Is the prediction of pKa values by constant-pH molecular dynamics being hindered by inherited problems?

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · DECEMBER 2011

Impact Factor: 2.63 · DOI: 10.1002/prot.23115 · Source: PubMed

2 AUTHORS:

Miguel Machuqueiro
University of Lisbon
36 PUBLICATIONS **679** CITATIONS
SEE PROFILE

Antonio Baptista
New University of Lisbon
90 PUBLICATIONS **1,977** CITATIONS
SEE PROFILE

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Is the prediction of pK_a values by constant-pH molecular dynamics being hindered by inherited problems?

Miguel Machuqueiro[1] and António M. Baptista[2]*

[1] Centro de Química e Bioquímica, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, C8, 1749-016 Lisboa, Portugal

[2] Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. da República, EAN, 2780-157 Oeiras, Portugal

ABSTRACT

In this study, we investigate two factors that can hinder the performance of constant-pH molecular dynamics methods in predicting protein pK_a values, using hen egg white lysozyme as a test system. The first factor is related to the molecular definition and pK_a value of model compounds in the Poisson-Boltzmann framework. We address this by defining the model compound as a molecular fragment with an associated pK_a value that is calibrated against experimental data, which results in a decrease of 0.12 units in pK_a errors. The second addressed factor is the possibility that detrimental structural distortions are being introduced in the simulations by the underlying molecular mechanics force field. This issue is investigated by analyzing how the gradual structural rearrangements affect the predicted pK_a values. The two GROMOS force fields studied here (43A1 and 53A6) yield good pK_a predictions, although a time-dependent performance is observed: 43A1 performs better after a few nanoseconds of structural reorganization (pK_a errors of ∼0.45), while 53A6 gives the best prediction right at the first nanosecond (pK_a errors of 0.42). These results suggest that the good performance of constant-pH molecular dynamics methods could be further improved if these force field limitations were overcome.

Proteins 2011; 79:3437–3447.

Key words: pentapeptides; hen egg white lysozyme; simulation; Poisson-Boltzmann; Monte Carlo; GROMOS; 43A1; 53A6.

INTRODUCTION

The importance of pH in the structure and function of proteins is well illustrated by the fact that about 25% of its residues contain ionizable side chains.[1] The ionizations that arise from changes in pH can generate strong electrostatic interactions, which will inevitably have a direct influence on molecular structure and binding. It is well accepted that the pH effect on proteins is of complex nature. This is mainly due to the multiplicity of titrable sites, which are not only subjected to different environments but also coupled to one another in complex ways.

As described in the first article of this special issue, the development and improvement of methods to model protonation equilibrium and to predict pK_a values in proteins has been a continued effort by many research groups over the years but problems still remain. One of the major factors affecting the modeling of protonation processes is certainly the coupling between protonation and conformation, which is explicitly addressed by constant-pH molecular dynamics (MD) methods.[2–25] Still, constant-pH MD methods necessarily inherit the problems of the underlying molecular mechanics (MM)/MD simulations concerning force field accuracy and sampling efficiency. Furthermore, several of these methods rely on Poisson-Boltzmann (PB) or generalized Born methods to compute the protonation free energies, and thus may inherit also some of the parametrization problems of these methods, concerning dielectric constants, charges, radii, model compound pK_as, etc.

The stochastic titration method[4–9] developed at our group to perform constant-pH MD simulations tries to derive all parameters required for PB calculations from the underlying MM/MD force field. Thus, atomic partial charges are the MM/MD ones, radii are directly computed from the Lennard-Jones interactions with the water oxygen, and the dielectric constant of the "instantly-frozen" protein is taken as 2 to account only for electronic polarization. Nonetheless, its treatment of model compounds retains some of the theoretical vagueness present in PB methods with respect to their molecular definition and pK_a value. Thus, the first

The authors state no conflict of interest.
Grant sponsor: Fundação para a Ciência e Tecnologia, Portugal (fellowship SFRH/BPD/29358/2006 and project PTDC/BIA-PRO/104378/2008)
*Correspondence to: António M. Baptista, Instituto de Tecnologia Química e Biológica, Av. da República, EAN, 2780-157 Oeiras, Portugal. E-mail: baptista@itqb.unl.pt.
Received 10 January 2011; Revised 27 May 2011; Accepted 9 June 2011
Published online 15 July 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.23115

aim of the present study is to provide a clear procedure to derive model compound pK_a values directly from experimental data, following a rationale similar to the one adopted in the parametrization of MM/MD force fields.

As seen in this special issue, current constant-pH MD methods predict pK_a values with an accuracy similar to that obtained with other good methods, but recent evidences on some structural distortions induced by MM/MD force fields[26–29] may lead us to ask whether such problems may be actually limiting their performance. Thus, the second aim of the present study is to study the cumulative effect of conformational relaxation during the simulations, analyzing whether it consistently improves predictions or not.

Although we started this study using two of the Staphylococcal nuclease mutants from Garcia-Moreno's Lab, serious issues were found involving the parametrization of neutral Arg and Lys groups, which are currently being investigated in detail. Therefore, we decided to use hen egg white lysozyme (HEWL), which has become, over the years, a standard test system for pK_a prediction methodologies,[7,14–17,22,24,30–50] because of the availability of accurate experimental data, and because it has a number of sites with pK_a values that differ markedly from their reference values, mainly in the acidic range.

THEORY AND METHODS

Model compounds and their pK$_a$ values

The protonation state of a protein with N protonatable sites is represented here as $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, where x_i denotes the protonation state of site i. For concreteness, we consider a framework using a (unique) ionized reference state and proton tautomerism,[31] but the discussion and conclusions in this section are easily extended to other frameworks (e.g., without tautomerism). Thus, $x_i = 0$ refers to the ionized state, while the remaining values of x_i refer to the alternative neutral tautomers with different proton positions. The protein in state $\boldsymbol{x}$ is denoted as $\mathrm{P}(\boldsymbol{x})$ and its probability of occurrence follows a semigrand canonical distribution[4]

$$p(\boldsymbol{x}) = \exp\{-\ln(10)\mathrm{pH}\Delta n(\boldsymbol{x}) - \Delta G^{\mathrm{P}}(\boldsymbol{x})/kT\}\Xi^{-1}, \quad (1)$$

where $\Delta G^{\mathrm{P}}(\boldsymbol{x})$ is the Gibbs free energy change for the reaction leading from $\mathrm{P}(\mathbf{0})$ to $\mathrm{P}(\boldsymbol{x})$, $\Delta n(\boldsymbol{x})$ is the associated change in the number of protons, k is Boltzmann's constant, T is the absolute temperature, and Ξ is the pH-dependent partition function normalizing the probabilities. This relation is exact, meaning that an estimation of the reaction free energies $\Delta G^{\mathrm{P}}(\boldsymbol{x})$ allows for a complete characterization of the protonation equilibrium of the protein, either from an exact or approximate (e.g., Monte Carlo) calculation.

