Published in final edited form as:

J Phys Chem B. 2012 August 2; 116(30): . doi:10.1021/jp303385x.

pH-Replica Exchange Molecular Dynamics In Proteins Using A Discrete Protonation Method

Danial Sabri Dashti¹, Yilin Meng², and Adrian E. Roitberg^{3,*}

¹Department of Physics and Quantum Theory Project, University of Florida, Gainesville, Florida, 32611-8435

²Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois, 60637

³Department of Chemistry and Quantum Theory Project, University of Florida, Gainesville, Florida, 32611-8435

Abstract

Protonation equilibria in biological molecules modulates structure, dynamics and function. A pH-Replica Exchange Molecular Dynamics (pH-REMD) method is described here to improve the coupling between conformational and protonation sampling. Under a Hamiltonian replica exchange setup, conformations are swapped between two neighboring replicas, which themselves are at different pHs. The method has been validated on a series of biological systems. We applied pH-REMD to a series of model compounds, to an terminally-charged ADFDA pentapeptide, and to a hepta-peptide derived from the ovomucoid third domain (OMTKY3). In all those systems, the predicted pK_a by pH-REMD is very close to the experimental value and almost identical to the ones obtained by Constant pH Molecular dynamics (CpH MD). The method presented here, pH-REMD, has the advantage of faster convergence properties, due to enhanced sampling of both conformation and protonation spaces.

Keywords

Hamiltonian Replica Exchange molecular Dynamics; Constant pH molecular Dynamics; pK_a prediction; Generalized Born; Metropolis Monte Carlo

Introduction

Solution pH plays a very important role in protein function, dynamics and structure. Many important biological phenomena function only at a certain range of pHs, including protein folding/misfolding¹⁻³, enzyme catalysis⁴⁻⁶ and ligand-substrate docking^{7,8}. In those cases, a pH change leads to a change in the ratio of protonation states of different titratable residues, which is usually coupled to a change in the conformation and dynamics of the protein itself.

The pK_a of a titratable residue is the pH value at which the ratio of deprotonated to protonated concentrations of that residue is equal to one.⁹⁻¹² The pK_a of an ionizable residue in a protein is highly dependent on its electrostatic environment, which is coupled to the protonation states of other titratable groups and the conformation state of the protein. Most current simulations of pH dependent properties do not, in fact, change protonation states during molecular dynamics, but rather pick a certain set of protonation states using educated

*to whom correspondence should be addressed. roitberg@ufl.edu.

guesses, or guided by simplified algorithms, and then keep it constant for the remainder of the simulation. These constant protonation MD methods suffer from two big disadvantages.¹³ First, at a pH near the pK_a of any of the titratable residues, any possible choice of constant protonation would clearly be wrong. Second, the choice of protonation state is coupled for neighboring residues, which itself couples to the conformational space of the system.

Like concentration and temperature, the solution pH is a very useful external and controllable variable in experimental methods. Hence, the importance of constant pH MD (CpH MD) has been recognized. In the last two decades, many constant pH MD methods have been developed.^{14,15} The goal of CpH MD is to describe correctly the protonation equilibrium coupled by conformation equilibrium at a certain pH. The majority of CpH MD methods can be divided into continuous^{14–22} and discrete^{13,23–32} protonation state methods.

Continuous protonation methods use a continuous protonation parameter to perturb ionizable residue between protonated and deprotonated states. In 1994, Mertz and Pettitt²¹ developed a grand canonical method for simulating a simple chemical reaction. They applied the method for exchanging a proton between water molecules and an ionizable side chain. In 1997, Baptista et al.¹⁴ introduced continuous constant pH method in implicit solvent based on a mean-field approximation. In 2001, Börjesson et al.¹⁵ used a weakly coupled proton bath to continuously adjust the protonation fraction of each titratable group towards equilibrium. More recently, the Brooks group has developed the continuous protonation state method further.^{16–20,33} In the case of highly coupled titration groups, where cooperativity effects are non-negligible, this model leads to inappropriate estimation of physical variables. To alleviate this problem, Lee et al.³³ added a biasing potential, centered at equal 0.5, to help drive the protonation coordinate value to fully protonated/deprotonated states, and away from the mid- unphysical states.

