Supporting Information

All-Atom Continuous Constant pH Molecular Dynamics With Particle Mesh Ewald and Titratable Water

Yandong Huang,[†] Wei Chen,[†] Jason A. Wallace,[‡] and Jana Shen[*,†]

† *Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD*

‡ *University of Oklahoma College of Dentistry, Oklahoma City, OK*

E-mail: jshen@rx.umaryland.edu

Calculation of the finite-size pK_{a} corrections

We calculated the pK_{a} corrections using the following equations (Eqs. 15 and 16 in the main text).

$$\Delta\Delta G^{\mathrm{offset}} = \frac{2\pi}{3}\kappa\gamma^{\mathrm{solv}}Q\left(\frac{N^{\mathrm{solv}}_{\mathrm{prot}}}{V_{\mathrm{prot}}} - \frac{N^{\mathrm{solv}}_{\mathrm{mod}}}{V_{\mathrm{mod}}}\right) \quad (15)$$

$$\Delta pK_a^{\mathrm{corr}} = \pm\frac{\Delta\Delta G^{\mathrm{offset}}}{ln(10)RT}, \quad (16)$$

where γ^{solv} is 0.764 e· Å^2; Q is -1 for acidic groups and +1 for basic groups; V for protein and model compound and N^{solv} are given in Table S1.

It would be desirable to calculate $\Delta pK_a^{\mathrm{corr}}$ before simulations using the the lattice-parameter-based volume (see Table S1, $V^{\mathrm{est}} = L^3$ for a cubic box and $V^{\mathrm{est}} = 0.77L^3$ for a truncated octahedron box). However, the actual simulation box volume V$^{\mathrm{sim}}$ is consistently smaller than V$^{\mathrm{est}}$, because after the equilibration water interacts with the solute (protein or model compound) and takes up less volume (see Table S1, last column). The ratio between V$^{\mathrm{sim}}$ and V$^{\mathrm{est}}$ is about 0.93 and slightly varies depending on the system (see Table S1, last column). Thus, we decided to estimate $\Delta pK_a^{\mathrm{corr}}$ using 0.93$\times V^{\mathrm{est}}_{prot}$, while the volume for the model compound system is taken from the simulation. Table S2 compares the estimated $\Delta pK_a^{\mathrm{corr}}$ with that calculated using the simulation volume for the four test proteins. As can be seen, the estimated pK_{a} corrections are very close to those calculated using the actual simulation volume if rounding to the first decimal place. The largest difference is 0.2 units. This suggests that the finite-size pK_{a} corrections can be pre-calculated and incorporated in the reference pK_{a}'s in the future.

Table S1: Volume and number of solvent in the simulations of proteins and model compounds

System	L (Å)	V^{sim} (Å^3)	V^{est} (Å^3)	N^{solv}	$V^{\mathrm{sim}}/V^{\mathrm{est}}$
HP36	54	$1.1242\times10^5 \pm 302$	1.2125×10^5	3602	0.92715
BBL	59	$1.4666\times10^5 \pm 334$	1.5814×10^5	4733	0.92741
HEWL	69	$2.3561\times10^5 \pm 404$	2.5295×10^5	7343	0.93145
SNase	70	$2.4186\times10^5 \pm 447$	2.6411×10^5	7515	0.91575
Model Asp	30	$2.5159\times10^4 \pm 220$	2.7000×10^4	842	0.93181
Model Glu	30	$2.5217\times10^4 \pm 220$	2.7000×10^4	843	0.93396
Model His	30	$2.5142\times10^4 \pm 227$	2.7000×10^4	838	0.93119

L denotes the unit-cell lattice parameter of the truncated octahedron (for proteins) and cubic box (for model compounds). The simulation box volume V^{sim} was taken to be the average of the values from the highest and lowest pH replicas. The last 1 ns (per replica) data was used. Note, as expected, the volume fluctuation is very small, around 0.1%. The estimated box volume V^{est} was calculated using L^3 for a cubic box and $0.77L^3$ for a truncated octahedron box. N^{solv} is the number of solvent.

Table S2: Finite-size corrections of the protein pK_{a}'s

System	Asp		Glu		His	
	est	sim	est	sim	est	sim
HP36	-0.6	-0.5	-0.6	-0.5	-	-
BBL	-0.5	-0.5	-0.5	-0.4	-0.4	-0.4
HEWL	-0.9	-0.9	-0.9	-0.9	-0.8	-0.8
SNase	-1.1	-0.9	-1.1	-0.9	-1.0	-0.8

Column est gives the pK_{a} corrections calculated using $0.93\times V^{\mathrm{est}}_{\mathrm{prot}}$ for the protein system, while column sim gives the pK_{a} corrections calculated using the simulation volume for the protein system. See text for further explanation.

Table S3: Validation of the finite-site pK_a correction: comparison of the estimated pK_a shifts and those from the actual simulations

L (Å)	ΔG^{offset} (kcal/mol)	V^{sim} (Å^3)	ΔpK_a^{est}	ΔpK_a^{sim}
HP36				
43	$\pm$16.17	5.60320×10^4	0.61	0.41$\pm$0.06
54*	$\pm$17.02	1.12416×10^5	0.00	0.00
BBL				
49	$\pm$16.58	8.21480×10^4	0.41	0.39$\pm$0.19
59*	$\pm$17.15	1.46660×10^5	0.00	0.00

L denotes the unit-cell lattice parameter of the truncated octahedron box. The reference box (box used in the simulations reported in the main text) is denoted with an asterisk. ΔG^{offset} denotes the calculated offset energy in kcal/mol (Eq. 14 in the main text). V^{sim} refers to the average volume from the lowest and highest pH replicas. ΔpK_a(est) refers to the estimated pK_a shift due to the decrease in box dimension relative to that of the reference box. ΔpK_a(sim) refers to the actual average pK_a shift from two independent sets of 5-ns simulations. The standard deviations for the actual pK_a shifts are based on the shifts of all residues.

