Assessing the histidine tautomer fractions in proteins. Test on diisopropylfluorophosphatase, a large all-β protein, from *Loligo vulgaris*

Yury N Vorobjev [1], **Harold A Scheraga** [2], **Jorge A Vila** Corresp. [2, 3]

[1] Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Science, Novosibirsk State University, Novosibirsk, Russia

[2] Chemistry and Chemical Biology, Cornell University, Ithaca, New York, USA

[3] IMASL-CONICET, Universidad Nacional de San Luis, San Luis, Argentina

Corresponding Author: Jorge A Vila
Email address: jorgevila84@gmail.com

The importance of histidine tautomerism for an accurate drug/protein structure determination and their relevance for the interaction between biological systems is recognized. Consequently, here we used a recently introduced method to determine the pKa values of ionizable residues and fractions of ionized and tautomeric forms of histidine and acid residues in protein, as a function of pH, to analyze a histidine-rich protein, specifically, to study the accuracy of the prediction of the tautomeric fractions of the imidazole ring of each of the 6 histidine residues of diisopropylfluorophosphatase (pdb id 1E1A), a 314-residue all beta-protein, from *Loligo vulgaris*. The average tautomeric fractions of the His was compute by using an approach that includes, but is not limited to, molecular dynamics simulations coupled with calculations of the ionization states for all 94 ionizable residues of the protein 1E1A in water at pH 6.5 and 300 K. The results of the analysis on protein 1E1A indicate that: (*i*) the averaged calculated fractions of histidine tautomers, namely for the N^{d1}-H and N^{e2}-H forms, respectively, are in good agreement, within ~11%, with the prediction of an NMR-based methodology; and (*ii*) the tautomer fractions among the six His residues are very different, mainly because the tautomer preferences are determined by the protein environment. This work provides critical insight about the limitations of continuum dielectric models to reproduce NMR-based tautomeric fractions, and may spur significant progress in our effort to develop fast and accurate methods to determine the forms of the imidazole ring of His in proteins as a function of pH.

Assessing the Histidine Tautomer Fractions in Proteins. Test on Diisopropylfluorophosphatase, a large all-β protein, from *Loligo vulgaris*.

Yury N. Vorobjev,[1,2] Harold A. Scheraga[2] and Jorge A. Vila[2,3]

[1] Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Science, Lavrentiev Avenue 8, Novosibirsk 630090, Russia; Novosibirsk State University, Novosibirsk 630090, Russia;

[2] Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301;

[3] IMASL-CONICET, Universidad Nacional de San Luis, Ejército de Los Andes 950, 5700 San Luis, Argentina.

Corresponding authors:

Yury N. Vorobjev,[1,2]

Jorge A. Vila[2,3]

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2816v1 | CC BY 4.0 Open Access | rec: 18 Feb 2017, publ: 18 Feb 2017

Abstract

The importance of histidine tautomerism for an accurate drug/protein structure determination and their relevance for the interaction between biological systems is recognized. Consequently, here we used a recently introduced method to determine the pKa values of ionizable residues and fractions of ionized and tautomeric forms of histidine and acid residues in protein, as a function of pH, to analyze a histidine-rich protein, specifically, to study the accuracy of the prediction of the tautomeric fractions of the imidazole ring of each of the 6 histidine residues of Diisopropylfluorophosphatase (pdb id 1E1A), a 314-residue all β-protein, from *Loligo vulgaris*. The average tautomeric fractions of the His was compute by using an approach that includes, but is not limited to, molecular dynamics simulations coupled with calculations of the ionization states for all 94 ionizable residues of the protein 1E1A in water at pH 6.5 and 300 K. The results of the analysis on protein 1E1A indicate that: (*i*) the averaged calculated fractions of histidine tautomers, namely for the $N^{\delta 1}$-H and $N^{\varepsilon 2}$-H forms, respectively, are in good agreement, within ~11%, with the prediction of an NMR-based methodology; and (*ii*) the tautomer fractions among the six His residues are very different, mainly because the tautomer preferences are determined by the protein environment. This work provides critical insight about the limitations of continuum dielectric models to reproduce NMR-based tautomeric fractions, and may spur significant progress in our effort to develop fast and accurate methods to determine the forms of the imidazole ring of His in proteins as a function of pH.

Introduction

Histidine (His) is a unique residue for a number of reasons: *i*) about 50% of all enzymes use His in their active sites Shimba et al., (2003); *ii*) His, with a pK^{o} of 6.6 at T = 300 K (Demchuk & Wade, 1996), titrates around neutral pH, enabling the deprotonated nitrogen of its imidazole ring to serve as an effective ligand for metal binding; *iii*) the imidazole ring of His (see Figure 1) includes two neutral, chemically distinct forms, and a protonated form, referred to as $N^{\delta 1}$-H and $N^{\epsilon 2}$-H tautomers, and His^{+}, respectively; one form is favored over the others by, mainly, the local protein environment and the pH. In particular, it has been suggested that tautomerization of the neutral state and variations of χ^{1} of His are crucial parts of the proton-transfer process (Hass et al., 2008).

