Supporting Information

Constant pH Molecular Dynamics of Proteins in Explicit Solvent with Proton Tautomerism

Garrett B. Goh,[†] Benjamin S. Hulbert,[†] Huiqing Zhou,[†] and Charles L. Brooks III[*,†,‡]

† Department of Chemistry, University of Michigan, 930 N. University, Ann Arbor, Michigan 48109, United States

‡ Biophysics Program, University of Michigan, 930 N. University, Ann Arbor, Michigan 48109, United States

Supplementary Discussion on Modeling Interactions between Coupled Titrating Residues

Further validation of the CPHMD$^{MS\lambda D}$ framework was performed on model dipeptide sequences Asp-Asp, Glu-Glu and Lys-Lys at zero ionic strength, where both residues were titrated simultaneously. The calculated pK_a values are summarized in Table S2 and Figure S1. For the aspartic acid dipeptide, we observed that the pK_a values were 3.1 and 4.6, with the N-terminus Asp having a consistently lower pK_a in all 3 simulations runs, suggesting that the two Asp residues are in a different electrostatic environment. An analysis of the hydrogen bonding contacts that each Asp side chain forms with the backbone of the dipeptide (Figure S2) indicated that the N-terminus Asp had 3 hydrogen bond donors within a ~5 Å radius, compared to the C-terminus Asp that had only 2 hydrogen bond donors. Thus, the increased presence of hydrogen bond donors around the N-terminus Asp facilitated the stabilization of its charged unprotonated state, explaining the decrease of its calculated pK_a value. By contrast, the calculated pK_a values for the glutamic acid dipeptide was 4.3 for both residues with no apparent pK_a shift. Similarly, the pK_a values for the lysine dipeptide was ~10.4 for both residues. The identical pK_a for both N- and C-terminus residues of both Glu-Glu and Lys-Lys dipeptides suggest that the electrostatic environment around each residue is similar. This is supported by the observation that no

hydrogen bonding capable backbone atom was present in a ~5 Å proximity from the titrating functional group, and so the backbone interactions that were responsible for creating an asymmetric environment in Asp-Asp is significantly reduced in both Glu-Glu and Lys-Lys dipeptides.

The calculated Hill coefficients of 0.7 suggest that anti-cooperative coupling is the dominant mode of interaction between the two adjacent titrating residues of these dipeptide systems. Prior work by Bashford and Karplus have demonstrated that when two residues titrate in the same pH region and have the same intrinsic (microscopic) pK_a, such as the Glu-Glu and Lys-Lys dipeptides in this analysis, the magnitude of their coupled interaction can raise/lower the apparent (macroscopic) pK_a of the system.[1] The existing HH-equation (i.e. eqn. 7) to which we fitted our data to calculate a pK_a value is a rearranged form of the equation first proposed by Tanford and Roxby.[2] When there is no coupling with other titrating residues (i.e. $n = 1$), eqn. 7 reduces to a form that can be derived from a mean-field approximation.[1] When there is coupling with other titrating residues (i.e. $n \neq 1$), the convention is to add the Hill coefficients to describe the anti-cooperative proton binding behavior. However, prior work by Onufriev *et. al.* in their derivation of the decoupled site representation (DSR) framework has shown that this approach may not give the best fit to experimental macroscopic pK_a values.[3] Consequently, it is not unexpected that our analysis was unable to obtain the macroscopic pK_a of the Glu-Glu and Lys-Lys dipeptides, where one would expect to see two distinct pK_a values. If one wishes to elucidate the coupled pK_a behavior for these two dipeptides, the pK_a values can be recalculated by fitting it to a modified version of the HH-equation (see *Method*s eqn. 8 for details), which can be derived from the DSR approach.[3] In this revised fitting method, where we analyzed the net proton uptake without pre-assigning the identity of each residue, the apparent pK_a values calculated cannot be

assigned to a specific titrating site (i.e. the calculated pK_a values are not the microscopic pK_a of specific residues). Using this approach, two clear and distinct pK_a values emerge for Glu-Glu (3.6, 5.0) and Lys-Lys (9.8, 11.0), which is consistent with the perturbation of one protonated residue on the other.

