

RESEARCH ARTICLE

A Virtual Mixture Approach to the Study of Multistate Equilibrium: Application to Constant pH Simulation in Explicit Water

Xiongwu Wu*, Bernard R. Brooks

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

* wuxw@nhlbi.nih.gov



Abstract

Chemical and thermodynamic equilibrium of multiple states is a fundamental phenomenon in biology systems and has been the focus of many experimental and computational studies. This work presents a simulation method to directly study the equilibrium of multiple states. This method constructs a virtual mixture of multiple states (VMMS) to sample the conformational space of all chemical states simultaneously. The VMMS system consists of multiple subsystems, one for each state. The subsystem contains a solute and a solvent environment. The solute molecules in all subsystems share the same conformation but have their own solvent environments. Transition between states is implicated by the change of their molar fractions. Simulation of a VMMS system allows efficient calculation of relative free energies of all states, which in turn determine their equilibrium molar fractions. For systems with a large number of state transition sites, an implicit site approximation is introduced to minimize the cost of simulation. A direct application of the VMMS method is for constant pH simulation to study protonation equilibrium. Applying the VMMS method to a heptapeptide of 3 ionizable residues, we calculated the pKas of those residues both with all explicit states and with implicit sites and obtained consistent results. For mouse epidermal growth factor of 9 ionizable groups, our VMMS simulations with implicit sites produced pKas of all 9 ionizable groups and the results agree qualitatively with NMR measurement. This example demonstrates the VMMS method can be applied to systems of a large number of ionizable groups and the computational cost scales linearly with the number of ionizable groups. For one of the most challenging systems in constant pH calculation, SNase Δ +PHS/V66K, our VMMS simulation shows that it is the state-dependent water penetration that causes the large deviation in lysine66's pKa.

OPEN ACCESS

Citation: Wu X, Brooks BR (2015) A Virtual Mixture Approach to the Study of Multistate Equilibrium: Application to Constant pH Simulation in Explicit Water. *PLoS Comput Biol* 11(10): e1004480. doi:10.1371/journal.pcbi.1004480

Editor: Qiang Cui, University of Wisconsin-Madison, UNITED STATES

Received: March 23, 2015

Accepted: July 29, 2015

Published: October 27, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was supported by the Intramural Research Programs of National Heart, Lung, and Blood Institute (Z01 HL001027-30). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

Computer simulation plays an important role to understand molecular systems and has been applied to problems of increasing complexity. Multistate equilibrium is a fundamental concept behind the structure and function of biological systems. Due to the limit in

computing resources and lack of good alternative methods, computer simulation has been conducted for systems in a single state, sampling from one state to another to infer equilibrium properties. This sequential approach has been successful in many cases such as protonation equilibrium with implicit solvation model. However, state transition is difficult when explicit solvent is used for more accurate solvation description. Many efforts have been dedicated to overcome this difficulty. Analogous to real multistate systems, we proposed a virtual mixture of multiple states (VMMS) to directly simulate the equilibrium. State transitions are replaced by changes in state molar fractions. Mimicking a test tube environment, all states are simulated in parallel to equilibrate with each other. Application to constant pH simulation in explicit water demonstrates the capability of this method. It is expected that the VMMS method will find more applications in biological problems related to the equilibrium of competing states.

“This is a *PLOS Computational Biology* Methods paper”

Introduction

Chemical and thermodynamic equilibrium of multiple states is a fundamental phenomenon in biological systems. Typical examples of multiple state equilibria are protonation equilibrium, ligand binding equilibrium, and phosphorylation equilibrium. For example, through the equilibrium of different protonation states, protein can fold or unfold in different pH environments. By changing phosphorylation states, protein can activate or deactivate certain functions. Ligand binding equilibrium controls the inhibition of enzyme activities.

Computer simulation has played an increasingly important role in understanding biological systems. However, due to the limit of computing resources and lack of good alternative methods, equilibrium of multiple states is often studied through simulations of individual states in order to minimize size and complexity of simulation systems. Equilibrium properties are derived indirectly through free energy calculation. For example, ligand binding equilibrium is indirectly studied by calculating binding affinities.

Some methods directly addressing state equilibrium have been developed recently. For example, constant pH simulation methods [1–20] can simulate equilibrium between different protonation states. In these methods, equilibrium between states is simulated through visiting or sampling different states sequentially. In other words, through frequently changing from one state to another, a simulation samples the equilibrium distributions of various states. Such a sequential sampling process works well if the transition between states is easy. For example, with implicit solvation models, transition between protonated and deprotonated states has no difficulty since implicit water reorganizes instantly, and proper sampling can be achieved through, for example, Monte Carlo [5] or replica-exchange [12]. However, when transition between states is difficult, such as in explicit solvent where large barriers exist between states, special efforts are needed to improve the sampling efficiency. Typical examples are the λ -dynamics [14] and the Enveloping Distribution Sampling [19,21–23].

In this work, we present a new simulation approach to address multi-state equilibrium in a parallel way without direct transition between states. All states exist simultaneously in a simulation with populations defined by their molar fractions. The transition between states is implicated by changes of their molar fractions. By avoiding direct transitions between states, this

method is designed for problems where transitions between states are difficult, such as deprotonation in explicit solvent. We call this new approach the virtual mixture of multi-states (VMMS) method, because this method simulates multiple states explicitly and treats explicit states as though in a virtual mixture to equilibrate with each other.

Results

The VMMS simulation method uses multiple techniques to achieve efficient simulation. Through several examples, we validate the techniques and demonstrate the application of the VMMS method in the study of protonation equilibrium.

Free energy calculation with IPS

Using the IPS potential for long range energy calculation is a key factor for VMMS efficiency. The IPS potential converts long-range interactions to pairwise interactions within a certain range, typically within a cutoff. This cutoff region is called the local region. Everywhere beyond the local region is called the remote region. Interactions with the remote region are replaced by interactions with the isotropic periodic image of the local region. The summation over the IPS image is an analytic function of the distance between particles within a local region. Ewald sum uses lattice images to calculate long range interactions. It has contributions from all atom pairs and depends not only on the distances, but also on the orientations of all atom pairs. Due to these complexities, Ewald summation needs special numerical methods such as the Particle Meshed Ewald (PME) technique to improve efficiency. Apparently, IPS and PME are based on similar concepts, but with different local regions and images. It has been demonstrated that IPS produces very similar results as PME [24–29]. However, PME is not pairwise and to separate solute and solvent interactions requires multiple PME calculations, thus, making it less suitable for the VMMS simulation.

Although many cutoff methods are pairwise, they cannot provide accurate free energies, which is crucial for the study of state equilibrium. IPS is as simple as any cutoff method, but, as accurate as PME [25–32]. To demonstrate the accuracy of IPS potential, we calculated the deprotonation free energies of the model compounds. The free energy profiles during the calculations are shown in Fig 1. For comparison, Fig 1 also shows the results from PME and another cutoff method.

The deprotonation free energies were calculated through thermodynamic integration with the PERT module of CHARMM. Each of the model compounds was dissolved in a box of TIP3P water. The box size was $31.1 \times 31.1 \times 31.1 \text{ \AA}^3$. The cutoff method [33] used force shift for electrostatic potential and energy switch for Lennard-Jones potential with $r_{\text{on}} = 10 \text{ \AA}$ and $r_{\text{off}} = 12 \text{ \AA}$. The 3D IPS used a local radius of $R_c = 12 \text{ \AA}$. For PME, a real space cutoff of 12 \AA was used and the grid dimension for Fast Fourier transform (FFT) calculation was $32 \times 32 \times 32$. There were total 190 λ windows, $\Delta\lambda = 0.01$ for $0 \leq \lambda \leq 0.9$ and $\Delta\lambda = 0.001$ for $0.9 < \lambda \leq 1$. For each window a 10 ps LD simulation was performed with a friction constant of 1/ps.

From Fig 1 we can see that the IPS results converge well with the PME results at $\lambda = 1$, with an average deviation of 0.37 kcal/mol. The cutoff results have an average deviation of 4.4 kcal/mol from the PME results. These results demonstrate that IPS is an efficient and competitive alternative to PME when pairwise properties are needed. The cutoff methods can provide pairwise terms, but their deviation from PME makes them a poor choice for quantitative studies.

Note that at intermediate λ the IPS and the PME results are differ, but at the significant end points ($\lambda = 0$ and $\lambda = 1$) the variation of free energy differences between IPS and PME is very small. During the free energy calculation, the energy at intermediate λ was a combination of the two end states. The difference at intermediate λ values is due to the different treatments of

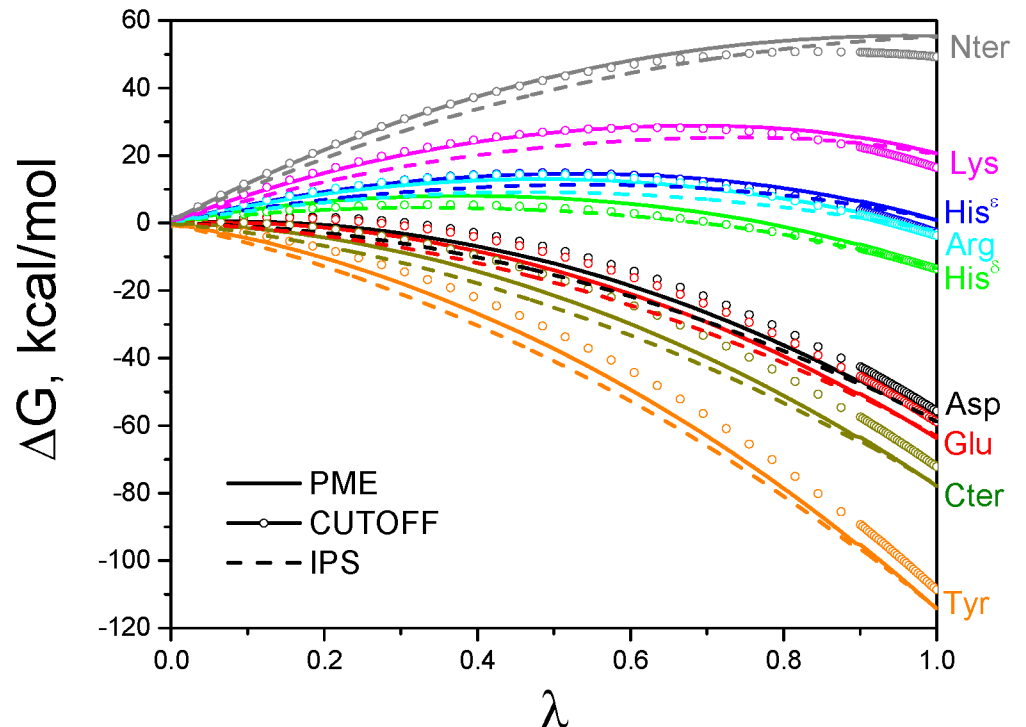


Fig 1. Deprotonation free energy profiles during the thermodynamic integration calculations. The PME and IPS results are shown as solid and dashed lines, respectively. The cutoff results are shown with open circles. Results for different residues are colored differently as labeled.

doi:10.1371/journal.pcbi.1004480.g001

boundary interactions in the two methods. What matters is the free energy difference between the end states, which is used to determine the state equilibrium.

VMMS ensemble distribution

Reweighting a VMMS conformation distribution to obtain the pure state conformational distributions is the key to simulate multistate equilibrium. The partition function of the VMMS ensemble, Eq (6), describes the VMMS conformational distribution and provides a reweighting relation, Eqs (9) and (10), to convert a VMMS distribution to a pure state distribution. Here we use two simple systems to demonstrate this reweighting relation.

First, we use a model compound, ACE-ASP-NME, with the generalized Born (GB) implicit solvation model[34] to examine the VMMS ensemble. With an implicit solvation model, the VMMS system contains only the solute in two charge states, the protonated state (state 0) and the deprotonated state (state 1). Fig 2 shows the energy distributions from the VMMS simulation at $x_H = x_D = 0.5$ and from the Langevin dynamics (LD) simulations of the pure states. As can be seen, the two ensembles have quite different energy distributions, with the VMMS ensemble shifting toward higher energy. After reweighting with Eq (10), the VMMS results match very well with the LD results for the pure states, validating the reweighting relation, Eq (10) and verifying indirectly that Eq (6) describes the conformational distribution of the VMMS ensemble. It should be noted that the reweighting process often has large noise at high energy region. A peak around -58 kcal/mol in the reweighted distribution (green line) is due to such noise.