An accurate method for calculating the fractions of the tautomeric forms of His as a function of pH, provided that the observed $^{13}C^{\gamma}$ and $^{13}C^{\delta 2}$ chemical shifts and the protein structure, or the fraction of the H^{+} form, are known, has been developed recently by Vila et al. (2011). However, in NMR spectroscopy, it is not a common practice to determine the $^{13}C^{\gamma}$ chemical shift of the imidazole ring of histidine. In fact, only 106 $^{13}C^{\gamma}$, versus 4,703 $^{13}C^{\delta 2}$, chemical shifts of the imidazole ring of histidine have been deposited in the Biological Magnetic Resonance data Bank (BMRB) (Ulrich et al, 2005). Overall, the sparse numbers of $^{13}C^{\gamma}$ chemical shifts prevents the application of the quantum-chemical-based method (Vila et al., 2011). For this reason, a new methodology with which to determine the fractions of the tautomeric forms of the imidazole ring of histidine in proteins accurately as a function of pH, provided that *only* the 3D structure of the protein is known, is necessary.

As a contribution to the problem of accurate computation of the ionization phenomenon and the fractions of ionized and neutral tautomer forms of histidine and acid residues in proteins, from *only* their 3D structures, was recently presented (Vorobjev, Scheraga & Vila, 2017). This method is based on a Molecular Dynamics (MD) simulation, to account for the protein conformational dynamics, together with an effective version of a continuum electrostatic method, for an accurate prediction of the pKa's of the residues in a protein in solution as a function of pH. In particular, this new method was validated by using a large-set of 297 experimental pKa dataset from 34 X-ray and NMR-determined proteins. Here, this new method is applied to estimate, at pH 6.5 and 300 K, the tautomeric fractions of 6 histidine residues of a 314-residue all β-protein, namely 1E1A (Scharff et al., 2001), for which the observed tautomeric fractions of the histidine forms, namely the $N^{\delta 1}$-H and $N^{\varepsilon 2}$-H forms respectively, were obtained from an NMR-based method (Vila et al., 2011).

There are a number of reasons as to why the use of protein 1E1A is interesting. Among others, we can mention the following: (*i*) this protein, to the best of our knowledge, possesses the larger number of His residues (6) for which the $^{13}C^{\gamma}$ and $^{13}C^{\delta 2}$ chemical shifts have been experimentally determined by NMR spectroscopy and, hence, for which the tautomeric fractions of the imidazole ring of His can be accurately predicted (Vila et al., 2011); (*ii*) the structure of 1E1A has been accurately determined by X-ray crystallography at 1.8 Å resolution; (*iii*) this protein possesses a large number (94) of ionizable residues and, hence, this work contributes to an understanding of how the *ionized* environment may affect the tautomeric distribution of the imidazole ring of His at a given fixed pH; and (*iv*) kinetic studies in combination with site-directed

mutagenesis reveal that His287, at the active site, is crucial for the diisopropylfluorophosphatase (pdb id 1E1A) hydrolysis (Scharff et al., 2001).

Materials and Methods

Calculation of the fraction of tautomeric and charged form of His

For the computation of the energies of ionization microstates and the free energy of ionization, see Vorobjev, Scheraga & Vila (2017). As already mentioned, in the latter manuscript, it is common practice (Bashford et al., 1993; Couch & Stuchebrukhov, 2011; Machuqueiro & Baptista, 2009; Wihtam et al., 2011; Anandakrishn, Aguilar & Onufriev, 2012; Nielsen, Gunner & Garcia-Moreno, 2011; Bashford & Karplus, 1990; Beroza et al., 1991; Yang et al., 1993; Yan & Honig, 1993) to use a continuum dielectric model to calculate the free energy of solvation, i.e., the free energy of solvent polarization, by solving the Poisson equation either by the finite difference method (Alexov et al., 2011; Son, Mao & Gunner, 2009; Yang et al., 1993; Yang & Honig, 1993) or by the Fast Adaptative Multi-Grid Boundary Element (FAMBE) method (Vorobjev, Almagro & Hermans, 1998; Vorobjev & Hermans, 1997, 1999; Vorobjev & Scheraga, 1997; Vorobjev, 2011, 2012) for which an accurate version of the Generalized Born Model, namely GB-MSR6C, has been developed (Vorobjev, 2011, 2012). Consequently, and for the reason already discussed (see Vorobjev, Scheraga & Vila, 2017), here we will use the GB-MSR6C for all the pKa calculations.

There is evidence (Vorobjev, Vila & Scheraga, 2008; Loffler, Schreiber & Steinhauser, 1997; Vorobjev, Scheraga & Vila, 2017) indicating that a calculation of ionization equilibria by using a large value of $D_i = 16$, for a fixed protein conformation, accounts for solvent structure mobility and reorganization due to nonstructural

responses, such as the charge redistribution. For this reason, in this manuscript a $D_i = 16$ was adopted as the internal dielectric constant (Vorobjev, Scheraga & Vila, 2017). Moreover, a consistent set of atomic charges for protein residues in neutral and ionized states was computed by fitting the electrostatic potential of the atomic charges to the reference electrostatic potential, which was calculated by a high-quality quantum mechanical method (Arnautova et al., 2009).

Molecular dynamics simulations

Protein dynamics in water was taken into account by carrying out MD simulations with implicit solvent, namely by using the Lazaridis-Karplus solvent model (Lazaridis & Karplus, 1999) with the BioPASED program (Popov & Vorobjev, 2010). For the MD simulations the following standard protocol was used: (*i*) the starting conformation was the deposited crystal structure after energy optimization; (*ii*) the system was heated slowly from 1K to 300K during 150 ps; (*iii*) a final equilibration at 300K, during ~0.5 ns, was carried out; and (*iv*) the equilibrium MD trajectories of the protein dynamics, during 10 ns, was collected as snapshots taken every 50 ps time-interval. Calculation of the average degrees of ionization and the tautomeric fractions of the imidazole ring of His, for each snapshot, was carried out as described in the Material and Methods section of Vorobjev, Scheraga & Vila (2017).

