ChemPubSoc Europe

DOI: 10.1002/cbic.201800628

CHEMBIOCHEM Full Papers

# Experimental p$K_a$ Value Determination of All Ionizable Groups of a Hyperstable Protein

Heiner N. Raum and Ulrich Weininger*[a]

Electrostatic interactions significantly contribute to the stability and function of proteins. The stabilizing or destabilizing effect of local charge is reflected in the perturbation of the p$K_a$ value of an ionizable group from the intrinsic p$K_a$ value. Herein, the charge network of a hyperstable dimeric protein (ribbon–helix–helix (rhh) protein from plasmid pRN1 from *Sulfolobus islandicus*) is studied through experimental determination of the p$K_a$ values of all ionizable groups. Transitions were monitored by multiple NMR signals per ionizable group between pH 0 and 12.5, prior to a global analysis, which accounted for the effects of neighboring residues. It is found that for several residues involved in salt bridges (four Asp and one Lys) the p$K_a$ values are shifted in favor of the charged state. Furthermore, the p$K_a$ values of residues C40 and Y47, both located in the hydrophobic dimer interface, are shifted beyond 13.7. The necessary energy for such a shift is about two-thirds of the total stability of the protein, which confirms the importance of the hydrophobic core to the overall stability of the rhh protein.

## Introduction

Protein stability results from several different forces, such as the hydrophobic effect[1] or polar interactions, such as hydrogen bonds[2] or aromatic interactions,[3] each of which stabilize the folded state. Contrastingly, the loss in conformational entropy destabilizes the folded state relative to the unfolded state. Charge–charge interactions can have both stabilizing and destabilizing effects. Typically, charge–charge interactions result in a net stabilization effect[4] and are responsible for the pH-dependent stability profile of proteins with unfolding at high or low pH. Beyond protein stabilization, they play a key role in protein structure, dynamics, and function.[5–7]

The degree of electrostatic interaction between charged (ionizable) groups is reflected in deviations of p$K_a$ values from their intrinsic (or unperturbed) values.[8] If the p$K_a$ value is shifted towards a lower (higher) p$K_a$ for negative charges or higher (lower) p$K_a$ for positive charges, then the presence of a charge is favorable (unfavorable). For a detailed quantification of the energetic contribution of a charge to protein stability, the determination of its p$K_a$ value in the unfolded state is necessary.[9] However, if p$K_a$ values do not fall close ($\pm$0.3 units) to their intrinsic values, the p$K_a$ values of the native state alone can be used in a semiquantitative way.

Seven amino acid side chains contain ionizable groups between pH 0 and 14. Four (Asp, Glu, Tyr, and Cys) are negatively charged above their p$K_a$ or uncharged below. Three (His, Lys, and Arg) are positively charged below their p$K_a$ or uncharged above. Additionally, the termini of a protein are charged at pH 7. The protonation behavior (and the presence and absence of charges) of ionizable groups can be monitored in a site-specific manner by means of NMR spectroscopy.[10–13] Chemical shifts of certain atoms in amino acids are strongly dependent on the degree of protonation and are consequently suitable probes for p$K_a$ value determination.[14] However, medium-range titrating events can also cause moderate changes in chemical shifts, which leads to additional transitions in titration curves and the need for a more involved analysis.[10] By using an NMR spectroscopy based approach, about 540 p$K_a$ values have been determined experimentally in folded proteins as of 2009.[15] Furthermore, for Asp and Glu residues, kinetic parameters were recently determined and linked to their p$K_a$ values.[16]

The small dimeric ribbon–helix–helix (rhh) protein[17] from plasmid pRN1 of *Sulfolobus islandicus* displays an extremely high thermodynamic stability over a wide pH range.[18] It consists of a $\beta$ strand formed by the two monomers followed by two $\alpha$ helices per monomer. The native dimeric state can be stabilized relative to the monomeric unfolded state by increasing the protein concentration. Furthermore, no partially folded states are populated,[18] which could lead to a misinterpretation of native-state p$K_a$ values. Thus, it is an ideal target for a complete experimental p$K_a$ value determination over a wide pH range. The thermodynamic stability of the rhh protein is unusually high (85 kJ mol$^{-1}$[18]), which must be a result of significantly more stabilizing than destabilizing interactions. For comparison, in the homologue Arc repressor, a R31M/E36Y/R40L triple mutant increased the stability from 40[19] to

[a] *H. N. Raum, Dr. U. Weininger*
*Institute of Physics, Biophysics*
*Martin-Luther-University Halle-Wittenberg*
*06120 Halle/Saale (Germany)*
*E-mail: ulrich.weininger@physik.uni-halle.de*

*Supporting information and the ORCID identification numbers for the authors of this article can be found under https://doi.org/10.1002/cbic.201800628.*

60 kJ mol$^{-1}$.[20] The rhh protein used here resembles the Arc repressor mutant at these positions. Due to the significantly higher stability of the rhh protein, it is extremely unlikely that it contains any interactions that significantly decrease its thermodynamic stability. Consequently, all significantly perturbed p$K_a$ values can be interpreted as positive or at least neutral for the overall stability in a first approximation. It contains 27 ionizable groups in a monomer: four Asp, four Glu, one Cys, three Tyr, two His, eight Lys, three Arg, and the N and C termini. A structure-based analysis suggested a high hydrophobicity of the dimer interface as one of the main reasons for stability.[17]

In this study, we determine the full set of experimental macroscopic p$K_a$ values (p$K_a$ values that describe the collective behavior of a coupled system, in contrast to microscopic p$K_a$ values, which describe the behavior of one ionizable group) of a protein. We analyze the reflection of all (short and medium distance) p$K_a$ values in the NMR spectroscopy titration curves by using multiple NMR signals per ionizable group in a global manner. Using this approach, we were able to unambiguously decompose all broadened transitions to residues with neighboring ionizable groups. Furthermore, we address the impact of electrostatic contributions to the extreme thermodynamic stability of the rhh protein evaluated by large perturbations of experimentally determined p$K_a$ values compared with the intrinsic p$K_a$ values. Four possible positive Coulomb interactions ("salt bridges"), present in the crystal structure (PDB ID: 3FT7), could be confirmed in solution because five of the residues involved (four Asp, one Lys) displayed shifted p$K_a$ values in favor of the charged state. However, the contribution of the salt bridges to overall stability is minor because the protein remains folded in their absence (pH 0). In contrast, p$K_a$ values of C40 and Y47, located in the hydrophobic dimer interface, are shifted to and beyond 13.7. Around pH 12.5 (at which a low percentage of C40 and Y47 become titrated) the protein unfolds; this confirms the importance of the hydrophobicity of the protein core in the overall thermodynamic stability of the rhh protein.

## Results

NMR spectroscopy signals of all 27 ionizable groups of the rhh protein could be monitored by chemical shift changes over a pH range from 0 to 12.5. Around pH 12.5, there is a continuous unfolding transition of the protein, that is, at pH 12.9 there are no signals remaining from the folded protein. Additionally, at that high pH, a much slower degradation process could be observed (Figure S1 in the Supporting Information). For all ionizable groups (between pH 0 and 14), multiple well-resolved pH titration curves could be acquired and analyzed (Figures S2–S19). In many cases, other effects than the main transition (transition with the highest amplitude, strong impact on intraresidual probes) influenced the chemical shifts as well.

