## *Supporting information*

# *Site-specific protonation kinetics of acidic side chains in proteins determined by pH-dependent carboxyl $^{13}C$ NMR relaxation*

Johan Wallerstein,[†] Ulrich Weininger,[†] M. Ashhar I. Khan,[†,§] Sara Linse,[‡] and Mikael Akke[†,*]

[†] Department of Biophysical Chemistry and [‡]Department of Biochemistry and Structural Biology, Center for Molecular Protein Science, Lund University, P.O. Box 124, SE-221 00 Lund, Sweden

[§] Department of Chemistry, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, 110 016 India

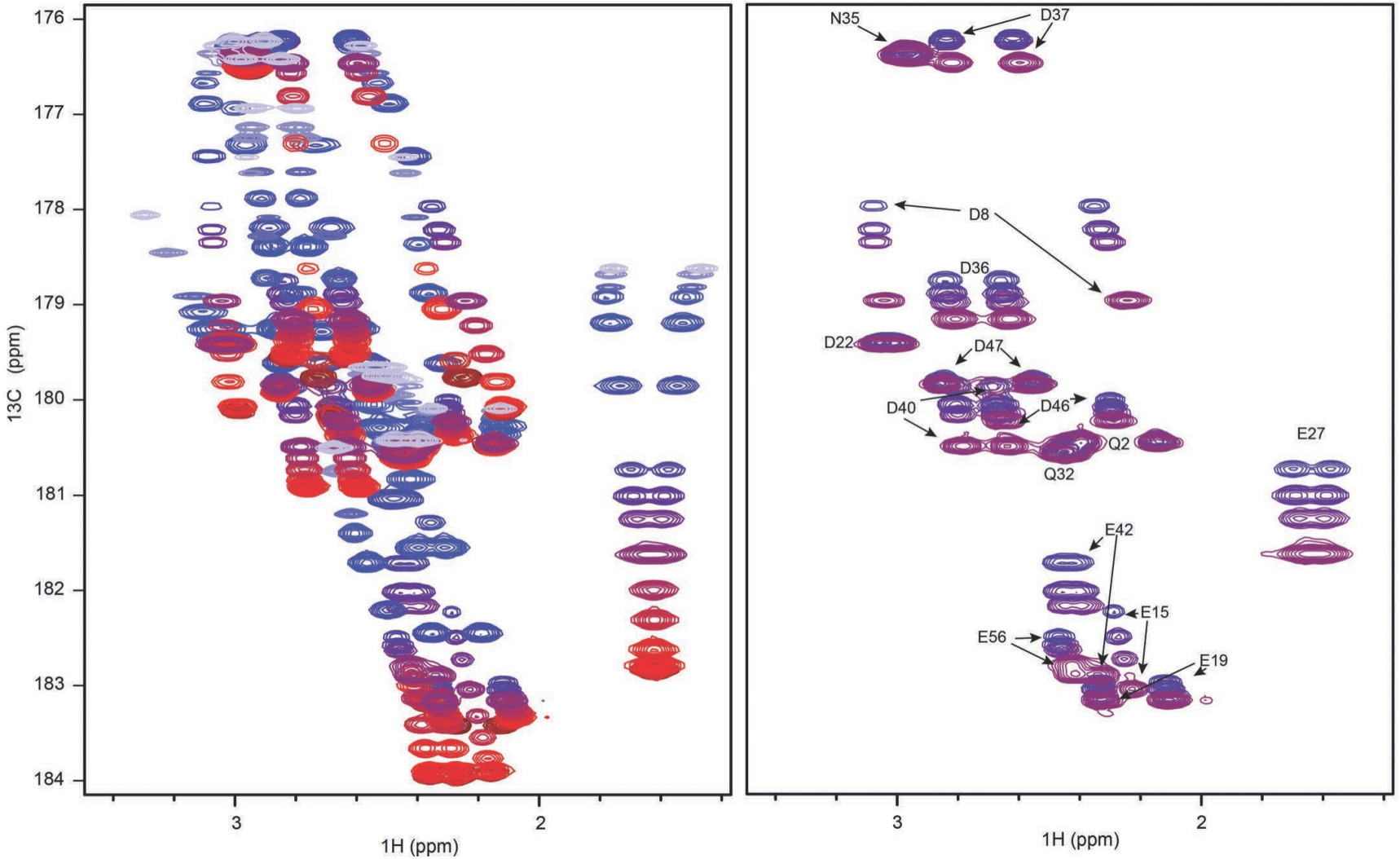


**Figure S1**. $^{13}C^{\gamma/\delta}$–$^{1}H^{\gamma/\delta}$ correlation spectra of PGB1-QDD acquired at 25 °C. The left-hand panel shows all 21 spectra are acquired at pH values ranging from 1.91 to 8.55. For clarity, resonance assignments are indicated in the right-hand panel, which shows a subset of 4 spectra acquired at pH values between 4.80 and 5.34.

