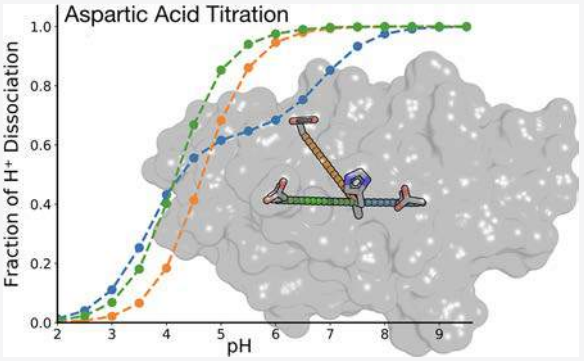
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Charge Interactions in a Highly Charge-Depleted Protein

Stefan Hervø-Hansen, Casper Højgaard, Kristoffer Enøe Johansson, Yong Wang, Khadija Wahni, David Young, Joris Messens, Kaare Teilum, Kresten Lindorff-Larsen,* and Jakob Rahr Winther*

Cite This: J. Am. Chem. Soc. 2021, 143, 2500−2508
Read Online

ABSTRACT: Electrostatic forces are important for protein folding and are favored targets of protein engineering. However, interactions between charged residues are difficult to study because of the complex network of interactions found in most proteins. We have designed a purposely simple system to investigate this problem by systematically introducing individual and pairs of charged and titratable residues in a protein otherwise free of such residues. We used constant pH molecular dynamics simulations, NMR spectroscopy, and thermodynamic double mutant cycles to probe the structure and energetics of the interaction between the charged residues. We found that the partial burial of surface charges contributes to a shift in pK_a value, causing an aspartate to titrate in the neutral pH range. Additionally, the interaction between pairs of residues was found to be highly context dependent, with some pairs having no apparent preferential interaction, while other pairs would engage in coupled titration forming a highly stabilized salt bridge. We find good agreement between experiments and simulations and use the simulations to rationalize our observations and to provide a detailed mechanistic understanding of the electrostatic interactions.



INTRODUCTION

The most fundamental biochemical reactions in proteins like enzymatic catalysis,[1] redox reactions,[2] H^+ transfer,[3] and ion homeostasis[4] are governed by electrostatics. Electrostatic properties are substantially modulated by the protonation status of the titratable residues yielding positive and negative charges. By changing the solvent pH, we effectively change the electrostatic interactions which results in adaptation of protein function and/or protein structure. Consequently, from a protein engineering point of view, it is desirable to understand and predict the effects of changing the pH of the solution. Furthermore, the insertion/removal of titratable residues from proteins affects the pK_a values of other residues nearby which may impact on features like catalysis and structural stability. Over the past decades, a number of studies have addressed these problems utilizing theoretical, computational, and structure-based methods for the calculation of pK_a values and benchmarking against experimental studies. In a similar fashion, many studies have been conducted on various model systems with either artificial or biologically relevant salt bridges. The possibility of predicting the effect of a charge perturbation on stability and pK_a has resulted in many "rational modifications" by, for example, insertion of titratable residues causing charge−charge interactions,[5] the filling of cavities by hydrophobic residues causing dehydration,[6] and the insertion of polar residues causing hydrogen bonding.[7] Despite the many important findings by these studies, many supposedly rational substitutions cause unpredictable and complicated results. These have been attributed to either long-range electrostatic effects, the alteration of several interactions upon mutagenesis, or denatured-state ensemble stabilization.[8]

Because of the complication of predicting the outcome of inserting or removing titratable residues in proteins, we have turned to a model protein that we have previously redesigned to be devoid of the ionizable amino acid residue types Asp, Glu, Arg, Lys, His, Tyr, and (free) Cys.[9] This protein is based on the cellulose binding domain EXG:CBM[10] and is well folded, functional, and stable and makes it possible to reduce the complexity of the multiple electrostatic interactions seen in almost all model proteins studied hitherto. The charge-depleted protein consequently permitted us to study interactions between pairs of charged residues individually and at the same time measure the effect of partial burial of the charge pairs. We modeled the effect of introduced charges in the EXG:CBM system using discrete constant pH molecular dynamics (CpHMD) to provide pH-dependent atomistic structural details about the system and protonation states. To complement and validate our simulations experimentally, we followed the protonation states using nuclear magnetic

Received: October 12, 2020
Published: February 2, 2021

https://dx.doi.org/10.1021/jacs.0c10789
J. Am. Chem. Soc. 2021, 143, 2500−2508

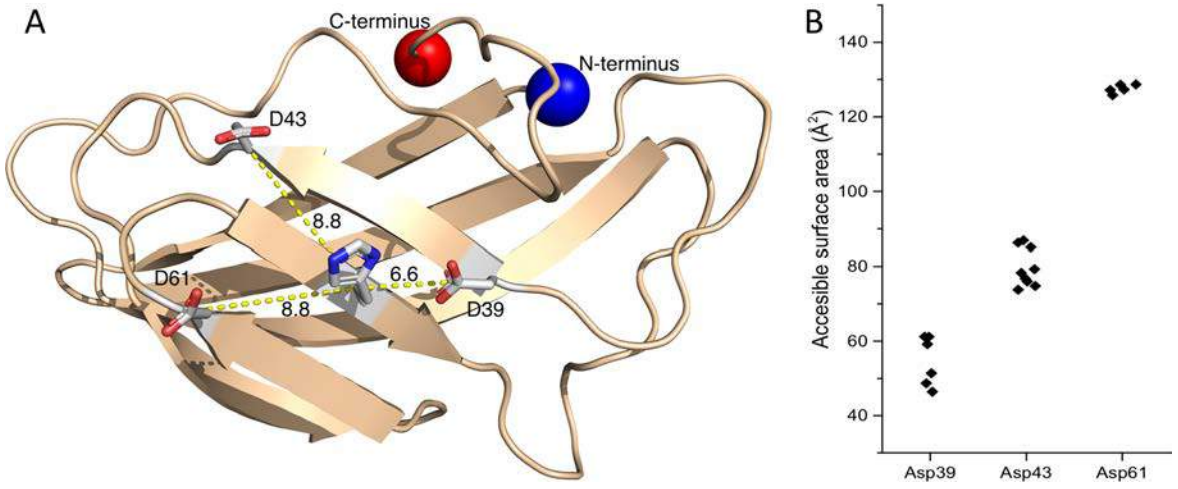


Figure 1. (A) Structure of EXG:CBM^{QQQW} with the introduced Asp and His residues shown in stick representation and the charged N- and C-termini clearly visualized. These Asp and His residues are not present in the crystal structure but are instead introduced here for illustrative purposes. Distances (in Å) between C_α atoms of the aspartate residues and the histidine residue are shown. (B) Solvent accessible surface area (SASA) of the carboxylic acid group of an aspartate in the three single aspartate containing variants, with points representing the side-chain conformations from the Pymol rotamer library.

resonance (NMR) spectroscopy, which, due to its high resolution, allows for monitoring of site-specific effects. We further used double mutant cycles to estimate pH-dependent coupling energies between charge pairs and to validate the protonation cycles obtained from simulations.

RESULTS AND DISCUSSION

Structure Determination of a Charge-Depleted Protein. We have previously described a variant of the cellulose binding domain protein EXG:CBM from a *Cellulomonas fimi* cellulase, in which we substituted four ionizable amino acid residues with nontitratable residue types (K28M, D36Q, R68M, and H90W).[9] In most respects, this variant performed like the wild-type protein; however, protein yields from expression in *E. coli* were relatively low. For this study, we identified a new charge-depleted variant, EXG:CBM^{QQQW}, from a screen for higher protein yields. The new variant differs from the former by two amino acid substitutions (M28Q and M68Q) but has a protein yield that is 4−5 times higher. We did not observe any ill effects of these substitutions, and this variant was chosen as starting point for our further analyses.

We determined a crystal structure of EXG:CBM^{QQQW} to enable us to interpret our biochemical experiments more precisely and to use it as starting point for our molecular dynamics simulations. In the published NMR solution structure of EXG:CBM, the first 3−4 residues were unstructured,[10] and so to aid crystallization of the new charge-depleted variant, we produced a truncated version with the four residues deleted from the N-terminal. Our crystals diffracted to 2.2 Å resolution (PDB ID 6QFS) with detailed data collection and refinement statistics found in the Supporting Information (Table S1). The eight molecules of EXG:CBM^{QQQW} in the asymmetric unit were conformationally homogeneous and well-superimposable (Figure S1A,C) and displayed close structural similarity to the NMR structure of wild-type EXG:CBM[10] with an aligned backbone RMSD of 1.5 Å. This supports our previous assertion that removing all charged residues from EXG:CBM had a minimal effect on the structure. We have used this crystal structure, which remarkably represents the first in the PDB of a protein of more than 100 amino acids completely free of ionizable side chain amino acid residues, throughout the present work.

Systematic Insertion of Ionizable Residues. We aimed to analyze the interaction between pairs of charged residues on the surface of EXG:CBM^{QQQW}, whose only formal charges before insertion are on the N- and C-termini. To monitor charge interactions, we introduced a histidine (T66H mutation) at the protein surface in the middle of a β-sheet, assuming this would result in minimal conformational changes when introducing other charges (Figure 1A). We then inserted, one at a time, three aspartic acid residues around H66 creating D39-H66 (T39D mutation), D43-H66 (S43D mutation), and D61-H66 (S61D mutation). The three aspartates were placed with comparable C^α−C^α distances of 7−9 Å (Figure 1A), typical for aspartate−histidine salt bridges,[11,12] but with different degree of solvent exposure (Figure 1B). Neither the N- nor C-terminal charges were expected to interact significantly with this part of the protein. The N-terminus is unstructured, placed on the opposite face of the H66, and did not show any signs of interactions to the rest of the protein in wild-type EXG:CBM.[9] The C-terminus is located around 20 Å away from H66 on the opposite face of the protein and should therefore also have negligible interaction with all the introduced charges.

Excellent Agreement in pK_a Determination by CpHMD and NMR Spectroscopy. We determined the pK_a values of all the titratable histidine and aspartate residues for all the designed EXG:CBM variants using both CpHMD and NMR spectroscopy (Figure 2 and Table S2, Figures S2 and S3), except for the experimental pK_a measurement of the D39 variant which showed low expression levels. For this variant we instead inferred the pK_a-value using stability data (see the Materials and Methods section). Comparing the pK_a values predicted from the CpHMD with the values obtained by NMR revealed a very good agreement with a RMSE of 0.39 pH units, which was lower than or equal to many model development studies[13] and studies aimed toward protein characterization.[14] With the exception of D39 in the D39-H66 variant, we found a small but systematic overestimation of pK_a values for the

https://dx.doi.org/10.1021/jacs.0c10789
J. Am. Chem. Soc. 2021, 143, 2500−2508

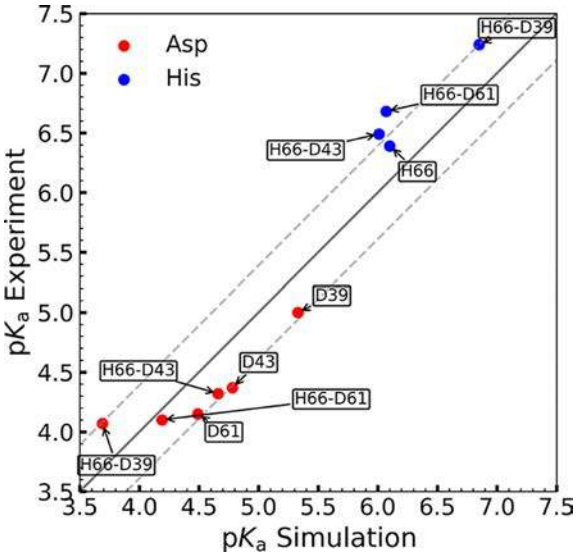


Figure 2. Comparing pK_a values from CpHMD with experimental pK_a values from NMR spectroscopy. Each point corresponds to a specific residue (red aspartate; blue histidine) in a specific EXG variant (label) so that, for example, the blue point labeled H66-D43 represents the pK_a values of H66 in the EXG variant H66-D43. The full line is the diagonal, and the dashed lines represent the RMSE (0.36 pK_a units) between experiment and computation. In all cases the experimental and computational errors are within extent of the symbols used (actual values are shown in Table S2).

aspartate residues and underestimation for the histidine residues by CpHMD.

As expected, we found that the presence of H66 shifted the pK_a values of the three Asp residues downward compared to having a Thr at position 66. Examining the magnitude of this effect, we found that while both D43 and D61 were relatively weakly affected by the presence of H66, the pK_a value of D39 was shifted substantially by the presence of H66 (Figure 2). Looking at the experimental pK_a values, we found both D43 and D61 were shifted down by 0.05 units, whereas D39 was shifted by ∼1 unit. These observations are corroborated by the CpHMD results, thus validating that simulations can be used to interpret the structural origin of the pK_a shifts. We note, however, also that the effect of H66 appeared to be larger in the simulations compared to the experiments, suggesting that the charge−charge interactions in computation might be overestimated (see below). Turning to H66, we found a similar picture. Specifically, we found that while the pK_a value of H66 was only mildly affected by D43 and D61, the pK_a value of H66 was upshifted by 0.85 pK_a units (equivalent to 4.9 kJ/mol) in the presence of D39.

Inspecting the titration profiles for Asp C$^\gamma$ obtained from NMR spectroscopy, we observed that all the charged residues in the single-charged variants showed single-step titrations with Hill coefficients estimated to be ∼1. We also observed monophasic titration profiles for Asp C$^\gamma$ from both the D43-H66 and D61-H66 variants, while Asp C$^\gamma$ in the signals from the D39−H66 variant showed biphasic titration profiles (Figure S2). We note that we observe unusually high chemical shifts for the C$^\gamma$ of D39 in this D39−H66 variant, in particular at low pH when the residue becomes protonated. In the simulations at pH 2, we observed in 23% of the structures that the protonated carboxylic acid of D39 acted as a hydrogen bond donor with Q68 being the acceptor and also observed a minor population of hydrogen bonds from D39 to N103 (3.7%), T41 (0.5%), and H66 (0.4%). While the origin of the unusual chemical shifts of D39 remains unclear, we speculate that these hydrogen bond interactions with D39 could play a role in explaining this observation.