The use of linear PB models to compute $\Delta G^{\mathrm{P}}(\boldsymbol{x})$ values is based on the use of model compounds. A model compound is a single-site compound containing the same chemical group as a protonatable site in the protein, and whose pK_a value is presumably known. To each protein site corresponds a model compound, making possible to write the following thermodynamic cycle[33,51]:

$$\begin{array}{ccc} \sum_i \mathrm{M}_i(0) + \Delta n(\boldsymbol{x})\mathrm{H}^+ & \xrightarrow{\Delta G^{\mathrm{M}}(\mathbf{x})} & \sum_i \mathrm{M}_i(x_i) \\ \downarrow & & \downarrow \\ \mathbf{P}(\mathbf{0}) + \Delta n(\boldsymbol{x})\mathrm{H}^+ & \xrightarrow{\Delta G^{\mathrm{P}}(\boldsymbol{x})} & \mathrm{P}(\boldsymbol{x}) \end{array} \quad (2)$$

where $\mathrm{M}_i(x_i)$ is the model compound of site i in state x_i, and $\Delta G^{\mathrm{M}}(\boldsymbol{x})$ is the counterpart of $\Delta G^{\mathrm{P}}(\boldsymbol{x})$ for the N model compounds in solution; a negative $\Delta n(\boldsymbol{x})$ corresponds to deprotonation. The fundamental assumption behind the use of model compounds is that the quantum contribution for the (de)protonation of a site in the protein is the same as in its corresponding model compound, so that only classical contributions (e.g., from a PB model) need to be considered when addressing this thermodynamic cycle. Thus, we can write

$$\Delta G^{\mathrm{P}}(\boldsymbol{x}) = \Delta G^{\mathrm{M}}(\boldsymbol{x}) + U^{\mathrm{P}}(\boldsymbol{x}) - U^{\mathrm{P}}(\mathbf{0}) - \sum_i U_{ii}^{\mathrm{M}}(x_i) + \sum_i U_{ii}^{\mathrm{M}}(0), \quad (3)$$

where $U^{\mathrm{P}}(\boldsymbol{x})$ and $\mathrm{U}_{ii}^{\mathrm{M}}(x_i)$ are the PB energies of, respectively, $\mathrm{P}(\boldsymbol{x})$ and $\mathrm{M}_i(x_i)$. The N model compounds in solution contribute independently to $\Delta G^{\mathrm{M}}(\boldsymbol{x})$, which is thus given by

$$\Delta G^{\mathrm{M}}(\boldsymbol{x}) = \ln(10)kT\sum_i{}' \gamma_i \mathrm{p}K_i^{\mathrm{M}}(x_i), \quad (4)$$

where the prime indicates that the sum \does not include i values for which $x_i = 0$, $\gamma_i = \pm 1$ is the charge of the ionized form of site i, and $\mathrm{p}K_i^{\mathrm{M}}(x_i)$ is the pK_a value of the dissociation reaction involving the neutral form $\mathrm{M}_i(x_i)$,[31,32] related to its global pK_a, $\mathrm{p}K_i^{\mathrm{M}}$, as

$$\mathrm{p}K_i^{\mathrm{M}}(x_i) = \mathrm{p}K_i^{\mathrm{M}} - \gamma_i \log f_i(x_i), \quad (5)$$

where $f_i(x_i)$ is the fraction of tautomer x_i among all neutral tautomers of M_i. Furthermore, if we conceptually split the protein into $N+1$ nonoverlapping fragments corresponding to the N protonatable sites plus the remaining nonprotonatable background (b), the linearity of the PB equation implies that the superposition principle[52] holds for these fragments, giving

$$U^{\mathrm{P}}(\boldsymbol{x}) = U_{\mathrm{bb}}^{\mathrm{P}} + \sum_i U_{i\mathrm{b}}^{\mathrm{P}}(x_i) + \sum_i U_{ii}^{\mathrm{P}}(x_i) + \sum_i \sum_{j<i} U_{ij}^{\mathrm{P}}(x_i, x_j), \quad (6)$$

where U_{uv}^{P} denotes the PB interaction term between fragments u and v. The previous equations, together with the relation $\Delta n(\boldsymbol{x}) = -\Sigma'_i \gamma_i$, give

$$p(\boldsymbol{x}) = \exp\left\{ \ln(10) \sum\nolimits'_i [\gamma_i (\mathrm{p}K_i^{\mathrm{M}} - \mathrm{pH}) - \log f_i(x_i)] + \beta \sum\nolimits_i [U_{ib}^{\mathrm{P}}(x_i) + U_{ii}^{\mathrm{P}}(x_i) - U_{ii}^{\mathrm{M}}(x_i)] - \beta \sum\nolimits_i [U_{ib}^{\mathrm{P}}(0) + U_{ii}^{\mathrm{P}}(0) - U_{ii}^{\mathrm{M}}(0)] + \beta \sum\nolimits_i \sum\nolimits_{j \neq i} [U_{ij}^{\mathrm{P}}(x_i, x_j) - U_{ij}^{\mathrm{P}}(0,0)] \right\} \Xi^{-1} \qquad (7)$$

where $\beta = 1/kT$, and whose terms in the second and third sums can be regarded as the effect of the protein environment on the model compound i in states x_i and 0, respectively.

The above derivation could have considered a background fragment for each model compound, which would result in additional subtractive terms $U_{ib}^{M}(x_i)$ and $U_{ib}^{M}(0)$ in the second and third sums, respectively. Instead, we intentionally considered each model compound to be identical to the corresponding fragment in the protein, as it is usual practice and implemented in most software tools. This procedure reduces the computational needs but the model compound becomes an unphysical "molecule" whose pK_a value cannot be measured. Moreover, most sets of pK_a values of model compounds currently in use seem to derive from the Nozaki and Tanford set,[53] which is a set of "typical" values inferred from various chemical species, rather than a set of experimental values of real compounds. This imprecise treatment of model compounds is further complicated by the fact that their properties, $\mathrm{p}K_i^{\mathrm{M}}$ and $f_i(x_i)$, must be modified if they are considered fixed in a conformation that is differently favored by the different protonation states. More exactly, the pK_a of M_i in conformation c is[3]

$$\mathrm{p}K_i^{\mathrm{M}}(c) = \mathrm{p}K_i^{\mathrm{M}} + \gamma_i \log \frac{p_i(c|x_i = 0)}{p_i(c|\text{ any } x_i \neq 0)}, \qquad (8)$$

where the numerator and denominator are the conditional probabilities of conformation c when M_i is, respectively, charged and neutral. Similarly, the fraction of tautomer x_i when M_i is in conformation c is

$$f_i(x_i|c) = \frac{p_i(x_i|c)}{p_i(\text{any } x_i' \neq 0|c)} = f_i(x_i) \frac{p_i(c|x_i)}{p_i(c|\text{ any } x_i' \neq 0)}. \qquad (9)$$

These conformational dependences have direct implications to the constant-pH MD method used here, which assumes that the PB calculations are done at fixed protein conformation[4] and thus should use $\mathrm{p}K_i^{\mathrm{M}}(c)$ and $f_i(x_i|c)$ values. These dependences could in principle be avoided by restricting the model compound to a small-enough fragment essentially devoid of conformational flexibility,[54] but this cannot be achieved with charge sets where (de)protonation affects the charge of atoms located at the flexible region of the residue side chain.

In this study, we address the problems described in the previous paragraph by regarding the model compound simply as a conceptual device to account for the contributions to (de)protonation that cannot be captured by PB energies. The adopted rationale is essentially to view the $\mathrm{p}K_i^{\mathrm{M}}$ and $f_i(x_i)$ as conformation-independent parameters whose values must be determined from accurate experimental data. Since, as noted above, effectively rigid model fragments are generally not possible to obtain, we require instead that the selected flexible fragments have no functional groups whose positional change may significantly affect (de)protonation. Given that the usual protonatable side chains of amino acids are monofunctional, we take them as the model compounds. For the sites corresponding to the N- and C-terminal sites, we use the fragments —$COC^{\alpha}NH_3$ and —COOH, respectively. These model compound definitions ensure that a peptide or protein is split into nonoverlapping fragments, as required by Eq. (6) (note that defining the model compound as the whole residue containing the site[34] may lead to overlapping fragments and invalidate the superposition principle). The conformation-independent parameters $\mathrm{p}K_i^{\mathrm{M}}$ and $f_i(x_i)$ corresponding to these fragments are thus expected to provide a good modeling of the system over all the conformations explored by the protein during a constant-pH MD simulation. This is similar to the modular approach adopted in MM force fields, where the conformation-independence of the bonded and nonbonded parameters is obviously an approximation. The suitability of this approach can be inferred from the results obtained.