Next, we examine a very simple model system with explicit solvent. Because a regular system with explicit solvent is often overwhelmed by solvent-solvent interactions, making it difficult

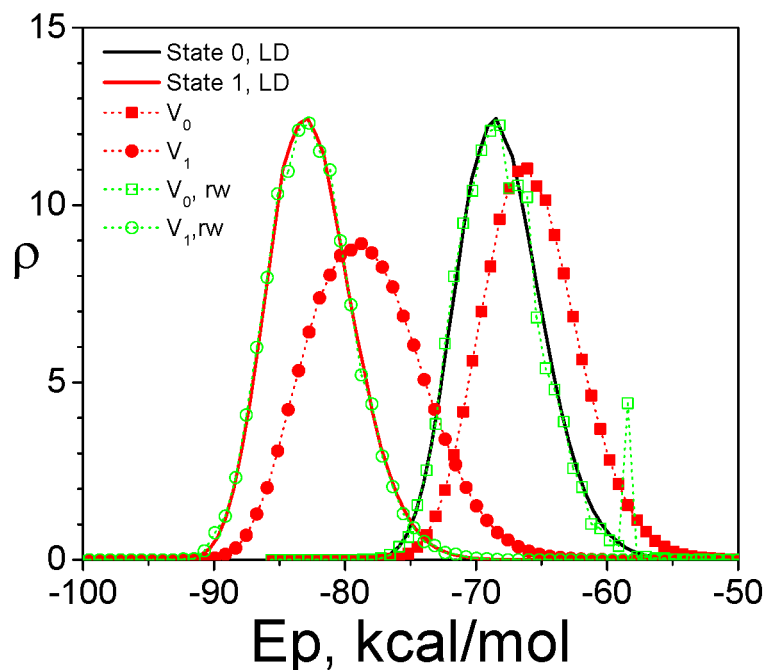


Fig 2. Energy distributions of the ASP model compound with the GB implicit solvation model in different states. The solid lines are the protonated state (state 0, black) and deprotonated state (state 1, red). The red dotted lines with filled symbols are in the V_0 (square) state and V_1 (circle) state. The green dotted lines with open symbols are the reweighted results in the V_0 (square) and V_1 (circle) states.

doi:10.1371/journal.pcbi.1004480.g002

to see the difference between a VMMS state and a pure state, we built a model solution with just one water molecule as the solute and 10 water molecules as the solvent in a cubic periodic box for each state. To make the two states more distinct, the solute charges were changed to have $q_O = 2.64e$ and $q_{H1} = 0$ and $q_{H2} = -0.64e$ for state 0 and $q_O = -2.64e$, $q_{H1} = 0.64e$, and $q_{H2} = 0$ for state 1. As a result, the solute had a net charge of $+2e$ in state 0 and a net charge of $-2e$ in state 1. The solvent charges were scaled by 0.1 to have $q_O = -0.0834e$ and $q_{H1} = q_{H2} = 0.0417e$. The solvent charge scaling made this solution a homogeneous system at $T = 300$ K and made the two states distinct in energy distribution. The cubic box size was $15.55 \times 15.55 \times 15.55 \text{ \AA}^3$. The 3D IPS potentials, Eqs (11)–(16), with a cutoff radius of 7.5 \AA , which was less than half of the box size, were used for both Lennard-Jones and electrostatic interactions.

We performed a VMMS simulation at $x_H = x_D = 0.5$ and LD simulations for both pure states. All simulation were done at $T = 300$ K and $\xi = 1/\text{ps}$. The energy distributions of this model solution are shown in Fig 3. The VMMS distributions are quite different from the LD distributions for both states. After reweighting, the VMMS energy distributions match very well with the LD distributions. This example validates the VMMS reweighting function, Eq (10), for explicit solvent.

VMMS simulations of model compounds

The VMMS simulation converts a state transition problem to a free energy calculation between the two states and pushes all contributions, such as chemical bonding and proton creation, into a reference value derived from the experimental pKa and simulated free energy difference of model compounds, as described by Eq (24). Therefore, we need to estimate the free energy difference of the modeling compounds before we can study state equilibrium of interested

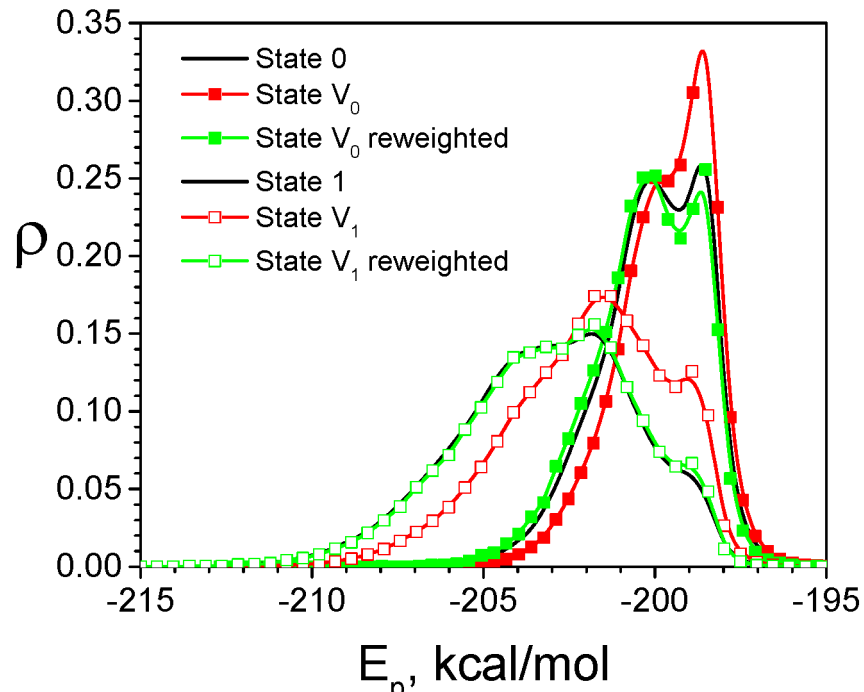


Fig 3. Energy distributions of the model HOH solution in the canonical ensemble and in the VMMS ensemble. The black lines are the LD simulation results for state 0 (solid) and state 1 (dashed). The red lines with squares are the VMMS simulation results for state V_0 (filled) and state V_1 (open). The green lines with circles are the reweighted VMMS simulation results for state V_0 (filled) and state V_1 (open).

doi:10.1371/journal.pcbi.1004480.g003

systems. To calculate the deprotonation free energies of the model compounds, we performed equal-molar VMMS simulations at fixed composition, $x_H = x_D = 0.5$. Fig 4 shows deprotonation free energies evaluated during the equal-molar VMMS simulations. The deprotonation free energies and the experimental pKas of these model compounds are used as reference free energies and reference pKas of the corresponding titration sites and are listed in Table 1.

Using these reference deprotonation free energies, we performed VMMS simulations for the model compounds at constant pH values to identify the equilibrium molar fractions. Fig 5 shows the molar fractions of the deprotonated states as functions of pH. The pKa should be the pH where $x_D = 0.5$. The pKa values estimated this way are listed in Table 1. As can be seen, the pKa estimated from the VMMS simulations agree well with the experimental values. Large fluctuations in the pKa results are caused by the large fluctuation in molar fractions and the short simulation lengths.

VMMS simulations of a heptapeptide with 3 titration sites

Now we are ready to apply the VMMS method in constant pH simulations. Here, we choose a peptide of 7 residues with a sequence: ACE-SDNKTYG-NME. This heptapeptide was derived from OMTKY3 and has been studied experimentally[35] and computationally[3,15,35]. The peptide was dissolved with 966 TIP3P water molecules in a $31.1\text{\AA} \times 31.1\text{\AA} \times 31.1\text{\AA}$ cubic box.

First, we present the results with all titration sites explicitly simulated. There are three ionizable residues, D, K, and Y and a total of $2^3 = 8$ titration states. For easy discussion, we denote the 8 states by the states of these three residues, H for a protonated residue and D for a deprotonated residue. For example, HHH stands for a state where all the three residues are in

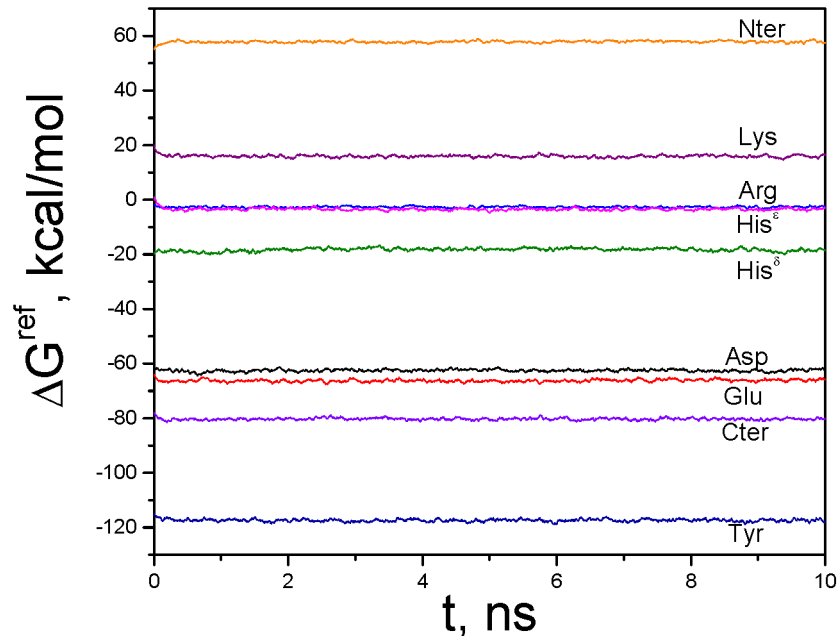


Fig 4. Deprotonation free energies evaluated during the equal-molar VMMS simulations of the model compounds.

doi:10.1371/journal.pcbi.1004480.g004

protonated states and DHD stands for a state where residues D and Y are in the deprotonated state and residue K in the protonated state.

As described in the method section, with 3 explicit sites, the VMMS system contains 8 simulation boxes with each state in one box. The peptide in the 8 boxes has the same conformation and moves the same way, but water molecules in each box interact with the peptide in one of the eight states and are independent of water in other boxes.

The VMMS simulations were performed for the peptide at constant pH of 0, 2, 4, 6, 8, 10, 12, and 14. All states initially had a molar fraction of 0.125 and were allowed to change to reach equilibrium. The state molar fractions during these simulations are shown in Fig 6. At pH = 0 and 2, all states approached zero except state HHH. At pH = 4, state DHH became dominant followed by state HHH. At pH = 6 and 8, DHH was the only noticeable state. At pH = 10, DHH and DHD were the competitive dominant forms. At pH = 12, DDD became the dominant form with significant amount of DHD. At pH = 14, DDD was the only dominant form.

The equilibrium molar fractions of all states are plotted against pH in Fig 7. There are in total 12 deprotonation processes between these 8 states. The pKa of these deprotonation processes can be determined by the pH values where the molar fractions of both the protonated state and deprotonated state are equal. There are four deprotonation processes for each residue and their equal molar positions are circled in Fig 7. As can be seen, the intersections of protonated state and deprotonated state are around pH = 4 for Asp, around 11 for Lys, and around 10 for Tyr. The pKa values of the ionizable residues are determined by the most dominant states and the results are listed in Table 2. These results are reasonably close to the pH-REMD and CpH results obtained with an implicit solvation model[15].

Next, we show the results of VMMS-1 simulations where at any moment only one site was explicit and the other two sites were implicit. As described in the method section, when a site is implicit, the atomic charges of this site will be calculated with Eq (27) based on the atomic

Table 1. Deprotonation free energies calculated using equal-molar VMMS simulations and pKa values calculated from constant pH VMMS simulations. Experimental pKa values are listed as reference pKas.

Compounds	ΔG^{ref} , kcal/mol	pKa(VMMS)	pKa ^{ref} (exp)[12]
ASP	-62.4±0.4	4.0±0.2	4.0
GLU	-66.2±0.4	4.4±0.2	4.4
HIS ^δ	-18.0±0.4	6.4±0.4	6.5
HIS ^ε	-3.4±0.4	7.4±0.4	7.1
ARG	-2.6±0.3	12.4±0.3	12.5
LYS	15.9±0.4	10.4±0.2	10.4
TYR	-117.4±0.5	9.6±0.2	9.6
Nter	57.9±0.3	7.4±0.5	7.5
Cter	-80.2±0.4	3.8±0.6	3.8

doi:10.1371/journal.pcbi.1004480.t001

charges in protonated and deprotonated states and their molar fractions. When a site changes from implicit to explicit, atomic charges return to the values in the corresponding states.

In the VMMS-1 simulations, each site was explicitly simulated for 10 ps and then became implicit. The first 2 ps simulation was used for equilibrium and the following 8 ps were used to evaluate free energy differences and to update molar fractions of explicit states. After becoming implicit, the ionizable sites maintain their state molar fractions until they become explicit again.

Fig 8 shows the molar fractions of the three ionizable sites during the VMMS-1 simulations at various pH values. All simulations started with the deprotonated molar fractions, x_D , of 0.5 for all three sites. The molar fractions of these three sites responded differently to pH values. Large fluctuations in molar fractions were observed when the pH value was close to the pKa value.