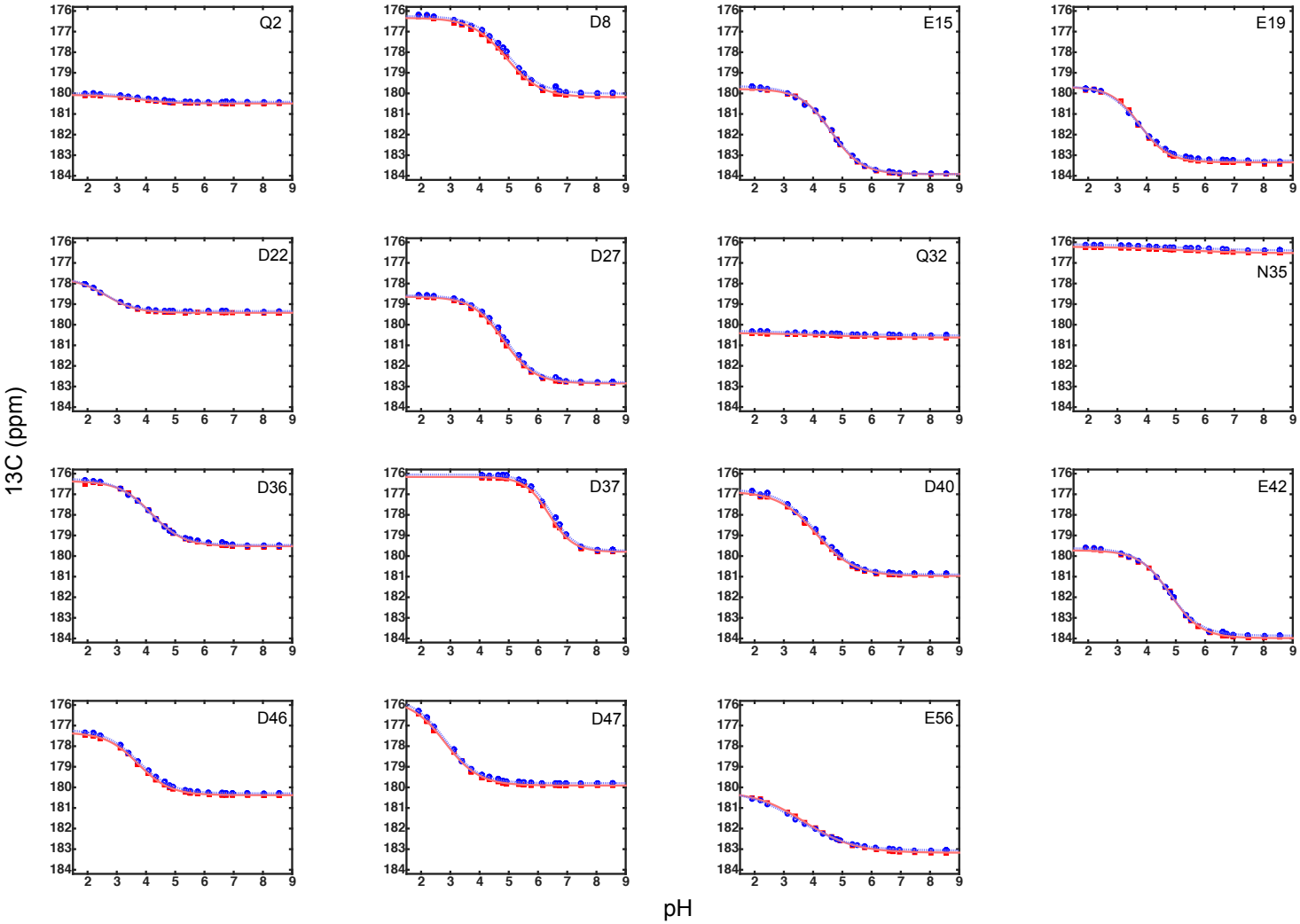


**Figure S2**. Carboxyl and carbonyl $^{13}C$ chemical shifts as a function of pH. Data are shown for all 15 D, E, N, or Q residues in PGB1-QDD. Data were acquired at two temperatures: 15 °C (blue circles) and 25 °C (red squares). The blue (15 °C) and red (25 °C) curves represent best fits of eqs 8–9 to the data. For D37, the data points at pH < 4 are excluded from the fit (see the main text). $\Delta\delta$ is to a very good approximation the difference between the two endpoint shifts at low ($\delta_{HA}$) and high ($\delta_{A}$) pH. Error bars indicate 1 standard deviation and are typically smaller than the symbol size.

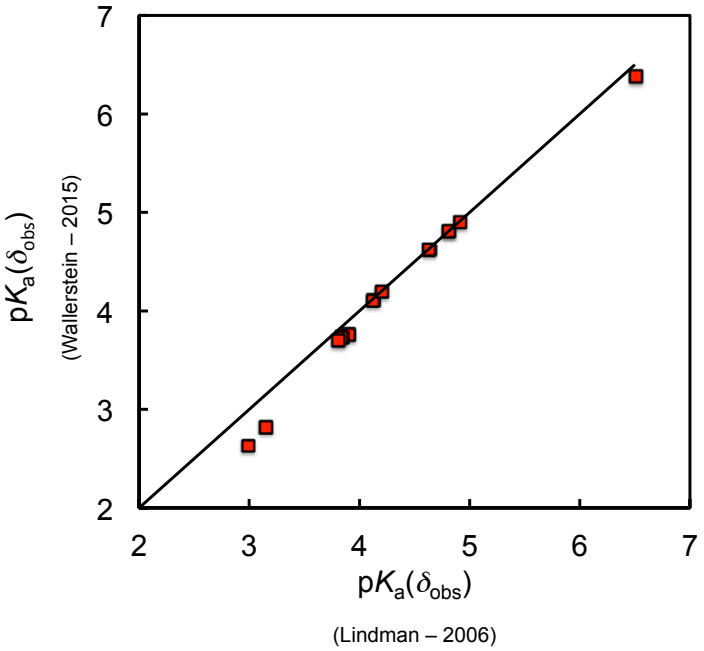


**Figure S3**. Comparison of p$K_a$ values determined for carboxyl groups in PGB1-QDD from chemical shift titrations at 25 °C in the present study (vertical axis, labeled Wallerstein – 2015) and by Lindman et al.[1] (horizontal axis, labeled Lindman – 2006). All 12 acidic residues are included. E27 and E42 have identical coordinates (4.81, 4.81). RMSD = 0.16.

.