In contrast to the continuous protonation state methods, the discrete protonation models, define the protonation state of the ionizable group as either zero or one during the simulation, corresponding to protonated and deprotonated states only. These models use a hybrid MD-MC scheme; the MD is used to sample conformational space for a number of steps, after which a Metropolis MC³⁴ attempt will be done for changing the protonation state/states. A new set of MD steps is then done with the protonation state chosen by the MC step, and the process is repeated. Many versions of the constant pH/discrete protonation MD method have been developed.^{13,24–29,31,32,35,36} The Baptista group^{23–26,36} used explicit solvent for the propagation of coordinates and the Poisson-Boltzmann (PB) method for calculating the energy in the MC section of the algorithm, Walczak et al.³² employed Langevin Dynamics for MD and the PB method for both of the MC and MD steps. Bürgi et al.²⁷ applied the Thermodynamic Integration method (TI) to calculate the transition energy between the protonated and deprotonated species in explicit solvent MD. The drawback for this approach is the large computational cost of TI calculation, which limits the amount of sampling in protonation space. In 2004, Mongan et al. developed a discrete protonation MD method by using the GB implicit solvent method to both the MD (structure) and MC (protonation state) sampling sections.¹³ A more detailed explanation of the method can be found in the methods section. This method is implemented in the AMBER Molecular Dynamics Package.³⁷

It has become clear in recent years that accurate modeling of protonation space also requires enhanced sampling of conformational space.^{38–40} Accurate sampling of conformational space of proteins remains a challenging area. Many theoretical methods have been proposed to overcome the free energy barriers in conformational space. Among those, generalized ensemble methods^{41–43} such as Replica Exchange MD (REMD)⁴⁴, Simulated Tempering⁴⁵

(ST), multicanonical ensemble^{46,47}, metadynamics⁴⁸ and Orthogonal Space Random Walk⁴⁹ (OSRW) are well established. In some of those methods, in addition to conformational space (i.e., regular MD) the system is made to perform a random walk in energy or temperature space, reducing conformational trapping.

To account for the coupling of protonation and conformational sampling, Khandogin et al.⁵⁰ have recently combined the continuous protonation Constant pH MD with the REMD method (REX-CPHMD). They applied it to the problems of pK_a prediction protein folding and pH dependent conformation, among others^{50,51}. Recently, Meng and Roitberg⁵² utilized a hybrid method by combining the Temperature REMD (T-REMD) and discrete protonation Constant pH MD.

In the present article, we introduce a method for pH-Exchange MD by combining the discrete protonation Constant pH MD (proposed by Mongan et al.¹³) and Hamiltonian REMD (H-REMD)⁵³.

We tested our method by applying it to five model dipeptides, to an uncapped pentapeptide with sequence ⁺H₃N-Ala-Asp-Phe-Asp-Ala-COO⁻ (ADFDA), and to an heptapeptide derived from the ovomucoid third domain OMTKY3 protein. Because the two ends of ADFDA are not capped, its two Asp residues have slightly different electrostatic environments and their pK_as deviate in different directions from the pK_a of unperturbed Asp. We will show that the pH-REMD MD improves the sampling efficiency in both protonation and conformation spaces.

In the rest of this paper, Constant pH MD (CpH) refers to Mongan's CpH MD approach, unless mentioned otherwise.