An NMR-based methodology to estimate the tautomeric fraction of His

The tautomeric fraction of the imidazole ring of His will be determined by an NMR-based methodology (Vila et al., 2011), namely by using the following equation:

$$f^{\varepsilon} = \frac{\Delta^{obs}(1 - \langle f^{H+} \rangle)}{\Delta^{\varepsilon}} \qquad (1)$$

where $\langle f^{H+} \rangle$ is the average fraction of the protonated form computed by using a continuum-electrostatic based method (Vorobjev, Scheraga & Vila, 2017); $\Delta^{obs} = |^{13}C^{\delta 2} - {}^{13}C^{\gamma}|$, where $^{13}C^{\delta 2}$ and $^{13}C^{\gamma}$ are the NMR-observed chemical shifts for each His residue in the sequence (see Δ^{obs} values in Table 1); Δ^{ε} is the single-valued first-order shielding difference computed for the $N^{\varepsilon 2}$-H tautomer, namely $\Delta^{\varepsilon} = 31.7$ ppm (Vila et al., 2011); and f^{ε} is the fraction of the $N^{\varepsilon 2}$-H tautomer. Naturally, the f^{δ} fraction, *viz.*, for the $N^{\delta 1}$-H tautomer, is obtained straightforwardly as:

$$f^{\delta} = 1 - \langle f^{H+} \rangle - f^{\varepsilon} \qquad (2)$$

It is worth noting that the tautomeric fractions determined by using this NMR-based methodology depends, *critically*, on an accurate determination of the average fraction of the protonated form of the His residue ($\langle f^{H+} \rangle$) and, hence, for consistency we will compute $\langle f^{H+} \rangle$ here by using a new methodology (Vorobjev, Scheraga & Vila, 2017) rather than use the $\langle f^{H+} \rangle$ values previously reported for protein 1E1A (Vila et al., 2011). The reason for this choice is that the new methodology, but not the previous one (Vorobjev, Vila & Scheraga, 2008), takes into account explicitly the protein dynamics which is crucial for an accurate estimation of the average fraction of the ionized His form ($\langle f^{H+} \rangle$) (Vorobjev, Scheraga & Vila, 2017).

Results and discussion

Analysis of the tautomeric distribution of the imidazole ring of His

The protein 1E1A (Scharff et al., 2001) contains 6 His residues, namely, His181, His219, His224, His248, His274 and His287, for which the fractions of the three tautomeric forms of each His can be determined accurately by using NMR data and a

quantum-chemical-based methodology (Vila et al., 2011). Consequently, in the absence of observed values for the tautomeric fractions of the imidazole ring of each His of 1E1A, the computed values obtained by using Eq. (1) will be adopted as a *reference* value with which to compare the electrostatic-based predictions, of the tautomeric forms for the imidazole ring of His, obtained with the GB-MSR6c-pK+MD methodology (Vorobjev, Scheraga & Vila, 2017). In addition, to take into account the conformational dynamics of protein 1E1A in a solvent, an MD simulation in water solution was carried out by using the BioPASED program (Popov & Vorobjev, 2010) with the AMBER force field (Cornell et. al., 1995). Calculations of pKa are carried out at pH = 6.5 by considering all ionizable residues with variable ionization degrees, as previously described (Vorobjev, Scheraga & Vila, 2017).

The computed pKa values for all HIS residues along a 10 ns MD trajectory are shown in Figure 2. As shown in this Figure, the range of pKa values, along the MD trajectory, for each of the six HIS residues of the protein 1E1A fluctuate within 0.12-0.21 pK units around the average values (see Table 1). Moreover, the variation, along the 10 ns MD trajectories, of the fractions of protonated His (f^{H+}) and the tautomeric forms ($N^{\delta 1}$-H and $N^{\varepsilon 2}$-H) for His181, His219, His224, His248, His274 and His287 residues, at pH = 6.5 and T = 300 K, are shown in Figure 3a-3d. Among all the panels of Figure 3 it is worth noting a considerable variation of the computed fraction (~0.2) for the $N^{\varepsilon 2}$-H tautomer of His219 (see green-solid line in Figure 3b). The fact, that a large variation of the $N^{\varepsilon 2}$-H tautomeric fraction accompany a variation of the fraction of the ionization degree (see black-solid line in Figure 3b), could indicates that the proton binding/release occurs preferentially on the $N^{\delta 1}$ nucleus. Over all, the large variation of

the computed fractions for His219 (see Figure 3b) could be a simple consequence of the fact that this residue is well exposed to the solvent (see Table 1) and, hence, exhibiting a large fluctuation of its surface area.