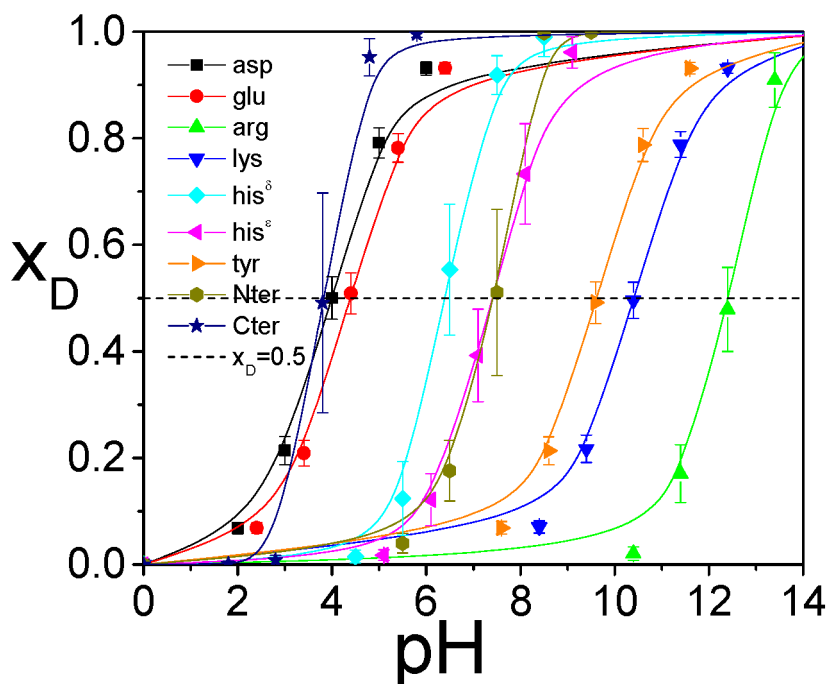


Fig 5. The titration curves of the model compounds obtained from the VMMS simulations.

doi:10.1371/journal.pcbi.1004480.g005

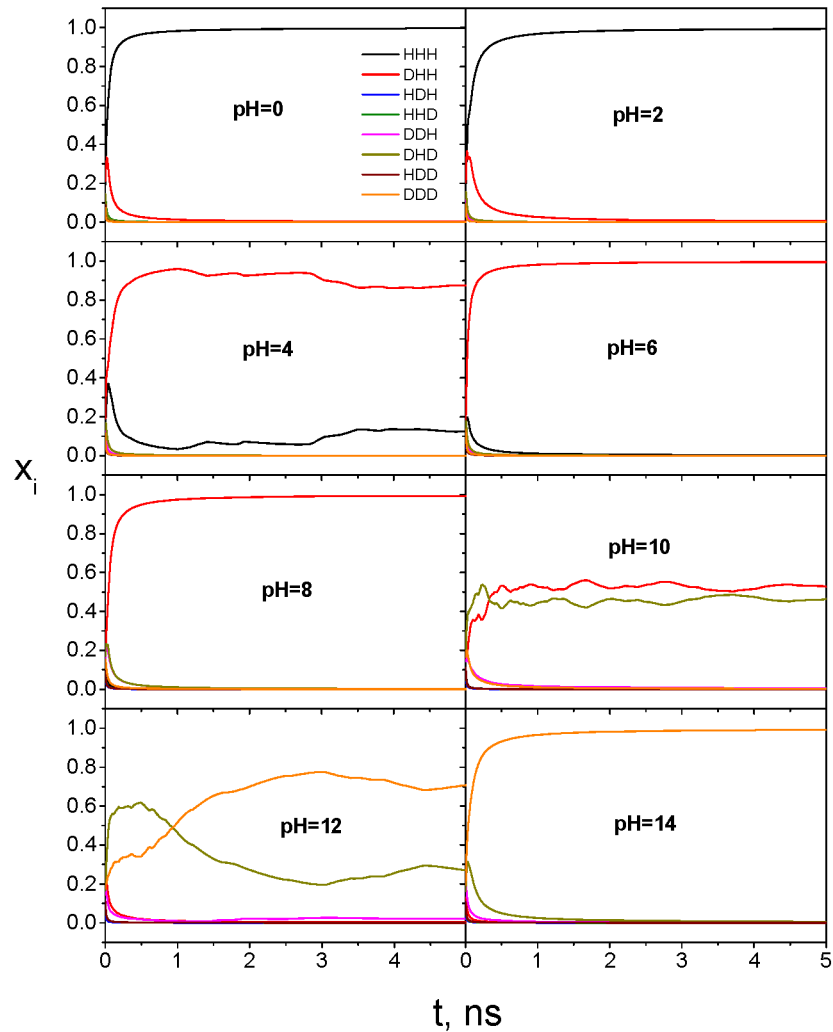


Fig 6. State molar fractions of the heptapeptide, Ace-SDNKTYG-NME, during constant pH VMMS simulations. pH values are labeled in each panel. All states have an initial molar fraction of 0.125.

doi:10.1371/journal.pcbi.1004480.g006

Fig 9 plots the molar fractions as functions of pH for the three ionizable residues. The pKa values determined from the pH values where $x_D = 0.5$, are listed in Table 2. Within the statistic ranges, the VMMS and VMMS-1 results agree with each other. The consistency of the results from the fully explicit VMMS and the partially implicit VMMS-1 demonstrates that the implicit site approximation is acceptable for this system.

Mouse epidermal growth factor with 9 ionizable groups

To further illustrate the application of the VMMS method in constant pH study, we applied the VMMS method to mouse epidermal growth factor (EPG) in explicit water, for which pKa's of 9 ionizable groups have been determined by NMR[36] (Table 3). It is impractical to explicitly simulate all protonation states of these 9 ionizable groups. Instead, we performed VMMS-1 simulations where only one site at a time was explicit and the rest were implicit.

EPG has 53 residues. Its structure is available in PDB (PDB code: 1EPI). This structure was dissolved in a $62.21 \times 46.65 \times 46.65 \text{ \AA}^3$ water box with 4273 TIP3P water molecules. Simulations were performed at constant temperature of 300 K and constant volume. 3D IPS with a local

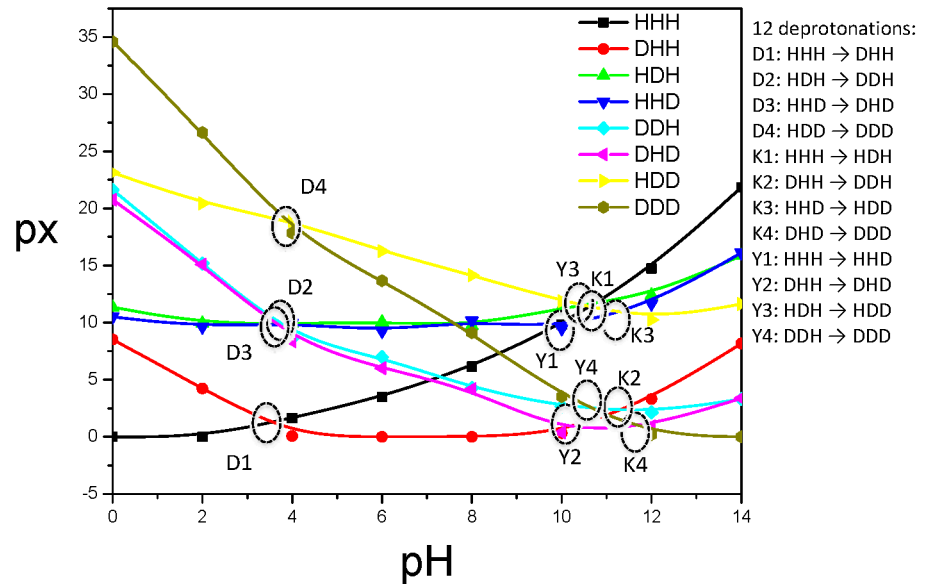


Fig 7. State molar fractions obtained from the constant pH VMMS simulations. There are 12 deprotonations and their pKa can be found from the intersection pH between curves of the protonated and deprotonated states, where $x_H = x_D$. The overall pKa will be determined by the dominant forms, i.e., the lowest intersections.

doi:10.1371/journal.pcbi.1004480.g007

region radius of 12 Å was used for both Lennard-Jones and electrostatic interactions. Every ionizable group was explicitly simulated for 10 ps in the sequential order.

The VMMS-1 simulations were performed at 10 pH values: 0, 2, 3, 4, 5, 6, 7, 8, 9, and 10. The molar fractions of the 9 ionizable groups are shown in Fig 10. As can be seen, starting from 0.5, the deprotonated molar fractions of these ionizable groups changed gradually to approach their equilibrium values. While 5 ns was not long enough for all states to reach their equilibrium molar fractions, the results show that the molar fractions of different groups responded differently to the pH changes and provide rough estimates of the pKas.

For the N-terminal group at Asn1 the deprotonated molar fraction approached 0 when $pH < 9$ and 1 when $pH = 10$. The four Asp residues, Asp11, Asp27, Asp40, and Asp46, were all different from each other. For Asp11, the titration point falls between $pH = 2$ and $pH = 3$. For Asp27, the titration point is between $pH = 4$ and $pH = 5$. Whereas for Asp40, it is between $pH = 3$ and $pH = 4$, and for Asp46, it is between $pH = 5$ and $pH = 6$.

We used the average molar fractions between 2 and 5 ns to plot the titration curves for each ionizable group in Fig 11. From the curves we can read the pKa values of these 9 groups, which are listed in Table 3. The agreement with NMR results is qualitatively good. To increase the accuracy, we need much longer simulations. Also, the simulation setup may also affect the results, such as salt concentration, force field, water model, etc. From this example, we

Table 2. pKa of the ionizable residues in the heptapeptide from the VMMS simulations and other studies.

residues	VMMS	VMMS-1	pH-REMD[15]	CpHMD[15]	Exp[35]
asp	3.6±0.5	4.1±0.3	3.6±0.2	3.7±0.2	3.6
lys	11.5±0.5	11.1±0.4	10.6±0.1	10.6±0.1	
tyr	10.2±0.5	9.9±0.5	10.1±0.1	9.9±0.1	

doi:10.1371/journal.pcbi.1004480.t002

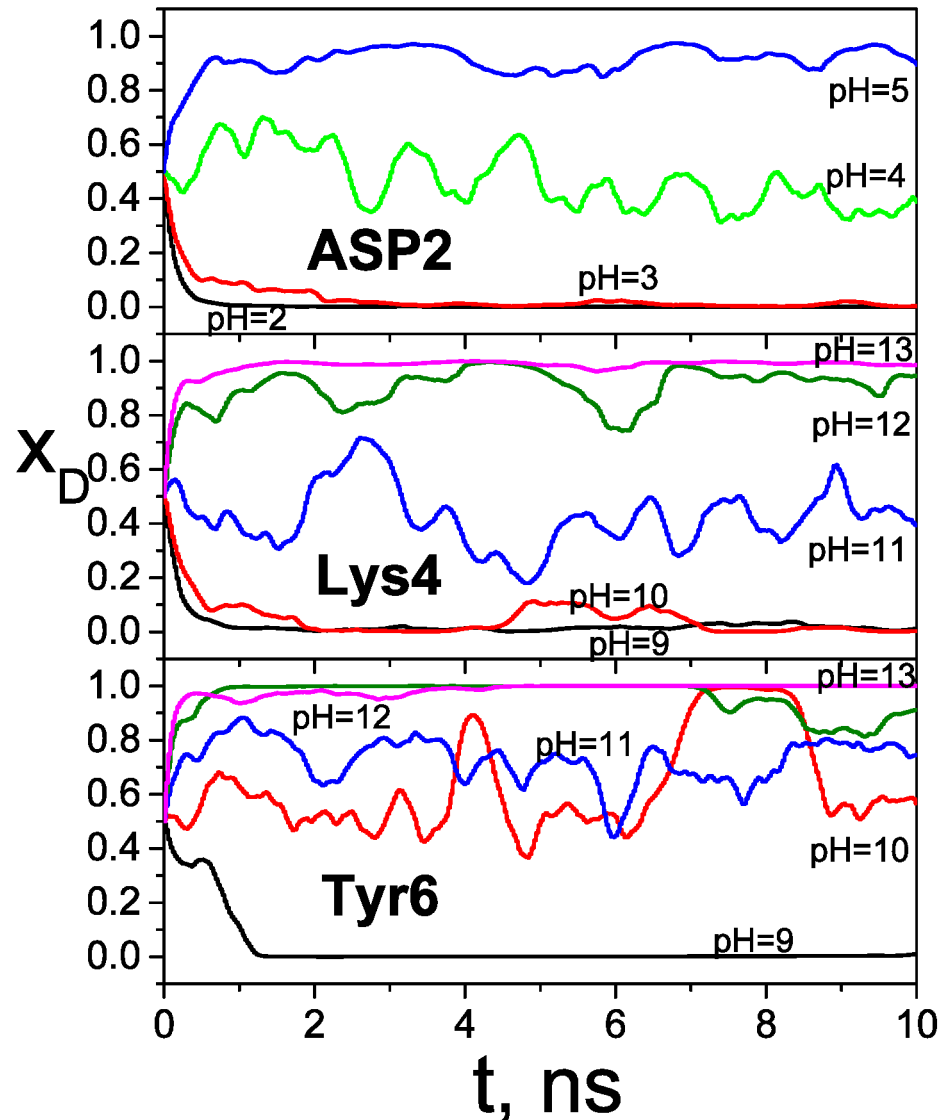


Fig 8. The molar fractions of deprotonated states during the VMMS-1 simulations of the heptapeptide.

doi:10.1371/journal.pcbi.1004480.g008

demonstrated that the VMMS-1 method can be applied to systems of many ionizable sites and the computing cost scales linearly with the number of ionizable sites.

VMMS simulation of SNase Δ +PHS/V66K in explicit water

To illustrate the capability of the VMMS method, we chose to simulate one of the most challenging systems in constant pH simulation, SNase Δ +PHS/V66K (PDB: 3HZX). This system is challenging because the Lys66 is buried inside the protein and has a large deviation in pKa from typical lysine residues. This system has been the focus of many studies[37–39]. Experimentally the pKa of this lysine is 5.6, very different from typical pKa of 10.4 for lysine.

In our VMMS simulation, only K66 has explicit protonated and deprotonated states. This protein was dissolved with 2830 TIP3P water in a $46.65 \times 46.65 \times 46.65 \text{ \AA}^3$ cubic box. 100 ns VMMS simulation was performed at constant molar fraction of $x_H = x_D = 0.5$.