Table S4: Calculated pK_a's for SNase in different time windows

Residue	Expt	0-5 ns	5-10 ns	10-16 ns
E10	2.8	3.2	3.2	3.3
D19*	2.2	4.5	3.3	3.2
D21*	6.5	5.5	6.0	6.3
D40	3.9	2.8	2.9	3.3
E43	4.3	4.3	4.1	4.2
E52*	3.9	5.1	4.7	4.9
E57	3.5	4.6	4.1	4.6
E67	3.8	3.7	4.0	4.1
E73	3.3	3.4	3.6	3.1
E75	3.3	3.5	2.7	2.5
D77	$<$2.2	$<$ 0	$<$0	$<$ 0
D83	$<$2.2	0.0	0.1	0.3
D95	2.2	2.9	3.0	3.0
E101*	3.8	5.0	4.7	4.7
E122	3.9	4.3	4.4	4.4
E129*	3.8	5.5	5.5	5.6
E135*	3.8	2.9	2.9	3.1
max		2.3	1.7	1.8
avg		0.83	0.70	0.69
rmsd		1.0	0.80	0.81

Residues denoted with an asterisk show an absolute deviation greater than 0.6 and are discussed in the main text.

Table S5: Parameters in the model PMF functions for PME-CpHMD simulations

	Asp	Glu	His	Lys	H_3O^+	OH^-
A	-	-	-	-80.32	-98.12	-117.86
B	-	-	-	0.63	0.10	1.03
A_1	-74.78	-75.21	-53.4	-	-	
B_1	0.11	0.080	0.38	-	-	-
A_0	-74.40	-74.69	-53.88	-	-	-
B_0	0.10	0.078	0.52	-	-	-
A_{10}	-11.11	-10.98	-47.04	-	-	-
B_{10}	0.50	0.50	0.34	-	-	-
a_0	0.0052	0.78	0.0	-	-	-
a_1	10.68	9.73	0.0	-	-	-
a_2	-10.67	-10.48	0.0	-	-	-
a_3	0.5	0.5	0.0	-	-	-
a_4	-74.67	-75.14	0.0	-	-	-
a_5	0.10	0.078	0.0	-	-	-
β	2.0	2.0	2.0	2.0	0.5	0.5

Coefficient names follow those in the Supplementary Materials of Ref.[1] The coefficients were initially obtained using a fitting procedure as detailed in Ref.[1]) followed by model compound simulations (5 for each model). The average pK_a's out of 5 sets of test simulations were: Lys 10.4; Asp 4.1; Glu 4.6 and His 6.2. Then, small adjustment was made to B_0 (a_5) to match the reference pK_a values of Asp and Glu. To match the reference pK_a of His, small adjustments were made to B_1,B_0 and B_{10}.

Table S6: Calculated pK_a's of four proteins with two λ cutoffs

Residue	Expt	Calc		Residue	Expt	Calc	
		0.1/0.9	0.2/0.8			0.1/0.9	0.2/0.8
HP36				**SNase**			
Asp44	3.1	2.6	2.6	Glu10	2.8	3.2	3.2
Glu45	4.0	3.8	3.8	Asp19	2.2	3.3	3.3
Asp46	3.5	3.9	4.0	Asp21	6.5	6.0	6.1
Glu72	4.4	4.5	4.6	Asp40	3.9	2.9	3.0
BBL				Glu43	4.3	4.1	4.1
Asp129	3.9	3.7	3.8	Glu52	3.9	4.7	4.8
Glu141	4.5	4.3	4.3	Glu57	3.5	4.1	4.1
His142	6.5	5.4	5.5	Glu67	3.8	4.0	4.0
Asp145	3.7	3.4	3.5	Glu73	3.3	3.6	3.6
Glu161	3.7	4.0	4.0	Glu75	3.3	2.7	2.7
Asp162	3.2	2.7	2.7	Asp77	$<$2.2	$<$-1	$<$-1
Glu164	4.5	4.3	4.3	Asp83	$<$2.2	0.1	0.0
His166	5.4	4.1	4.1	Asp95	2.2	3.0	2.9
HEWL				Glu101	3.8	4.7	4.7
Glu7	2.6	3.2	3.2	Glu122	3.9	4.4	4.4
His15	5.5	4.0	4.1	Glu129	3.8	5.5	5.5
Asp18	2.8	2.9	2.9	Glu135	3.8	2.9	2.9
Glu35	6.1	7.1	7.2				
Asp48	1.4	0.9	0.9				
Asp52	3.6	5.6	5.6				
Asp66	1.2	1.1	1.1				
Asp87	2.2	2.3	2.3	*rmsd*		0.77	0.76
Asp101	4.5	5.2	5.2	*regression slope*		0.98	1.01
Asp119	3.5	3.5	3.5	*regression interc*		0.06	0.11

Experimental pK_a values were determined by NMR titration for HP36,[2] BBL,[3,4] HEWL[5] and SNase.[6] The pK_a's were based on the last 5 ns of the 10-ns (per replica) simulations. Column 0.1/0.9 or 0.2/0.8 indicates the cutoffs $\lambda^{\mathrm{P}}/\lambda^{\mathrm{U}}$. For example, the cutoffs 0.1/0.9 define the protonated state as $\lambda < 0.1$ and the deprotonated state as $\lambda > 0.9$. In the presence of tautomer states, the same cutoffs are used for the tautomer variable x. For 0.1/0.9, the protonated state is defined as $\lambda < 0.1$ with $x < 0.1$ or $x > 0.9$; the deprotonated state is defined as $\lambda > 0.9$ with $x < 0.1$ or $x > 0.9$.