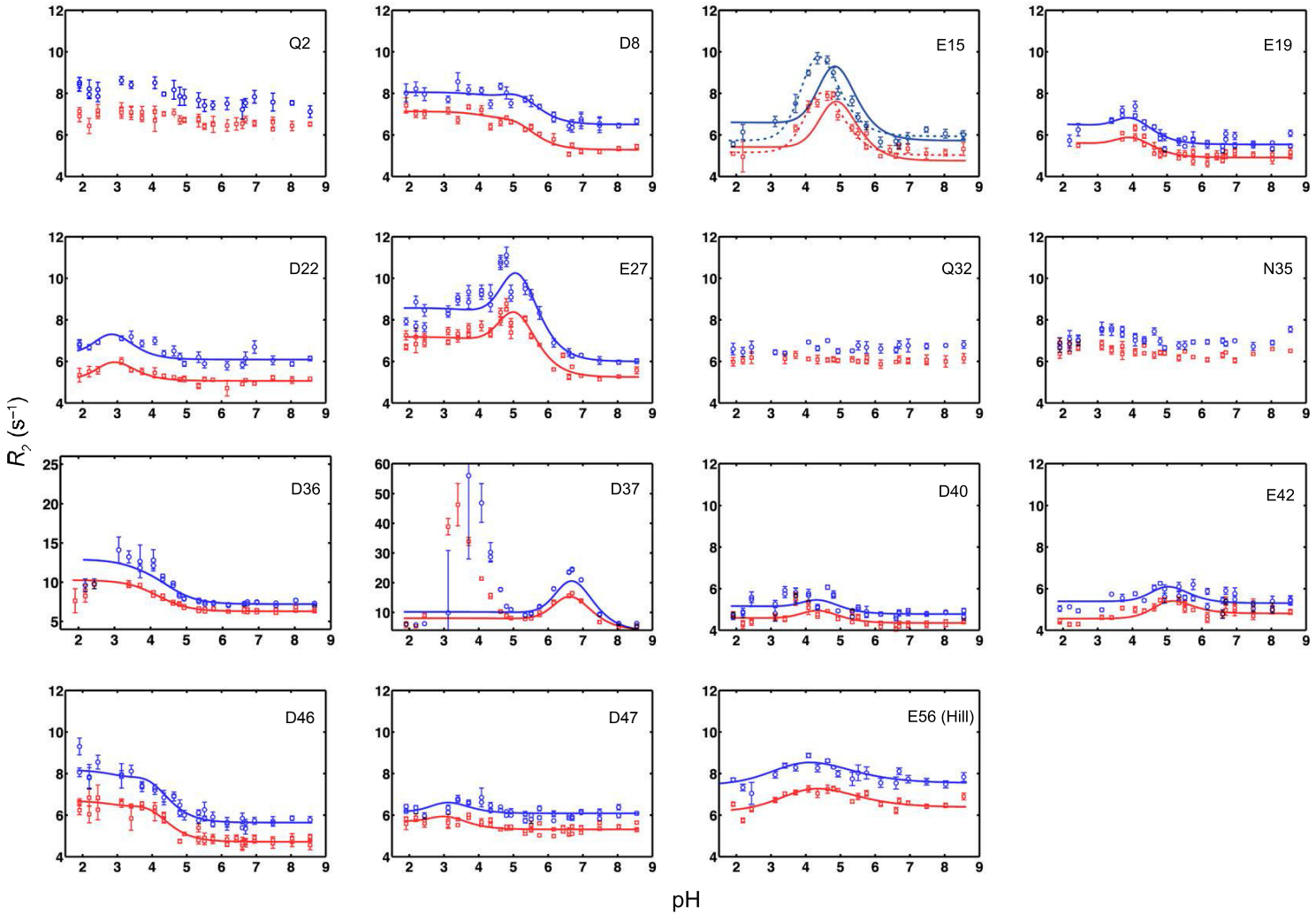


**Figure S4**. $R_2$ vs. pH relaxation dispersion profiles. Data are shown for all 15 D, E, N, or Q residues in PGB1-QDD. Blue circles refers to data at 15 °C and red squares at 25 °C. Two data points are available at some pH values for those cases where two separate $^1H^{\beta}$-$^{13}CO$ cross-peaks are observed. The continuous blue and red curves describe the best-fit model, eq 7, including the free parameters $k_{off}$, $R_{2,A}$ and $R_{2,HA}$ (3-parameter fit), see Table S2, while keeping $K_a$ and $\Delta\omega$ fixed at the values determined from the chemical shift titration. For E15, dotted lines represent fits to eq 7 with $K_a$ included as a free parameter (4-parameter fit). Error bars indicate 1 standard deviation.

