Published in final edited form as:

J Phys Chem B. 2017 April 20; 121(15): 3318–3330. doi:10.1021/acs.jpcb.6b08249.

Origin of pK_a Shifts of Internal Lysine Residues in SNase Studied Via Equal-Molar VMMS Simulations in Explicit Water

Xiongwu Wu*, Juyong Lee, and Bernard R. Brooks

Laboratory of Computational Biology, National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH), Bethesda, Maryland 20892, United States

Abstract

Protein internal ionizable groups can exhibit large shifts in pK_a values. Although the environment and interaction changes have been extensively studied both experimentally and computationally, direct calculation of pK_a values of these internal ionizable groups in explicit water is challenging due to energy barriers in solvent interaction and in conformational transition. The virtual mixture of multiple states (VMMS) method is a new approach designed to study chemical state equilibrium. This method constructs a virtual mixture of multiple chemical states in order to sample the conformational space of all states simultaneously and to avoid crossing energy barriers related to state transition. By applying VMMS to 25 variants of staphylococcal nuclease with lysine residues at internal positions, we obtained the pK_a values of these lysine residues and investigated the physics underlining the pK_a shifts. Our calculation results agree reasonably well with experimental measurements, validating the VMMS method for pK_a calculation and providing molecular details of the protonation equilibrium for protein internal ionizable groups. Based on our analyses of protein conformation relaxation, lysine side chain flexibility, water penetration, and the micro-environment, we conclude that the hydrophobicity of the microenvironment around the lysine side chain (which affects water penetration differently for different protonation states) plays an important role in the pK_a shifts.

INTRODUCTION

Protonation states of ionizable groups play critical roles in the structures and functions of biomolecules. During biomolecular function cycles, ionizable groups that experience microenvironment change may adopt different protonation states. While proteins tend to bury hydrophobic residues and expose those that are hydrophilic in general, some ionizable groups are buried inside proteins to achieve biological functions, such as enzyme catalysis, proton transport, and e-transfer, etc.^{1–5} These internal ionizable groups are often keys to understanding protein functions, and therefore, attract the attention of many studies.

pK_a is a property that governs the protonation states of ionizable groups. It is measured experimentally,^{6–10} calculated theoretically,^{11–43} and is an important observable property that bridges these experimental and theoretical studies to provide an understanding of

*Corresponding Author: wuxw@nhlbi.nih.gov; Tel: 301-451-6251; Fax: 301-480-6494.

The authors declare no competing financial interest.

ionizable groups in biological systems. Computational approaches to calculate pK_a serve three main purposes: (1) predicting the pK_a of ionizable groups such that their protonation states at experimental conditions can be predicted,^{15,18,21,25,26,29,44,45} (2) revealing the physics behind pK_a shifts at atomic levels,^{46–49} and (3) accessing protonation state dependent conformational transitions like protein folding and functioning.^{33,36,38–40,50–63} A series of computational methods have been developed for pK_a calculation.^{12–14,16,18–21,23–26,28,30–32,34–38,40,41,59, 64–70} For buried ionizable groups, it has been realized that explicit water molecules are needed to accurately describe the microenvironment;^{23,48,71,72} this presents challenges for most of the pK_a calculation methods.

Previously, variants of staphylococcal nuclease (SNase) had been studied via simulations.^{49,71,73} These simulations were performed at a single protonation state, either protonated or deprotonated, to examine the patterns of conformational reorganization triggered by the ionization of internal groups in the protein interior. Circular dichroism, steady-state tryptophan fluorescence, and nuclear magnetic resonance spectroscopy have shown that the variants of SNase with internal ionizable side chains exhibit local reorganization or global unfolding when the internal ionizable groups are charged.^{74–77} It was found that for most variants, the RMSD values of the neutral side chain were smaller than those of charged side chains, suggesting that the crystal structures have neutral side chains. Damjanovic et al. have shown that protonation and deprotonation of internal groups frequently trigger the penetration of water into the protein interior.^{49,71} Their studies of pK_a shifts of internal ionizable groups have been focused on conformational relaxation, exposing charged side chains, and water penetration. Significant correlations between the pK_a shifts and several structural parameters were found: polarity change of the microenvironment, hydration changes of the ionizable groups, and RMSD change of the ionizable groups.^{49,71,73} However, since these simulations were unable to calculate pK_a values, it is a question how well the simulations reproduce the behavior of these SNase variants. Most pK_a calculations for SNase variants were performed with implicit solvation models,^{21,25,46,66,78,79} which may miss the role of interior water molecules in the protein. There are many degrees of freedom in designing a simulation, such as protonation states of all ionizable residues in the protein, ion types and concentration, water model type and density, protein conformation, force fields, etc. These degrees of freedom may result in different observations from simulations. Thus, it is desired to conduct simulations that can reproduce experimentally observed pK_a shifts in order to provide convincing molecular models needed to decipher the origin of pK_a shifts.

A major difficulty in pK_a calculations with explicit water molecules is the reorientation of water upon changes in the protonation state. With explicit water molecules, different charge states are separated by large energy barriers related to the re-orientation of interacting water molecules, making any thorough conformational search a profound challenge. For internal ionizable groups, pK_a calculations suffer additional difficulties. In addition to the solvent reorientation energy barriers, the slow solvent penetration process further limits the convergence of conformational sampling. The environment of the protein interior is highly heterogeneous and creates local minimums trapping any conformational sampling. All these difficulties make pK_a calculations for interior residues in SNase variants tough tasks. Many

efforts have been dedicated to address these difficulties. For example, Ghosh and Cui⁸⁰ calculated the pK_a of residue 66 in two mutants (V66D and V66E) of SNase using a combined quantum mechanical/molecular mechanical (QM/MM) potential function and found that partial unfolding of a nearby β -sheet region is critical for the pK_a calculation of Asp66. To tackle this notoriously challenging problem of accurately predicting the pK_a value of a buried titratable residue, Zheng et al. developed an efficient free-energy calculation algorithm called orthogonal space random walk and successfully calculated the pK_a value of Asp66 in the SNase mutant with this method.⁸¹

The virtual mixture of multiple states (VMMS) method¹² is designed to directly simulate chemical state equilibrium in explicit water and can be applied to calculate pK_a values for these difficult systems. This method constructs a virtual mixture of multiple chemical states so as to sample the conformational space of all chemical states simultaneously. The VMMS system consists of multiple subsystems, one for each state. Each subsystem contains a solute of a given chemical state and a solvent environment. The solute molecules in all subsystems share the same conformation but have their own solvent environments. Transitions between states are implicated by changes of their molar fractions so that no solvent reorientation is needed. This design makes it convenient to estimate free energy differences between different chemical states so that pK_a can be calculated. This work applied the VMMS method to calculate the pK_a values of the lysine amine group in SNase variants and investigate the origin of pK_a shifts.

To examine the behavior of internal ionizable groups systematically, a family of SNase variants with mutations of ionizable side chain residues at various positions were studied.^{8,9,48,74,75,78} These variants provide an excellent benchmark with which to examine the performance of pK_a calculation methods. This work applied VMMS to 25 lysine mutation variants to demonstrate the application of VMMS in these difficult systems and to provide insights into the origin of pK_a shifts of interior ionizable groups.

METHODS

The details of the VMMS method are described thoroughly elsewhere.¹² In this study we applied a special case of the VMMS method: the equal-molar VMMS simulation for the pK_a calculation of a single ionizable group. Here, we describe the details of the equal-molar VMMS simulation method.

Equal-Molar VMMS System.

The equal-molar VMMS system is designed to simulate the protonation equilibrium of a solute with a single ionizable group at $pH = pK_a$ in an implicit or explicit solvent environment. Let us consider a solute, e.g., the SNase variant V66K, in solution as shown in Figure 1. The solution (shown at the top) contains a solute whose K66 is in both the protonated (H) and deprotonated (D) states. At $pH = pK_a$, both protonation states of V66K coexist in the solution at equilibrated concentrations, $x_H = x_D = 0.5$. Here, the relative concentrations are described as the state molar fractions, which are calculated with the solute only. At the low concentration limit, the solute molecules in different chemical states do not interact with each other. Therefore, the solute at each state can be considered as a

solvated molecule.⁸² In other words, every solute is solvated by its own solvent, and solvent molecules only interact with the solute of one state. This design avoids solvent reorientation upon state transitions and allows the solute to undergo a conformational search with an equilibrated solvation. The solvent can be either implicit or explicit and is always in a state-equilibrated condition. In this work, we chose explicit water to solvate the solute, because explicit water provides more accurate solvation than an implicit solvation model.^{48,71}

The equal-molar VMMS system has two subsystems, one for each state including the environment (solvent) around it. While the solute at each state has its own solvent environment, the solute molecules at both states share the same conformation such that they search the conformational space together. In other words, each subsystem is a solute—solvent simulation box containing a solute surrounded by its implicit/explicit solvent environment. While the solute molecules in all subsystems are constrained to have the same conformation and move in the same way, the solvent molecules are not constrained and move in their own interaction environment. The solvents sample their own conformational space to provide equilibrated solvation to the solute. This design makes it very convenient to calculate the relative free energies of the states using Bennett's acceptance ratio (BAR) method.⁸³ Let us imagine that the two states form an ideal solution and that each state has its own population defined as state molar fractions, $x_H = x_D = 0.5$. At equilibrium, all states have the same chemical potential, which determines the equilibrium molar fractions of all states. Because these states are not physically mixed to form a real solution as in mixture simulations,⁸⁴ the system is called virtual mixture of multiple states (VMMS).