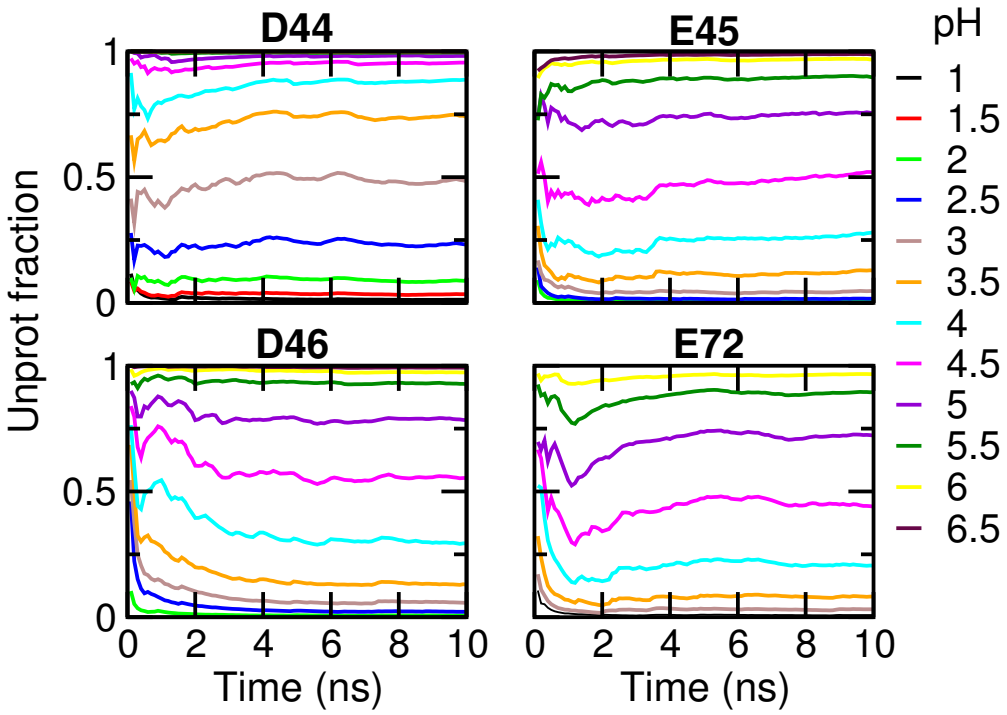


Figure S1: **Convergence of protonation-state sampling for HP36.** Unprotonated fractions cumulatively calculated as a function of time.

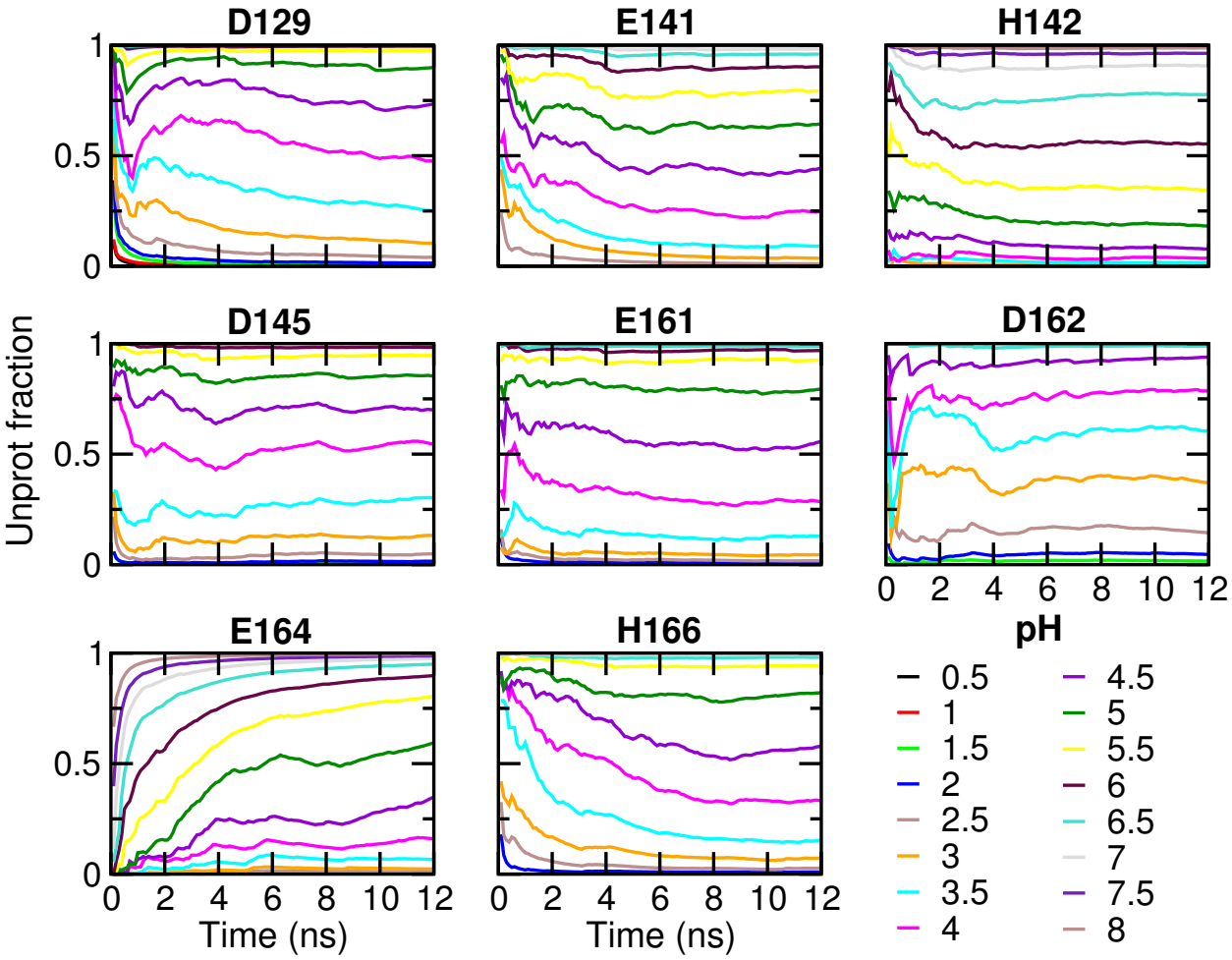


Figure S2: **Convergence of protonation-state sampling for BBL.** Unprotonated fractions cumulatively calculated as a function of time.

