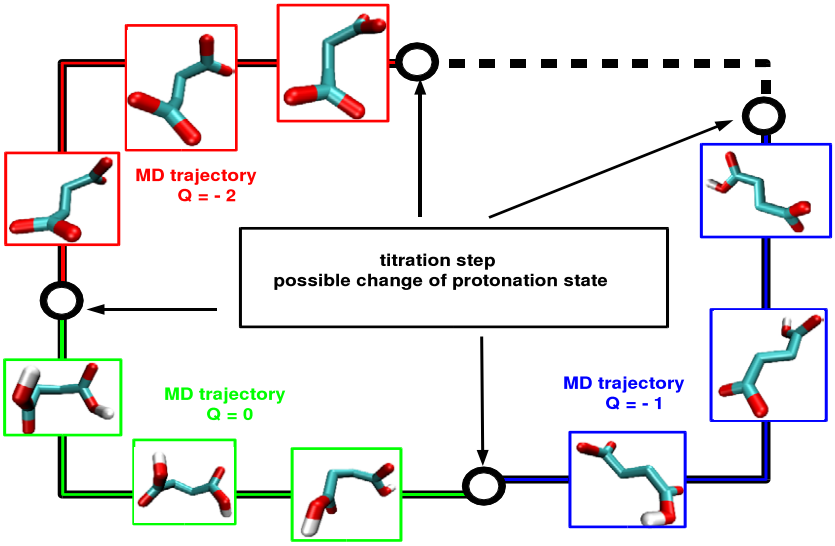

Figure 1. Outline of the MD simulation procedure at constant pH.

problem of proton exchange by titratable groups of the polymer can be successfully treated by running traditional, physically sound, MD simulations for different, most probable protonation patterns of the whole molecule and averaging the results with appropriate weights. A positive answer to this question would significantly simplify the modelling of pH-dependent properties of molecules.

This study presents a consideration of this problem based on an application of a constant-pH molecular dynamics algorithm developed by our group [39, 40] to the analysis of coupling between conformational and protonation equilibria in a heptapeptide, acetyl-Ser–Asp–Asn–Lys–Thr–Tyr-Gly-amide, derived from turkey ovomucoid third domain, OMTKY3 (residues 26-32) [41]. We show that the averaging of results of traditional MD simulations with fixed protonation states is rather not to be expected to correctly predict results of constant-pH MD simulations.

2. Methods

2.1. Constant-pH molecular dynamics

A detailed description of our algorithm for constant-pH MD has been presented elsewhere [39, 40], hence only a brief description is given here. The basic idea behind our approach is outlined in figure 1. According to figure 1, the whole constant-pH trajectory is divided into a set of sub-trajectories. Those sub-trajectories generated with standard constant temperature molecular dynamics protocol cover a predefined period of time (one picosecond in the present case). The last structure from the finished sub-trajectory undergoes titration. The protonation fractions and the most probable protonation patterns for the titratable residues are evaluated based on the free energies of the states, at a given pH, using a Monte Carlo procedure [40, 42], the protonation state of the molecule is eventually changed (by explicit proton addition or removal), and a new sub-trajectory is started. The whole procedure is repeated until the constant-pH trajectory covers a desired length of time.

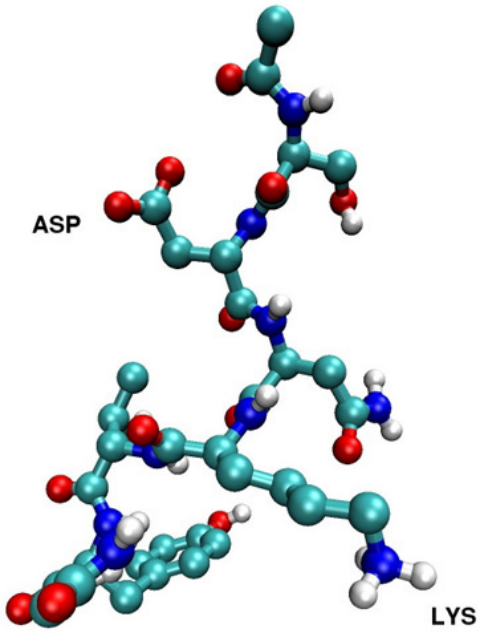

Figure 2. Structure of the heptapeptide, acetyl-Ser–Asp–Asn–Lys–Thr–Tyr–Gly-amide (residues 26-32 of turkey ovomucoid third domain, OMTKY3). (The drawing was done with program VMD [46].)