The fractions of the neutral tautomers, $f_i(x_i)$, were assigned as previously described.[31,32] In the case of carboxyl sites (Asp, Glu, and C-terminus), we consider four tautomers, corresponding to a proton bonded in either syn or anti geometry to each of the carboxyl oxygen atoms; each syn:anti pair was assigned the ratio 94.5:5.5 inferred from experimental and theoretical studies of acetic acid.[32] For His, we consider two tautomers, corresponding to have the proton in either the $N^{\epsilon 2}$ or $N^{\delta 1}$ atoms; the $N^{\epsilon 2}$:$N^{\delta 1}$ pair was assigned the ratio 70:30 measured for His with blocked termini.[55] For all the remaining sites, we considered equiprobable tautomers, corresponding to two in-plane proton positions for Tyr, three sp^3 proton positions for Cys, and three sp^3 proton vacancies for Lys and N-terminus.

To determine the $\mathrm{p}K_i^{\mathrm{M}}$ for each type of model compound, we consider cycle (2) again, but now ignoring tautomerism and regarding the top reaction as the unknown. We then select for P a simple protein-like molecule with a single site with known pK_a, perform

constant-pH MD simulations for it at different pH values, and then compute the corresponding pK_i^M value that gives the best Henderson–Hasselbalch fit to the resulting titration curve. For this purpose, we used the eight blocked pentapeptides studied by Pace and coworkers[56,57] (Ala_5-NH_2, Ac-Ala_5 and Ac-Ala_2-X-Ala_2-H_2, where Ac denotes acetyl and X = Asp, Glu, His, Cys, Tyr, and Lys), which should provide a suitable protein-like chemical environment for the fragments selected as model compounds. As seen in Eq. (7), the populations observed in the simulations would depend on pK_i^M only through the difference $pK_i^M - pH$, meaning that we can perform the simulations of each peptide using an arbitrary pK_i^M value which is corrected afterward by simply adding the same constant required to bring the pentapeptide titration curve to its experimental position. As noted above, the pK_i^M values thus obtained should already reflect the overall conformational effects on the M_i fragments.

Constant-pH MD methodology

All constant-pH MD simulations were performed using the previously described stochastic titration method.[4–8] The method relies on different sequential blocks. The first block is a PB/Monte Carlo (PB/MC) calculation where the protonation states resulting from the last MC step are assigned to the protein. The second block is the solvent relaxation dynamics, a short MM/MD (MM/MD) simulation of the system with frozen protein, which allows the solvent to adapt to the new protonation states (the duration of this block is hereafter designated as τ_{rlx}). The last block is a full MM/MD of the unconstrained system (the duration of this block is hereafter designated as τ_{prt}). A variant of the reduced titration method is used.[5,7,58]

PB and MC calculations were performed with the programs MEAD[59] and PETIT,[32,60] respectively. The atomic charges and radii used in the PB calculations were derived from the GROMOS 43A1 and 53A6 force fields,[61–63] as previously described.[33] All PB calculations consisted of finite-difference linear PB calculations performed with the program MEAD (version 2.2.0)[59] using a temperature of 300 K, a molecular surface defined with a solvent probe radius of 1.4 Å, and a Stern (ion exclusion) layer of 2.0 Å. The dielectric constants were 80 for solvent and 2 for the protein/peptide. The ionic strength used was 0.1 M. A two-step focusing procedure[64] was used, with consecutive grid spacings of 1.0 and 0.25 Å. The MC runs were performed using 10^5 MC cycles, one cycle consisting of sequential state changes over all individual sites and also all pairs of sites with at least one interaction term above 2.0 pK_a units.[60]

The MD simulations were performed with the GROMACS 3.2.1 distribution[65,66] (with some modifications).[5] Explicit water was treated with the single point charge (SPC) model.[67] Periodic boundary conditions were used with a rhombic dodecahedral unit cell. Long-range electrostatic interactions were treated using the generalized reaction field (GRF)[68] method. A twin-range cutoff was used, with short- and long-range cutoffs of 8 and 14 Å, respectively, and with neighbor lists updated every five steps. All bond lengths were constrained using the LINCS algorithm, and a time step of 2 fs was used. Two Berendsen's temperature couplings to baths at 300 K and with relaxation times of 0.1 ps[69] were used for both the solute and solvent. A Berendsen's pressure coupling[69] at 1 bar was used with a relaxation time of 0.5 ps and a compressibility of 4.5×10^{-5} bar^{-1}. The ionic strength in the simulations was set to 0.1 M using a GRF-modified version of the GROMACS 3.2.1 distribution.[5]

Simulation settings and analysis

The eight pentapeptides (Ala_5-NH_2, Ac-Ala_5, and Ac-Ala_2-X-Ala_2-NH_2, where Ac denotes acetyl and X = Asp, Glu, His, Cys, Tyr, and Lys) were built in an unstructured conformation using PyMOL.[70] The HEWL starting structure was the Brookhaven Protein Data Bank (PDB) entry 4LZT, obtained at pH 4.5.

The protein and the pentapeptides were placed at the center of different boxes and filled with ~5500 and ~1100 water molecules, respectively. The systems were minimized first with ~50 steps of steepest descent followed by ~10,000 steps using the low-memory Broyden-Fletcher-Goldfarb-Shanno algorithm. The initiations were achieved by harmonically restraining all atoms in a 50 ps MD simulation, followed by another 50 ps simulation with only the C^α atoms restrained. Runs of 30 ns were performed for lysozyme and of 50 ns for pentapeptides. The relaxation of the solvent (τ_{rlx}) was done for 0.2 ps while each full system dynamics segment (τ_{prt}) was 2.0 ps long.

The constant-pH MD titration of HEWL was simulated only in the acidic region, given that there are a high number of acidic sites with atypical pK_as. This also allowed us to increase the number of simulations and, therefore, to improve the sampling. The simulations were performed at pH values from 0.5 to 7.0 with increments of 0.5 units. For each pH value, three replicates were performed. In all simulations, the titrating sites were treated in their tautomeric form.[7] The titration was computed by averaging at each pH value the occupancy states of all titrable sites over the final equilibrated segment. The data was fit to a Hill equation to get the pK_a and Hill coefficient values for each titrable site. The fits were done with a nonlinear Levenberg-Marquardt least-squares algorithm[71] weighting the average protonations with their correlation-corrected errors.[72] The normal-asymptotic standard errors thus obtained for the pK_a parameters were used as a measure of their statistical uncertainty.