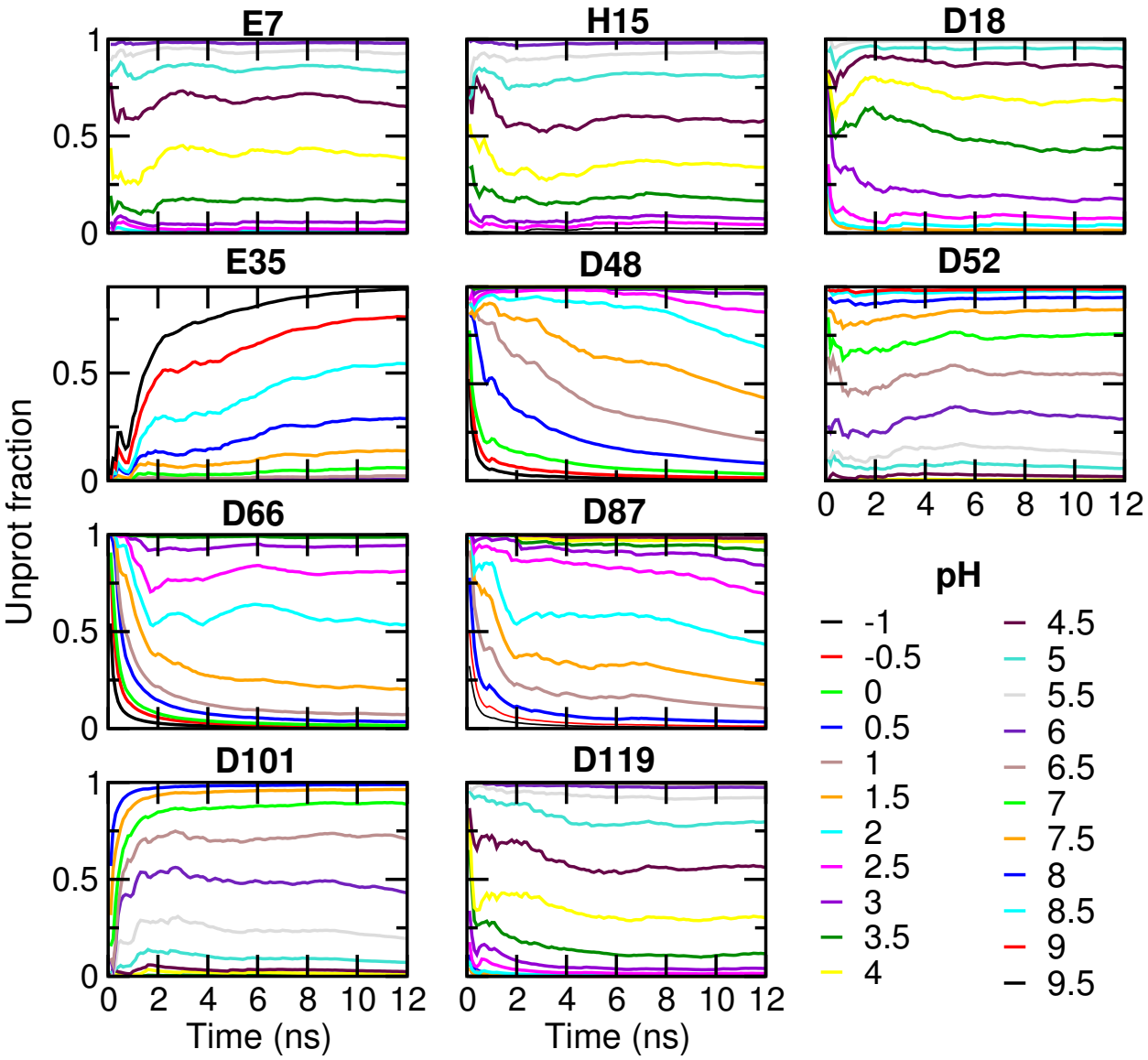


Figure S3: **Convergence of protonation-state sampling for HEWL.** Unprotonated fractions cumulatively calculated as a function of time.