Analysis of the data listed in Table 1 indicates that there are two His residues, His274 and His287, for which the computed tautomeric fractions show the best agreement with the *reference* values, namely with an absolute difference (Δ) of ~7% and ~0%, respectively. In particular, the computed tautomeric fraction of His274 shows a preference for the $N^{\delta 1}$-H rather than the $N^{\varepsilon 2}$-H tautomer (see Table 1). Notable, the His274 side-chain in the crystal structure show a direct coordination of the $N^{\delta 1}$-H tautomer to the calcium ion (Scharff et al, 2001). In addition, the $N^{\varepsilon 2}$-H form of His287 (at the active site of protein 1E1A), at pH = 6.5, is the most populated tautomer, ~80%, in line with the predicted value for the *reference*. A visual inspection of the 3D structure of protein 1E1A reveals that the H^+ form of His287 could be stabilized by the existence of an almost perfect hydrogen bond between the $N^{\varepsilon 2}$-H group of Hi287 and the backbone O atom of Gly19. Therefore, the neutral state of His287 keeps the $N^{\varepsilon 2}$ group protonated, i.e., in the $N^{\varepsilon 2}$-H form, while $N^{\delta 1}$ becomes mostly non-protonated.

Overall, we found among all the computed tautomeric fractions a clear preference for the $N^{\varepsilon 2}$-H rather than $N^{\delta 1}$-H tautomer (see Table 1), except for His274 (for which the coordination of the $N^{\delta 1}$-H tautomer to the calcium ion, as seen in the crystal structure, is remarkable because it usually coordinated to oxygen atoms [Scharff et al, 2001]). On the other hand, the preferences of the *reference* values are not so well defined. Indeed, for His219 both tautomeric forms are almost equally preferred; however, for His181 and His248 the $N^{\delta 1}$-H rather than the $N^{\varepsilon 2}$-H tautomer is the

preferred one among the neutral forms, although the reverse preferences are shown for His224, His274 and His287.

Conclusions

Calculation of the fractions of ionized H^+ and the tautomeric $N^{\delta1}$-H and $N^{\varepsilon2}$-H forms of the imidazole ring of His in proteins is a challenging task because it needs an accurate estimation of relative free energies of the tautomer forms, i.e., within a fraction of 1.0 kcal. A state of the art method for estimation of the free energy of a protein in a given ionization state was used here to calculate the fractions of the ionized H^+ and tautomeric forms of His in protein 1E1A. In particular, the fractions of the tautomeric forms of the imidazole ring of His in the sequence are accurately predicted within an average error of ~0.11. While this accuracy of the method is encouraging, it is worth noting that a recently-introduced quantum-mechanical-based method (Kromman et al., 2016) to estimate the pKa of small molecules shows very promising accuracy although it is, for the moment, limited to small compounds.

Overall, the only necessary requirement for an accurate prediction of the populations of the His forms in proteins, by using the coupled molecular dynamics and electrostatic-based methods used here, is the 3D structure. This advantage of this method could be a very valuable tool when the determination of the $^{13}C^{\gamma}$ and $^{13}C^{\delta2}$ chemical shifts by NMR spectroscopy is not possible or does not exist.

Table 1. pKa values and fractions of H^+, $N^{\delta 1}$-H and $N^{\varepsilon 2}$-H forms for the imidazole ring of His, computed at pH = 6.5 and T = 300 K, for protein 1E1A.

His number[a]	pKa computed values[b]	Fractions of the imidazole ring of His forms at pH = 6.5			
	<pKa>$_{MD}$	*Reference* values	Computed values		
			<fraction>[e]	RMSD[f]	Abs(Δ)[g]
His181 (11.2 Å^2)	6.97 (0.14)	0.90[c] 0.74[d] 0.25 0.01	0.74 0.11 0.15	0.05 0.03 0.03	n/a 0.14 0.14 **0.14**[h]
His219 (46.1 Å^2)	6.58 (0.12)	14.3 0.54 0.25 0.21	0.54 0.08 0.38	0.09 0.02 0.08	n/a 0.17 0.17 **0.17**
His224 **(5.1 Å^2)**	6.16 (0.21)	18.9 0.35 0.26 0.39	0.35 0.14 0.51	0.08 0.04 0.08	n/a 0.22 0.12 **0.17**
His248 (20.1 Å^2)	7.11 (0.13)	3.70 0.77 0.20 0.03	0.77 0.09 0.14	0.05 0.02 0.04	n/a 0.11 0.11 **0.11**
His274 (17.2 Å^2)	7.21 (0.16)	22.4 0.82 0.05 0.13	0.82 0.12 0.07	0.05 0.04 0.02	n/a 0.07 0.06 **0.07**
His287 (16.7 Å^2)	7.10 (0.14)	25.9 0.80 0.04 0.16	0.80 0.04 0.16	0.05 0.01 0.04	n/a 0.00 0.00 **0.00**
Averaged Error	–	–			**0.11**[i]

[a] in parentheses, for each His residue, the average smooth molecular surface area, computed by using the SIMS method (Vorobjev, Almagro & Hermans, 1998), is listed.

[b] pKa values computed by using *All ioniz* method (Vorobjev, Scheraga & Vila, 2017), i.e., all His, acid and base residues have variable ionizations. In parentheses, we listed the estimated statistical errors of the MD/MC pKa calculations obtained by using multiple MC runs with variable chain length, i.e., 2.0-4.0 x10^6, for each pH value, with pH step equal to 0.25 (see Vorobjev, Scheraga & Vila, 2017); the estimated statistical error was always computed as the pKa RMSD along the MD 10.5 ns trajectory, and the maximal RMSD obtained was 0.21 pH units for His224.

[c] for every His in the first column we listed, underlined, the Δ^{obs} value ($|^{13}C^{\delta 2} - {}^{13}C^{\gamma}|$) [see Table S1 of Vila et al., 2011].