The pentapeptides were titrated at different pH values. We chose 3–5 pH values close to the titration midpoint

Table I
pK_a Values for All Pentapeptides Studied Using Both 43A1 and 53A6 Force Fields

			43A1			53A6		
Residue	Tanford pK_a	Exp. pK_a	pK_a	Shift	pK^M	pK_a	Shift	pK^M
Asp	4.0	3.94	4.37	+0.43	**3.57**	4.34	+0.40	**3.60**
Glu	4.4	4.25	4.61	+0.36	**4.04**	4.53	+0.28	**4.12**
His	6.45	6.54	5.96	−0.58	**7.03**	6.21	−0.33	**6.78**
Cys	9.5	8.55	10.05	+1.50	**8.00**	9.47	+0.92	**8.58**
Tyr	9.6	9.84	9.93	+0.09	**9.51**	9.92	+0.08	**9.52**
Lys	10.4	10.40	10.19	−0.21	**10.61**	10.30	−0.10	**10.50**
CTer	3.8	3.67	4.34	+0.67	**3.13**	4.49	+0.82	**2.98**
NTer	7.5	8.00	6.92	−1.08	**8.58**	7.52	−0.48	**7.98**

The simulations were performed using the Tanford model compound pK_as.[33] Experimental pK_as were taken from Refs. 56 and 57. The new calibrated pK^M values are marked in bold in the table.

of each group. Three replicates were performed for every pH value. The titrations were computed by averaging at each pH value the occupancy states of each titrable site over the whole simulation. The pK_a values were obtained by fitting the data to the corresponding Henderson–Hasselbalch curve.

All analyses were done using GROMACS,[65,66] Gnuplot,[73] or in-house tools.

RESULTS AND DISCUSSION

pK^M calibration using pentapeptide simulations

The determination of model compound pK_a values from experimental data, following the procedure described in the Theory and Methods section, used the potentiometric pK_a measurements done by Pace and coworkers[56,57] on eight blocked pentapeptides (Ala_5-NH_2, Ac-Ala_5, and Ac-Ala_2-X-Ala_2-NH_2, where Ac denotes acetyl and X = Asp, Glu, His, Cys, Tyr, and Lys). We performed constant-pH MD simulations on these pentapeptides with two different force fields, GROMOS 43A1 and 53A6. The initial pK^M values used in these simulations, shown as "Tanford pK_a" in Table I, were the ones from Nozaki and Tanford,[53] except for His.[31] From the titration curves produced by these simulations, we obtained for each pentapeptide the pK_a value and shift to the experimental value (Table I). The new calibrated pK^M values, shown simply as "pK^M" in Table I, were obtained by subtracting those shifts from the Tanford values. The new pK^M values thus obtained already reflect the overall conformational effects on the model compound fragments.

Interestingly, we note that the pK^M values obtained with the 53A6 force field are, in all cases but CTer, closer to the experimental values for the residue in the pentapeptide template, indicating that the effect of the peptide environment is smaller in that case. When comparing with the previous Tanford pK^M values, the sites Cys and NTer seem to be the ones with larger shifts, especially with the 43A1 force field, meaning that they would be those more affected by systematic errors when using that set of values.

HEWL titration

The constant-pH MD simulations of HEWL were performed using both the 43A1 and 53A6 GROMOS force fields. In Table II, there is a collection of pK_a values obtained for the equilibrated segment of the simulations (10–30 ns). A comparison between the 43A1 results and the ones from previous work[7] allows for a measure of the improvement only due to the redefinition of the model compound fragments and to the pK^M calibration against experiment. With the new approach we obtained an overall pK_a root mean square error (RMSE) of 0.70, corresponding to an improvement of 0.12 pK units

Table II
pK_a Values for All Sites Titrating at Acidic pH Using the Equilibrated Segment of the Simulations (last 20 ns) in Both Force Fields

		Prev. 43A1[7]		43A1		53A6	
Residue	Exp. pK_a (avg.)	pK_a	Error	pK_a	Error	pK_a	Error
Glu-7	2.85	4.04	**1.19**	3.51 ± 0.03	0.66	3.35 ± 0.05	0.50
His-15	5.36	4.22	**1.14**	4.81 ± 0.16	0.55	5.68 ± 0.15	0.32
Asp-18	2.66	4.11	**1.45**	3.28 ± 0.06	0.62	3.32 ± 0.07	0.66
Glu-35	6.20	5.50	0.70	4.86 ± 0.10	**1.34**	5.04 ± 0.11	**1.16**
Asp-48	1.60	2.56	0.96	2.87 ± 0.10	**1.27**	2.66 ± 0.10	**1.06**
Asp-52	3.68	3.96	0.28	3.62 ± 0.07	0.06	2.78 ± 0.06	0.90
Asp-66	0.90	1.50	0.60	1.35 ± 0.11	0.45	2.40 ± 0.04	**1.50**
Asp-87	2.07	2.68	0.61	2.31 ± 0.05	0.24	2.37 ± 0.05	0.30
Asp-101	4.09	3.77	0.32	3.62 ± 0.06	0.47	3.59 ± 0.05	0.50
Asp-119	3.20	2.97	0.23	2.76 ± 0.06	0.44	3.04 ± 0.08	0.16
CTer-129	2.75	3.23	0.48	3.16 ± 0.04	0.41	3.15 ± 0.06	0.40
RMSE			0.82		0.70		0.79

All experimental pK_a values were taken from Ref. 74. Results from previous work using the 43A1 force field[7] are presented for a better comparison. All errors over 1 pK unit are marked in bold. RMSE was computed relative to the average value of the experimental range and is presented in pK units. The statistical uncertainties (shown as ± values) of the pK_a values were computed as described in the Theory and Methods section.