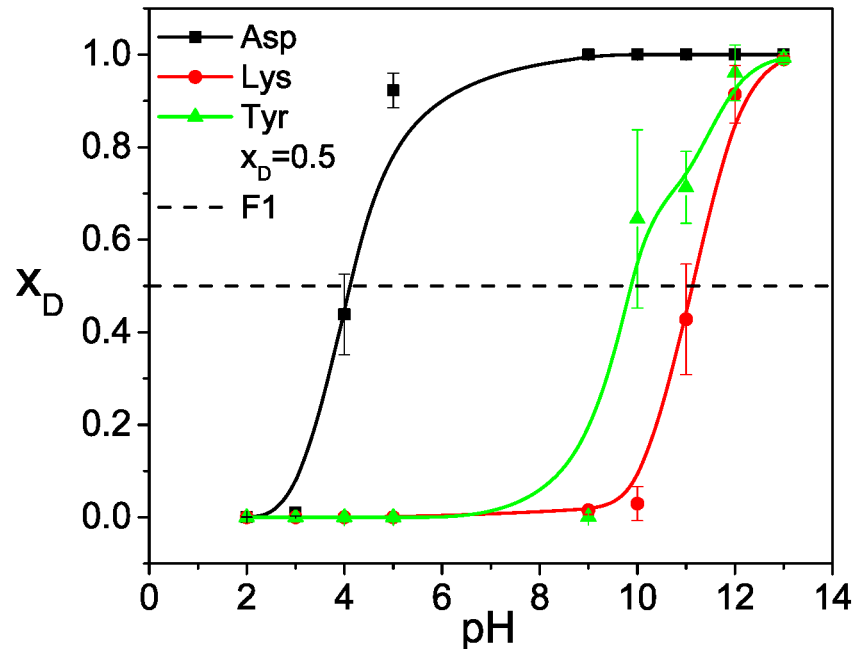


Fig 9. The molar fractions of deprotonated states as functions of pH for the heptapeptide obtained from the VMMS-1 simulations.

doi:10.1371/journal.pcbi.1004480.g009

The pKa of K66 obtained from the simulation is shown in Fig 12. Clearly, the pKa had large fluctuations, representing environment changes around K66. The pKa quickly reached 7 in 3 ns and remained there for more than 10 ns. The pKa dropped to below 4 at around 20 ns, then came back to above 7 at about 30 ns. Before 40 ns the pKa dropped to below 5.6 and fluctuated around 5.6 afterward.

We examined the simulation trajectory to determine what causes the large deviation and fluctuation in K66’s pKa. First, we found that the protein conformation remained stable throughout the simulation. As shown in the rmsd plot of Fig 12, after the 100 ns simulation, the rmsds of the protein backbone and all atoms from the initial structure were about 1.4 Å and 2 Å, respectively. While certain conformational fluctuations occurred, the sidechain of K66 remained inside its original place, indicating that the large deviation of K66’s pKa is not necessarily due to its sidechain flipping in and out at different protonation states. Next, we

Table 3. pKa of the ionizable residues in mouse epidermal growth factor from the VMMS simulations and other studies.

residues	group	VMMS-1	NMR[36]
Asn1	α -NH ₃	9.5±0.5	7.7±0.1
Asp11	β -COOH	2.7±0.4	3.9±0.05
His22	imidazole	7.5±0.5	6.8±0.1
Glu24	γ -COOH	4.6±0.3	4.1±0.1
Asp27	β -COOH	4.5±0.5	4.0±0.1
Asp40	β -COOH	3.5±0.5	3.6±0.1
Asp46	β -COOH	5.5±0.5	3.8±0.1
Glu51	γ -COOH	5.4±0.5	~4
Arg53	α -COOH	4.6±0.4	3.5±0.1

doi:10.1371/journal.pcbi.1004480.t003

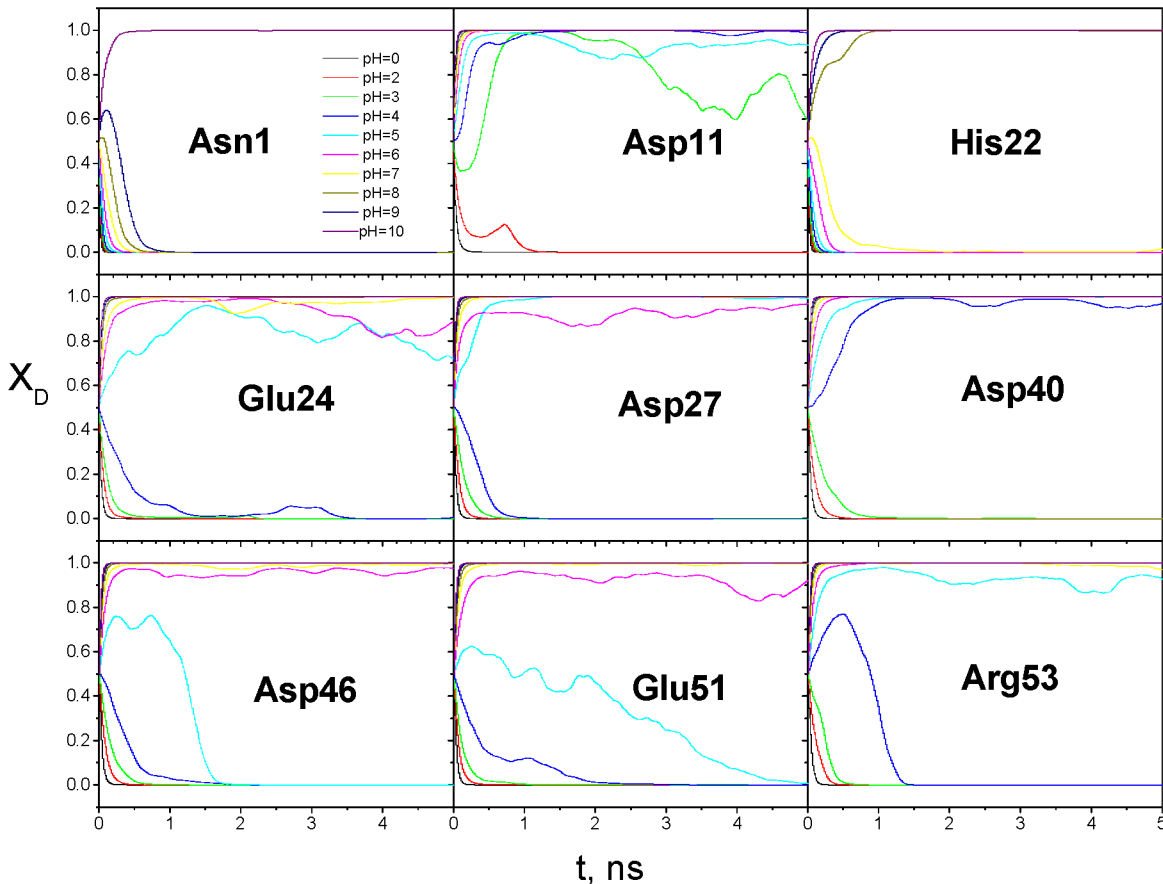


Fig 10. The molar fractions of deprotonated states for the 9 ionizable groups of EPG during the VMMS-1 simulations.

doi:10.1371/journal.pcbi.1004480.g010

examined the hydrogen bonding water molecules during the simulation. Because each state had its own solvent boxes, we can easily see the differences in solute-solvent interactions. In the middle panel of Fig 12, we plot the numbers of water molecules that hydrogen bond to the side-chain of K66. As can be seen, the VMMS simulation started with the same conformations for both states where there were 3 water molecules hydrogen bonding to K66's sidechain. At 10 ns, the hydrogen bonded water in the deprotonated state reduced to 2 and the pKa reduced to between 7 and 8. At 20 ns the hydrogen bonded water of the deprotonated state reduced further to 1 and the pKa decreased to below 4. Then at 30ns, the hydrogen bonding water of the deprotonated state increased to 2 and the pKa reached over 7 again. At 40 ns, the hydrogen bonding water of the deprotonated state reduced to 1 and the pKa value reduced to slightly below the experiment value of 5.6. This tight correlation between the hydrogen bonding water number and the pKa value strongly suggests that water penetration around the K66 sidechain plays an important role in the large deviation of K66's pKa. This result agrees with the high dielectric constant observed in the interior of this protein[40]. Fig 13 shows the protein and the hydrogen bonding water in both states.

The large fluctuation in the pKa of lysine66 shown in Fig 12 indicates that much longer simulation is needed to provide an accurate pKa value. This short simulation only provides a preliminary insight of the system. Thorough understanding of the system is beyond the scope of this paper.

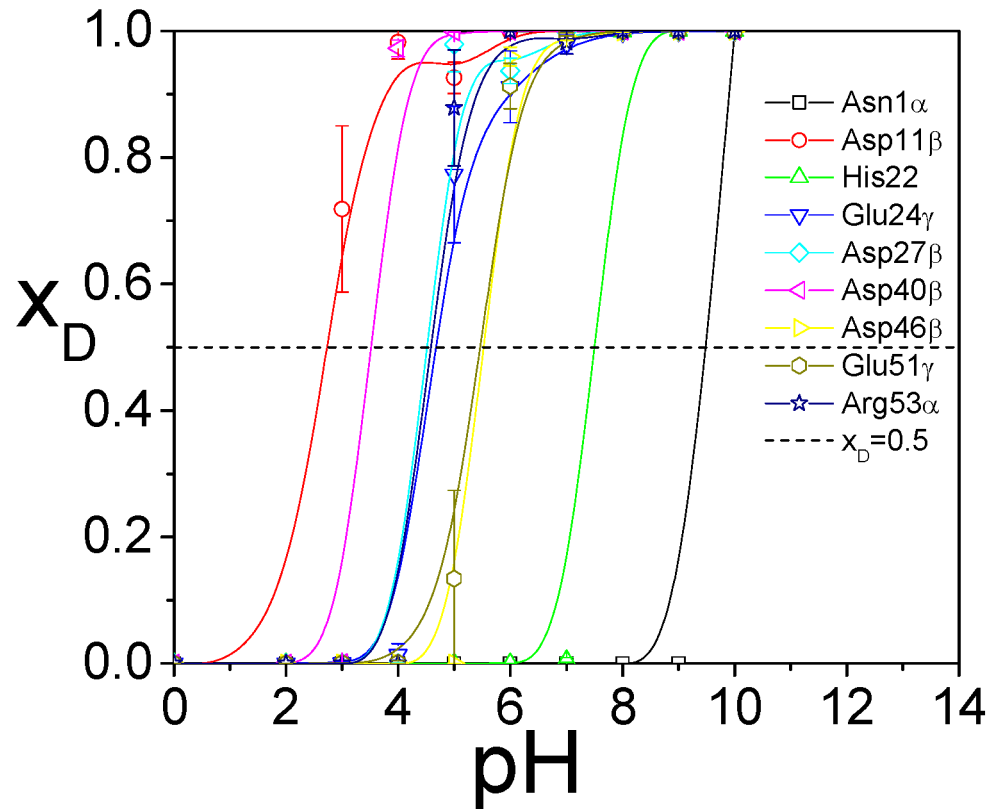


Fig 11. The titration curves of the 9 ionizable groups of EPG as functions of pH values obtained from the VMMS-1 simulations.

doi:10.1371/journal.pcbi.1004480.g011

This example shows the advantage to have explicit solvent for each protonation state in the VMMS method. For single state simulation methods, for example, λ -dynamics, conformations are sampled in a serial manner. The free energy difference between two states can only be calculated after a significant number of transitions between states. While in VMMS, both states are sampled simultaneously, and the free energy difference can be estimated right away. The most striking benefit is when the energy barrier for state transition is high, the VMMS method does not suffer the energy barrier crossing difficulty. In addition, the VMMS method can be combined with other existing accelerated simulation methods, such as the self-guided Langevin dynamics[41–43] to reach the conformation favoring dominant states quickly.

The protonation equilibrium is especially suitable for the VMMS method. By forcing the solute to sample the overlapping conformational space of different states, the VMMS method can accurately estimate the free energy differences between these states. However, when the solute conformation spaces of some states are not completely accessible by other states, alternate techniques or companion methods need to be employed to properly sample all important conformational spaces. Therefore, the VMMS method of current version is limited to cases where the solute conformation of one state is fully accessible by other states. In addition to protonation equilibrium, another such case is oxidation-reduction equilibrium where the oxidized and reduced states are different only by charges.

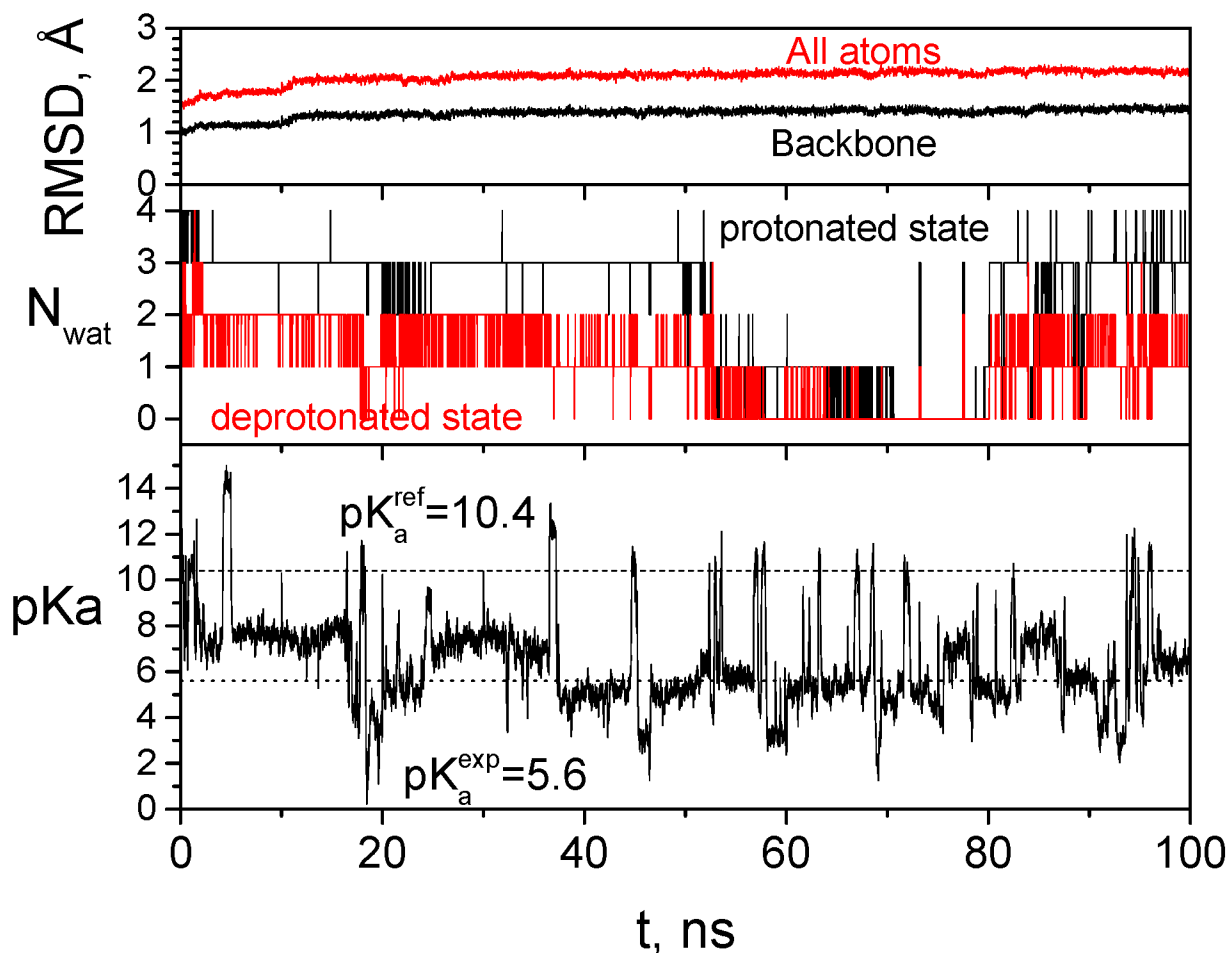


Fig 12. The pKa of K66 and conformational properties of SNase Δ +PHS/V66K during the VMMS simulation.