In summary, the equal molar VMMS system contains one solute distributed in 2 states, $S(x_H = x_D = 0.5)$, and 2 solvents, W_H and W_D , which provide solvation to the 2 states. Solvent W_H only interacts with the solute of state H, and solvent W_D only interacts with the solute of state D. The equal-molar VMMS system is denoted as $S(x_H = x_D = 0.5) + W_H + W_D$. We define the VMMS state H or D, denoted as V_H or V_D , as the solute at the mixed states, $S(x_H = x_D = 0.5)$, in solvent W_H or W_D , respectively. The difference between a pure state H or D and a VMMS state V_H or V_D resides in the solute only. In a pure state, the solute exists only in one chemical state, i.e., $S(x_H = 1, x_D = 0)$ for the pure H state and $S(x_H = 0, x_D = 1)$ for the pure D state, while in the VMMS state, V_H or V_D , the solute exists in both chemical states, i.e., $S(x_H = x_D = 0.5)$. At $x_H = 1$ and $x_D = 0$, the V_H state is the same as the H state and at $x_H = 0$ and $x_D = 1$, the V_D state is the same as the D state.

Conformational Distribution in the Equal-Molar VMMS Simulation.

The energy of a subsystem is the sum of the solute energy and the solvent energy:

$$E^{(H)} = E_S^{(H)} + E_W^{(H)} \quad (1a)$$

$$E^{(D)} = E_S^{(D)} + E_W^{(D)} \quad (1b)$$

where the superscripts denote the states and subscripts denote substances. The solute interaction is a sum of all pairwise interactions involving the solute, including solute—solute and solute-solvent interactions:

$$E_S^{(j)} = \sum_{a \in S} E_a^{(j)} = \frac{1}{2} \sum_{a \in S} \sum_i^{N^{(j)}} \epsilon_{ai}^{(j)} \quad j = H \text{ or } D \quad (2)$$

Here, $N^{(j)}$ is the number of all atoms, $E_a^{(j)}$ is the total interaction energy of atom a , and $\epsilon_{ai}^{(j)}$ is the interaction between atom a and atom i in subsystem j . Similarly, the solvent interaction is a sum of all pairwise interactions involving the solvent, including solvent—solvent and solute—solvent interactions:

$$E_W^{(j)} = \sum_{a \in W_j} E_a^{(j)} = \frac{1}{2} \sum_{a \in W_j} \sum_i^{N^{(j)}} \epsilon_{ai}^{(j)} \quad j = H \text{ or } D \quad (3)$$

For an equal-molar VMMS system, $S(x_H = x_D = 0.5) + W_H + W_D$, there are 2 subsystems, one for each state. The equal-molar VMMS system has the solvents for H and D subsystems, but only 1 solute, distributed in the two states equally. Therefore, the equal-molar VMMS potential energy has the following form:

$$E^{(em)} = 0.5(E_S^{(H)} + E_S^{(D)}) + E_W^{(H)} + E_W^{(D)} \quad (4)$$

Here, we use superscript “em” to denote quantities of the equal-molar VMMS system. The partition function of the equal-molar VMMS ensemble is

$$\begin{aligned} Q^{(em)} &= \sum_{\Omega^{(em)}} \exp(-\beta E^{(em)}) \\ &= \sum_{\Omega^{(em)}} \exp\left[-\beta\left(0.5(E_S^{(H)} + E_S^{(D)}) + E_W^{(H)} + E_W^{(D)}\right)\right] \end{aligned} \quad (5)$$

where $\Omega^{(em)}$ represents the conformational space of the VMMS system. The quantity, $\beta = \frac{1}{kT}$, depends on the Boltzmann constant k and temperature T .

The ensemble average of any conformational related property, P , in a pure state can be reweighted from an equal-molar VMMS simulation by

$$\langle P \rangle^{(H)} = \frac{\langle P \exp[-0.5\beta(E_S^{(H)} - E_S^{(D)})] \rangle^{(V_H)}}{\langle \exp[-0.5\beta(E_S^{(H)} - E_S^{(D)})] \rangle^{(V_H)}} \quad (6a)$$

$$\langle P \rangle^{(D)} = \frac{\langle P \exp[-0.5\beta(E_S^{(D)} - E_S^{(H)})] \rangle^{(V_D)}}{\langle \exp[-0.5\beta(E_S^{(D)} - E_S^{(H)})] \rangle^{(V_D)}} \quad (6b)$$

Here, the symbol, $\langle \dots \rangle^{(V_H)}$ or $\langle \dots \rangle^{(V_D)}$, represent a VMMS ensemble average based on the subsystem, V_H or V_D , respectively. The reweighting in VMMS simulations as shown in eq 6 needs solute energies, which can be efficiently estimated from pairwise interactions by eq 2. The isotropic periodic sum (IPS) method provides a convenient way to obtain pairwise long-range interactions accurately.⁸⁵⁻⁹¹ We have validated the IPS accuracy through free energy calculation of model compounds¹² and found that IPS and the particle mesh Ewald (PME) method produced consistent free energies.^{11,12} The use of IPS can avoid some of the problems encountered with Ewald methods, such as orientation dependent interactions with lattice images and poor scalability for massive parallel computing. Most importantly for the VMMS method, IPS allows the partition of energy components in a manner that is not possible with PME, unless a separate PME calculation is performed for each state and for solute and solvent separately.

Free Energy Differences between States.

The VMMS simulation samples conformations at all chemical states, providing sufficient information to calculate free energy differences between all states using BAR method.⁸³ Because the solute is not in the pure states during a VMMS simulation, reweighting is needed to obtain the conformational distributions at the pure states. After reweighting, the free energy difference between state H and state D has the following form:

$$\begin{aligned} \Delta G_{HD} = C_{HD} \\ - kT \ln \frac{\langle f[\beta(E^{(D)} - E^{(H)} - C_{HD})] \exp[-0.5\beta(E_S^{(H)} - E_S^{(D)})] \rangle^{(V_H)} \langle \exp[-0.5\beta(E_S^{(D)} - E_S^{(H)})] \rangle^{(V_D)}}{\langle f[\beta(E^{(H)} - E^{(D)} + C_{HD})] \exp[-0.5\beta(E_S^{(D)} - E_S^{(H)})] \rangle^{(V_D)} \langle \exp[-0.5\beta(E_S^{(H)} - E_S^{(D)})] \rangle^{(V_H)}} \end{aligned} \quad (7)$$

whereas the Fermi function is defined as

$$f(x) = \frac{1}{1 + \exp(x)} \quad (8)$$

This work uses self-guided Langevin dynamics via the generalized Langevin equation (SGLD-GLE)⁹² to enhance the conformational search. Because SGLD-GLE does not alter conformational distribution relative to MD, eq 7 can be used directly to calculate relative free energies. In eq 7, there is a constant C_{HD} , which is set to the local average free energy difference calculated in the following way:

$$\begin{aligned} C_{HD} &= \Delta \tilde{G}_{HD}(t) \\ &= \left(1 - \frac{1}{L}\right) \Delta \tilde{G}_{HD}(t - \Delta t) + \frac{1}{L} \Delta G_{HD}(t) \end{aligned} \quad (9)$$

where t is the time period to evaluate the ensemble averages in eq 7 and L is the free energy local averaging size.

Calculation of pK_a from Equal-Molar VMMS Simulations.