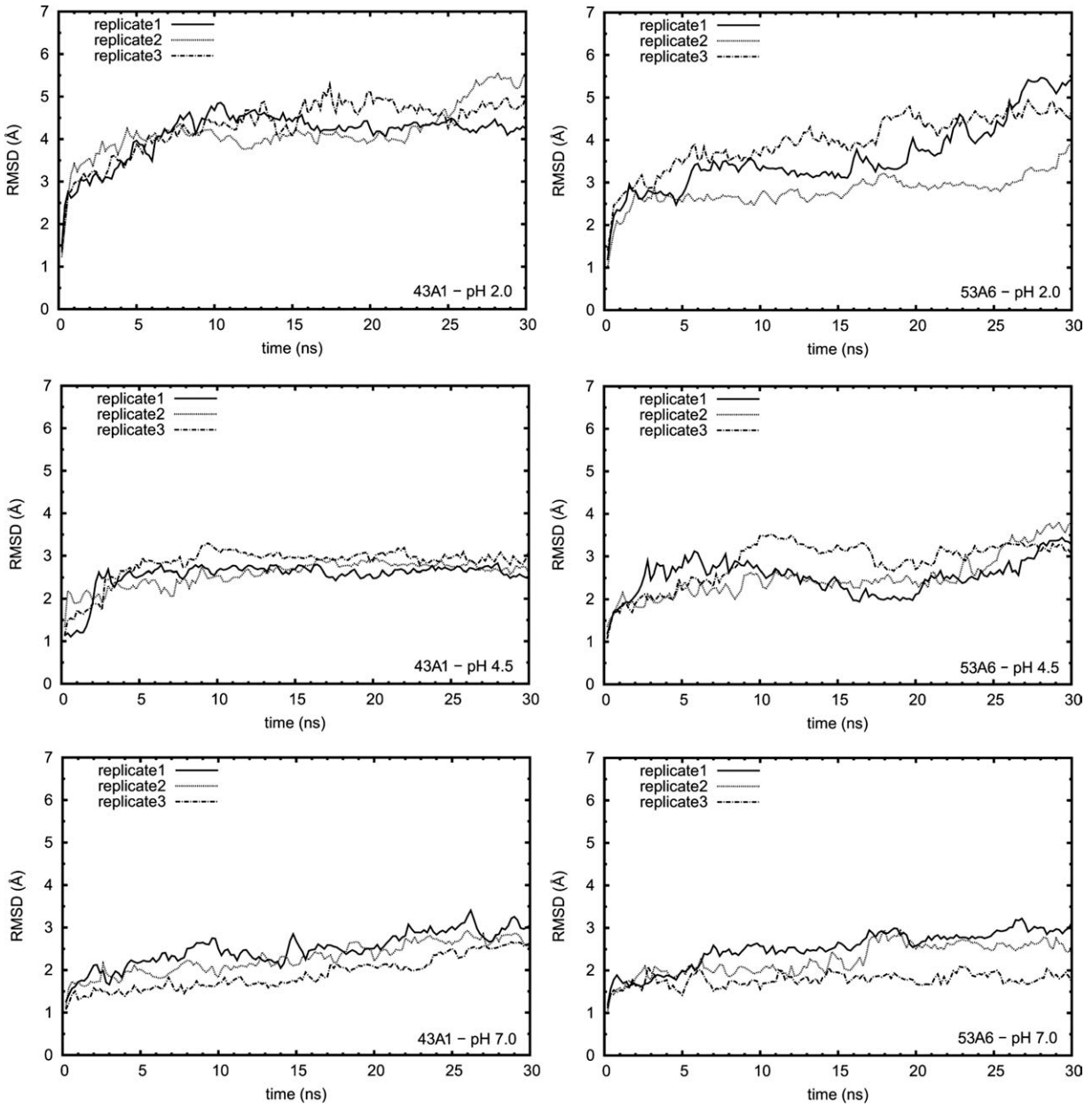

Figure 1

Temporal evolution of the RMSD of the simulations at three different pH values (2.0, 4.5, and 7.0) for both force fields used in this work. There are three replicates for each pH value. The fit and RMSD values were computed using the C^{α} atoms and taken relative to the X-ray structure.

relative to that previous study. From the individual pK_a values, we got significant improvements in some residues but the predictions for Glu 35 and Asp 48 became significantly worse. The case of Glu 35 is particularly important because this residue is the proton donor in the catalytic site of HEWL and has a highly shifted pK_a.

The pK_as obtained with the 53A6 force field were somewhat disappointing, not due to the RMSE obtained (0.79 is a good overall result), but because the method did a poor job dealing with the internalized residues. In fact, such residues seemed to lose some of their local environment during the constant-pH MD simulations, behaving more like solvated residues. Interestingly, this loss of environment effects using 53A6 was also observed in the pentapeptide simulations (see above). This suggests that the results for the pentapeptides and HEWL may be both reflecting some tendency of 53A6 to exaggerate solvent exposure.

These results raise several questions regarding the influence of these two force fields on the stability of

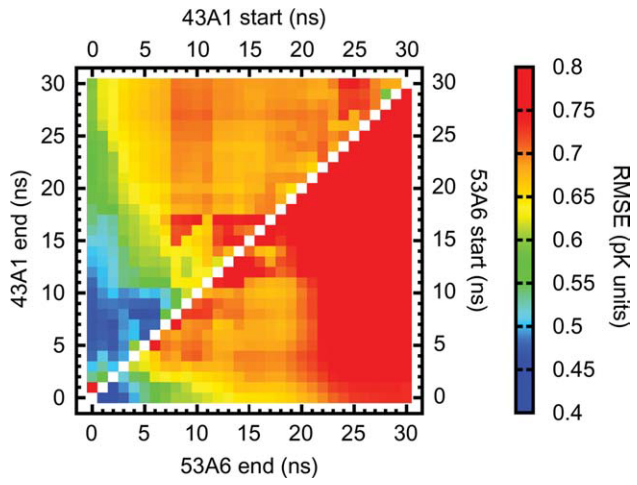

Figure 2

pK_a RMSE color map obtained with GROMOS force fields 43A1 (upper-left triangle) and 53A6 (lower-right triangle). The RMSE values were computed for all time segments multiple of 1 ns, with each colored square in the plot representing a segment of time comprised by the start and end points indicated in the axes. To present both force fields with a symmetrical orientation, the *x* and *y* axes were exchanged for 53A6.

HEWL and its impact on the prediction of correct protonation states. From the temporal evolution in the C^{α} RMSD plots (illustrated in Fig. 1) and from visual inspection of the structures, we do not see any major conformational transition that could seriously compromise the pK_a calculations. Nevertheless, the 53A6 simulations seem to exhibit more variations in the observed RMSDs (Fig. 1) and a slight decrease in helical content (data not shown). A key factor may be the significantly high C^{α} RMSD values observed after equilibration in the simulations with both force fields. Although the particularly high RMSD values at pH 2.0 certainly reflect the structural destabilization that HEWL starts to experience at very low pH,[75–77] even the values of 2–3 Å observed at less extreme pH values seem excessive considering that the 4LZT X-ray crystallographic structure was determined at pH 4.5–4.6 (close to the optimal pH of ~5 for the enzyme activity). Inspection of the RMSD plots reveals that most of the structural deviation takes place early in the simulations (as seen in the examples in Fig. 1), often during the first nanoseconds, suggesting that the conformations used in the pK_a calculations are all systematically deviated from the 4LZT structure. In other words, this raises the question of whether the structural rearrangements allowed by the force fields have created any significant distortions in addition to the intended structural reorganization. In particular, the aforementioned slight loss of helical content observed with 53A6 is in agreement with a previous study reporting a tendency of this force field to disrupt helical peptides.[28]

To investigate how the gradual structural rearrangements occurring along the simulations have affected the pK_a predictions, we split our 30 ns of simulations into all possible fragments with sizes multiple of 1 ns. For all these fragments ($(30^2-30)/2 = 435$), we did independent pK_a calculations of HEWL followed by calculation of the corresponding RMSE values. The results obtained for both force fields were combined into a single color map plot (Fig. 2) for easier comparison.

From Figure 2, it is evident that the two force fields have distinct time-dependent RMSE patterns. The GROMOS 43A1 simulations present somewhat poor pK_a predictions at the very initial time segments, especially the first one (0–1 ns) when the protein conformations sampled are still close to the crystal structure. Nevertheless, after some equilibration time, the RMSE values of the next short equilibrated segments decrease signifi-

Table III
Best pK_a Value Predictions

	Experimental pK_a		43A1 (5–8 ns)		53A6 (0–1 ns)	
Residue	Range	Avg	pK_a	Error	pK_a	Error
Glu-7	2.60–3.10	2.85	3.48 ± 0.06	0.63 (0.38)	3.34 ± 0.08	0.49 (0.24)
His-15	5.29–5.43	5.36	4.98 ± 0.11	0.38 (0.31)	5.43 ± 0.06	0.07 (0.00)
Asp-18	2.58–2.74	2.66	3.11 ± 0.08	0.45 (0.37)	3.43 ± 0.05	0.77 (0.69)
Glu-35	6.1–6.3	6.20	5.17 ± 0.09	**1.03** (0.93)	5.57 ± 0.12	0.63 (0.53)
Asp-48	1.2–2.0	1.60	1.73 ± 0.11	0.13 (0.00)	1.46 ± 0.10	0.14 (0.00)
Asp-52	3.60–3.76	3.68	3.65 ± 0.05	0.03 (0.00)	3.37 ± 0.10	0.31 (0.23)
Asp-66	0.4–1.4	0.90	1.02 ± 0.30	0.12 (0.00)	1.30 ± 0.34	0.40 (0.00)
Asp-87	1.92–2.22	2.07	2.12 ± 0.09	0.05 (0.00)	2.38 ± 0.09	0.31 (0.16)
Asp-101	4.02–4.16	4.09	3.82 ± 0.07	0.27 (0.20)	3.80 ± 0.04	0.29 (0.22)
Asp-119	3.11–3.29	3.20	2.90 ± 0.04	0.30 (0.21)	2.87 ± 0.04	0.33 (0.24)
CTer-129	2.63–2.97	2.75	3.21 ± 0.09	0.46 (0.24)	3.06 ± 0.04	0.31 (0.09)
RMSE				0.45 (0.36)		0.42 (0.30)

The segment chosen for each force field was the one with the lowest RMSE. All experimental pK_a ranges were taken from Ref. 74. All errors over 1 pH unit are marked in bold. The errors in parenthesis are calculated relative to the experimental range. RMSE was computed relative to the midpoint of the experimental range (or the range itself when presented in parenthesis). The statistical uncertainties (shown as ± values) of the pK_a values were computed as described in the Theory and Methods section.