Derivation of Equation 7 from Mean Field Approximation

Here, we show how eqn. 7 of the main text can be derived from eqn. 1b (i.e. mean field approximation) from Bashford and Karplus.[1] We start with eqn. 1b of ref 1:

$$\log\frac{\theta}{1-\theta} = pKa - pH \tag{1}$$

where θ is the probability that the site is protonated:

$$\theta = \frac{N_p}{N_p + N_u} \tag{2}$$

Therefore eqn. 1b can be rewritten as:

$$\log\frac{\theta}{1-\theta} = \log\frac{\left(\frac{N_p}{N_p + N_u}\right)}{\left(\frac{N_u}{N_p + N_u}\right)} = pKa - pH \tag{3}$$

$$\log\frac{N_p}{N_p + N_u} = \log\frac{N_u}{N_p + N_u} + pKa - pH \tag{4}$$

$$\log\frac{N_p}{N_p + N_u} = \log\frac{N_u}{N_p + N_u} + \log\left(10^{pKa-pH}\right) \tag{5}$$

$$\frac{N_p}{N_p + N_u} = \frac{N_u}{N_p + N_u}\left(10^{pKa-pH}\right) \tag{6}$$

$$\frac{N_u}{N_p + N_u} + \frac{N_u}{N_p + N_u}\left(10^{pKa-pH}\right) = 1 \tag{7}$$

$$\frac{N_u}{N_p + N_u}\left(1+10^{pKa-pH}\right)=1 \tag{8}$$

$$\frac{N_u}{N_p + N_u} = \frac{1}{1+10^{pKa-pH}} = \frac{1}{1+10^{-(pH-pKa)}} \tag{9}$$

which is the same form as eqn. 7 for n = 1:

$$S = \frac{N_u}{N_p + N_u} = \frac{1}{1+10^{-(pH-pKa)}} \tag{10}$$

Derivation of Equation 8 from Decoupled Site Representation

Here, we show how eqn. 8 of the main text can be derived from eqn. 15 (i.e. decoupled site representation) from Onufriev *et. al.*[3] We start with eqn. 15 of ref 3:

$$\sum_i^N \langle x_i \rangle = \sum_j^N \frac{10^{pKa_j - pH}}{1+10^{pKa_j - pH}} \tag{11}$$

Since x_i represents the fraction of protonated states for each titrating residue *i*:

$$\sum_i^N \langle x_i \rangle = \sum_i^N \left(1 - S_i^{unprot}\right) \tag{12}$$

$$\sum_i^N \left(1 - S_i^{unprot}\right) = \sum_j^N \frac{10^{pKa_j - pH}}{1+10^{pKa_j - pH}} \tag{13}$$

$$\sum_i^N S_i^{unprot} = \sum_j^N \left(1 - \frac{10^{pKa_j - pH}}{1+10^{pKa_j - pH}}\right) = \sum_j^N \left(\frac{1}{1+10^{pKa_j - pH}}\right) = \sum_j^N \left(\frac{1}{1+10^{-(pH-pKa_j)}}\right) \tag{14}$$

The DSR framework maps a set of *i* real sites to a set of *j* non-interacting quasi-sites. Assuming a one-to-one mapping of real to quasi sites ($i = j$), we obtain the following expression, which is the same expression as eqn. 8:

$$\sum_i^N S_i^{unprot} = \sum_i^N \left(\frac{1}{1+10^{-(pH-pKa_i)}}\right) \tag{15}$$

Table S1: Sampling characteristics of 2-state titration simulations performed at pH = pK_a.