In VMMS, all states form a virtual solution at equilibrium molar fractions. The chemical potential of each state depends on its molar fraction:

$$\mu_i = \frac{\partial G}{\partial n_i} = \mu_i^0 + kT \ln x_i \quad (10)$$

where μ_i^0 is the standard chemical potential of state i . At equilibrium, the chemical potential of the protonated and deprotonated states equal: $\mu_H = \mu_D$, giving us:

$$\frac{x_D}{x_H} = \exp\left(-\frac{\mu_D^0 - \mu_H^0}{kT}\right) = \exp\left(-\frac{\Delta\mu_{HD}^0}{kT}\right) \quad (11)$$

For protonation equilibrium, the standard chemical potential difference contains many contributions such as those from classic electrostatic interactions and those from quantum mechanics of protonation. All contributions other than those from classic electrostatic interactions are assumed to be constant for a given titration site and can be estimated from the equilibrium properties of model compounds. More details can be found elsewhere.^{31,32,34–36,38} At a given pH, the standard chemical potential difference of deprotonation can be calculated from the state free energy difference:

$$\Delta\mu_{HD}^0 = \Delta G_{HD} - kT(\text{pH} - \text{p}K_a^{\text{ref}}) \ln 10 - \Delta G_{HD}^{\text{ref}} \quad (12)$$

Here, pK_a^{ref} is the experimental pK_a of a model compound for the corresponding amino acid and $\Delta G_{\text{HD}}^{\text{ref}}$ is the free energy difference between the protonated state (H) and the deprotonated state (D) calculated from the simulation of the model compound. G_{HD} is the free energy difference in the studied system.

For equal-molar VMMS simulations where $\Delta\mu_{\text{HD}} = \Delta\mu_{\text{HD}}^0 = 0$ and $pK_a = \text{pH}$, if using eq 12, pK_a can be directly calculated from G_{HD} :

$$pK_a = pK_a^{\text{ref}} + \frac{\Delta G_{\text{HD}}(x_{\text{H}} = x_{\text{D}}) - \Delta G_{\text{HD}}^{\text{ref}}}{kT \ln 10} \quad (13)$$

SIMULATION DETAILS

Equal-Molar VMMS Systems.

First we need the reference free energy difference ($\Delta G_{\text{HD}}^{\text{ref}}$) of lysine to calculate the pK_a values of lysine in the SNase variants. Our model compound was a lysine residue with an acetyl group (ACE) at the N-terminal and a methyl amide group (NME) at the C-terminal: ACE-lysine-NME. The reference system was built by dissolving the model compound into a cubic water box with 990 TIP3P⁹³ water molecules. The box side was 31.1 Å.

For the 25 SNase variants, their initial conformations were taken from PDB database if available. For those variants without crystal structures, we mutated the -PHS SNase structure (PDB ID: 3bdc) to generate their initial conformation. The VMMS systems were constructed by dissolving the initial conformations into a box of TIP3P water together with a number of ions to produce a neutral system in a cubic box with a side of 46.655 Å. The number of water molecules and ions in protonated and deprotonated states are listed in Table 1.

The deprotonated state of lysine has a dummy hydrogen so that the SNase variants in both states have the same number of atoms. The dummy hydrogen atom is identical to a hydrogen atom except that it has no charge. Figure 2 shows the atomic charges of lysine in their protonated and deprotonated states.

Equal-molar VMMS Simulations.

An equal-molar VMMS system contains two subsystems, one for the protonated state and one for the deprotonated state. At every time step interaction forces at each subsystem are calculated independently. The solvent forces at each state are used to drive the motion of the solvent atoms. The solute forces from both states are averaged (50% from each state) in order to produce combined forces that drive the motion of the solute atoms. Therefore, solvent atoms in different subsystems experience different forces and sample their conformations differently, whereas solute atoms in all subsystems experience the same combined forces and sample their conformational space the same way. At every specific interval (10 fs in this work), the energy changes of the solute in transition from the current

state to another state were calculated in order to evaluate the quantities in eq 7: $f[\beta(E^{(D)} - E^{(H)} - C_{HD})]$, $f[\beta(E^{(H)} - E^{(D)} + C_{HD})]$, $\exp[-0.5\beta(E_S^{(D)} - E_S^{(H)})]$, and $\exp[-0.5\beta(E_S^{(H)} - E_S^{(D)})]$.

Their averages were calculated every 1 ps to determine free energy differences between states according to eq 7. The free energy differences were locally averaged according to eq 9 with a local average size of $L = 100$ (corresponding an average time of 100 ps). These local average free energy differences were used as C_{HD} in eq 7 for the following calculations of free energy differences. For equal-molar simulations performed in this work, molar fractions were fixed at $x_H = x_D = 0.5$ and pK_a was calculated directly from the local average free energy difference according to eq 13.

All simulations presented here were performed with a modified version 39 of CHARMM^{94,95} with the VMMS method implemented. The all-atom CHARMM36 force field⁹⁶ was used for energy calculation. Unless noted otherwise, all simulations were performed in a constant volume and a constant temperature of 300 K using the SGLD-GLE method⁹² with a local averaging time $t_L = 0.2$ ps, a guiding factor $\lambda = 1$, and a friction constant $\xi = 10/\text{ps}$. A time step of 1 fs was used and the SHAKE algorithm⁹⁷ was employed to fix the hydrogen connecting bond lengths.

RESULTS AND DISCUSSIONS

For the internal ionizable groups in the SNase variants, many factors have been identified to be correlated with pK_a shifts, such as conformational relaxation,^{49,71} side chain flexibility,⁴⁹ protein interaction,⁹ water penetration,^{71,73,78} local unfolding,^{8,47} hydro-phobic environment,⁷⁴ etc. Simulating these challenging systems in explicit water maybe affected by many arbitrary factors such as initial protein structures, crystal water, ion concentration, system density, or force fields. These factors can affect simulation results and may lead to different observations. To validate the relevance of our simulations we need reproduce experimental pK_a values. With simulations reproducing experimental pK_a values, we have confidence in analyzing the physics behind the pK_a shifts.

Equal-Molar VMMS Simulations.

In order to calculate pK_a from VMMS simulations, we first calculated the reference free energy difference of the model compound, ACE-lys-NME, from an equal-molar VMMS simulation. The reference pK_a for lysine is the experiment value ($pK_a^{\text{ref}} = 10.4$) and the reference free energy change calculated from the equal-molar VMMS simulation is $\Delta G_{\text{lys}}^{\text{ref}} = 17.4 \pm 0.2$ kcal/mol.

Equal-molar VMMS simulations of 60 ns were performed for each of the 25 SNase variants in explicit water. An equal-molar VMMS simulation has only two states, protonated state (H) and deprotonated state (D) with the state molar fractions being $x_H = x_D = 0.5$. During equal molar VMMS simulations, pK_a can be calculated directly from the free energy difference according to eq 13.

It should be noted that the equal-molar VMMS simulations did not take into account the protonation states of other ionizable residues and, it was assumed that they adopted their

default protonation states at the standard conditions. This simplification allows our simulations to focus only on the lysine residues of interest. Obviously, the protonation states of other ionizable residues could influence the pK_a of the studied one. Because the number of chemical states increases exponentially with the number of ionizable groups if they are all considered explicitly,¹² to minimize the computing expense, we adopted this simplification in this study and leave a full consideration of multiple ionizable groups to future studies.

To illustrate the equal-molar VMMS simulations, we selected several variants with different pK_a shifts to show simulation profiles during our equal-molar VMMS simulations.

1. N118K.—This variant has no pK_a shift for the lysine amine group ($pK_a = 10.4$).⁷⁵ Figure 3 shows the structure of this variant and the location of K118, as well as the distributions of the amine group and water molecules. As can be seen in Figure 3, K118 is at the surface of the protein. The large distribution volume of the amine group indicates that the side chain of K118 is freely floating in water. The large water distributions indicate an easy access of the amine group by water molecules. The fact that K118 has little contact with protein implies that little pK_a shift should be expected for this lysine.

From the simulation, we examined several properties of the system, including the pK_a values, backbone and lysine side chain conformations, and interacting water molecules and polar groups. These quantities during the simulation are plotted in Figure 4. From Figure 4, we can see that the pK_a of K118 fluctuate around 10.5 with a standard deviation of 0.7. This calculated pK_a agrees well with the experiment measurement of $pK_a = 10.4$.⁷⁵ The backbone root-mean-square deviation (RMSD) from the initial conformation is 1.14 ± 0.05 Å, indicating a stable structure of this variant. The lysine side chain is very flexible with RMSD as high as 8 Å. This large RMSD is due to the fact that the lysine side chain was largely exposed to the solvent and can float around freely. To examine the interaction of the amine group, we defined the interaction number as the number of polar groups including water within 4 Å from the amine nitrogen atom. The water interaction numbers and total polar group interaction numbers in both the protonated state and the deprotonated state are shown in the upper two panels of Figure 4. The protonated state has slightly more interacting waters than the deprotonated state at 4.4 ± 1.0 and 4.0 ± 1.2 , respectively. The quick fluctuation of water numbers indicates an easy access by water. The black and red lines show the number of total interacting groups and interacting water molecules, respectively. Most of the time only red lines can be seen, indicating that all of the interacting groups are water molecules and that interaction with protein polar groups is rare.