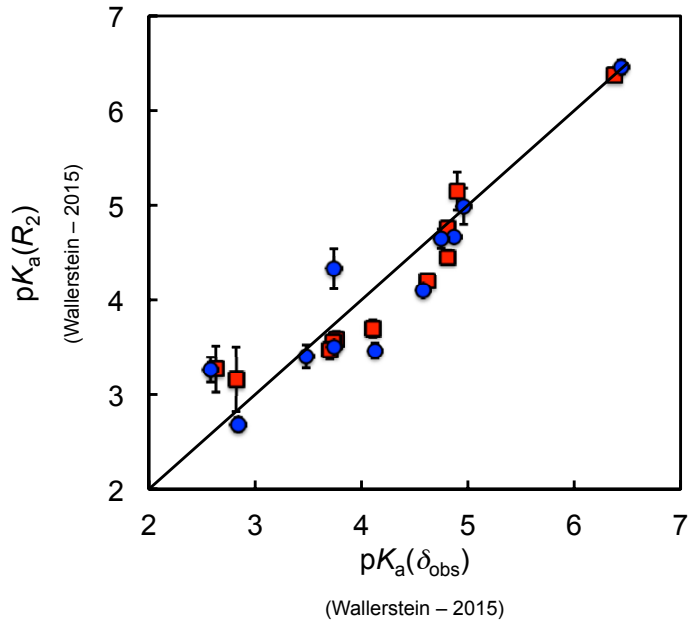


**Figure S5**. Correlation between p$K_a$ values determined in the present study from chemical shift titrations (horizontal axis, labeled p$K_a$($\delta_{obs}$); fit of $\delta_{obs}$ to eq 8) and from $R_2$ vs pH data (vertical axis, labeled p$K_a$($R_2$); fit of $R_2$ to eq 7) at 15 °C (blue circles) and 25 °C (red squares) for 11 out of 12 acidic residues (excluding D36; see the main text) in PGB1-QDD. The solid line represents perfect agreement (slope = 1) and is drawn to guide the eye. The correlation coefficient between the data sets is $r^2$ = 0.87 at 15 °C and 0.90 at 25 °C. The slope of the linear regression trendline (not shown) is 0.90 for both temperatures, indicating a tendency towards higher p$K_a$ for shift measurements.

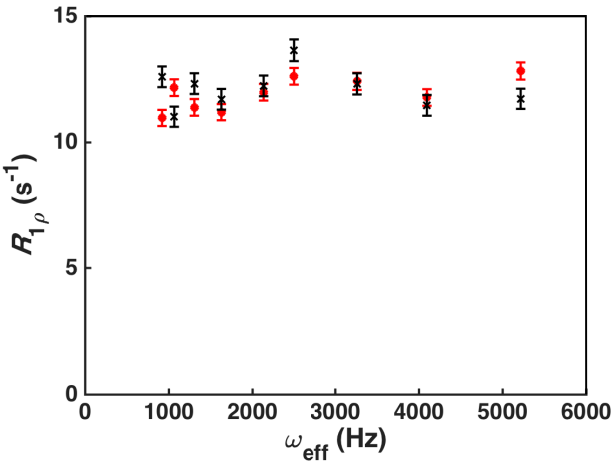


**Figure S6.** Carboxyl $^{13}C$ $R_{1\rho}$ relaxation dispersion profile for the side-chain carboxyl carbon in residue D37 at pH 7.48 and 15 °C. Error bars indicate 1 standard deviation. The carrier is placed on resonance at 179.6 ppm.