The above algorithm was applied to a short peptide derived from ovomucoid third domain (OMTKY3), acetyl-Ser–Asp–Asn–Lys–Thr–Tyr–Gly-amide (residues 26-32 of OMTKY3) shown in figure 2, investigated already in a previous study [39]. There are three titratable residues in this molecule—Asp 2, Lys 4, and Tyr 6. The new simulations, discussed in the present work, were focused on elucidation of the coupling between conformational and protonation equilibria in this molecule. The constant-pH MD simulations described here were performed for pH values of 3, 4, 5, and 6, for two forms of Lys 4 residue—one usual with its side group free to exchange protons with the environment, and the other with the side group neutralized by acetylation. For the pH values selected in this study, normal Lys 4 residue, and Tyr 6 residue should remain protonated, so their protonation states were fixed. Therefore, only Asp 2 residue was allowed to change protonation state during the constant-pH MD simulations. As the result, we consider three systems: heptapeptide with protonable Asp 2, heptapeptide with protonated Asp 2 and heptapeptide with deprotonated Asp 2. The second and the third system can be treated as two limiting cases of the first one (they correspond to extremely low and high values of pH). The main purpose of this work is to answer the question if the properties of the first system can be represented as a linear combination of the properties of the second and the third system. Based on our results we conclude that in general this is not possible.

2.2. Simulations details

For the generation of ensembles of the heptapeptide structures with fixed protonation states, we performed molecular dynamics simulations for both variants (normal and acetylated lysine residue) of the heptapeptide for two fixed protonation states of the Asp 2 side chain (protonated

and deprotonated). Those simulations resulted in four trajectories of 5 ns length each. From each of these trajectories 1000 structures, equally spaced in time, were recorded.

The procedure described above gave us two ensembles of 2000 structures each, corresponding to two variants of the Lys residue. We used 80 structures from each of those ensembles as a starting point for generation of constant-pH trajectories, generating 80 constant-pH trajectories, each of 50 ps length, for both variants of the lysine residue, at three pH values. The first half of each trajectory was rejected to assure that generated structures are independent of the initial one and last 25 structures equally spaced in time were saved for further analysis. As a result we constructed an ensemble of 2000 structures for each variant of the studied heptapeptide for each value of pH of interest, covering the total time of 2 ns. Because we used many different starting points, the sampling efficiency of these 2 ns is expected to be satisfactory.

As mentioned already above, an attempt of possible change of protonation state of the simulated heptapeptide was done every 1 ps. In the constant-pH molecular dynamics simulations referenced above this 'sampling' time was in the range 0.1–30 ps. Moreover, in some studies it was shown that 'sampling' times between 0.2 and 5 ps lead to reasonably similar results [36, 39, 40]. On the other hand, based on MD simulations of proton transport with quantum mechanically derived proton hopping rates, the proton transfer rates were estimated at about 0.5 ps^{-1} [43].

The constant-pH MD simulations presented in this work are quite time consuming. Collection of all the results took about three months of almost continuous work on two PC workstations each with an AMD Athlon XP 3200+ processor. However, it seems that the length of the simulated trajectories is sufficient to justify the conclusions described below. Ultrafast spectroscopy methods revealed sub-nanosecond conformational dynamics of a Phe–Gly–Asp–Cys–Thr–Ala–Cys–Ala peptide, backbone-cyclized with (4-aminophenyl)-azobenzoic acid [44], and molecular dynamics studies of the reversible folding/unfolding of short Ala and Gly-based peptides indicated that the formation of a first alpha-helical turn occurs within 0.1–1 ns [45].

