

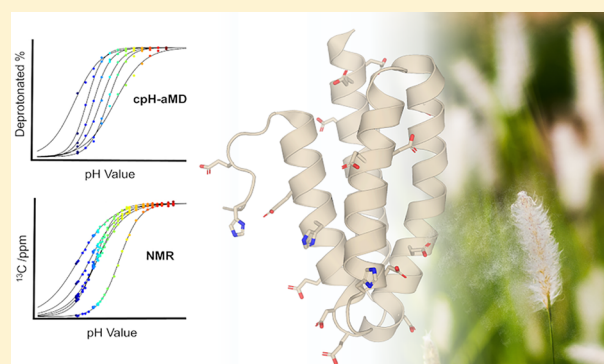
pH-Dependent Protonation of the Phl p 6 Pollen Allergen Studied by NMR and cpH-aMD

Florian Hofer,[†] Valentin Dietrich,[‡] Anna S. Kamenik,[†] Martin Tollinger,^{‡,‡} and Klaus R. Liedl^{*,†,‡}

[†]Institute for General, Inorganic and Theoretical Chemistry and [‡]Institute for Organic Chemistry, Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innrain 80/82, A-6020 Innsbruck, Austria

Supporting Information

ABSTRACT: We use state-of-the-art NMR experiments to measure apparent pK_a values in the native protein environment and employ a cutting-edge combination of enhanced sampling and constant pH molecular dynamics (MD) simulations to rationalize strong pK_a shifts. The major timothy grass pollen allergen Phl p 6 serves as an ideal model system for both methods due to its high number of titratable residues despite its comparably small size. We present a proton transition analysis as intuitive tool to depict the captured protonation state ensemble in atomistic detail. Combining microscopic structural details from MD simulations and macroscopic ensemble averages from NMR shifts leads to a comprehensive view on pH dependencies of protonation states and tautomers. Overall, we find striking agreement between simulation-based pK_a predictions and experiment. However, our analyses suggest subtle differences in the underlying molecular origin of the observed pK_a shifts. From accelerated constant pH MD simulations, we identify immediate proximity of opposite charges, followed by vicinity of equal charges as major driving forces for pK_a shifts. NMR experiments on the other hand, suggest only a weak relation of pK_a shifts and close contacts to charged residues, while the strongest influence derives from the dipolar character of α helices. The presented study hence pinpoints opportunities for improvements concerning the theoretical description of protonation state and tautomer probabilities. However, the coherence in the resulting apparent pK_a values from simulations and experiment affirms cpH-aMD as a reliable tool to study allergen dynamics at varying pH levels.



INTRODUCTION

The pH level is well known to be a critical environmental determinant of protein function and stability.^{1–3} Small changes in solution pH can be sufficient to completely destabilize a protein or change its activity profile.^{4,5} Proteins react to pH changes via protonation or deprotonation of titratable residues, thereby changing their charge distribution. The nature and strength of this effect depends on the number of affected titratable residues, on their structural environment, and on possible compensation of introduced charges. The acidity of a titratable group is represented by its pK_a value and directly dependent on its electrostatic surroundings.^{6,7} Thus, while for isolated amino acids in solution these pK_a values are easily measurable, they can be drastically perturbed and challenging to measure within the context of a protein.⁸ Yet, an accurate representation of protonation states and tautomers is paramount for reliable experimental and especially computational studies of protein structures.⁹

An additional challenge arises from the dynamic nature of proteins.¹⁰ The free-energy surface of proteins in solution is vast and rugged, where each minimum represents a different conformational state.^{11,12} From the multitude of constantly interchanging conformational states of a protein, each exhibits

individual solvent accessibility, charge distribution, etc., resulting in variations of the according microscopic pK_a values.^{13,14} The ensemble average of these microscopic pK_a values is then represented as the macroscopic or apparent pK_a value. NMR spectroscopy offers the unique possibility of directly measuring such apparent pK_a values experimentally. Numerous homonuclear and heteronuclear approaches can be employed for monitoring protonation probabilities of titratable groups in pH titrations and extracting their pK_a values in a native, dynamic protein environment.^{8,15}

Similar to NMR experiments, molecular dynamic (MD) simulations also capture an ensemble of protein conformations.^{16,17} While the accuracy of simulations is restricted by the accuracy of the applied force field and the captured sampling time, they provide a time-resolved representation of the configurational ensemble of proteins in atomistic detail. However, classic MD simulations do not allow the breaking of bonds and are thus not able to simulate changes in protonation states. Hence, the protonation states of each amino acid must be chosen during the structure preparation

Received: June 3, 2019

Published: September 2, 2019

steps and cannot change during the simulation. Choosing an appropriate protonation state can be very challenging and requires either available experimental pK_a values of the studied system or reliable means to predict the unknown pK_a values.⁹

Over the last decades, a plethora of pK_a prediction tools have been developed.¹⁸ Most commonly, static methods like PROPKA¹⁹ or H++²⁰ are applied, which estimate pK_a values from a single input structure and employ an implicit solvent model, making the calculation relatively fast and the input preparation straightforward. However, as discussed above, proteins do not occur as one single static structure. Hence, dynamic methods were developed, like the family of constant pH molecular dynamics (cpH-MD) methods, which rely on an ensemble of structures to estimate pK_a values. In contrast to static methods, these approaches are capable of capturing the interplay of changes in pK_a and changes in structure, yet at a higher computational cost.⁹

In the following, we will only briefly discuss the key points of cpH-MD; for more in-depth discussions, the reader is pointed to the respective works.^{21–34} There are mainly two different approaches to cpH-MD: using continuous²⁸ or discrete protonation states.^{22,27} In the first, originally implemented in the software package CHARMM,³⁵ the protonation states are sampled along a continuous titration coordinate employing the λ -dynamics approach. In contrast to that, discrete protonation states are sampled via Metropolis Monte Carlo (MC³⁶) moves, which happen at defined intervals over the course of the simulation. While originally implicit solvent models³⁷ were used, both methods have seen further adaptations to also make use of explicit solvent models.^{30,32,33,38} In this study, we employed the discrete protonation state approach, as implemented in the AMBER³⁹ simulation package.^{27,32} To achieve faster pK_a convergence as well as higher pK_a prediction accuracy, combinations of cpH-MD and enhanced sampling techniques like replica exchange (REMD⁴⁰) and accelerated MD (aMD⁴¹) have been reported.^{31,32,42,43} Especially, the coupling with aMD showed that an extensive conformational sampling significantly increased the accuracy of the predicted pK_a values.⁴³ In this study, we perform a detailed investigation on how the conformational states from cpH-aMD simulations and their respective microscopic pK_a values relate to macroscopic pK_a values from NMR experiments.

We employ the major timothy grass pollen allergen Phl p 6 as a model system for this study. Phl p 6 is one of the most important grass allergens with over 75% of grass pollen allergic patients having IgE antibodies recognizing Phl p 6.^{44,45} As it is the case for many allergen proteins, the biological function of Phl p 6, as well as the source of its allergenicity, is unknown.