doi:10.1371/journal.pcbi.1004480.g012

Discussions

This work presents the VMMS simulation method for directly simulating the equilibrium of multiple chemical states. Through including explicit solvent environments of all states, this method avoids the time-consuming solvent reorientation between states, allowing free energy differences between different states to be efficiently estimated during simulation. Using model systems, we examined the conformational distribution of the VMMS ensemble and validated the reweighting formula from the VMMS distributions to pure state distributions.

The VMMS method is applied here to study protonation equilibrium. At a given pH, the VMMS simulation produces equilibrium molar fractions of all states. These molar fractions in turn will affect the potential energy of the VMMS solute and the conformational search. pKa can be determined by identifying the pH where the protonated and deprotonated states have equal molar fractions.

For systems with many titration sites, we employ implicit sites to reduce the number of subsystems in a VMMS simulation. This implicit site treatment allows VMMS simulations scale linearly with the number of titration sites. Through the heptapeptide we demonstrated that the VMMS-1 simulation with one explicit site and two implicit sites produced very similar result to the VMMS simulation with all explicit sites. We applied VMMS-1 simulations to evaluated

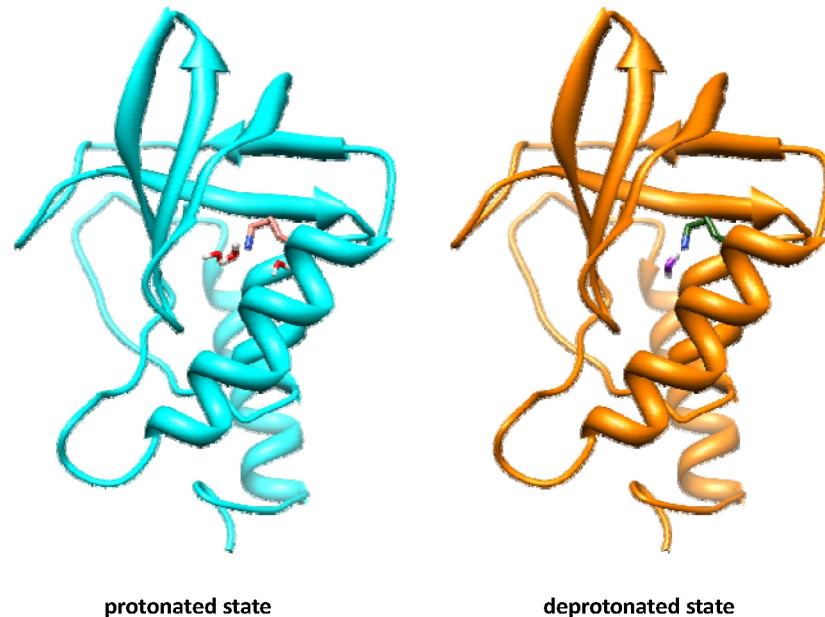


Fig 13. SNase Δ +PHS/V66K in the protonated and deprotonated states after 50 ns VMMS simulation. K66 sidechain has 3 hydrogen bonding water in the protonated state but only 1 hydrogen bonding water in the deprotonated state.

doi:10.1371/journal.pcbi.1004480.g013

pKas of 9 titration sites in mouse epidermal growth factor and the results agree qualitatively with the NMR measurement[36].

Applying the VMMS method to SNase Δ +PHS/V66K in explicit water, we found that the large deviation in lysine66's pKa is likely due to the state-dependent penetration of water into the protein interior. This example demonstrates the VMMS method can properly handle explicit water for this difficult case in pKa calculation.

The VMMS method provides a general approach to study multistate equilibrium. Unlike existing simulation methods that either study state equilibrium indirectly or sample different states sequentially, the VMMS method simulates all states simultaneously and utilizes experiment observable molar fractions as state variables in simulation. It is expected that the VMMS method will find more applications in biological problems related to the equilibrium of multiple states.

Methods

The VMMS system

We design a VMMS system to simulate state equilibrium of a solute in a solvent environment, implicit or explicit. Consider a solute, e.g., a protein, in solution as shown in Fig 14. The solution (shown in the center) contains a solute in various chemical states, e.g., protonated and deprotonated states for ionizable residues. All chemical states of the solute coexist in the solution at equilibrated concentration. 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 represented by a solvated molecule. In other words, every solute is solvated by its own solvent, and solvent molecules only interact with the solute of one state.

Because the conformational distribution and the chemical state of solute are interdependent, conformational sampling should be correlated to the equilibrium of chemical states. The state

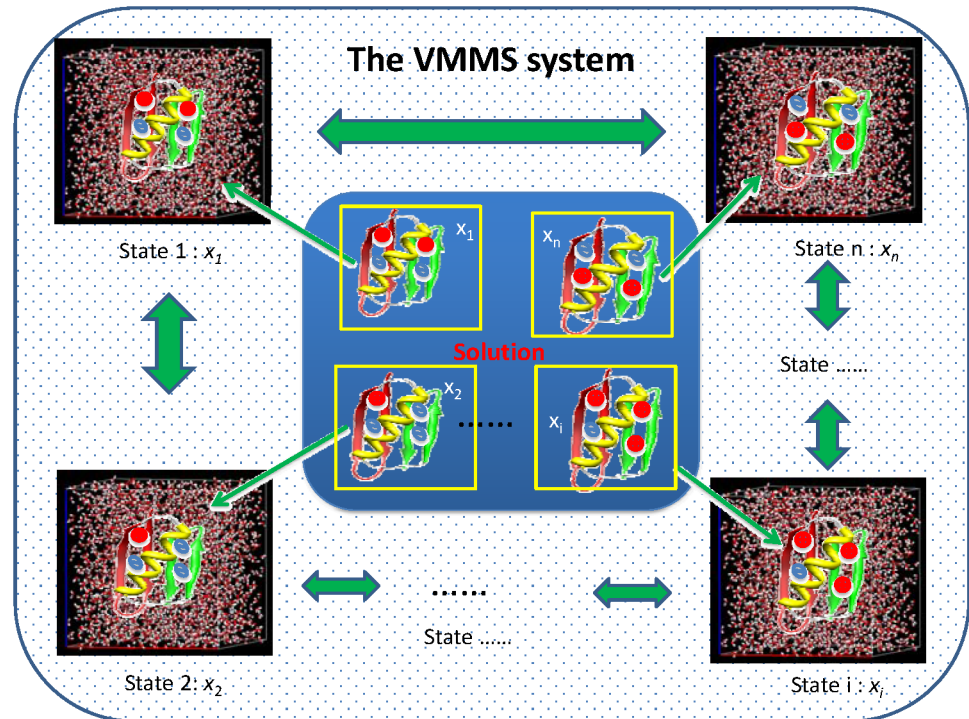


Fig 14. State equilibrium is represented as a virtual mixture of multiple states (VMMS). In a solution shown in the center, all states equilibrate with each other and exist by state molar fractions, x_1, x_2, \dots, x_n . In VMMS, every state is explicitly represented by an all-atom subsystem. Each state has its own solvent environment and solvent can be either implicit or explicit. All subsystems form a virtual ideal solution to equilibrate with each other.

doi:10.1371/journal.pcbi.1004480.g014

equilibrium depends on the state free energies, which can be calculated through adequate sampling of individual states, as well as the overlapping conformational space between states. In this work, we let solute of all states to share the same conformation to enhance the sampling on the overlapping region. On the other hand, we let solvents independent of each other to have adequate sampling of individual solvent states. Here, solvent plays a role to provide solvation to guide solute dynamic simulation. Even though solvent molecules explicitly surround the solute, they behave just like an implicit solvation model to provide solvation energy to the solute. In other words, we use the interaction from the explicit solvent to calculate solvation energy. Using explicit solvent to provide solvation effect for solute dynamic simulation has been explored previously[44]. This design avoids solvent reorientation upon state transition and allows the solute to undergo a state-equilibrated conformational search. The solvent is always in a state-equilibrated condition and provides more accurate solvation than an implicit solvation model.

If the solute has n chemical states, the VMMS system will have n 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 all states share the same conformation and 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. The solute molecules in all subsystems are constrained to have the same conformation and move the same way, but the solvent molecules are not constrained and move under their own interaction environment. The solvents sample their own conformational space to provide equilibrated solvation to their solute. This design makes it very convenient to calculate relative

free energies of the states using Bennett's ratio method [45]. Let's imagine that the n states form an ideal solution and that each state has its own population defined as state molar fractions, x_1, x_2, \dots, x_n , where $x_1 + x_2 + \dots + x_n = 1$. At equilibrium, all states have the same chemical potential, which determine the equilibrium molar fractions of all states. Because these states are not physically mixed to form a real solution as in a mixture simulation [46], the system is the so-called virtual mixture of multiple states (VMMS).

In summary, a VMMS system contains one solute distributing in n states, $P(x_1, x_2, \dots, x_n)$, and n solvents, $W_i, i = 1, 2, \dots, n$, which provide solvation to the n states. The solute exists in all n chemical states by the state molar fractions of x_1, x_2, \dots, x_n . Solvent W_i only interacts with the solute of state i . The VMMS system is denoted as: $P(x_1, x_2, \dots, x_n) + \sum_i^n W_i$. We define a VMMS state j , denoted as V_j , as the solute at the mixed states, $P(x_1, x_2, \dots, x_n)$, in solvent W_j . The difference between a pure state j and a VMMS state j lays at the solute only. In a pure state, the solute exists only in one chemical state, i.e., $P(x_j = 1, x_i \neq j = 0)$, while in the VMMS state, V_j , the solute exists in all chemical states, i.e., $P(x_1, x_2, \dots, x_n)$.

VMMS conformational distribution

For a subsystem j containing a solute, P , of state j and solvent W_j , the potential energy, $E^{(j)}$, can be separated into solute part, $E_p^{(j)}$, and solvent part, $E_w^{(j)}$:

$$E^{(j)} = E_p^{(j)} + E_w^{(j)} \tag{1}$$

The solute interaction is a sum of all pairwise interactions involving the solute, including solute-solute and solute-solvent interactions:

$$E_p^{(j)} = \sum_{a \in P} E_a^{(j)} = \frac{1}{2} \sum_{a \in P} \sum_i^{N^{(j)}} \epsilon_{ai}^{(j)} \tag{2}$$

Here, $N^{(j)}$ is the number of 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:

$$E_w^{(j)} = \sum_{a \in W_j} E_a^{(j)} = \frac{1}{2} \sum_{a \in W_j} \sum_i^{N^{(j)}} E_{ai}^{(j)} \tag{3}$$

For a VMMS system, $P(x_1, x_2, \dots, x_n) + \sum_i^n W_i$, there are n subsystems, one for each state. The VMMS system has the solvents of all subsystems, but only 1 solute, distributing in all states by the state molar fractions. Therefore, the VMMS potential energy has the following form:

$$E^{(VMMS)} = \sum_i^n (x_i E_p^{(i)} + E_w^{(i)}) \tag{4}$$

The VMMS state j , V_j , contains the VMMS solute and the solvent interacting with the solute at state j . The potential energy of V_j is a sum of the solute potential energy and the solvent

potential energy in subsystem j :

$$E^{(V_j)} = \sum_i^n (x_i E_p^{(i)}) + E_W^{(j)} \quad (5)$$

which is different from the potential energy of the pure state j , Eq (1), in the solute part.

The partition function of the VMMS ensemble is:

$$Q^{(VMMS)} = \sum_{\Omega^{(VMMS)}} \exp(-\beta(E^{(VMMS)})) = \sum_{\Omega^{(VMMS)}} \prod_i^n \exp(-\beta(x_i E_p^{(i)} + E_W^{(i)})) \quad (6)$$

where $\Omega^{(VMMS)}$ represents the conformational space of the VMMS system and $\beta = \frac{1}{kT}$.