Theoretical Method and Simulation Details

Constant pH Molecular Dynamics and pH-Replica Exchange Molecular Dynamics (pH-REMD)

The goal of CpH MD is to sample the equilibrium between protonated and deprotonated state of titratable sites at a given pH. The free energy difference between protonated and deprotonated states determines the ratio of their concentrations. This free energy difference can not be calculated by Molecular Mechanics (MM), because the change in protonation state involves a bond breaking/forming phenomena, which requires a series of highly accurate quantum calculations. To address this issue, a method^{13,35,54}, which uses a pre-calculated pK_a of reference compounds, has been developed. Reference/model compounds, in AMBER terminology, are represented by a capped dipeptide for each titratable residue (i.e., ACE-titratable residue-NME). The free energy for the protonation change to be used in the MC criteria is described by Mongan⁵⁴ et al as:

$$\Delta G_{protein} = \Delta G_{protein,MM} + k_B T (pH - pK_{a,ref}) \ln 10 - \Delta G_{ref,MM} \quad \text{Equation 1}$$

Where T is the temperature, and k_B is the Boltzmann constant. $G_{protein,MM}$ is the molecular mechanics part of free energy of the titratable site in the protein, and $G_{ref,MM}$ is the pre-computed deprotonation free energy for the reference compound, described above. Using Equation 1, there is no need to calculating the QM contribution to the free energy in the protein. This method is implemented in the AMBER MD suite³⁷, using the GB implicit solvent model. Every few MD steps a Metropolis MC attempt will be done³⁴ to change the protonation state of the titratable residue and $G_{protein}$ will be used to make a decision about accepting or rejecting the proposed MC move. In other word, the MC moves will sample

protonation space, and the MD steps sample configuration space. During the MD steps, the protonation state is kept constant.

Titration curve—The pK_a of a titratable residue is related to pH environment through the Henderson-Hasselbalch (HH) equation,

$$pK_a = pH - n \log \left(\frac{[A^-]}{[HA]} \right) \quad \text{Equation 2}$$

with $[A^-]$ and $[HA]$ being the deprotonated and protonated concentrations respectively; n is the Hill coefficient, which should approach 1 for non-interacting ionizable residues, deviating from one in the case of interacting titratable residues because of cooperativity⁵⁵.

Because of the ergodicity assumption underlying in MD, the ratio of the time that titratable residue spend at deprotonated states to the time that of it spend in protonated state, can be considered as a ratio of concentration of deprotonated state to that of protonation state.

On other hand, pH-REMD is a combination of CpH and H-REMD⁵³. In contrast to temperature replica exchange⁴⁴, in which each replica runs at a different temperature, in H-REMD, each replica runs in a distinct Hamiltonian but the same temperature. In pH-REMD each replica runs a constant pH MD at a unique pH, and periodically an exchange of conformation between two adjacent replicas is attempted (See Appendix A for more information about pH-REMD method).

We note that a recent publication⁴⁹ presented a pH replica exchange method using a different exchange criterion, that does not require recomputing energies for different replicas. In the future, we will compare the two formulations.

Simulation details

In order to validate and test the method presented here, we choose and ran simulations on three categories of systems. First, we studied the capped reference compounds (described in the methods section), consisting the ACE-titratable residue-NME.

Simulation times were 3 ns for both CpH MD and pH-REMD (for each replica) methods. We used eight pH-replicas for all the model compounds in pH-REMD.

We also tested our simulation method on a terminally charged pentapeptide model, ^+H_3N -Ala-Asp-Phe-Asp-Ala- COO^- (ADFDA). Because the two ends of ADFDA are oppositely charged, the two Asp experience different electrostatic environments and have different pK_a s. The simulation times of ADFDA were 90 ns for CpH and ten ns for each replica in pH-REMD. Eight replicas, at pH=2.5–6.0 with increment of 0.5 were used in pH-REMD method.

The third system was a heptapeptide derived from OMTKY3 (ACE-Ser-Asp-Asn-Lys-Thr-Tyr-Gly-NME). We simulated 100 ns for CpH and 10 ns for pH-REMD. Twelve replicas, at pH=2–13 with increment of 1 pH unit, were used in pH-REMD method.