Phl p 6 excels as a model system for our study on the one hand, because with 111 residues it is a rather small protein, which facilitates efficient sampling of its conformational space. On the other hand, based on major antigen processing pathways, pH stability is considered a decisive factor for a protein's allergenicity.^{46,47} In general, after uptake, e.g., by inhalation in the case of pollen allergens, the allergen enters an antigen-presenting cell via endocytosis. In the endosome, the proteins are proteolytically processed into peptides, which are then loaded onto class II major histocompatibility complex (MHC) molecules. The MHC peptide complexes are transported to the cell surface and presented to naive T-cells. Recognition of the linear T-cell epitope then triggers the immune response.^{48,49} The digestion of the proteins is tightly coupled to a strong acidification of the endosome during its

maturation, i.e., a drop in pH from around 7 to around 4. The higher the stability of the protein, the harsher the condition in terms of pH that is needed to digest the protein.^{46,47,50} As discussed by Scheiblhofer et al., the fold stability of a protein during this process determines the associated immune response. As already discussed above, a protein's reaction to pH changes is determined by its titratable, i.e., charged residues. With 28 residues, including histidines, out of the 104 residues present in the X-ray structure (PDB code 1NLX⁵¹), Phl p 6 shows a rather large number of charged residues for its small size. In Figure 1, all glutamic, aspartic, and histidine

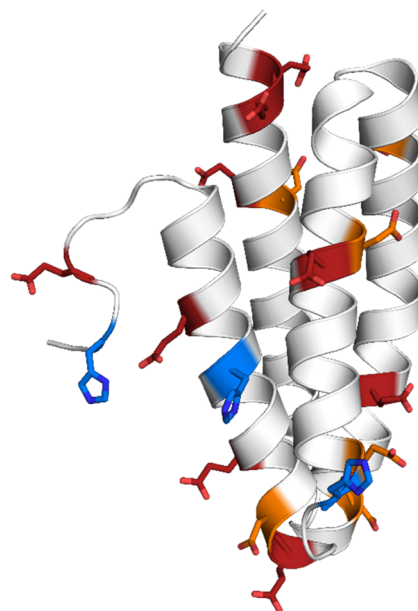


Figure 1. Crystal structure of the Phl p 6 allergen (PDB code 1NLX).⁵¹ Aspartic, glutamic, and histidine residues are shown as orange, red, and blue sticks, respectively. With 28 charged residues out of 111 residues in total (104 residues resolved in crystal structure), the protein has a very high ratio of charges relative to its small size. Of the 28 charged residues, the 18 residues shown in the picture were allowed to titrate in the simulation.

residues, which we consider titratable in the acidic pH range up to a pH of 8.0, are shown as sticks. Thus, predicting the redistribution of protonation states upon acidification is far from trivial.

Combining cpH-aMD simulations and NMR experiments, we present a strategy for detailed studies on the titration behavior of proteins. We envisage our approach as a reliable foundation to investigate the structural stability of antigens during endolysosomal degradation.

■ METHODS

Simulation Setup. The starting structure for the simulation was prepared from the wildtype X-ray structure of Phl p 6 (PDB code 1NLX, chain A⁵¹) with the program MOE (molecular operating environment⁵²). Of the 111 residues, only 104 residues were resolved in the crystal structure. Specifically, four residues on the C-terminus and three residues at the N-terminus of the protein are missing, including the starting methionine. However, the missing residues did not include any aspartates, glutamates, or histidines. In the NMR experiments, all 111 amino acids were present.

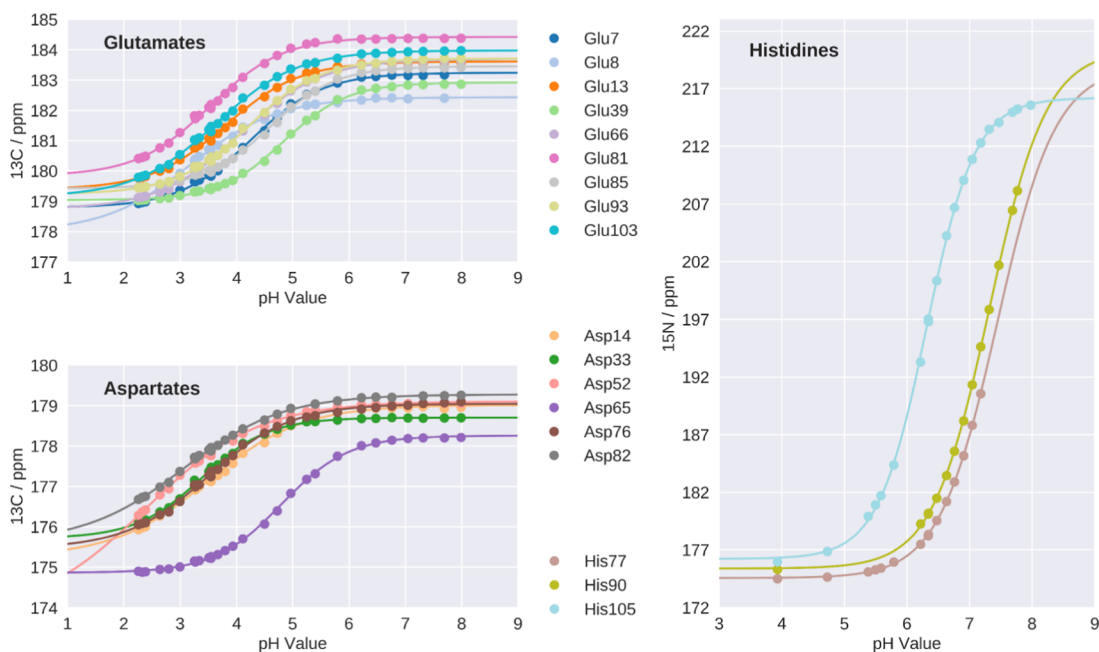


Figure 2. Experimental titration curves with chemical shifts of side chain $^{13}\text{C}'$ for ASP and GLU, $^{15}\text{N}\delta 1$ for HIS77 and HIS90 and $^{15}\text{N}\epsilon 2$ for HIS105.

The protein shows exclusively α -helical structure elements, connected by short loops (see Figure 1). Notably, the C-terminal residues 96–104 are not part of a helix but in fact adopt a looplike conformation.

