Quantitative prediction of pH-controlled ionization of oligoampholytes

Raju Lunkad, Anastasiia Murmiliuk, Pascal Hebbeker, Milan Boublík, Zdeněk Tošner, Miroslav Štěpánek, and Peter Košovan*

Department of Physisal and Macromolecular Chemistry, Charles University, Hlavova 8, 128 43 Prague, Czech Republic

E-mail: peter.kosovan@natur.cuni.cz

June 24, 2020

Abstract

Weak ampholytes are ubiquitous in nature and commonly found in artificial pH-responsive systems. However, our limited understanding of their ionisation response and the lack of predictive capabilities hinder the bottom-up design of such systems. Here, we used a coarse-grained model of a flexible polymer with weakly ionisable monomer units to quantitatively analyse the ionisation behaviour of two oligopeptides. Differences in ionisation response between oligopeptides and monomeric amino acids showed that electrostatic interactions between weak acid and base side chains play a key role in oligopeptide ionisation, as predicted by our model. Moreover, by comparing our simulations with experimental results from potentiometric titration, capillary zone electrophoresis and NMR, we demonstrated that our model reliably predicts the ionisation response and electrophoretic mobilities of various peptide sequences. Ultimately, our model is the first step towards using predictive bottom-up design of responsive ampholytes to tailor their properties as a function of charge and pH.

Introduction

The ionisation behaviour of molecules with multiple titratable groups is very different from that of low-molecular weak acids or bases. For acids with few titratable groups (e.g., oxalic acid or phosphoric acid), different pK_A values can be assigned to each ionisation state because the titration curve contains several distinct inflection points. In contrast, titration curves of synthetic polymers, such as poly(acrylic) acid, vary smoothly across a broad range of pH, thus making it impossible to assign a specific pK_A to each state. A molecule with n different ionisable groups has 2^n distinct ionisation states. At a rather small n, this number of states becomes too large. In addition, electrostatic interactions between ionised groups affect the overall ionisation response in a complicated manner.[1–6] The general notion is that repulsion between like-charged groups suppresses ionisation, while attraction between oppositely charged groups enhances ionisation. The net result depends on the pK_A of each individual group and on their distribution in space. This electrostatic effect on the ionisation should be distinguished from the effect of replacing local substituents upon formation of the peptide bond, which conditions the pK_A by affecting the electron density in neighbouring chemical bonds.

Changes in pH can be used to control enzyme activity or protein aggregation,[7,8] to trigger the release of anti-cancer drugs[9] or to control protein sequestration in polyelectrolyte brushes, gels and complexes,[10–12] as shown by the rapid development of pH-responsive materials, and their applications.[13] However, the ionisation behaviour of these systems remains poorly understood and therefore is often interpreted only qualitatively, based on the pK_A of their parent monomers. Accordingly, quantitative studies or predictions are either very scarce or nonexistent.

Computational models based on the Poisson-Boltzmann approach have explained the effect of these interactions on charge regulation in various proteins and peptides with rigid secondary and tertiary structures.[14–19] However, their conformational flexibility has been neglected because it does not significantly affect the ionisation of these rigid proteins and

peptides. In addition, their ionisation has been rarely studied in a broad range of pH, far from the isoelectric point, because it is usually not relevant for their biological function. Nevertheless, recent advances in the synthesis of polyampholytes, polyzwitterions[20–23] and polyampholyte gels[24,25] have underscored the need to further understand their behaviour in the whole range of pH and ionisation states, particularly for peptide-based pH-responsive materials[26,27] and for therapeutic peptides used in biomedical applications.[28]

Computer simulations using coarse-grained models can account for ionisation-conformation coupling, which plays a key role in long flexible polymers. Similar models have helped us understand the ionisation of synthetic polymers.[3,29–32] Yet, despite the simplicity of these models, comparisons with experimental results have been surprisingly quantitative,[29,33–35] thus suggesting that they correctly capture the most important effects. The challenge is to assess whether such simulations, using suitably designed coarse-grained models, are able to predict the ionisation of molecules more complex than long homopolymer chains.

To answer this question, we used two short oligopeptides with weak acid and weak base side-chains as a model system with a complex ionisation response in which acid and base groups mutually influence each other. The importance of coupling charge distribution and conformation in intrinsically disordered proteins has been previously demonstrated in both simulations and experiments,[36–39] albeit disregarding their ionisation response. Conversely, our model considers the ionisation response of short oligopeptides and its relation to molecular geometry and conformation, presumably for the first time. By using similar but simpler peptides as a model system, we can easily design ampholytes with a well-defined sequence and with sufficient purity, which allows us to choose suitable combinations of $\mathrm{p}K_\mathrm{A}$ values.[40] Ultimately, modeling short peptides paves the way towards more complex biomolecules, such as intrinsically disordered proteins, and engineering pH-responsive materials, including therapeutic peptides, for a wide range of applications.[28,36,40,41]

The ionisation response of acid and base groups can be described by the following reac-

tions:

$$\mathrm{HA} \rightleftharpoons \mathrm{A}^- + \mathrm{H}^+ \tag{1}$$

$$\mathrm{BH}^+ \rightleftharpoons \mathrm{B} + \mathrm{H}^+ \tag{2}$$

where A and B stands for the weak acid or base group, while A^- and BH^+ represent their ionised forms. In the ideal case (in the absence of interactions), the degree of ionisation, α, of each ionisable group depends only on the difference $\mathrm{pH} - \mathrm{p}K_\mathrm{A}$ via the Henderson-Hasselbalch equation (ideal titration curve):

$$\mathrm{pH} - \mathrm{p}K_\mathrm{A}^\mathrm{acid} = \log_{10} \frac{\alpha}{1-\alpha}\,, \qquad \mathrm{pH} - \mathrm{p}K_\mathrm{A}^\mathrm{base} = \log_{10} \frac{1-\alpha}{\alpha} \tag{3}$$

where $\mathrm{p}K_\mathrm{A}^\mathrm{acid}$ and $\mathrm{p}K_\mathrm{A}^\mathrm{base}$ are the acidity constants of the acid and base groups, respectively.[29] The total charge of the peptide at a given pH is then given by

$$z(\mathrm{pH}) = \sum_i n_i \alpha_i(\mathrm{pH}) z_i \tag{4}$$

where n_i is the number of groups of type i in the peptide, α_i is their average degree of ionisation, and $z_i = \pm 1$ is their charge in the ionised state. If $\Delta\mathrm{p}K_\mathrm{A} = \mathrm{p}K_\mathrm{A}^\mathrm{base} - \mathrm{p}K_\mathrm{A}^\mathrm{acid}$ is large, then the peptide responds to a change in pH by varying the charge of one group or the other. If the $\Delta\mathrm{p}K_\mathrm{A}$ is small, then the peptide responds to a change in pH by simultaneously varying the charge of both groups.

To assess the effect of $\Delta\mathrm{p}K_\mathrm{A}$, we composed model peptides consisting of two blocks: Five amino acids with weak acid side chains and five with weak base side chains (see Fig. 1). The $\mathrm{Glu}_5 - \mathrm{His}_5$ peptide has $\mathrm{p}K_\mathrm{A}^\mathrm{acid} = 4.25$, $\mathrm{p}K_\mathrm{A}^\mathrm{base} = 6.00$ and a small $\Delta\mathrm{p}K_\mathrm{A} = 1.75$. The $\mathrm{Lys}_5 - \mathrm{Asp}_5$ peptide has $\mathrm{p}K_\mathrm{A}^\mathrm{acid} = 3.65$, $\mathrm{p}K_\mathrm{A}^\mathrm{base} = 10.53$, and a large $\Delta\mathrm{p}K_\mathrm{A} = 6.88$. Both termini on each peptide were protected to ensure that the ionisation response is not affected by free carboxyl and amino groups. The reported $\mathrm{p}K_\mathrm{A}$ values correspond to side-chains of

free amino acids, with slightly different values reported in the literature.[42,43] Furthermore, these values can also change when the amino acid is incorporated into the peptide bond.

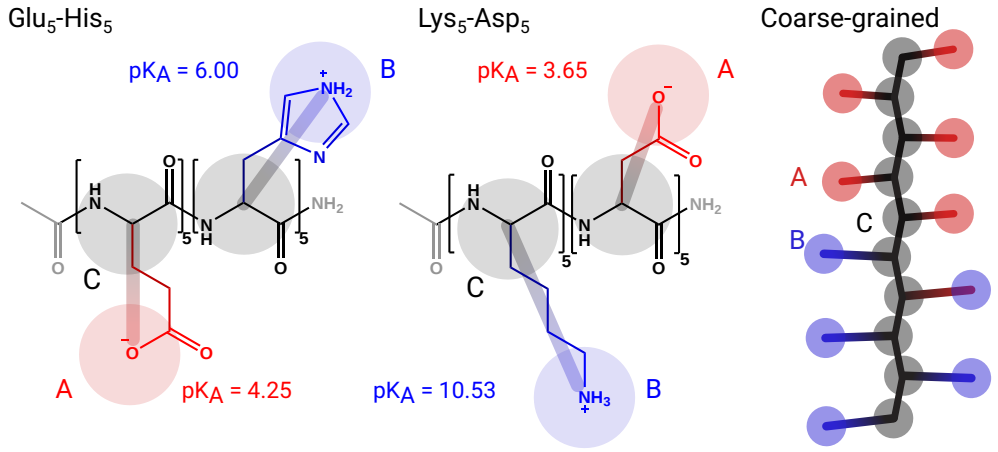


Figure 1: Chemical structure of the real peptides and a schematic representation of the coarse-grained bead-spring model consisting of A, B and C beads.