All calculations in this paper are done using the AMBER 10 molecular simulation suite. The AMBER ff99SB force field⁵⁶ and OBC Generalized Born implicit solvent⁵⁷ (igb=2) were used in all simulations. The Shake algorithm⁵⁸ was applied to constrain the bonds between heavy and hydrogen atoms, which allowed the use 2 fs MD steps. The cutoff of 30 Å for non-bonded interactions was chosen in all calculations.

In all pH-REMD simulations, except the heptapeptide derived from OMTKY3, replicas attempted a pH exchange every 1000 MD steps. In the case of the heptapeptide, replicas attempted a pH exchange every 500 MD step to accelerate sampling⁵⁹. The exchange acceptance ratios between replicas for all system were between 0.2 and 1.0. For calculating the error bars and the uncertainty of pK_a s, every 4000 MC steps (in protonation space), deprotonation fraction has been calculated.

Results and Discussion

I. Titratable model compounds

When we apply pH-REMD to the model compounds that were initially used to parametrize the method, it is not surprising to find that it produces the correct pK_a s. It is, however, an important calculation to perform to check the method and to gauge its efficiency versus constant pH runs. The results are shown in Table 1 where all pK_a values have been calculated by a fit to the linearized version of the Henderson-Hasselbalch equation.

In Figure 1, the Hill plots for capped Lysine has been plotted for both pH-REMD and CpH MD methods, it is showing agreement among them over a large pH range.

II. ADFDA model compound

We now apply the method to the model peptide, ^+H_3N -Ala-Asp-Phe-Asp-Ala- COO^- (ADFDA), that has charged ends. This system has an intrinsic but subtle asymmetry in the electrostatic environment for the two Asp ionizable side chains. Asp2 is close to the NH_3^+ , which makes its pK_a to be shifted slightly below 4.0. Asp4 is close to the COO^- terminal, which makes its pK_a to be shifted above 4.0.

The plotted titration curves of Asp2 and Asp4 in Figure 2 show that of curves of Asp2 is shifted to the left of the model compound, while the curve for Asp4 is shifted to the right. There is a good agreement between the pH-REMD and CpH curves. The predicted pK_a s and Hill coefficients for both methods have been calculated by linear fitting on a Hill plot. Table 2 shows that the results of both methods are identical. It is worth noting that the free energy difference associated with the two different pK_a s is 0.96 kcal/mol, which highlights the sensitivity of the method to very small environmental changes.

To gauge the comparative efficiency of CpH and pH-REMD, we show, in Figure 3, the cumulative average protonation fractions of Asp2 and Asp4 for both methods at pH=4.0 has been plotted. The data clearly shows that for the both titratable groups, pH-REMD shows a faster convergence in protonation space when compared to the regular CpH method.

III. Heptapeptide derived from OMTKY3

We applied the pH-REMD method to a capped heptapeptide derived from OMTKY3 (ACE-Ser-Asp-Asn-Lys-Thr-Tyr-Gly-NME). Dlugosz and Antosiewicz^{28,29} studied this heptapeptide and they predicted the pK_a of 4.24 for Asp3 using their CpH MD method. According to NMR experiments^{30,31}, the pK_a of Asp in this heptapeptide is about 3.6. From a Hill plot, the pK_a of the three titratable groups has been calculated with the results presented in Table 3. Our computed for the Asp3 pK_a of 3.7 (CpH) and 3.6 (pH-REMD) are in excellent agreement with the experimental value. The titration curves are shown in Figure 4.

There are two more titratable groups in the heptapeptide, i.e., Lys5 and Tyr7, which we also titrated. Figure 5 presents the titration curves of Lys5 and Tyr7, with the computed pK_a s listed in Table 3.

To compare the convergence speed between CpH and pH-REMD, we studied the cumulative average protonation/deprotonation fraction as function of MC titration steps for all ionizable residues. Figure 6 shows the data for Tyr 7. It is evident that pH-REMD converges faster (and smoother) to the final protonation fraction.