2. V104K.—The experimental pK_a value of the K104 amine group is 7.7,⁷⁵ which is a moderate shift from free lysine. Figure 5 shows the structure of V104K, the side chain of the lysine, as well as the distributions of the lysine amine group and the water molecules at the protonated and deprotonated states. K104 is half buried in the protein. The side chain of K104 is restricted by the surrounding protein residues as evidenced by the small volume of its distribution. Water molecules are mainly distributed on one side of the amine group with the other side against the protein.

The equal-molar VMMS simulation profiles are shown in Figure 6. The average pK_a from our VMMS simulation is 9.3 with a standard deviation of 1.7. Compared with the experimental value of 7.7, our VMMS result is 1.6 higher. Such difference could be due to an insufficient simulation length as the pK_a curve tends to move lower over time. Also, the simplification of protonation states for all other ionizable groups may also contribute to this error. The RMSD of the backbone remains around 1.02 ± 0.04 Å, while the lysine side chain has a RMSD of 1.59 ± 0.12 Å, indicating that the protein is quite stable and the lysine is restricted to its initial conformation. The water interaction numbers at both states fluctuate between 2 and 3 with the deprotonated state visiting 1 more frequently. The average interaction numbers for water and for all polar groups are 2.3 ± 0.4 and 3.3 ± 0.4 at the protonated state and 2.1 ± 0.5 and 3.1 ± 0.5 at the deprotonated state. Because one side of the lysine amine group is exposed to the bulk water, we do not observe a large difference between the numbers of water molecules of the two states. The other side interacts with protein, as evidenced by the fact that the total interaction number is significantly larger than the water interaction number. Even though K104 is half buried, there are protein polar groups replacing some of interaction water. As a result, the pK_a shift is moderate.

3. L125K.—The experimental pK_a of K125 amine group is 6.2.⁷⁵ Figure 7 shows the structure of L125K and the location of K125 side chain, as well as the distributions of the lysine amino group and the surrounding water molecules. The lysine amine group is again half buried, similar to V104K discussed above. The amine group is restricted by surrounding protein residues as apparent by the small distribution volume. Water access to the amino group is easy as evidenced by the large amount of water on top of the amine group.

The equal-molar VMMS simulation profiles are shown in Figure 8. From Figure 8 we can see that pK_a initially fluctuated widely and reached equilibrium after 20 ns. The equilibrium pK_a has an average of 4.7 and a standard deviation of 1.5. The calculated pK_a is 1.5 lower than experimental value. The backbone RMSD is around 1 Å, while the side chain RMSD is around 3 Å, indicating that this variant is stable and the lysine side chain is fairly rigid. The interaction number for the protonated state is around 3, fluctuating between 2 and 4 frequently, while the interaction number for the deprotonated state is between 1 and 2, fluctuating between 0 and 3 frequently. Because we do not see many black lines in the interaction plots, the lysine amine group interacts with few protein polar groups, indicating that the burying pocket is quite hydrophobic. Compared with V104K, which is also half buried, the difference is that K125 is buried in a hydrophobic environment, while K104 is buried in a more hydrophilic pocket. Therefore, the hydrophobicity of the burying pocket may account for the larger pK_a shift of L125K than V104K. From the fluctuation frequency of the interaction numbers, we can also see that L125K shows a slower change in interaction numbers. It is reasonable to assume that the rate of water interaction number change is related to the hydrophobicity of the burying pocket.

4. I92K.—Among the 25 SNase variants, I92K has the largest experimental pK_a shift, $pK_a = 5.3$.⁷⁵ Figure 9 shows the structure of I92K, the lysine side chain, and the distributions of the amine group and the surrounding water molecules. The lysine side chain is buried inside. The lysine amine group has two distribution regions as evidenced by the two density

volumes. Water access to the amine group is through a narrow channel between the α -helix $\alpha 1$ and the β -strand $\beta 1$. After the channel, the water distribution volumes become larger because of the void space around the lysine side chain. It is also because of the large void space that the amine group of the lysine can have two distribution densities, which causes a large pK_a fluctuation during the simulation.

Figure 10 shows the VMMS simulation profiles for the SNase variant I92K. As can be seen from Figure 10, the pK_a fluctuates widely throughout the simulation period with the average value after 20 ns being 4.8, very close to the experiment value of 5.3. The standard deviation of pK_a is 2.6. The backbone RMSD remains around 1.0 Å, while the lysine side chain RMSD quickly jumps to 4 Å and remains there. The protonated state interacts with 3.2 water molecules on average, while the deprotonated state interacts with 2.1 water molecules on average. There are few protein polar groups interacting with lysine, as not many black lines can be seen in Figure 10, indicating that the burying pocket is quite hydrophobic. The change in interacting water numbers is very slow, which may relate to the hydrophobicity of the pocket.

Comparison of VMMS Results and Experimental Measurements.

From the equal molar VMMS simulations, we calculated the pK_a values of the lysine amine groups in the 25 SNase variants. The pK_a values of the 25 variants are listed in Table 1. The error ranges in the calculated pK_a values are statistical measurements of the data distribution during the simulation, which do not include the errors caused by incomplete sampling. Figure 11 shows the correlation between the VMMS results and experiment measurements.⁷⁵ Even though large deviations are observed between the VMMS and experiment results, the data distributes well around the ideal correlation, $y = x$, line. A linear fitting equation has the following form:

$$pK_a^{\text{calc}} = (0.81 \pm 0.14)pK_a^{\text{exp}} + (1.74 \pm 1.17) \quad (14)$$

This fitting has a coefficient of determination (COD) of $R = 0.77$. This reasonable agreement between our VMMS simulations and experiment measurements implies that our VMMS simulation systems can grasp the characteristics of these SNase variants, which provides us with the confidence necessary to examine the origin of the pK_a shifts through the VMMS simulations.

Origin of pK_a Shifts.

The difficulties in evaluating the pK_a of internal ionizable groups with explicit water have been a barrier to investigating the origin of the pK_a shifts. While simulating these systems is straightforward, without a validation of pK_a values, there are many factors that could lead to misrepresentation of the system. For example, is a conformational change at different protonation states the cause of a pK_a shift or is it because of a defect of a force field? Is water penetration enough to explain the pK_a shift? The reasonable reproduction of experimental pK_a values for the 25 SNase variants gave us confidence that these simulations

captured the physics of these variants and allowed us to investigate the true origin of the pK_a shifts.

In previous simulation studies, Damjanovic et al.^{49,71,73} examined several factors to identify their correlations with pK_a shifts. These parameters include conformational relaxation, side chain flexibility, and water penetration. Here, we examine these factors in our simulations.

First, let us examine the conformational relaxation. The global conformational change can be measured by the RMSD of backbone atoms. From the above four example variants, we can see that the backbone RMSDs are all small, around 1 Å. To examine its correlation with pK_a , we plotted the RMSDs of all 25 variants against their experimental pK_a values in the lower panel of Figure 12. The RMSDs range between 0.8 and 1.2 Å and show no correlation with pK_a . The small backbone RMSDs of the 25 variants confirm that the structures were quite stable during the VMMS simulation period and a structural relaxation is not necessary for their pK_a shifts. Furthermore, we examined the RMSDs of the lysine side chains and plotted them against pK_a in the middle panel of Figure 12. Again, while the lysine side chains have RMSD as high as 8 Å, little correlation is observed. The flexibility of the lysine side chain is reduced when the lysine is buried inside SNase. We can use the volume of the amino group distribution map to quantitatively measure the flexibility. The top panel of Figure 12 plots the volume against experimental pK_a values, from which we can see little correlation between them. Therefore, we believe that the lysine side chain flexibility is not a cause of pK_a shift. The fact that no major pH-sensitive conformational reorganization of the backbone of SNase was detected using NMR spectroscopy⁷² supports our simulation observation. Goh et al. calculated the dielectric constant of +PHS and the Lys-66, Asp-66, and Glu-66 mutants from first-principles using the Kirkwood-Frohlich equation and discovered that SNase has a naturally high dielectric constant ranging from 20 to 30.⁷⁸ Such high protein dielectric constants are required to reproduce the pK_a shifts of the interior ionizable groups. They concluded that conformational fluctuations and relaxations are not necessary for the pK_a shifts; this agrees with our observations.

Water penetration is widely recognized to be a cause of the pK_a shift^{8,71} and can be reflected by the change in water molecule numbers around the amine group. We define the interaction change rate by the following equation:

$$k_g = \left| \frac{dn_g}{dt} \right| \quad (15)$$

where n_g is the number of interaction groups. We calculated the interaction change rates from the 25 variant simulations and plotted them against the experimental pK_a values as shown in Figure 13. The top panel of Figure 13 shows the protein interaction changing rate, the middle panel shows the water interaction changing rate, and the bottom panel shows the total interaction changing rate. As can be seen, the protein interaction changing rates, water interaction changing rates, and the total interaction changing rates all correlate positively with the experimental pK_a values. However, the protein interaction rates and the water interaction rates are deviated much more than the total interaction rates, indicating that there

are a significant amount of changes due to exchanges between protein interactions and water interactions. The positive correlations indicate that slower interaction changing rates correspond to lower pK_a values, or larger downward pK_a shifts. Because the interaction changing rate reflects water penetration speed, we can infer that a slower water penetration is related to a larger pK_a shift. However, due to water-protein exchanges, water penetration itself only accounts for a certain part of interaction changes, as evidenced by the improved correlation with total interaction changes over water or protein interaction changes.