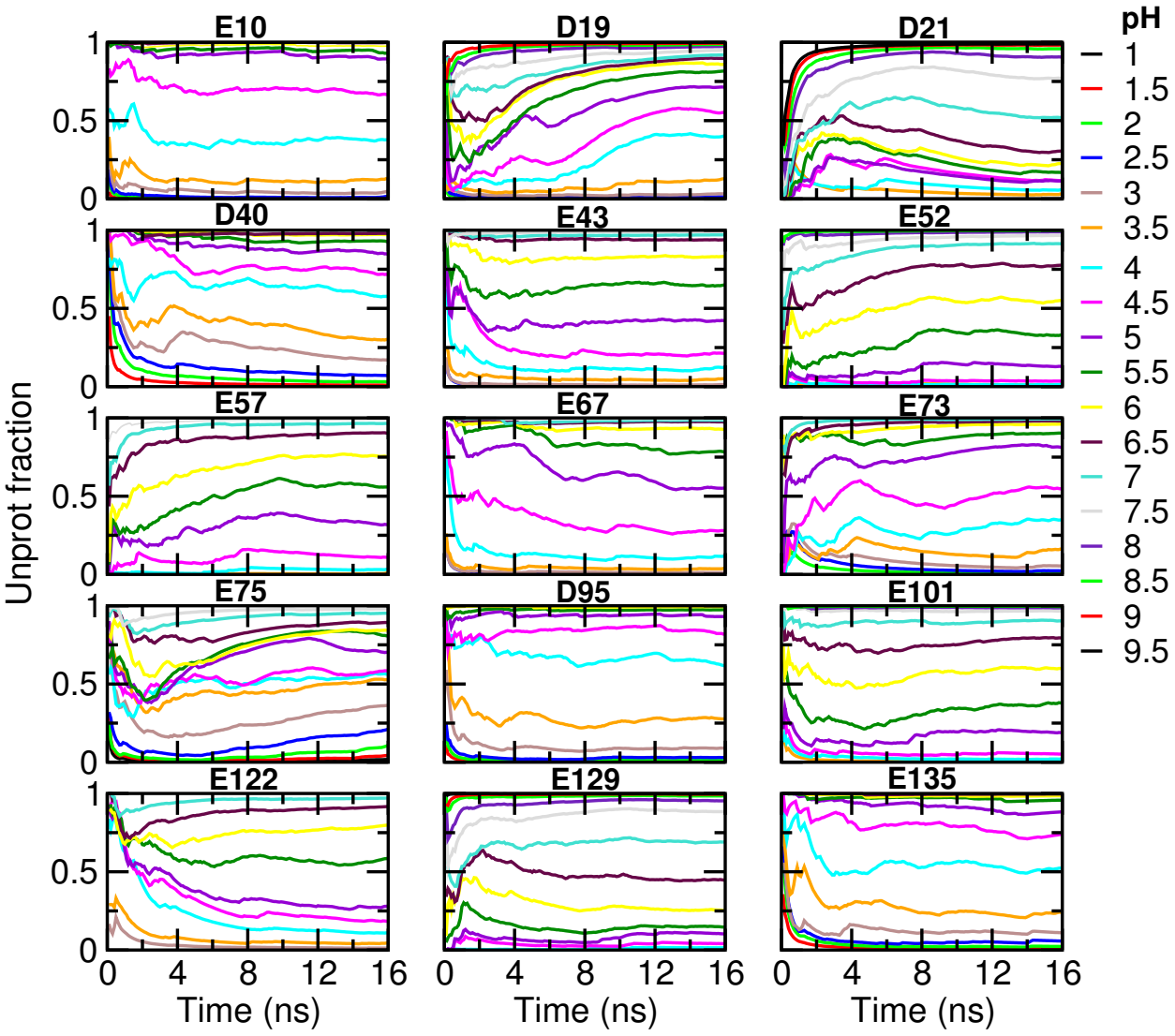


Figure S4: **Protonation-state sampling for SNase is converged with 16-ns (per replica) sampling.** Unprotonated fractions cumulatively calculated as a function of time. Note simulation was carried out to 16 ns per replica.

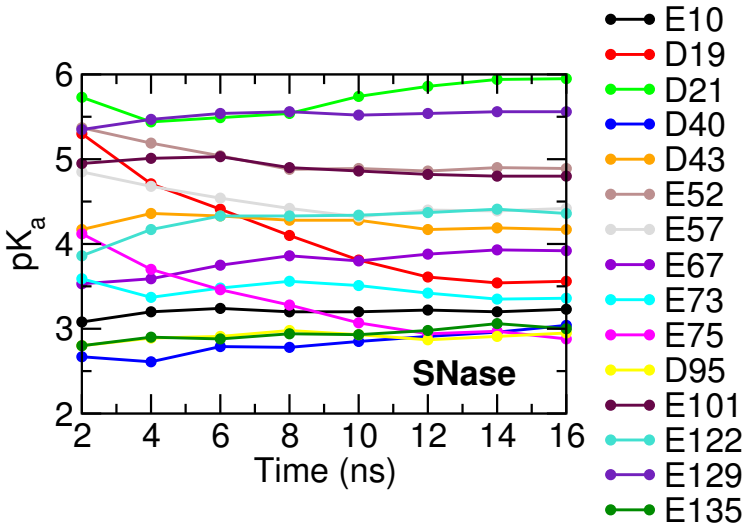


Figure S5: **Calculated pK_a's for SNase are converged with 16-ns per replica sampling.** pK_a calculation was performed every 2 ns per replica based on the cumulative values of the unprotonated fractions at all pH.

Analysis of pK_a's errors

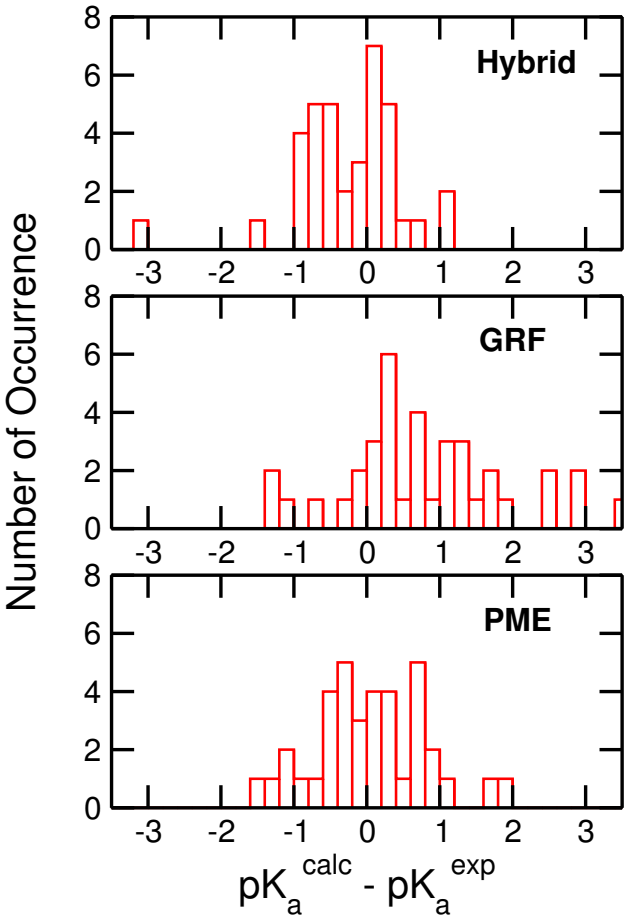
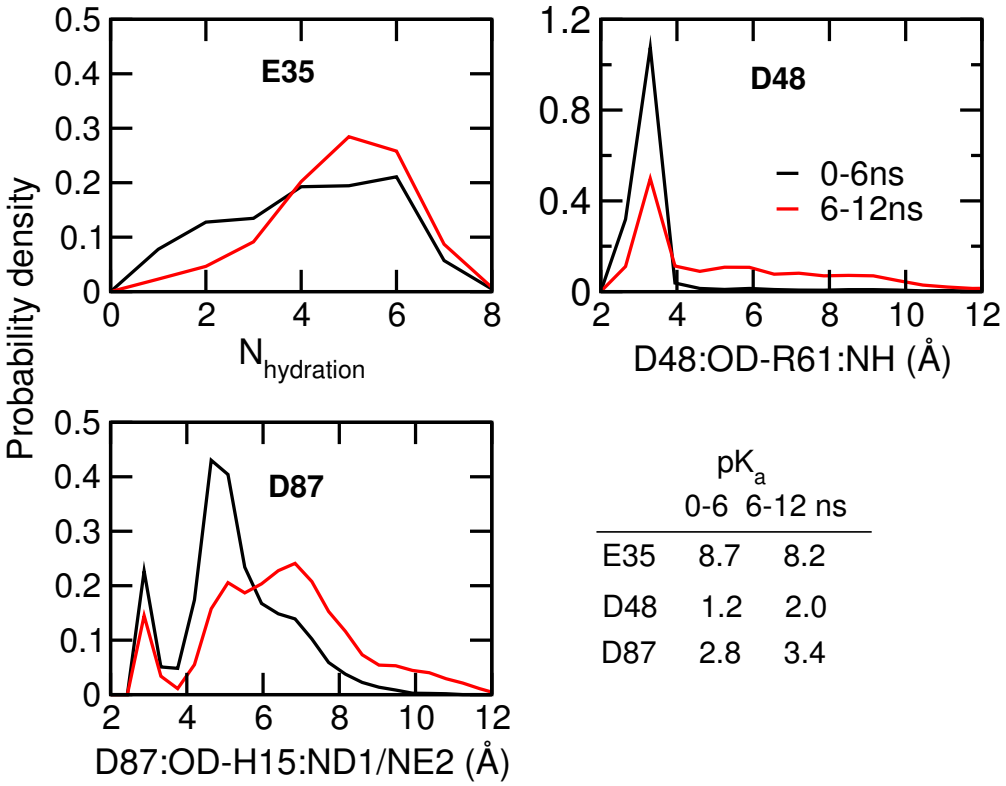