While the convergence of protonation equilibrium is crucial for the proper computation of a pK_a , the convergence of structural properties is also important. To consider this issue, we calculated the RMSD of α -Carbons for all pHs (For both CpH and pH-remd simulations) respect to the average structure of CpH simulation at pH 10. For both CpH and pH-REMD methods, the conformational convergence has been studied by calculation of The Kullback-Leibler divergence (see the Appendix) of RMSD cumulative distributions vs. time. This is a measure of the rate of convergence to the final conformational ensemble. We presented results for both CpH and pH-REMD at pH 10 for every 100 ps (Figure 7). We used the Final RMSD distribution of CpH simulation at pH 10 as a reference for the plot. As the inset of Figure 7 shows both simulations converge to the same RMSD distribution. It is clear that pH-REMD shows a faster and smoother conformational convergence than CpH. We also compared the rate of visiting distinct structures by calculating the RMSD autocorrelation at all pH for both methods. According to Figure 8, the correlation time of RMSDs in pH-REMD simulations are significantly shorter than that of CpH simulations, which implies that pH-REMD visits distinct conformations more often than CpH.

Conclusions

In the present work we use a combination of Hamiltonian REMD with Constant pH MD to create what we call pH-REMD. The predicted pK_a for a number of systems using pH-REMD is in excellent agreement with experimental data. Compared to CpH MD, pH-REMD converges faster in both conformational and protonation spaces. In contrast the Temperature Replica Exchange Molecular Dynamics methods (T-REMD), in pH-REMD the replica ladder (pH) is very limited, so the overlap of energy and consequently the exchange ratio between neighboring replicas are always high. This new method is expected to perform very well for biosystems with highly coupled conformational and protonation states, like proteins with pH-dependent structure and dynamics.

Acknowledgments

This work is supported from National Institute of Health under Contract 1R01 AI073674. Computer resources and support were provided by the Large Allocations Resource Committee through grant TG-MCA05S010 and the University of Florida High-Performance Computing Center.

References

1. Bierzyński A, Kim PS, Baldwin RL. Proc Natl Acad Sci U S A. 1982; 79:2470–2472. [PubMed: 6283528]
2. Shoemaker KR, Kim PS, Brems DN, Marqusee S, York EJ, Chaiken IM, Stewart JM, Baldwin RL. Proc Natl Acad Sci U S A. 1985; 82:2349–2353. [PubMed: 3857585]
3. Schaefer M, Van Vlijmen HWT, Karplus M. Adv Protein Chem. 1998;51,1–57.
4. Demchuk E, Genick UK, Woo TT, Getzoff ED, Bashford D. Biochemistry. 2000; 39:1100–1113. [PubMed: 10653656]
5. Dillet V, Dyson HJ, Bashford D. Biochemistry. 1998; 37:10298–10306. [PubMed: 9665738]
6. Harris TK, Turner GJ. IUBMB Life. 2002; 53:85–98. [PubMed: 12049200]
7. Antosiewicz J, Briggs JM, McCammon JA. Eur Biophys J. 1996; 24:137–141. [PubMed: 8852560]
8. Hunenberger PH, Helms V, Narayana N, Taylor SS, McCammon JA. Biochemistry. 1999; 38:2358–2366. [PubMed: 10029529]