In agreement with the NMR spectroscopy data, CpHMD also predicted the titration profile for aspartate and histidine to conform fairly well to the Henderson−Hasselbalch equation for all variants with the exception of the D39−H66 pair, which exhibited biphasic titration curves (Figure S3). The biphasic shape and the symmetry of the D39−H66 titration curves clearly showed that in this variant the aspartate protonation state was dependent on the histidine protonation state, and vice versa. We used a diprotic acid model, in which the two titratable sites together could adopt four microstates (thus ignoring the histidine tautomerism),[15] to fit the CpHMD data from D39-H66 (Figure 3) and derived a coupling energy

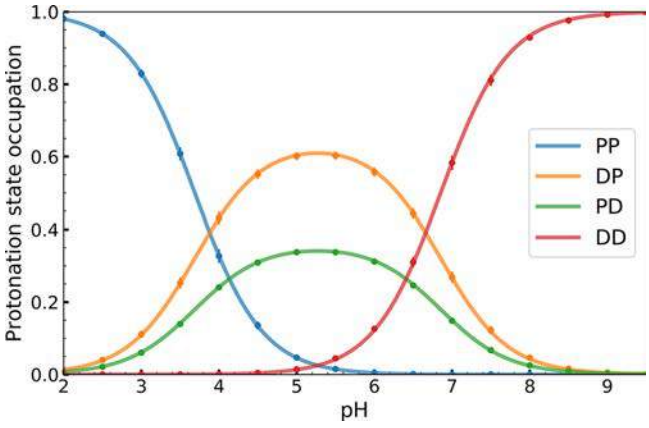


Figure 3. Bjerrum diagram of the diprotic D39-H66 EXG variant with the legend specifying the protonation state of D39 (first) and H66 (second) to be either protonated (P) or deprotonated (D). The points are sampled by CpHMD, while the solid line is the best fit obtained from the diprotic acid model. The microscopic pK_a values (where the subscript is the reactant and the superscript is the product) obtained from the fit are pK_{PP}^{DP} = 3.88 ± 0.04, pK_{PP}^{PD} = 4.13 ± 0.03, pK_{DP}^{DD} = 6.66 ± 0.04, pK_{PD}^{DD} = 6.41 ± 0.03, and a coupling energy $\Delta\Delta G^\circ$ = 14.6 ± 0.4 kJ/mol.

between the two nonindependent titrations of the D39−H66 pair to be 14.6 kJ/mol. The large coupling energy emphasized that the condition of both the aspartate and histidine being simultaneously protonated or deprotonated was much more energetically unfavorable than the zwitterion and doubly neutral state. Because of strong coupling between D39 and H66, the pK_a value for the two residues was obtained from the diprotic acid model through macroscopic pK_a values (Figure 2), yielding a pK_a of 3.69 and 6.85 for the first and second titration event, respectively. These two pK_a values corresponded to the pH of the first and second transitions of the biphasic titration curve and should not be confused with single site-specific pK_a values (known as p$K_{1/2}$), which cannot be obtained from the D39−H66 pair. It has previously been shown that the difference between p$K_{1/2}$ and macroscopic pK_a values for diprotic acids is zero when the coupling energy is either zero or very strong;[15] thus, we could safely assume the macroscopic pK_a values obtained from CpHMD was comparable to the experimentally measured pK_a values for the D39−H66 pair.

https://dx.doi.org/10.1021/jacs.0c10789
J. Am. Chem. Soc. 2021, 143, 2500−2508

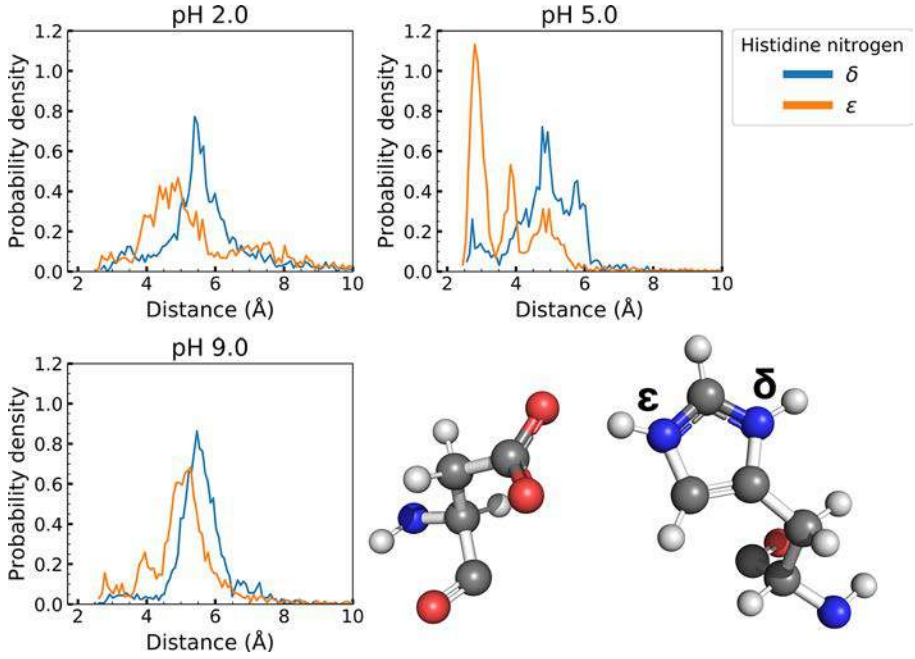


Figure 4. Distance distribution between the N^{δ} or N^{ϵ} (visualized in the bottom right) of His66 and the center of charge of the oxygen in the carboxylate group in the side chain of Asp39 for the Asp39−His66 pair at pH 2.0, 5.0, and 9.0 as obtained from the molecular dynamics simulation.

D39−H66 Pair Adopts Protonation-State-Dependent Conformations Optimal for Salt Bridging. To investigate the molecular details of the strong interaction of the D39−H66 pair, we used the CpHMD simulations to perform a structural analysis. We calculated the distribution of distances between the N^{δ} and N^{ϵ} atoms in H66 and the center of the two side chain oxygens of D39 (Figure 4). At pH 5, we observed a distinct peak around 2.8 Å between the N^{ϵ} and the aspartate carboxylate center of charge. At this pH, the His−Asp system was almost fully in a single-protonated state with ∼60% having a protonated His and negatively charged Asp and ∼40% having a protonated Asp and a neutral His (Figure 2), leading to strong direct interactions between the two residues. Examining the structure corresponding to the short-distance peak, we indeed observed that the N^{ϵ} and associated titratable proton of the histidine side chain were facing the carboxylate group (Figure 4). Both the zwitterionic and neutral states possess hydrogen-bonding capacity; however, the zwitterionic state additionally possessed formal charges for the aspartate and histidine residue, thus causing the interaction to resemble a combination of Coulombic interaction and hydrogen bonding, that is, a salt bridge.

At pH 2.0 and 9.0, where the two residues were either both protonated or deprotonated, respectively, we instead observed a distribution of longer distances to N^{ϵ} between ∼4 and ∼6 Å, which were also found as a minor population at pH 5.0 (Figure 4). Examining the joint distribution of χ_1 angles of D39 and H66, we found that all nine combinations of the three χ_1 rotamers for each residue were sampled, though we saw varying populations that depended substantially on pH (Figure S7). At pH 2.0 and 9.0 D39 was mostly found with χ_1 in the p-rotamer ($\chi_1 \sim 60°$),[16] whereas at pH 5.0 it mostly sampled the t-rotamer ($\chi_1 \sim 180°$). The histidine was mostly found in the t-rotamer, sampling also the m-rotamer ($\chi_1 \sim -60°$) at low pH. When D39 was in the p-rotamer, the aspartate side chain faced the solvent, whereas the t-rotamer state populated at pH 5.0 corresponded to the conformation from the short-distance peak at pH 5 in Figure 4, with the aspartate side chain facing the side chain of H66 (Figure S8). These pH-dependent conformations and their exchange can be explained from a simple charge−charge perspective; an unpaired charge will face the solvent to minimize its burial, while the paired charge of D39 and H66 will face one another to engage in strong interactions, predominantly salt bridging.

The NMR experiments also provided some structural insight into the aspartate−histidine pairs because the tautomeric state of a protonated histidine could be determined from its C^{δ} chemical shift.[17] In the D39−H66 variant the C^{δ} chemical shift of His66 decreased at the beginning of the H66 titration, consistent with the $N^{\epsilon}H$ tautomer being dominant, but at higher pH the signal disappeared (Figure S2). In the other variants, the C^{δ} peak also disappeared at pH coinciding with the histidine titration, suggesting that the variants had a mixture of both tautomers in intermediate exchange. The signals from H66−C^{δ} in the three different Asp variants were also affected differently by the protonation of the aspartates. Interestingly, they seemed to react differently at the beginning of the H66 titration, indicating that the three aspartates influenced the tautomer distribution differently.

As an alternative to the CpHMD, we found PROPKA3,[18] in general, to predict the pK_a of Asp and His fairly well; however, for interactions affected by multiple rotameric states (e.g., of the 39D−66H variant) some predictions differed >1 pK_a unit from the experimental value (Figure S11). While averaging with the weight of each rotameric state obtained from the MD simulations, improved the agreement between PROPKA3 and experimental values, such a feature is unavailable in PROPKA3.

Partial Burial of Surface-Exposed Ionizable Residues Influence pK_a Values. The pK_a values obtained from the titration of the single aspartate variants, by both CpHMD and NMR spectroscopy, differed by up to ∼1 pK_a unit despite all being on the protein surface. We attribute the differences in pK_a values to the difference in partial burial (partial dehydration) of the aspartate residues, which corresponds

https://dx.doi.org/10.1021/jacs.0c10789
J. Am. Chem. Soc. 2021, 143, 2500−2508

well with the upshift in pK_a values and is due to favoring of the carboxylic acid group over the carboxylate anion in a low dielectric environment. As a consequence, the pK_a of D39 was higher than the pK_a of D43, which again was higher than that of D61. The pK_a value of D61 was very close to the reference value for aspartate in the CpHMD model, which could be considered a reference of a fully solvated aspartate. The presence of hydrogen bonds could be another possible contribution to the perturbed pK_a values. However, from the CpHMD simulations we did not find other residues to engage in long-lived hydrogen bonding with the aspartates, thus supporting the hypothesis that pK_a differences are due to the different degrees of hydration. These effects appear to be captured by the generalized Born solvation model and correspond well to SASA data obtained from the crystal structure of EXG:CBM^{QQQW}.

The CpHMD simulation also provided some interesting insight into the different conformational heterogeneities of the three aspartate residues. Being located at the protein surface, the aspartate side chains were free to align along the protein backbone or to stretch out into the solvent. This conformational rearrangement could influence the pK_a value of a titratable group due to the energetic relief upon a change in protonation state. By looking at the first dihedral angle of the aspartate side chain (χ_1), we could observe this pH-dependent conformational change for the D39 variant, while it was much less pronounced in the D43 and D61 variants (Figure S9). The absence of substantially populated hydrogen bonds between the charged residues and the backbone protons, and the fact that no significant structural rearrangements take place in the simulations, also agrees with the experimental observations. This is also in agreement with the observation that the H^N chemical shifts in all the EXG:CBM variants (except for the H^N-shifts of the charged residues) changed with <0.2 ppm in response to the pH titration.

Thermodynamic Cycles Reveal Different Modes of His−Asp Interaction. The results described above demonstrated substantial perturbation of the pK_a values of several of the residues and also highlighted strong coupling between the titration behavior of Asp39 and His66. To get an alternative view of the coupling between the residues in the three charge pairs, we turned to thermodynamic double mutant cycles (DMC). In DMC, the change of the stability from removing a charged residue in a charge pair was subtracted from the change of the stability from removing the same residue in a protein variant lacking the charged partner (Figure 5). In our analysis, we treated the charge pair as the reference state and mutated the protein sequence back to the original EXG:CBM^{QQQW} sequence. We performed the DMC experiments at several different values of pH and also in some cases in the presence of high concentrations of salt.

All the tested variants displayed classical two-state equilibrium unfolding behavior (Figure S4; for individual unfolding free energies see Table S3). Also, comparing mutant variants, we observed only very few changes in the chemical shifts in ^{15}N-HSQC NMR spectra, suggesting the structures to be very similar (Figure S5). The stability data and thermodynamic cycles are presented in Figures S4 and S6, and the summarized results of apparent coupling energies are shown in Table 1. The DMC data confirmed that although all three aspartates are placed at potential salt bridge distances to the histidine, their interactions were very different.

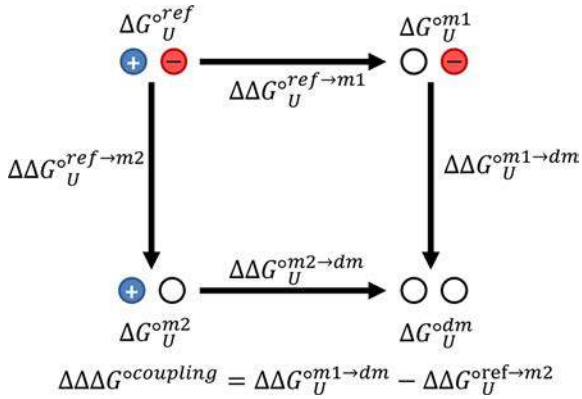


Figure 5. Double mutant cycle (DMC) relating the unfolding energies, ΔG°_U, of a protein with two charged residues (ref), two single mutants (m1 and m2), and a double mutant with no charged residues (dm). The blue (+), red (−), and white (neutral) circles represent the charge state of the two residues investigated.