The correlation between the rate of interaction changes and the pK_a values indicates that the environmental changes around the lysine side chains play important roles in the pK_a shifts. To quantitatively examine the microenvironment, we define the microenvironment by all groups, including water (N_w), protein polar (N_{polar}) and nonpolar groups (N_{nonp}), within 6 Å of the lysine amine group. Here, we examined four quantities in related to the microenvironment: (1) the total number of polar groups, including water, N_{polar} ; (2) the difference of water molecules between the two protonation states, δN_w ; (3) the hydration ratio, $\rho_w = N_w/(N_{polar} + N_{nonp})$; and (4) the hydrophobic ratio, $\rho_{nonp} = N_{nonp}/(N_{polar} + N_{nonp})$.

These four quantities describe the microenvironment in different aspects. A large number of polar groups indicate a hydrophilic microenvironment. The difference in water numbers is directly related to the equilibrium between the two states. The ratio of hydration describes how deeply it is buried, with a low hydration ratio representing a deeply buried pocket. The hydrophobic ratio measures the hydrophobicity of the burying pocket. Obviously, these quantities are related to each other. Figure 14 plots these quantities against the pK_a values for the 25 SNase variants. While all these four properties correlate with experimental pK_a values, the CODs are different. The highest COD is found to be 0.76 between N_{polar} and pK_a , shown in Figure 14a, indicating that the hydrophilic degree of the microenvironment is likely responsible for the pK_a shifts.

The second highest COD is 0.74, between the hydrophobic ratio and pK_a , shown in Figure 14d. This strong correlation, combined with the stronger correlation between the number of polar groups and pK_a , supports that the hydrophobicity of the environment is likely the main cause of pK_a shifts.

The interaction difference between the two protonation states determines their free energy difference, which in turn determines the pK_a . Therefore, it is logical to consider that the difference in water numbers between the two protonation states contributes to the pK_a shifts. Figure 14b shows water number difference around the lysine amine groups against experimental pK_a values. A negative correlation is observed with COD = 0.70, indicating that the larger the difference, the smaller the pK_a value, or the larger the pK_a shift. The water number difference depends on the hydrophobicity of the microenvironment. A more hydrophobic microenvironment will make water harder to penetrate for the deprotonation (neutral) state than the protonated (charged) state, resulting in a larger water number difference.

The hydration ratio, $\rho_w = N_w/(N_{\text{polar}} + N_{\text{nonp}})$, measures how close the amine group is to the bulk water, or how deeply the amine group is buried. Figure 14c shows ρ_w against pK_a . A low hydration ratio means that the amine group of the lysine is deeply buried. Therefore, deeply buried lysine has a larger pK_a shift in general. The relatively low R value, 0.42, indicates that the hydration ratio is only a partial factor for the pK_a shift. A buried pocket could be deep like L103 K where hydration ratio is less than 0.04, but because its pocket is rich in polar groups, its pK_a shift is moderate. Therefore, the burying depth inside protein cannot completely account for the pK_a shifts.

Based on the fact that the highest COD is between N_{polar} and pK_a , we derived the following relation to estimate the pK_a values of lysine residues in SNase variants:

$$pK_a^{\text{calc}} = (0.26 \pm 0.03)N_{\text{polar}} + (1.57 \pm 0.92) \quad (16)$$

To estimate the pK_a of a lysine in a variant of SNase, we solvate the variant with explicit water and calculate N_{polar} within 6 Å of the lysine amine group, from which we can estimate pK_a using eq 16.

CONCLUSIONS

As an application of the VMMS method, we studied the 25 SNase variants with lysine mutations at various positions. The pK_a values obtained from the VMMS simulations agree reasonably well with experiment measurements. This agreement demonstrates that the VMMS method can calculate the pK_a of internal ionizable groups in explicit water.

Through these simulations, we examined the correlation between the pK_a values and protein conformation, lysine side chain flexibility, water penetration, and microenvironment. Conformational relaxation and lysine side chain flexibility were found to have little correlation with the pK_a values, indicating that they are not the main reason for the pK_a shifts. On the other hand, the interaction changing rate and the interaction numbers were found to have strong correlations with the pK_a values. These correlations point us to the hydrophobicity of the microenvironment as the main cause of the pK_a shifts. The hydrophobicity of the microenvironment affects water penetration differently at different protonation states and water penetration could be a consequence of the microenvironment. Therefore, among many factors that could affect the pK_a shifts, the hydrophobicity of the microenvironment is likely the main cause of the pK_a shifts for internal ionizable groups in SNase.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Programs of National Heart, Lung, and Blood Institute (Z01 HL001050-19). The authors thank Eunice Wu and Andrew Brooks for proofreading the manuscript.

REFERENCES

- (1). von Ballmoos C; Wiedenmann A; Dimroth P Essentials for atp synthesis by f1f0 atp synthases. Annu. Rev. Biochem 2009, 78, 649–672. [PubMed: 19489730]

- (2). Ho MC; Menetret JF; Tsuruta H; Allen KN The origin of the electrostatic perturbation in acetoacetate decarboxylase. *Nature* 2009, 459, 393–397. [PubMed: 19458715]
- (3). Luecke H; Schobert B; Richter HT; Cartailler JP; Lanyi JK Structural changes in bacteriorhodopsin during ion transport at 2 angstrom resolution. *Science* 1999, 286, 255–260. [PubMed: 10514362]
- (4). Sagnella DE; Voth GA Structure and dynamics of hydronium in the ion channel gramicidin a. *Biophys. J* 1996, 70, 2043–2051. [PubMed: 9172729]
- (5). McIntosh LP; Hand G; Johnson PE; Joshi MD; Korner M; Plesniak LA; Ziser L; Wakarchuk WW; Withers SG The pka of the general acid/base carboxyl group of a glycosidase cycles during catalysis: A ¹³c-nmr study of bacillus circulans xylanase. *Biochemistry* 1996, 35, 9958–9966. [PubMed: 8756457]
- (6). Bezencon J; Wittwer MB; Cutting B; Smiesko M; Wagner B; Kansy M; Ernst B Pka determination by (1)h nmr spectroscopy - an old methodology revisited. *J. Pharm. Biomed. Anal* 2014, 93, 147–155. [PubMed: 24462329]
- (7). McDougal OM; Granum DM; Swartz M; Rohleder C; Maupin CM Pka determination of histidine residues in alpha-conotoxin mii peptides by 1h nmr and constant ph molecular dynamics simulation. *J. Phys. Chem. B* 2013, 117, 2653–2661. [PubMed: 23336579]
- (8). Takayama Y; Castaneda CA; Chimenti M; Garcia-Moreno B; Iwahara J Direct evidence for deprotonation of a lysine side chain buried in the hydrophobic core of a protein. *J. Am. Chem. Soc* 2008, 130, 6714–6715. [PubMed: 18454523]
- (9). Harms MJ; Schlessman JL; Chimenti MS; Sue GR; Damjanovic A; Garcia-Moreno B A buried lysine that titrates with a normal pka: Role of conformational flexibility at the protein-water interface as a determinant of pka values. *Protein Sci* 2008, 17, 833–845. [PubMed: 18369193]
- (10). Joshi MD; Hedberg A; McIntosh LP Complete measurement of the pka values of the carboxyl and imidazole groups in bacillus circulans xylanase. *Protein Sci* 1997, 6, 2667–2670. [PubMed: 9416621]
- (11). Lee J; Miller BT; Brooks BR Computational scheme for ph-dependent binding free energy calculation with explicit solvent. *Protein Sci* 2016, 25, 231–243. [PubMed: 26189656]
- (12). Wu X; Brooks BR A virtual mixture approach to the study of multistate equilibrium: Application to constant ph simulation in explicit water. *PLoS Comput. Biol* 2015, 11, e1004480. [PubMed: 26506245]
- (13). Chen Y; Roux B Constant-ph hybrid nonequilibrium molecular dynamics-monte carlo simulation method. *J. Chem. Theory Comput* 2015, 11, 3919–3931. [PubMed: 26300709]
- (14). Swails JM; York DM; Roitberg AE 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. [PubMed: 24803862]
- (15). Mitsuoka K A pka calculation of residues in a proton pump, bacteriorhodopsin, from structures determined by electron crystallography. *Microscopy* 2014, 63 (1), i30–i31.
- (16). Lee J; Miller BT; Damjanovic A; Brooks BR Constant ph molecular dynamics in explicit solvent with enveloping distribution sampling and hamiltonian exchange. *J. Chem. Theory Comput* 2014, 10, 2738–2750. [PubMed: 25061443]
- (17). Goh GB; Hulbert BS; Zhou H; Brooks CL, 3rd Constant ph molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: Struct. Funct., Genet* 2014, 82, 1319–1331. [PubMed: 24375620]
- (18). Sharma I; Kaminski GA Calculating pka values for substituted phenols and hydration energies for other compounds with the first-order fuzzy-border continuum solvation model. *J. Comput. Chem* 2012, 33, 2388–2399. [PubMed: 22815192]
- (19). Kilambi KP; Gray JJ Rapid calculation of protein pka values using rosetta. *Biophys. J* 2012, 103, 587–595. [PubMed: 22947875]
- (20). Witham S; Talley K; Wang L; Zhang Z; Sarkar S; Gao D; Yang W; Alexov E Developing hybrid approaches to predict pka values of ionizable groups. *Proteins: Struct., Funct. Genet* 2011, 79, 3389–3399. [PubMed: 21744395]