9. Hill TL. *J Am Chem Soc.* 1956; 78:1577–1580.
10. Simonson T, Carlsson J, Case D. *J Am Chem Soc.* 2004; 126:4167–4180.
11. Tanford C, Kirkwood JG. *J Am Chem Soc.* 1957; 79:5333–5339.
12. Warshel A. *Nature.* 1987; 330:15–16. [PubMed: 3670392]
13. Mongan J, Case DA, McCammon J. *J Comput Chem.* 2004; 25:2038–2048.
14. Baptista AM, Martel PJ, Petersen SB. *Proteins.* 1997; 27:523–544. [PubMed: 9141133]
15. Borjesson U, Hunenberger PH. *J Chem Phys.* 2001; 114:9706.
16. Khandogin J, Brooks CL. *Biophys J.* 2005; 89:141–157. [PubMed: 15863480]
17. Khandogin J, Brooks CL. *Biochemistry.* 2006; 45:9363–9373. [PubMed: 16878971]
18. Khandogin J, Brooks CL. *Proc Natl Acad Sci U S A.* 2007; 104:16880–16885. [PubMed: 17942695]
19. Khandogin J, Chen JH, Brooks CL. *Proc Natl Acad Sci U S A.* 2006; 103:18546–18550. [PubMed: 17116871]
20. Khandogin J, Raleigh DP, Brooks CL. *J Am Chem Soc.* 2007; 129:3056–3057. [PubMed: 17311386]
21. Mertz JE, Pettitt BM. *Int J Supercomput Ap.* 1994; 8:47–53.
22. Borjesson U, Hunenberger PH. *J Phys Chem B.* 2004; 108:13551–13559.
23. Baptista AM, Teixeira VH, Soares CM. *J Chem Phys.* 2002; 117:4184.
24. Machuqueiro M, Baptista AM. *Biophys J.* 2006; 92:1836–1845. [PubMed: 17172294]
25. Machuqueiro M, Baptista AM. *Proteins: Struct, Funct, Bioinf.* 2008; 72:289.
26. Machuqueiro M, Baptista AM. *J Am Chem Soc.* 2009; 131:12586–12594. [PubMed: 19685871]
27. Burgi R, Kollman PA, van Gunsteren WF. *Proteins.* 2002; 47:469–480. [PubMed: 12001225]
28. Dlugosz M, Antosiewicz JM. *Chem Phys.* 2004; 302:161–170.
29. Dlugosz M, Antosiewicz JM. *J Phys Chem B.* 2005; 109:13777–13784. [PubMed: 16852726]
30. Dlugosz M, Antosiewicz JM. *J Phys: Condens Matter.* 2005; 17:S1607.
31. Dlugosz M, Antosiewicz JM, Robertson AD. *Phys Rev E.* 2004; 69:021915–021924.
32. Walczak AM, Antosiewicz JM. *Phys Rev E.* 2002; 66:051911–051918.
33. Lee MS, Salsbury FR, Brooks CL. *Proteins: Struct, Funct, Bioinf.* 2004; 56:738–752.
34. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. *J Chem Phys.* 1953; 21:1087.
35. Baptista AM. *J Chem Phys.* 2002; 116:7766–7768.
36. Machuqueiro M, Baptista AM. *J Phys Chem B.* 2006; 110:2927–2933. [PubMed: 16471903]
37. Case, DA.; Darden, TA.; Cheatham, I.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Crowley, M.; Walker, RC.; Zhang, W., et al. *AMBER 10*. University of California; San Francisco: 2008.
38. Johnson, ML.; Brand, L., editors. *Methods in Enzimology*. Academic Press; San Diego, Burlington, London: 2009. p. 467
39. Warwicker J. *Protein Sci.* 2004; 13:2793–2805. [PubMed: 15388865]
40. Barth P, Alber T, Harbury PB. *Proc Natl Acad Sci U S A.* 2007; 104:4898–4903. [PubMed: 17360348]
41. Li HZ, Fajer M, Yang W. *J Chem Phys.* 2007; 126:024106. [PubMed: 17228942]
42. Mitsutake A, Sugita Y, Okamoto Y. *Biopolymers.* 2001; 60:96–123. [PubMed: 11455545]
43. Zheng LQ, Chen MG, Yang W. *J Chem Phys.* 2009; 130:234105. [PubMed: 19548709]
44. Sugita Y, Okamoto Y. *Chem Phys Lett.* 1999; 314:141–145.
45. Lyubartsev AP, Martsinovski AA, Shevkunov SV, Vorontsov-Velyaminov PN. *J Chem Phys.* 1992; 96:1776.
46. Berg BA, Neuhaus T. *Phys Lett B.* 1991; 267:249–253.
47. Berg BA, Neuhaus T. *Phys Rev Lett.* 1992; 68:9–12. [PubMed: 10045099]
48. Laio A, Parrinello M. *Proc Natl Acad Sci U S A.* 2002; 99:12562–12566. [PubMed: 12271136]
49. Itoh SG, Damjanovi A, Brooks BR. *Proteins: Structure, Function, and Bioinformatics.* 2011; 79:3420–3436.

