1 Supporting Information

1.1 Imidazole and histidine side chain force field atomic charges

Table s1: Atomic charges of histidine side chain (HIS, from GROMOS96 53a6[1])
and imidazole (IM) in different protonation states (Figure 4 main text).

atom	HIS (00)	IM (00)	HIS (10)	IM (10)	HIS (01)	IM (01)	HIS/IM (11)
CB/HG	0.00	0.10	0.00	0.11	0.00	0.11	0.11
CG	-0.05	-0.10	0.00	-0.02	0.00	-0.02	-0.02
ND1	0.38	0.31	-0.05	-0.04	-0.54	-0.54	-0.64
HD1	0.30	0.30	0.31	0.31	-	-	-
CE1	-0.34	-0.32	0.00	-0.02	0.00	-0.02	-0.02
HE1	0.10	0.10	0.14	0.11	0.14	0.11	0.11
NE2	0.31	0.31	-0.54	-0.54	-0.05	-0.04	-0.64
HE2	0.30	0.30	-	-	0.31	0.31	-
CD2	-0.10	-0.10	0.00	-0.02	0.00	-0.02	-0.02
HD2	0.10	0.10	0.14	0.11	0.14	0.11	0.11

1.2 Contribution of bonded terms to deprotonation free energy of glutamic acid

Table s2: Free energy of deprotonation ΔG (kJ mol^{-1}) of glutamic acid with neutral termini in explicit solvent from 5 ns free energy simulation. It is indicated if nonbonded (charges, Lennard-Jones) and bonded (bonds, angles, torsions) force field parameters are mantained in the protonated state (AH) or are perturbed to the deprotonated state (A$^-$).

nonbonded	bonded	ΔG
AH$\rightarrow$A$^-$	AH$\rightarrow$AH	-220.8$\pm$1.9
AH$\rightarrow$A$^-$	AH$\rightarrow$A$^-$	-219.1$\pm$2.0

1.3 Coefficients of polynomial fits for the reference states

Table s3: Coefficients (a_0, a_1, a_2, and a_3) of the third-order polynomial fits in λ for the reference states. For imidazole (Im), where the N_δ and N_ϵ sites are chemically coupled, each coefficient a_i is described by a second third-order polynomial (with coefficients $a_{i,0}$, $a_{i,1}$, $a_{i,2}$, and $a_{i,3}$) in λ of the respective coupled site. The parameters for Im-N_ϵ (not listed) are the same as for Im-N_δ.

titratable site	a_0	a_1	a_2	a_3
Glu	-129.1	788.0	-262.6	174.0
N-terminus	-574.2	556.1	-216.3	141.6
C-terminus	-117.8	766.6	-207.6	139.6
Im-$N_\delta(a_{i,0})$	-453.7	746.4	-684.8	613.1
Im-$N_\delta(a_{i,1})$	542.1	-667.6	1244.1	-745.2
Im-$N_\delta(a_{i,2})$	-101.8	567.6	-888.6	306.4
Im-$N_\delta(a_{i,3})$	7.5	-195.6	121.2	71.3

1.4 Error estimate for the average protonation state

For a constant pH trajectory of length T, the statistical accuracy of the average protonation increases with increasing number of observed transitions between protonated state (*e.g.*, $\lambda < 0.1$) and deprotonated state (*e.g.*, $\lambda > 0.9$). Given a number of n_0 deprotonation and n_1 protonation transitions, defined respective off and on rates $\kappa_0 : 0 \to 1$ and $\kappa_1 : 1 \to 0$, a mean protonation $\langle \lambda \rangle$, as well as its statistical uncertainty were estimated via a Bayesian approach as follows.

With probabilities $\kappa_0 \Delta t$ and $\kappa_1 \Delta t$ for a deprotonation and protonation event, respectively, to occur during a small time interval Δt, the total probability of finding a given number of protonation/deprotonation events amongst all possible protonation/deprotonation sequences is

$$p(n_0, n_1 | \kappa_0, \kappa_1) = \alpha \, (\kappa_0 \Delta t)^{n_0} (1 - \kappa_0 \Delta t)^{N_0 - n_0} (\kappa_1 \Delta t)^{n_1} (1 - \kappa_1 \Delta t)^{N_1 - n_1} \quad , \tag{1}$$

with α as combinatorial pre-factor that will cancel out further below upon normalization, and $N_0 = T_0 / \Delta t$, $N_1 = T_1 / \Delta t$, with T_0 and T_1 the total time spent in states 0 and 1, respectively ($T = T_0 + T_1$).