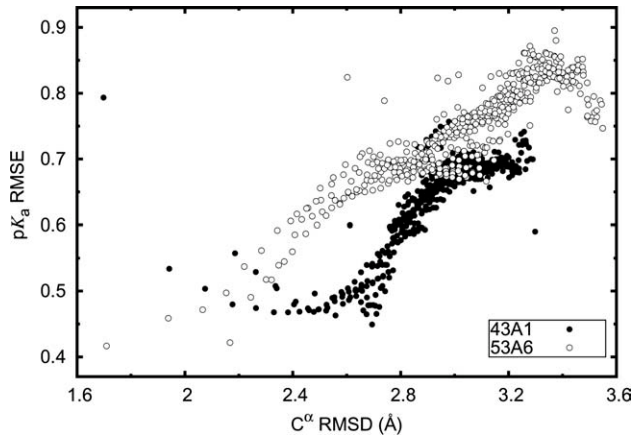

Figure 3

Scatter plot of C^{α} RMSD versus pK_a RMSE calculated from simulations with GROMOS force fields 43A1 and 53A6. Both RMSD and RMSE were obtained over different time segments following the procedure used in Figure 2. The RMSE values are presented in pK units. The RMSD values were computed using the C^{α} atoms and taken relative to the X-ray structure.

cantly. The best pK_a predictions were obtained in one of these segments (5–8 ns) with a remarkable RMSE value of 0.45 pK units. For longer segments or segments starting after ~10 ns, the RMSE values increased consistently, reaching values close to the one reported above for the whole equilibrated segment (0.70).

The GROMOS 53A6 simulations present a very uniform gradient of RMSE values, with the best pK_a predictions at the very initial time segments. In fact, the best predictions occur at the first segment (0–1 ns) with the even more remarkable RMSE value of 0.42 pK units, which is probably the lowest RMSE value ever reported for HEWL. Unlike the 43A1 force field, the 53A6 seems to perform outstandingly well when the protein conformations are still very close to the crystal structure. This becomes even more obvious from the fact that the RMSE values also increase dramatically with increasing structural RMSDs (see Fig. 1). Segments after the first ~10 ns result in poorer pK_a predictions.

Table III presents the pK_a values calculated for the best simulation segments of each GROMOS force field. With the 43A1 force field and using the segment 5–8 ns, the method is able to correctly predict four Asp residues inside the experimental ranges, while still showing a relatively high error (>1 pK unit) for the Glu 35 residue. When we used the first nanosecond of the 53A6 simulations, we were able to predict all residues under 1 pK unit error and, in three cases, we obtained pK_a values inside the experimental range. All the RMSE values mentioned till now were obtained using the usual procedure of taking the midpoint of the measured range as "the" experimental value. As this implies that even computed values lying within the measured experimental range have an associated "error," it can be argued that a more meaningful approach is to compute the error to the nearest experimental bound[31] (which obviously gives lower errors). Therefore, Table III also presents the results obtained with this second approach (0.36 and 0.30 for GROMOS 43A1 and 53A6, respectively), which is probably more directly comparable to the accuracy of experimental measurements. It must be stressed that, by showing these best-segment results, we are not arguing that these are the actual predictions of the simulations, but rather pointing how significantly those predictions seem to be affected by the apparently excessive structural rearrangements imparted by the force fields. It is also important to note that the trends seen in Figure 2 clearly indicate a systematic effect and not a mere statistical dependence on the selected segment.

In Figure 2, it is clear that, especially for the simulations using the 53A6 force field, the pK_a RMSE values are increasing with simulation time and might thus be correlated with the C^{α} RMSD. To investigate this, we did scatter plots of these two properties (Fig. 3). Although the 43A1 simulations present a negative correlation at low RMSD values, they exhibit a high positive correlation for RMSD values above ~2.7 Å. The simulations using 53A6 force field exhibit a strong correlation between the two properties. These results support the idea that, although structural reorganization is obviously crucial in a constant-pH MD simulation, the significant deviations from the crystal structure that eventually take place in the long run may be detrimental to its ability to predict pK_a values.

It is also instructive to examine how the length of the 435 MD segments considered in Figures 2 and 3 affects the frequency of protonation/deprotonation transitions and the precision of the subsequent pK_a calculations. As the number of transitions is expected to be very low near complete protonation or deprotonation, we consider for each site and MD segment the pH value that brings that site closer to its titration midpoint, computing the percentage of transitions relative to the maximum allowed

Table IV

Average of the Percentage of Protonation/Deprotonation Transitions Observed for Each Site

Residue	43A1	53A6
Glu-7	26.2	24.6
His-15	15.1	23.7
Asp-18	14.7	8.3
Glu-35	9.0	3.7
Asp-48	11.2	8.8
Asp-52	3.6	1.5
Asp-66	0.5	5.7
Asp-87	11.0	10.5
Asp-101	32.8	32
Asp-119	33.7	33.9
CTer-129	30.9	25.8

Each average was computed over the 435 MD segments considered in Figure 2, taking for each segment the pH value for which the site being considered is closer to midpoint titration.

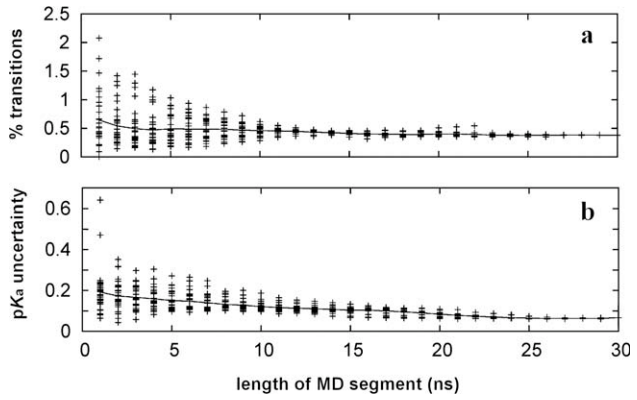

Figure 4

Percentage of protonation/deprotonation transitions (**a**) and pK_a statistical uncertainty (**b**) versus the length of each of the 435 MD segments, for site Asp-66 using the 43A1 force field. Each percentage of transitions corresponds to the pH value for which the site being considered is closer to midpoint titration for that particular MD segment. Lines indicate the average values.

by the deprotonation fraction obtained in that segment; this results in 435 values of the percentage of transitions for each site, which should be representative of the actively titrating site. The averages of those percentages, shown in Table IV, are not strongly dependent on the force field, but they vary widely from "fast" sites like Asp-119 to very "slow" sites like Asp-66. Interestingly, there is not a marked dependence of the percentage of transitions with the length of the MD segment, even for extremely slow sites, as illustrated in Figure 4(a) for Asp-66 using the 43A1 force field—although there is a larger spread of values for short segments, the same average percentage is kept. As seen in Figure 4(b), a similar trend is observed for the pK_a statistical uncertainties, which already reflect the fast or slow nature of the transitions through the correlation time of the protonation states (see Theory and Methods section). Similar features are observed for the other sites. This indicates that the precision of the pK_a calculations is remarkably robust to variations in the length of the MD segment being considered. Furthermore, we do not find any general relation between the percentage of transitions and the accuracy of the calculated pK_a values, which may be positively or negatively correlated, depending on the site. For example, more frequent transitions of Asp-48 are usually associated with an excessive solvent exposure that results in a worse prediction (the transition percentages and the errors of the pK_a prediction have a correlation around 0.6 using either force field). These results seem to indicate that, as in other Monte Carlo simulations,[72,78] there is no reason to expect that a high frequency of transitions will result in a better sampling of states. Each site probably has a typical transition frequency that simply reflects its local environment, and trying to increase that frequency may be actually inadvisable—forcing a site to be constantly titrating may end up favoring conformations that are equally favored by both the protonated and deprotonated states, eventually missing conformations that are more characteristic of either state and crucial in determining the titration behavior of that site.