- (21). Wallace JA; Wang Y; Shi C; Pastoor KJ; Nguyen BL; Xia K; Shen JK Toward accurate prediction of pka values for internal protein residues: The importance of conformational relaxation and desolvation energy. *Proteins: Struct. Funct., Genet* 2011, 79,3364–3373. [PubMed: 21748801]
- (22). Machuqueiro M; Baptista AM Is the prediction of pka values by constant-ph molecular dynamics being hindered by inherited problems? *Proteins: Struct. Funct. Genet* 2011, 79, 3437–3447. [PubMed: 22072522]
- (23). Czodrowski P Blind, one-eyed, or eagle-eyed? Pka calculations during blind predictions with staphylococcal nuclease. *Proteins: Struct., Funct. Genet* 2011, 79, 3299–3305. [PubMed: 22072517]
- (24). Carstensen T; Farrell D; Huang Y; Baker NA; Nielsen JE On the development of protein pka calculation algorithms. *Proteins: Struct. Funct. Genet* 2011, 79, 3287–3298. [PubMed: 21744393]
- (25). Arthur EJ; Yesselman JD; Brooks CL, 3rd Predicting extreme pka shifts in staphylococcal nuclease mutants with constant ph molecular dynamics. *Proteins: Struct., Funct., Genet* 2011, 79, 3276–3286. [PubMed: 22002886]
- (26). Wallace JA; Shen JK Predicting pka values with continuous constant ph molecular dynamics. *Methods Enzymol* 2009, 466, 455–475. [PubMed: 21609872]
- (27). Li L; Vorobyov I; Allen TW Potential of mean force and pka profile calculation for a lipid membrane-exposed arginine side chain. *J. Phys. Chem. B* 2008, 112, 9574–9587. [PubMed: 18636765]
- (28). Stern HA Molecular simulation with variable protonation states at constant ph. *J. Chem. Phys* 2007, 126, 164112. [PubMed: 17477594]
- (29). Lu H; Chen X; Zhan CG First-principles calculation of pka for cocaine, nicotine, neurotransmitters, and anilines in aqueous solution. *J. Phys. Chem. B* 2007, 111, 10599–10605. [PubMed: 17691837]
- (30). 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. [PubMed: 16471903]
- (31). Mongan J; Case DA Biomolecular simulations at constant ph. *Curr. Opin. Struct. Biol* 2005, 15, 157–163. [PubMed: 15837173]
- (32). Khandogin J; Brooks CL, 3rd Constant ph molecular dynamics with proton tautomerism. *Biophys. J* 2005, 89, 141–157. [PubMed: 15863480]
- (33). Ploemen JP; Kelder J; Hafmans T; van de Sandt H; van Burgsteden JA; Saleminck PJ; van Esch E Use of physicochemical calculation of pka and clogp to predict phospholipidosis-inducing potential: A case study with structurally related piperazines. *Exp. Toxicol. Pathol* 2004, 55, 347–355. [PubMed: 15088636]
- (34). Mongan J; Case DA; McCammon JA Constant ph molecular dynamics in generalized born implicit solvent. *J. Comput. Chem* 2004, 25, 2038–2048. [PubMed: 15481090]
- (35). Lee MS; Salsbury FR, Jr.; Brooks CL, 3rd Constant-ph molecular dynamics using continuous titration coordinates. *Proteins: Struct., Funct., Genet* 2004, 56, 738–752. [PubMed: 15281127]
- (36). 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. Stat. Nonlin. Soft. Matter Phys* 2004, 69, 021915. [PubMed: 14995499]
- (37). Walczak AM; Antosiewicz JM Langevin dynamics of proteins at constant ph. *Phys. Rev. E: Stat. Phys. Plasmas, Fluids, Relat. Interdiscip. Top* 2002, 66,051911.
- (38). Burgi R; Kollman PA; Van Gunsteren WF Simulating proteins at constant ph: An approach combining molecular dynamics and monte carlo simulation. *Proteins: Struct., Funct., Genet* 2002, 47, 469–480. [PubMed: 12001225]
- (39). Fukamizo T; Juffer AH; Vogel HJ; Honda Y; Tremblay H; Boucher I; Neugebauer WA; Brzezinski R Theoretical calculation of pka reveals an important role of arg205 in the activity and stability of streptomyces sp. N174 chitosanase. *J. Biol. Chem* 2000, 275, 25633–25640. [PubMed: 10829022]
- (40). Baptista AM; Martel PJ; Petersen SB Simulation of protein conformational freedom as a function of ph: Constant-ph molecular dynamics using implicit titration. *Proteins: Struct., Funct., Genet* 1997,27, 523–544. [PubMed: 9141133]

- (41). Gibas CJ; Subramaniam S Explicit solvent models in protein pka calculations. *Biophys. J* 1996, 71, 138–147. [PubMed: 8804597]
- (42). Pickard FC, IV; Konig G; Tofoleanu F; Lee J; Simmonett AC; Shao Y; Ponder JW; Brooks BR Blind prediction of distribution in the SAMPL5 challenge with QM based protomer and pKa corrections. *J. Comput.-Aided Mol. Des* 2016, DOI:10.1007/s10822-016-9955-7.
- (43). Lee J; Tofoleanu F; Pickard FC, IV; Konig G; Huang J; Damjanovic A; Baek M; Seok C; Brooks BR Absolute binding free energy calculations of cbclip host–guest systems in the sampl5 blind challenge. *J. Comput.-Aided Mol. Des* 2016, DOI: 10.1007/s10822-016-9968-2.
- (44). Williams SL; Blachly PG; McCammon JA Measuring the successes and deficiencies of constant ph molecular dynamics: A blind prediction study. *Proteins: Struct., Funct., Genet* 2011, 79, 3381–3388. [PubMed: 22072520]
- (45). Rosales T; Xu J; Wu X; Hodoscek M; Callis P; Brooks BR; Knutson JR Molecular dynamics simulations of perylene and tetracene librations: Comparison with femtosecond upconversion data. *J. Phys. Chem. A* 2008, 112, 5593–5597. [PubMed: 18507364]
- (46). Pathak AK Effect of a buried ion pair in the hydrophobic core of a protein: An insight from constant ph molecular dynamics study. *Biopolymers* 2015, 103, 148–157. [PubMed: 25363335]
- (47). Chimenti MS; Khangulov VS; Robinson AC; Heroux A; Majumdar A; Schlessman JL; García-Moreno B Structural reorganization triggered by charging of lys residues in the hydrophobic interior of a protein. *Structure* 2012, 20, 1071–1085. [PubMed: 22632835]
- (48). Harms MJ; Castaneda CA; Schlessman JL; Sue GR; Isom DG; Cannon BR; Garcia-Moreno EB The pk(a) values of acidic and basic residues buried at the same internal location in a protein are governed by different factors. *J. Mol. Biol* 2009, 389, 34–47. [PubMed: 19324049]
- (49). Damjanovic A; Wu X; Garcia-Moreno EB; Brooks BR Backbone relaxation coupled to the ionization of internal groups in proteins: A self-guided langevin dynamics study. *Biophys. J* 2008, 95, 4091–4101. [PubMed: 18641078]
- (50). Brocklehurst K; Willenbrock SJ; Salih E Effects of conformational selectivity and of overlapping kinetically influential ionizations on the characteristics of ph-dependent enzyme kinetics. Implications of free-enzyme pka variability in reactions of papain for its catalytic mechanism. *Biochem. J* 1983, 211, 701–708. [PubMed: 6309137]
- (51). Machuqueiro M; Baptista AM The ph-dependent conformational states of kyotorphin: A constant-ph molecular dynamics study. *Biophys. J* 2007, 92, 1836–1845. [PubMed: 17172294]
- (52). Morrow BH; Koenig PH; Shen JK Self-assembly and bilayer-micelle transition of fatty acids studied by replica-exchange constant ph molecular dynamics. *Langmuir* 2013, 29, 14823–14830. [PubMed: 24215478]
- (53). Henriques J; Costa PJ; Calhorda MJ; Machuqueiro M Charge parametrization of the dvh-c3 heme group: Validation using constant-(ph,e) molecular dynamics simulations. *J. Phys. Chem. B* 2013, 117, 70–82. [PubMed: 23199023]
- (54). Chu WT; Zhang JL; Zheng QC; Chen L; Wu YJ; Xue Q; Zhang HX Constant ph molecular dynamics (cphmd) and molecular docking studies of cquiobp1 ph-induced ligand releasing mechanism. *J. Mol. Model* 2013, 19, 1301–1309. [PubMed: 23179772]
- (55). Carvalheda CA; Campos SR; Machuqueiro M; Baptista AM Structural effects of ph and deacylation on surfactant protein c in an organic solvent mixture: A constant-ph md study. *J. Chem. Inf. Model* 2013, 53, 2979–2989. [PubMed: 24106805]
- (56). Bhattacharjee N; Rani P; Biswas P Capturing molten globule state of alpha-lactalbumin through constant ph molecular dynamics simulations. *J. Chem. Phys* 2013, 138, 095101. [PubMed: 23485328]
- (57). Ben-Shimon A; Shalev DE; Niv MY Protonation states in molecular dynamics simulations of peptide folding and binding. *Curr. Pharm. Des* 2013, 19, 4173–4181. [PubMed: 23170889]
- (58). Vila-Vicosa D; Campos SR; Baptista AM; Machuqueiro M Reversibility of prion misfolding: Insights from constant-ph molecular dynamics simulations. *J. Phys. Chem. B* 2012, 116, 8812–8821. [PubMed: 22803931]
- (59). Mondal J; Zhu X; Cui Q; Yethiraj A Sequence-dependent pka shift induced by molecular self-assembly: Insights from computer simulation. *J. Phys. Chem. B* 2012, 116, 491–495. [PubMed: 22121949]