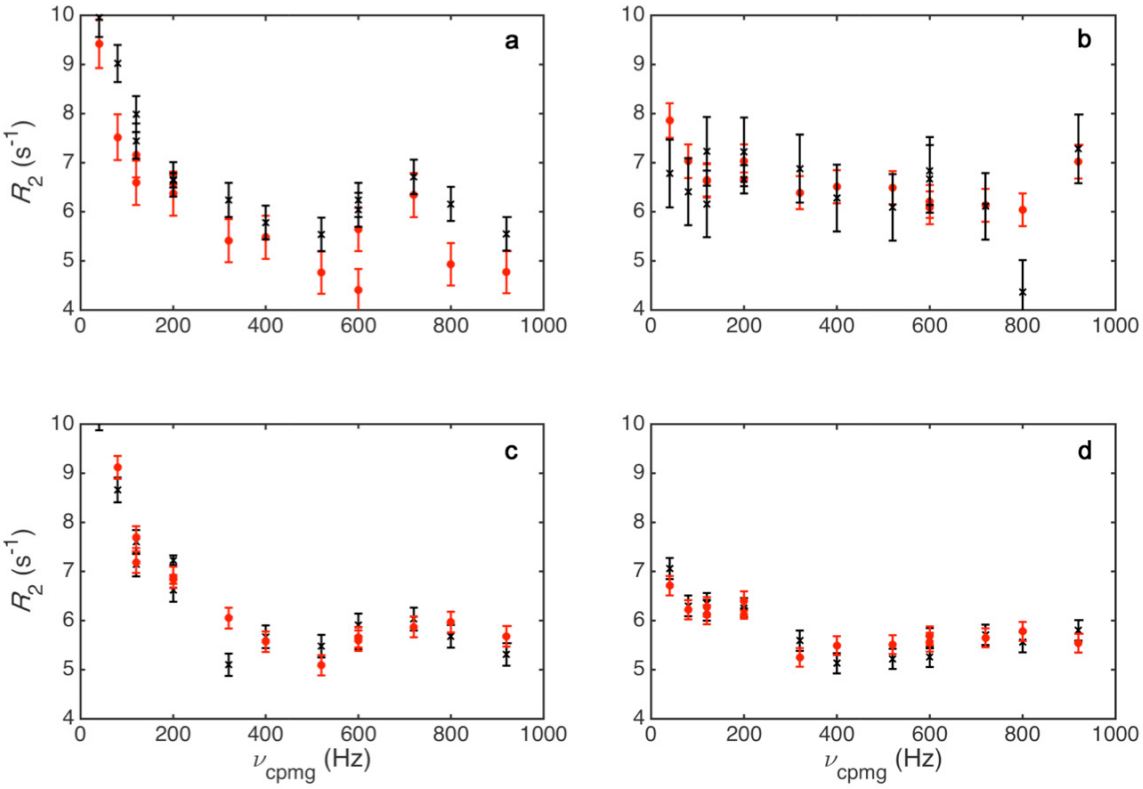


**Figure S7**. Carboxyl $^{13}C$ CPMG relaxation dispersion profiles for D8 (a and b) and N35 (c and d), acquired at pH 2.19 (a and c) and 4.20 (b and d) and 25 °C. Red circles and black crosses denote data obtained from the two separate $^{1}H^{\beta}$-$^{13}CO$ cross-peaks. Error bars indicate 1 standard deviation. These data demonstrate that a refocusing frequency ($\nu_{CPMG}$) of 800 Hz mitigates the contribution from slow conformational exchange observed at low pH; $\nu_{CPMG}$ = 800 Hz was used to acquire the $R_2$ data shown in Fig. 2 and S4.

**Table S1. Fitted parameters and model selection in fitting eqs (7) and (8) to the data in Figures S4 and S2, respectively.**

| Residue | p$K_a$ ($R_2$)[a] | | p$K_a$ ($\delta_{obs}$)[b] | | $\Delta\delta$ (ppm)[c] | | $n_H$[d] | | $p$-value from $F$-test[e] | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 25 °C | 15 °C | 25 °C | 15 °C | 25 °C | 15 °C | 25 °C | 15 °C | 25 °C | 15 °C |
| D8 | 5.15±0.20 | 4.99±0.19 | 4.90±0.01 | 4.96±0.01 | 3.86±0.02 | 3.77±0.02 | 0.74±0.01 | 0.77±0.01 | < 0.001 | < 0.001 |
| E15 | 4.20±0.03 | 4.10±0.04 | 4.62±0.01 | 4.58±0.01 | 4.13±0.02 | 4.28±0.02 | 0.84±0.01 | 0.77±0.01 | < 0.001 | < 0.001 |