In principle, the chemical shift can be affected by three processes (other than the intraresidual ionizable group) in a pH-dependent way: 1) pH-induced global structural changes. This can be safely neglected for the studied protein. No protein folding intermediates or pH-dependent changes in secondary structure were found in a previous study.[18] Furthermore, such global structural changes usually happen in the slow NMR exchange regime. No such processes other than global unfolding at high pH were observed in the NMR spectra. 2) Impact of remote ionizable groups (up to 10 Å) on the chemical shift. These were treated as additional p$K_a$ values in the fit. We cross-checked that indeed an ionizable group with matching p$K_a$ value was located in structural proximity, and that resulting chemical shift changes by these remote events were low, especially for chemical shifts that predominantly depended on the charged state of the intraresidual ionizable group.[14] 3) The impact of remote ionizable groups (up to 10 Å) on the charge of the intraresidual ionizable group. Ionizable groups with similar p$K_a$ values can influence each other and cause a broadened pH transition between the uncharged and charged states. Points 2) and 3) cannot be distinguished easily: the impact of 2) on the chemical shift is low, whereas 3) can only occur for two charges with similar p$K_a$ values (up to 1.3 units difference for the lower or upper microscopic p$K_a$ values[21]). Additionally, point 2) should affect multiple chemical shifts that report on the studied transition differently (because the influence of the nearby charge on chemical shifts is different), whereas 3) should affect all chemical shifts that report on the studied transition in the same way. This should result in the same kind of broadening caused by Coulomb interactions. In this case (3), if the chemical shift is not dependent on the ionization state of the other charge, the titration curves of the coupled system show mirror images (ideally mirrored vertically on the middle of both macroscopic p$K_a$ values). Furthermore, it must be taken into consideration that chemical shifts near both charges can be a combination of both transitions, resulting in a monophasic or non-monotonic transition (Figure S20).

### Global p$K_a$ value determination by multiple chemical shifts

To address the complications mentioned above, multiple chemical shifts were used per residue across the pH titration and a global macroscopic model was used to allow multiple p$K_a$ values (but no cooperativity factor, Hill coefficient) per pH titration curve.[21,22] The macroscopic p$K_a$ values generally reflect the total protonation/deprotonation state of a coupled system and in our protein all monitored transitions could be explained. We found that five main transitions were outside the studied pH range (R3, R24, C40, R41, and Y47). Here, only lower borders of the p$K_a$ values could be established. From the remaining 22 ionizable residues that could be quantitatively studied across the pH range, 6 main transitions can be described by a single p$K_a$ value alone (N terminus, H22, K32, D38, H51, and C terminus). For one residue, a second p$K_a$ value with a low chemical shift amplitude was necessary (K30). We interpret this example as a case in which a remote ionizable group (Y5) has an impact on the chemical shift (case 2) because some probes display monophasic behavior.

The pH transitions of the remaining 15 ionizable groups (Y5/K6, K12/Y16, E20/D23/E27, E31/D52, E46/D49, and K45/K53/K55/K66) are more complex. They can be described by two macroscopic p$K_a$ values, both with significant chemical shift

amplitudes, whereby the major p$K_a$ value (macroscopic p$K_a$ value that belongs to the transition with the highest amplitude) is a close reflection of the charged state of the ionizable group. Alternatively, they can be described with one p$K_a$ value and a significantly reduced Hill coefficient (Table S1 and Figure S21). However, in the case of Y5/K6 and K12/Y16, the approach with two p$K_a$ values is significantly better because this is the only approach that enables all of the monitored transition curves to be explained (Figure 1). This approach produces two different "major" p$K_a$ values for Tyr and Lys, which are in the same range as that of the unperturbed p$K_a$ values for Tyr and Lys. Both Y5/K6 and K12/Y16 are solvent exposed with Y5/K6 located in the unstructured N terminus. A small shift from the unperturbed p$K_a$ value to lower (Tyr) and higher (Lys) values may be caused by mutual stabilization of the opposite charges. A possible coupling between K12 in each monomer predicted by the structure was not observed experimentally. For the remaining 11 ionizable groups, both approaches give similar results.

For residues E20/D23/E27, E31/D52, and E46/D49, we likely observe an impact of the remote charge on the intraresidual charge itself. E20, D23, and E27 are located on one side of the first helix (Figure 2 A). D23 is in close proximity to both E20 and E27. All main transitions for all three residues are broadened. Describing them with two macroscopic p$K_a$ values results in a significant amplitude for chemical shift changes of the second p$K_a$ value. This has to be interpreted not only as an effect on the chemical shift, but also on the charge. Describing them with one p$K_a$ value and a Hill coefficient results in a low Hill coefficient, which can be interpreted as a significant influence of one charge by the other. The major p$K_a$ value of D23 is

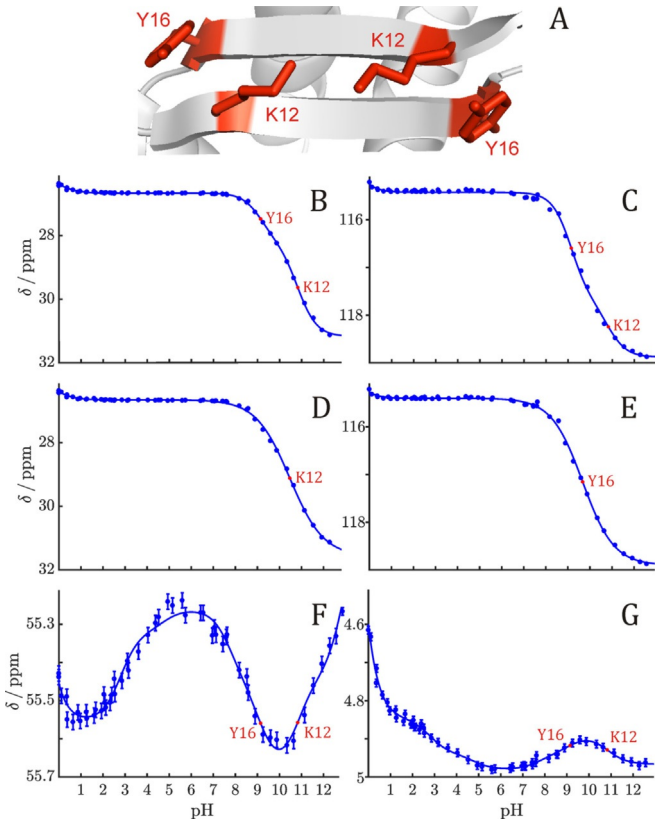


**Figure 1.** Deconvolution of multiple p$K_a$ values by using multiple probes for the example of K12 and Y16. A) A section of the reported structure (PDB ID: 3FT7) with K12 and Y16 shown as sticks in red and labeled. K12 $^{13}$Cδ (B, D), Y16 $^{13}$Cε (C, E), $^{13}$Cα (F), and $^{1}$Hα (G) shifts are plotted versus pH. Fits are shown as solid lines and p$K_a$ values are shown as red dots and labeled. B), C), F), and G) are fitted in a global way with multiple probes (Figure S5) to two p$K_a$ values for Y16 (9.14 ± 0.09) and K12 (10.83 ± 0.10). D) and E) are fitted to one p$K_a$ value (9.67 ± 0.05) and Hill coefficient (0.62 ± 0.03) for Y16 and for K12 (10.47 ± 0.08/0.57 ± 0.03).