- (60). Chu WT; Wu YJ; Zhang JL; Zheng QC; Chen L; Xue Q; Zhang HX Constant pH molecular dynamics (cphmd) and mutation studies: Insights into aegobp1 pH-induced ligand releasing mechanism. *Biochim. Biophys. Acta, Proteins Proteomics* 2012, 1824, 913–918.
- (61). Zimmermann T; Leszczynski J; Burda JV Activation of the cisplatin and transplatin complexes in solution with constant pH and concentration of chloride anions; quantum chemical study. *J. Mol. Model* 2011, 17, 2385–2393. [PubMed: 21523556]
- (62). Yang L; Zhang J; Ho B; Ding JL Histidine-mediated pH-sensitive regulation of m-ficolin:GlcNAc binding activity in innate immunity examined by molecular dynamics simulations. *PLoS One* 2011, 6, e19647. [PubMed: 21573188]
- (63). Sharma S; Bandyopadhyay P Investigation of the acylation mechanism of class C beta-lactamase: pKa calculation, molecular dynamics simulation and quantum mechanical calculation. *J. Mol. Model* 2012, 18, 481–492. [PubMed: 21541744]
- (64). Lee J; Miller BT; Damjanovic A; Brooks BR Enhancing constant-pH simulation in explicit solvent with a two-dimensional replica exchange method. *J. Chem. Theory Comput* 2015, 11, 2560–2574. [PubMed: 26575555]
- (65). Kim MO; Blachly PG; Kaus JW; McCammon JA Protocols utilizing constant pH molecular dynamics to compute pH-dependent binding free energies. *J. Phys. Chem. B* 2015, 119, 861–872. [PubMed: 25134690]
- (66). Polydorides S; Simonson T Monte Carlo simulations of proteins at constant pH with generalized Born solvent, flexible sidechains, and an effective dielectric boundary. *J. Comput. Chem* 2013, 34, 2742–2756. [PubMed: 24122878]
- (67). Chen W; Wallace JA; Yue Z; Shen JK Introducing titratable water to all-atom molecular dynamics at constant pH. *Biophys. J* 2013, 105, L15–L17. [PubMed: 23972860]
- (68). Itoh SG; Damjanovic A; Brooks BR pH replica-exchange method based on discrete protonation states. *Proteins: Struct., Funct., Genet* 2011, 79, 3420–3436. [PubMed: 22002801]
- (69). Donnini S; Tegeler F; Groenhof G; Grubmüller H Constant pH molecular dynamics in explicit solvent with lambda-dynamics. *J. Chem. Theory Comput* 2011, 7, 1962–1978. [PubMed: 21687785]
- (70). Meng Y; Roitberg AE Constant pH replica exchange molecular dynamics in biomolecules using a discrete protonation model. *J. Chem. Theory Comput* 2010, 6, 1401–1412. [PubMed: 20514364]
- (71). Damjanovic A; Brooks BR; García-Moreno EB Conformational relaxation and water penetration coupled to ionization of internal groups in proteins. *J. Phys. Chem. A* 2011, 115, 4042–4053. [PubMed: 21428436]
- (72). Castaneda CA; Fitch CA; Majumdar A; Khangulov V; Schlessman JL; Garcia-Moreno BE Molecular determinants of the pKa values of Asp and Glu residues in staphylococcal nuclease. *Proteins: Struct., Funct., Genet* 2009, 77, 570–588. [PubMed: 19533744]
- (73). Damjanovic A; Garcia-Moreno B; Lattman EE; Garcia AE Molecular dynamics study of water penetration in staphylococcal nuclease. *Proteins: Struct., Funct., Genet* 2005, 60, 433–449. [PubMed: 15971206]
- (74). Stites WE; Gittis AG; Lattman EE; Shortle D In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core. *J. Mol. Biol.* 1991, 221, 7–14.
- (75). Isom DG; Castaneda CA; Cannon BR; Garcia-Moreno B Large shifts in pKa values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. U. S. A* 2011, 108, 5260–5265.
- (76). Karp DA; Stahley MR; García-Moreno EB Conformational consequences of ionization of Lys, Asp, and Glu buried at position 66 in staphylococcal nuclease. *Biochemistry* 2010, 49, 4138–4146. [PubMed: 20329780]
- (77). Richman DE; Majumdar A; García-Moreno EB Conformational reorganization coupled to the ionization of internal Lys residues in proteins. *Biochemistry* 2015, 54, 5888–5897. [PubMed: 26335188]
- (78). Goh GB; García-Moreno EB; Brooks CL The high dielectric constant of staphylococcal nuclease is encoded in its structural architecture. *J. Am. Chem. Soc* 2011, 133, 20072–20075. [PubMed: 22085022]