Results and discussion

To represent the peptides in a simulation, we created a simple bead-spring model, shown in Fig. 1, derived from the well established Kremer-Grest polymer model.[29,44] Each amino acid is represented by one central bead C for the backbone and one side-chain bead A or B for the acid or base group. All beads have the same size and are connected by harmonic springs. Distances between ionisable groups are crucial to quantify the ionisation response. Therefore, we used all-atom simulations to determine the distances r_{AC}, r_{BC}, r_{CC} in our model. The beads are neutral in the non-ionised state and carry an elementary charge in the ionised state ($\pm 1e$ for acid or base, respectively). To account for the ionisation response, we simulated the ionisation reactions (Eq. 1 and 2) in the constant-pH ensemble, using the pK_{A} of side-chains of the corresponding free amino acids.

The simulations predict a significantly lower charge of the model peptides, in comparison to the free amino acids, as shown in Fig. 2. The weak acid and weak base groups in our model have identical interaction parameters, except for pK_{A} and geometry (bond lengths

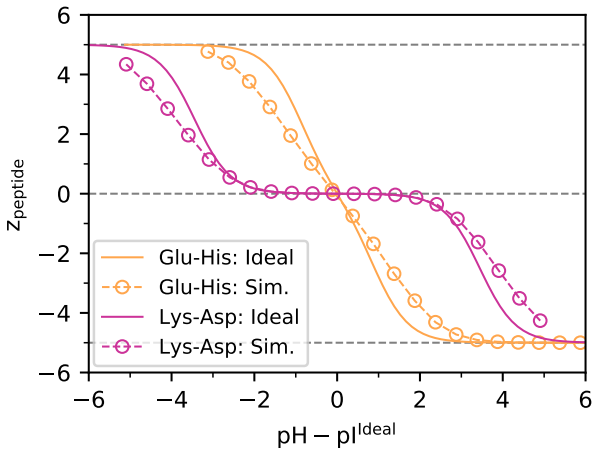


Figure 2: Simulation predictions of the total charge of the $Glu_5 - His_5$ and $Lys_5 - Asp_5$ peptides (data points), Compared with the respective ideal titration curves obtained from Eq. 3 using pK_A values of free amino acids (solid lines).

r_{AC}, r_{BC}). Therefore, the isoelectric point derived from simulations is very close to the ideal value $pI_{sim} \approx pI_{ideal} = (pK_A^{acid} + pK_A^{base})/2$. The two distinct inflection points in the ionisation of $Lys_5 - Asp_5$ indicate that the two blocks change their ionisation in different pH ranges because of the large ΔpK_A. Conversely, the ionisation of $Glu_5 - His_5$ lacks a clear inflection point and is nearly linear across a broad range of pH, indicating that the ionisation of the acid and base blocks varies simultaneously because the ΔpK_A is much smaller. Thus, the following question emerges: To what extent does such a simple simulation model predict the behaviour of real ampholytes if it accounts only for the pK_A values of the amino acids, disregarding all chemical complexity, which should be essential for biological functions?

To answer this question, we followed the charge of the peptides using several independent experiments: capillary zone electrophoresis (CZE), potentiometric titration and NMR. *In principle*, these measurements yield quantities proportional to the charge or degree of ionisation. In practice, each method has its own limitations and pitfalls, which hinder quantitative analysis.

In CZE, we directly measure the effective electrophoretic mobility

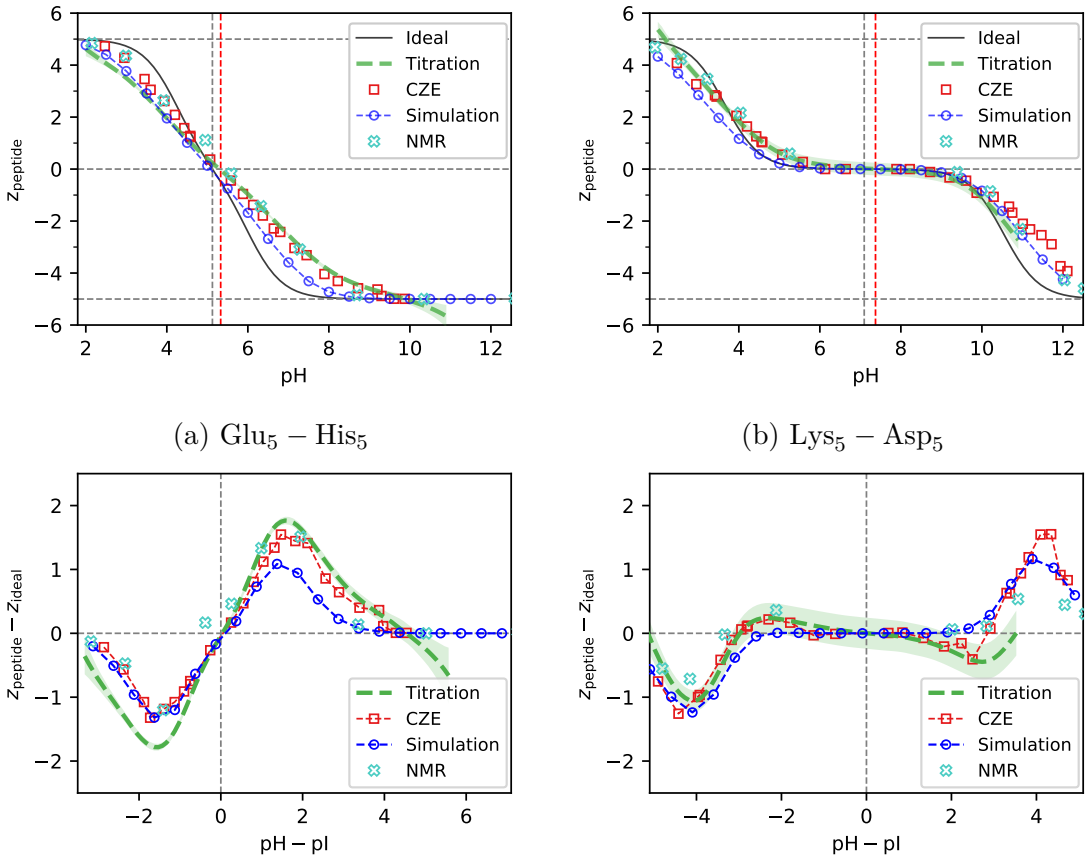
$$\mu(\mathrm{pH}) = \frac{v(\mathrm{pH})}{E} \approx \frac{ez(\mathrm{pH})}{6\pi\eta R_{\mathrm{H}}} \tag{5}$$

where v is the velocity of charged particle moving in an electric field E. The relation between μ, charge on the peptide z, its hydrodynamic radius, R_{H}, and solvent viscosity, η, is only approximate because electrophoretic mobility yields the effective charge, which is lower than the bare charge.[45–47] Nevertheless, CZE yields the exact isoelectric point, defined by $\mu(\mathrm{pI}) = 0$. We estimated the bare charge on the peptides by observing that R_{H} of both peptides was virtually independent of pH and then renormalising the mobility $\mu(\mathrm{pH})$ by the maximum absolute mobility $\mu^{\mathrm{max}}(\mathrm{pH}^{\mathrm{max}})$, and by the charge $z_{\mathrm{sim}}(\mathrm{pH}^{\mathrm{max}})$ determined from the simulations (see ESI, Section 2.2.2).

To complement the CZE with an independent experiment, we performed potentiometric titration of peptides in 10 mM HCl titrated by 10 mM NaOH. Using the electroneutrality condition, we determined the charge on the peptide from the known volume of HCl and amount of added NaOH and from the measured pH. *In principle*, the potentiometric titration should yield the absolute value of z. In practice, each peptide sample contained a slight excess of TFA counterions that shifted the apparent $z(\mathrm{pH})$ to lower values. To correct for the unknown amount of TFA, we used it as an adjustable parameter to match the pI determined from CZE (see ESI, Section 2.3.2).

Lastly, we used the NMR chemical shifts to determine the degree of ionisation of each type of amino acid, averaged over amino acids of the same type in the peptide, and calculated charge on the peptide using Eq. 4 (see ESI, Section 2.4.2). Unlike the other two methods, determination of the degree of ionisation from NMR did not require additional independent inputs.

Despite the simplicity of the model, the simulations agree with all experiments (NMR, CZE and potentiometric titration) in a broad range of pH, as shown in Fig. 3(a,b). All ap-



(a) $Glu_5 - His_5$

(b) $Lys_5 - Asp_5$

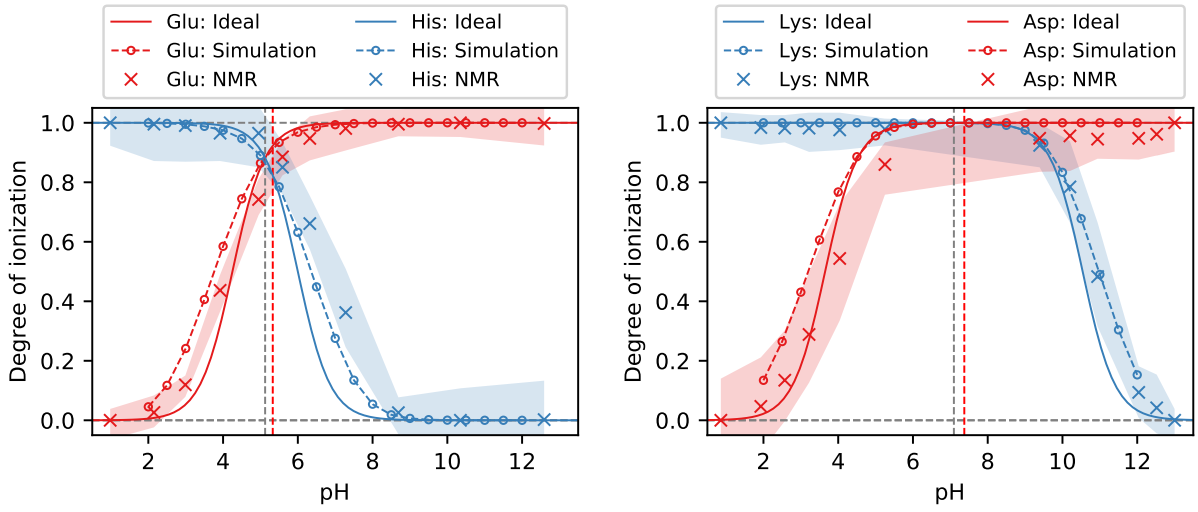
(c) $Glu_5 - His_5$

(d) $Lys_5 - Asp_5$

Figure 3: Panels (a) and (b): Total charge on the peptides as a function of pH from ideal titration curve, simulations, NMR, potentiometric titration and capillary zone electrophoresis (CZE). The gray vertical line indicates the ideal isoelectric point whereas the red vertical line indicates the isoelectric point determined from CZE. Panels (c) and (d): Differences between the determined charge on the peptide and the ideal titration curve as a function of pH − pI.