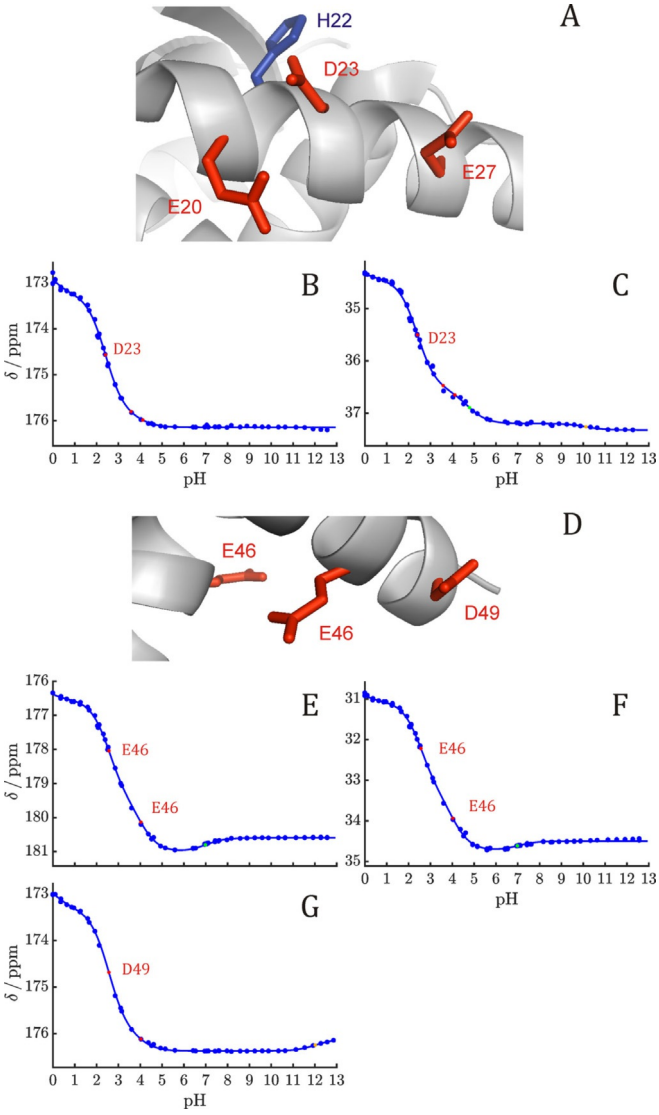


**Figure 2.** Coupling systems of the acidic residues. A) D23 is located between E20 and E27 (red) and next to H22 (blue) on the first helix; they are labeled and shown as sticks. For D23, the $^{13}$Cγ chemical shift (B) is weakly influenced by the charges of E20 and E27, but not by H22. Its impact on the chemical shift can be seen in D23 $^{13}$Cβ (C). E46 and D49 are located at a close distance on the second helix, whereby E46 is also close to E46 of the other monomer chain (D); they are labeled and shown as sticks. For E46, $^{13}$Cδ and $^{13}$Cγ chemical shifts (E, F) are broadened as a result of "self-coupling", whereas, for D49, E46 has a weak impact on the $^{13}$Cγ chemical shift (G). Red points represent the p$K_a$ values used as global parameters by global analysis of multiple probes, green points represent known p$K_a$ values of neighboring ionizable groups, and orange points represent unknown p$K_a$ values that ensure a suitable baseline.

significantly reduced compared with the intrinsic value (Figure 2B, C). This can be explained structurally by a Coulomb interaction with H22. Furthermore, some transitions clearly look not only biphasic, but also like mirror images, which cannot be described by one p$K_a$ value and Hill coefficient.

For E31/D52, we likely observe an impact of the remote charge on the inter-residual charge itself. E31 is located on the first helix facing the second helix, where D52 is located. Both main transitions are broadened and show mirror images. An analysis with two macroscopic p$K_a$ values results in a significant amplitude for the second p$K_a$ value. Also, an individual fit with the Hill equation produces a similar Hill coefficient of about 0.6. The shift of the major p$K_a$ value of D52 to a lower value, relative to the unperturbed value, can be explained by a Coulomb interaction with R24.

E46/D49 are located at a close distance on the second helix (Figure 2D). E46 is also close to E46 of the other monomer chain. Although the main transition of D49 shows only a small additional transition (see case 2), E46 appears to be broader; this indicates a charge effect across the dimer interface (Figure 2E–G). Because both titration curves show the opposite of mirror images, coupling between E46 and D49 is improbable. Even though both curves can be described by the same two macroscopic p$K_a$ values (2.56 and 4.03), the lower microscopic p$K_a$ value of E46 and the major p$K_a$ value of D49 (2.56) are not separable by fitting procedures. For example, inspection of the main transition of E46 (both transitions have the same chemical shift difference) gives 2.39 and 3.60 as microscopic p$K_a$ values, which correspond to an electrostatic coupling of about 1.6 kcal mol$^{-1}$. Because E46 and D49 are close to each other in space and chemical shifts are influenced by both transitions, a more detailed analysis was not possible. Through the use of the approach with one p$K_a$ value and Hill coefficient, a p$K_a$ value of 2.62 is obtained for the D49 transition and 2.97 for the E46 transition. The latter is equivalent to the mean value of the microscopic p$K_a$ values (3.00). In both approaches, the p$K_a$ value of D49 is significantly reduced compared with its unperturbed value. This can be explained structurally by a positive Coulomb interaction of D49 with K45.

The remaining system (K45/K53/K55/K56) also displays a weak impact of the remote charge on the intraresidual charge itself. K45 is located on the second helix, pointing towards the solvent and the unstructured C terminus, where K53, K55, and K56 are located. All three Lys residues show a mirror image to K53. An approach with two macroscopic p$K_a$ values in each case (K45/K53, K53/K55, and K53/K56) describes the titration curves very well, although coupling between K45, K55, and K56 was neglected because of their comparable titration curve shapes. However, some uncertainties remain. Firstly, the distance between charges is not known due to the unstructured nature of this region and the proximity to the solvent weakens Coulomb interactions. Secondly, both the baseline and titration curve shape in the basic pH range is poorly defined. In this case, the p$K_a$ value estimation with the Hill equation is a good alternative. The differences between the major macroscopic p$K_a$ value and the apparent p$K_a$ value from the Hill equation are quite low (from 0.02 to 0.17 for K45 and K53, respectively).

In summary, both approaches (macroscopic model and Hill coefficient) result in very similar p$K_a$ values that fall within 0.4 units (0.25 on average), except for E31 and Y16, for which the differences are 0.68 and 0.53 units. In the following sections, we only focus on p$K_a$ values derived from the macroscopic model, and consequently, we refer to a major macroscopic p$K_a$ value of an ionizable group simply as p$K_a$ value.

### pH titration curves below pH 1

Nearly all transition curves show an additional transitions below pH 1. 4,4-Dimethyl-4-silapentane-1-sulfonic acid (DSS) referencing, which was not necessary for all other pH regions, was unable to correct this effect completely because of possible pH deviation from the protein samples. We attribute this effect to a drastic change of the solvent. Water becomes significantly more protonated, and thus, many counterions (salt) are introduced. This was treated as an additional p$K_a$ value in the fit, but ultimately did not interfere with our results in any way because none of the p$K_a$ values of the protein were shifted below two.

### Experimental p$K_a$ values of acidic residues

There are nine acidic residues per monomer in the rhh protein. Four Asp residues (D23, D38, D49, and D52), four Glu residues (E20, E27, E31, and E46) and the C terminus. The p$K_a$ values of all four Asp are significantly (one unit or more) shifted to lower values (Figures 3A–D, S6, 10, 12, and 16, and Table 1) compared with the intrinsic p$K_a$ value;[14] thus revealing a stabilization of the negatively charged state. This is in agreement with structural findings that they are within 4 Å of a positive charge (H22, R41, K45, and R24), indicating distinct positive Coulomb interactions, at least under low-salt conditions. In the case of D49, the titration event of the Coulomb partner (K45) can also be seen in the titration curve (Figure 3A). In contrast, the titration events of the Coulomb partners of D38 and D52 (R41 and R24, respectively) are not covered in the range of the titration curves. The titration event of the Coulomb partner of D23 (H22) is not seen in the titration curve. In contrast to the four Asp, all Glu except E46 (Figures 3E–H, S6, S10, and S16, and Table 1) and the C terminus (Figures 4A and S19, and Table 1) are within 0.8 units of their intrinsic p$K_a$ values;[14] thus showing almost ideal behavior of surface-exposed residues. Three Glu show negative Coulomb interactions with Asp (E20 and E27 to D23, E31 to D52), but have also positive charges in their close environment. R24 is about 5 Å away from E20, K32 and H51 about 7 Å away from E31, and K30 and R24 less than 10 Å away from E27. These positive Coulomb interactions enable them to energetically couple with Asp, which, in turn, causes these major p$K_a$ values near the intrinsic ones. Similarly, E46 has two positive charges (6 Å to H51 and about 9 Å to K32) in its environment, but, in this case, it couples with the E46 residue in the other rhh monomer across the dimerization interface.