Residue	Fraction of Physical States	Transition (ns^{-1})
Asp	0.74 ± 0.04	35 ± 5
Glu	0.75 ± 0.02	35 ± 5
His-δ	0.72 ± 0.03	60 ± 10
His-ε	0.71 ± 0.04	64 ± 14
Lys	0.76 ± 0.03	50 ± 8

Table S2: Calculated pK_a of various model dipeptide sequences. Values reported in the top table were calculated using equation 7 (identity of residue was pre-assigned), and those reported in the bottom table were calculated using equation 8 (identity of residue was not pre-assigned).

Residue Identity Pre-Assigned					
Residue	**Ref pK_a (of amino acid)**	**Site1**		**Site2**	
		pK_a	**n**	**pK_a**	**n**
Asp-Asp	4.0	3.1 ± 0.2	0.7 ± 0.1	4.6 ± 0.1	0.7 ± 0.0
Glu-Glu	4.4	4.3 ± 0.1	0.7 ± 0.0	4.3 ± 0.1	0.7 ± 0.0
Lys-Lys	10.4	10.3 ± 0.0	0.7 ± 0.0	10.5 ± 0.1	0.7 ± 0.0
Residue Identity Not Pre-Assigned					
Residue	**Ref pK_a (of amino acid)**	**Site1[a]**		**Site2[a]**	
		pK_a	**n**	**pK_a**	**n**
Glu-Glu	4.4	3.6 ± 0.0	-	5.0 ± 0.0	-
Lys-Lys	10.4	9.8 ± 0.1	-	11.0 ± 0.1	-

[a] Site1 and Site2 pK_a values are defined as the residue that produces the lower and higher "instantaneous" pK_a value. When averaged across the entire trajectory, they would correspond to the two macroscopic pK_a values recorded by experiments.

Table S3: pK_a values of HEWL calculated using explicit solvent pH-REX CPHMD$^{MS\lambda D}$ simulations show good convergence within a 20 ns trajectory.

Residue	Exp pK_a	Explicit Solvent CPHMD$^{MS\lambda D}$ pK_a			
		0-5ns	6-10ns	11-15ns	16-20ns
GLU-7	2.6 ± 0.2	3.3 ± 0.2	3.1 ± 0.3	2.9 ± 0.1	2.7 ± 0.1
HIS-15	5.5 ± 0.2	5.6 ± 0.2	5.6 ± 0.1	5.8 ± 0.1	6.0 ± 0.2
ASP-18	2.8 ± 0.3	1.9 ± 0.0	2.1 ± 0.1	2.0 ± 0.2	2.1 ± 0.2
GLU-35	6.1 ± 0.4	7.3 ± 0.3	7.3 ± 0.2	6.8 ± 0.4	7.0 ± 0.3
ASP-48	1.4 ± 0.2	1.3 ± 0.0	1.3 ± 0.0	1.3 ± 0.0	1.3 ± 0.0
ASP-52	3.6 ± 0.3	4.6 ± 0.2	4.8 ± 0.4	4.5 ± 0.3	4.5 ± 0.3
ASP-66	1.2 ± 0.2	1.4 ± 0.0	1.4 ± 0.0	1.4 ± 0.1	1.5 ± 0.1
ASP-87	2.2 ± 0.1	1.3 ± 0.0	1.3 ± 0.0	1.3 ± 0.0	1.3 ± 0.0
ASP-101	4.5 ± 0.1	5.5 ± 0.4	5.6 ± 0.5	5.6 ± 0.6	5.1 ± 0.5
ASP-119	3.5 ± 0.3	1.9 ± 0.1	1.9 ± 0.1	1.7 ± 0.1	1.6 ± 0.0
RMSE		**0.91**	**0.92**	**0.85**	**0.84**

Figure S1: Titration curves of (a) Asp-Asp, (b) Glu-Glu and (c) Lys-Lys dipeptide sequences. The first two curves of each dipeptide correspond to the titration curve for the N and C-terminal residue respectively, which were fitted using eqn 7. The third curve corresponds to the titration curve of the dipeptide fitted using eqn 8. Colors represent the results from the triplicate runs.