Table 1. Apparent Coupling Standard Free Energies from Double Mutant Cycles

EXG variant	pH 2.5 (kJ/mol)	pH 5 (kJ/mol)	pH 8 (kJ/mol)	pH 10 (kJ/mol)
D39−H66	−0.3 ± 1.1	5.2 ± 1.1	4.6 ± 1.0	3.7 ± 0.9
D39−D66 w/ salt		5.2 ± 1.4	4.2 ± 1.3	
D43−H66	0.1 ± 1.2	2.4 ± 1.2	2.0 ± 1.1	
D43−H66 w/ salt		−1.1 ± 1.5		
D61−H66	−1.0 ± 1.2	0.3 ± 1.2	0.9 ± 1.1	

We focus first on the results at pH 5, where the His and Asp residues together will typically have taken up a single proton, generally associated with H66. In line with the strong coupling of the titration behavior of D39−H66, we found that the coupling energy from the DMC was also higher for this pair than the other two (Table 1). Protonating the aspartates at pH 2.5 decreased the coupling energies to close to zero, suggesting that the charge interactions between the Asp and His pairs contributed substantially to the coupling energies measured at pH 5.

Increasing pH to 8.0 or adding salt, however, revealed a more complex picture. While the couplings generally become smaller in these cases, they remained large, particularly for D39−H66. At pH 5.0 the coupling standard free energy of D39−H66 remained constant at 5.2 kJ/mol even after addition of 1.5 M NaCl and only dropped to 4.6 and 3.7 kJ/mol at pH 8 and 10, respectively. Thus, these results suggest a strong interaction between D39−H66 relative to the reference state, even when both residues were deprotonated.

Overall, the data therefore suggest three different cases in the three aspartate−histidine pairs. D61 does not seem to interact with H66 to a measurable degree with any of the approaches used in this study. The interaction between D43 and H66 is relatively weak and can be screened by salt, pointing at it being classical Coulombic in nature. The reason this interaction is not more pronounced in the NMR and CpHMD data could perhaps be the higher salt concentrations (0.1 M NaCl) in these experiments. The strong interaction between D39 and H66, on the other hand, cannot be explained purely by a Coulombic interaction since the interaction persist at pH 10 where H66 is not positively charged and cannot be

https://dx.doi.org/10.1021/jacs.0c10789
J. Am. Chem. Soc. 2021, 143, 2500−2508

screened by 1.5 M salt. The latter is however, not unusual for solvated[19] and partially solvated[20] salt bridges. We cannot rule out that these coupling energies contain contributions from denatured state effects,[8,21] though we expect that such effects would dominate in D61−H66 (due to the short sequence separation of the two charged residues) and be smallest in D39−H66.

CONCLUDING REMARKS

We have designed and explored a simple model system to study protein electrostatics, looking at three different positions for Asp residues surrounding a His residue. Even for residues that one would typically categorize as solvated, we found differences in pK_a of up to ∼1 unit, equivalent to ∼5.7 kJ/mol, most likely due to differences in hydration. However, the perturbation by partial burial was only secondary compared to the impact of the interaction between the histidine and aspartate residues. From the three Asp−His pairs studied here, only one engaged in a salt bridge, causing strongly coupled titration curves. Thus, CpHMD simulations suggested the interaction between the two strongly coupled titration sites to favor salt bridging over hydrogen bonding obtained (60:40) from the proton tautomerization. This was most likely due to the charge pair being partially solvated on the surface of EXG. Finally, because of the generally good agreement between the CpHMD simulations and NMR experiments, we were able to gain detailed mechanistic insight into the relationship between protein interactions and titrations. Specifically, the biphasic titration observed here was caused by strong electrostatic interactions, and the observed pH-dependent coupling energies became easier to interpret. As such, we anticipate this work will expand the general knowledge of protein electrostatics studies.

MATERIALS AND METHODS

Molecular dynamics simulations were performed in the Amber16 PMEMD (Particle Mesh Ewald Molecular Dynamics) program[22,23] with long-range electrostatics handled by particle mesh Ewald summation[24] and default LJ and short-range electrostatic cutoffs by using the AMBER constph force field, derived from ff14SB,[25] in explicit TIP3P water.[26] All simulations were run with pH replica exchange enhanced sampling[27,28] in the pH range 2.0−9.5 for the double titratable variants and single histidine variant and the pH range 0.5−8.0 for the single aspartate variants with all replica interspaced by 0.5 pH units with each replica running 100 ns. The systems were first minimized and next equilibrated in terms of temperature and pressure to 300 K and 1 bar pressure separately by Langevin dynamics[29,30] with a collision frequency of 5.0 ps^{-1} and the Berendsen barostat[31,32] with a rise time (coupling constant) of 1 ps and isothermal compressibility of 44.6×10^{-6} bar^{-1}. A time step of 2 fs was utilized and the SHAKE algorithm[33] applied to constrain hydrogen-heavy atom bond lengths. The equilibration was followed by a production run in the semigrand canonical ensemble of protons ($NV\Delta\mu_{H^+}$) where only histidine and aspartate residues were allowed to titrate every 0.2 ps through a Monte Carlo protonation state move for each titratable residue and every 0.4 ps attempt to perform a replica exchange between the simulations to swap pH. The Monte Carlo protonation state move used generalized Born solvation[13,34] with a salt concentration (Debye−Hückel based) of 0.1 M monovalent salt,[35] and upon a successful change in protonation state 0.2 ps of solvent relaxation was conducted after reinstating the explicit water.[13] Configurations were gathered every 50 ps for statistical evaluation.

For single-phase titration curves, pK_a values and Hill coefficients were obtained through averaging the deprotonated factions for each pH of the simulation and fitted to the Hill-modified Henderson−Hasselbalch equation (eq 1),[36−38] yielding a single pK_a value and Hill coefficient for each titratable residue. The Hill-modified Henderson−Hasselbalch equation is given by

$$y_a = \frac{1}{1 + 10^{n_H(pH - pK_{a1})}} \quad (1)$$

where y_a is the deprotonated faction and n_H is the Hill coefficient. This method is similar to titration experiments and thus effectively measures the ensemble-averaged deprotonated fractions. In these analyses we included a Hill coefficient as a control to see whether the transitions would be broadened but found $n_H \sim 1$ in all cases. For biphasic titration curves, the populations of the protonation species for each pH were fitted to a diprotic acid model, yielding macroscopic and microscopic pK_a values as explained in detail by Ullmann.[15] Error estimation was conducted by block averaging the using four equally sized blocks from which standard deviations, and standard errors were obtained.

Protein expression and purification of all variants were done following the same protocol as described for EXG:CBM,[9] although the gel filtration step and second lyophilization were omitted, and the temperature needed to elute the protein from Avicel (Sigma-Aldrich) ranged from 50 to 70 °C. The protein yields after elution from Avicel were ∼30 mg/L *E. coli* culture for EXG:CBM^{QQQW} and EXG:CBM^{QQQW,66H}, 11−18 mg/L for EXG:CBM^{QQQW,43D}, EXG:CBM^{QQQW,61D}, EXG:CBM^{QQQW,39D,66H}, EXG:CBM^{QQQW,43D,66H}, and EXG:CBM^{QQQW,61D,66H}, and 3−6 mg/L for EXG:CBM^{QQQW,39D}. After purification, the proteins were buffer exchanged into the appropriate buffer by using an Illustra NAP-5 column (GE Healthcare).

Protein NMR spectroscopy was performed on a Bruker 600 MHz Avance IIIHD spectrometer with a quadruple resonance cryoprobe. The amide chemical shifts of EXG:CBM^{QQQW,39D,66H} were assigned from NH(CA)CO, HNCO, HNCOCACB, and HNCACB spectra at pH 3.6. Signals for all residues except Ala1 and Cys108 could be identified. The combined assignment of EXG:CBM^{QQQW,39D,66H} and EXG:CBM[9] made it possible to assign the amide signals in the remaining variants. To get the $C^{\epsilon 1}$, $H^{\epsilon 1}$, $C^{\delta 2}$, and $H^{\delta 2}$ chemical shifts of His, spectra of the aromatic carbon region were measured (using the Bruker pulse sequence *trosyargpphwg*), and a H2(C)CO experiment, correlating C^{γ} chemical shifts to H^{β}, was measured to obtain the pK_a of the aspartates (Bruker pulse sequence *hcacogp3d* optimized for carboxylic side chains[39]). Because the C^{α} and the carboxylic acid of the C-terminal glycine resemble an aspartate side chain, the chemical shifts of the C-terminus was also directly measured in the H2(C)CO experiment. Side chain assignment of the relevant residues was easy because each spectrum only contained one set of Asp, His, or C-terminal signals. For the titrations, NMR samples (0.3 mM protein, 100 mM KCl, 10% D_2O, 1 mM DDS) were split in two (one titrating up, one titrating down) by addition of NaOH or HCl, respectively. Appropriate side chain spectra together with a ^{15}N-HSQC were measured at 25 °C. The pH was adjusted in steps of 1/3 pH units, and the pH of the sample was measured (immediately before and after recording the NMR spectra) by using a pH meter calibrated at room temperature. Titration curves were fitted to the modified Hill equation,[36−38] while chemical shifts with two titration events were fitted to

$$\delta_{obs} = \left(\frac{\delta_a}{1 + 10^{pH - pK_{a1}}}\right) + \left(\frac{\delta_b}{1 + 10^{pH - pK_{a2}}}\right) + \delta_{off} \quad (2)$$

where δ_{obs} is the observed chemical shift, δ_a and δ_b are the change in chemical shift due to the first and second titration with pK_{a1} and pK_{a2}, respectively, and δ_{off} is the offset. For fitting, the chemical shifts of C^{γ} (Asp), $C^{\epsilon 1}$ (His), or C (C-term) were used. In the case of incomplete titrations of the side chain signals, the amide and side chain ^{13}C signal of a changed residue were globally fit to obtain the pK_a value. The pK_a value of EXG:CBM^{QQQW,66H} was fitted from amide signals.

Protein Stability and Double Mutant Cycles. All experiments were conducted in 5 mM buffer (pH 2.5, sodium phosphate; pH 5, sodium acetate; pH 8, MOPS; pH 10, boric acid). Urea solutions were made fresh daily and incubated with AG 501-X8 resin (Biorad)

https://dx.doi.org/10.1021/jacs.0c10789
J. Am. Chem. Soc. 2021, 143, 2500−2508

for a minimum of 5 h to remove ionic contaminants. The refractive index was measured after the addition of buffer compound to determine the exact concentration of each urea stock. Urea, buffer, and protein were mixed to the appropriate concentrations and left to equilibrate at room temperature overnight before measurements. The fluorescence of each sample was measured on a PerkinElmer LS55 at 25 °C. Because of relatively weak transitions between folded and unfolded state, fits to the linear extrapolation method[40] were made globally for six fluorescence signals of each sample (Ex: 280 and 295 nm; Em: 360, 370, and 380 nm) to obtain C_m values and apparent m values, m_{app} (for details see the Supporting Information). For the standard Gibbs free energies, $\Delta G°$, we define the standard state as a hypothetical 1 M solution of the solute present as a dimensionless uncharged point particle. $\Delta G°$ values were calculated from the linear extrapolation model[40] $\Delta G° = C_m\langle m_{app}\rangle$, with $\langle m_{app}\rangle$ being the average of m_{app} from all measurements. Apparent coupling energies (Figure S10) were calculated as the difference between opposite edges in a thermodynamic cycle (i.e., $\Delta\Delta\Delta G^{\circ coupling} = \Delta\Delta G^{\circ mI\rightarrow dm}_{U} - \Delta\Delta G^{\circ wt\rightarrow m2}_{U}$). The error on individual stability determinations reflects the error of the mean m_{app} value, and coupling uncertainties were propagated from these.

For the D39 variant, a pK_a value by NMR titration experiments could not be determined due to low expressions levels; however, the pK_a has instead been estimated from unfolding data. The pK_a of the folded state of D39 can be estimated to be 1.1 pK_a unit higher than in the unfolded state. Assuming that the value in the unfolded state is close to the reference value of aspartate, the pK_a of D39 would be upshifted to ∼5.

Protein Crystallization and Structure Determination by X-ray Diffraction. A truncated EXG:CBM^{QQQW} variant missing the N-terminal residues 2−5 was crystallized by a vapor diffusion hanging drop experiment. EXG:CBM^{QQQW} at a concentration of 32 mg/mL was mixed at a 1:2 ratio with the crystallization precipitant solution (450 mM MES monohydrate pH 6.5, 55 mM HEPES pH 7.5, 45 mM $ZnSO_4$ heptahydrate, 27.5 mM $CdSO_4$ hydrate, 550 mM Na acetate trihydrate, 11.25% PEG monomethyl ether 550) and incubated at 10 °C. Crystals appeared within a week and were cryoprotected prior to cryostorage by mixing 1:1 with a cryoprotectant of 28% ethylene glycol, 19.3% sucrose, 120 μg/mL 3-(1-pyridinio)-1-propane-sulfonate. X-ray diffraction data were collected on a PILATUS 6M detector (Dectris) at the ESRF beamline ID30B at 100 K and a wavelength of 0.954 Å.

Diffraction data were indexed and integrated in XDS,[41] scaled, and merged by using Aimless,[42] and phases were solved by molecular replacement in Phaser[43] with the NMR structure of EXG:CBM (PDB ID1EXG) as a search model.[10] Maximum likelihood refinement was performed in *phenix.refine*[44] against data to a maximum resolution of 2.2 Å with manual model correction in COOT.[45] No disallowed peptide torsion angles were modeled, with 95.64% Ramachandran favored and no outliers. Data collection and refinement statistics are summarized in Table S1. The final coordinates have been deposited in the PDB with the accession code6QFS.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.0c10789.