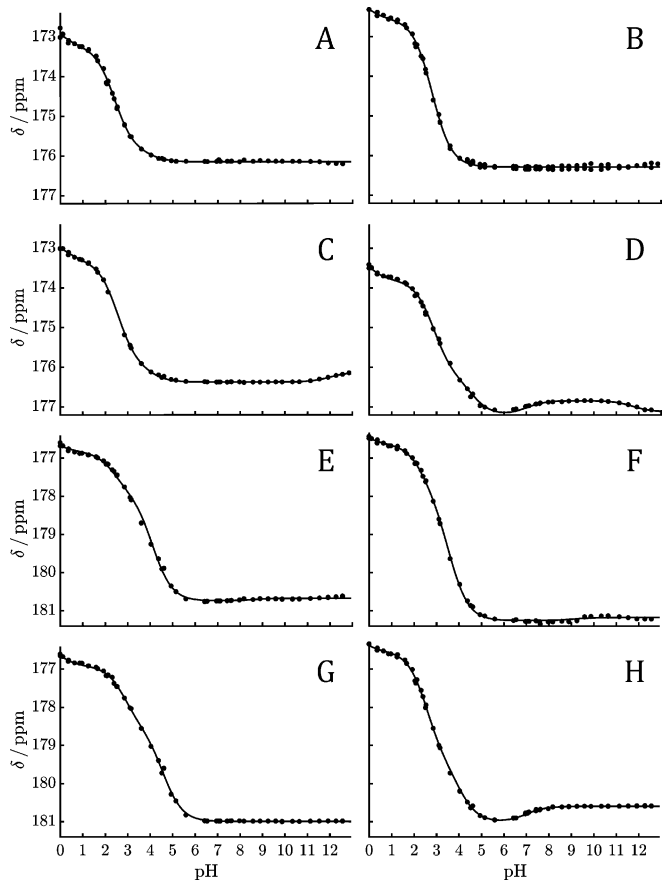


**Figure 3.** Experimental pH titration curves of acidic residues. For D23 (A), D38 (B), D49 (C), and D52 (D), $^{13}$COγ chemical shifts are plotted versus pH, whereas for E20 (E), E27 (F), E31 (G), and E46 (H) $^{13}$COδ chemical shifts are used. Fits estimated by global analysis of multiple probes (see Figures S6, S10, S12, and S16, and Table S2) are shown as solid lines and results are summarized in Table 1. All measurements were conducted at 25 °C on 1 mM protein in water (low salt, self-buffering, <60 mM).

**Table 1.** Macroscopic p$K_a$ values determined and the corresponding difference in chemical shift (Δδ) of the main transition.[a] The major p$K_a$ values of coupled residues are underlined.

| Residue | p$K_a$ | Δδ (ppm) | Residue | p$K_a$ | Δδ (ppm) |
|---|---|---|---|---|---|
| D23 | <u>2.41 ± 0.04</u> | 2.69 ± 0.13 | K6 | <u>10.78 ± 0.13</u> | 4.18 ± 0.39 |
| | 3.58 ± 0.05 | 0.05 ± 0.24 | | 9.66 ± 0.11 | 0.78 ± 0.45 |
| | 4.12 ± 0.04 | 0.20 ± 0.13 | K12 | <u>10.83 ± 0.10</u> | 3.01 ± 0.16 |
| D38 | 2.80 ± 0.03 | 3.73 ± 0.03 | | 9.14 ± 0.08 | 1.49 ± 0.18 |
| D49 | <u>2.56 ± 0.04</u> | 2.73 ± 0.10 | K30 | 11.18 ± 0.09 | 4.69 ± 0.15 |
| | 4.03 ± 0.08 | 0.33 ± 0.09 | K32 | 10.43 ± 0.06 | 2.75 ± 0.15 |
| D52 | <u>2.88 ± 0.03</u> | 2.45 ± 0.06 | K45 | <u>11.84 ± 0.09</u> | 6.04 ± 0.38 |
| | 4.59 ± 0.04 | 0.95 ± 0.07 | | 10.38 ± 0.11 | 0.94 ± 0.29 |
| E20 | <u>4.12 ± 0.04</u> | 2.84 ± 0.07 | K53 | <u>10.38 ± 0.11</u> | 4.65 ± 0.64 |
| | 2.41 ± 0.04 | 1.08 ± 0.08 | | 11.63 ± 0.11[b] | 1.06 ± 0.35 |
| E27 | <u>3.58 ± 0.05</u> | 3.32 ± 0.16 | K55 | <u>11.77 ± 0.10</u> | 4.35 ± 0.30 |
| | 2.41 ± 0.04 | 1.31 ± 0.18 | | 10.38 ± 0.11 | 1.12 ± 0.29 |
| E31 | <u>4.59 ± 0.04</u> | 2.44 ± 0.08 | K56 | <u>11.44 ± 0.10</u> | 4.57 ± 0.46 |
| | 2.88 ± 0.03 | 1.63 ± 0.08 | | 10.38 ± 0.11 | 0.70 ± 0.57 |
| E46 | 2.56 ± 0.04 | 2.93 ± 0.16 | R3[a] | >13.7 | |
| | 4.03 ± 0.08 | 1.56 ± 0.16 | R24[a] | >12.5 | |
| C-term | 3.23 ± 0.03 | 1.80 ± 0.10 | R41[a] | >13.3 | |
| Y5 | <u>9.66 ± 0.11</u> | 2.52 ± 0.30 | H22 | 4.79 ± 0.04 | 0.61 ± 0.02 |
| | 10.78 ± 0.13 | 0.79 ± 0.31 | H51 | 6.94 ± 0.03 | 1.88 ± 0.01 |
| Y16 | <u>9.14 ± 0.08</u> | 2.27 ± 0.15 | N-term | 7.76 ± 0.04 | 3.13 ± 0.01 |
| | 10.83 ± 0.10 | 1.18 ± 0.15 | | | |
| Y47[b] | >13.7 | | | | |
| C40[c] | >14.9 | | | | |

[a] Reported errors are from Monte-Carlo simulations. Intrinsic p$K_a$ values taken from the literature are 3.86 (D), 4.34 (E), 3.55 (C-terminus), 10.34 (K), 8.23 (N-terminus), 13.9 (R), 6.45 (H), 9.76 (Y), and 8.49 (C).[14] A comparison between the macroscopic model and fit procedures with the Hill equation can be seen in Figure S21. Most important positions for the p$K_a$ value determination are listed in Table S2. All included titration curves are given in Figures S2–S19. [b] Estimations have been made by using 50% of the reported chemical shift differences.[14] [c] p$K_a$ value is a mean of the major p$K_a$ values of K45/K53/K56.