Figure S6: Histogram of the deviations between calculated and experimental pK_a's.



	pK$_a$ 0-6	6-12 ns
E35	8.7	8.2
D48	1.2	2.0
D87	2.8	3.4

Figure S7: Improved pK_a's in the second half of the simulation of HEWL. Left. The hydration number of Glu35 at pH 9.5. Right. The distance between Asp48 and Arg61 at pH 2. The selected pH conditions are slightly higher than the uncorrected pK_a values. The first and second half of the simulation are displayed in black and red, respectively.

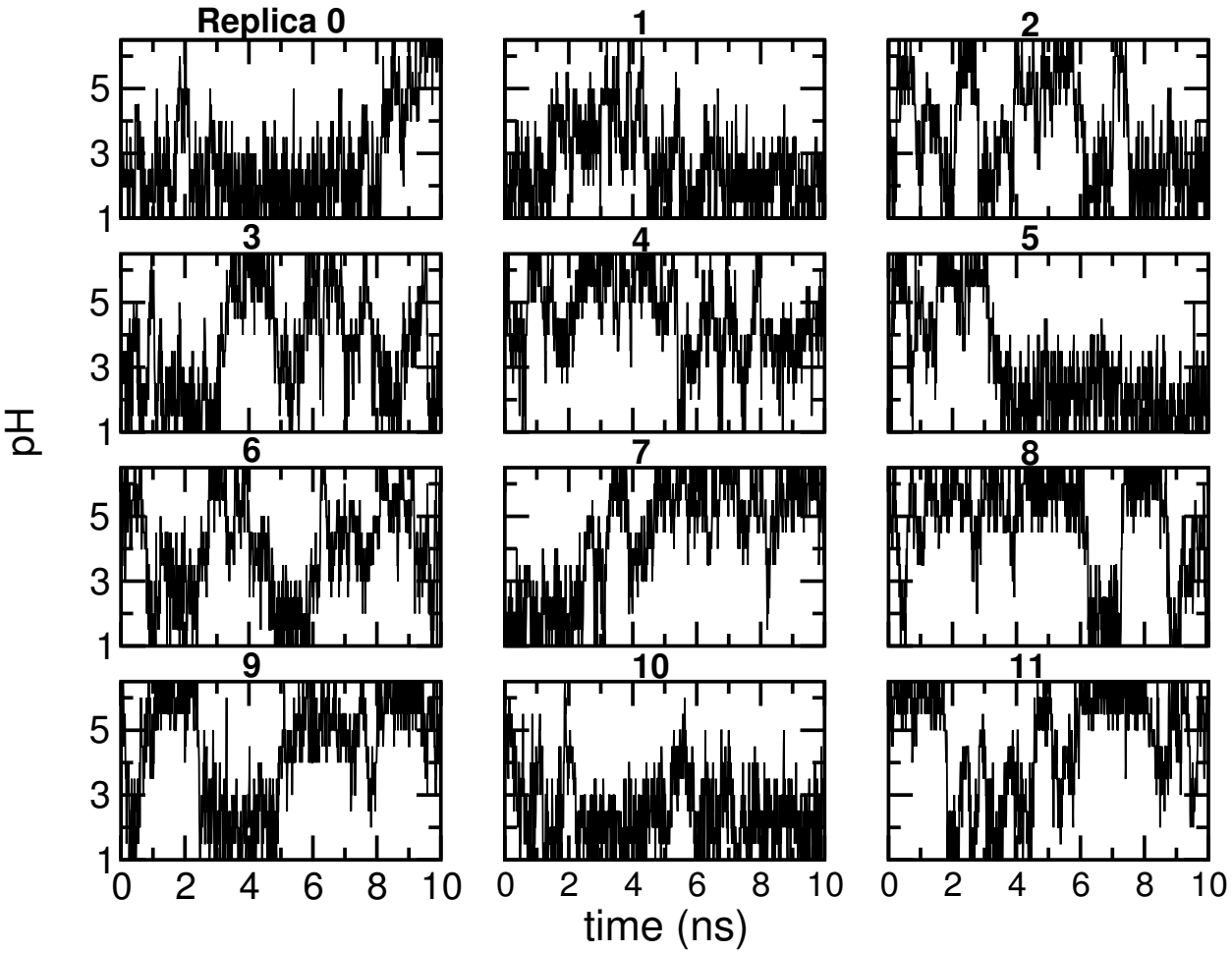


Figure S8: Replica walk in the pH-REX CpHMD simulation of HP36.