- (79). Mardis KL; Luo R; Gilson MK Interpreting trends in the binding of cyclic ureas to hiv-1 protease. *J. Mo/. Bio/* 2001, 309, 507–517.
- (80). Ghosh N; Cui Q Pka of residue 66 in staphylococcal nuclease. I. Insights from qm/mm simulations with conventional sampling. *J. Phys. Chem. B* 2008, 112, 8387–8397. [PubMed: 18540669]
- (81). Zheng L; Chen M; Yang W Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Nat/.Acad. Sci. U.S.A* 2008, 105, 20227–20232.
- (82). Wu XW; Sung SS Simulation of peptide folding with explicit water - a mean solvation method. *Proteins: Struct., Funct., Genet* 1999, 34, 295–302. [PubMed: 10024017]
- (83). Bennett CH Efficient estimation of free energy differences from monte carlo data. *J. Comput. Phys* 1976, 22, 245–268.
- (84). Wu X-W; Sung S-S An efficient approach to estimating the thermodynamic properties of fluid mixtures in molecular simulation. *J. Chem. Phys* 1996, 104, 3709–3717.
- (85). Wu X; Brooks BR Molecular simulation with discrete fast fourier transform In *Fourier transform - materials analysis*; Salih SM, Ed.; InTech, 2012; pp 137–164.
- (86). Wu X; Brooks BR Isotropic periodic sum of electrostatic interactions for polar systems. *J. Chem. Phys* 2009, 131, 024107. [PubMed: 19603970]
- (87). Wu X; Brooks BR Using the isotropic periodic sum method to calculate long-range interactions of heterogeneous systems. *J. Chem. Phys* 2008, 129, 154115. [PubMed: 19045184]
- (88). Klauda JB; Wu X; Pastor RW; Brooks BR Long-range lennard-jones and electrostatic interactions in interfaces: Application of the isotropic periodic sum method. *J. Phys. Chem. B* 2007, 111, 4393–4400. [PubMed: 17425357]
- (89). Wu X; Brooks BR Isotropic periodic sum: A method for the calculation of long-range interactions. *J. Chem. Phys* 2005, 122, 044107.
- (90). Ojeda-May P; Pu J Assessing the accuracy of the isotropic periodic sum method through madelung energy computation. *J. Chem. Phys* 2014, 140, 164106. [PubMed: 24784252]
- (91). Ojeda-May P; Pu J Isotropic periodic sum treatment of long-range electrostatic interactions in combined quantum mechanical and molecular mechanical calculations. *J. Chem. Theory Comput* 2014, 10, 134–145. [PubMed: 26579897]
- (92). Wu X; Brooks BR; Vanden-Eijnden E Self-guided langevin dynamics via generalized langevin equation. *J. Comput. Chem* 2016, 37, 595–601. [PubMed: 26183423]
- (93). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys* 1983, 79, 926–935.
- (94). Brooks BR; Brooks CL, 3rd; Mackerell AD, Jr.; Nilsson L; Petrella RJ; Roux B; Won Y; Archontis G; Bartels C; Boresch S; et al. Charmm: The biomolecular simulation program. *J. Comput. Chem* 2009, 30, 1545–1614. [PubMed: 19444816]
- (95). Brooks BR; Brucoleri RE; Olafson BD; States DJ; Swaminathan S; Karplus M Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem* 1983, 4, 187–217.
- (96). MacKerell AD; Bashford D; Bellott M; Dunbrack RL; Evanseck JD; Field MJ; Fischer S; Gao J; Guo H; Ha S; et al. Allatom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 1998, 102, 3586–3616. [PubMed: 24889800]
- (97). Ryckaert JP; Ciccotti G; Berendsen HJC Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes. *J. Comput. Phys* 1977, 23, 327–341.

Figure 1.

Virtual mixture of multiple states (VMMS) system for a two state equilibrium at $\text{pH} = \text{pK}_a$. In a solution with $\text{pH} = \text{pK}_a$ shown in the top, the two states, protonated state (H) and deprotonated state (D), equilibrate with each other and have equal molar fractions, $x_H = x_D = 0.5$. The solute molecules at the two states share the same conformation and search their conformational spaces together. Each state is represented by a subsystem and has its own solvent environment, which can be either implicit or explicit. All subsystems form a virtual ideal solution and equilibrate with each other.

Figure 2.

Atomic charges of lysine in its protonated/deprotonated state.

Figure 3. Structure of N118K variant of SNase. K118 is shown as sticks. The distribution of the amine group in the side chain of K118 is shown as a blue volume. The distributions of water molecules around the K118 amine group are shown as red and yellow volumes for the protonated state and the deprotonated state, respectively.

Figure 4. VMMS simulation profiles of N118K. From bottom up are pKa, RMSD, interaction group numbers at the protonated state ($N_g^{(H)}$) and the deprotonated state ($N_g^{(D)}$).

Figure 5. Structure of the SNase variant V104K. K104 is shown as sticks. The distribution of K104 side chain amine group is shown as a blue volume. The distributions of water molecules around K104 amine group are shown as red and yellow volumes for the protonated state and the deprotonated state, respectively.

Figure 6. VMMS simulation profiles of the SNase variant V104K. From bottom up are pK_a RMSD, number of interaction group numbers at the protonated state ($N_g^{(H)}$) and the deprotonated state ($N_g^{(D)}$).

Figure 7. Structure of L125K variant of SNase. K125 is shown as sticks. The distribution of K125 side chain amine group is shown as a blue volume. The distributions of water molecules around K125 amine group are shown as red and yellow volumes for the protonated state and the deprotonated state, respectively.

Figure 8. VMMS simulation profiles of L125K. From bottom up are pKa, RMSD, interaction group numbers at the protonated state ($N_g^{(H)}$) and the deprotonated state ($N_g^{(D)}$).

Figure 9. Structure of I92K variant of SNase. K92 is shown as sticks. The distribution of K92 side chain ionizable group is shown as a blue volume. The distributions of water molecules around K92 amine group are shown as red and yellow volumes for the protonated state and the deprotonated state, respectively.

Figure 10.

VMMS simulation profiles of I92K. From bottom up are pK_a , RMSD, interaction group numbers at the protonated state ($N_g^{(H)}$) and the deprotonated state ($N_g^{(D)}$).

Figure 11.

Comparison of the pK_a values obtained from the equal-molar VMMS simulations and experimental measurements.

Figure 12.

Comparison of conformational relaxation and the lysine side chain flexibilities of the 25 SNase variants against their pK_a values. Bottom: RMSD of the backbone; middle: RMSD of the lysine amine group; top: distribution volume of the lysine amine group.

Figure 13.

Correlations between the rates of interaction changes and the pK_a values. Top: protein polar group interaction changes; middle: water interaction changes; bottom: total interaction changes.

Figure 14.

Microenvironment properties of the lysine side chain against pK_a experimental values for the 25 SNase variants. The microenvironment is quantified by groups within 6 Å of the lysine amine group. (a) N_{polar} is the total number of polar groups in the microenvironment. (b) δN_w is the difference of water molecule numbers between the protonated and deprotonated states. (c) $\rho_w = N_w / (N_{\text{polar}} + N_{\text{nonp}})$ defines the water ratio of the microenvironment. (d) $\rho_{\text{nonp}} = N_{\text{nonp}} / (N_{\text{polar}} + N_{\text{nonp}})$ defines the hydrophobicity of the microenvironment.

Table 1.

Simulation Systems for the 25 Variants of SNase

mutates	PDB(mutation)	protonated		deprotonated		$pK_a^{(VMMS)}$	
		pK_a	water	(Ca^{2+}, Na^+, Cl^-)	water		(Ca^{2+}, Na^+, Cl^-)
A109K	4yji(e109k)	9.2	2786	(1,4,15)	2785	(1,5,15)	9.51 ± 0.03
A132K	3bdc	10.4	2809	(1,6,15)	2808	(1,7,15)	9.71 ± 0.02
A58K	3bdc	10.4	2816	(1,6,15)	2815	(1,7,15)	10.08 ± 0.05
A90K	3dhq(r90k)	8.6	2817	(1,6,15)	2816	(1,7,15)	8.12 ± 0.03
F34K	3itp	7.1	2811	(1,6,15)	2810	(1,7,15)	6.23 ± 0.05
G20K	5e1f(e20k)	10.4	2786	(1,6,15)	2785	(1,7,15)	10.93 ± 0.02
I72K	2rbm	8.6	2782	(1,5,15)	2781	(1,6,15)	8.35 ± 0.03
I92K	1tt2	5.3	2800	(1,5,15)	2799	(1,6,15)	4.77 ± 0.06
L103K	3e5s	8.2	2808	(1,6,15)	2807	(1,7,15)	9.47 ± 0.03
L125K	3cle	6.2	2775	(1,6,15)	2774	(1,7,15)	4.67 ± 0.03
L25K	3erq	6.3	2804	(1,6,15)	2803	(1,7,15)	6.02 ± 0.05
L36K	3eji	7.2	2798	(1,6,15)	2797	(1,7,15)	8.45 ± 0.05
L37K	3bdc	10.4	2811	(1,6,15)	2810	(1,7,15)	9.65 ± 0.03
L38K	3bdc	10.4	2826	(1,6,15)	2825	(1,7,15)	8.51 ± 0.05
N100K	5deh(d100k)	8.6	2802	(1,6,15)	2801	(1,7,15)	8.53 ± 0.04
N118K	3bdc	10.4	2804	(1,6,15)	2803	(1,7,15)	10.53 ± 0.02
T41K	3bdc	9.3	2810	(1,6,15)	2809	(1,7,15)	10.00 ± 0.06
T62K	3dmu	8.1	2819	(1,8,15)	2818	(1,9,15)	7.56 ± 0.05
V104K	3clf	7.7	2803	(1,6,15)	2802	(1,7,15)	9.35 ± 0.04
V23K	3qoj	7.3	2818	(1,6,15)	2817	(1,7,15)	9.34 ± 0.04
V39K	3sk5(d39k)	9	2807	(1,6,15)	2806	(1,7,15)	7.27 ± 0.02
V66K	3owf	5.6	2796	(1,6,15)	2795	(1,7,15)	7.22 ± 0.03
V74K	3ruz	7.4	2799	(1,6,15)	2798	(1,7,15)	7.23 ± 0.04
V99K	4hmi	6.5	2785	(1,5,15)	2784	(1,6,15)	8.26 ± 0.05
Y91K	3d4d(e91k)	9	2799	(1,6,15)	2798	(1,7,15)	10.73 ± 0.03