### Experimental p$K_a$ values of basic residues

In contrast, there are 14 basic residues per monomer in the rhh protein. Two His (H22 and H51), eight Lys (K6, K12, K30, K32, K45, K53, K55, and K56), three Arg (R3, R24, and R41), and the N terminus. The N terminus (Figures 4B and S2, and Table 1), H51 (Figures 4D, F and S17, and Table 1), and five Lys (K6, K12, K30, K32, and K53; Figures 5A–D and F, S4, S5, S9, S11, and 18, and Table 1) are within 0.5 (except K30 within 0.8) units to the intrinsic p$K_a$ values,[14] and even closer to experimental p$K_a$ values in apo-calmodulin.[23] All of these groups are solvent exposed. H51 is not solvent exposed, but also does not display any close contacts to other charges. The p$K_a$ value of all three Arg residues (R3, R24, and R41; Figures 5I–K, S3, S8, and S14, and Table 1) are not lower than the intrinsic p$K_a$ value,[14,24] although their p$K_a$ values are outside the studied pH region, and expected shift changes are low. In case of R41, it is unclear whether the start of the transition is observed or not, due to the proximity to Y47. Despite all of the uncertainties, they describe the lowest possible p$K_a$ value accurately. Therefore, p$K_a$ must to be close to their intrinsic values or shifted higher, which together indicate favoring of the charged state.

The p$K_a$ value of H22 (Figures 4C, E, and S7, and Table 1) is significantly (1.7 units) shifted to lower values, which indicates that the positive charge is unfavorable. Additionally, it is found that more than 50% are in the uncommon HNδ1 tautomer in its neutral form, which can be evaluated by an increasing $^{13}$Cδ2 shift with higher pH, at which the neutral form becomes populated. High $^{13}$Cδ2 shifts are very indicative of the HNδ1 tautomer, whereas the shifts are much lower and equal among the HNε2 tautomer and the protonated state.[25–28] Long-range $^{1}$H,$^{15}$N spectra of the His side chain confirm this finding (Figure S7H, I). Because H22 is not exclusively in one tautomeric neutral form, and no potential hydrogen bonds stand out in the structure, it is likely that the shift in the p$K_a$ value is caused by the unfavored effect of the positively charged state and not by stabilizing interactions of one of the neutral tautomeric states. The p$K_a$ values of K45, K55, and K56 (Figures 5E, G, and H, S18, and Table 1) are shifted more than one unit to higher values, which indicates a positive contribution to the protein stability of the positive charge. K45 forms a Coulomb interaction with D49 (within 4 Å distance). K55 and K56 are located in the unstructured C-terminal region, so a difference from the

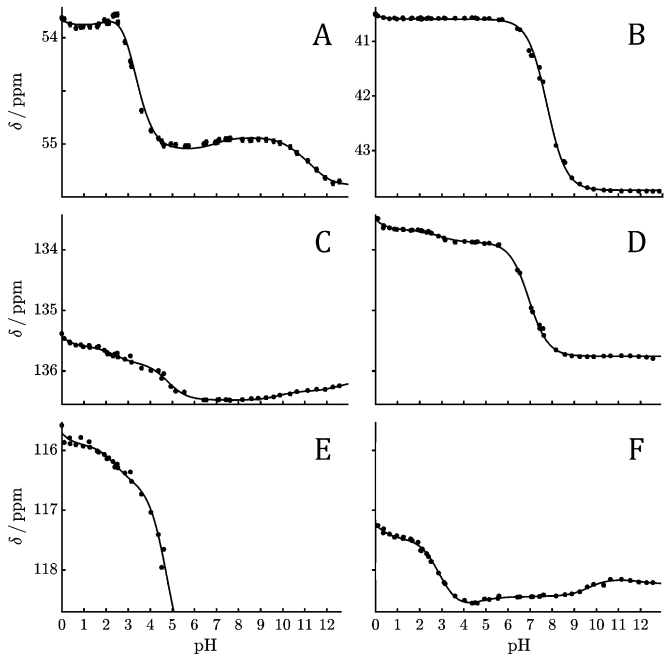


**Figure 4.** Experimental pH titration curves of the protein termini and His. For the C terminus (A) and N terminus (B) $^{13}C\alpha$ chemical shifts are plotted versus pH, while for H22 (C) and H51 (D) $^{13}C\varepsilon1$ chemical shifts are used. The difference in the amplitude (C, D) and the different direction of shifts (E, F) of the histidines arises from the different tautomeric forms. Fits estimated by global analysis of multiple probes (Figures S2, S7, S17, and S19, Table S2) are shown as solid lines and results are summarized in Table 1. All measurements were conducted at 25 °C on 1 mM protein in water (low-salt, self-buffering, < 60 mM).

intrinsic p$K_a$ value is unexpected. However, this could be explained by charge compensation with the negative charged C terminus (position 56).

### Experimental p$K_a$ values of Tyr and Cys

There are three Tyr (Y5, Y16, and Y47) and one Cys (C40) residues per monomer in the rhh protein. They are neutral at acidic and neutral pH, but can be weak acids at high pH. Y5 and Y16 (Figures 6A, B, S4, and S5, and Table 1) are solvent exposed and are within 0.6 units of their intrinsic p$K_a$ values[14] and close to usually reported values in proteins.[29] In contrast, Y47 and C40 (Figures 6C, D, S15, and S13, and Table 1) are located in the hydrophobic domain interface and show dramatically shifted p$K_a$ values to higher values, which indicates a negative impact of the negative charges. For Y47, the transition starts at pH 12, and thus, is not completed at pH 12.5, whereas, for C40, no signs of a transition could be observed up to pH 12.5. The p$K_a$ values could be estimated to be larger than 13.7, assuming unaffected cooperativity and −50% divergence of the usual amplitude.[14] The drastic shift of these p$K_a$ values located in the hydrophobic core thus reflects the strength of the hydrophobicity of the domain interface, which is responsible for such a shift.

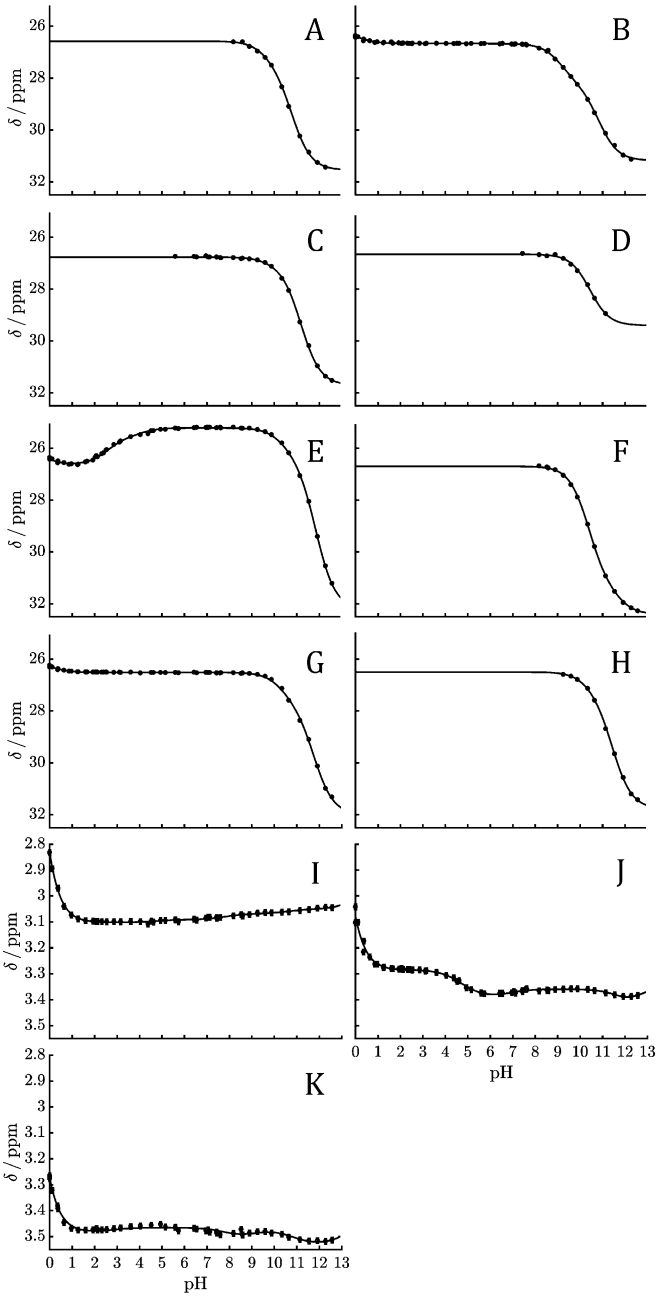


**Figure 5.** Experimental pH titration curves of basic residues. For K6 (A), K12 (B), K30 (C), K32 (D), K45 (E), K53 (F), K55 (G), and K56 (H), $^{13}C\delta$ chemical shifts are plotted versus pH, whereas, for R3 (I), R24 (J), and R41 (K), $^{1}H\delta$ chemical shifts are used. Fits estimated by global analysis of multiple probes (Figures S4, S5, S9, S11, and S18, Table S2) are shown as solid lines and results are summarized in Table 1. All measurements were conducted at 25 °C on 1 mM protein in water (low-salt, self-buffering, < 60 mM).