Experimental details, structure elucidation, titration and NMR, stability determinations, and details on computational data (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Jakob Rahr Winther − *Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark;* orcid.org/0000-0001-6995-9154; Email: jrwinther@bio.ku.dk

Kresten Lindorff-Larsen − *Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark;* orcid.org/0000-0002-4750-6039; Email: lindorff@bio.ku.dk

Authors

Stefan Hervø-Hansen − *Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark; Division of Theoretical Chemistry, Department of Chemistry, Lund University, SE 221 00 Lund, Sweden;* orcid.org/0000-0002-9629-9195

Casper Højgaard − *Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark;* orcid.org/0000-0002-1258-1717

Kristoffer Enøe Johansson − *Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark;* orcid.org/0000-0001-6054-0461

Yong Wang − *Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark;* orcid.org/0000-0001-9156-0377

Khadija Wahni − *VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie, Brussels Center for Redox Biology, and Structural Biology Brussels, Vrije Universiteit Brussel, B-1050 Brussels, Belgium*

David Young − *VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie, Brussels Center for Redox Biology, and Structural Biology Brussels, Vrije Universiteit Brussel, B-1050 Brussels, Belgium*

Joris Messens − *VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie, Brussels Center for Redox Biology, and Structural Biology Brussels, Vrije Universiteit Brussel, B-1050 Brussels, Belgium;* orcid.org/0000-0002-2128-8264

Kaare Teilum − *Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark;* orcid.org/0000-0001-6919-1982

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.0c10789

Author Contributions

S.H.-H. and C.H. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from the Velux Foundations and Novo Nordisk Foundation. Support from the University of Copenhagen for a scholarship to C.H. is gratefully acknowledged. S.H.H. thanks LUNARC in Lund, Sweden, for computational resources. This work was also supported by the Vlaams Instituut voor Biotechnologie and the Strategic Research Programme Grant SRP34 of the Vrije Universiteit Brussel (to J.M.).

■ REFERENCES

(1) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106*, 3210−3235.

(2) Varadarajan, R.; Zewert, T.; Gray, H.; Boxer, S. Effects of buried ionizable amino acids on the reduction potential of recombinant myoglobin. *Science* **1989**, *243*, 69−72.

(3) Luecke, H. Proton Transfer Pathways in Bacteriorhodopsin at 2.3 Angstrom Resolution. *Science* **1998**, *280*, 1934−1937.
(4) Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. The Structure of the Potassium Channel: Molecular Basis of K+ Conduction and Selectivity. *Science* **1998**, *280*, 69−77.
(5) 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. U. S. A.* **2013**, *110*, 11337−11342.
(6) Isom, D. G.; Castaneda, C. A.; Cannon, B. R.; Garcia-Moreno, E. B. Large shifts in pKa values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 5260−5265.
(7) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. Hydrogen Bonding Markedly Reduces the pK of Buried Carboxyl Groups in Proteins. *J. Mol. Biol.* **2006**, *362*, 594−604.
(8) Cho, J.-H.; Raleigh, D. P. Mutational Analysis Demonstrates that Specific Electrostatic Interactions can Play a Key Role in the Denatured State Ensemble of Proteins. *J. Mol. Biol.* **2005**, *353*, 174−185.
(9) Højgaard, C.; Kofoed, C.; Espersen, R.; Johansson, K. E.; Villa, M.; Willemoës, M.; Lindorff-Larsen, K.; Teilum, K.; Winther, J. R. A Soluble, Folded Protein without Charged Amino Acid Residues. *Biochemistry* **2016**, *55*, 3949−3956.
(10) Xu, G.-Y.; Ong, E.; Gilkes, N. R.; Kilburn, D. G.; Muhandiram, D. R.; Harris-Brandts, M.; Carver, J. P.; Kay, L. E.; Harvey, T. S. Solution Structure of a Cellulose-Binding Domain from Cellulomonas fimi by Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **1995**, *34*, 6993−7009.
(11) Kumar, S.; Nussinov, R. Close-Range Electrostatic Interactions in Proteins. *ChemBioChem* **2002**, *3*, 604.
(12) Donald, J. E.; Kulp, D. W.; DeGrado, W. F. Salt bridges: Geometrically specific, designable interactions. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 898−915.
(13) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10*, 1341−1352.
(14) Hofer, F.; Dietrich, V.; Kamenik, A. S.; Tollinger, M.; Liedl, K. R. pH-Dependent Protonation of the Phl p 6 Pollen Allergen Studied by NMR and cpH-aMD. *J. Chem. Theory Comput.* **2019**, *15*, 5716−5726.
(15) Ullmann, G. M. Relations between Protonation Constants and Titration Curves in Polyprotic Acids: A Critical View. *J. Phys. Chem. B* **2003**, *107*, 1263−1271.
(16) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 389−408.
(17) Platzer, G.; Okon, M.; McIntosh, L. P. pH-dependent random coil 1H, 13C, and 15N chemical shifts of the ionizable amino acids: a guide for protein pKa measurements. *J. Biomol. NMR* **2014**, *60*, 109−129.
(18) Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(19) Pylaeva, S.; Brehm, M.; Sebastiani, D. Salt Bridge in Aqueous Solution: Strong Structural Motifs but Weak Enthalpic Effect. *Sci. Rep.* **2018**, 8 DOI: 10.1038/s41598-018-31935-z.
(20) Luisi, D. L.; Snow, C. D.; Lin, J.-J.; Hendsch, Z. S.; Tidor, B.; Raleigh, D. P. Surface Salt Bridges, Double-Mutant Cycles, and Protein Stability: An Experimental and Computational Analysis of the Interaction of the Asp 23 Side Chain with the N-Terminus of the N-Terminal Domain of the Ribosomal Protein L9. *Biochemistry* **2003**, *42*, 7050−7060.
(21) Cho, J. H.; Meng, W.; Sato, S.; Kim, E. Y.; Schindelin, H.; Raleigh, D. P. Energetically significant networks of coupled interactions within an unfolded protein. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (33), 12079−12084.
(22) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.
(23) Le Grand, S.; Gotz, A. W.; Walker, R. C. SPFP: Speed without Compromise- Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374−380.
(24) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(25) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(26) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(27) Itoh, S. G.; Damjanović, A.; Brooks, B. R. pH replica-exchange method based on discrete protonation states. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 3420−3436.
(28) Swails, J. M.; Roitberg, A. E. Enhancing Conformation and Protonation State Sampling of Hen Egg White Lysozyme Using pH Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 4393−4404.
(29) Pastor, R. W.; Brooks, B. R.; Szabo, A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* **1988**, *65*, 1409−1419.
(30) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **2001**, *114*, 2090−2098.
(31) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(32) Morishita, T. Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath. *J. Chem. Phys.* **2000**, *113*, 2976−2982.
(33) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(34) Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **2004**, *25*, 2038−2048.
(35) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **1999**, *101*, 426−434.
(36) Henderson, L. J. Concerning the relationship between the strength of acids and their capacity to preserve neutrality. *Am. J. Physiol.* **1908**, *21*, 173−179.
(37) Hill, T. *Cooperativity Theory in Biochemistry, Steady-State and Equilibrium Systems*; Springer-Verlag: New York, 1985.
(38) Markley, J. L. Observation of histidine residues in proteins by nuclear magnetic resonance spectroscopy. *Acc. Chem. Res.* **1975**, *8*, 70−80.
(39) Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson. (1969-1992)* **1990**, *89*, 496−514.
(40) Greene, R. F.; Pace, C. N. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α-chymotrpsin, and β-lactoglobumin. *J. Biol. Chem.* **1974**, *249*, 5388−5394.
(41) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125−132.
(42) Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1204−1214.
(43) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phasercrystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658−674.

https://dx.doi.org/10.1021/jacs.0c10789
J. Am. Chem. Soc. 2021, 143, 2500−2508

(44) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213−221.
(45) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.

https://dx.doi.org/10.1021/jacs.0c10789
J. Am. Chem. Soc. 2021, 143, 2500−2508

SUPPORTING INFORMATION

Charge Interactions in a Highly Charge-depleted Protein

Stefan Hervø-Hansen[a,b,1], Casper Højgaard[a,1], Kristoffer Enøe Johansson[a], Yong Wang[a], Khadija Wahni[c,d,e], David Young[c,d,e,§], Joris Messens[c,d,e], Kaare Teilum[a], Kresten Lindorff-Larsen[a**], and Jakob Rahr Winther[a*]

[a]Linderstrøm-Lang Centre of Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark, [b]Division of Theoretical Chemistry, Department of Chemistry, Lund University, SE 221 00 Lund, Sweden, [c]VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie - Vrije Universiteit Brussel, B-1050 Brussels, Belgium, [d]Brussels Center for Redox Biology, Vrije Universiteit Brussel, B-1050 Brussels, Belgium, [e]Structural Biology Brussels, Vrije Universiteit Brussel, B-1050 Brussels, Belgium

[*]To whom correspondence may be addressed jrwinther@bio.ku.dk

[**]Co-corresponding author, lindorff@bio.ku.dk

[1] These authors contributed equally to this work.

[§] Current address: The Francis Crick Institute, London, UK.

Contents:

Table S1 – Data collection and refinement statistics

	6QFS
Data collection	
Space group	$P3_221$
Cell dimensions	
a, *b*, *c* (Å)	137.38, 137.38, 148.55
α, β, γ (°)	90, 90, 120
Resolution (Å)	74.27-2.20 (2.24-2.20)*
R_{merge}	0.093 (0.937) *
$I / \sigma I$	16.5 (1.52) *
Completeness (%)	99.9 (99.9) *
Redundancy	14.1 (13.1) *
Refinement	
Resolution (Å)	63.01-2.2
No. reflections	82385
R_{work} / R_{free}	0.177 / 0.202
No. atoms	
Protein	6199
Ligand/ion	271
Water	213
B-factors	
Protein	66.2
Ligand/ion	76.2
Water	54.8
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.181

*Values in parentheses are for highest-resolution shell.

Table S2 – pK_a values and Hill coefficients from titrations of EXG:CBM variants.

pK_a-values from CpHMD simulation and experiments.

	Asp39	Asp43	Asp61	His66
CpHMD/Experimental values				
EXG:CBM^{QQQW,66H}				6.10±0.09/6.39±0.06
EXG:CBM^{QQQW,39D}	5.33±0.04/~5b			
EXG:CBM^{QQQW,43D}		4.78±0.01/4.37±0.01		
EXG:CBM^{QQQW,61D}			4.49±0.01/4.15±0.01	
EXG:CBM^{QQQW,39D,66H}	3.69 ±0.04[a]/4.07±0.02			6.85±0.04[a]/7.24±0.03
EXG:CBM^{QQQW,43D,66H}		4.66±0.02/4.32±0.02		6.01±0.10/6.49±0.04
EXG:CBM^{QQQW,61D,66H}			4.19±0.02/4.1±0.02	6.07±0.04/6.68±0.02

[a]Macroscopic pK_a from the diprotic acid model (see Materials and Methods)

[b]This pK_a-value is an estimate based on stability data (See Figure S6).

Hill coefficients from CpHMD simulation and experiments.

	Asp39	**Asp43**	**Asp61**	**His66**
CpHMD/Experimental values				
EXG:CBM^{QQQW,66H}				1.0/0.98
EXG:CBM^{QQQW,39D}	1.01/-			
EXG:CBM^{QQQW,43D}		1.0/1.08		
EXG:CBM^{QQQW,61D}			0.96/0.98	
EXG:CBM^{QQQW,39D,66H}	-/-			-/-
EXG:CBM^{QQQW,43D,66H}		0.97/-		0.98/0.91
EXG:CBM^{QQQW,61D,66H}			0.94/-	0.99/1.1

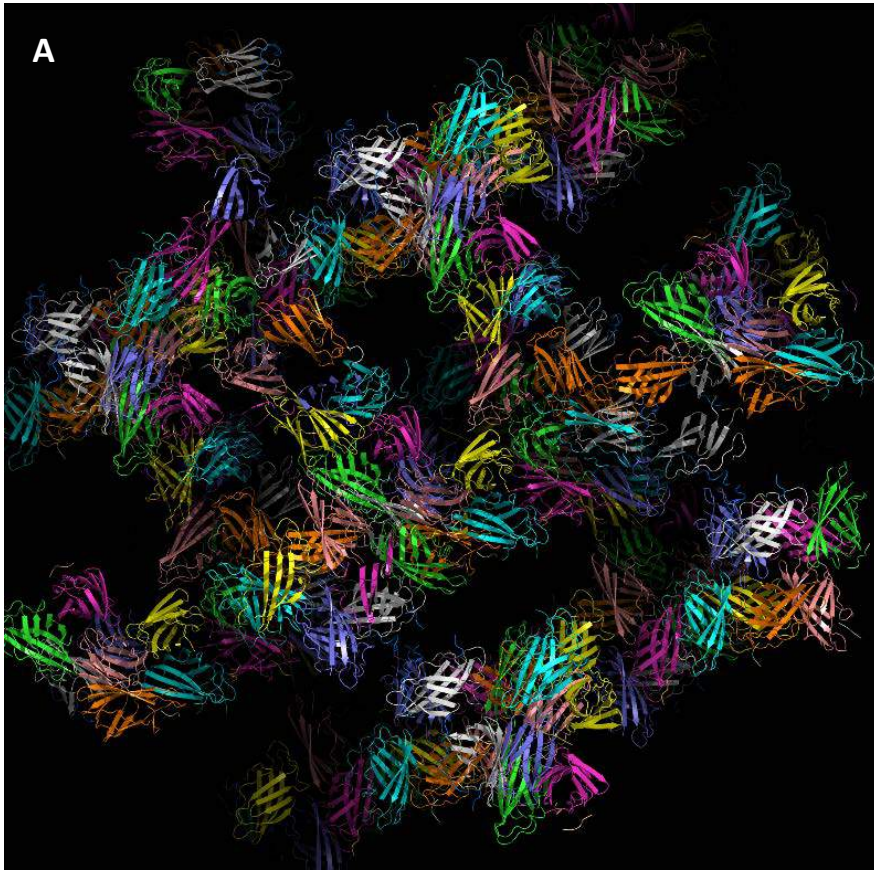
"-" Hill coefficient was not determined due to double titration in signal.