proaches yield titration curves with similar shapes, clearly deviating from the ideal titration curve. To better visualise these deviations, we plotted the difference from the ideal curve as a function of pH − pI, using pI_{ideal} for the ideal curve, $pI_{sim} \approx pI_{ideal}$ for the simulations, and pI_{cze} for all experimental data. The use of different pI values eliminates the uncertainty in choosing the right pK_A values as input parameters for the ideal titration curve and for the simulations. Generally, the NMR results almost coincide with CZE. For $Glu_5 - His_5$ the simulations quantitatively match CZE at pH < pI, while the titration data deviate slightly more from the ideal curve, as shown in Fig. 3c. At pH > pI, the difference between

simulation and CZE is slightly larger than between CZE and titration. Nevertheless, the deviations of all three curves from the ideal curve have the same trend and a similar magnitude. Eventually, at high pH, the titration yields nonphysical values of $z_{\text{titration}} < -|z^{\max}|$, caused by the numerical instability of the calculation of $z_{\text{titration}}(\text{pH})$ using Eq. 8 due to its sensitivity to temperature and to the precision of the pH measurement, as indicated by the broad error band. For $\text{Lys}_5 - \text{Asp}_5$, the simulation results very closely follow the CZE in the whole pH range, except pH $\approx$ pI, as shown in Fig. 3d. Around pI, both CZE and titration exhibit an undulation that is not captured by the simulation. The titration agrees well with both simulation and CZE at pH < 10 but then follows a different trend from the other two methods. This difference can also be ascribed to numerical difficulties in calculating $z_{\text{titration}}(\text{pH} \gtrsim 10)$ using Eq. 8. Considering the whole pH range, we conclude that, despite some minor discrepancies, all three methods consistently depict the ionisation response of both peptides. This demonstrates that the dominant contribution to non-ideal ionisation of weak ampholytes originates from electrostatic interactions.



(a) $\text{Glu}_5 - \text{His}_5$ (b) $\text{Lys}_5 - \text{Asp}_5$

Figure 4: The degree of ionisation of acid and base groups on the $\text{Glu}_5 - \text{His}_5$ and $\text{Lys}_5 - \text{Asp}_5$ peptides from simulations, NMR measurements and ideal titration curves. Red and grey vertical lines represent the isoelectric point determined from CZE and from the simulations. Shaded areas indicate the spread of 5 NMR signals corresponding to 5 amino acids of the same type.

It is tempting to explain the lower charge on the peptide in Fig. 2 by the like-charge repulsion within the weak acid and weak base block, in analogy to the ionisation response of polyacid or polybase homopolymers.[48] However, a closer look at ionisation of acid and base blocks in Fig. 4 suggests otherwise. In both cases, the peptide is fully ionised at the isoelectric point, although its total charge is zero. In $\mathrm{Lys}_5 - \mathrm{Asp}_5$ (Fig. 4b), ionisation within each block is higher than the expected ideal behaviour. The fully ionised oppositely charged block enhances the ionisation and overrides the effect of like-charge repulsion within the partly ionised block. Thus, even though the ionisation of each block of $\mathrm{Lys}_5 - \mathrm{Asp}_5$ varies in a different pH range, the two blocks strongly affect each other. In $\mathrm{Glu}_5 - \mathrm{His}_5$ (Fig. 4a), ionisation within the base block is suppressed at pH $<$ pI but enhanced at pH $>$ pI, and the opposite holds true for the acid block. This demonstrates that the small $\Delta \mathrm{p}K_\mathrm{A}$ enables both blocks to interact strongly, whilst simultaneously affecting each other's ionisation. The simulated average degree of ionisation of individual amino acids in Fig. 4 matches the degree of ionisation determined by NMR. Generally, the NMR curves adopt the same shape as the simulated curves, but some of them are shifted towards higher pH. These shifts can be attributed to using $\mathrm{p}K_\mathrm{A}$ of free amino acids in the model, which does not account for the change in $\mathrm{p}K_\mathrm{A}$ when the amino acid is incorporated into the peptide. The values of these shifts are not known a priori but can be qualitatively assessed based on the amino acid structure, that is, the shift should decrease with the increase in the distance between the charged group on the side chain and the peptide bond. More specifically, we expect a small shift in Lys, larger shifts in Glu and His and the largest shift in Asp. Accordingly, both NMR curves of $\mathrm{Glu}_5 - \mathrm{His}_5$ are shifted towards higher pH in comparison to simulations, suggesting a similar shift in the $\mathrm{p}K_\mathrm{A}$ values of Glu and His. In the case of $\mathrm{Lys}_5 - \mathrm{Asp}_5$, only the Asp curve is shifted, while the Lys curves almost overlap, suggesting that the $\mathrm{p}K_\mathrm{A}$ of Asp is affected by its incorporation into the peptide, while Lys remains almost unaffected. These effects may also explain the difference between pI calculated from simulations and CZE. To test this hypothesis, we ran a new set of simulations, increasing the $\mathrm{p}K_\mathrm{A}$ of Glu

and His by ΔpI of $Glu_5 - His_5$ and the pK_A of Asp by 2ΔpI of $Lys_5 - Asp_5$ so that the pI of the new simulations would match the pI of CZE. Simulations using the modified pK_A values almost perfectly match the NMR results for each individual amino acid, as shown in Fig. S18. Thus, the NMR results confirm that the simulations quantitatively capture the deformation of titration curves and support the hypothesis that this deformation is primarily caused by electrostatic interactions between charged groups, while the incorporation of amino acids into the peptide slightly increases their pK_A values in comparison to free amino acids.

NMR spectroscopy provides local information on the atomic structure and, in principle, on the degree of ionisation of each individual amino acid, as long as its signal can be resolved in the spectrum. Thus, NMR chemical shifts reflect not only the degree of ionisation but also other changes in the local environment, e.g., due to conformational changes. Specific carbon atoms have been selected according to the literature as good reporters of ionisation, based on previous studies on proteins.[49] However, we did not always observe 5 peaks, as expected for 5 amino acids (see ESI, Section 2.4.2). Therefore, we used the centre of mass of the NMR peaks corresponding to the same carbon atoms within the amino acid to determine the average degree of ionisation of amino acids of the same type. Thus, NMR appears to be the best method for determining local ionisation because its only limitation is the need to resolve individual peaks in the spectrum of a peptide composed of nearly equivalent amino acids.

Conclusion

By combining coarse-grained simulations with several experimental methods, we were able to quantitatively analyse the pH-controlled ionisation of two model peptides. The large ΔpK_A of the $Lys_5 - Asp_5$ peptide results in a titration curve with two inflection points, each corresponding to ionisation changes within one of the blocks. In contrast, the small ΔpK_A of the $Glu_5 - His_5$ peptide results in an almost linear titration curve in a broad pH

range around the isoelectric point, indicating simultaneous changes in the ionisation of both blocks. We predicted the ionisation response of both peptides using a simple coarse-grained model which only accounts for basic aspects of molecular geometry and for electrostatic interactions, using pK_A of free amino acids as input parameters. Our simulation predictions agreed with the experimentally determined charge of the peptides, thus suggesting that their ionisation response is predominantly controlled by electrostatic interactions. Capillary zone electrophoresis yielded the isoelectric point but required renormalisation using the simulation results as the input. Potentiometric titration yielded a systematically shifted absolute charge due to the presence of other ions, thus requiring independently determining the isoelectric point to yield the actual charge of the peptide, and even then the determination was unreliable at high pH. Lastly, NMR provided detailed information about the ionisation of individual acid and base groups and allowed us to estimate how the pK_A of each amino acid was affected by its incorporation into the peptide. Thus, our results suggest a novel route to bottom-up engineering the ionisation response of peptides and weak ampholytes by choosing suitable pK_A values of the side-chains and by optimising the sequence using numerical simulations. These simulations can be completed within a few hours of computer time (see ESI, Section 1.4). As a result, this modeling is considerably less expensive than experiments. Although a more refined model would be needed for modeling more complicated peptide sequences or disordered proteins, such models have been previously designed, accounting for hydrophobicity, hydrogen bonding and other effects. Therefore, our results provide the missing piece of the puzzle by showing that the ionisation-conformation coupling can be modeled quantitatively. Thanks to these developments, our coarse-grained modeling may become a key predictive tool in the design of peptides and ampholytes with tunable ionisation.

Methods

Full details of all simulations and experiments, including the data analysis and simulation/experimental protocol are provided in supporting information (ESI).

Simulations

Coarse-grained simulations

The simulation contained 10 peptide chains and 100 additional salt ion pairs to keep the ionic strength approximately constant. The box length $L = 25.513\,\mathrm{nm}$ was chosen to match the concentrations used in the titration experiments. We used the full Coulomb potential to account for electrostatic interactions, the WCA potential for steric interactions, the harmonic potential for connections between particles, with no restrictions on bond angles or dihedrals.