| E19 | 3.58±0.08 | 3.50±0.05 | 3.76±0.01 | 3.74±0.01 | 3.66±0.02 | 3.63±0.02 | 0.90±0.01 | 0.82±0.01 | < 0.001 | < 0.001 |
| D22 | 3.27±0.24 | 3.26±0.13 | 2.63±0.10 | 2.58±0.08 | 1.69±0.14 | 1.65±0.09 | 0.83±0.07 | 0.85±0.07 | < 0.001 | < 0.001 |
| E27 | 4.76±0.02 | 4.67±0.04 | 4.81±0.01 | 4.87±0.01 | 4.21±0.01 | 4.21±0.01 | 0.82±0.01 | 0.82±0.01 | < 0.001 | < 0.001 |
| D36 | 4.04±0.12 | 3.14±0.34 | 4.19±0.01 | 4.09±0.01 | 3.16±0.02 | 3.23±0.02 | 0.79±0.01 | 0.75±0.01 | < 0.001 | < 0.001 |
| D37 | 6.38±0.06 | 6.46±0.07 | 6.38±0.01 | 6.44±0.01 | 3.64±0.01 | 3.69±0.01 | 1.05±0.02 | 1.05±0.02 | < 0.01 | < 0.01 |
| D40 | 3.69±0.09 | 3.46±0.08 | 4.10±0.01 | 4.12±0.01 | 4.10±0.02 | 4.14±0.02 | 0.67±0.01 | 0.68±0.01 | < 0.001 | < 0.001 |
| E42 | 4.45±0.08 | 4.65±0.10 | 4.81±0.01 | 4.75±0.01 | 4.27±0.02 | 4.25±0.02 | 0.83±0.01 | 0.79±0.01 | < 0.001 | < 0.001 |
| D46 | 3.55±0.08 | 4.33±0.21 | 3.73±0.01 | 3.74±0.01 | 3.04±0.02 | 3.09±0.02 | 0.79±0.01 | 0.77±0.01 | < 0.001 | < 0.001 |
| D47 | 3.16±0.34 | 2.68±0.07 | 2.82±0.03 | 2.84±0.02 | 4.17±0.08 | 4.18±0.07 | 0.75±0.02 | 0.77±0.02 | < 0.001 | < 0.001 |
| E56 | 3.47±0.10 | 3.40±0.12 | 3.70±0.04 | 3.48±0.06 | 3.01±0.05 | 2.96±0.07 | 0.51±0.02 | 0.48±0.02 | < 0.001 | < 0.001 |