## Discussion

### Deviations of experimental p$K_a$ values from intrinsic values

Sixteen of 27 p$K_a$ values are within one unit to the intrinsic values. In these cases, the charged state is neither especially favored or unfavored. In eight cases, the charged state is favored, namely, in D23, D38, D49, D52, and K45, all involved in four distinct positive Coulomb interactions, in E46, and in K55

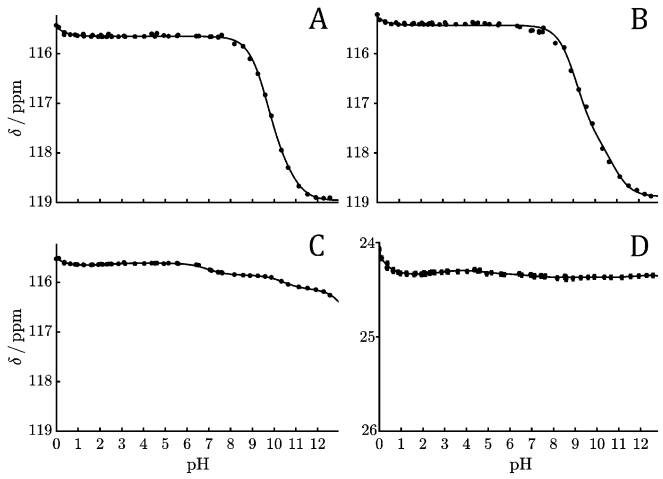


**Figure 6.** Experimental pH titration curves of Tyr and Cys. For Y5 (A), Y16 (B), and Y47 (C), $^{13}C\varepsilon$ chemical shifts are plotted versus pH, whereas for C40 (D) $^{13}C\beta$ is used. Fits estimated by global analysis of multiple probes (Figures S4, S5, S13, and S15, and Table S2) are shown as solid lines and the results are summarized in Table 1. All measurements were conducted at 25 °C on 1 mM protein in water (low salt, self-buffering, <60 mM).

and K56 in close proximity to the negatively charged C terminus. In three cases, the charged state is unfavored, namely, in H22, C40, and Y47. H22 is involved in hydrophobic contacts; C40 and Y47 are located in the hydrophobic domain interface.

We calculated p$K_a$ values from the crystal structure (PDB ID: 3FT7) using H++,[30,31] ROSIE,[32,33] and DEPTH,[34] to see how much these shifts in p$K_a$ values could be reproduced from calculations (Table S3). H++ was able to capture the shift in p$K_a$ values of C40 and Y47 to higher values (>12). ROSIE failed on Y47 (9.4; and does not offer a calculation on C40), whereas DEPTH did not offer a calculation for C40 and Y47. Both H++ and ROSIE capture the shift in p$K_a$ values of D23, D38, and D49 to lower values, but do not provide a close match of p$K_a$ values (<0.5 units), nor the correct ranking. DEPTH only qualitatively reproduces D38, but fails on D23 and D49. ROSIE captures the shift in p$K_a$ value of K45 to a higher value (and the trends in other Lys), H++ predicts the p$K_a$ value of K45 accurately (but failed on K32). DEPTH did not reproduce p$K_a$ values in Lys. All programs failed to capture the shift in p$K_a$ value of H22 to a lower value. D52, K55, and K56 are not (well resolved) in the crystal structure. E46 was a complicated case with a charge effect across the dimer interface and was therefore excluded from the calculations. In summary, H++ was able to reproduce most shifts in p$K_a$ values qualitatively, but failed on H22.

### Salt bridges

All four Asp (D23, D38, D49, and D52) are involved in distinct positive Coulomb interactions, which could be interpreted as salt bridges. They have a positive charge (H22, R41, K45, and R24) within 4 Å (Figures 7 and S22) and their p$K_a$ values are shifted to lower values (Figure 3 A–D and Table 1); thus experimentally demonstrating a stabilization of the negative charge and a direct contribution to stability under low-salt conditions

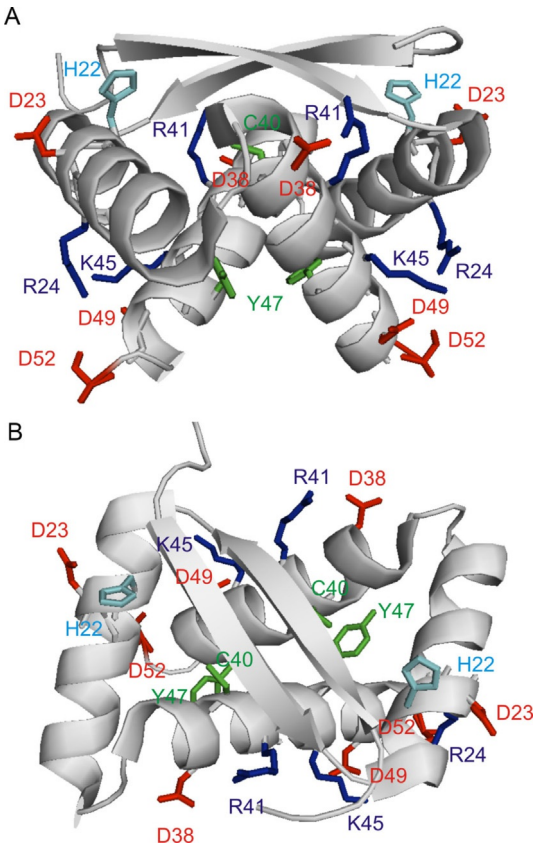


**Figure 7.** Ribbon representation of the reported structure (PDB ID: 3FT7) in two different orientations (A, B). Negatively charged residues involved in positive Coulomb interactions (D23, D38, D49, and D52) are shown as sticks, labeled and colored in red; positively charged residues are shown as sticks, labeled and colored in blue (R24, R41, and K45) and cyan (H22). Neutral residues with p$K_a$ values ≥14 (C40 and Y47) are shown as sticks, labeled and colored in green.

and low pH. For two salt bridges (D23–H22 and D49–K45), the distance is significantly shorter (about 3.5 Å); this indicates potentially stronger Coulomb interactions. This effect is also seen from further shifted p$K_a$ values of the two Asp involved (2.41 and 2.56, compared with 2.80 and 2.88, respectively); thus confirming two stronger and two weaker Coulomb interactions in solution. Three of the Coulomb pairs (D23-H22, D38-R41, and D49–K45) are intra-monomer; the fourth (D52–R24) is inter-monomer. On the side of the positively charged state, the results are not as uniform. The increased p$K_a$ value (Figure 5 E and Table 1) of K45 (salt bridge to D49) reports a stabilization of the positive charge and indicates a positive Coulomb interaction. Thus, this Coulomb interaction (D49–K45) contributes to the protein stability over a large pH range. For R24 and R41, no change in p$K_a$ values could be observed, although neither transition was well captured by experiments (Figure 5 J, K and Table 1). It is not clear if the unchanged p$K_a$ values are a result of experimental uncertainties caused by their high p$K_a$ value (out of the pH window of protein stability) or if the Coulomb interactions (D38–R41, D52–R24) simply contribute more to protein stability at low pH (at which a shift of the p$K_a$ value is observed) than that at high pH (at which there might not be a shift of the p$K_a$ value). In the fourth Coulomb interaction (D23–