We performed a Langevin dynamics simulation in implicit solvent, using the constant-pH ensemble to account for the ionisation reactions (Eq. 1 and 2).[29,50] $\mathrm{p}K_\mathrm{A}$ and pH were input parameters, while degree of ionisation was the output of the simulation. The reaction was implemented as a Monte Carlo procedure, changing the state from non-ionised to ionised in the forward direction of the reaction, or vice versa in the reverse direction.[29,50] In both reactions, H^+ was inserted or deleted at a random position in the system, and the new state was accepted with the probability[29,50]

$$P_\mathrm{acc}^\xi = \min[1, \exp(-\beta \Delta U_\mathrm{on} + \xi(\mathrm{pH} - \mathrm{p}K_\mathrm{A}) \ln(10))] \tag{6}$$

where $\beta = 1/k_\mathrm{B}T$, $\Delta U_\mathrm{on} = U_\mathrm{n} - U_\mathrm{o}$ is the energy difference between the new state (n) and the old state (o), and $\xi = \pm 1$ for the forward or reverse direction of the reaction.

All-atom simulations

Equilibrium bond lengths for the CG model, r_{CC}, r_{AC} and r_{BC} were determined using all-atom (AA) molecular dynamics (MD) simulations of amino acid tetramers Glu_4, His_4, Lys_4 and Asp_4, with fully ionised side chains and protective groups on the C and N-ends. They were solvated by water and neutralised by Cl^- and Na^+ ions, with additional Na^+ and Cl^- ions to fix the ionic strength. The use of tetramers allowed us to assess whether these distances were affected by position of the amino acid in the sequence: N-terminus, middle, C-terminus.

Experiments

Capillary zone electrophoresis (CZE)

Samples for CZE with the concentration of the peptide chains $c = 0.1\,\mathrm{mM}$, equivalent to the concentration of the monomeric units [Glu] = [His] = [Asp] = [Lys] = 0.5 mM, were prepared by dissolving the peptide directly in running buffers, specified in ESI, Table S5. We determined R_{H} using the Diffusion Ordered spectroscopy (DOSY) NMR (ESI, Section 2.4.3), showing that R_{H} is almost constant throughout the pH range in both peptides. This allowed us to eliminate R_{H} from the function $\mu(\mathrm{pH})$ and to determine the charge on the peptides using the maximum mobility, μ^{max}, measured at $\mathrm{pH}^{\mathrm{max}}$. If the peptide was fully ionized at $\mathrm{pH}^{\mathrm{max}}$, this approach should re-scale the electrophoretic mobilities to yield the bare charge because the factors relating bare and effective charge cancel each other. Unfortunately, we were able to measure CZE only in a limited range, preventing us from reliably assuming full ionisation at $\mathrm{pH}^{\mathrm{max}}$. We corrected for this limitation by renormalising $\mu^{\mathrm{max}}(\mathrm{pH}^{\mathrm{max}})$ by the peptide charge determined from our simulations, $z_{\mathrm{sim}}(\mathrm{pH}^{\mathrm{max}} + \Delta\mathrm{pI})$ (ESI Section 2.2.2 for full details of data analysis)

$$z_{\mathrm{cze}}(\mathrm{pH}) = \frac{\mu(\mathrm{pH} - \mathrm{pI}_{\mathrm{cze}})}{\mu^{\mathrm{max}}(\mathrm{pH}^{\mathrm{max}})} z_{\mathrm{sim}}(\mathrm{pH}^{\mathrm{max}} + \Delta\mathrm{pI}) \tag{7}$$

where $\Delta\mathrm{pI} = \mathrm{pI_{cze}} - \mathrm{pI_{sim}}$ is the difference between isoelectric points determined from CZE and from the simulations.

Potentiometric titration

The peptides were dissolved in 0.01 M standardised HCl to prepare solutions to a final concentrations of monomeric units: [Glu] = [His] = [Asp] = [Lys] = 5 mM. Sample volumes of approximately 2 ml were weighed to determine the precise amount and then titrated with standardised 0.01 M NaOH. The charge on the peptide was calculated from the pH and from the known volumes of HCl and NaOH

$$z_{\mathrm{titration}} = \frac{(V_{\mathrm{HCl}}((c_{\mathrm{H}} - c_{\mathrm{OH}}) + c_{\mathrm{HCl}}) - V_{\mathrm{NaOH}}(c_{\mathrm{H}} - c_{\mathrm{OH}} + c_{\mathrm{NaOH}}))}{(c_{\mathrm{peptide}} V_{\mathrm{NaOH}})} - x_{\mathrm{TFA}} \tag{8}$$

where c is the concentration, z is the charge on the peptide, and $x_{\mathrm{TFA}} \gtrsim 1$ is the mole fraction of trifluoroacetate (TFA) counterions contained in the peptide sample, relative to the basic side chains on the peptide. To correct for the unknown value of x_{TFA}, we used it as an adjustable parameter to match the isoelectric point determined from CZE (ESI, Section 2.3.2). The concentrations of H^+ and OH^- ions, used in Equation 8, were calculated from the measured pH and $\mathrm{p}K_{\mathrm{w}}$. Because $\mathrm{p}K_{\mathrm{w}}$ is sensitive to temperature, $z_{\mathrm{titration}}(\mathrm{pH})$ from Eq. 8 was also sensitive to temperature and to the precision of the pH measurement, yielding reliable results only at intermediate pH, $3 \gtrsim \mathrm{pH} \gtrsim 11$. We quantified the reliability of $z_{\mathrm{titration}}(\mathrm{pH})$ by comparing titrations of peptide samples with blank titrations of HCl stock solutions.

NMR

Samples for NMR were prepared by dissolving each peptide in 0.01 M HCl to a final concentration of 15 gL^{-1} and by titrating the solutions with NaOH to adjust the pH to the desired value. 2D NMR spectra, COSY and ^{1}H-^{13}C HSQC, (Fig. S16 and S17) at pH 2 were used

for peak assignment. The degrees of ionisation were determined from the chemical shift of specific atoms, which were identified in the literature as good reporters of ionisation

$$\alpha_{\text{base}}(\text{pH}) = \frac{\delta_{max} - \delta(\text{pH})}{\delta_{max} - \delta_{min}}\,, \qquad \alpha_{\text{acid}}(\text{pH}) = \frac{\delta(\text{pH}) - \delta_{min}}{\delta_{max} - \delta_{min}} \tag{9}$$

Acknowledgement

The authors would like to thank Dr. Carlos V. Melo for proofreading the manuscript and Dr. Pavel Dubský for advice regarding the CZE experiments. This work was supported by the Czech Science foundation, grant 19-10429S, and by the Grant agency of the Charles University, grant GAUK 978218. Miroslav Štěpánek acknowledges support from the Ministry of Education, Youth and Sports of the Czech Republic, Operational Programme Research, Development and Education: "Excellent Research Teams", Project No. CZ.02.1.01/0.0/0.0/15_003/0000417-CUCAM. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

Supporting Information Available

Technical details of CG and AA simulations; Determination of model parameters; Comparison of different pK_{A} values from literature; Comparison of computational demands and costs; Experimental materials; Technical details, and data analysis of CZE, potentiometric titration and NMR.

References

(1) Lund, M.; Jönsson, B. Charge regulation in biomolecular solution. *Quarterly reviews of biophysics* **2013**, *46*, 265–8.

(2) Katchalsky, A.; Gillis, J. Theory of the potentiometric titration of polymeric acids. *Rec. Trav. Chim.* **1949**, *68*, 879.

(3) Israëls, R.; Leermakers, F. A. M.; Fleer, G. J. On the Theory of Grafted Weak Polyacids. *Macromolecules* **1994**, *27*, 3087–3093.

(4) Hartley, G. S.; Roe, J. W. Ionic concentrations at interfaces. *Trans. Faraday Soc.* **1940**, *35*, 101–109.

(5) Arnold, R. The titration of polymeric acids. *Journal of Colloid Science* **1957**, *12*, 549–556.

(6) Ninham, B. W.; Parsegian, V. A. Electrostatic potential between surfaces bearing ionizable groups in ionic equilibrium with physiologic saline solution. *Journal of Theoretical Biology* **1971**, *31*, 405–428.

(7) Ikebuchi, M.; Kashiwagi, A.; Asahina, T.; Tanaka, Y.; Takagi, Y.; Nishio, Y.; Hidaka, H.; Kikkawa, R.; Shigeta, Y. Effect of medium pH on glutathione redox cycle in cultured human umbilical vein endothelial cells. *Metabolism* **1993**, *42*, 1121 – 1126.

(8) Ainis, W. N.; Boire, A.; Solé-Jamault, V.; Nicolas, A.; Bouhallab, S.; Ipsen, R. Contrasting Assemblies of Oppositely Charged Proteins. *Langmuir* **2019**, *35*, 9923–9933.

(9) Ulbrich, K. Polymeric anticancer drugs with pH-controlled activation. *Advanced Drug Delivery Reviews* **2004**, *56*, 1023–1050.

(10) van Lente, J. J.; Claessens, M. M. A. E.; Lindhoud, S. Charge-Based Separation of Proteins Using Polyelectrolyte Complexes as Models for Membraneless Organelles. *Biomacromolecules* **2019**, *20*, 3696–3703.

(11) Freudenberg, U.; Atallah, P.; Limasale, Y. D. P.; Werner, C. Charge-tuning of glycosaminoglycan-based hydrogels to program cytokine sequestration. *Faraday Discussions* **2019**, *219*, 244–251.

(12) Schirmer, L.; Chwalek, K.; Tsurkan, M. V.; Freudenberg, U.; Werner, C. Glycosaminoglycan-based hydrogels with programmable host reactions. *Biomaterials* **2020**, *228*, 119557.

(13) Jana, S.; Uchman, M. Poly(2-oxazoline)-based Stimulus-Responsive (Co)polymers: An Overview of their Design, Solution Properties, Surface-chemistries and Applications. *Progress in Polymer Science* **2020**, 101252.