50. Wallace JA, Shen JK. *J Chem Theory Comput.* 2011; 7:2617–2629.
51. Wallace JA, Wang Y, Shi C, Pastoor KJ, Nguyen B, Xia K, Shen JK. *Proteins: Structure, Function, and Bioinformatics.* 2011; 79:3364–3373.
52. Meng Y, Roitberg AE. *J Chem Theory Comput.* 2010; 6:1401–1412. [PubMed: 20514364]
53. Meng Y, Sabri Dashti D, Roitberg A. *J Chem Theory Comput.* 2011; 7:2721–2727. [PubMed: 22125475]
54. Mongan J, Case DA. *Curr Opin Struct Biol.* 2005; 15:157–163. [PubMed: 15837173]
55. Williams DH, Stephens E, O'Brien DP, Zhou M. *Angew Chem, Int Ed.* 2004; 43:6596–616.
56. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. *Proteins: Struct, Funct, Bioinf.* 2006; 65:712–725.
57. Ryckaert JP, Ciccotti G, Berendsen HJC. *J Comput Phys.* 1977; 23:327–341.
58. Onufriev A, Bashford D, Case DA. *J Phys Chem B.* 2000; 104:3712–3720.
59. Sindhikara D, Meng Y, Roitberg AE. *J Chem Phys.* 2008; 128:024103. [PubMed: 18205439]
60. McClendon CL, Hua L, Barreiro G, Jacobson MP. *J Chem Theory Comput.* 2012; 8:2115–2126. [PubMed: 23316121]
61. Hamacher K. *J Comput Chem.* 2007; 28:2576–2580. [PubMed: 17486549]

Appendix A: pH-REMD

According to H-REMD algorithm, each replica runs a distinct Hamiltonian. In pH-REMD each replica runs a CpH MD simulation at a unique pH, and regularly a swap of conformations between two adjacent replicas is attempted. Using the detailed balance condition, we can write an equilibrium proposition for the ensemble before and after an exchange is attempted.

$$\wp(X)w(X \rightarrow X') = \wp(X')w(X' \rightarrow X) \quad \text{Equation A1}$$

Where $\wp(X)$ is the probability of being at state X and $w(X \rightarrow X')$ is the probability of transition from state X to that of X' .

At the exchange moment, if we only swap the conformations between two adjacent replicas and keep the protonation state unchanged, then the generalized states of X and X' can be written as

$$X = \begin{pmatrix} q_1 & \cdots & q_i & q_j & \cdots & q_N \\ n_1 & \cdots & n_i & n_j & \cdots & n_N \end{pmatrix} \text{ and } X' = \begin{pmatrix} q_1 & \cdots & q_j & q_i & \cdots & q_N \\ n_1 & \cdots & n_i & n_j & \cdots & n_N \end{pmatrix}.$$

Here q represents a conformation and n represents protonation states. The i^{th} column is related to the i^{th} replica.