The LEaP module of AmberTools 17³⁹ was used to add missing hydrogens, as well as create topology and starting coordinate files. The AMBER ff99SB force field⁵³ was used, along with the necessary force field modifications for constant pH simulations.^{27,32} The GB radii of the aspartate and glutamate oxygens were reduced to 1.3 Å, as suggested by Swails et al.³² The protein was soaked in a truncated octahedral TIP3P water box with a minimum wall distance of 10 Å.⁵⁴

Before production simulations were carried out, the systems were minimized and relaxed with an elaborate protocol previously developed in our group.⁵⁵

All simulations were performed with the graphics processing unit implementation of the pmemd module of AMBER 17.⁵⁶ The Berendsen barostat⁵⁷ with a relaxation time of 2 ps was used to maintain atmospheric pressure, as well as the Langevin thermostat with a collision frequency of 5 ps⁻¹ to keep the system at 300 K.⁵⁸ The SHAKE⁵⁹ algorithm was used to restrain all bonds involving hydrogens, allowing the use of a 2 fs timestep. Long-range electrostatics were treated with the particle-mesh Ewald method,⁶⁰ and a nonbonded cutoff of 8 Å was used. Simulations were performed from pH 2.5 to 8.0 with a 0.5 spacing. Protonation state changes were performed every 200 steps, followed by 200 steps of solvent relaxation after a successful exchange. For the GB calculations a salt concentration of 0.1 was used. Acceleration was achieved with the dihedral boost algorithm of AMBER 17, appropriate boosting parameters were calculated according to the work of Pierce et al. from short classical MD simulations and can be found in the Supporting Information.⁶¹ Frames were collected every 1000 steps. All simulations were run for 1 μs , resulting in a total simulation time of 12 μs .

Analysis. All analyses were performed using the programs cptraj and pytraj from the AmberTools 17 package,⁶² as well as inhouse python scripts. All structural visualizations were produced with PyMol.⁶³

For all analyses, the trajectories were reweighted using a McLaurin series to the 10th order.^{61,64}

Titration data from constant pH simulations were collected with the program cphstats from AmberTools.³⁹ The modified Hill equation was used to estimate pK_a values. Histidine tautomer distributions were calculated directly from the simulation output. The δ -tautomer will be denoted as HID in the following sections, ϵ -tautomer as HIE and the doubly protonated, i.e., positively charged form as HIP. Shifts in pK_a were calculated using the pK_a values for free tripeptides of the form acetyl-GXG-amide (N- and C-terminally blocked tripeptides), as measured by Platzer and McIntosh as references.⁸ Convergence of the calculated pK_a values was monitored by computing the cumulative averages with cphstats.

Distance histograms were calculated between titrated residues and nontitrated basic residues. We used the heavy atom centers of mass of the titratable head groups of glutamate, aspartate, and histidine, as well as for the guanidinium group of arginine as reference points. For lysine only, the position of the side-chain nitrogen was used. Furthermore, distances to helix termini were measured for all titrated residues. Reference points for helix termini were defined as the center of mass of the $\text{C}\alpha$ carbon atoms of the first three residues of the respective helix. Angles between the titrated residues and the helix were calculated by calculating the angle between the helix axis, defined as the vector from the C- to the N-terminal end of the helix, and the distance vector from the center of the helix to the titrated residue.

To analyze the transition probabilities between strongly coupled protonation states, we calculated transition matrices for all pH values. For this, we focused on the close interaction of residues GLU81, ASP82, and GLU85. We defined 8 states

based on the titration state of these residues, as shown in Table 2. For this purpose, we considered all 4 protonated states of glutamate and aspartate as one state. The transition matrices were visualized as network plots, in which the circle size relates to state populations and arrow sizes to transition probabilities. State positions were chosen so that transitions on the edges correspond to single protonation state changes, diagonal transitions encode a double transition, and finally a transition over the main diagonal relates to a change in all three protonation states at the same time.

NMR Spectroscopy. Phl p 6 (residues 1–111) was recombinantly expressed in *Escherichia coli* BL21(DE3) Star cells using M9 minimal media supplemented with $^{13}\text{C}_6\text{-D}$ -glucose and $^{15}\text{NH}_4\text{Cl}$ as carbon and nitrogen sources, respectively. The protein was purified by hydrophobic interaction and size exclusion chromatography employing HiTrap Phenyl FF and HiLoad 16/600 Superdex 75 columns (GE Healthcare). NMR samples contained 1 mM protein, 10 mM citrate-phosphate buffer, and 10% D_2O . All NMR experiments were performed at 25 °C on 500 MHz Agilent DirectDrive 2 and 700 MHz Bruker Avance Neo spectrometers. Backbone amide resonance assignment was obtained by use of two-dimensional ^1H – ^{15}N -HSQC and three-dimensional HNCACB, CBCA(CO)NH experiments at pH 7.0.

Side-chain pK_a values of Asp and Glu were determined using two-dimensional spectra that correlate side-chain C' and backbone amide $^1\text{H}^{\text{N}}$ chemical shifts as reported.⁶⁵ For His side chains, ^{15}N chemical shifts were recorded in two-dimensional $^1\text{H}^{15}\text{N}$ HSQC spectra where the INEPT transfer delay was adjusted to $^2\text{J}_{\text{HN}}$ couplings. Histidine $\text{H}^{\beta 2}$ assignments were obtained from a HBCBCGCDHD experiment,⁶⁶ and the corresponding $\text{H}^{\beta 1}$ were identified in the two-dimensional $^1\text{H}^{15}\text{N}$ HSQC spectra. In titration experiments, pH values were adjusted between pH 2.2 and 8.5 by adding small aliquots of HCl or NaOH. All pH values were measured using ^{15}N imidazole and formic acid ^{13}C chemical shifts as internal pH references in two-dimensional $^1\text{H}^{15}\text{N}$ and $^1\text{H}^{13}\text{C}$ correlation spectra as described.⁶⁷ Side-chain $^{13}\text{C}'$ (Asp, Glu) and ^{15}N (His) chemical shift data were fit by standard equations for a single ionizable group¹⁵ to obtain pK_a and limiting chemical shift values (fitting equations are shown in the Supporting Information). Tautomeric distributions of His side chains were determined from the so-obtained $^{15}\text{N}^{\beta 1}$ and $^{15}\text{N}^{\beta 2}$ limiting chemical shifts and pK_a values as described.⁶⁸

RESULTS

pK_a Values. Side-chain pK_a values of all six aspartates, all nine glutamates, and the three histidines in Phl p 6 were determined experimentally using heteronuclear two-dimensional NMR spectroscopy (Figure 2). The data reveal pK_a values for aspartates and glutamates between 2.4 and 4.8, while for histidines, experimental pK_a values are in the range between 6.5 and 7.4. In cpH-aMD simulations, pK_a values for all 18 titratable residues were calculated. The respective titration curves can be found in the Supporting Information (Figure S1). To estimate the accuracy of the simulation-derived protonation state ensemble, we benchmarked these pK_a values against the experimental (NMR) pK_a values (see Table 1). The correlation between both methods is visualized in Figure 3 with a root mean square error of 0.89 and a Pearson correlation coefficient of 0.81. Out of 18 titrated residues, 11 predictions lie within an error of 1 pK_a unit, visualized as gray lines in Figure 3.