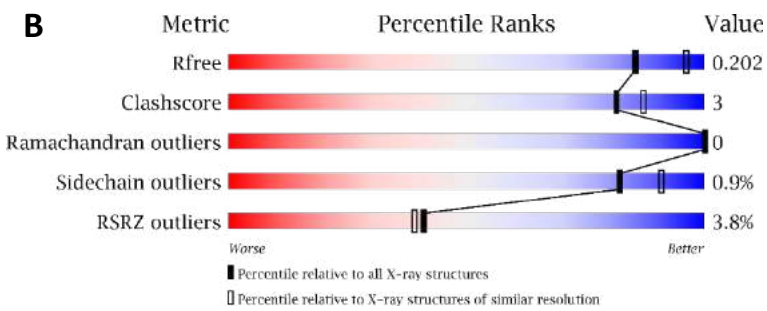
Table S3 – Stability parameters from urea unfolding

		m_{app} (kJ/mol)[a]	C_m (M)[a]	$\Delta G°_U$ (kJ/mol)[b]
EXG:CBM^{QQQW}	pH 2.5	4.24	5.01	26.4
	pH 5	4.79	5.16	27.1
	pH 8	5.22	4.95	26.0
	pH 10	5.17	4.89	25.7
	pH 5, 1.5M NaCl	4.47	6.33	33.3
	pH 8, 1.5M NaCl	4.64	6.43	33.8
EXG:CBM^{QQQW,66H}	pH 2.5	3.91	4.16	21.9
	pH 5	4.51	4.04	21.2
	pH 8	4.38	3.96	20.8
	pH 10	4.53	3.91	20.6
	pH 5, 1.5M NaCl	4.77	5.56	29.2
	pH 8, 1.5M NaCl	7.46	5.64	29.7
EXG:CBM^{QQQW,39D}	pH 2.5	5.58	4.15	21.9
	pH 5	4.95	3.56	18.7
	pH 8	6.20	2.85	15.0
	pH 10	6.75	2.77	14.6
	pH 5, 1.5M NaCl	5.13	4.47	23.5
	pH 8, 1.5M NaCl	4.87	4.20	22.1
EXG:CBM^{QQQW,43D}	pH 2.5	5.54	4.81	25.3
	pH 5	5.06	4.66	24.5
	pH 8	5.38	4.54	23.9
	pH 5, 1.5M NaCl	4.37	6.26	32.9
EXG:CBM^{QQQW,61D}	pH 2.5	5.40	5.11	26.9
	pH 5	5.16	4.96	26.1
	pH 8	4.77	4.83	25.4
EXG:CBM^{QQQW,66H,39D}	pH 2.5	5.21	3.25	17.1
	pH 5	5.67	3.42	18.0
	pH 8	5.87	2.75	14.4
	pH 10	5.77	2.50	13.1
	pH 5, 1.5M NaCl	5.22	4.69	24.6
	pH 8, 1.5M NaCl	4.35	4.23	22.2
EXG:CBM^{QQQW,66H,43D}	pH 2.5	6.35	3.97	20.9
	pH 5	5.43	3.99	21.0
	pH 8	5.43	3.93	20.7
	pH 5, 1.5M NaCl	4.20	5.26	27.7
EXG:CBM^{QQQW,66H,61D}	pH 2.5	5.32	4.06	21.4
	pH 5	6.08	3.90	20.5
	pH 8	5.07	4.01	21.1

[a]Parameters from fitting the data in Figure S4 to the linear extrapolation method
[b]Calculated from $\Delta G_U = C_M \cdot \langle m_{app} \rangle$

Figure S1 – Crystal structure of EXG:CBM$^{QQQW,\Delta 2\text{-}5}$. (A) Crystal packing in the solved structure. (B) Validation summary from wwPDB comparing the solved structure to other deposed structures. C) Alignment of the eight chains in the asymmetric unit.





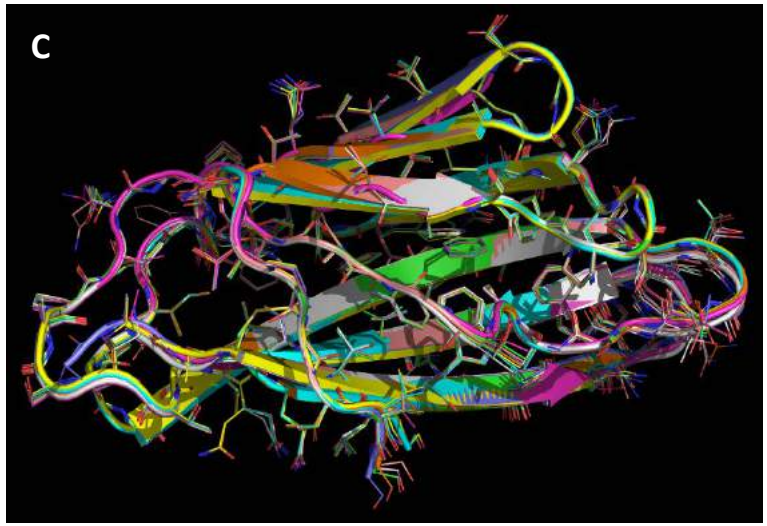


Figure S2 - pH titration followed by NMR chemical shifts. Selected chemical shifts as function of pH. Histidine and C-terminus titrations could be followed entirely by ^{13}C-chemical shifts, while the $^{13}C^{\gamma}$-shifts of Asp often disappeared at low pH. Therefore, Asp pK_a-values were fitted globally from $^{13}C^{\gamma}$ and amide signals. Color legend from Asp C^{γ} titration curve also applies to Asp H and Asp N nuclei.

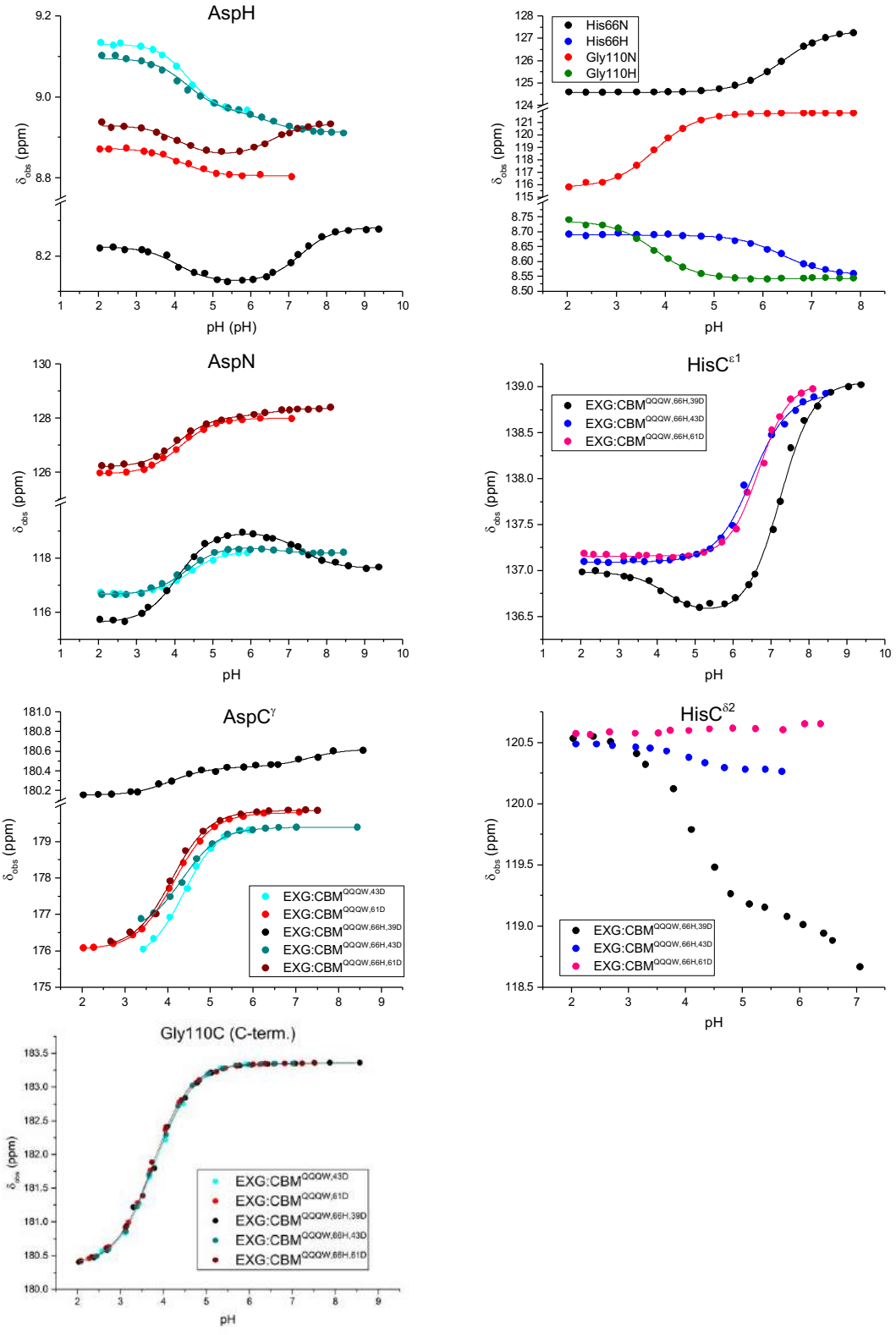


Figure S3 – pK_a prediction from the protonation state sampling by CpHMD. (A) Titration curves for the three single aspartate EXG variants. The degree of proton dissociation at various pH was fitted to the Hill equations. (B) Titration curve of (**top**) the aspartate residues (D39, D43, and D61) in the double titratable EXG variants fitted to the Henderson-Hasselbalch equation (solid line. Titration curve of (**bottom**) of the histidine residue (His66) in the double titratable EXG variants fitted to the Henderson-Hasselbalch equation (solid line). For Asp39 and His66 in the presence of Asp39 the data cannot be fitted to the Henderson-Hasselbalch equation to obtain pK_a values and Hill coefficients.