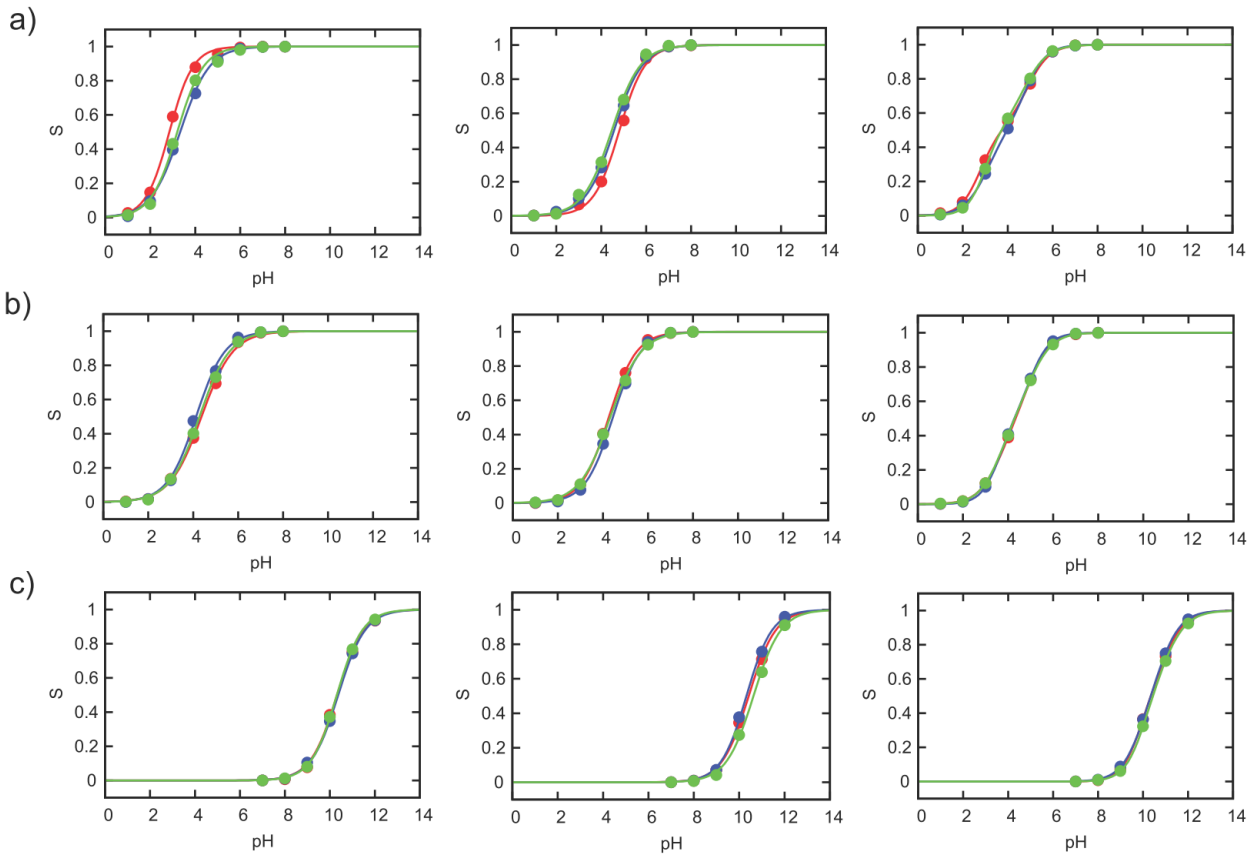


Figure S2: Average distances between the OD1/OD2 atoms on Asp and the N atoms on the N-H groups of the peptide backbone for the (a) N-terminus and (c) C-terminus Asp residues in aspartic acid dipeptide as a function of pH.