[d] for each of the three rows of every His in the first column, we listed the corresponding *reference* values for the His forms, namely for the H^+, $N^{\delta 1}$-H and $N^{\varepsilon 2}$-H forms, respectively; with the average fraction of the H^+ computed as described below in footnote (*e*) and the fractions of the tautomeric form for the imidazole ring of His were computed by using Eq.(1) and Eq.(2).

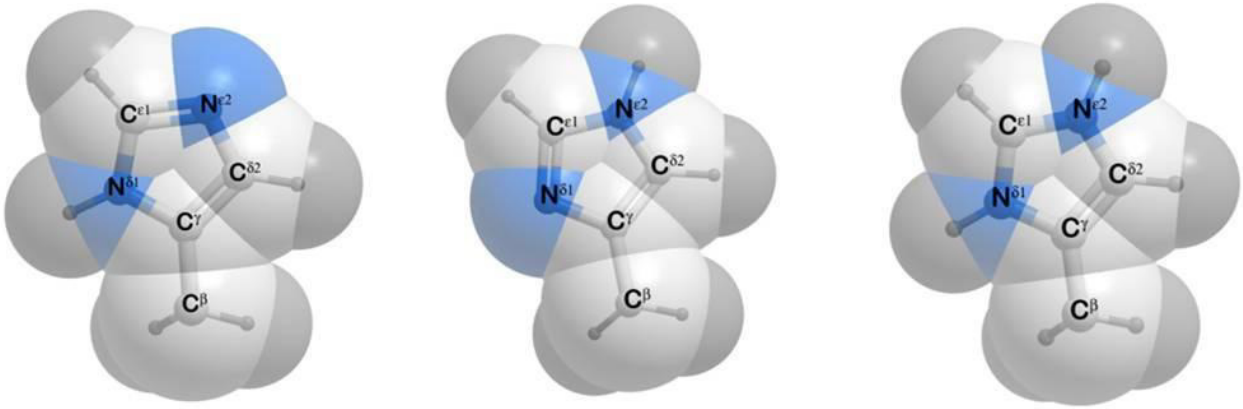
[e] in the first three rows the computed averaged fractions of each His form, namely for the H^+, $N^{\delta 1}$-H and $N^{\varepsilon 2}$-H form, respectively, by using the Vorobjev, Scheraga & Vila (2017) method, are listed; the average fractions were calculated over 10 ns MD trajectory.

[f] RMSD of fluctuations along the MD trajectory for fractions of ionized and tautomer forms of HIS residues.

[g] in the first three rows the absolute error (Δ) for each His form computed as the difference between the *reference* and the computed tautomeric fraction, respectively, are listed, except for the ionized form H^+ because it was determined only once, as shown in footnote *e*; this fact is highlighted as n/a (not apply).

[h] average absolute error among *all* Δ's of each His, i.e., computed as: $1/3\ \Sigma_{i=1,3}\ \Delta_i$, are listed in the last row.

[i] averaged error value, computed as: $1/5\ \Sigma_{i=1,6}\ \langle\Delta\rangle_i$, with $\langle\Delta\rangle$ defined in item (g) above, are listed in the last row.



(a) (b) (c)

Figure 1. (a) Structure of the $N^{\delta 1}$-H form of the His residue; (b) same as (a) for the $N^{\varepsilon 2}$-H form; and (c) same as (a) for the ionized H^+ form.