2.3. Analysis of molecular dynamics trajectories

Based on the structures recorded from molecular dynamics trajectories at constant pH, we collected predicted protonation probabilities of the Asp 2 side group and distances between atom CG of Asp 2 and atom NZ of Lys 4 (the atom names are taken as they appear in the Protein Data Bank structural files). Both these quantities constitute two of many possible characteristics of conformational preferences of the heptapeptide chain. We present them in a form of distributions. In the results section we compare these distributions obtained from constant-pH MD simulations with those obtained from MD simulations with fixed protonation state of Asp 2 residue of the heptapeptide. We perform this comparison for both forms of Lys 4 residue: the normal one—permanently charged, and the acetylated one—permanently neutral.

3. Results

3.1. Distributions of protonation fractions of Asp 2 side group

Figure 3 presents distributions of the protonation fractions of the Asp 2 side group obtained from constant-pH MD simulations at selected pH values. Because these protonation fractions depend on the molecular environment of the titratable groups, they are undoubtedly related to the conformational preferences of the polypeptide chain. It can be seen that both distributions

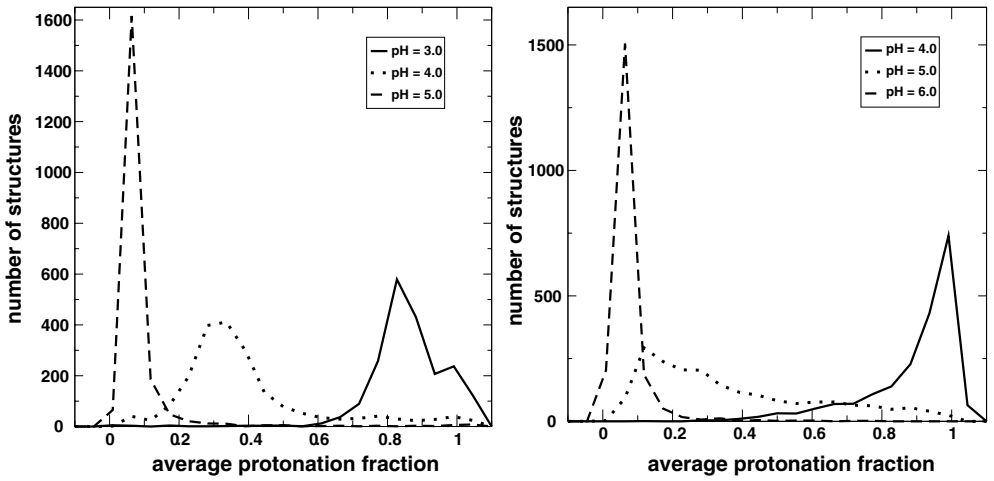

Figure 3. Distribution of protonation fractions of the Asp 2 side group for normal (left-hand side) and acetylated (right-hand side) Lys 4 residue, for selected pH values, obtained from constant-pH MD simulations.