(14) Srivastava, D.; Santiso, E.; Gubbins, K.; Barroso da Silva, F. L. Computationally Mapping pKa Shifts Due to the Presence of a Polyelectrolyte Chain around Whey Proteins. *Langmuir* **2017**, *33*, 11417–11428.

(15) Lund, M. Electrostatic chameleons in biological systems. *Journal of the American Chemical Society* **2010**, *132*, 17337–17339.

(16) de Vos, W. M.; Leermakers, F. A. M.; de Keizer, A.; Cohen Stuart, M. A.; Kleijn, J. M. Field Theoretical Analysis of Driving Forces for the Uptake of Proteins by Like-Charged Polyelectrolyte Brushes: Effects of Charge Regulation and Patchiness. *Langmuir* **2010**, *26*, 249–259.

(17) Avni, Y.; Andelman, D.; Podgornik, R. Charge regulation with fixed and mobile charged macromolecules. *Current Opinion in Electrochemistry* **2019**, *13*, 70–77.

(18) de Vries, R.; Cohen Stuart, M. Theory and simulations of macroion complexation. *Current Opinion in Colloid & Interface Science* **2006**, *11*, 295–301.

(19) Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chemical Reviews* **2018**, *118*, 1691–1741.

(20) von der Lühe, M.; Weidner, A.; Dutz, S.; Schacher, F. H. Reversible Electrostatic Adsorption of Polyelectrolytes and Bovine Serum Albumin onto Polyzwitterion-Coated

Magnetic Multicore Nanoparticles: Implications for Sensing and Drug Delivery. *ACS Applied Nano Materials* **2018**, *1*, 232–244.

(21) Biehl, P.; von der Lühe, M.; Schacher, F. H. Reversible Adsorption of Methylene Blue as Cationic Model Cargo onto Polyzwitterionic Magnetic Nanoparticles. *Macromolecular Rapid Communications* **2018**, *39*, 1800017.

(22) Vrbata, D.; Uchman, M. Preparation of lactic acid- and glucose-responsive poly(ε-caprolactone)-b-poly(ethylene oxide) block copolymer micelles using phenylboronic ester as a sensitive block linkage. *Nanoscale* **2018**, *10*, 8428–8442.

(23) Ďorďovič, V.; Vojtová, J.; Jana, S.; Uchman, M. Charge reversal and swelling in saccharide binding polyzwitterionic phenylboronic acid-modified poly(4-vinylpyridine) nanoparticles. *Polymer Chemistry* **2019**, *10*, 5522–5533.

(24) Dyakonova, M. A.; Gotzamanis, G.; Niebuur, B.-J.; Vishnevetskaya, N. S.; Raftopoulos, K. N.; Di, Z.; Filippov, S. K.; Tsitsilianis, C.; Papadakis, C. M. pH Responsiveness of hydrogels formed by telechelic polyampholytes. *Soft Matter* **2017**, *13*, 3568–3579.

(25) Xu, W.; Rudov, A.; Oppermann, A.; Wypysek, S.; Kather, M.; Schroeder, R.; Richtering, W.; Potemkin, I. I.; Wöll, D.; Pich, A. Synthesis of Polyampholyte Janus-like Microgels by Coacervation of Reactive Precursors in Precipitation Polymerization. *Angewandte Chemie International Edition* **2020**, *59*, 1248–1255.

(26) Praveen, K.; Das, S.; Dhaware, V.; Pandey, B.; Mondal, B.; Gupta, S. S. pH-Responsive "Supra-Amphiphilic" Nanoparticles Based on Homoarginine Polypeptides. *ACS Applied Bio Materials* **2019**, *2*, 4162–4172.

(27) Martens, A. A.; Portale, G.; Werten, M. W. T.; de Vries, R. J.; Eggink, G.; Cohen Stuart, M. A.; de Wolf, F. A. Triblock Protein Copolymers Forming Supramolecular Nanotapes and pH-Responsive Gels. *Macromolecules* **2009**, *42*, 1002–1009.

(28) Du, A. W.; Stenzel, M. H. Drug Carriers for the Delivery of Therapeutic Peptides. *Biomacromolecules* **2014**, *15*, 1097–1114.

(29) Landsgesell, J.; Nová, L.; Rud, O.; Uhlík, F.; Sean, D.; Hebbeker, P.; Holm, C.; Košovan, P. Simulations of ionization equilibria in weak polyelectrolyte solutions and gels. *Soft Matter* **2019**, *15*, 1155–1185.

(30) Ullner, M.; Jönsson, B.; Widmark, P. Conformational properties and apparent dissociation constants of titrating polyelectrolytes: Monte Carlo simulation and scaling arguments. *The Journal of Chemical Physics* **1994**, *100*, 3365.

(31) Ullner, M.; Jönsson, B.; Söderberg, B.; Peterson, C. A Monte Carlo study of titrating polyelectrolytes. *The Journal of Chemical Physics* **1996**, *104*, 3048–3057.

(32) Borisov, O. V.; Zhulina, E. B.; Leermakers, F. A.; Ballauff, M.; Müller, A. H. E. In *Self Organized Nanostructures of Amphiphilic Block Copolymers I*; Müller, A. H. E., Borisov, O., Eds.; Adv. Polym. Sci.; Springer Berlin Heidelberg, 2011; Vol. 241; pp 1–55.

(33) Uhlík, F.; Košovan, P.; Limpouchová, Z.; Procházka, K.; Borisov, O. V.; Leermakers, F. A. M. Modeling of Ionization and Conformations of Starlike Weak Polyelectrolytes. *Macromolecules* **2014**, *47*, 4004–4016.

(34) Murmiliuk, A.; Košovan, P.; Janata, M.; Procházka, K.; Uhlík, F.; Štěpánek, M. Local pH and Effective pK of a Polyelectrolyte Chain: Two Names for One Quantity? *ACS Macro Letters* **2018**, *7*, 1243–1247.

(35) Laguecir, A.; Ulrich, S.; Labille, J.; Fatin-Rouge, N.; Stoll, S.; Buffle, J. Size and pH effect on electrical and conformational behavior of poly(acrylic acid): Simulation and experiment. *European Polymer Journal* **2006**, *42*, 1135–1144.

(36) Zheng, W.; Borgia, A.; Buholzer, K.; Grishaev, A.; Schuler, B.; Best, R. B. Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment. *Journal of the American Chemical Society* **2016**, *138*, 11702–11713, tex.ids: zheng2016a.

(37) Soranno, A.; Holla, A.; Dingfelder, F.; Nettels, D.; Makarov, D. E.; Schuler, B. Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations. *Proceedings of the National Academy of Sciences* **2017**, *114*, E1833–E1839, tex.ids: soranno2017a.

(38) Kyrychenko, A.; Blazhynska, M. M.; Kalugin, O. N. Protonation-dependent adsorption of polyarginine onto silver nanoparticles. *Journal of Applied Physics* **2020**, *127*, 075502.

(39) Statt, A.; Casademunt, H.; Brangwynne, C. P.; Panagiotopoulos, A. Z. Model for disordered proteins with strongly sequence-dependent liquid phase behavior. *The Journal of Chemical Physics* **2020**, *152*, 075101, tex.ids: statt2020a.

(40) Chang, L.-W.; Lytle, T. K.; Radhakrishna, M.; Madinya, J. J.; Vélez, J.; Sing, C. E.; Perry, S. L. Sequence and entropy-based control of complex coacervates. *Nature Communications* **2017**, *8*, 1723.

(41) Zhang, P.; Gao, Z.; Cui, J.; Hao, J. Dual-Stimuli-Responsive Polypeptide Nanoparticles for Photothermal and Photodynamic Therapy. *ACS Applied Bio Materials* **2019**, *3*, 561–569.

(42) Haynes, W. *CRC Handbook of Chemistry and Physics*, 96th ed.; CRC Press: New York, 2015.

(43) Lide, D. R. *CRC Handbook of Chemistry and Physics*, 72nd ed.; CRC Press: New York, 1991.

(44) Grest, G. S.; Kremer, K. Molecular dynamics simulation for polymers in the presence of a heat bath. *Physical Review A* **1986**, *33*, 3628–31.

(45) Lobaskin, V.; Dünweg, B.; Holm, C. Electrophoretic mobility of a charged colloidal particle: a computer simulation study. *Journal of Physics: Condensed Matter* **2004**, *16*, S4063–S4073.

(46) Hill, R. J.; Saville, D.; Russel, W. Electrophoresis of spherical polymer-coated colloidal particles. *Journal of Colloid and Interface Science* **2003**, *258*, 56–74.

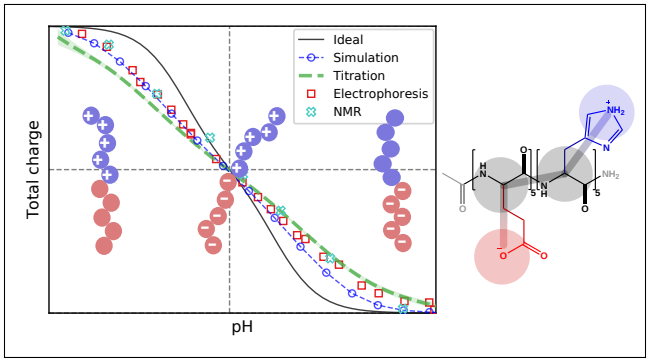
(47) Griffiths, P. C.; Paul, A.; Stilbs, P.; Petterson, E. Charge on Poly(ethylene imine): Comparing Electrophoretic NMR Measurements and pH Titrations. *Macromolecules* **2005**, *38*, 3539–3542.

(48) Nová, L.; Uhlík, F.; Košovan, P. Local pH and effective pK_{A} of weak polyelectrolytes - insights from computer simulations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 14376–14387.

(49) Hass, M. A.; Mulder, F. A. Contemporary NMR Studies of Protein Electrostatics. *Annual Review of Biophysics* **2015**, *44*, 53–75, tex.ids: hass2015a.