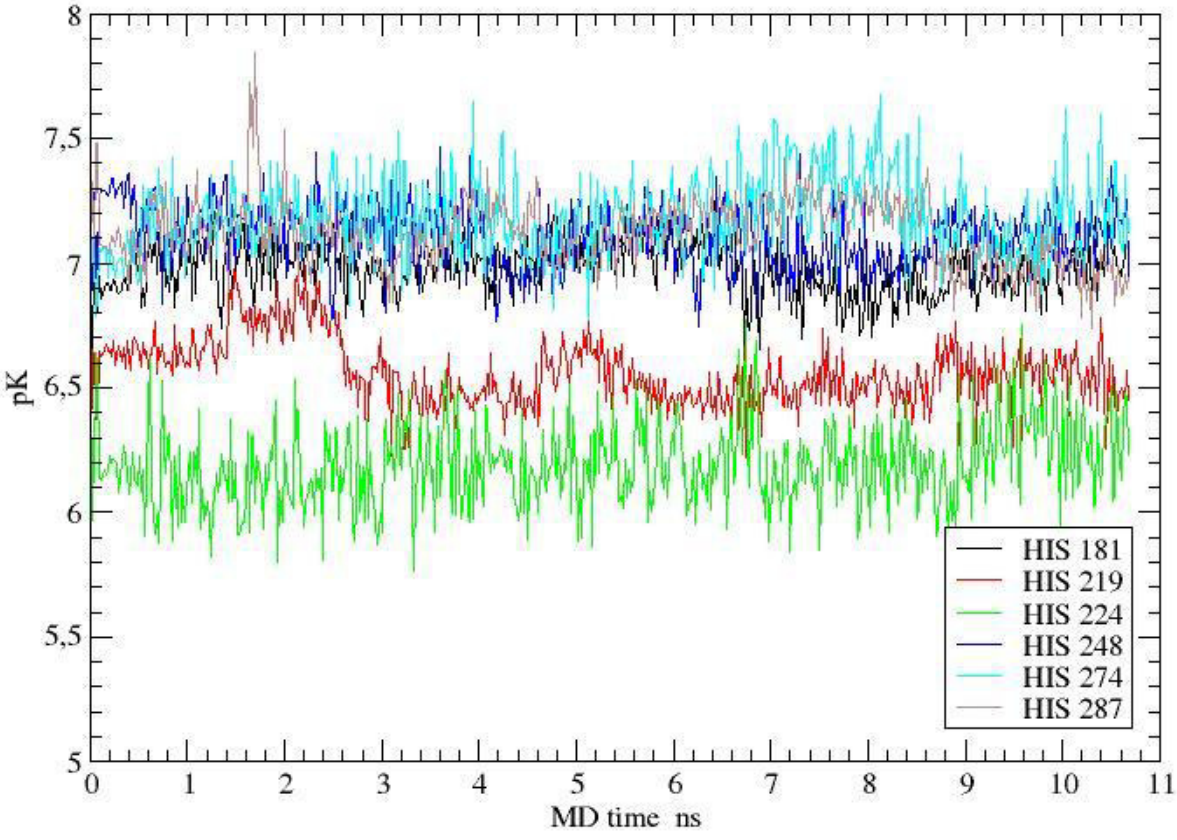
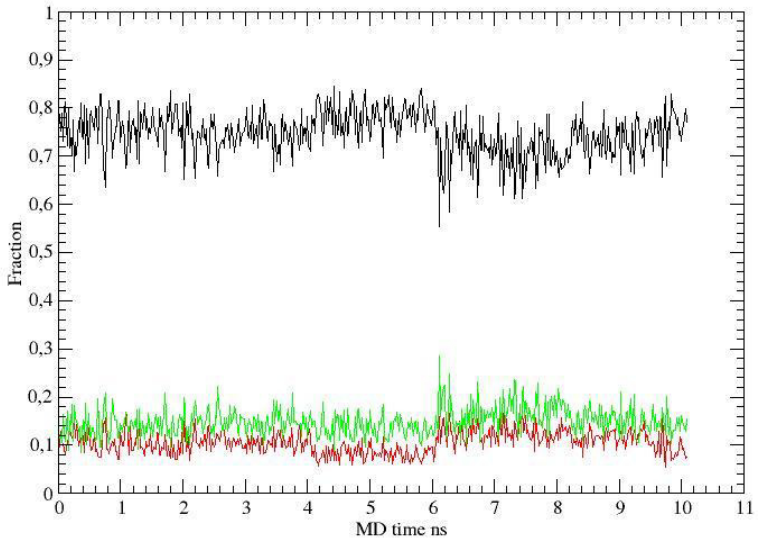
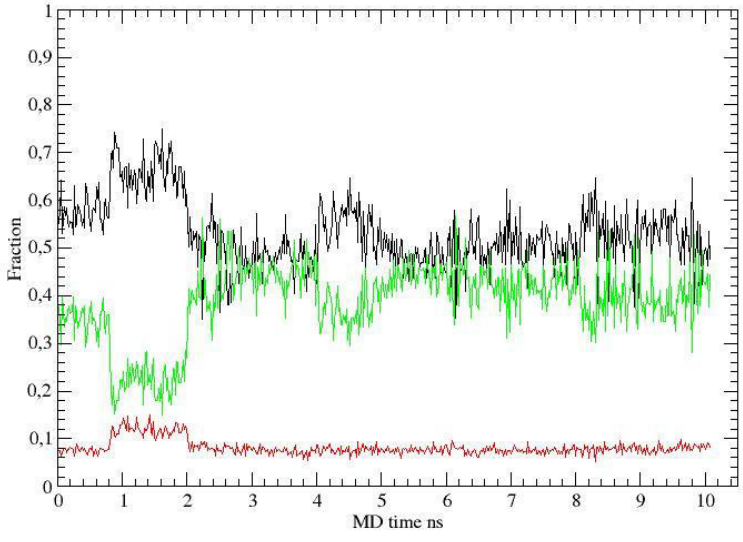


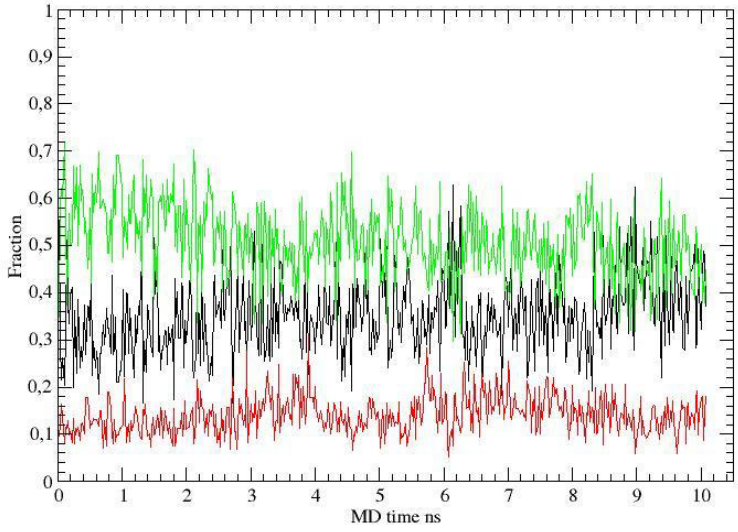
Figure 2 pKa variations of *all* His residues, of protein 1E1A, along the MD trajectory.