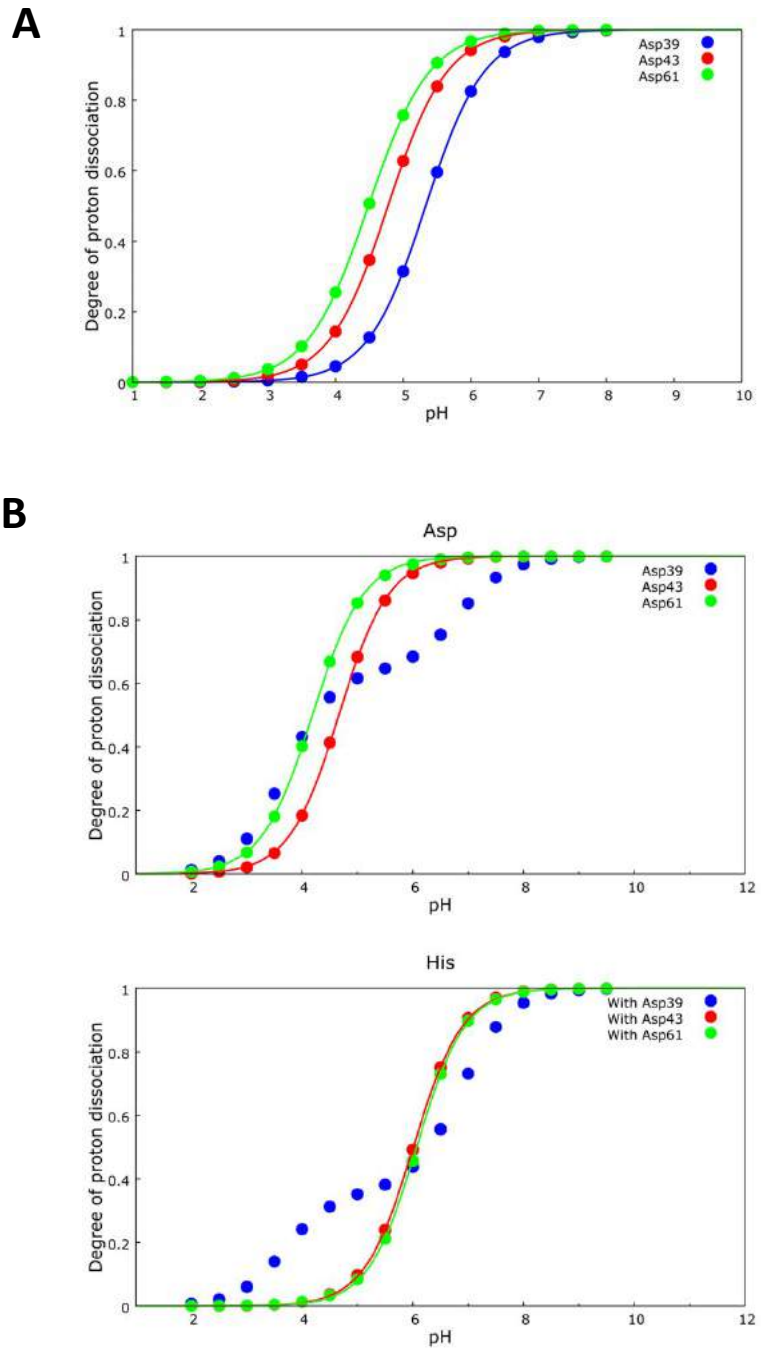
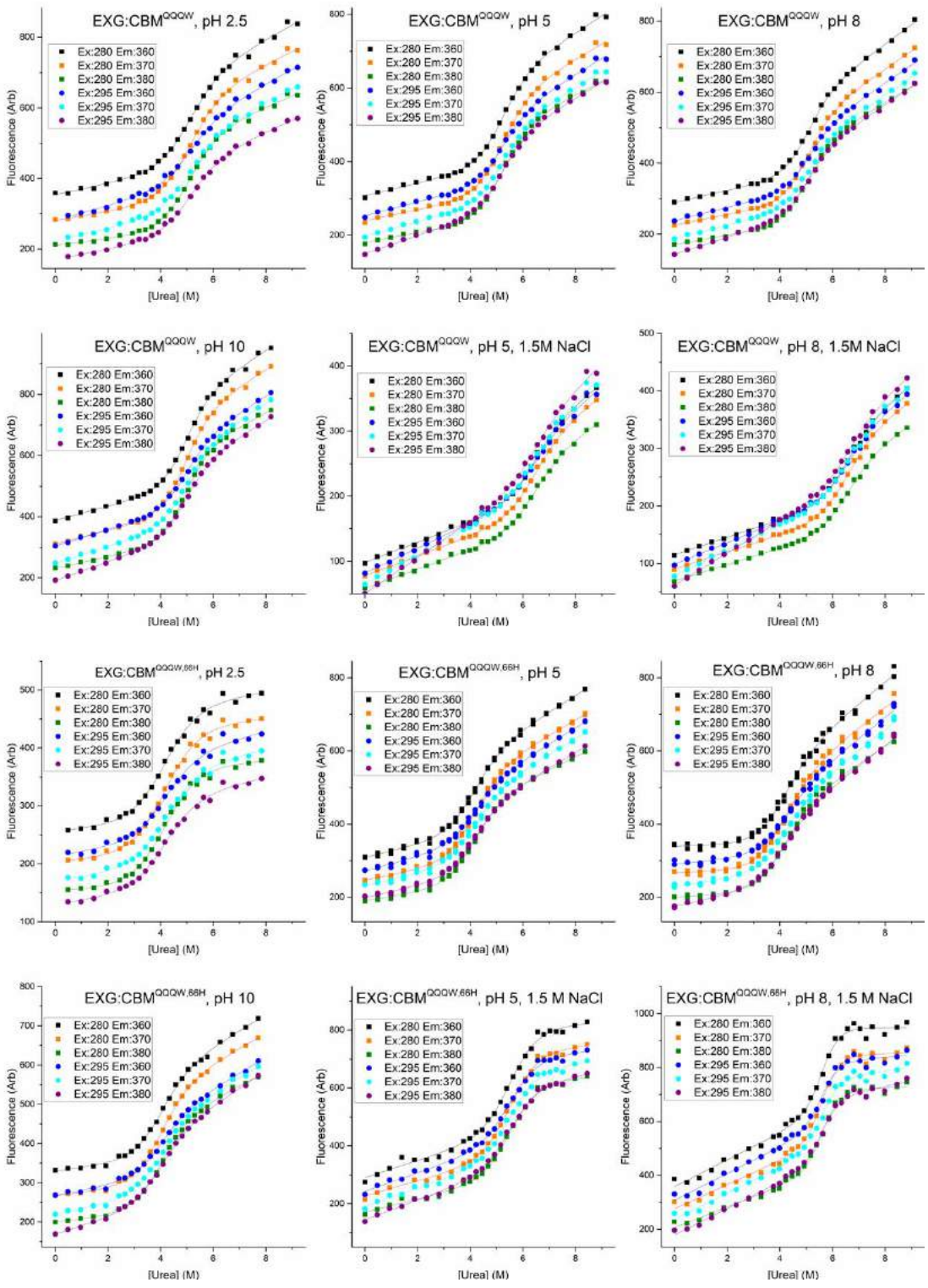
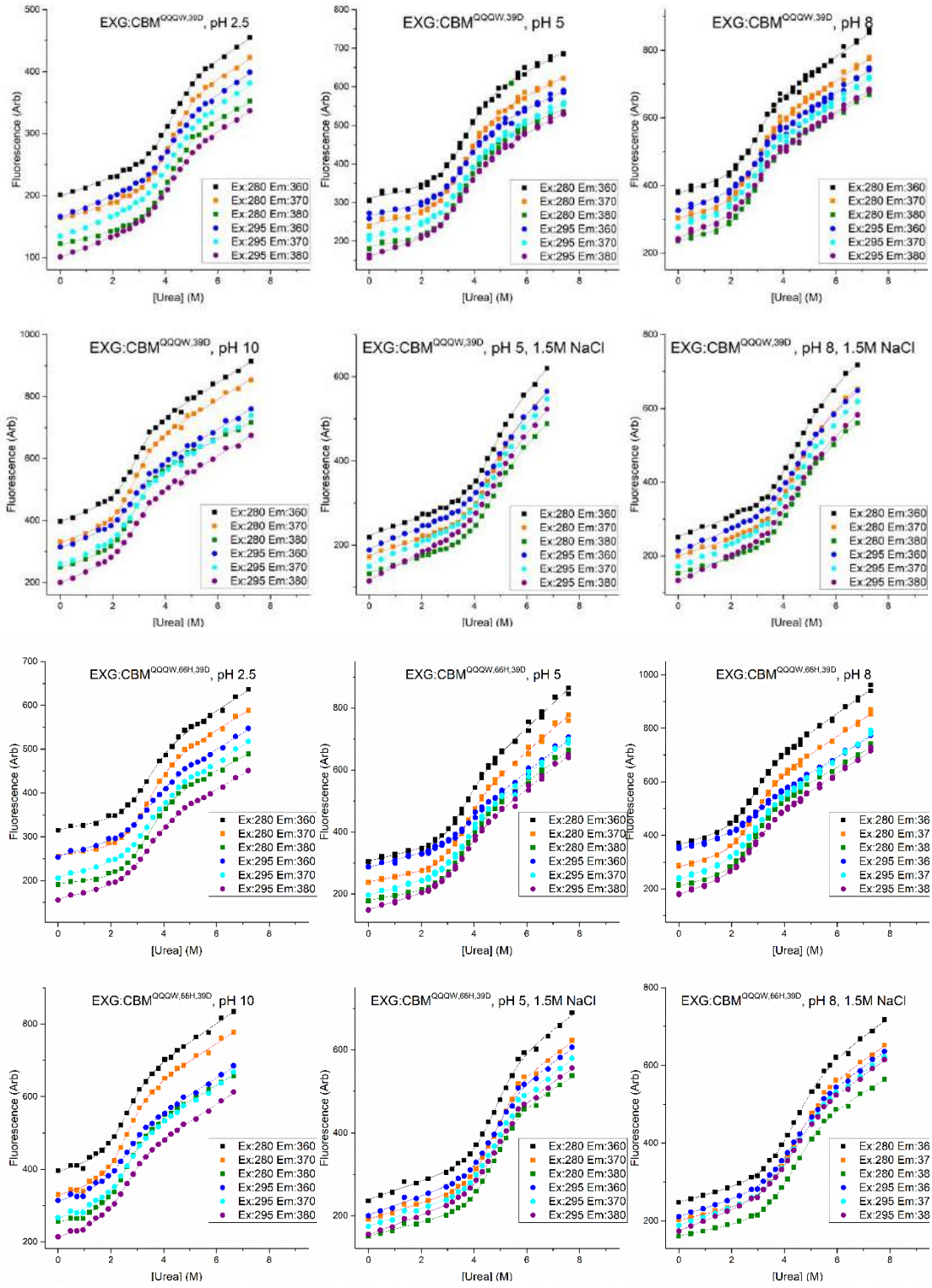


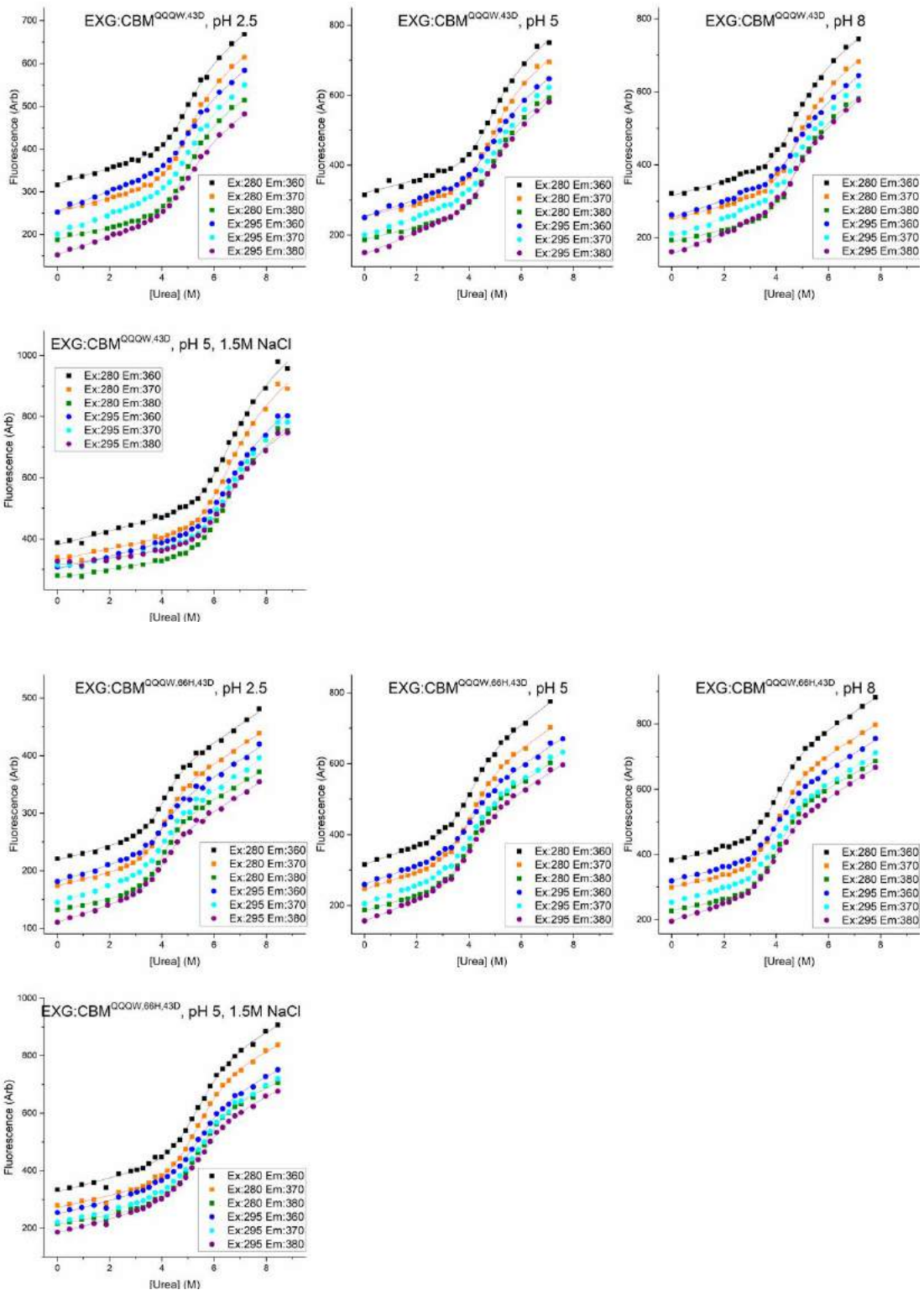
Figure S4 – Urea induced unfolding data. Raw data from fluorescence measurements. Lines represent global fits to the linear extrapolation method, using the following equation:

$$y = \frac{(y_f + m_f\,[\mathrm{urea}]) + (y_u + m_u\,[\mathrm{urea}]) \cdot \exp -(\frac{\Delta G^\circ(\mathrm{H_2O}) - m[\mathrm{urea}]}{RT})}{1 + \exp -(\frac{\Delta G^\circ(\mathrm{H_2O}) - m[\mathrm{urea}]}{RT})}$$

where y_f and y_u are the intercepts and m_f and y_u the slopes of the pre- and post-transition baselines. $\Delta G^\circ(\mathrm{H_2O})$ and m are the standard free energy of folding in water and linear dependency if that parameter with denaturant concentration.


EXG:CBM^QQQW, pH 2.5
EXG:CBM^QQQW, pH 5
EXG:CBM^QQQW, pH 8
EXG:CBM^QQQW, pH 10
EXG:CBM^QQQW, pH 5, 1.5M NaCl
EXG:CBM^QQQW, pH 8, 1.5M NaCl
EXG:CBM^QQQW,66H, pH 2.5
EXG:CBM^QQQW,66H, pH 5
EXG:CBM^QQQW,66H, pH 8
EXG:CBM^QQQW,66H, pH 10
EXG:CBM^QQQW,66H, pH 5, 1.5 M NaCl
EXG:CBM^QQQW,66H, pH 8, 1.5 M NaCl
Ex:280 Em:360
Ex:280 Em:370
Ex:280 Em:380
Ex:295 Em:360
Ex:295 Em:370
Ex:295 Em:380
Fluorescence (Arb)
[Urea] (M)


EXG:CBM^QQQW,39D, pH 2.5
EXG:CBM^QQQW,39D, pH 5
EXG:CBM^QQQW,39D, pH 8
EXG:CBM^QQQW,39D, pH 10
EXG:CBM^QQQW,39D, pH 5, 1.5M NaCl
EXG:CBM^QQQW,39D, pH 8, 1.5M NaCl
EXG:CBM^QQQW,66H,39D, pH 2.5
EXG:CBM^QQQW,66H,39D, pH 5
EXG:CBM^QQQW,66H,39D, pH 8
EXG:CBM^QQQW,66H,39D, pH 10
EXG:CBM^QQQW,66H,39D, pH 5, 1.5M NaCl
EXG:CBM^QQQW,66H,39D, pH 8, 1.5M NaCl
Fluorescence (Arb)
[Urea] (M)
Ex:280 Em:360
Ex:280 Em:370
Ex:280 Em:380
Ex:295 Em:360
Ex:295 Em:370
Ex:295 Em:380
Ex:280 Em:36
Ex:280 Em:37
Ex:280 Em:38
Ex:295 Em:36
Ex:295 Em:37
Ex:295 Em:38


EXG:CBM^QQQW,43D, pH 2.5
EXG:CBM^QQQW,43D, pH 5
EXG:CBM^QQQW,43D, pH 8
EXG:CBM^QQQW,43D, pH 5, 1.5M NaCl
EXG:CBM^QQQW,66H,43D, pH 2.5
EXG:CBM^QQQW,66H,43D, pH 5
EXG:CBM^QQQW,66H,43D, pH 8
EXG:CBM^QQQW,66H,43D, pH 5, 1.5M NaCl
Fluorescence (Arb)
[Urea] (M)
Ex:280 Em:360
Ex:280 Em:370
Ex:280 Em:380
Ex:295 Em:360
Ex:295 Em:370
Ex:295 Em:380