Equation 1 can be approximated for $n_0 << N_0$ and $n_1 << N_1$ by the Poisson distribution

$$p(n_0, n_1 | \kappa_0, \kappa_1) = \frac{\alpha}{N_0^{n_0} N_1^{n_1}} \, (T_0 \kappa_0)^{n_0} (T_1 \kappa_1)^{n_1} \exp\left[-T_0 \kappa_0 - T_1 \kappa_1\right] \quad . \tag{2}$$

Applying Bayes Theorem, and assuming uniform *a priori* probabilities $p(\kappa_0)$ and $p(\kappa_1)$, one obtains for the probability distribution of protonation and deprotonation rates, given a number of protonation/deprotonation events,

$$p(\kappa_0, \kappa_1 | n_0, n_1) = \frac{T_0^{n_0+1} T_1^{n_1+1}}{n_0! n_1!} \kappa_0^{n_0} \kappa_1^{n_1} \exp\left[-T_0 \kappa_0 - T_1 \kappa_1\right] \quad . \tag{3}$$

3

From this distribution, the expectation value for the average deprotonation ($\langle\lambda\rangle = \frac{\kappa_0}{\kappa_0+\kappa_1}$) is readily obtained,

$$\langle\lambda\rangle = \int_0^\infty \int_0^\infty p(\kappa_0, \kappa_1|n_0, n_1)\frac{\kappa_0}{\kappa_0 + \kappa_1}d\kappa_0 d\kappa_1 \quad , \tag{4}$$

and similarly its variance,

$$\sigma_\lambda^2 = \int_0^\infty \int_0^\infty p(\kappa_0, \kappa_1|n_0, n_1)\left(\frac{\kappa_0}{\kappa_0 + \kappa_1} - \langle\lambda\rangle\right)^2 d\kappa_0 d\kappa_1 \quad . \tag{5}$$

To estimate the uncertainty of $\langle\lambda\rangle$, the upper and lower bounds of the confidence interval (95%) of the $\langle\lambda\rangle$ distribution are calculated. The $\langle\lambda\rangle$ distribution is approximated by a density histogram (of bin-with 0.01), which is obtained by solving numerically the integral in Equation 4.

1.4.1 Imidazole and Histidine

Imidazole and histidine are described by two titration coordinates, as there are two titratable sites on the imidazole moiety (Figure 4). When estimating the error of $\langle\lambda\rangle$ during the first deprotonation reaction of one site, κ_0 and κ_1 described the forward and backward transitions, respectively, between the fully protonated, positive state (state 00) and the respective neutral form (state 10 or 01). The portion of the simulation where the other neutral form occurred was not included in the analysis. Similarly was done for the error of the second microscopic deprotonation reaction. In contrast, to estimate the error in the overall first deprotonation of imidazole and histidine, the whole simulation was considered. In this case κ_0 and κ_1 described the forward and backward transition rates, respectively, for the deprotonation of the fully protonated, positive state (state 00) to yield any two of the neutral forms (state 10 and 01). The λ-coordinate corresponded in this case to a *macroscopic* titration coordinate. The error for the second deprotonation reaction was estimated

4

similarly.

1.5 Force field comparison: GROMOS96 and OPLSA

To assess the sensitivity of the constant pH molecular dynamics approach to the chosen force field, we calculated titration curves and $p\mathrm{K_a}$ values of glutamic acid and of the tri-peptide Ala-Glu-Ala with two different force fields. Figure s1 compares the titration curves of glutamic acid described with GROMOS96[1] and SPC[2] water model (in black), and with OPLSA and TIP4P[3] water model (in red). Figure s2 shows the overall (left) and site-specific (right) titration curves of the tripeptide Ala-Glu-Ala calculated using GROMOS96[1] and SPC[2] water model (in black), and OPLSA[4] and SPC[2] water model (in red).

References

(1) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(2) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydratation. In *Intermolecular Forces*. Pullman, B., Ed.; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1981; pp 331–342.