Table 1. Comparison of the pK_a Values Measured by NMR and Predicted by Simulation with Absolute Differences

ResID	NMR	simulation	difference
GLU7	4.25	4.20	0.05
GLU8	3.19	3.70	0.51
GLU13	3.80	4.83	1.03
ASP14	3.50	3.06	0.44
ASP33	3.38	4.37	0.99
GLU39	4.84	6.04	1.20
ASP52	2.4	2.30	0.10
ASP65	4.80	3.92	0.88
GLU66	4.06	5.25	1.19
ASP76	3.50	3.48	0.02
HIS77	7.4	6.96	0.44
GLU81	3.52	4.76	1.24
ASP82	3.06	4.81	1.75
GLU85	4.59	5.30	0.71
HIS90	7.3	6.00	1.30
GLU93	4.21	4.27	0.06
GLU103	3.65	4.66	1.01
HIS105	6.5	6.52	0.02

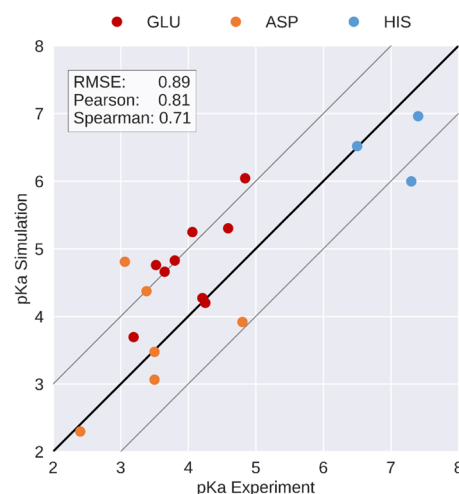


Figure 3. Correlation plot between the pK_a values measured by NMR and predicted by simulation. Pearson and Spearman correlation coefficients and RMSE are shown. The black line denotes the ideal correlation and gray lines denote the prediction error margin of ± 1 pK_a unit, as typically reported in the literature.

In both experiment and simulation, we identified residues with pK_a values strongly deviating from reported pK_a values of the corresponding tripeptides.⁸ Figure 4 shows the absolute deviations of the measured and predicted pK_a values from the reference values for free tripeptides. Notably, ASP52 shows the largest negative pK_a shift of all titratable residues, i.e., a strong acidification, consistently in both experiment and simulation. Other notable acidic (GLU8 and ASP14) and basic shifts (GLU39, HIS77 and GLU85) were identified with acceptable agreement between simulation and NMR. However, some shifts seen in the experiment were significantly mispredicted in the simulation (GLU13, GLU81, ASP82, HIS90) in the opposite direction. The strongest disagreement is shown by ASP82 with an unsigned error of 1.75 pK_a units. Also, its neighboring residue GLU81 shows a significant prediction error of 1.24 pK_a units. Of the three titrated histidines, HIS90 shows a notable prediction error of 1.30, while HIS77 only

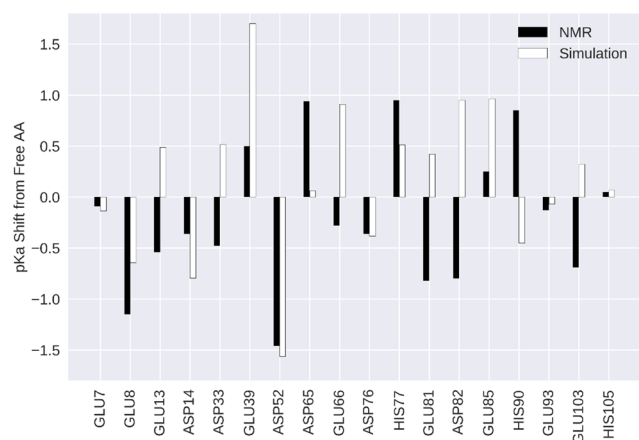


Figure 4. Deviations of measured and predicted pK_a values from the respective tripeptide pK_a values. Reference values are 3.86 for ASP and 4.34 for GLU and 6.45 for HIS as published by Platzer et al.⁸ Black and white bars represent experiment and simulation respectively.

shows a moderate error of 0.44 pK_a units and HIS105 is in almost perfect agreement with experiment.

Transition Analysis. A generally poor pK_a prediction can be seen for residues GLU81, ASP82 (both mispredicted), and GLU85 (shift strongly overestimated). For both ASP82 and its flanking residue GLU85 the fit to the Hill equation is not optimal (see [Supporting Information](#)). The structural interplay of these residues is visualized in [Figure 5A](#), which shows that all three residues are oriented in the same direction, pointing into the solvent. To ensure that our sampling of the protonation state space is sufficient and to investigate the correlation of protonation states of spatially close residues, we performed a protonation state transition analysis. For this, we focused on the close arrangement of GLU81, ASP82, and GLU85. We defined eight states based on the titration state of these residues, as shown in [Table 2](#). To this end, we considered all four protonated states of glutamate and aspartate as one state. The transition matrices were visualized as network plots, in which circle sizes relate to state populations and arrow sizes, to transition probabilities. State positions were chosen so that transitions on the edges correspond to single protonation state

Table 2. State Definition of the Model System for the Protonation State Transition Analysis

state	GLU81	ASP82	GLU85
0	0	0	0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1

^aProtonation states of the residues are denoted as 0 (deprotonated, negatively charged) or 1 (protonated, neutral). State number represents binary encoding of the protonation of the three residues.

changes, diagonal transitions encode a double transition, and finally a transition over the main diagonal relates to a change in all three protonation states at the same time.

We found a distinct, pH-dependent pattern in their protonation states as well as in the transitions between the states we defined in [Table 2](#). The transition count matrices show high numbers of transition between the different states across all pH values. At extreme pH values, i.e., 2.5 and 8.0, the completely protonated and completely unprotonated states, respectively, are populated almost exclusively. However, a few transitions to sparsely populated states are still visible. As the pH value starts to increase from 2.5, state populations start to shift, with states 1 (GLU85 protonated) and 5 (GLU81 and GLU85 protonated) being the prominent ones at moderate pH values (between 4.0 and 5.5). As the pH further increases, the populations shift first to the monoprotonated states 1 and 2 and finally almost completely to the unprotonated state 0 at pH 8. An exemplary visualization of the data of pH 5.0 is shown in [Figure 5B](#). Visualizations for all pH values can be found in the [Supporting Information](#) ([Figure S2](#)).

Tautomer Estimation. From both the experimental as well as the simulation data, we calculated the tautomer distribution of the histidines at each measured and simulated pH value ([Figure 6](#)). Clearly, the prediction is poor for HIS77 and HIS90. In both cases, the fraction of the δ -tautomer is significantly overestimated in the simulations, whereas in the experiment, a nearly 1:1 ratio of δ - and ϵ -tautomer is observed.