(50) Reed, C. E.; Reed, W. F. Monte Carlo study of titration of linear polyelectrolytes. *The Journal of Chemical Physics* **1992**, *96*, 1609–1620.

Graphical TOC Entry



Supporting Information: Quantitative prediction of pH-controlled ionization of oligoampholytes

Raju Lunkad, Anastasiia Murmiliuk, Pascal Hebbeker, Milan Boublík, Zdeněk Tošner, Miroslav Štěpánek, and Peter Košovan*

Department of Physisal and Macromolecular Chemistry, Charles University, Hlavova 8, 128 43 Prague, Czech Republic

E-mail: peter.kosovan@natur.cuni.cz

June 24, 2020

Contents

1 Simulations

1.1 Coarse Grained (CG) Simulations

1.1.1 Interaction Potentials

Bonds in the coarse-grained model were represented by the harmonic potential

$$U_{\mathrm{h}}(r) = -\frac{k_{\mathrm{h}}}{2}(r-b)^2 \tag{1}$$

with the stiffness constant $k_{\mathrm{h}} = 400\, k_{\mathrm{B}}T\mathrm{nm}^{-1}$ common to all bonds. The bond stiffness was chosen arbitrarily. We tested that different values of this parameter yield very similar simulation results. The parameter b determines the equilibrium bond length. The bond length between central beads was $b_{\mathrm{CC}} = 0.382\,\mathrm{nm}$ for the all amino acid pairs. The bond length between central and side chain beads was $b_{\mathrm{AC}} = 0.435\,\mathrm{nm}$ for Glutamic acid, $b_{\mathrm{AC}} = 0.329\,\mathrm{nm}$ for Aspartic acid, $b_{\mathrm{BC}} = 0.452\,\mathrm{nm}$ for Histidine, and $b_{\mathrm{BC}} = 0.558\,\mathrm{nm}$ for Lysine. These values were determined from all-atom simulations described in Section1.2.

The short-range excluded volume interaction was modeled using the fully-repulsive Weeks-Chandler-Andersen (WCA) potential between all bead pairs

$$U_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon & r \le r_{\mathrm{cut}} \\ 0 & r > r_{\mathrm{cut}} \end{cases} \tag{2}$$

where $\epsilon = 1\, k_{\mathrm{B}}T$, $\sigma = 0.35\,\mathrm{nm}$ and $r_{\mathrm{cut}} = 0.4\,\mathrm{nm}$, which defines the effective particle size.

Electrostatic interaction between point charges i and j were represented by the Coulomb potential:

$$U_{ij}(r) = z_i z_j k_{\mathrm{B}}T\frac{l_{\mathrm{B}}}{r} = \frac{1}{4\pi\epsilon_0\epsilon_{\mathrm{r}}}\frac{z_i z_j e^2}{r} \tag{3}$$

where z is the valency, e is the elementary charge, ϵ_0 is the permittivity of free space and ϵ_{r} is the relative permittivity. We set the Bjerrum length to its approximate value in aqueous

solutions at ambient temperature, $l_B \approx 0.71\,\mathrm{nm}$. Electrostatic interactions were calculated using the P3M method, tuned to relative accuracy 0.001 using the tuning algorithm implemented in the ESPResSo simulation software that we used for simulations of the CG model.[S1]

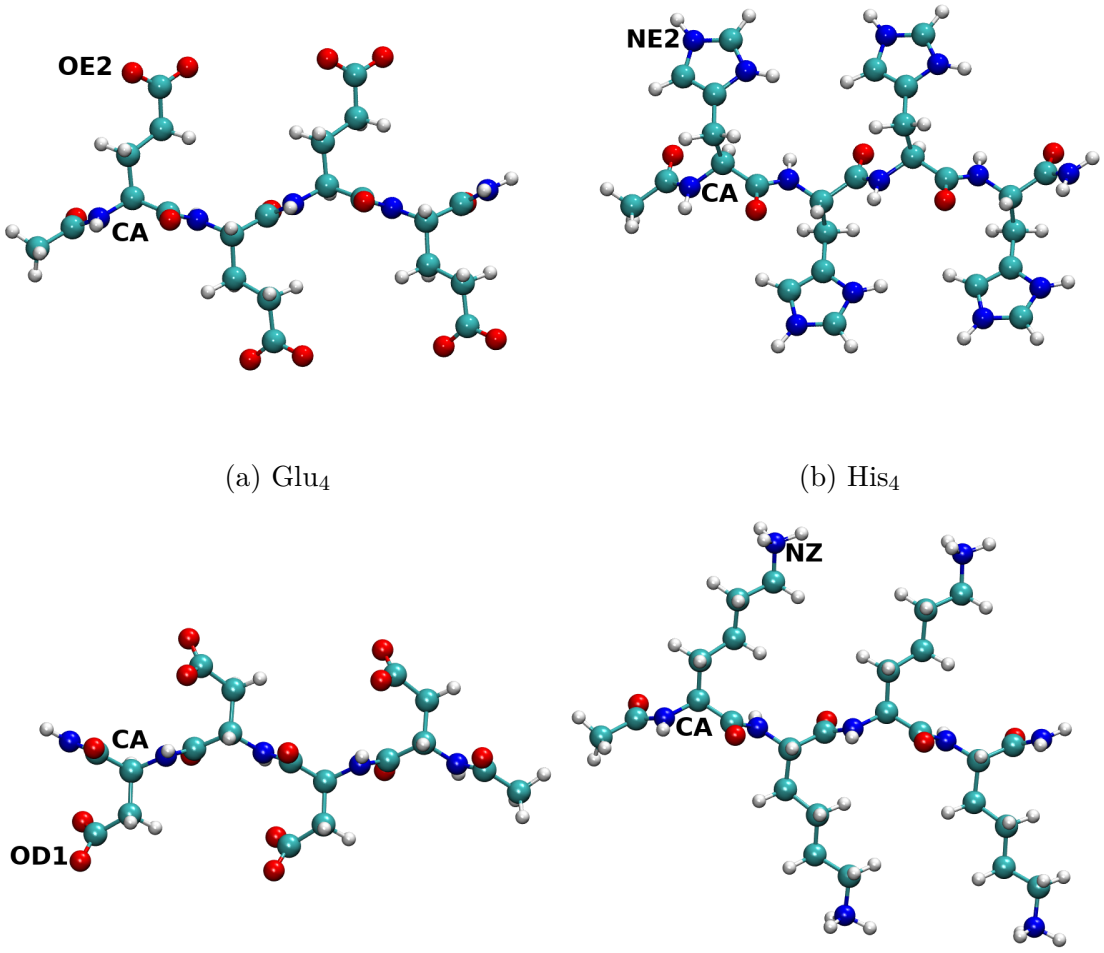
1.1.2 Simulation Protocol and data analysis

All Simulations were performed using time step $\delta t = 0.01\tau$, where $\tau = \sigma\sqrt{m/\epsilon}$. The particle mass m is arbitrary and has no effect on the results. The system was kept at a constant temperature $T = 300\,\mathrm{K}$ via a Langevin thermostat with a damping constant $\gamma = 1.0\tau^{-1}$. The duration of each simulation was 10^5 cycles, and each cycle consisted of 10 reaction moves followed by 100 integration steps of the Langevin dynamics. First 20% of all cycles were discarded as the equilibration. The remaining part was treated as production run and used for analysis. The productive run typically produced approximately 10^3 uncorrelated samples of peptide conformations measured by the autocorrelation time of the radius of gyration, and twice the number of uncorrelated samples of the degree of ionisation. We used the correlation-corrected error estimates to assess the statistical accuracy of our data.[S2]

1.2 All Atom (AA) Simulations

1.2.1 Simulation Model and Setup

We simulated the tetramers solvated with 4028 water molecules, neutralised by adding Cl^- and Na^+ ions, with additional Na^+ and Cl^- ions to represent the added salt, which determined the ionic strength. In total, $11Cl^-$ and $7Na^+$ ions were present for the Glu_4 and Asp_4, while $7Cl^-$ and $11Na^+$ ions were present for the His_4 and Lys_4. The system was simulated in a cubic box with an edge length of $L = 6.00\,\mathrm{nm}$, yielding the salt concentration of 0.05 M. Gromacs 2018.6 package was used for AA MD simulations.[S3,S4]



(a) Glu_4 (b) His_4

(c) Asp_4 (d) Lys_4

Figure S1: Initial configurations of the peptides used in AA simulations. Labels mark the atoms which we used to measure various intra-molecular distances.

1.2.2 Interaction potentials

We used AMBER99sb-ILDN force filed for the peptide and TIP3P force field for the water molecules. The LINCS algorithm was used to impose the constraints on the bond lengths. The Particle Mesh Ewald (PME) method was used for long-range electrostatic interactions. The Van der Waals interactions were truncated at 1.2 nm.

1.2.3 Simulation protocol and data analysis

We performed 5×10^4 energy minimisation steps using the steepest descent method to remove high-energy contacts. After energy minimisation, we performed an $[NVT]$ run of 500 ps using the velocity re-scaling algorithm at a temperature of $T = 300$ K with the thermostat coupling constant $\tau = 0.1$ ps. Last, we performed an $[NPT]$ run of 100 ns using Parrinello-Rahman algorithm at pressure 1 bar with a pressure coupling constant $\tau = 2.0$ ps. The integration time step used was 2 fs for all simulations. The last 90 ns of the $[NPT]$ run were used for production and data analysis.