When $x_j = 1$ and $x_i \neq j = 0$, the VMMS system becomes a pure state j plus all other solvents, $W_i \neq j$. The partition function can be written to the following form:

$$Q^{(VMMS)}(x_j = 1, x_{i \neq j} = 0) = \sum_{\Omega^{(VMMS)}} \exp(-\beta(E_p^{(j)} + E_W^{(j)})) \prod_{i \neq j}^n \exp(-\beta E_W^{(i)}) \quad (7)$$

Therefore, we have the relation between VMMS ensembles at (x_1, x_2, \dots, x_n) and at $(x_j = 1, x_i \neq j = 0)$:

$$\begin{aligned} \frac{Q^{(VMMS)}(x_j = 1)}{Q^{(VMMS)}(x_1, x_2, \dots, x_n)} &= \frac{\sum_{\Omega^{(VMMS)}} \exp(-\beta(E_p^{(V_j)} + \sum_i^n E_W^{(i)}))}{\sum_{\Omega^{(VMMS)}} \exp(-\beta \sum_i^n (x_i E_p^{(i)} + E_W^{(i)}))} \\ &= \frac{\sum_{\Omega^{(VMMS)}} \exp(-\beta(E_p^{(j)} - \sum_i^n x_i E_p^{(i)})) \exp(-\beta \sum_i^n (x_i E_p^{(i)} + E_W^{(i)}))}{\sum_{\Omega^{(VMMS)}} \exp(-\beta \sum_i^n (x_i E_p^{(i)} + E_W^{(i)}))} \\ &= \langle \exp(-\beta(E_p^{(j)} - \sum_i^n x_i E_p^{(i)})) \rangle_{VMMS} \\ &= \langle \exp(-\beta(E^{(j)} - E^{(V_j)})) \rangle_{VMMS} \end{aligned} \quad (8)$$

Eq (8) shows that the VMMS ensemble average of property P at $(x_j = 1, x_i \neq j = 0)$ can be obtained in a VMMS simulation at (x_1, x_2, \dots, x_n) through reweighting by a factor:

$$w_j = \exp(-\beta(E^{(j)} - E^{(V_j)})) = \exp(-\beta \sum_i^n x_i (E_p^{(j)} - E_p^{(i)})) \quad (9)$$

Because at $(x_j = 1, x_i \neq j = 0)$ the solute energy depends only on solvent W_j and is independent of all other solvents, it is easy to prove that the VMMS ensemble average of any state j

related property, P_j , at ($x_j = 1, x_i \neq j = 0$) equals the canonical ensemble average at pure state j :

$$\begin{aligned} \langle P_j \rangle_{\text{VMMS}(x_j=1)} &= \frac{\sum_{\Omega(\text{VMMS})} P_j \exp(-\beta(E_p^{(j)} + E_w^{(j)})) \prod_{i \neq j}^n \exp(-\beta E_w^{(i)})}{\sum_{\Omega(\text{VMMS})} \exp(-\beta(E_p^{(j)} + E_w^{(j)})) \prod_{i \neq j}^n \exp(-\beta E_w^{(i)})} \\ &= \frac{\sum_{\Omega(\text{VMMS})} P_j \exp(-\beta(E_p^{(j)} + E_w^{(j)})) \prod_{i \neq j}^n \exp(-\beta E_w^{(i)})}{\sum_{\Omega(\text{VMMS})} \exp(-\beta(E_p^{(j)} + E_w^{(j)})} \frac{\sum_{\Omega(\text{VMMS})} \exp(-\beta(E_p^{(j)} + E_w^{(j)}))}{\sum_{\Omega(\text{VMMS})} \exp(-\beta(E_p^{(j)} + E_w^{(j)})) \prod_{i \neq j}^n \exp(-\beta E_w^{(i)})} \\ &= \frac{\langle P_j \prod_{i \neq j}^n \exp(-\beta E_w^{(i)}) \rangle_j}{\langle \prod_{i \neq j}^n \exp(-\beta E_w^{(i)}) \rangle_j} = \frac{\langle P_j \rangle_j \langle \prod_{i \neq j}^n \exp(-\beta E_w^{(i)}) \rangle_j}{\langle \prod_{i \neq j}^n \exp(-\beta E_w^{(i)}) \rangle_j} = \langle P_j \rangle_j \end{aligned}$$

The separation of P_j and $\prod_{i \neq j}^n \exp(-\beta E_w^{(i)})$ is because they are independent of each other at ($x_j = 1, x_i \neq j = 0$). Therefore, the ensemble average of any state j related property, P_j , in pure state j can be reweighted from a VMMS simulation by:

$$\langle P_j \rangle_j = \langle P_j \rangle_{\text{VMMS}(x_j=1)} = \frac{\langle P_j w_j \rangle_{\text{VMMS}}}{\langle w_j \rangle_{\text{VMMS}}} \quad (10)$$

Using 3D IPS in VMMS simulation

The reweighting in VMMS simulation as shown in Eqs (9) and (10) need solute energies, which can be efficiently estimated from pairwise interactions by Eq (2). The isotropic periodic sum (IPS) method provides a pairwise way to accurately calculate long range interactions [24,47–50]. The use of IPS here can simplify or eliminate some of the problems encountered with Ewald methods. Specifically, the use of IPS allows the partition of energy components in a manner that is not possible with PME, unless a separate PME calculation is done for each state and for solute and solvent separately. In the result section, we validate the IPS accuracy through free energy calculation of model compounds.

The IPS potential for polar electrostatic interaction is[24]:

$$\epsilon_{\text{ele}}^{\text{IPSP}}(r, R_c) = \begin{cases} \frac{q_i q_j}{r} \left(1 - \frac{19}{16} \left(\frac{r}{R_c} \right) + \frac{35}{16} \left(\frac{r}{R_c} \right)^3 - \frac{21}{16} \left(\frac{r}{R_c} \right)^5 + \frac{5}{16} \left(\frac{r}{R_c} \right)^7 \right) & r \leq R_c \\ 0 & r > R_c \end{cases} \quad (11)$$

where R_c is the local region radius, commonly called cutoff distance. q_i and q_j are charges at atoms i and j , respectively. The total electrostatic interaction energy can be calculated as simple as typical cutoff methods:

$$E_{\text{ele}}^{(\text{IPS})}(\mathbf{r}^N, R_c) = \frac{1}{2} \sum_i^N \sum_j^N \epsilon_{\text{ele}}^{(\text{IPSP})}(r_{ij}, R_c) = \frac{1}{2} \sum_i^N \sum_{r_{ij} < R_c} \epsilon_{\text{ele}}^{(\text{IPSP})}(r_{ij}, R_c) \quad (12)$$

Here, we use the neutral charge assumption so that the total IPS boundary electrostatic energy is zero. For Lennard-Jones (L-J) energy, the IPS potentials for the dispersion and repulsive

parts are:

$$\epsilon_{\text{disp}}^{3\text{D-IPS}}(r, R_c) \approx -\frac{C}{r^6} \left(1 + \frac{7}{16} \left(\frac{r}{R_c}\right)^6 + \frac{9}{14} \left(\frac{r}{R_c}\right)^8 - \frac{3}{28} \left(\frac{r}{R_c}\right)^{10} + \frac{6}{7} \left(\frac{r}{R_c}\right)^{12} \right) \quad (13)$$

$$\epsilon_{\text{rep}}^{3\text{D-IPS}}(r, R_c) \approx \frac{A}{r^{12}} \left(1 + \frac{5}{787} \left(\frac{r}{R_c}\right)^{12} + \frac{9}{26} \left(\frac{r}{R_c}\right)^{16} - \frac{3}{13} \left(\frac{r}{R_c}\right)^{20} + \frac{27}{26} \left(\frac{r}{R_c}\right)^{24} \right) \quad (14)$$

Here C and A are dispersion and repulsion constants, respectively, for L-J potential. The pairwise L-J IPS potential is:

$$\epsilon_{\text{LJ}}^{3\text{D-IPSp}}(r, R_c) = \begin{cases} \epsilon_{\text{disp}}^{3\text{D-IPSp}}(r, R_c) + \epsilon_{\text{rep}}^{3\text{D-IPSp}}(r, R_c) - \epsilon_{\text{disp}}^{3\text{D-IPSp}}(R_c, R_c) - \epsilon_{\text{rep}}^{3\text{D-IPSp}}(R_c, R_c) & r \leq R_c \\ 0 & r > R_c \end{cases} \quad (15)$$

And the total L-J interaction is a sum over all atom pairs:

$$E_{\text{LJ}}^{(\text{IPS})}(\mathbf{r}^N, R_c) = \frac{1}{2} \sum_i^N \sum_{r_{ij} < R_c} \epsilon_{\text{LJ}}^{(\text{IPS})}(r_{ij}, R_c) + \frac{4\pi R_c^3}{3NV} \sum_i^N \sum_j^N (\epsilon_{\text{disp}}^{(\text{IPS})}(R_c, R_c) + \epsilon_{\text{rep}}^{(\text{IPS})}(R_c, R_c)) \quad (16)$$

The first term is the pairwise IPS potential and the second term is the L-J IPS boundary energy, which is constant during a NVT simulation.

Using pairwise potentials is not only convenient for separating solute and solvent interactions, required by Eqs (1-3), but also helps avoid repeated calculation for the same conformation at different states, which is essential for free energy calculation. For the same conformation, energies of different states can be calculated efficiently with a small amount of overhead:

$$E_{\text{ele}}^{(\text{IPSp})}(q^{(k)}) = \frac{1}{2} \sum_i^{N_{\text{nc}}} \sum_j^{N_{\text{nc}}} \epsilon_{\text{ele}}^{(\text{IPSp})}(r_{ij}, R_c) + \sum_i^{N_c} q_i^{(k)} \sum_j^{N_{\text{nc}}} q_j \hat{\epsilon}_{\text{ele}}^{(\text{IPSp})}(r_{ij}, R_c) + \frac{1}{2} \sum_i^{N_c} q_i^{(k)} \sum_j^{N_c} q_j^{(k)} \hat{\epsilon}_{\text{ele}}^{(\text{IPSp})}(r_{ij}, R_c) \quad (17)$$

where N_{nc} is the number of non-changing atoms and N_c is the number of changing atoms, $\hat{\epsilon}_{ij}$ is the unit charge interaction, $\hat{\epsilon}_{ij} = \epsilon_{ij}/(q_i q_j)$, and $q_i^{(k)}$ is the charge of atom i at state k . The first term is for atom pairs of unchanged charges, which costs the most, and does not need to be recalculated for different charge states.

Free energy differences between states

The VMMS simulation samples conformations at all states, providing sufficient information to calculate free energy differences between all states. Because the solute is not in either of the pure states, reweighting is needed to obtain conformational distributions at the pure states.

Fig 15 shows the free energy calculation diagram. To calculate free energy between state 0 and state 1, we reweight the VMMS conformation distribution to state 0 and state 1 so that we can apply Bennett's ratio method[45]. Here, we modify Bennett's Ratio method for the VMMS simulation:

$$\Delta G_{01} = C_{01} - kT \ln \frac{\langle f(\beta(E^{(1)} - E^{(0)} - C_{01})) \exp(-\beta(E^{(0)} - E^{(V0)}) \rangle_{\text{VMMS}} \langle \exp(-\beta(E^{(1)} - E^{(V1)}) \rangle_{\text{VMMS}}}{\langle f(\beta(E^{(0)} - E^{(1)} + C_{01})) \exp(-\beta(E^{(1)} - E^{(V1)}) \rangle_{\text{VMMS}} \langle \exp(-\beta(E^{(0)} - E^{(V0)}) \rangle_{\text{VMMS}}} \quad (18)$$

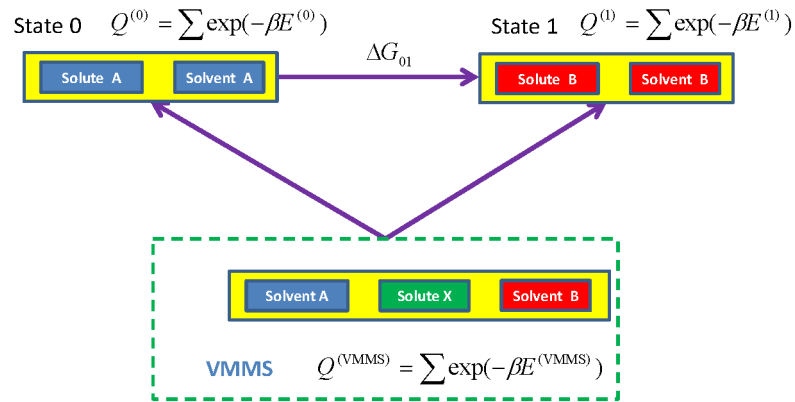


Fig 15. Calculation diagram for state free energy differences. The conformational distributions of state 0 and 1 can be obtained through reweighting from the VMMS simulation.

doi:10.1371/journal.pcbi.1004480.g015

where the Fermi function is defined as:

$$f(x) = \frac{1}{1 + \exp(x)} \quad (19)$$

The VMMS simulation can be performed with enhanced sampling methods, such as self-guided Langevin dynamics[41,42,51,52], to accelerate the conformation search. When applying SGLD, the SGLD reweighting factors, w_{SG0} and w_{SG1} , should be incorporated into Eq (18):

$$\Delta G_{01} = C_{01} - kT \ln \frac{\langle f(\beta(E^{(1)} - E^{(0)} - C_{01}))w_{SG0} \exp(-\beta(E^{(0)} - E^{(V0)})) \rangle_{VMMS} \langle w_{SG1} \exp(-\beta(E^{(1)} - E^{(V1)})) \rangle_{VMMS}}{\langle f(\beta(E^{(0)} - E^{(1)} + C_{01}))w_{SG1} \exp(-\beta(E^{(1)} - E^{(V1)})) \rangle_{VMMS} \langle w_{SG0} \exp(-\beta(E^{(0)} - E^{(V0)})) \rangle_{VMMS}} \quad (20)$$

This equation provides a useful connection between an accelerated VMMS simulation and the desired canonical ensemble. If choosing SGLDfp[42] or SGLD-GLE[43] where $w_{SG} = 1$, one can simply use Eq (18) to calculate relative free energies.