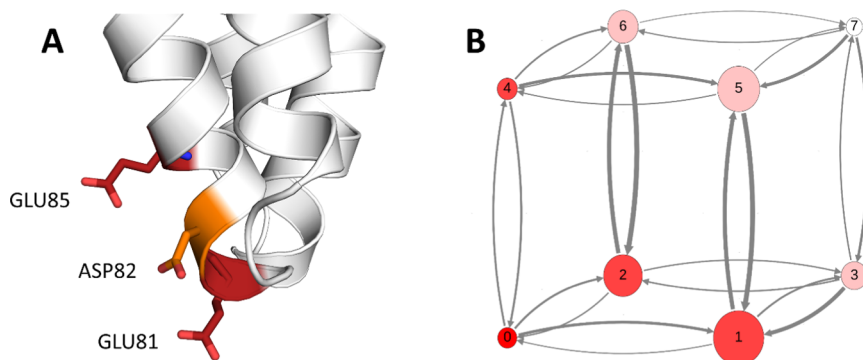


Figure 5. (A) Interaction and spatial vicinity of the residues GLU81, ASP82, and GLU85 in PhI p 6 the crystal structure, which was used as input for the simulations. (B) Analysis of protonation state population and transitions of the 3-residue model system. Analysis at pH 5 is shown as representative. States denoted as 0–7 (see [Table 2](#)), circle size denotes state population, and arrow thickness indicates transition probability from one state to the other (cutoff at 0.01). At pH 5, states 0 and 7, i.e., fully deprotonated and fully protonated, are predicted to be least populated, with high transition probabilities to different states. No transitions over a diagonal, i.e., 2 protonation state change, can be observed at this pH within the cutoff.

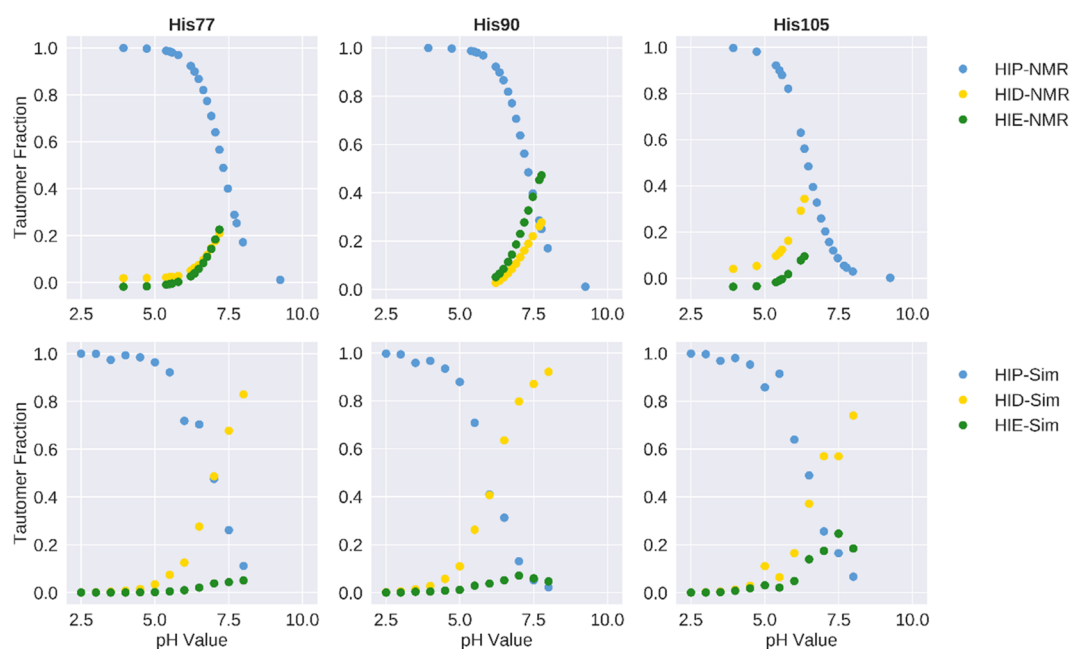


Figure 6. Tautomer distributions for all three histidines in the system were calculated for experiment (top) and simulation (bottom). Clear overestimation in the prediction of the δ -tautomer for HIS77 and HIS90 can be observed, whereas a good agreement can be seen for HIS105.

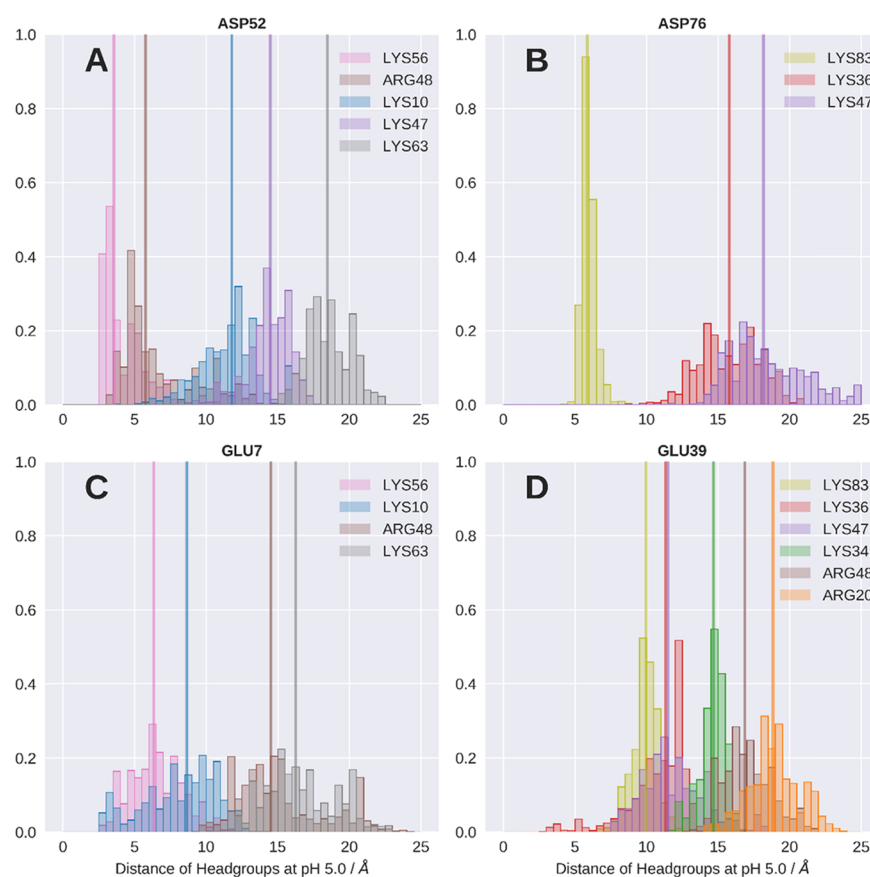


Figure 7. Normalized histograms of the distances between the carboxyl group of selected acidic residues and the head groups (amino and guanidinium group for LYS and ARG, respectively) of near positively charged residues. Analysis at pH 5.0 is shown representatively. Vertical lines denote the median of the respective distance distribution.