[a] p$K_a$ values determined from $R_2$ vs pH data (Fig. S4) using eq (7) with 4 free parameters: $k_{off}$, $K_a$, $R_{2,A}$ and $R_{2,HA}$.

[b] p$K_a$ values determined from $\delta_{obs}$ vs pH data (Fig. S2) using eq (8) with 4 free parameters: $K_a$, $\delta_A$, $\delta_{HA}$, and $n_H$.

[c] $\Delta\delta$ values ($\Delta\delta = \delta_A - \delta_{HA}$) determined from $\delta_{obs}$ vs pH data (Fig. S2) using eq (8) with 4 free parameters: $K_a$, $\delta_A$, $\delta_{HA}$ and $n_H$.

[d] Hill parameters, $n_H$, determined from $\delta_{obs}$ vs pH data (Fig. S2) using eq (8) with 4 free parameters: $K_a$, $\delta_A$, $\delta_{HA}$ and $n_H$.

[e] Comparing fits to eq (8) with 3 ($K_a$, $\delta_A$, and $\delta_{HA}$) or 4 ($K_a$, $\delta_A$, $\delta_{HA}$, and $n_H$) free parameters. The $p$-value indicates the probability of falsely rejecting the null hypothesis that the 4-parameter model does not provide a significantly better fit than the 3-parameter model.

**Table S2. Results from *F*-statistics comparing fits of $R_2$ vs pH-profiles to eq (7) with 2, 3, or 4 free parameters.**