State equilibrium

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

$$\mu_i = \frac{\partial G}{\partial n_i} = \mu_i^0 + kT \ln x_i \quad (21)$$

Where μ_i^0 is the standard chemical potential of state i. At equilibrium, the chemical potential of all states equal: $\mu_i = \mu_j$, therefore, we have:

$$\frac{x_j}{x_i} = \exp\left(-\frac{\mu_j^0 - \mu_i^0}{kT}\right) = \exp\left(-\frac{\Delta\mu_{ij}^0}{kT}\right) \quad (22)$$

For multiple states, we use the following formula to derive molar fractions from the standard chemical potential differences:

$$x_k = \left(\sum_j^n \exp\left(-\frac{\Delta\mu_{kj}^0}{kT}\right) \right)^{-1} / \sum_i^n \left(\sum_j^n \exp\left(-\frac{\Delta\mu_{ij}^0}{kT}\right) \right)^{-1} \quad (23)$$

For protonation equilibrium, the standard chemical potential difference contains many contributions such as those from molecular interactions and those from quantum mechanics of protonation. All contributions other than those from molecular 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. [1,3–6,53] At a given pH, the standard chemical potential difference of deprotonation can be calculated from the state free energy difference:

$$\Delta\mu_{ij}^0 = \Delta G_{ij} - kT(pH - pK_a^{\text{ref}})\ln 10 - \Delta G_{ij}^{\text{ref}} \quad (24)$$

Here, pK_a^{ref} is the experimental pKa of a model compound for the corresponding amino acid to change from the protonated state, i , to the deprotonated state, j , and $\Delta G_{ij}^{\text{ref}}$ is the free energy difference calculated from simulation of the model compound. Using Eqs (23) and (24), one can calculate equilibrium molar fractions of all states. Based on the definition of pKa, from equilibrium molar fractions we can calculate pKa:

$$pKa = -\lg \frac{x_D[H]}{x_H} = pH + px_D - px_H \quad (25)$$

where x_H and x_D are the molar fractions of the protonated and deprotonated states, respectively. Typically, in constant pH simulation, pKa is determined by the pH value where $x_H = x_D$. pKa can also be directly calculated from a VMMS simulation at equal state molar fractions ($x_H = x_D$) where $\Delta\mu_{ij} = \Delta\mu_{ij}^0 = 0$ and $pKa = pH$. From Eq (24), we have:

$$pKa = pKa^{\text{ref}} + \frac{\Delta G(x_H = x_D) - \Delta G^{\text{ref}}}{kT \ln 10} \quad (26)$$

Using implicit sites to handle systems with a large number of titration sites

VMMS uses explicit chemical states to efficiently sample the conformations at these states and avoid difficulties related to state transitions. Because each titration site has two states, for a system with m titration sites there would be as many as 2^m states. Obviously, it is impractical to include explicitly all protonation states in a VMMS simulation when m is large, say, more than 6. To circumvent this burden, we propose a divide-and-conquer approach to explicitly consider only one or a few titration sites at a time and treat the remaining sites implicitly to have combined states. This implicit site approximation is the same, in spirit, to the titration coordinate when running λ -dynamics for constant pH simulations [4,6], where the titration coordinate λ represents a combination of protonated and deprotonated states. In other words, in λ -dynamics all ionizable sites are treated as a mixture of $1-\lambda$ protonated and λ deprotonated states, just like an implicit site described here. The titration of an ionizable site in λ -dynamics corresponds

to a transition of λ between 0 and 1. But in VMMS, the titration of an ionizable site corresponds to a change of molar fractions, which only happens when an ionizable site is explicit.

A titration site is explicit when it has distinct protonated and deprotonated states, which differ by their atomic charges. To make a titration site implicit, its protonated and deprotonated states are converted to a single state where the atomic charges are a linear combination of the charges in the two explicit states.

$$q_a^{(im)} = (1 - x_D)q_a^{(H)} + x_Dq_a^{(D)}. \quad (27)$$

Here, $q_a^{(H)}$, $q_a^{(D)}$ are atomic charges of atom a in the protonated and deprotonated states, respectively, and $q_a^{(im)}$ is the atomic charge in the implicit site. When a site is implicit, all atoms in the site are implicit and their charges are defined by Eq (27). When a titration site changes from implicit to explicit, its atomic charges return to $q_a^{(H)}$ for the protonated state and to $q_a^{(D)}$ for the deprotonated state.

During a VMMS simulation, one or a few titration sites are chosen to be explicit and the rest implicit. The molar fractions of the explicit sites are updated during simulation according to Eqs (23) and (24). After this explicit simulation period, the explicit sites are converted to implicit based on Eq (27) and one or a few other implicit sites are chosen to be converted to explicit for the next explicit simulation period. This process can loop repeatedly through all titration sites until the molar fractions of all sites converge.

For a system of m titration sites, a VMMS simulation can be carried out with k explicit titration sites ($0 < k \leq m$) and $m - k$ implicit titration sites. We use VMMS- k to characterize a VMMS simulation by the number of explicit sites. Without mentioning the explicit site number, a VMMS simulation means a VMMS- m simulation where all titration sites are explicit, $k = m$. For example, for a protein of 3 titration sites, a VMMS-1 simulation means that at any moment during the simulation, there is one explicit site and two implicit sites, while a VMMS-3 or VMMS simulation means that all 3 sites are explicit. For a VMMS- k simulation, the system contains 2^k subsystems, one for each explicit state. Because the k explicit sites will loop through the m sites, the cost of a VMMS- k simulation is $O(2^k m / k)$, which scales linearly with the number of titration sites, m . This implicit site treatment is optimal for parallel computing where different explicit sites can be simulated in parallel and their results, x_D , will be used to update the atomic charges according to Eq (27) for simulations where the sites are implicit.

Simulation details

The VMMS method is carried out as parallel simulations of a series of subsystems, one for each explicit chemical state. Energies and forces are calculated as in normal dynamic simulations. Before integrating the equation of motion, the solute forces are replaced by the weighted combination according to the state molar fractions, while the solvent forces remain unmodified:

$$f_a^{(VMMS)} = \begin{cases} \sum_i^n (x_i f_a^{(i)}) & a \in P \\ f_a^{(i)} & a \notin P \end{cases} \quad (28)$$

Using the VMMS forces to integrate the equation of motion for either MD, LD, or SGLD simulations, we propagate the conformation to sample the conformational space. The VMMS forces shown in Eq (28) will not conserve the total energy for each subsystem. To maintain

energy conservation for each subsystem, we scale the solvent velocities in the following way:

$$\mathbf{v}_a^{(i)} = \begin{cases} \mathbf{v}'_a & a \in P \\ \mathbf{v}'_a \sqrt{1 - \frac{\delta^{(i)}}{\sum_a \frac{1}{2} m_a v_a^2}} & a \notin P \end{cases} \quad (29)$$

where \mathbf{v}'_a is the velocity after current time step and $\delta^{(i)} = \sum_{a \in P} (\mathbf{f}_a^{(VMMS)} - \mathbf{f}_a^{(i)}) \mathbf{v}'_a \delta t$ is the extra work of the VMMS forces do to subsystem i . The solute experiences the same forces and moves exactly the same way in all subsystems. To avoid numerical errors, solute movement is performed only in one subsystem and the coordinates are broadcasted to all subsystems. Other than these, all procedures are the same as that in regular dynamics simulations.

All simulations presented here were performed with CHARMM[54,55]. The all-atom CHARMM force field[56] was used for energy calculation. Excepted noted otherwise, all simulations were performed in a constant volume and a constant temperature of 300K using the SGLD-GLE method[43] with a local averaging time of $t_L = 0.2$ ps, a guiding factor of $\lambda = 1$, and a friction constant of $\xi = 10$ /ps. A time step of 1 fs was used and SHAKE algorithm[57] was employed to fix the hydrogen connecting bond lengths.

Six typical ionizable residues as well as two terminal groups were examined here. They are aspartic acid (ASP or D), glutamic acid (GLU or E), lysine (LYS or K), arginine (ARG or R), histidine (HIS $^\delta$ and HIS $^\epsilon$ or H $^\delta$ and H $^\epsilon$), and tyrosine (TYR or Y), N-terminal group (Nter) and C-terminal group (Cter). Their model compounds were built by adding an acetyl (ACE) group to the N-terminal and a methyl amide (NME) group to the C-terminal, except for the terminal groups whose model compounds were built by attaching them to an alanine residue at one terminal and a block group at the other terminal.

The deprotonated states of ionizable residues have a dummy hydrogen so that both states have the same number of atoms. The dummy hydrogen atom is identical to a hydrogen atom except that it has no charge. Fig 16 shows the atomic charges of these ionizable residues in their protonated and deprotonated states.

The VMMS systems were built by immersing a solute molecule in a box of water and any water whose oxygen distance to a solute atom was less than 2.4 angstroms was removed. It should be noted that different states can have different number of solvent molecules and/or ions. In this work, the same starting conformation was used for all states. During VMMS simulations, the Fermi functions and related weighting factors in Eq (18) were evaluated every 10 fs and free energies were calculated every 0.5 ps. During simulations, we used the following scheme to obtain local average free energies:

$$\tilde{P}_n = (1 - \frac{1}{L})\tilde{P}_{n-1} + \frac{1}{L}P_n \quad (30)$$

where L is the local averaging period. In the simulations reported here, we used $L = 200$.

For VMMS simulations at constant pH, molar fractions were updated every 0.5 ps according to Eq (23). For VMMS-1 simulations, one ionizable site was explicitly simulated for 10 ps before being converted to implicit. Of the 10 ps explicit simulation period, 2 ps were used for equilibrium and the remaining 8 ps were used for free energy and state molar fraction calculation. When a site was implicit, its state molar fractions remained unchanged.

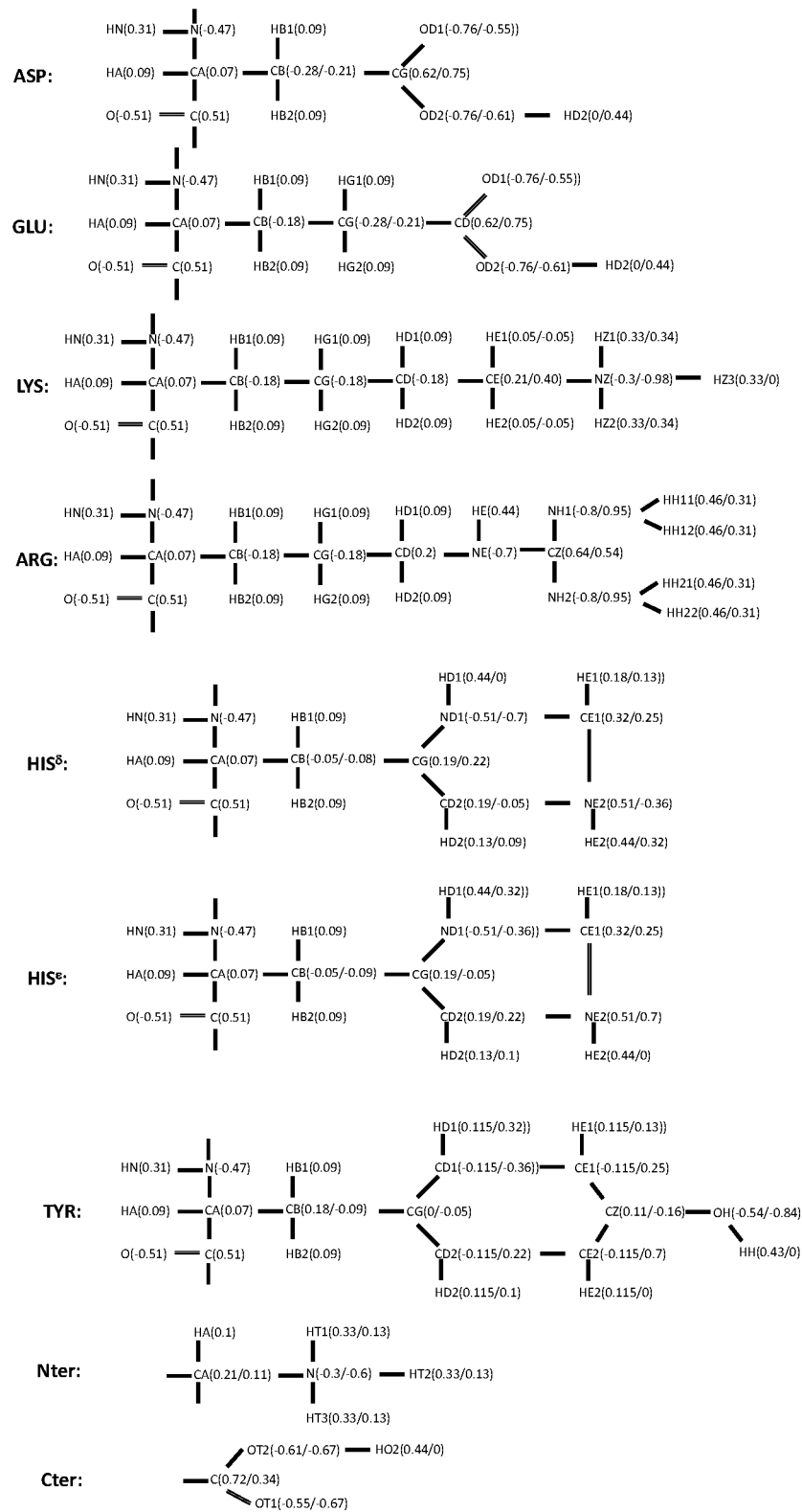


Fig 16. Atomic charges in the ionizable residues in the protonated/deprotonated states. Dummy hydrogen atoms with zero charge are added in the deprotonated states.

doi:10.1371/journal.pcbi.1004480.g016

Acknowledgments

We thank Dr. Juyong Lee, Dr. Ana Damjanovic, and Prof. Wei Yang for thoughtful discussions, and thank Eunice Wu for proof reading the manuscript.