As shown above, also the pK_a prediction for these two residues is not optimal (see Table 1), especially for HIS90. However, for HIS105, we find perfect agreement between simulation and

experiment for both the pK_a value (see Table 1) and the tautomer ratio.

Distance Analysis. To elucidate the origin of pK_a perturbances, we analyze the microscopic chemical environ-

ment of each titratable acidic residue. We calculated distance histograms of nearby positively charged residues. We find that a few of the acidic residues show strong and close interactions with nearby positively charged residues with the medians of the distributions below 5 Å across all pH values. The histogram of ASP52 at pH 5.0 is shown in Figure 7A as an example. A close interaction with LYSS6 with the median (vertical lines) of the distance distribution at 3.5 Å is visible. Another, more distant contact is formed with ARG48 with a distance median of 5.8 Å. On the other hand, we also find residues that only form less pronounced and more distant interactions; exemplary histograms are shown in Figure 7B–D.

DISCUSSION

In the presented work, we use NMR experiments and MD simulations to profile protonation probabilities within the folded protein environment of the major timothy grass allergen Phl p 6. The interplay of both methods not only unveils the molecular origin of several strongly shifted residue pK_a values but also outlines the current limitations of protonation state predictions.

As can be seen from Figure 3, we find a remarkable overall correlation between experiment and simulation in terms of pK_a values indicating that both approaches capture similar protonation state ensembles. For 11 out of the 18 titrated residues, the difference between experiment and simulation is less than 1 pK_a unit, an error margin typically reported in the literature. Also, the predicted pK_a values of the remaining residues do not exceed this margin substantially, with the exception of ASP82.

Furthermore, the convergence analysis of the predicted pK_a values (see Figure S3) suggests that while a few residues reach a converged pK_a value within 50 ns of simulation time, other residues need significantly longer to reach a converged pK_a value, with the most prominent example being ASP52, which reaches a converged pK_a value after about 800 ns. The majority of the residues show converged pK_a values after about 100–200 ns of simulation time. Moreover, we generally see an improvement in the pK_a prediction with longer simulation times.

As highlighted in Figure 4, both methods identify strong pK_a shifts, i.e., large deviations of pK_a values within the protein from the pK_a values of the respective tripeptides. This analysis further illustrates the overall remarkable consistency between the predicted pK_a shifts and the NMR experiments. Notably, the strong acidic shift of ASP52 was found with both approaches with near perfect agreement. Also, the basic shift of GLU39 was found with both techniques but was overestimated in the simulation—an effect that has been reported previously.³² Despite the overall agreement between NMR and cpH-aMD in terms of pK_a values, we can note a few residues with significant prediction errors. In the following text, we will discuss in detail analyses and hypotheses rationalizing potential driving forces for the observed inaccuracies.

The strongest discrepancy is found for ASP82 with an unsigned error of 1.75 pK_a units. With a measured pK_a value of 3.06, the residue is found to be distinctly more acidic than the free tripeptide (pK_a of 3.86⁸). However, in the simulation the opposite is the case, in that a more basic pK_a value of 4.81 is predicted. In the crystal structure (visualized in Figure 5A), ASP82 is flanked by two other acidic residues GLU81 and GLU85, packed closely together with the side chains oriented in the same direction. Unintuitively, the pK_a of ASP82 is

experimentally determined to be the most acidic one of these three residues. The pK_a of the flanking residue GLU81 is also measured to be considerably more acidic than that of the free tripeptide (3.52 vs 4.34⁸), while the pK_a of GLU85 (4.59) shows no strong perturbation. As can clearly be seen from Table 1, the simulation fails to predict the pK_a values for all three residues.

To evaluate the ability to capture the correct titration behavior of strongly coupled residues, we performed a protonation state-based transition analysis for the GLU81, ASP82, and GLU85 residue pack. This method allowed a very detailed yet intuitive representation of the protonation state probabilities and exchange rates between them at each simulated pH level. From the high numbers of transitions between the states across all pH levels, we conclude that at no pH level the simulation gets trapped in a protonation state configuration. Also, the shifts in the state probabilities and correlation of protonation states in dependence of the change in pH value appear to be conclusive. On the basis of these observations, we can exclude insufficient sampling of protonation states as a cause of the large prediction error.

As mentioned above, the transition analysis additionally provided us with insights into the underlying kinetics of the protonation state changes. From the transition probabilities (Figure 5B), we can clearly see that within our cutoff of 0.01 no transitions occur at any pH value over the principal diagonal, i.e., all protonation states change at the same constant pH step. Also, other diagonal transitions, i.e., 2-proton transitions are very rare. The fact that transitions occur almost exclusively by changing one protonation state at a time, despite the fact that multititrations would be possible, suggests that a short local structural equilibration of the new protonation state is necessary to adapt to the new electrostatic potential and to make the next change possible.

While we surmise that the reason for the prediction error of the three discussed residues is not an inefficient protonation state sampling, we do see a few other possible reasons for the disagreement of experiment and simulation.

First, the correct description of tightly packed acidic residues is a known issue of implicit solvent models.^{69,70} Paired with the lack of counter ions in the simulations, this might lead to mispredicted pK_a values. Furthermore, the acidification of GLU81 and ASP82 coupled with their exposed position in a loop between two helices in contrast to GLU85 indicates that they are stabilized by a cation during the experiments and GLU85 is not. This effect would be completely missed in the simulations since they are not considered in the implicit solvent steps.

Secondly, although the protonation state space is well covered, the sampled conformational ensemble might still be insufficient. Limited conformational sampling in turn also limits the protonation state ensemble and hence affects the accuracy of apparent pK_a values estimated from this limited ensemble.

This assessment is supported by the pK_a and tautomer predictions of the histidines. As can be seen in the crystal structure in Figure 1, HIS77 and HIS90, for which both the pK_a and the tautomer estimation are quite poor (Table 1 and Figure 6), are located within stable α -helices. In contrast, HIS105, for which both estimations are in perfect agreement with experiment, is located in the highly flexible C-terminal loop. We surmise that the conformational sampling of HIS77 and HIS90 will be hindered compared to that of HIS105,

However, ASP33, which is also located at the N-terminus of a helix, was found to be acidified in the NMR experiments, yet the simulation fails to predict this shift. Similar coherence is found for the experimentally determined acidification of GLU81 and ASP82, both close to the N-terminal helix dipole. Also, here, the simulation fails to predict the acidification of both residues, as can clearly be seen from Figures 4 and 8. Since the treatment of polarization is a known weakness of classical force fields, we expect to be prone to entirely miss its effects with the applied simulation setup.^{69,72} We further illustrate the impact of the helix dipole by defining distinct observables representing the dipolar character of the helices, i.e., the smallest distances of the titrated group to the closest N-terminal helix end and the angle between the titrated group and the helix axis. Figure 8D depicts the striking concurrence of acidification in all residues that show small distances as well as small angles, yet only for the NMR pK_a values. As anticipated, however, this trend is not reproduced for pK_a values predicted from the cpH-aMD simulations. We surmise that the incorporation of polarizable force fields might allow for proper treatment of this effect.