1.3 Determination and validation of parameters for the CG model

To determine the bond lengths for the CG model of the peptides, we calculated the distributions of distances between the atoms of each tetramer from AA simulations. Furthermore, we calculated additional distributions from AA simulations, which we compared to the corresponding distributions from CG simulations to verify the validity of our model. Lastly, we determined the same set of distances from the peptide structures after simple energy minimization using the Avogadro software.[S5]

Specifically, we calculated the distance distributions between CA atoms in the peptide backbones from the AA simulations, which we then used to set the equilibrium bond length, r_{CC} between the C beads in the CG simulations. Subsequently, we calculated the distance distributions between the CA atom in the peptide backbone and the charged atom in the side chain. The charged atom was OE2 for Glu_4, NE2 for His_4, OD1 for Asp_4 and NZ for

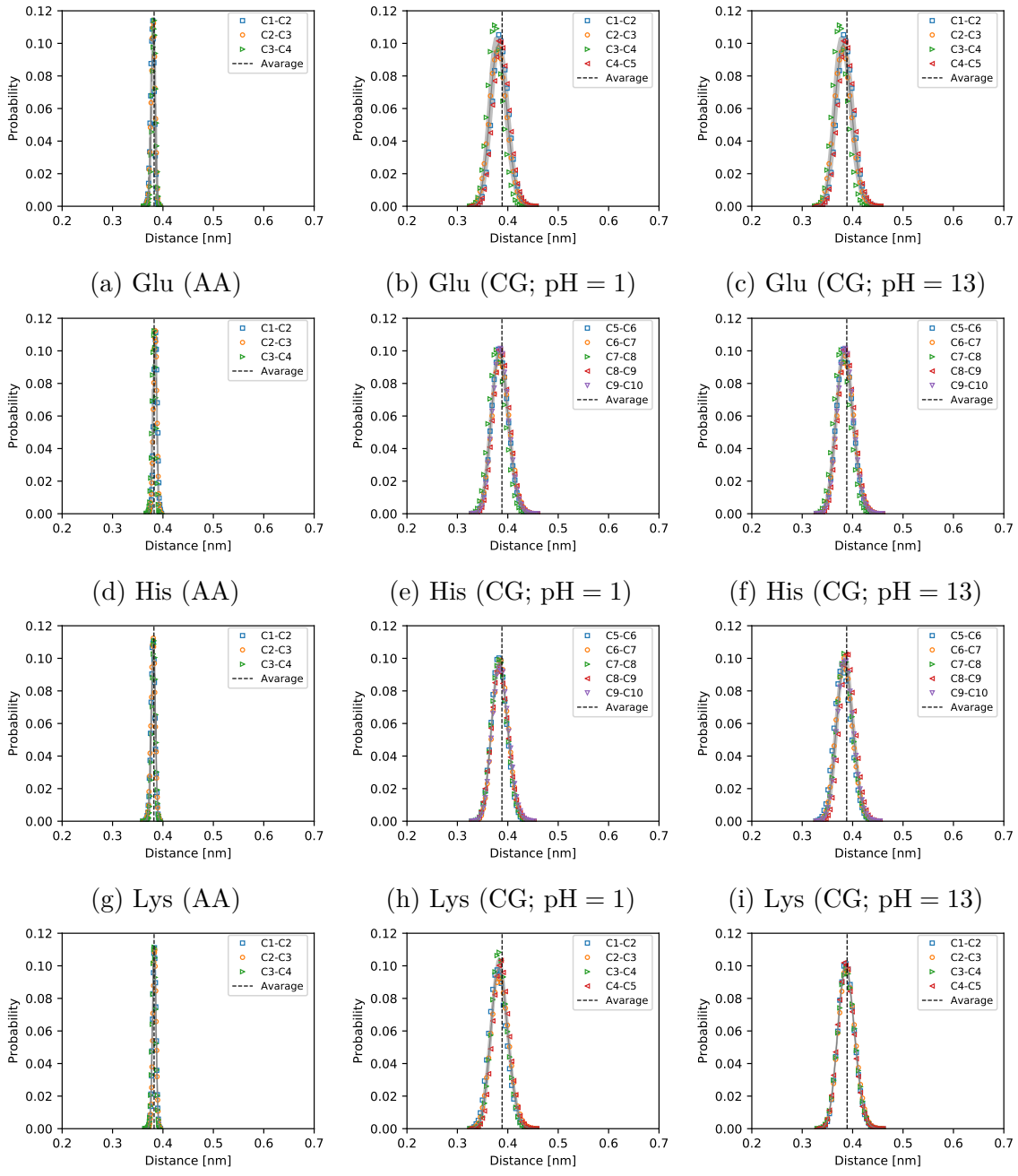
Table S1: Average distance between the CA atoms on neighbouring amino acids in the AA simulations; Distance between the same atoms determined using energy minimisation in the Avogadro software; Average distance between the central beads in the CG simulations, r_{CC}, at the extreme pH values.

Peptide	Acid Amino acid	distance [nm]	Base Amino acid	distance [nm]
$\mathrm{Glu_5 - His_5}$	Glu (AA)	0.382 ± 0.004	His (AA)	0.382 ± 0.004
	Glu (CG; pH = 1)	0.389 ± 0.016	His (CG; pH = 1)	0.388 ± 0.016
	Glu (CG; pH = 13)	0.389 ± 0.016	His (CG; pH = 13)	0.388 ± 0.016
	Glu (Avogadro)	0.388	His (Avogadro)	0.389
$\mathrm{Lys_5 - Asp_5}$	Asp (AA)	0.382 ± 0.006	Lys (AA)	0.382 ± 0.006
	Asp (CG; pH = 1)	0.389 ± 0.016	Lys (CG; pH = 1)	0.388 ± 0.016
	Asp (CG; pH = 13)	0.389 ± 0.016	Lys (CG; pH = 13)	0.388 ± 0.016
	Asp (Avogadro)	0.391	Lys (Avogadro)	0.389

$\mathrm{Lys_4}$, as indicated in Fig.S1. We used these distances to set the CG equilibrium bond lengths between the C and A beads of the acidic side chains r_{AC}, and between the C and B beads of the basic side-chains, r_{BC}. Furthermore, to verify the charge-charge distance predicted from the CG simulations, we also calculated the distribution between the charged groups on the nearest-neighbour and next-nearest-neighbour amino acid side chains, and compared them between the AA and CG simulations. To verify whether any of the above distances depend on pH, we used CG simulations at two extreme pH values: pH = 1 and pH = 13. At pH = 1 the basic groups are fully charged, while the acidic groups are uncharged. These results should match the AA simulations of the fully charged basic peptides. At pH = 13 this situation is reversed, and these results should match the AA simulations of the fully charged acidic peptides.

1.3.1 Distances between the central beads.

Table S1 shows that the distances between CA atoms on the peptide backbone from AA simulations were very well reproduced by the distances between the C beads in the CG



(a) Glu (AA) (b) Glu (CG; pH = 1) (c) Glu (CG; pH = 13)

(d) His (AA) (e) His (CG; pH = 1) (f) His (CG; pH = 13)

(g) Lys (AA) (h) Lys (CG; pH = 1) (i) Lys (CG; pH = 13)

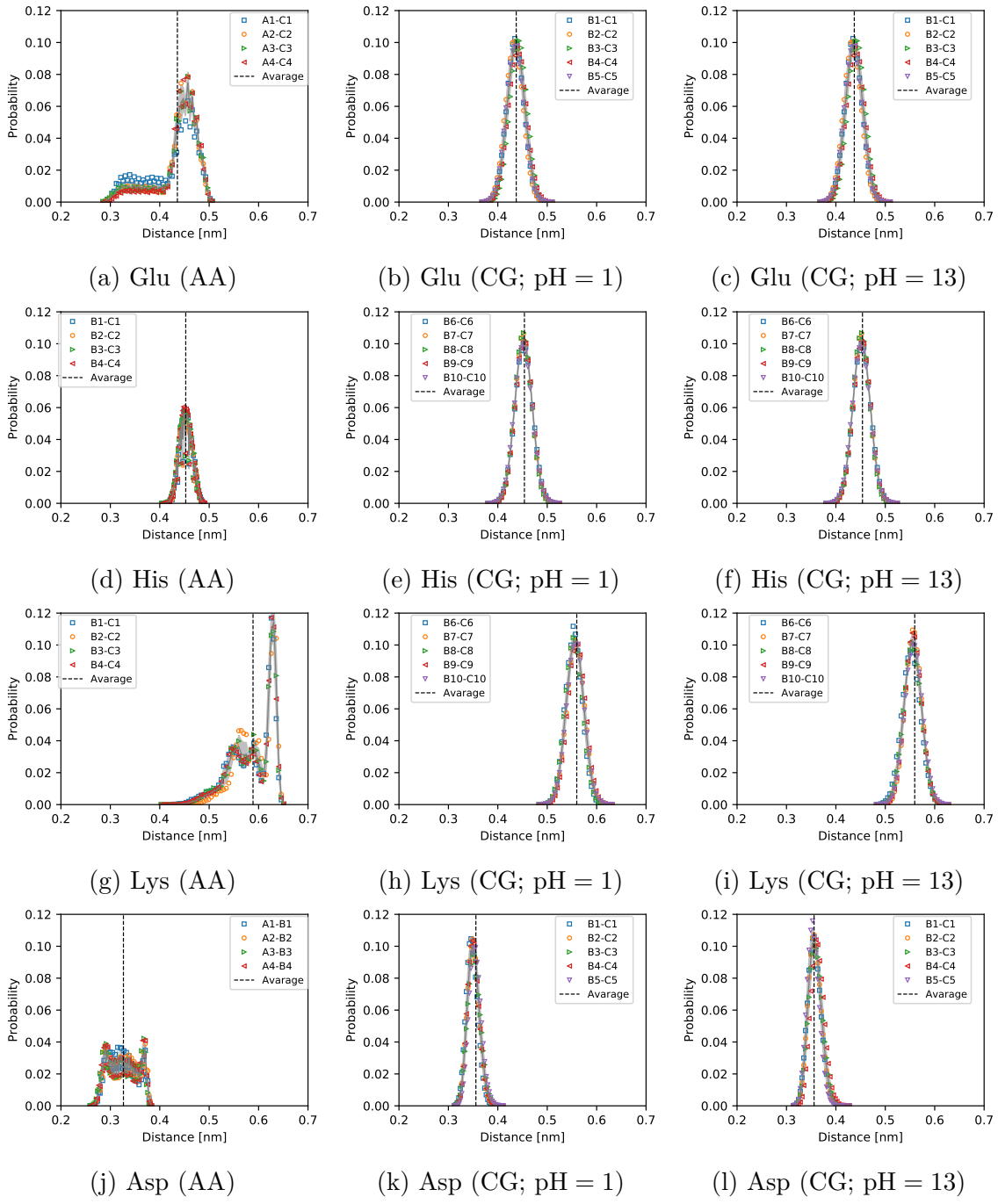
(j) Asp (AA) (k) Asp (CG; pH = 1) (l) Asp (CG; pH = 13)

Figure S2: Distribution of distances between CA atoms on neighbouring amino acids from AA and CG simulations (the latter at the extreme pH values). The vertical line shows the average distance after averaging over individual pairs. These average values are listed in Table S1.

models. The differences between AA and CG models within approx. 1% are well below the statistical uncertainty. The distances measured using the Avogadro software agree very well with all simulations. The distributions of distances in Fig. S2 reveal that all distributions consist of a single peak. The AA distributions are very narrow, while the CG distributions are slightly broader. However, the average distances from these distributions show no visible dependence on the type of the amino acid, or on the pH.