(3) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

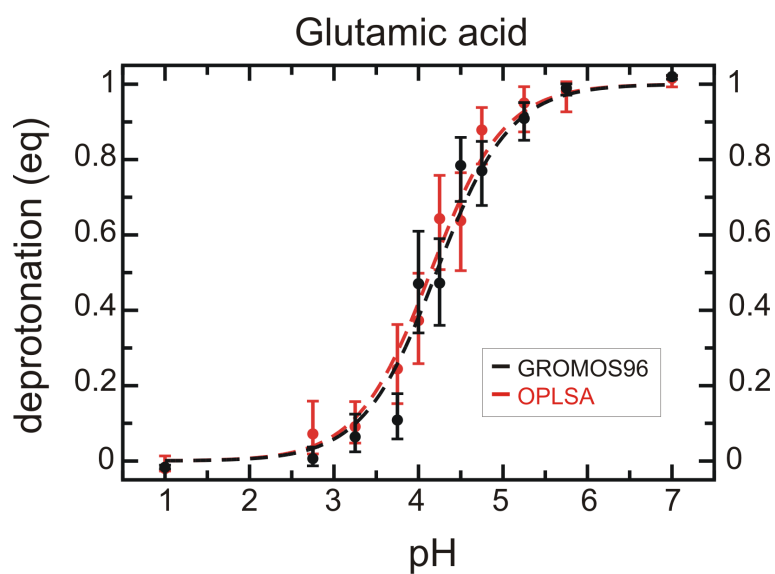(4) Jorgensen, W. L.; Maxwell, S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

Figure s1: Calculated titration curve of glutamic acid with neutral termini described with GROMOS96[1] (and SPC[2] water model) and OPLSA[4] (and TIP4P[3] water model) in black and red, respectively. At each of 11 pH values, the average deprotonation (in equivalents, eq) from 4 simulations of 4.5 ns each is shown (filled dots). The dashed line is a Henderson-Hasselbalch fit to the average data. Error bars denote estimates from the statistics of the observed transitions.
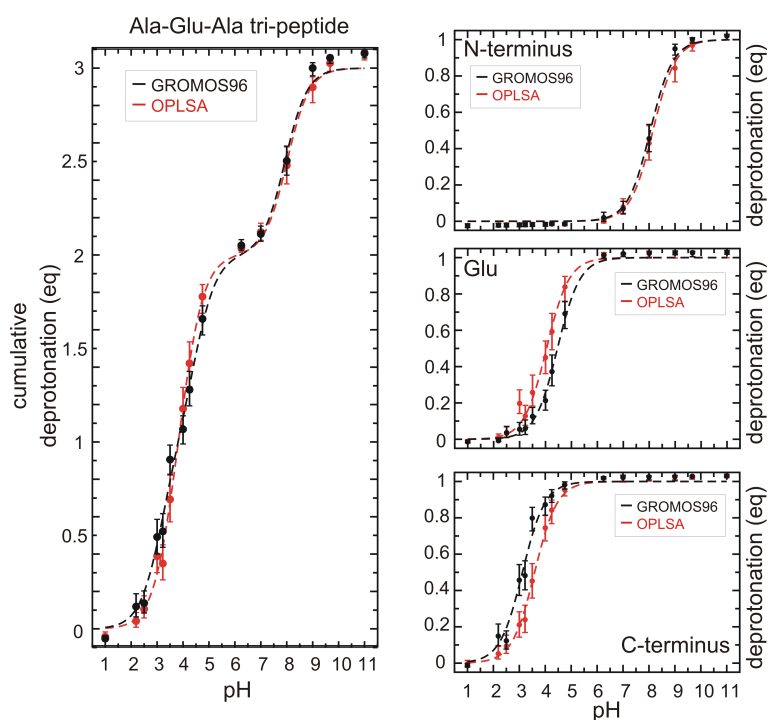
Figure s2: Calculated titration curves of Ala-Glu-Ala tri-peptide described with GROMOS96[1] (and SPC[2] water model) and OPLSA[4] (and SPP[2] water model) in black and red, respectively. Left: titration curve of the tri-peptide; right: from top to bottom, site-specific titration curves of the N-terminus, Glu, and C-terminus, respectively. At each of 15 pH values, the average deprotonation (in equivalents, eq) from 4 simulations of 7.5 ns each is shown (filled dots). The dashed line is a Henderson-Hasselbalch fit to the average data. Error bars denote estimates from the statistics of the observed transitions.