However, there are numerous other and maybe more severe shortcomings we face with our simulation approach. First and foremost, the utilization of implicit solvent models at the constant pH step, coupled with the necessary removal of water molecules and ions, excludes the incorporation of ionic stabilizations and thereby distorts the predicted pK_a values, as already discussed above. This limitation might be overcome with the use of continuous cpH techniques, as they allow a fully explicit solvent approach, including ions.

Furthermore, as already stated above, the pK_a values measured in experiments correspond to a much higher timescale compared to our simulations, which, in turn, could mean that the conformational ensemble behind the measured pK_a values could be much broader and diverse than the simulated one.

Nevertheless, despite the discussed shortcomings, the overall agreement in predicted and experimentally measured protonation state probabilities is strongly convincing for the Phl p 6 pollen allergen.

CONCLUSIONS

The combination of structural as well as dynamical information from the cpH-aMD simulations and NMR experiments allowed us to unveil the sources of the observed perturbations in the protonation state ensembles. In particular, we identify formation of ion pairs to cause the strongest acidifications. Persistent hydrogen bonds between charged residues as well as the N-terminal dipole moment of α -helices lead to weaker but still notable perturbations. However, the effect of the helix dipole on the pK_a values was only captured in our NMR experiment and not in our simulations. We surmise that this is due to the use of classical force fields and might be resolved with the use of polarizable force fields. Our results concerning the tautomer distribution of the three histidines in the system underlines the importance of conformational sampling to obtain a reasonable protonation state distribution.

The reliable modeling of protonation state probabilities at varying pH levels is specifically crucial for allergen proteins and will be essential for further studies on their dynamics and endolysosomal degradation mechanism. However, the presented processing and interpretation of NMR and cpH-aMD data is designed to find a broad applicability.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00540.

Boosting parameters; fitting equations for obtaining titration curves; residue-wise pK_a values; pK_a convergence analysis; H-bond analysis for histidines; transition analysis for pH values (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: Klaus.Liedl@uibk.ac.at.

ORCID

Martin Tollinger: 0000-0002-2177-983X

Klaus R. Liedl: 0000-0002-0985-2299

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Austrian Science Fund (FWF) via the grant P30737 "Protein Dynamics and Proteolytic Susceptibility". The authors acknowledge support by the Austrian Research Promotion Agency FFG (project 858017—West-Austrian BioNMR).

REFERENCES

- (1) Cornish-Bowden, A. J.; Knowles, J. The pH-dependence of pepsin-catalysed reactions. *Biochem. J.* **1969**, *113*, 353–362.
- (2) White, F. H., Jr.; Anfinsen, C. B. Some relationships of structure to function in ribonuclease. *Ann. N. Y. Acad. Sci.* **1959**, *81*, 515–523.
- (3) Garcia-Moreno, B. Adaptations of proteins to cellular and subcellular pH. *J. Biol.* **2009**, *8*, 98.
- (4) Tanford, C.; Kirkwood, J. G. Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. *J. Am. Chem. Soc.* **1957**, *79*, 5333–5339.
- (5) Perutz, M. Electrostatic effects in proteins. *Science* **1978**, *201*, 1187–1191.
- (6) Gunner, M. R.; Mao, J.; Song, Y.; Kim, J. Factors influencing the energetics of electron and proton transfers in proteins. What can be learned from calculations. *Biochim. Biophys. Acta Bioenerg.* **2006**, *1757*, 942–968.
- (7) Harris, T. K.; Turner, G. J. Structural Basis of Perturbed pK_a Values of Catalytic Groups in Enzyme Active Sites. *IUBMB Life* **2002**, *53*, 85–98.
- (8) Platzer, G.; Okon, M.; McIntosh, L. P. pH-dependent random coil 1H, 13C, and 15N chemical shifts of the ionizable amino acids: a guide for protein pK_a measurements. *J. Biomol. NMR* **2014**, *60*, 109–129.
- (9) Chen, W.; Morrow, B. H.; Shi, C.; Shen, J. K. Recent development and application of constant pH molecular dynamics. *Mol. Simul.* **2014**, *40*, 830–838.
- (10) Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964.
- (11) Keller, B. G.; Prinz, J.-H.; Noé, F. Markov models and dynamical fingerprints: Unraveling the complexity of molecular kinetics. *Chem. Phys.* **2012**, *396*, 92–107.
- (12) Chodera, J. D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135–144.
- (13) Tanford, C. Ionization-linked Changes in Protein Conformation. I. Theory. *J. Am. Chem. Soc.* **1961**, *83*, 1628–1634.
- (14) Di Russo, N. V.; Estrin, D. A.; Marti, M. A.; Roitberg, A. E. pH-Dependent Conformational Changes in Proteins and Their Effect on Experimental pK_a s: The Case of Nitrophorin 4. *PLoS Comput. Biol.* **2012**, *8*, No. e1002761.