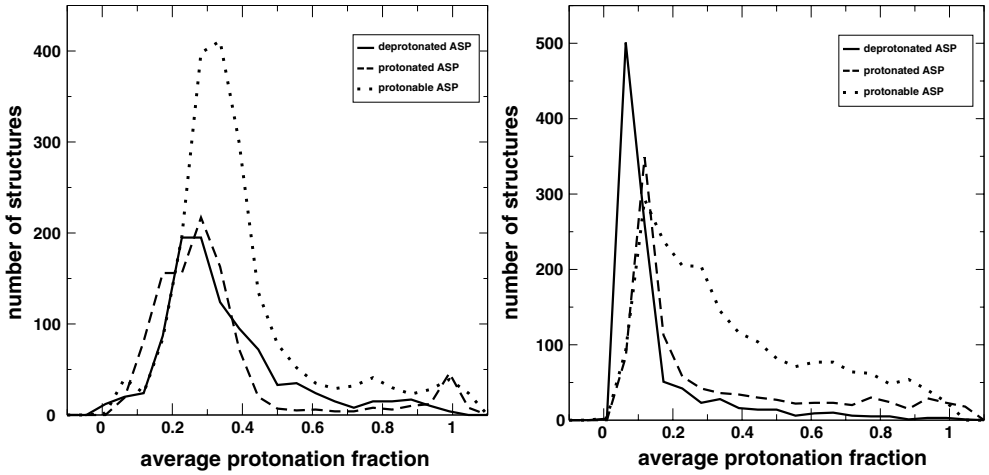

Figure 4. Distribution of protonation fractions of the Asp 2 side group from constant pH and from fixed protonation states MD simulations. Left: normal Lys 4 simulations, pH = 4. Right: acetylated Lys 4 simulations, pH = 5.

change their position from a high proton occupancy to a low proton occupancy with an increase in the pH value, and that for the peptide with normal Lys 4 residue this transition occurs approximately at pH values that are higher by 1 unit. Subsequently, figure 4 presents a comparison between distributions of the protonation fractions of the Asp 2 side group obtained at pH 4, for normal Lys 4, and pH 5, for acetylated Lys 4, and the corresponding distributions obtained from MD simulations with fixed protonation states of Asp 2, either permanently protonated or permanently deprotonated. It should be noted at this point, that although the ensemble of structures was generated with fixed protonation states of Asp 2, the probability of protonation is still a function of the pH. The fixed protonation state MD simulations lead to structures with a variety of protonation probabilities of the residues because

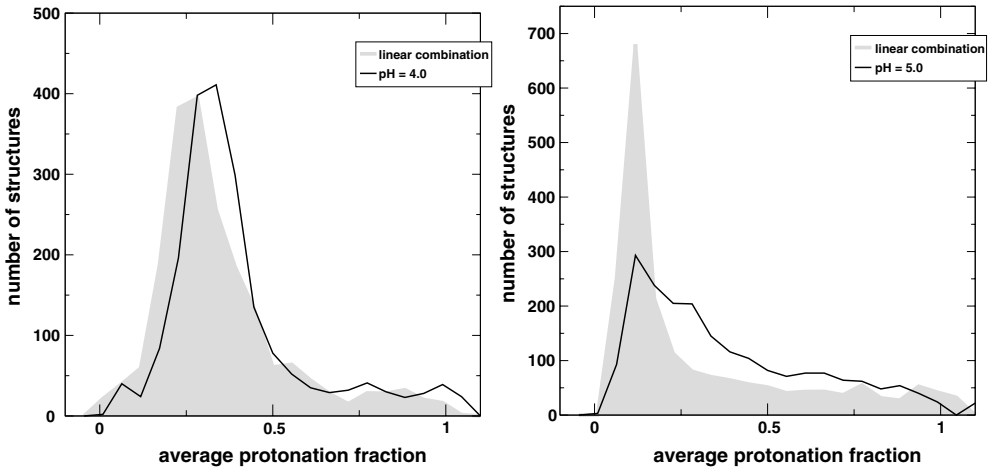

Figure 5. Distribution of protonation fractions of the Asp 2 side group obtained from constant-pH simulations (black solid curve) compared with the closest, regarding the sum of squared deviations, linear combination of distributions obtained from MD simulations with fixed protonation states (grey contour filled as a polygon). Left: normal Lys 4 simulations, pH = 4. Right: acetylated Lys 4 simulations, pH = 5.

of thermal fluctuations in the structures, which need not necessarily always be consistent with the imposed protonation states. It can be seen that distributions obtained from constant-pH MD simulations extend beyond areas covered by distributions obtained from MD simulations with fixed protonation states of the residues. Therefore, the distribution obtained from constant-pH MD simulation is not an intermediate distribution between two limiting cases, one with protonated Asp and the other with deprotonated Asp. Figure 5 presents a comparison between the distribution computed as the best linear combination, regarding the sum of squared deviations, of the two limiting distributions obtained with the fixed protonation states, and the distribution obtained from constant-pH MD simulation. We must conclude that the probability distribution of the heptapeptide structures in the conformational phase space available for the molecule cannot be reproduced if we know the probability distributions for the two limiting cases.