H22), the p$K_a$ value of H22 (Figure 4C, E and Table 1) is significantly shifted to a lower value; thus reporting a preference of the uncharged state over the positively charged state, which is participating in the Coulomb interaction. The Coulomb interaction therefore is only favorable and possibly contributes to protein stability at low pH, although it is one of the two potentially stronger interactions (in terms of shorter distance and further shifted p$K_a$ value of D23). In the process of losing the four positive Coulomb interactions by pH titration at around pH 2.5, the protein loses significant thermodynamic stability (Figure 8), which is reflected in a decrease of melting temperature from 97 °C at pH 4 to 61.7 °C at pH 2.[18] Because the protein is a two-state folder and no changes in the cooperativity of the unfolding transitions are observed, transition midpoints are proportional to the free energy of unfolding.[18] Despite being destabilized, the protein remains folded, even at pH 0.

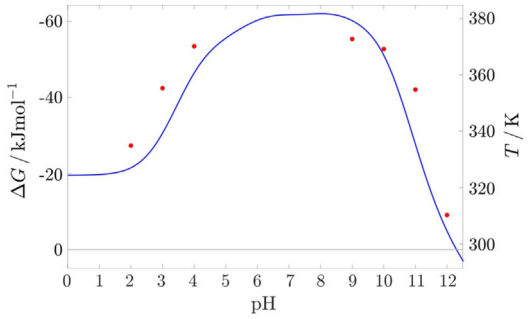


**Figure 8.** Dependence of the thermodynamic stability, $\Delta G$, and melting temperature, $t_m$, on the pH value. $\Delta G$ (blue line) was calculated by using the major macroscopic p$K_a$ values and the reported intrinsic p$K_a$ values as an approximation for the unfolded state.[12,13] It should be noted that the melting temperature (red points) was determined with 5 μM ORF56 and 20 mM sodium phosphate and 20 mM sodium citrate as buffer.[17] The gray line marks a temperature of 298 K and $\Delta G = 0$ kJ mol$^{-1}$.

### p$K_a$ values of buried Tyr and Cys are shifted to extremely high values

Cysteines and tyrosines are usually not charged at regular pH and prefer the uncharged state. In C40 and Y47, both located in the hydrophobic core/domain interface, the p$K_a$ values are shifted more than 6 and 3.5 units to higher values, which implies an additional disfavor of the charged state. The energy necessary for this must be provided from the hydrophobic domain interface.

This confirms the importance of the hydrophobicity of the domain interface for protein stability. This was concluded earlier on a structural comparison of different domain interfaces of rhh proteins with different stability.[17] Here, the experimentally found shifted p$K_a$ values of C40 and Y47 correspond to an energy, which has to be provided by the domain interface, of more than 57 kJ mol$^{-1}$ ($RT(\ln 10^{3.9} + \ln 10^{6.1})$).[12] This is a significant portion of 85 kJ mol$^{-1}$[18] total stability per monomer. Also, the rhh protein is stable down to pH 0, but unfolds over pH 12, at which the deprotonation of Y47 also occurs (Figure 8). All other charge transitions are at least 90% completed at this pH, with the exception of the 3 Arg. However, the p$K_a$ values are close to the intrinsic p$K_a$ value of Arg[14] or shifted to higher values. This leaves Y47 (and C40) with p$K_a$ values of around 14 as potential reasons for the loss of stability. They are also drastically shifted compared with their intrinsic values.[14] In fact, with p$K_a$ values of $>14$ for C40 and $>13.7$ for Y47, these values are significantly higher than those of the highest reported values for Cys (11.1) and Tyr (12.1).[35] Taken together, the optimized hydrophobic domain interface/hydrophobic core is the main reason for hyper-thermostability of the rhh protein from plasmid pRN1 of *S. islandicus*.[18]

## Conclusion

We have determined (or estimated a lower value) all p$K_a$ values in a protein following transitions on multiple probes over a wide range of pH. Combining the results of multiple chemical shifts in a global analysis, we were able to decompose complex and broad transition curves. In four Asp (D23, D38, D49 and D52) and one Lys (K45), all involved in the four positive Coulomb interactions (salt bridges), we found significant shifts in the p$K_a$ values in favor of the charged state. Additionally, a stabilized charged state was detected in K55 and K56, which were in close proximity to the C terminus. In contrast, H22, C40, and Y47 disfavored the charged state. In particular, in C40 and Y47, both located in the hydrophobic core, the p$K_a$ values were shifted to 14 or beyond.

## Experimental Section

**Protein samples**: The rhh protein (UniProtKB Q54323) was expressed in M9 minimal media, by using 1.5 g L$^{-1}$ $^{15}$N $NH_4Cl$ and 2 g L$^{-1}$ $^{13}C_6$ glucose, and purified as described previously,[36] with the modification of cell lysis directly by the heat step. It was concentrated to 1 mM. The pH was adjusted with NaOH or HCl directly on each sample in water (10% $D_2O$) and checked right before and after measurement in the NMR tube. pH values were measured at 22 °C by using an inoLab pH 720 pH meter with a Hamilton Spintrode pH electrode. Between pH 2 and 12, the total salt concentration (counterions of protein, $H^+$, and $OH^-$, $NaN_3$) was below 60 mM.

**NMR spectroscopy**: All experiments were performed on a Bruker DRX 500 NMR spectrometer at a static magnetic field strength of 11.7 T at 23 °C. $^1H^{13}C$ constant-time HSQC experiments optimized for aliphatic and aromatic regions and a $Hx_2CO$ experiment for Asp and Glu[37,38] were recorded at various pH values. Water suppression was achieved through gradient selection for the HSQC experiments and WATERGATE for the $Hx_2CO$ experiment. Assignments were taken from ref. [17] and confirmed through HNCACB, H(CCO)NH, (H)C(CO)NH and aromatic (HB)CB(COCD)HD, and TOCSY experiments. All spectra were processed with NMRPipe[39] and analyzed with NMRView.[40] The spectra were not referenced with DSS. External DSS referencing revealed a $^1H$ chemical shift of $\delta = -0.104$ ppm.

**Data analysis**: Only measurement points with a difference of pH value before and after measurement lower than 0.1 units, which did not show long-term variations in the NMR spectra (e.g., degradation), were included in the analysis. pH titration curves monitored by multiple chemical shifts (Figures S2–S19 and Table S2) were fitted with nonlinear least-squares regression analysis (Levenberg–Marquardt algorithm) to standard Hendersson–Hasselbalch equations for a variable number of p$K_a$ values in a global way by using MATLAB.. Convergence to the global minimum was ensured