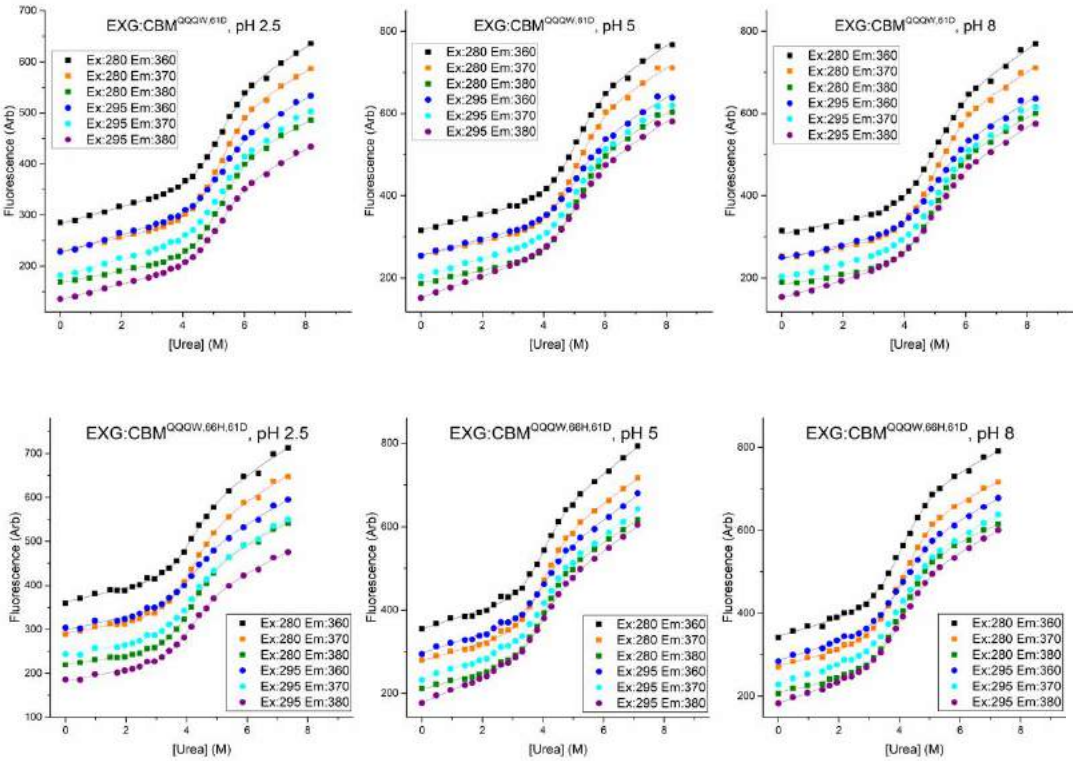

EXG:CBM^{QQQW,61D}, pH 2.5
EXG:CBM^{QQQW,61D}, pH 5
EXG:CBM^{QQQW,61D}, pH 8
EXG:CBM^{QQQW,66H,61D}, pH 2.5
EXG:CBM^{QQQW,66H,61D}, pH 5
EXG:CBM^{QQQW,66H,61D}, pH 8
Ex:280 Em:360
Ex:280 Em:370
Ex:280 Em:380
Ex:295 Em:360
Ex:295 Em:370
Ex:295 Em:380
Fluorescence (Arb)
[Urea] (M)

Figure S5a – ^{15}N-HSQC spectra of the investigated EXG:CBM variants. The colors correspond to EXG:CBM^{QQQW,43D} (green), EXG:CBM^{QQQW,61D} (orange), EXG:CBM^{QQQW,66H,39D} (red), EXG:CBM^{QQQW,66H,43D} (blue), and EXG:CBM^{QQQW,66H,61D} (black). The spectra were measured at pH ~4.

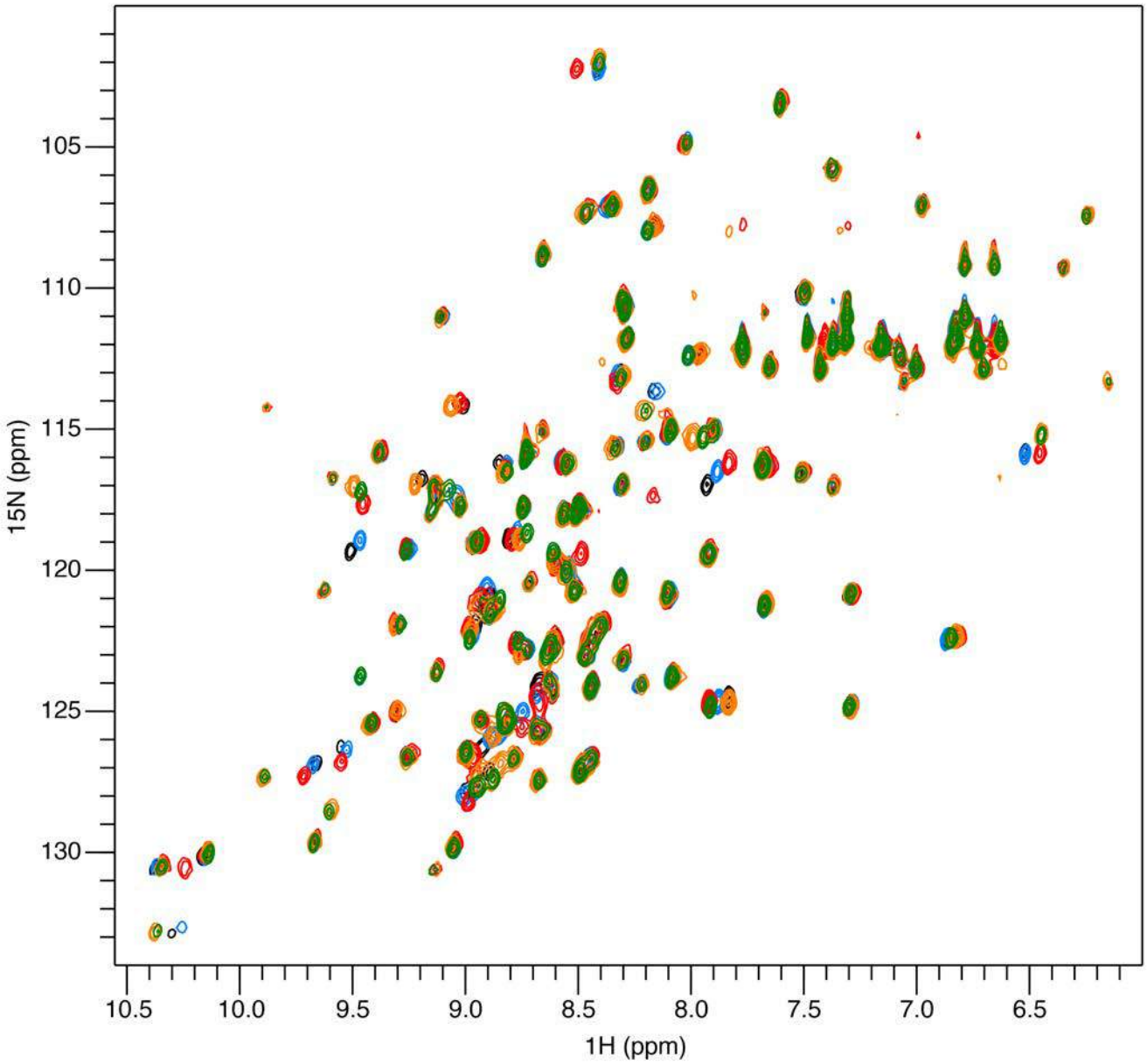


Figure S5b – Chemical shift perturbations for weighted backbone amide chemical shifts. The perturbations are calculated relative to EXG:CBM^{QQQW,66H} and shown with the same color code as S5a.

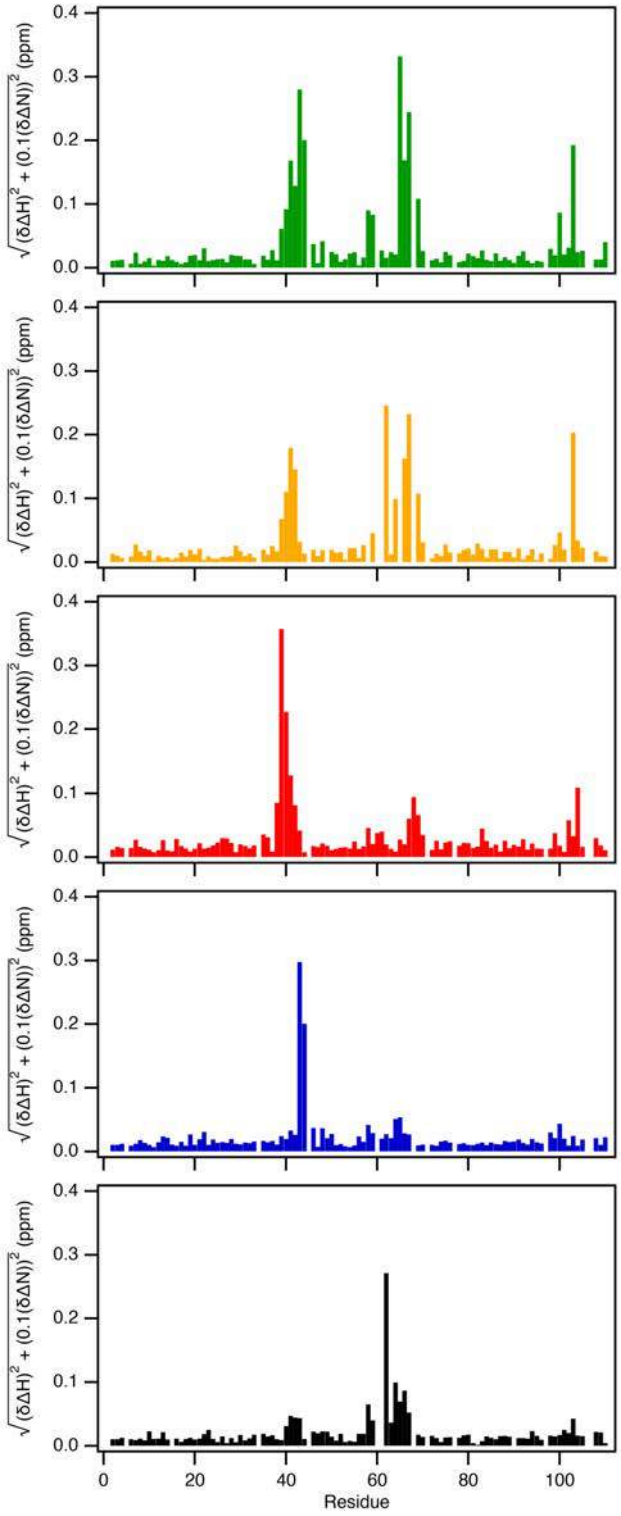


Figure S6. Double-mutant cycles.

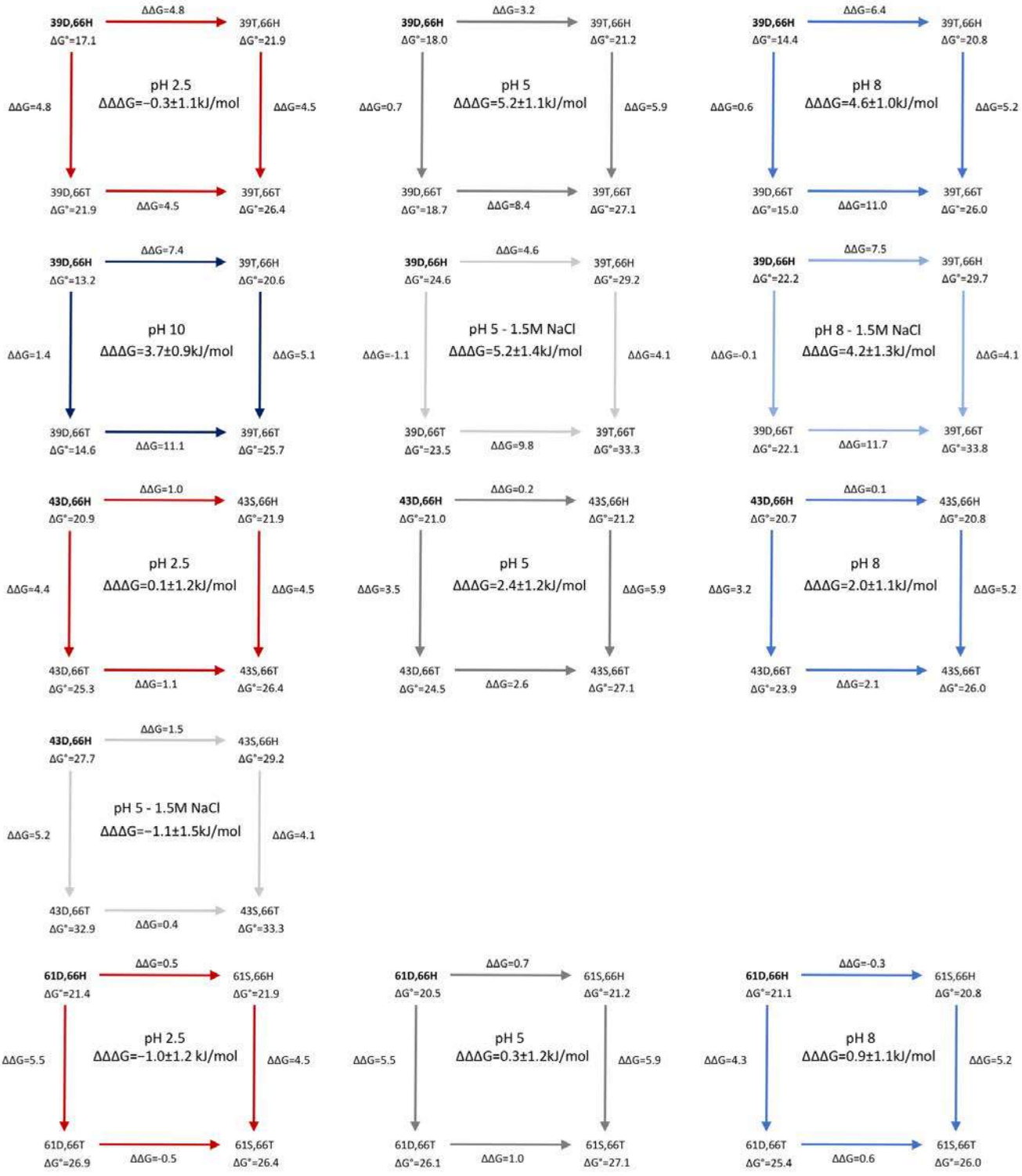


Figure S7 – Rotamer population of Asp and His in double titratable EXG:CBM variants. Histograms of the sampled χ_1 dihedrals of Asp and His residues in the EXG:CBM^{QQQW,39D,66H} (top row), EXG:CBM^{QQQW,43D,66H} (middle row) EXG:CBM^{QQQW,61D,66H} (bottom row) from CpHMD at pH 2.0, 5.0 and 9.0.