Because all N replicas are independent, the probability of the system being in the

generalized state X can be written as $\wp(X) = \prod_{i=1}^N p_i(q_i, n_i)$, where $p_i(q_i, n_i)$ is the probability of replica i at conformation q_i and protonation state of n_i . Substituting those in Equation A will result

$$\frac{w(X \rightarrow X')}{w(X' \rightarrow X)} = \frac{p_i(q_j, n_i)p_j(q_i, n_j)}{p_i(q_i, n_i)p_j(q_j, n_j)}. \quad \text{Equation A2}$$

In the canonical ensemble regime (NVT), and using the Boltzmann distribution as a limiting ensemble, this can be written as

$$\frac{w(X \rightarrow X')}{w(X' \rightarrow X)} = \frac{\exp(-\beta H(q_j, n_i))\exp(-\beta H(q_i, n_j))}{\exp(-\beta H(q_i, n_i))\exp(-\beta H(q_j, n_j))} = \exp(-\Delta) \quad \text{Equation A3}$$

; where $\Delta = - (H(q_0, n_0) - H(q_0, n_1) - H(q_1, n_1) + H(q_1, n_0))$ and $\beta = 1/k_B T$.

This setup can be realized by setting the exchange probability according to the Metropolis MC criteria is

$$w(X \rightarrow X') = w((q_i, n_i); (q_j, n_j) \rightarrow (q_j, n_i); (q_i, n_j)) = \min(1, \exp(-\Delta))$$

In computing Δ , only the potential energies are required. Since the two exchanging replicas are running at the same temperature and have the same number and mass of particles, the kinetic energy terms will cancel each other.

Appendix B: Kullback-Leibler divergence

Kullback-Leibler divergence^{60,61} or relative entropy is defined as

$$D_{K-L}(X||Y) = \sum_i X(i) \log(X(i)/Y(i)) \quad \text{Equation B1}$$

; where X and Y are the probability distributions of a discrete random variable. It is the average, over the distribution X , of the logarithmic difference between the probabilities X and Y . It measures the degree to which X is distinguishable from Y . Small value of D indicates that X and Y are highly overlapped. This metric can be used as a measure of convergence rate; where Y is a target (or reference) distribution and X is changing by time.

Figure 1.
Comparison of Hill plots between pH-REMD and CpH methods for Lys reference model.

Figure 2. The titration curves of Asp side chains in ADFDA computed by both constant pH MD (blue and purple) and pH-REMD (red and green) methods; The titration curve of Asp for the reference compound (light blue) is also shown.

Figure 3. Cumulative average protonation fractions of Asps' side chains in ADFDA VS MC titration steps at pH=4.0.

Figure 4.
The titration curves of Asp3 in the heptapeptide derived from OMTKY3.

Figure 5.
The titration curves of Lys5 and Tyr7 in the heptapeptide derived from OMTKY3.

Figure 6.

Cumulative average protonation fraction for TYR7 versus MC titration steps at pH=8.0,9.0,10.0,11.0,12.0 and 13.0, a Comparison between the CpH (A) and the pH-REMD (B) methods.

Figure 7. The Kullback-Leibler divergence measure of RMSD distributions of CpH (Green) and pH-REMD (Red) respect to the final RMSD distribution in CpH. The inset shows the RMSD distributions of CpH (Green) and pH-REMD (Red) respectively after 100 and 10 ns.

Figure 8.
RMSD Autocorrelation for CpH (Green) and pH-REMD (Red) at all pHs.

Table 1

pK_as of the reference compounds computed by different methods

	ASP	GLU	HIS	TYR	LYS
Experimental value	4.0	4.4	6.3	9.6	10.5
CpH MD	3.9±0.1	4.4±0.1	6.3±0.1	9.6±0.03	10.4±0.03
pH-REMD	3.9±0.1	4.4±0.1	6.4±0.1	9.7±0.02	10.4±0.04

Table 2

pK_a prediction and Hill coefficient of fitted from the HH equation.

	ASP2		ASP4	
	pH	Hill coefficient	pH	Hill coefficient
pH-REMD	3.8±0.1	0.89	4.5±0.1	0.88
CpH	3.8±0.1	0.91	4.5±0.1	0.86

Table 3

pK_a values of the titratable residues in the heptapeptide derived from OMKTY3

	Asp3	Lys5	Tyr7
pH-REMD	3.6±0.2	10.6±0.1	10.1±0.1
CpH	3.7±0.2	10.6±0.1	9.9±0.1