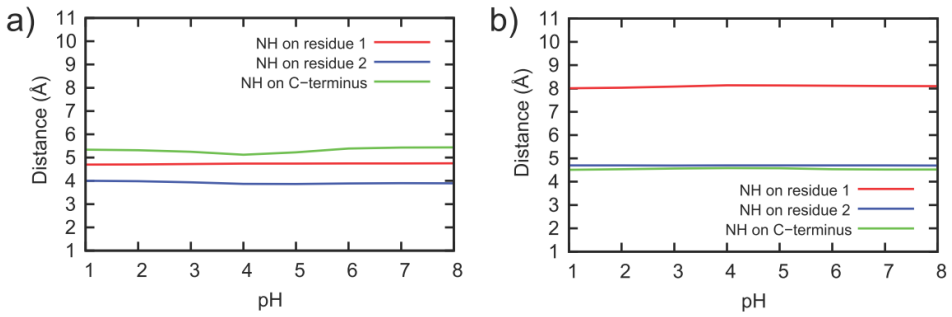


Figure S3: Titration curves of Asp/Glu/His residues of HEWL from the last 5 ns segment of the trajectory. Colors represent the results from the triplicate runs.

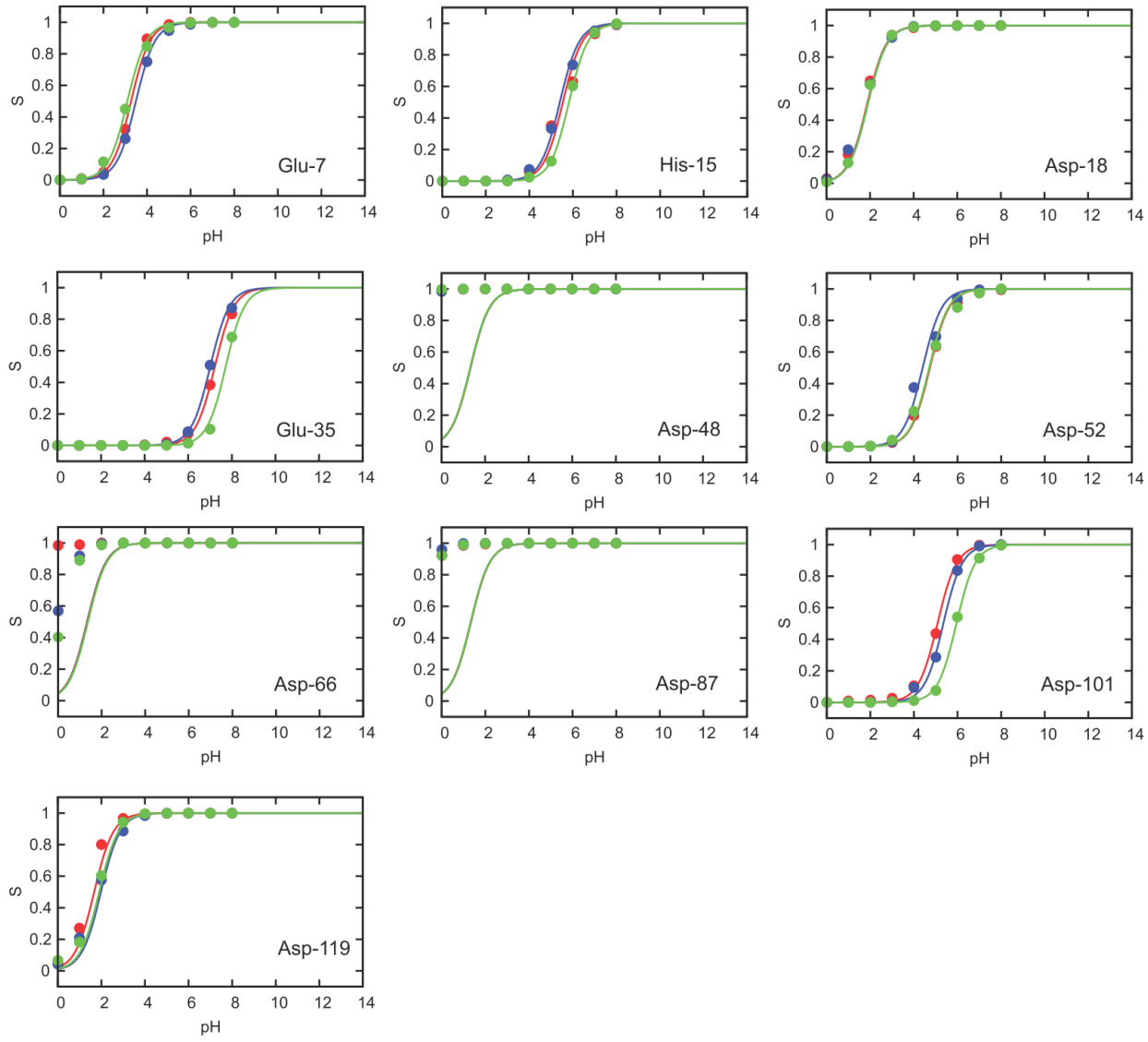


Figure S4: Histogram of hydrogen bonding distances between the OD1/OD2 atom of Asp-119 with the amide backbone hydrogen (HN) of Gln-121 and Ala-121 indicates the presence of a persistent hydrogen bond at all pH environments.