In order to see if this conclusion is not limited to the case of protonation fractions as a particular 'mirror' of conformational preferences of the polypeptide chain, in the next subsection we perform a similar analysis for another quantity related to these preferences.

3.2. Distributions of Asp–Lys distances

Another property reflecting conformational preferences of the heptapeptide chain is the distance between atom CG of the Asp 2 residue and atom NZ of the Lys 4 residue. This choice seems to be obvious, as in the case of the peptide with normal Lys residue one can expect an influence of Asp and Lys side chain interactions (in charged states) on simulated structures. Moreover, the distances represent a different kind of characteristic, as those generated for fixed protonation states of Asp 2 are pH independent. Figure 6 presents distributions of the Asp–Lys distances obtained at several pH values from constant pH trajectories, as well as those from the fixed state simulations. Generally, it can be seen that in both discussed cases (normal and acetylated Lys 4) the average Asp–Lys distances are shifted towards smaller values for distributions obtained from constant-pH simulations. As in the previous subsection, figure 7 shows that the results

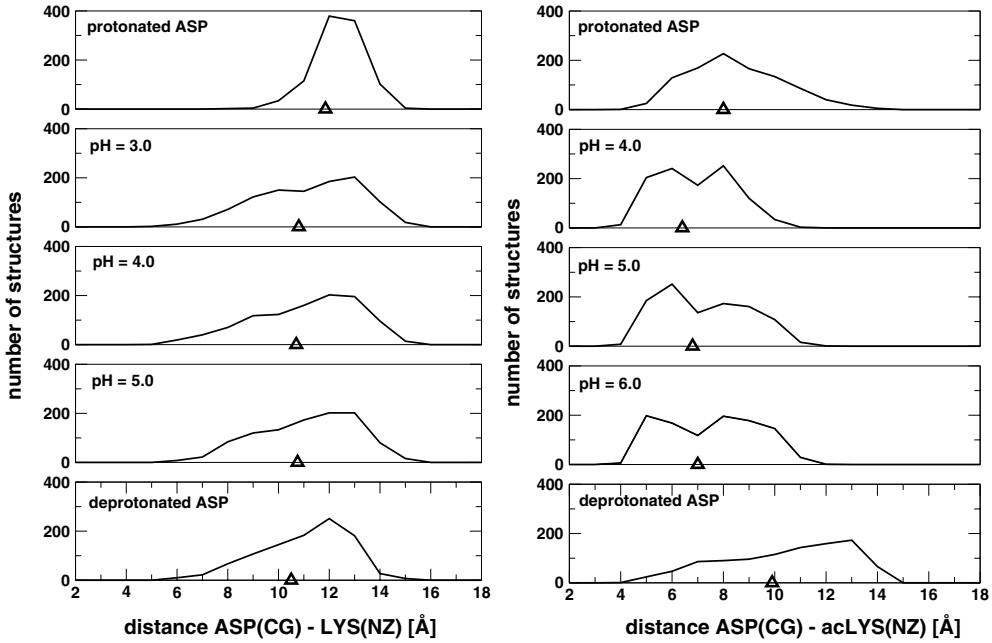

Figure 6. Distribution of distances between atom CG of Asp 2 and atom NZ of Lys 4, for normal (left-hand side) and acetylated (right-hand side) Lys 4 residue simulations, obtained from constant-pH MD simulations, for selected pH values, and for simulations with fixed protonation state of Asp 2 residue, as marked on the figure. The black triangles indicate average values of the presented distributions.