CONCLUSIONS

Two limitations to the performance of our constant-pH MD methodology were investigated in this work. The first is the ill-defined nature of model compounds in PB calculations, both in terms of molecular definition and pK_a value. The second limitation is the eventual distortion that MM force fields may impart into the structure in addition to the intended reorganization to (de)protonation events.

The model compound issue was addressed here by inverting the typical thermodynamic cycle (euqation 2) in such a way that, using experimental pK_a values from a set of pentapeptides, we were able to obtain calibrated pK^M values that properly model our unphysical model compound fragments. Using this calibration, we calculated new pK_a values for the acidic range of HEWL and obtained a significant improvement on the RMSE value (0.70 instead of the previous 0.82 pK units) using the GROMOS 43A1 force field. Another calculation was done using the newer GROMOS 53A6 force field, which yielded a RMSE value of 0.79 pK units.

The force field influence on the constant-pH MD method's ability to predict pK_a values was studied here in detail by looking individually at all simulation segments with a length multiple of 1 ns, for the 43A1 and 53A6 GROMOS force fields. We observed that the method performed very distinctly using either force field, especially at the initial nonequilibrated segments of the simulations. With the 43A1 force field, the method requires an initial equilibration period after which it performs with very high precision (RMSE of 0.45 pK units) during short nanosecond segments. In contrast, with the 53A6 force field the performance decreases continually throughout the MD simulation, although the RMSE obtained using only nonequilibrated 1 ns simulations is remarkably low (0.42 pK units). These results seem to indicate that, despite allowing for structural reorganization to (de)protonation, MM force fields may also be inducing some structural distortions, as indicated by recent studies.[26–29] In particular, one of those studies[28] reported that the 53A6 force field led to a destabilization of helical peptides, while no such effect was observed for 43A1; this could explain the slight loss of helical content of HEWL observed here using 53A6 and was presumably solved by a recent reparameterization that we plan to investigate.[79] Thus, although current constant-pH MD methods are among the best approaches to predict pK_a

values, the problems of the underlying MM force fields may be hindering their progression toward an unprecedented level of accuracy. On the other hand, these results suggest that very short constant-pH MD simulations may be a suitable approach to predict pK_a values for proteins whose structure is not expected to be markedly dependent on pH.

REFERENCES

1. Stryer L. Biochemistry, 4th ed. New York: Freeman; 1995.
2. Mertz JE, Pettitt BM. Molecular-dynamics at a constant pH. Int J Supercomput Appl 1994;8:47–53.
3. Baptista AM, Martel PJ, Petersen SB. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. Proteins 1997;27:523–544.
4. Baptista AM, Teixeira VH, Soares CM. Constant-pH molecular dynamics using stochastic titration. J Chem Phys 2002;117:4184–4200.
5. Machuqueiro M, Baptista AM. Constant-pH molecular dynamics with ionic strength effects: Protonation-conformation coupling in decalysine. J Phys Chem B 2006;110:2927–2933.
6. Machuqueiro M, Baptista AM. The pH-dependent conformational states of kyotorphin: a constant-pH molecular dynamics study. Biophys J 2007;92:1836–1845.
7. Machuqueiro M, Baptista AM. Acidic range titration of HEWL using a constant-pH molecular dynamics method. Proteins-Struct Funct Bioinform 2008;72:289–298.
8. Machuqueiro M, Baptista AM. Molecular dynamics at constant-pH and reduction potential: application to cytochrome c_3. J Am Chem Soc 2009;131:12586–12594.
9. Campos SRR, Machuqueiro M, Baptista AM. Constant-pH molecular dynamics simulations reveal a β-rich form of the human prion protein. J Phys Chem B 2010;114:12692–12700.
10. Borjesson U, Hunenberger PH. Explicit-solvent molecular dynamics simulation at constant-pH: methodology and application to small amines. J Chem Phys 2001;114:9706–9719.
11. Walczak AM, Antosiewicz JM. Langevin dynamics of proteins at constant-pH. Phys Rev E 2002;66:051911.
12. Dlugosz M, Antosiewicz JM, Robertson AD. Constant-pH molecular dynamics study of protonation–structure relationship in a heptapeptide derived from ovomucoid third domain. Phys Rev E 2004;69:021915.
13. Dlugosz M, Antosiewicz JM. Constant-pH molecular dynamics simulations: a test case of succinic acid. Chem Phys 2004;302:161–170.
14. Burgi R, Kollman PA, van Gunsteren WF. Simulating proteins at constant-pH: an approach combining molecular dynamics and Monte Carlo simulation. Proteins 2002;47:469–480.
15. Mongan J, Case DA, McCammon JA. Constant-pH molecular dynamics in generalized Born implicit solvent. J Comput Chem 2004;25:2038–2048.
16. Lee MS, Salsbury FR, Brooks CL. Constant-pH molecular dynamics using continuous titration coordinates. Proteins 2004;56:738–752.
17. Khandogin J, Brooks CL. Toward the accurate first-principles prediction of ionization equilibria in proteins. Biochemistry 2006;45:9363–9373.
18. Khandogin J, Brooks CL. Constant-pH molecular dynamics with proton tautomerism. Biophys J 2005;89:141–157.
19. Borjesson U, Hunenberger PH. pH-dependent stability of a decalysine α-helix studied by explicit-solvent molecular dynamics simulations at constant-pH. J Phys Chem B 2004;108:13551–13559.
20. Baptista AM. Comment on "Explicit-solvent molecular dynamics simulation at constant-pH: methodology and application to small amines". J Chem Phys 2002;116:7766.
21. Stern HA. Molecular simulation with variable protonation states at constant-pH. J Chem Phys 2007;126:164112.
22. Wallace JA, Shen JK. Predicting pK_a values with continuous constant-pH molecular dynamics. Methods Enzymol: Biothermodyn, Pt B 2009;466:455–475.
23. Meng YL, Roitberg AE. Constant-pH replica exchange molecular dynamics in biomolecules using a discrete protonation model. J Chem Theory Comput 2010;6:1401–1412.
24. Williams SL, de Oliveira CAF, McCammon JA. Coupling constant-pH molecular dynamics with accelerated molecular dynamics. J Chem Theory Comput 2010;6:560–568.
25. Aleksandrov A, Polydorides S, Archontis G, Simonson T. Predicting the acid/base behavior of proteins: a constant-pH Monte Carlo approach with generalized Born solvent. J Phys Chem B 2010;114:10634–10648.
26. Best RB, Buchete NV, Hummer G. Are current molecular dynamics force fields too helical? Biophys J 2008;95:L07–L09.
27. Project E, Nachliel E, Gutman M. Force field-dependant structural divergence revealed during long time simulations of calbindin d9k. J Comput Chem 2009;31:1864–1872.
28. Matthes D, de Groot BL. Secondary structure propensities in peptide folding simulations: a systematic comparison of molecular mechanics interaction schemes. Biophys J 2009;97:599–608.
29. Lange OF, van der Spoel D, de Groot BL. Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. Biophys J 2010;99:647–655.
30. Alexov EG, Gunner MR. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. Biophys J 1997;72:2075–2093.
31. Baptista AM, Soares CM. Some theoretical and computational aspects of the inclusion of proton isomerism in the protonation equilibrium of proteins. J Phys Chem B 2001;105:293–309.
32. Teixeira VH, Cunha CA, Machuqueiro M, Oliveira ASF, Victor BL, Soares CM, Baptista AA. On the use of different dielectric constants for computing individual and pairwise terms in Poisson-Boltzmann studies of protein ionization equilibrium. J Phys Chem B 2005;109:14691–14706.
33. Tanford C, Roxby R. Interpretation of protein titration curves—application to lysozyme. Biochemistry 1972;11:2192.
34. Bashford D, Karplus M. pK_a's of ionizable groups in proteins—atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
35. Delbuono GS, Figueirido FE, Levy RM. Intrinsic pK_a's of ionizable residues in proteins—an explicit solvent calculation for lysozyme. Proteins 1994;20:85–97.
36. Yang AS, Honig B. On the pH-dependence of protein stability. J Mol Biol 1993;231:459–474.
37. Oberoi H, Allewell NM. Multigrid solution of the non-linear Poisson-Boltzmann equation and calculation of titration curves. Biophys J 1993;65:48–55.
38. Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. J Mol Biol 1994;238:415–436.
39. You TJ, Bashford D. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. Biophys J 1995;69:1721–1733.
40. Antosiewicz J, McCammon JA, Gilson MK. The determinants of pK_a's in proteins. Biochemistry 1996;35:7819–7833.
41. Beroza P, Case DA. Including side chain flexibility in continuum electrostatic calculations of protein titration. J Phys Chem 1996;100:20156–20163.
42. Demchuk E, Wade RC. Improving the continuum dielectric approach to calculating pK_a's of ionizable groups in proteins. J Phys Chem 1996;100:17373–17387.
43. Gibas CJ, Subramaniam S. Explicit solvent models in protein pK_a calculations. Biophys J 1996;71:138–147.
44. van Vlijmen HWT, Schaefer M, Karplus M. Improving the accuracy of protein pK_a calculations: conformational averaging versus the average structure. Proteins 1998;33:145–158.