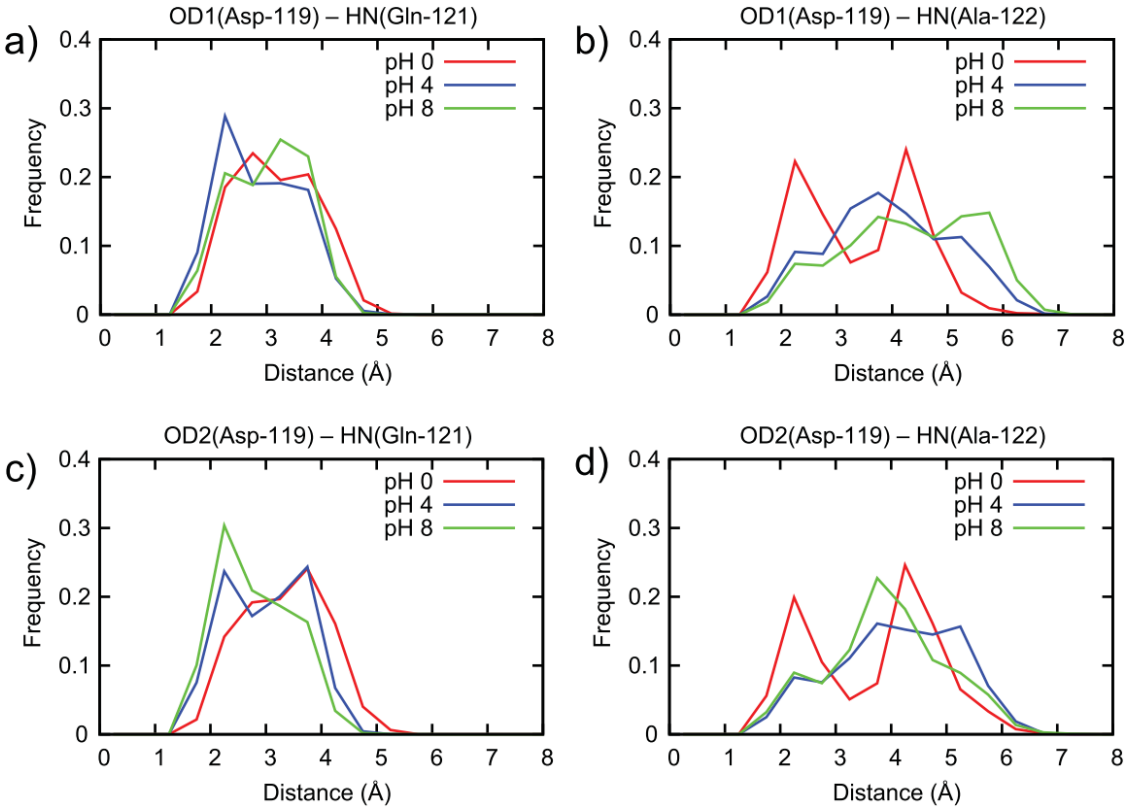
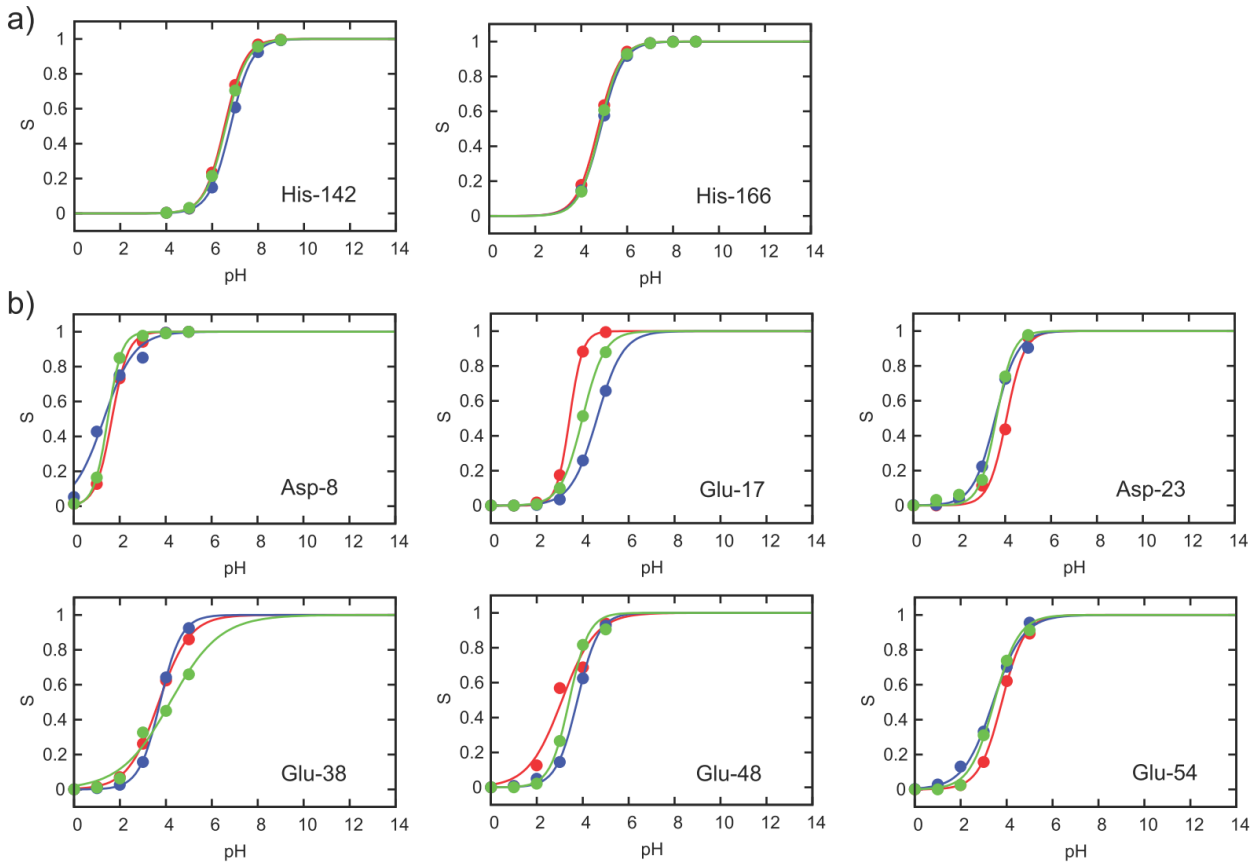


Figure S5: Titration curves of Asp/Glu/His residues of (a) BBL and (b) NTL9. Colors represent the results from the triplicate runs.



References

(1) Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 9556.
(2) Tanford, C.; Roxby, R. *Biochemistry* **1972**, *11*, 2192.
(3) Onufriev, A.; Case, D. A.; Ullmann, G. M. *Biochemistry* **2001**, *40*, 3413.