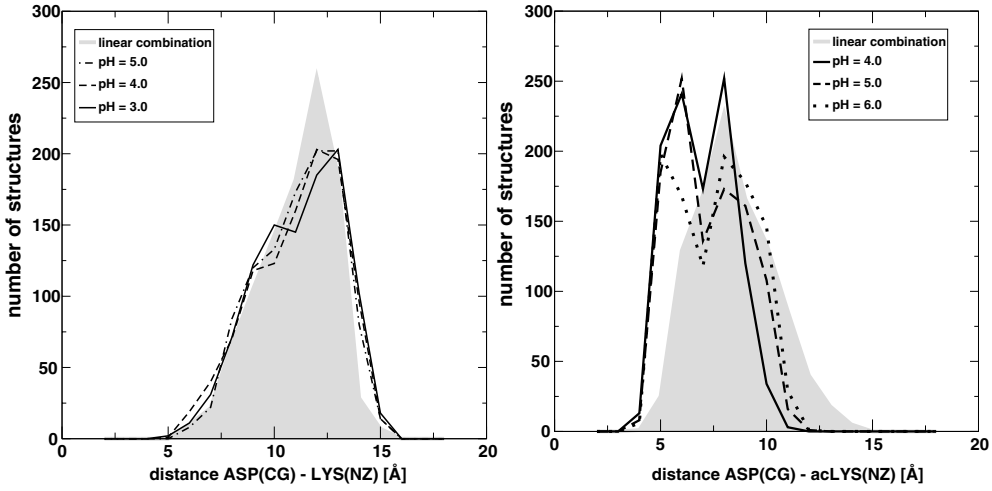

Figure 7. Distribution of distances between atom CG of Asp 2 and atom NZ of Lys 4, for normal (left-hand side) and acetylated (right-hand side) Lys 4 residue obtained from simulations at several pH values compared with the closest, regarding the sum of squared deviations, linear combination of distributions obtained from simulations with a fixed protonation state of Asp 2 residue (grey contour filled as a polygon, the same for all three values of pH used in constant-pH MD, for each of the Lys 4 type).

obtained by constant-pH simulations cannot be reproduced by taking linear combinations of distributions obtained from simulations with the permanently charged or permanently neutral side chain of Asp 2. This is further evidence that justifies our conclusions.

4. Conclusions

The results obtained in this study, although our simulations were devoted to a relatively small polypeptide chain, indicate that molecular dynamics simulations of proteins with fixed protonation states of their titratable residues, even if they are done for a representative set of protonation patterns of the molecule, may lead to physically meaningless results. This is expected when the rate constants characterizing the process of proton exchange by titratable residues with the environment are comparable to the rate constants characterizing conformational dynamics. This condition can be easily satisfied by some of titratable residues in real proteins, at least at certain solution conditions. This supports the significance of efforts to develop physically realistic molecular dynamics algorithms for simulations including protonation degrees of freedom.

Acknowledgments

This work was supported by European Union (Contract No. QLRI-CT-2002-90383), and by Warsaw University (BST-833/BF).