- (15) McIntosh, L. P.; Naito, D.; Baturin, S. J.; Okon, M.; Joshi, M. D.; Nielsen, J. E. Dissecting electrostatic interactions in *Bacillus circulans* xylanase through NMR-monitored pH titrations. *J. Biomol. NMR* **2011**, *51*, 5.
- (16) Fenwick, R. B.; Esteban-Martín, S.; Salvatella, X. Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur. Biophys. J.* **2011**, *40*, 1339–1355.
- (17) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, No. 71.
- (18) Alexov, E.; Mehler, E. L.; Baker, N. M.; Baptista, A.; Huang, Y.; Milletti, F.; Erik Nielsen, J.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. Progress in the prediction of pKa values in proteins. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3260–3275.
- (19) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (20) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40*, W537–W541.
- (21) Baptista, A. M.; Martel, P. J.; Petersen, S. B. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. *Proteins: Struct., Funct., Bioinf.* **1997**, *27*, 523–544.
- (22) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **2002**, *117*, 4184–4200.
- (23) Machuqueiro, M.; Baptista, A. M. Constant-pH Molecular Dynamics with Ionic Strength Effects: Protonation–Conformation Coupling in Decalysine. *J. Phys. Chem. B* **2006**, *110*, 2927–2933.
- (24) Bürgi, R.; Kollman, P. A.; van Gunsteren, W. F. Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 469–480.
- (25) Börjesson, U.; Hünenberger, P. H. Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. *J. Chem. Phys.* **2001**, *114*, 9706–9719.
- (26) Börjesson, U.; Hünenberger, P. H. 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.
- (27) Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **2004**, *25*, 2038–2048.
- (28) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III Constant-pH molecular dynamics using continuous titration coordinates. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 738–752.
- (29) Khandogin, J.; Brooks, C. L. Constant pH Molecular Dynamics with Proton Tautomerism. *Biophys. J.* **2005**, *89*, 141–157.
- (30) Goh, G. B.; Knight, J. L.; Brooks, C. L. Constant pH Molecular Dynamics Simulations of Nucleic Acids in Explicit Solvent. *J. Chem. Theory Comput.* **2012**, *8*, 36–46.
- (31) Swails, J. M.; Roitberg, A. E. Enhancing Conformation and Protonation State Sampling of Hen Egg White Lysozyme Using pH Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 4393–4404.
- (32) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10*, 1341–1352.
- (33) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks, C. L., III Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 1319–1331.
- (34) Huang, Y.; Harris, R. C.; Shen, J. Generalized Born Based Continuous Constant pH Molecular Dynamics in Amber: Implementation, Benchmarking and Analysis. *J. Chem. Inf. Model.* **2018**, *58*, 1372–1383.
- (35) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caffisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (36) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (37) Bashford, D.; Case, D. A. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
- (38) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmüller, H. Constant pH Molecular Dynamics in Explicit Solvent with λ -Dynamics. *J. Chem. Theory Comput.* **2011**, *7*, 1962–1978.
- (39) Case, D.; Cerutti, D.; Cheatham, T., III; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Greene, D.; Homeyer, N. AMBER; University of California: San Francisco, 2017.
- (40) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (41) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929.
- (42) Itoh, S. G.; Damjanović, A.; Brooks, B. R. pH replica-exchange method based on discrete protonation states. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3420–3436.
- (43) Williams, S. L.; de Oliveira, C. A. F.; McCammon, J. A. Coupling Constant pH Molecular Dynamics with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2010**, *6*, 560–568.
- (44) Vrtala, S.; Fischer, S.; Grote, M.; Vangelista, L.; Pastore, A.; Sperr, W. R.; Valent, P.; Reichelt, R.; Kraft, D.; Valenta, R. Molecular, Immunological, and Structural Characterization of Phl p 6, a Major Allergen and P-Particle-Associated Protein from Timothy Grass (*Phleum pratense*) Pollen. *J. Immunol.* **1999**, *163*, 5489–5496.
- (45) Vrtala, S.; Focke, M.; Kopec, J.; Verdino, P.; Hartl, A.; Sperr, W. R.; Fedorov, A. A.; Ball, T.; Almo, S.; Valent, P.; Thalhamer, J.; Keller, W.; Valenta, R. Genetic Engineering of the Major Timothy Grass Pollen Allergen, Phl p 6, to Reduce Allergenic Activity and Preserve Immunogenicity. *J. Immunol.* **2007**, *179*, 1730–1739.
- (46) Machado, Y.; Freier, R.; Scheiblhofer, S.; Thalhamer, T.; Mayr, M.; Briza, P.; Grutsch, S.; Ahammer, L.; Fuchs, J. E.; Wallnoefer, H. G.; Isakovic, A.; Kohlbauer, V.; Hinterholzer, A.; Steiner, M.; Danzer, M.; Horejs-Hoecck, J.; Ferreira, F.; Liedl, K. R.; Tollinger, M.; Lackner, P.; Johnson, C. M.; Brandstetter, H.; Thalhamer, J.; Weiss, R. Fold stability during endolysosomal acidification is a key factor for allergenicity and immunogenicity of the major birch pollen allergen. *J. Allergy Clin. Immunol.* **2016**, *137*, 1525–1534.
- (47) Freier, R.; Dall, E.; Brandstetter, H. Protease recognition sites in Bet v 1a are cryptic, explaining its slow processing relevant to its allergenicity. *Sci. Rep.* **2015**, *5*, No. 12707.
- (48) Watts, C. The exogenous pathway for antigen presentation on major histocompatibility complex class II and CD1 molecules. *Nat. Immunol.* **2004**, *5*, 685–692.
- (49) Scheiblhofer, S.; Laimer, J.; Machado, Y.; Weiss, R.; Thalhamer, J. Influence of protein fold stability on immunogenicity and its implications for vaccine design. *Expert Rev. Vaccines* **2017**, *16*, 479–489.
- (50) Compeer, E. B.; Flinsenberger, T. W. H.; van der Grein, S. G.; Boes, M. Antigen Processing and Remodeling of the Endosomal Pathway: Requirements for Antigen Cross-Presentation. *Front. Immunol.* **2012**, *3*, No. 37.
- (51) Fedorov, A.; Ball, T.; Fedorov, E.; Vrtala, S.; Valenta, R.; Almo, S. Crystal structure of Phl p 6, a major timothy grass pollen allergen, co-crystallized with Zn.

(52) Chemical Computing Group. Molecular Operating Environment (MOE); 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2017.

(53) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950–1958.

(54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(55) Wallnoefer, H. G.; Handschuh, S.; Liedl, K. R.; Fox, T. Stabilizing of a Globular Protein by a Highly Complex Water Network: A Molecular Dynamics Simulation Study on Factor Xa. *J. Phys. Chem. B* **2010**, *114*, 7405–7412.

(56) Case, D. A.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER; University of California: San Francisco, 2017.

(57) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(58) Adelman, S. A.; Doll, J. D. Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *J. Chem. Phys.* **1976**, *64*, 2375–2388.

(59) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(60) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(61) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.

(62) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(63) The PyMOL Molecular Graphics System, version 1.8.6.0; Schrodinger LLC, 2017.

(64) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10*, 2677–2689.

(65) Tollinger, M.; Forman-Kay, J. D.; Kay, L. E. Measurement of Side-Chain Carboxyl pKa Values of Glutamate and Aspartate Residues in an Unfolded Protein by Multinuclear NMR Spectroscopy. *J. Am. Chem. Soc.* **2002**, *124*, 5714–5717.

(66) Yamazaki, T.; Forman-Kay, J. D.; Kay, L. E. Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. *J. Am. Chem. Soc.* **1993**, *115*, 11054–11055.

(67) Baryshnikova, O. K.; Williams, T. C.; Sykes, B. D. Internal pH indicators for biomolecular NMR. *J. Biomol. NMR* **2008**, *41*, 5–7.

(68) Pelton, J. G.; Torchia, D. A.; Meadow, N. D.; Roseman, S. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIIGlc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. *Protein Sci.* **1993**, *2*, 543–558.

(69) Williams, S. L.; Blachly, P. G.; McCammon, J. A. Measuring the successes and deficiencies of constant pH molecular dynamics: A blind prediction study. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3381–3388.

(70) Olsson, M. H. M. Protein electrostatics and pKa blind predictions; contribution from empirical predictions of internal ionizable residues. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3333–3345.

(71) Lodi, P. J.; Knowles, J. R. Direct evidence for the exploitation of an .alpha.-helix in the catalytic mechanism of triosephosphate isomerase. *Biochemistry* **1993**, *32*, 4338–4343.

(72) Kramer, C.; Spinn, A.; Liedl, K. R. Charge Anisotropy: Where Atomic Multipoles Matter Most. *J. Chem. Theory Comput.* **2014**, *10*, 4488–4496.