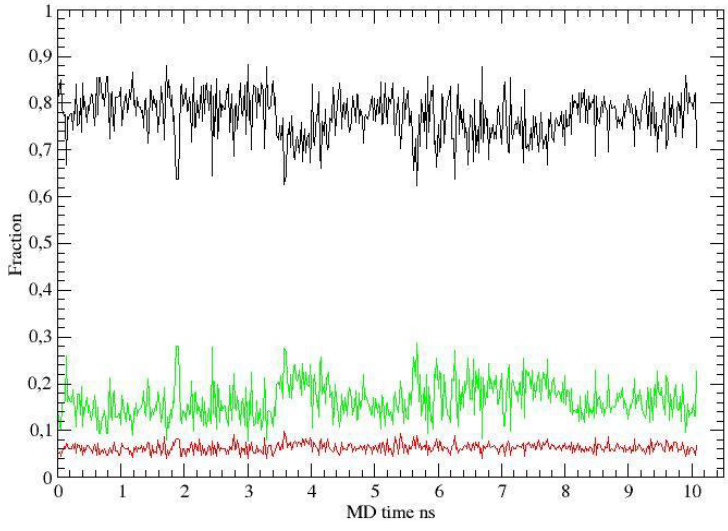
(a)



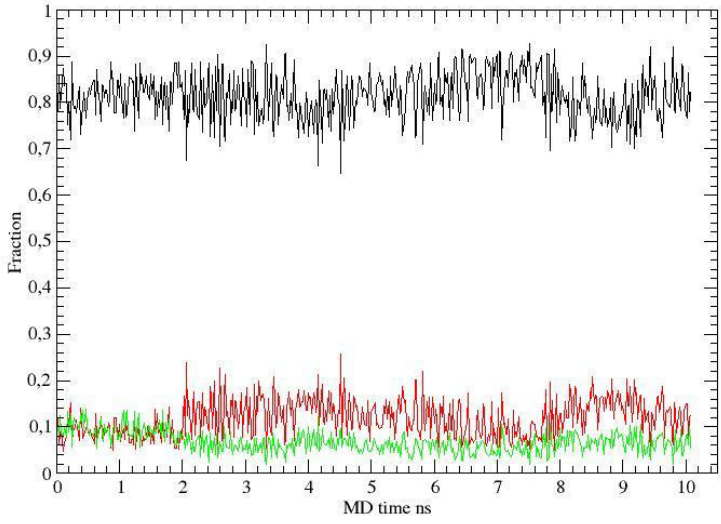
(b)



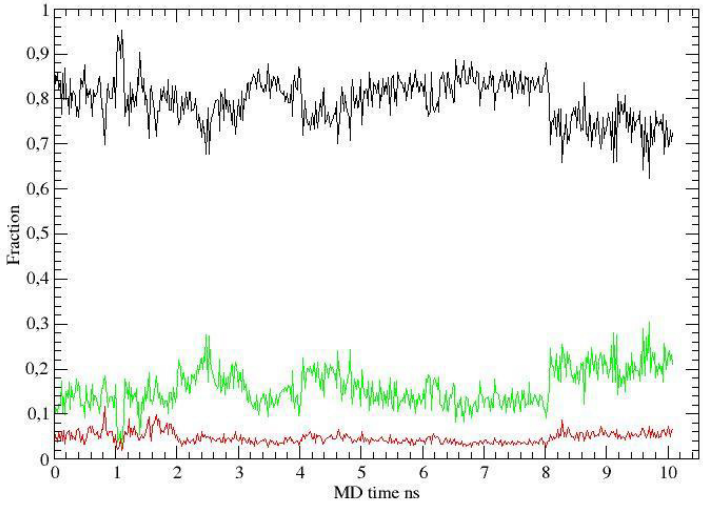
(c)



(d)



(e)



(f)

Figure 3. (a) Fractions of the HIS 181 forms along the MD trajectory: *black* line for the H^+ form; *red* and *green* line for the $N^{\delta 1}$-H and $N^{\varepsilon 2}$-H tautomer, respectively; (**b**) same as (**a**) for HIS219; (**c**) same as (**a**) for HIS224; (**d**) same as (**a**) for HIS248; (**e**) same as (**a**) for HIS274; and (**f**) same as (**a**) for HIS287.

References

Alexov E Mehler EL, Baker N, Baptista AM, Huang Y, Mille F, Nielsen JE, Farrell D, Carstensen T, Olson MHM, Shen JK, Warwicker J, Williams S, Word JM. 2011. Progress in the prediction of pKa values in proteins. *Proteins*. 79:3260-3275.

A**nandakrishn R, Aguilar B, Onufriev AV. 2012.** H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*. 40: W537–W541.

Arnautova YA, Vorobjev YN, Vila JA, Scheraga HA. 2009. Identifying native-like protein structures with scoring functions based on all-atom ECEPP force fields, implicit solvent models and structure relaxation. *Proteins*. 77:38-51.

Bashford D, Case DA, Dalvit C, Tennant L, Wright PE. 1993. Electrostatic calculations of side-chain pKa values in myoglobin and comparison with NMR data for histidines. *Biochemistry*. 32:8045-8056.

Bashford D, Karplus M. 1990. pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry*. 29:10219-10225.

Beroza P, Fredkin D R, Okamura M Y, Feher G. 1991. Protonation of interacting residues in a protein by a Monte Carlo method: Application to lysozyme and the photosynthetic reaction center of Rhodobacter sphaeroides. *Proc Natl Acad Sci USA*. 88:5804-5808.

Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W., Kollman P.A. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc*. 117:5179–5197.

Couch V, Stuchebrukhov A. 2011. HIS in continuum electrostatics protonation state calculations. Proteins. 79:3410-3419.

Demchuk E Wade RC. 1996. Improving the Continuum Dielectric Approach to Calculating pKas of Ionizable Groups in Proteins *J Phys Chem*. 100:17373–17387.

Hass MAS, Hansen DF, Christensen HEM, Led JJ, Kay LE. 2008. Characterization conformational exchange of a histidine side chain: protonation, rotamerization, and tautomerization of His61 plastocyanin from Anabaena variabilis. *J Am Chem Soc*. 130:8460–8470.

Kilmbi AP, Gray JJ. 2012. Rapid calculation of protein pK_a values using Rosetta *Biophysical J.* 112:587-593.

Kromman JC, Larsen F, Moustafa H, Jensen JH. 2016. Prediction of pKa values using the PM6 semiempirical method. *PeerJ* 4:e2335, DOI 10.7717/peerj.2335

Lazaridis T, Karplus M. 1999. Effective energy functions for protein in solvent. *Proteins*. **35**:133-154.

Löffler G, Schreiber H, Steinhauser O. 1997. Calculation of the dielectric properties of aprotein and its solvent: theory and a Case study. *J Mol Biol*. 270:520-534.

Machuqueiro M, Baptista A M. 2009. Molecular Dynamics at Constant pH and Reduction Potential: Application to Cytochrome c3. *J Am Chem Soc*. 131:12586-12594.

Nielsen JE, Gunner MR, Garcia-Moreno BE. 2011. The pKa cooperative: A collaborative effort to advance structure-based calculation of pKa values and electrostatic effects in proteins. *Proteins*. 79:3249-3259.

Nielsen JE, Gunner, MR, Garcia-Moreno BE 2011. The pKa Cooperative: A collaborative effort to advance structure-based calculation of pKa values and electrostatic effects in proteins. *Proteins*. 79:3249-3259.

Popov AV, Vorobjev YN. 2010. GUI-BioPASED: A Program for Molecular Dynamics Simulations of Biopolymers with a Graphical User Interface. *Molecular Biology*. 44:648-654.

Scharff EI, Koepke J, Fritzsch G, Lücke C, Rüterjans H. 2001. Crystal structure of diisopropylfluorophosphatase from Loligo vulgaris. *Structure*. 9:493-502.

Shimba N, Serber Z, Ledwidge R, Miller SM, Craik CS, Dötsch V. 2003. Quantitative identification of the protonation state of histidine in vitro and in vivo. *Biochemistry*. 42:9227–9234.

Song W, Mao J Gunner MR. 2009. MCCE2: Improving Protein pK Calculations with Extensive Side Chain Rotamer Sampling. *J Comput Chem*. 30:2231-2247.

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao H, Markley JL. 2008. BioMagResBank *Nucleic Acids Res*. 36: D402-D408.

Vila JA, Arnautova YA, Vorobjev YN, Scheraga HA. 2011. Assessing the fractions of tautomeric forms of the imidazole ring of histidine in proteins as a function of pH. *Proc Natl Acad Sci USA*.108:5602-5607.

Vorobjev YN, Almagro JC, Hermans J. 1998. Discrimination between native and intentionally misfolded conformation of proteins: ES/IS new method for calculating conformational free energy that uses both dynamic s simulations with an explicit solvent and implicit solvent continuum model. *Proteins*. 32:399-413.

Vorobjev YN, Hermans J. 1997. SIMS, computation of a smooth invariant molecular surface. *Biophys J*. 73:722-732.

Vorobjev YN, Hermans J. 1999. ES/IS: Estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. *Biopys Chem*. 78:195-205.

Vorobjev YN, Scheraga HA. 1997. A Fast Adaptive Multigrid Boundary Element Method for Macromolecula Electrostatic Computations in a Solvent. *J Comput Chem*. 18:569-583.

Vorobjev YN, Vila JA, Scheraga HA. 2008. FAMBE-pH: A fast and accurate method to compute the total solvation free energies of proteins. *J Phys Chem B*. 112:11122-11136.

Vorobjev YN. 2011. Advances in Implicit Models of Water Solvent to Compute Conformational Free Energy and Molecular Dynamics of Proteins a Constant pH. *Adv Prot Chem Struct Biology*. 85:282-322.

Vorobjev YN. 2012. Potential of Mean Force of Water–Proton Bath and Molecular Dynamic Simulation of Proteins at Constant pH. *J Comput Chem*. 33:832-842.

Vorojev YN, Scheraga HA, Vila JA. 2017. Coupled molecular dynamic and continuum electrostatic to compute the ionization pKa's of proteins as a function of pH. Test on a large set of proteins. *J Biomol Struct Dyn*, *in press* (http://dx.doi.org/10.1080/07391102.2017.1288169).

Wihtam S, Talley K, Wang L, Zhang Z, Sarkar S, Gao D, Yang W, Alexov E. 2011. Developing of hybrid approaches to predict pKa values of ionizable groups. *Proteins*. 79:3389-3399.

Yang A S, Gunner M R, Sampogna R, Sharp K, Honig B. 1993. On the calculation of pKa's in proteins. *Proteins*. 15:252-265.

Yang SA, Honig B. 1993. On the pH-dependence protein stability. *J Mol Biol*. 231:459-474.