Author Contributions

Conceived and designed the experiments: XW BRB. Performed the experiments: XW. Analyzed the data: XW BRB. Contributed reagents/materials/analysis tools: XW. Wrote the paper: XW BRB.

References

- Burgi R, Kollman PA, Van Gunsteren WF (2002) Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *Proteins* 47: 469–480. PMID: [12001225](#)
- Walczak AM, Antosiewicz JM (2002) Langevin dynamics of proteins at constant pH. *Phys Rev E Stat Nonlin Soft Matter Phys* 66: 051911. PMID: [12513527](#)
- Dlugosz M, Antosiewicz JM, Robertson AD (2004) 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* 69: 021915. PMID: [14995499](#)
- Lee MS, Salsbury FR Jr., Brooks CL 3rd (2004) Constant-pH molecular dynamics using continuous titration coordinates. *Proteins* 56: 738–752. PMID: [15281127](#)
- Mongan J, Case DA, McCammon JA (2004) Constant pH molecular dynamics in generalized Born implicit solvent. *J Comput Chem* 25: 2038–2048. PMID: [15481090](#)
- Khandogin J, Brooks CL 3rd (2005) Constant pH molecular dynamics with proton tautomerism. *Biophys J* 89: 141–157. PMID: [15863480](#)
- Machuqueiro M, Baptista AM (2006) Constant-pH molecular dynamics with ionic strength effects: protonation-conformation coupling in decalysine. *J Phys Chem B* 110: 2927–2933. PMID: [16471903](#)
- Stern HA (2007) Molecular simulation with variable protonation states at constant pH. *J Chem Phys* 126: 164112. PMID: [17477594](#)
- Campos SR, Machuqueiro M, Baptista AM (2010) Constant-pH molecular dynamics simulations reveal a beta-rich form of the human prion protein. *J Phys Chem B* 114: 12692–12700. doi: [10.1021/jp104753t](#) PMID: [20843095](#)
- Meng Y, Roitberg AE (2010) Constant pH replica exchange molecular dynamics in biomolecules using a discrete protonation model. *J Chem Theory Comput* 6: 1401–1412. PMID: [20514364](#)
- Donnini S, Tegeler F, Groenhof G, Grubmuller H (2011) Constant pH Molecular Dynamics in Explicit Solvent with lambda-Dynamics. *J Chem Theory Comput* 7: 1962–1978. PMID: [21687785](#)
- Itoh SG, Damjanovic A, Brooks BR (2011) pH replica-exchange method based on discrete protonation states. *Proteins* 79: 3420–3436. doi: [10.1002/prot.23176](#) PMID: [22002801](#)
- Chu WT, Wu YJ, Zhang JL, Zheng QC, Chen L, et al. (2012) Constant pH molecular dynamics (CpHMD) and mutation studies: insights into AaegOBP1 pH-induced ligand releasing mechanism. *Biochim Biophys Acta* 1824: 913–918. doi: [10.1016/j.bbapap.2012.03.008](#) PMID: [22575088](#)
- Goh GB, Knight JL, Brooks CL 3rd (2012) Constant pH Molecular Dynamics Simulations of Nucleic Acids in Explicit Solvent. *J Chem Theory Comput* 8: 36–46. PMID: [22337595](#)
- Sabri Dashti D, Meng Y, Roitberg AE (2012) pH-replica exchange molecular dynamics in proteins using a discrete protonation method. *J Phys Chem B* 116: 8805–8811. doi: [10.1021/jp303385x](#) PMID: [22694266](#)
- Chen W, Wallace JA, Yue Z, Shen JK (2013) Introducing titratable water to all-atom molecular dynamics at constant pH. *Biophys J* 105: L15–17. doi: [10.1016/j.bpj.2013.06.036](#) PMID: [23972860](#)
- Carvalheda CA, Campos SR, Machuqueiro M, Baptista AM (2013) Structural effects of pH and deacylation on surfactant protein C in an organic solvent mixture: a constant-pH MD study. *J Chem Inf Model* 53: 2979–2989. doi: [10.1021/ci400479c](#) PMID: [24106805](#)
- Goh GB, Hulbert BS, Zhou H, Brooks CL 3rd (2014) Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins* 82: 1319–1331. doi: [10.1002/prot.24499](#) PMID: [24375620](#)
- Lee J, Miller BT, Damjanovic A, Brooks BR (2014) Constant pH Molecular Dynamics in Explicit Solvent with Enveloping Distribution Sampling and Hamiltonian Exchange. *J Chem Theory Comput* 10: 2738–2750. PMID: [25061443](#)

20. Swails JM, York DM, Roitberg AE (2014) Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J Chem Theory Comput* 10: 1341–1352. PMID: [24803862](#)
21. Christ CD, van Gunsteren WF (2007) Enveloping distribution sampling: a method to calculate free energy differences from a single simulation. *J Chem Phys* 126: 184110. PMID: [17508795](#)
22. Christ CD, van Gunsteren WF (2008) Multiple free energies from a single simulation: extending enveloping distribution sampling to nonoverlapping phase-space distributions. *J Chem Phys* 128: 174112. doi: [10.1063/1.2913050](#) PMID: [18465915](#)
23. Christ CD, Van Gunsteren WF (2009) Comparison of three enveloping distribution sampling Hamiltonians for the estimation of multiple free energy differences from a single simulation. *J Comput Chem* 30: 1664–1679. doi: [10.1002/jcc.21325](#) PMID: [19504591](#)
24. Wu X, Brooks BR (2009) Isotropic periodic sum of electrostatic interactions for polar systems. *J Chem Phys* 131: 024107. doi: [10.1063/1.3160730](#) PMID: [19603970](#)
25. Ojeda-May P, Pu J (2013) Isotropic Periodic Sum Treatment of Long-Range Electrostatic Interactions in Combined Quantum Mechanical and Molecular Mechanical Calculations. *Journal of Chemical Theory and Computation* 10: 134–145.
26. Takahashi KZ, Narumi T, Yasuoka K (2011) Cutoff radius effect of the isotropic periodic sum and Wolf method in liquid-vapor interfaces of water. *J Chem Phys* 134: 174112. doi: [10.1063/1.3578473](#) PMID: [21548678](#)
27. Takahashi K, Narumi T, Yasuoka K (2011) Cut-off radius effect of the isotropic periodic sum method for polar molecules in a bulk water system. *Molecular Simulation Online*.
28. Takahashi K, Narumi T, Yasuoka K (2010) Cutoff radius effect of the isotropic periodic sum method in homogeneous system. II. Water. *J Chem Phys* 133: 014109. doi: [10.1063/1.3462241](#) PMID: [20614961](#)
29. Venable RM, Chen LE, Pastor RW (2009) Comparison of the extended isotropic periodic sum and particle mesh Ewald methods for simulations of lipid bilayers and monolayers. *J Phys Chem B* 113: 5855–5862. doi: [10.1021/jp900843x](#) PMID: [19351117](#)
30. Ojeda-May P, Pu J (2014) Assessing the accuracy of the isotropic periodic sum method through Madelung energy computation. *J Chem Phys* 140: 164106. doi: [10.1063/1.4871871](#) PMID: [24784252](#)
31. Takahashi K, Yasuoka K, Narumi T (2007) Cutoff radius effect of isotropic periodic sum method for transport coefficients of Lennard-Jones liquid. *J Chem Phys* 127: 114511. PMID: [17887861](#)
32. Takahashi KZ (2014) Design of a reaction field using a linear-combination-based isotropic periodic sum method. *J Comput Chem* 35: 865–875. doi: [10.1002/jcc.23562](#) PMID: [24615639](#)
33. Steinbach PJ, Brooks BR (1994) New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *Journal of Computational Chemistry* 15: 667–683.
34. Lee MS, Feig M, Salsbury FR Jr., Brooks CL 3rd (2003) New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J Comput Chem* 24: 1348–1356. PMID: [12827676](#)
35. Dlugosz M, Antosiewicz JM (2005) Effects of solute-solvent proton exchange on polypeptide chain dynamics: a constant-pH molecular dynamics study. *J Phys Chem B* 109: 13777–13784. PMID: [16852726](#)
36. Kohda D, Sawada T, Inagaki F (1991) Characterization of pH titration shifts for all the nonlabile proton resonances a protein by two-dimensional NMR: the case of mouse epidermal growth factor. *Biochemistry* 30: 4896–4900. PMID: [2036358](#)
37. Chimenti Michael S, Khangulov Victor S, Robinson Aaron C, Heroux A, Majumdar A, et al. (2012) Structural Reorganization Triggered by Charging of Lys Residues in the Hydrophobic Interior of a Protein. *Structure* 20: 1071–1085. doi: [10.1016/j.str.2012.03.023](#) PMID: [22632835](#)
38. Damjanovic A, Wu X, Garcia-Moreno EB, Brooks BR (2008) Backbone relaxation coupled to the ionization of internal groups in proteins: a self-guided Langevin dynamics study. *Biophys J* 95: 4091–4101. doi: [10.1529/biophysj.108.130906](#) PMID: [18641078](#)
39. Takayama Y, Castaneda CA, Chimenti M, Garcia-Moreno B, Iwahara J (2008) Direct evidence for deprotonation of a lysine side chain buried in the hydrophobic core of a protein. *J Am Chem Soc* 130: 6714–6715. doi: [10.1021/ja801731g](#) PMID: [18454523](#)
40. Goh GB, Garcia-Moreno E B, Brooks CL (2011) The High Dielectric Constant of Staphylococcal Nuclease Is Encoded in Its Structural Architecture. *Journal of the American Chemical Society* 133: 20072–20075. doi: [10.1021/ja2084866](#) PMID: [22085022](#)
41. Wu X, Damjanovic A, Brooks BR (2012) Efficient and Unbiased Sampling of Biomolecular Systems in the Canonical Ensemble: A Review of Self-Guided Langevin Dynamics. In: Rice SA, Dinner AR, editors. *Advances in Chemical Physics*. Hoboken: John Wiley & Sons, Inc. pp. 255–326.

42. Wu X, Brooks BR (2011) Force-momentum-based self-guided Langevin dynamics: a rapid sampling method that approaches the canonical ensemble. *J Chem Phys* 135: 204101. doi: [10.1063/1.3662489](https://doi.org/10.1063/1.3662489) PMID: [22128922](https://pubmed.ncbi.nlm.nih.gov/22128922/)
43. Wu X, Brooks BR, Vanden-Eijnden E (2015) Self-guided Langevin dynamics via generalized Langevin equation. *Journal of Computational Chemistry*: n/a-n/a. doi: [10.1002/jcc.24015](https://doi.org/10.1002/jcc.24015) PMID: [26183423](https://pubmed.ncbi.nlm.nih.gov/26183423/)
44. Wu XW, Sung SS (1999) Simulation of peptide folding with explicit water—A mean solvation method. *Proteins: Structure, Function and Genetics* 34: 295–302.
45. Bennett CH (1976) Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* 22: 245–268.
46. Wu X-W, Sung S-S (1996) An efficient approach to estimating the thermodynamic properties of fluid mixtures in molecular simulation. *The Journal of Chemical Physics* 104: 3709–3717.
47. Wu X, Brooks BR (2012) Molecular Simulation with Discrete Fast Fourier Transform. In: Salih SM, editor. *Fourier Transform—Materials Analysis: InTech*. pp. 137–164.
48. Wu X, Brooks BR (2008) Using the isotropic periodic sum method to calculate long-range interactions of heterogeneous systems. *J Chem Phys* 129: 154115. doi: [10.1063/1.2992601](https://doi.org/10.1063/1.2992601) PMID: [19045184](https://pubmed.ncbi.nlm.nih.gov/19045184/)
49. Klauda JB, Wu X, Pastor RW, Brooks BR (2007) Long-range Lennard-Jones and electrostatic interactions in interfaces: application of the isotropic periodic sum method. *J Phys Chem B* 111: 4393–4400. PMID: [17425357](https://pubmed.ncbi.nlm.nih.gov/17425357/)
50. Wu X, Brooks BR (2005) Isotropic periodic sum: a method for the calculation of long-range interactions. *J Chem Phys* 122: 44107. PMID: [15740235](https://pubmed.ncbi.nlm.nih.gov/15740235/)
51. Wu X, Brooks BR (2011) Toward canonical ensemble distribution from self-guided Langevin dynamics simulation. *J Chem Phys* 134: 134108. doi: [10.1063/1.3574397](https://doi.org/10.1063/1.3574397) PMID: [21476744](https://pubmed.ncbi.nlm.nih.gov/21476744/)
52. Wu X, Brooks BR (2003) Self-guided Langevin dynamics simulation method. *Chemical Physics Letters* 381: 512–518.
53. Mongan J, Case DA (2005) Biomolecular simulations at constant pH. *Curr Opin Struct Biol* 15: 157–163. PMID: [15837173](https://pubmed.ncbi.nlm.nih.gov/15837173/)
54. Brooks BR, Brooks CL 3rd, Mackerell AD Jr., Nilsson L, Petrella RJ, et al. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30: 1545–1614. doi: [10.1002/jcc.21287](https://doi.org/10.1002/jcc.21287) PMID: [19444816](https://pubmed.ncbi.nlm.nih.gov/19444816/)
55. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *JComputChem* 4: 187–217.
56. MacKerell AD Jr., Bashford D, Bellott M, Dunbrack RL Jr., Evanseck JD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *JPhysChemB* 102: 3586–3616.
57. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical Intergration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *JComputPhys* 23: 327–341.