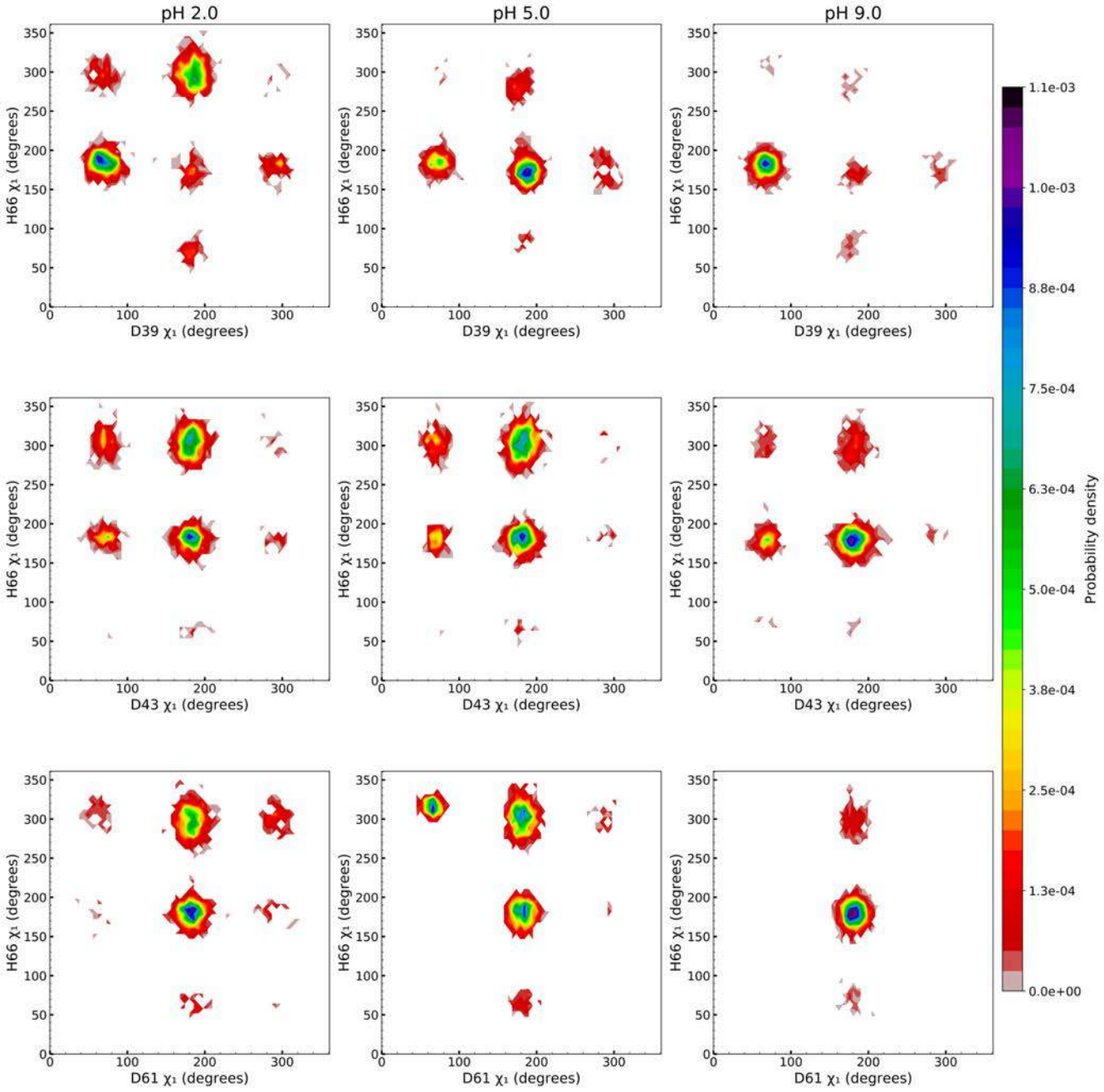


Figure S8 – Relative orientation of Asp and His rotamers sampled by CpHMD. Histogram of the sampled χ_1 dihedrals of Asp and His residues in the EXG:CBM^{QQQW,39D,66H} at pH 2.0 with structures indicating the relative orientation of the two residues compared to the each other, the remaining protein, and solvent. Note the aspartate and histidine are illustrated with additional protons associated with each of their titratable sites; in total two for histidine and four for aspartate (one *anti* and one *syn* protonation state for each oxygen in the carboxylic acid side chain).

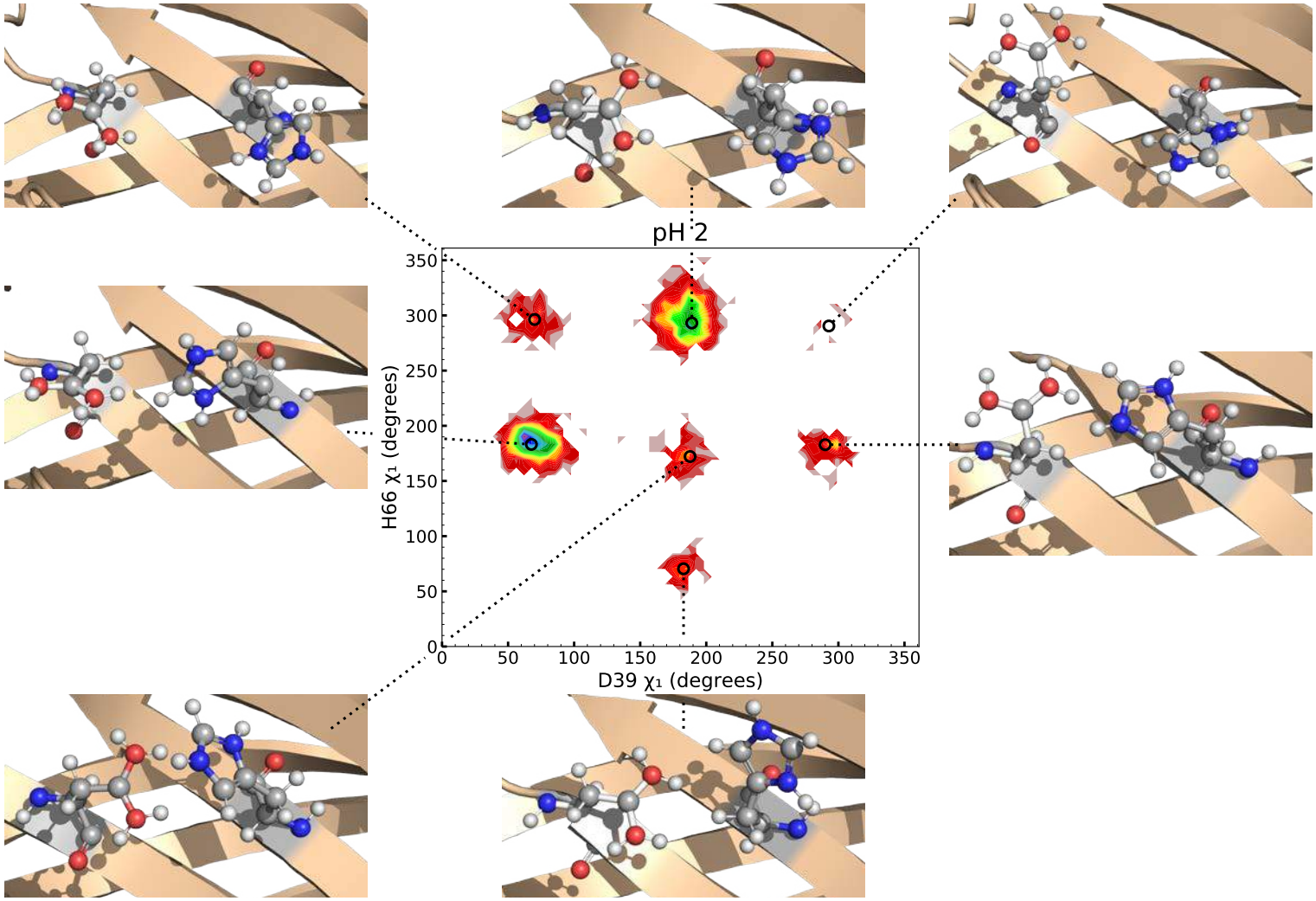


Figure S9 – Rotamer population of Asp and His in double titratable EXG:CBM variants. Histograms of the sampled χ_1 dihedrals of Asp in EXG:CBM^{QQQW,39D} (left), EXG:CBM^{QQQW,43D} (middle) and, EXG:CBM^{QQQW,61D} (right) from CpHMD at pH 0.5 (blue) and 8.0 (orange).

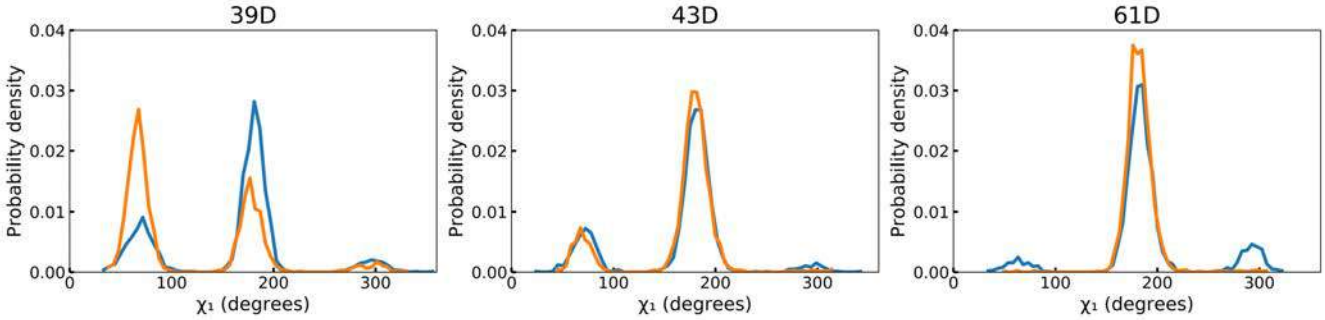
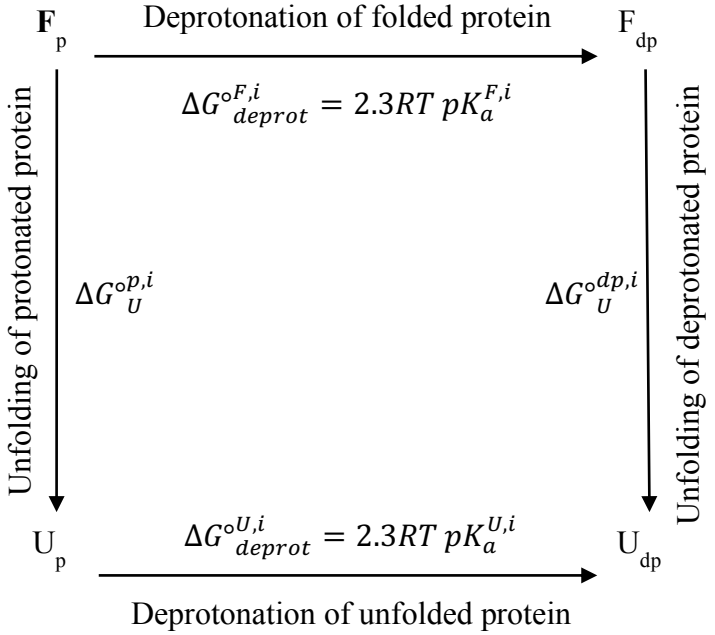


Figure S10 – Thermodynamic cycle for estimating pK$_a$ shifts from stability data. Thermodynamic cycle relating the energy of deprotonation the residue *i* (ΔG^{i}_{deprot}) in the folded (F) and unfolded (U) state of a protein, to the residues contribution to protein stability (ΔG^{i}_{U}) at pH where the residue is fully protonated (p) and fully deprotonated (dp) respectively.



From Figure S10 it follows that:

$$\Delta G^{\circ F,i}_{deprot} + \Delta G^{\circ dp,i}_{U} = \Delta G^{\circ p,i}_{U} + \Delta G^{\circ U,i}_{deprot} \quad \text{Eq. S1}$$

$$\Delta G^{\circ dp,i}_{U} - \Delta G^{\circ p,i}_{U} = 2.3RT\left(pK_a^{U,i} - pK_a^{F,i}\right) \quad \text{Eq. S2}$$

$$\Delta pK_a^{U-F,i} = (\Delta G^{\circ dp,i}_{U} - \Delta G^{\circ p,i}_{U})/2.3RT \quad \text{Eq. S3}$$

In a protein with multiple charged residues, the total change in stability, $\Delta\Delta G^{\circ}{}_{U}$, integrated over the entire pH range is thus related to difference in pK$_a$-values between folded and unfolded state of all its residues (Bosshard et al, 2004):

$$\Delta\Delta G^{\circ}{}_{U} = \Delta G^{\circ dp}_{U} - \Delta G^{\circ p}_{U} = 2.3RT \sum_{i=1}^{n} (pK_a^{U,i} - pK_a^{F,i}) \quad \text{Eq. S4}$$

The total change in protein stability between two pH is determined by the sum of the contributions from all charged residues (Eq. S4). In EXG:CBM^{QQQW,39D}, there are only three potential charges: Asp39 and the C- and N-terminus. The effects of the C- and N-termini on $\Delta\Delta G^{\circ}{}_{U}$ can be calculated from $\Delta G^{\circ dp}_{U} - \Delta G^{\circ p}_{U}$ in EXG:CBM^{QQQW}. Subtracting the found energy difference from the $\Delta\Delta G^{\circ}{}_{U}$ of EXG:CBM^{QQQW,39D} leaves the contribution of Asp39, assuming the contribution of the C- and N-termini are the same in the two proteins. This energy was used in Eq. S3 to estimate the pK$_a$ difference of Asp39 in the folded and unfolded state. Stability data from pH 2 (Asp39 fully protonated) and pH 8 (Asp39 fully deprotonated) were used for the calculation.

Figure S11 – Static structure calculation using PROPKA3 on two populated conformations extracted from the pH-REMD simulation on the EXG39D,66H variant at pH 5.0. This shows the strong rotamer dependence of the PROPKA3 predictions.