| Residue | $\chi^2(2)$[a] | $\chi^2(3)$[b] | $\chi^2(4)$[c] | *N*, data points[d] | *F*-ratio[e] | | *p*-value[f] | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 2 vs. 3 | 3 vs. 4 | 2 vs. 3 | 3 vs. 4 |
| **25 °C** | | | | | | | | |
| **D8** | **924.6** | **98.2** | **98.0** | **36** | **277.68** | **0.97** | **< 0.0001** | **0.53** |
| **E15** | **274.1** | **204.4** | **49.7** | **19** | **5.68** | **3.86** | **< 0.001** | **< 0.01** |
| **E19** | **167.5** | **95.5** | **82.9** | **32** | **23.10** | **1.11** | **< 0.0001** | **0.39** |
| **D22** | **37.8** | **37.7** | **35.4** | **21** | **0.06** | **1.01** | **1** | **0.49** |
| **E27** | **1581.4** | **147.3** | **139.5** | **31** | **221.46** | **1.02** | **< 0.0001** | **0.48** |
| **D36** | **1552.3** | **236.5** | **268.4** | **37** | **189.15** | **0.86** | **< 0.0001** | **0.67** |
| **D37[g]** | **3670.1** | **769.0** | **12584.6** | **19** | **60.37** | **0.06** | **< 0.0001** | **1** |
| **D40** | **210.2** | **180.3** | **147.0** | **38** | **5.81** | **1.19** | **< 0.0001** | **0.3** |
| **E42** | **122.0** | **101.3** | **73.3** | **31** | **5.73** | **1.33** | **< 0.0001** | **0.23** |
| **D46** | **625.6** | **63.3** | **57.4** | **38** | **316.68** | **1.07** | **< 0.01** | **0.42** |
| **D47** | **288.1** | **263.4** | **292.6** | **40** | **3.28** | **0.88** | **0.206** | **0.65** |
| **E56** | **82.0** | **87.2** | **87.2** | **21** | **1.47** | **0.94** | **< 0.0001** | **0.55** |
| **15 °C** | | | | | | | | |
| **D8** | **429.3** | **64.1** | **62.2** | **34** | **182.39** | **1.00** | **< 0.0001** | **0.54** |
| **E15** | **449.6** | **383.2** | **46.2** | **19** | **2.71** | **7.77** | **0.0261** | **< 0.001** |
| **E19** | **228.3** | **145.4** | **113.2** | **35** | **17.88** | **1.24** | **< 0.0001** | **0.27** |
| **D22** | **40.8** | **37.2** | **23.1** | **21** | **1.74** | **1.52** | **0.11** | **0.2** |
| **E27** | **1405.0** | **203.1** | **137.8** | **32** | **165.84** | **1.42** | **< 0.0001** | **0.17** |
| **D36** | **685.3** | **369.1** | **781.7** | **36** | **37.14** | **0.46** | **< 0.0001** | **0.99** |
| **D37[g]** | **4429.1** | **1065.0** | **15045.7** | **20** | **45.30** | **0.07** | **< 0.0001** | **1** |
| **D40** | **369.6** | **320.0** | **212.3** | **40** | **5.37** | **1.47** | **< 0.0001** | **0.12** |
| **E42** | **237.5** | **232.6** | **223.3** | **31** | **0.35** | **1.00** | **0.91** | **0.5** |
| **D46** | **539.9** | **63.8** | **62.9** | **36** | **179.32** | **0.98** | **< 0.0001** | **0.52** |
| **D47** | **212.9** | **211.9** | **162.3** | **40** | **0.27** | **1.27** | **1** | **0.23** |
| **E56** | **51.1** | **50.4** | **49.9** | **22** | **0.24** | **0.96** | **1** | **0.54** |

[a] $\chi^2(2)$ denotes $\chi^2$ resulting from a fit of eq (7) with 2 free parameters: $k_{off}$ and $R_2 = R_{2,A} = R_{2,HA}$ (keeping $\Delta\omega$ and $K_a$ fixed).

[b] $\chi^2(3)$ denotes $\chi^2$ resulting from a fit of eq (7) with 3 free parameters: $k_{off}$, $R_{2,A}$ and $R_{2,HA}$ (keeping $\Delta\omega$ and $K_a$ fixed).

[c] $\chi^2(4)$ denotes $\chi^2$ resulting from a fit of eq (7) with 4 free parameters: $k_{off}$, $R_{2,A}$, $R_{2,HA}$ and $K_a$ (keeping $\Delta\omega$ fixed).

[d] The number of data points (*N*) depends on whether one or two resolved $^1H^{\beta/\gamma}$–$^{13}CO$ cross-peaks are observed in the spectrum. The maximum number of points is 42, because 21 samples (at different pH) were analyzed.

[e] The *F*-ratio was calculated as follows.[2] For the comparision of the 2- and 3-parameters models, which are nested, *F* was calculated as:

$$F = [\chi^2(2) - \chi^2(3)]/\chi^2(3)/(N-3)$$

For the comparison of the 3- and 4-parameter models, which are not nested, *F* was calculated as:

$$F = \chi^2(3)/(N-3)/\chi^2(4)/(N-4)$$

[f] The *p*-value indicates the probability of falsely rejecting the null hypothesis that the 3 (4)-parameter model does not provide a significantly better fit than the 2 (3)-parameter model.

[g] Data for residue D37 covers pH > 5 (see the main text).

**References**

(1) Lindman, S.; Linse, S.; Mulder, F. A. A.; Andre, I. *Biochemistry* **2006**, *45*, 13993.
(2) Bevington, P. R.; Robinson, D. K. *Data reduction and error analysis for the physical sciences*; WCB/McGraw-Hill: New York, 1992.