through visual inspection of the fitted curves and that the estimated results agreed with the protein structure (coupled ionizable groups shared common p$K_a$ values and all titration curves were explainable through ionizable groups at a short distance). Altogether, more than 140 different shifts were analyzed and, excluding the extrapolated p$K_a$ values, on average, 10 chemical shifts were fitted with a global p$K_a$ value. For neighboring residues with a p$K_a$ value in a comparable range (Y5/K6, K12/Y16, E20/D23/E27, E31/D52, E46/D49, and K45/K53/K55/K56), at which the transitions showed coupled behavior (mirror images), the p$K_a$ values were fitted simultaneously with the macroscopic model.[21,22] In the case of E20/D23/E27, the system was not fitted to three macroscopic p$K_a$ values in a direct way, but to two times two macroscopic p$K_a$ values for which D23 served as the interface. The K45/K53/K55/K56 system was treated similarly. In the case of known transitions in the chemical shift, this further information was added for the global p$K_a$ value determination (with an uncertainty of 0.3 for D, E, H, and the termini, and 0.9 for Y and K considered for error estimation). In unclear cases, local variable p$K_a$ values were added to ensure a suitable baseline. For transitions not fully covered by experimental pH values, assumptions of the chemical shift changes have been made according to ref. [14] by using $-50\%$ of standard chemical shift differences caused by the titration. Effects on the chemical shift at low pH have been fitted to Hendersson–Hasselbalch curves by using p$K_a$ values of $<0$. This reflected the water hydronium transition or salt concentration effects. For error estimation, Monte-Carlo simulations, with 1000 steps and a random variation of pH ($\pm 0.1$), as suggested and discussed in ref. [41] and $^{1}H$ ($\pm 0.01$ ppm) and $^{13}C$ ($\pm 0.02$ ppm) chemical shifts were executed. The latter was determined by analyzing the chemical shifts without any transitions. Please note that possible systematic errors might not be captured by the data and error analysis, so the true error might be higher. The estimated p$K_a$ values refer to 10% $D_2O$.

## Acknowledgements

*We thank Kathrin Waldheim for excellent technical assistance. This research was supported by the Deutsche Forschungsgemeinschaft (WE 5587/1-1).*

## Conflict of Interest

*The authors declare no conflict of interest.*

**Keywords:** electrostatic interactions · ionization potentials · NMR spectroscopy · pH titrations · proteins

[1] W. Kauzmann, *Adv. Protein Chem.* **1959**, *14*, 1.
[2] G. A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Structures*, SpringerLink (Online service), Study Edition, Springer, Berlin, **1991**, p. 1 online resource.
[3] S. K. Burley, G. A. Petsko, *Science* **1985**, *229*, 23.
[4] C. N. Pace, J. M. Scholtz, G. R. Grimsley, *FEBS Lett.* **2014**, *588*, 2177.
[5] T. E. Creighton, *The Biophysical Chemistry of Nucleic Acids & Proteins*, Helvetian Press, [S.l.], **2010**, pp. 1 online resource xxxvii.
[6] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. B. Liu, M. H. M. Olsson, *Chem. Rev.* **2006**, *106*, 3210.
[7] A. Warshel, J. Aqvist, *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267.
[8] R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz, C. N. Pace, *Protein Sci.* **2006**, *15*, 1214.
[9] M. Tollinger, K. A. Crowhurst, L. E. Kay, J. D. Forman-Kay, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4545.
[10] M. A. S. Hass, F. A. A. Mulder, *Annu. Rev. Biophys.* **2015**, *44*, 53.
[11] S. Lindman, S. Linse, F. A. A. Mulder, I. Andre, *Biophys. J.* **2007**, *92*, 257.
[12] D. G. Isom, C. A. Castaneda, B. R. Cannon, B. E. Garcia-Moreno, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5260.
[13] J. D. Forman-Kay, G. M. Clore, A. M. Gronenborn, *Biochemistry* **1992**, *31*, 3442.
[14] G. Platzer, M. Okon, L. P. McIntosh, *J. Biomol. NMR* **2014**, *60*, 109.
[15] G. R. Grimsley, J. M. Scholtz, C. N. Pace, *Protein Sci.* **2009**, *18*, 247.
[16] J. Wallerstein, U. Weininger, M. A. Khan, S. Linse, M. Akke, *J. Am. Chem. Soc.* **2015**, *137*, 3093.
[17] U. Weininger, M. Zeeb, P. Neumann, C. Low, M. T. Stubbs, G. Lipps, J. Balbach, *Biochemistry* **2009**, *48*, 10030.
[18] M. Zeeb, G. Lipps, H. Lille, J. Balbach, *J. Mol. Biol.* **2004**, *336*, 227.
[19] M. E. Milla, R. T. Sauer, *Biochemistry* **1994**, *33*, 1125.
[20] Z. S. Hendsch, T. Jonsson, R. T. Sauer, B. Tidor, *Biochemistry* **1996**, *35*, 7621.
[21] L. P. McIntosh, D. Naito, S. J. Baturin, M. Okon, M. D. Joshi, J. E. Nielsen, *J. Biomol. NMR* **2011**, *51*, 5.
[22] J. T. Edsall, J. Wyman, *Thermodynamics, Electrostatics and the Biological Significance of the Properties of Matter*, Academic Press, New York, **1958**.
[23] I. André, S. Linse, F. A. A. Mulder, *J. Am. Chem. Soc.* **2007**, *129*, 15805.
[24] C. A. Fitch, G. Platzer, M. Okon, B. Garcia-Moreno, L. P. McIntosh, *Protein Sci.* **2015**, *24*, 752.
[25] J. A. Vila, Y. A. Arnautova, Y. Vorobjev, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5602.
[26] J. G. Pelton, D. A. Torchia, N. D. Meadow, S. Roseman, *Protein Sci.* **1993**, *2*, 543.
[27] U. Weininger, K. Modig, A. J. Geitner, P. A. M. Schmidpeter, J. R. Koch, M. Akke, *Biochemistry* **2017**, *56*, 334.
[28] K. Saraboji, M. Hakansson, S. Genheden, C. Diehl, J. Qvist, U. Weininger, U. J. Nilsson, H. Leffler, U. Ryde, M. Akke, D. T. Logan, *Biochemistry* **2012**, *51*, 296.
[29] N. A. Oktaviani, T. J. Pool, H. Kamikubo, J. Slager, R. M. Scheek, M. Kataoka, F. A. A. Mulder, *Biophys. J.* **2012**, *102*, 579.
[30] R. Anandakrishnan, B. Aguilar, A. V. Onufriev, *Nucleic Acids Res.* **2012**, *40*, W537.
[31] J. Myers, G. Grothaus, S. Narayanan, A. Onufriev, *Proteins Struct. Funct. Bioinf.* **2006**, *63*, 928.
[32] K. P. Kilambi, J. J. Gray, *Biophys. J.* **2012**, *103*, 587.
[33] S. Lyskov, F. C. Chou, S. O. Conchuir, B. S. Der, K. Drew, D. Kuroda, J. Q. Xu, B. D. Weitzner, P. D. Renfrew, P. Sripakdeevong, B. Borgo, J. J. Havranek, B. Kuhlman, T. Kortemme, R. Bonneau, J. J. Gray, R. Das, *Plos One* **2013**, *8*, e63906.
[34] K. P. Tan, T. B. Nguyen, S. Patel, R. Varadarajan, M. S. Madhusudhan, *Nucleic Acids Res.* **2013**, *41*, W314.
[35] C. N. Pace, G. R. Grimsley, J. M. Scholtz, *J. Biol. Chem.* **2009**, *284*, 13285.
[36] G. Lipps, M. Stegert, G. Krauss, *Nucleic Acids Res.* **2001**, *29*, 904.
[37] A. L. Hansen, L. E. Kay, *J. Biomol. NMR* **2011**, *50*, 347.
[38] Y. Oda, T. Yamazaki, K. Nagayama, S. Kanaya, Y. Kuroda, H. Nakamura, *Biochemistry* **1994**, *33*, 5275.
[39] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277.
[40] B. A. Johnson, *Methods Mol. Biol.* **2004**, *278*, 313.
[41] H. Webb, B. M. Tynan-Connolly, G. M. Lee, D. Farrell, F. O'Meara, C. R. Sondergaard, K. Teilum, C. Hewage, L. P. McIntosh, J. E. Nielsen, *Proteins Struct. Funct. Bioinf.* **2011**, *79*, 685.

Manuscript received: October 18, 2018
Accepted manuscript online: December 4, 2018
Version of record online: February 11, 2019

 