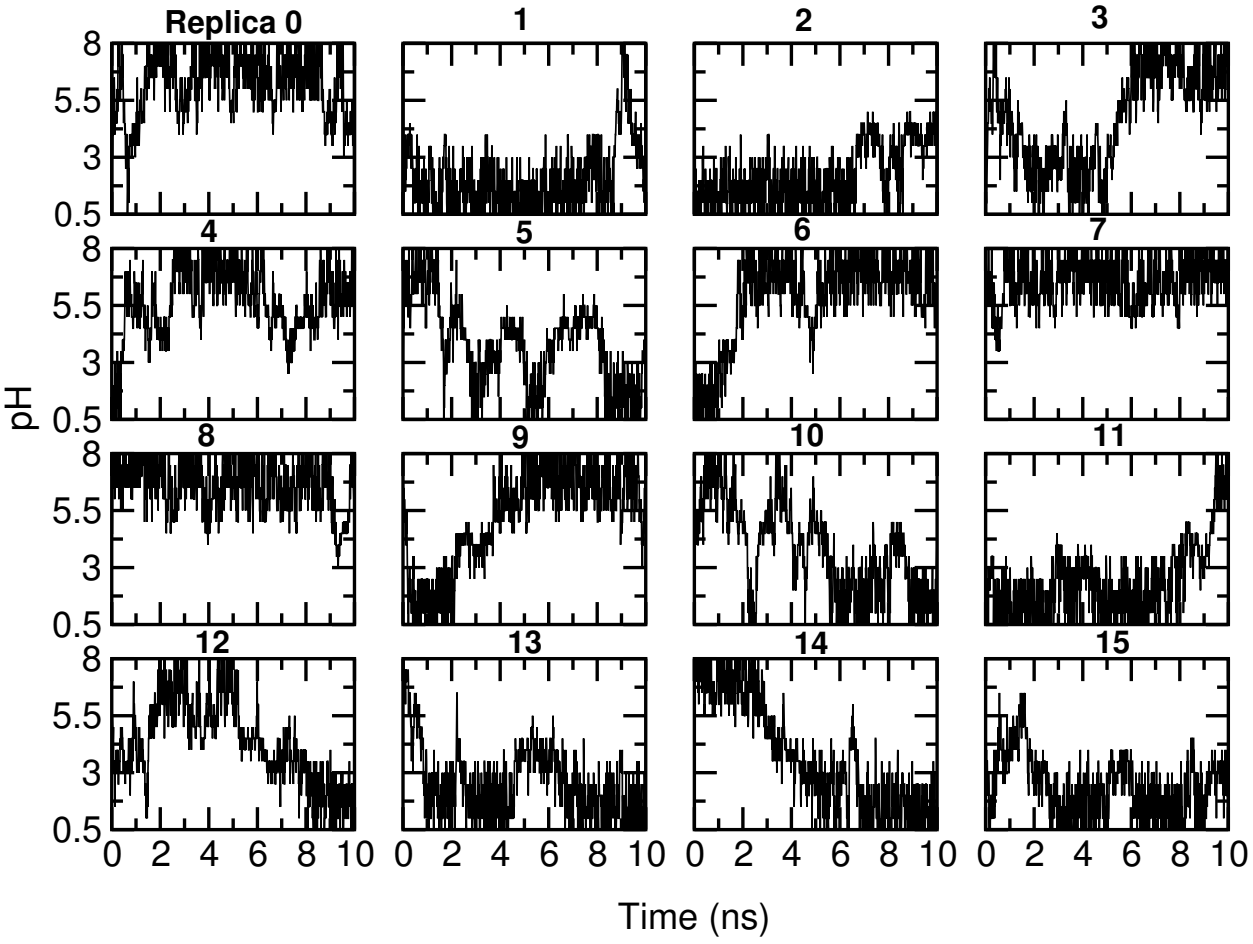


Figure S9: Replica walk in the pH-REX CpHMD simulation of BBL.

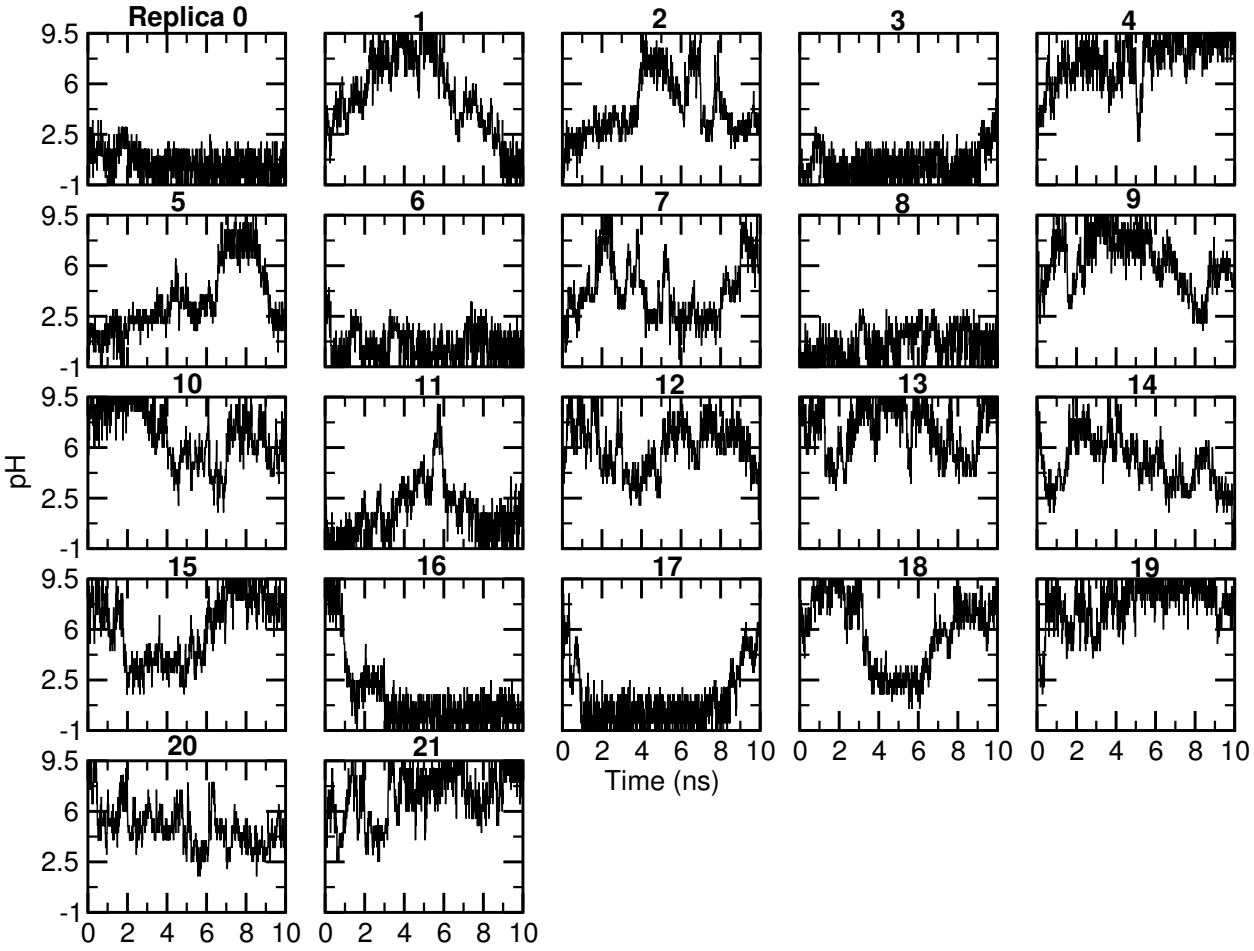


Figure S10: Replica walk in the pH-REX CpHMD simulation of HEWL.