References

[1] Stryer L 1995 *Biochemistry* 4th edn (New York: Freeman)
[2] Pace C N, Laurents D V and Thomson J A 1990 *Biochemistry* **29** 2564
[3] Pace C N, Laurents D V and Erickson R E 1992 *Biochemistry* **31** 2728
[4] Tollinger M, Crowhurst K A, Kay L E and Forman-Kay J D 2003 *Proc. Natl Acad. Sci. USA* **100** 4545
[5] Huang Q, Opitz R, Knapp E-W and Herrmann A 2002 *Biophys. J.* **82** 1050
[6] Ripoll D R, Vorobjev Y N, Liwo A, Vila J A and Scheraga H A 1996 *J. Mol. Biol.* **264** 770
[7] Muth G W, Chem L, Kosek A B and Strobel S A 2001 *RNA* **7** 1403
[8] Saksmerprome V and Burke D H 2004 *J. Mol. Biol.* **341** 685
[9] Martinez R E and Bowler B E 2004 *J. Am. Chem. Soc.* **126** 6751
[10] Wüthrich K 1986 *NMR of Proteins and Nucleic Acids* (New York: Wiley)
[11] Bartik K, Redfield C and Dobson C M 1994 *Biophys. J.* **66** 1180
[12] Schaller W and Robertson A D 1995 *Biochemistry* **34** 4714
[13] Bashford D and Karplus M 1990 *Biochemistry* **29** 10219
[14] Bashford D, Case D A, Dalvit C, Tennant L and Wright P E 1993 *Biochemistry* **32** 8045
[15] Yang A S, Gunner M R, Sampogna R, Sharp K and Honig B 1993 *Proteins: Struct. Funct. Genet.* **15** 252
[16] Antosiewicz J, McCammon J A and Gilson M K 1994 *J. Mol. Biol.* **238** 415
[17] Antosiewicz J, Briggs J M, Elcock A E, Gilson M K and McCammon J A 1996 *J. Comput. Chem.* **17** 1633
[18] Alexov E G and Gunner M R 1997 *Biophys. J.* **74** 2075
[19] Warshel A 1981 *Biochemistry* **20** 3167
[20] Warshel A and Russell S 1984 *Q. Rev. Biophys.* **17** 283
[21] Sham Y Y, Chu Z T and Warshel A 1997 *J. Phys. Chem.* **101** 4458
[22] Warshel A and Papazyan A 1998 *Curr. Opin. Struct. Biol.* **8** 211
[23] Briggs J M and Antosiewicz J 1999 *Rev. Comput. Chem.* **13** 249
[24] Schutz C N and Warshel A 2001 *Proteins: Struct. Funct. Genet.* **44** 400
[25] Simonson T, Carlsson J and Case D A 2004 *J. Am. Chem. Soc.* **126** 4167
[26] McCammon J A and Harvey S C 1987 *Dynamics of Proteins and Nucleic Acids* (Cambridge: Cambridge University Press)
[27] Lim C, Bashford D and Karplus M 1991 *J. Phys. Chem.* **95** 5610
[28] Chipman D M 2002 *J. Phys. Chem.* A **106** 7413

[29] Li H, Robertson A D and Jensen J H 2002 *J. Phys. Chem.* B **106** 3486
[30] Li H, Robertson A D and Jensen J H 2004 *Proteins: Struct. Funct. Bioinf.* **55** 689
[31] Baptista A M, Martel P J and Petersen S B 1997 *Proteins: Struct. Funct. Genet.* **27** 523
[32] Börjesson U and Hünenberger P H 2001 *J. Chem. Phys.* **114** 9706
[33] Börjesson U and Hünenberger H 2004 *J. Phys. Chem.* B **108** 13551
[34] Lee M S, Salsburyy F R Jr and Brooks C L III 2004 *Proteins: Struct. Funct. Genet.* **56** 738
[35] Bürgi R, Kollman P A and van Gunsteren W F 2002 *Proteins: Struct. Funct. Genet.* **47** 469
[36] Baptista A M, Teixeira V H and Soares C M 2002 *J. Chem. Phys.* **117** 4184
[37] Mongan J, Case D A and McCammon J A 2004 *J. Comput. Chem.* **25** 2038
[38] Walczak A M and Antosiewicz J M 2002 *Phys. Rev.* E **60** 051911
[39] Dlugosz M, Antosiewicz J M and Robertson A D 2004 *Phys. Rev.* E **69** 021915
[40] Dlugosz M and Antosiewicz J M 2004 *Chem. Phys.* **302** 161
[41] Read R J, Fujinaga M, Sielecki A R and James M N G 1983 *Biochemistry* **22** 4420
[42] Antosiewicz J 1995 *Biophys. J.* **69** 1344
[43] Lill M A and Helms V 2001 *J. Chem. Phys.* **115** 7993
[44] Spörlein S, Carstens H, Satzger H, Renner C, Berrendt R, Moroder L, Tavan P, Zinth Z and Wachtveitl J 2002 *Proc. Natl Acad. Sci. USA* **99** 7998
[45] Hummer G, Garcia A E and Garde S 2001 *Proteins: Struct. Funct. Genet.* **42** 77
[46] Humphrey W, Dalke A and Schulten K 1996 *J. Mol. Graph.* **14** 33