45. Georgescu RE, Alexov EG, Gunner MR. Combining conformational flexibility and continuum electrostatics for calculating pK_a's in proteins. Biophys J 2002;83:1731–1748.
46. Beroza P, Fredkin DR, Okamura MY, Feher G. 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 1991;88:5804–5808.
47. Gilson MK. Multiple-site titration and molecular modeling—2 rapid methods for computing energies and forces for ionizable groups in proteins. Proteins 1993;15:266–282.
48. Nielsen JE, Vriend G. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK_a calculations. Proteins 2001;43:403–412.
49. Nielsen JE, McCammon JA. On the evaluation and optimization of protein X-ray structures for pK_a calculations. Protein Sci 2003;12:313–326.
50. Fisher SJ, Wilkinson J, Henchman RH, Helliwell JR. An evaluation review of the prediction of protonation states in proteins versus crystallographic experiment. Crystallogr Rev 2009;15:231–259.
51. Warshel A. Calculations of enzymatic-reactions—calculations of pK_a, proton-transfer reactions, and general acid catalysis reactions in enzymes. Biochemistry 1981;20:3167–3177.
52. Jackson, J. D. Classical electrodynamics, 2nd ed. New York: Wiley, 1975.
53. Nozaki Y, Tanford C. Examination of titration behavior. Methods Enzymol 1967;11:715–734.
54. Grycuk T. Revision of the model system concept for the prediction of pK_a's in proteins. J Phys Chem B 2002;106:1434–1445.
55. Tanokura M. ^{1}H-NMR study on the tautomerism of the imidazole ring of histidine-residues: 1. microscopic pK values and molar ratios of tautomers in histidine-containing peptides. Biochim Biophys Acta 1983;742:576–585.
56. Thurlkill RL, Grimsley GR, Scholtz JM, Pace CN. pK values of the ionizable groups of proteins. Protein Sci 2006;15:1214–1218.
57. Grimsley GR, Scholtz JM, Pace CN. A summary of the measured pK values of the ionizable groups in folded proteins. Protein Sci 2009;18:247–251.
58. Bashford D, Karplus M. Multiple-site titration curves of proteins—an analysis of exact and approximate methods for their calculation. J Phys Chem 1991;95:9556–9561.
59. Bashford D, Gerwert K. Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. J Mol Biol 1992;224:473–486.
60. Baptista AM, Martel PJ, Soares CM. Simulation of electron-proton coupling with a Monte Carlo method: application to cytochrome c_3 using continuum electrostatics. Biophys J 1999;76:2978–2998.
61. Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF. The GROMOS biomolecular simulation program package. J Phys Chem A 1999;103:3596–3607.
62. van Gunsteren WF, Berendsen HJC. Computer-simulation of molecular-dynamics—methodology, applications, and perspectives in chemistry. Angew Chem Int Ed 1990;29:992–1023.
63. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 2004;25:1656–1676.
64. Gilson MK, Sharp KA, Honig BH. Calculating the electrostatic potential of molecules in solution—method and error assessment. J Comput Chem 1988;9:327–335.
65. Berendsen HJC, Vanderspoel D, Vandrunen R. GROMACS—a message-passing parallel molecular-dynamics implementation. Comput Phys Commun 1995;91:43–56.
66. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 2001;7:306–317.
67. Hermans J, Berendsen HJC, Vangunsteren WF, Postma JPM. A consistent empirical potential for water–protein interactions. Biopolymers 1984;23:1513–1518.
68. Tironi IG, Sperb R, Smith PE, Vangunsteren WF. A generalized reaction field method for molecular-dynamics simulations. J Chem Phys 1995;102:5451–5459.
69. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
70. DeLano WL. The PyMOL Molecular graphics system. San Carlos, CA: DeLano Scientific; 2002. Available at: http://www.pymol.org.
71. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in C, the art of scientific computing, 2nd ed. New York: Cambridge University Press; 1992.
72. Allen MP, Tildesley DJ. Computer simulations of liquids. New York: Oxford University Press; 1987.
73. Williams T, Kelley C (2007) Gnuplot: An Interactive Plotting Program. Available at: http://www.gnuplot.info.
74. Bartik K, Redfield C, Dobson CM. Measurement of the individual pK_a values of acidic residues of hen and turkey lysozymes by 2-dimensional H^1-NMR. Biophys J 1994;66:1180–1184.
75. Chen LL, Hodgson KO, Doniach S. A lysozyme folding intermediate revealed by solution X-ray scattering. J Mol Biol 1996;261:658–671.
76. Sasahara K, Demura M, Nitta K. Partially unfolded equilibrium state of hen lysozyme studied by circular dichroism spectroscopy. Biochemistry 2000;39:6475–6482.
77. Sasahara K, Demura M, Nitta K. Equilibrium and kinetic folding of hen egg-white lysozyme under acidic conditions. Proteins 2002;49:472–482.
78. Frenkel D, Smit B. Understanding molecular simulation: from algorithms to applications. San Diego, CA: Academic Press; 2001.
79. Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J 2011;40:843–856.