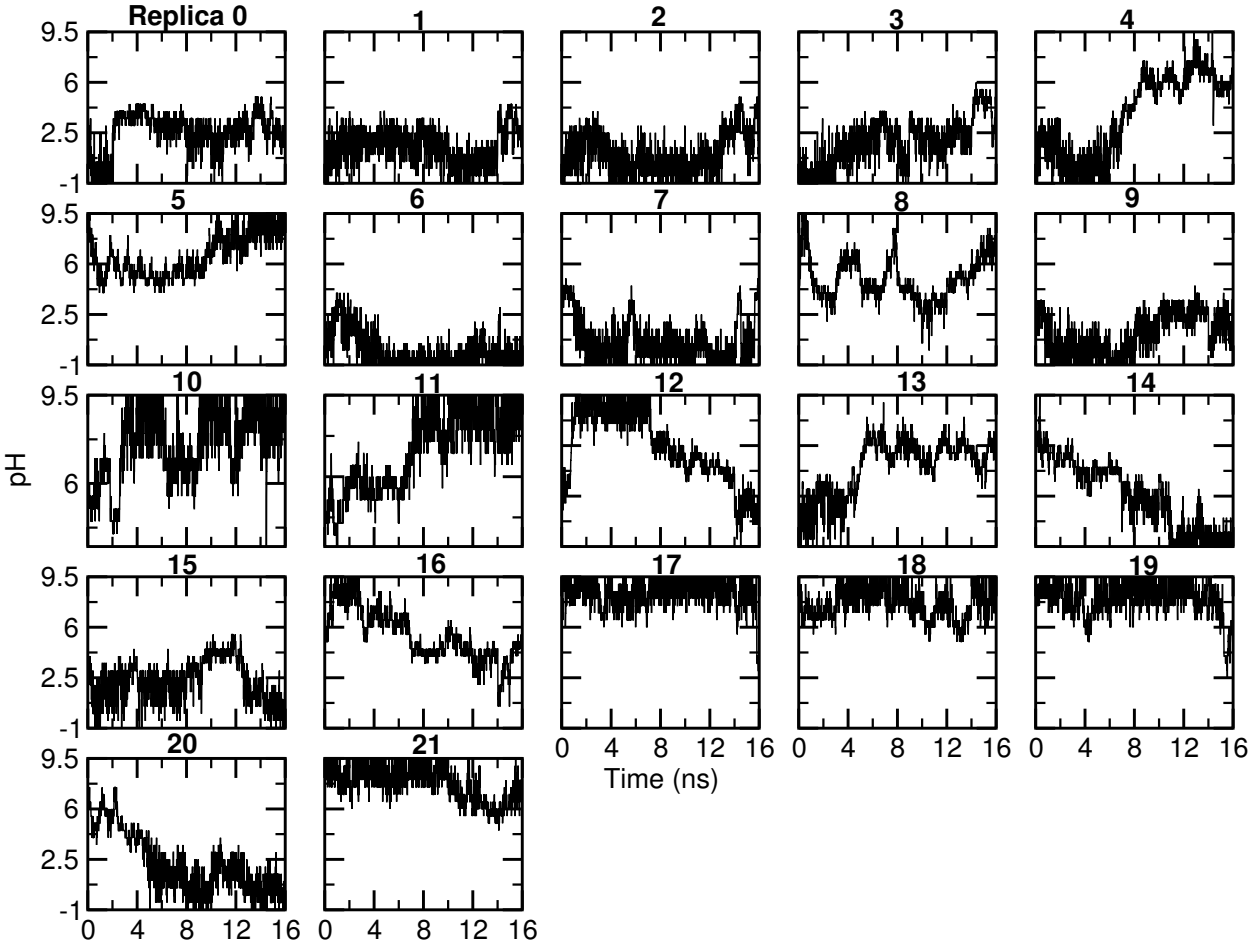


Figure S11: Replica walk in the pH-REX CpHMD simulation of SNase.

References

(1) Khandogin, J.; Brooks III, C. L. Constant pH molecular dynamics with proton tautomerism. *Biophys. J.* **2005**, *89*, 141–157.

(2) Xiao, S.; Patsalo, V.; Shan, B.; Bi, Y.; Green, D. F.; Raleigh, D. P. Rational modification of protein stability by targeting surface sites leads to complicated results. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 11337–11342.

(3) Arbely, E.; Rutherford, T. J.; Sharpe, T. D.; Ferguson, N.; Fersht, A. R. Downhill versus barrier-limited folding of BBL 1: energetic and structural pertubation effects upon protonation of a histidine of unusually low pKa. *J. Mol. Biol.* **2009**, *387*, 986–992.

(4) Arbely, E.; Rutherford, T. J.; Neuweiler, H.; Sharpe, T. D.; Ferguson, N.; Fersht, A. R. Carboxyl pK_a values and acid denaturation of BBL. *J. Mol. Biol.* **2010**, *403*, 313–327.

(5) Webb, H.; Tynan-Connolly, B. M.; Lee, G. M.; Farrell, D.; O'Meara, F.; Søndergaard, C. R.; Teilum, K.; Hewage, C.; McIntosh, L. P.; Nielsen, J. E. Remeasuring HEWL pK_a values by NMR spectroscopy: methods, analysis, accuracy, and implications for theoretical pK_a calculations. *Proteins* **2011**, *79*, 685–702.

(6) Castañeda, C. A.; Fitch, C. A.; Majumdar, A.; Khangulov, V.; Schlessman, J. L.; García-Moreno E., B. Molecular determinants of the pK_a values of Asp and Glu residues in staphylococcal nuclease. *Proteins* **2009**, *77*, 570–588.