1.3.2 Distances between central beads C and side-chain beads A or B.

Distances between the CA atoms on peptide backbone and the charged atoms on the respective side chains show a similar trend to the distances between CA atoms. Table S2 reveals that the differences between the AA and CG simulations are slightly larger in Asp and Lys, but they remain within the estimated statistical error; thus, we consider them insignificant. Also the distances measured using the Avogadro software agree with all simulations, although the differences are slightly beyond the estimated statistical error of the simulation data. Fig. S3 reveals that these differences could be attributed to a more complex shape of the AA distributions, which was not fully reproduced by the CG simulations. This could be attributed to cis-trans conformational transitions, hydrogen bonds, or other specific interactions, which were not explicitly included in the CG model. Nevertheless, the average values of the distances are reproduced within the statistical uncertainty of approx. 5%.



(a) Glu (AA) (b) Glu (CG; pH = 1) (c) Glu (CG; pH = 13)

(d) His (AA) (e) His (CG; pH = 1) (f) His (CG; pH = 13)

(g) Lys (AA) (h) Lys (CG; pH = 1) (i) Lys (CG; pH = 13)

(j) Asp (AA) (k) Asp (CG; pH = 1) (l) Asp (CG; pH = 13)

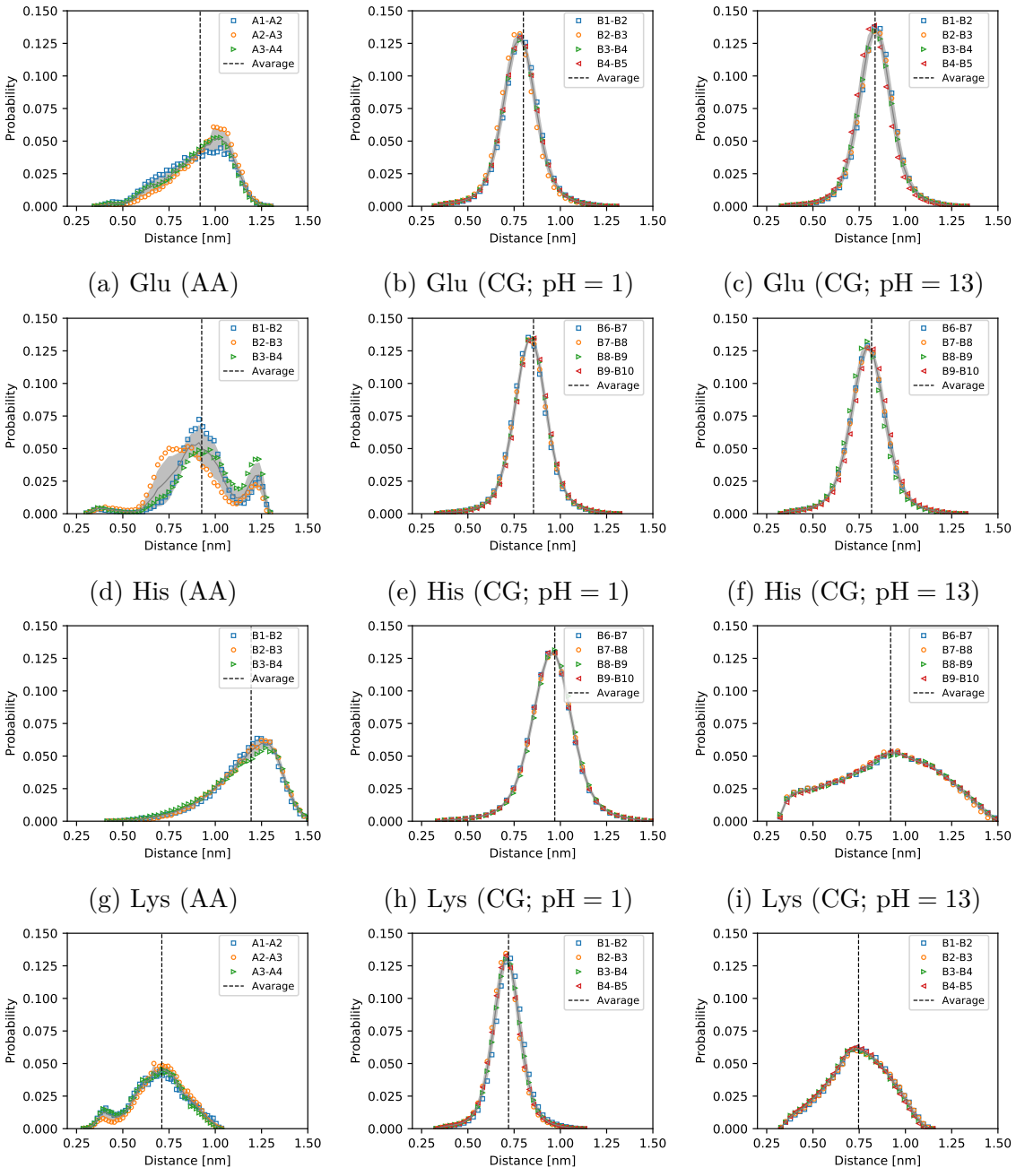
Figure S3: Distribution of distances between the CA atoms and the charged group on the amino acids from the AA and CG simulations (the latter at the extreme pH values). The vertical line shows the average distance after averaging over individual pairs. These average values are listed in Table S2.

Table S2: Average distance between the CA atoms and the charged group on the amino acids in the AA simulations; Distance between the same atoms determined using energy minimisation in the Avogadro software; Average distance between the central bead and the A or B bead in the CG simulations, r_{AC} and r_{BC}, at the extreme pH values;

Peptide	Acid Amino acid	distance [nm]	Base Amino acid	distance [nm]
$\mathrm{Glu_5 - His_5}$	Glu (AA)	0.436 ± 0.044	His (AA)	0.453 ± 0.013
	Glu (CG; pH = 1)	0.437 ± 0.018	His (CG; pH = 1)	0.454 ± 0.018
	Glu (CG; pH = 13)	0.437 ± 0.018	His (CG; pH = 13)	0.454 ± 0.018
	Glu (Avogadro)	0.453	His (Avogadro)	0.462
$\mathrm{Lys_5 - Asp_5}$	Asp (AA)	0.327 ± 0.029	Lys (AA)	0.589 ± 0.042
	Asp (CG; pH = 1)	0.356 ± 0.012	Lys (CG; pH = 1)	0.560 ± 0.018
	Asp (CG; pH = 13)	0.356 ± 0.012	Lys (CG; pH = 13)	0.560 ± 0.018
	Asp (Avogadro)	0.385	Lys (Avogadro)	0.639

1.3.3 Distances between the nearest-neighbour side-chain beads.

All previously discussed distances were used as inputs in constructing the CG model; therefore, as expected, the CG model reproduces well their values calculated from the AA simulations. To verify the validity of the CG model, we compared the distributions of distances between neighbouring charged groups on the peptides. The ability of the CG model to reproduce these distances is crucial to quantitatively account for the effect of electrostatic interactions within the peptide. Table S3 reveals that the average distances between the nearest-neighbour charged groups are well reproduced, but the statistical uncertainty has increased to approximately 10%. Interestingly, the distances determined from Avogadro reasonably agree with the simulations. The average distances are weakly affected by the pH, albeit still below the estimated statistical error. In contrast, the shape of the distributions in Fig. S4 clearly depends on pH and on the type of the amino acid. For Glu and Lys, the AA simulations yield a skewed distribution with a single peak; inc contrast, for His and Asp, they yield a double-peak distribution. The CG simulations yield a single-peaked distribution



(a) Glu (AA) (b) Glu (CG; pH = 1) (c) Glu (CG; pH = 13)

(d) His (AA) (e) His (CG; pH = 1) (f) His (CG; pH = 13)

(g) Lys (AA) (h) Lys (CG; pH = 1) (i) Lys (CG; pH = 13)

(j) Asp (AA) (k) Asp (CG; pH = 1) (l) Asp (CG; pH = 13)

Figure S4: Distribution of distances between charged group on neighbouring amino acids from the AA and CG simulations (the latter at the extreme pH values). The vertical line shows the average distance after averaging over individual pairs. These average values are listed in Table S3.

Table S3: Average distance between the charged group on the nearest-neighbour amino acids in the AA simulations; Distance between the same atoms determined using energy minimisation in the Avogadro software; Average distance between the A and the nearest-neighbour A bead, or B and the nearest-neighbour B bead, in the CG simulations at the extreme pH values.

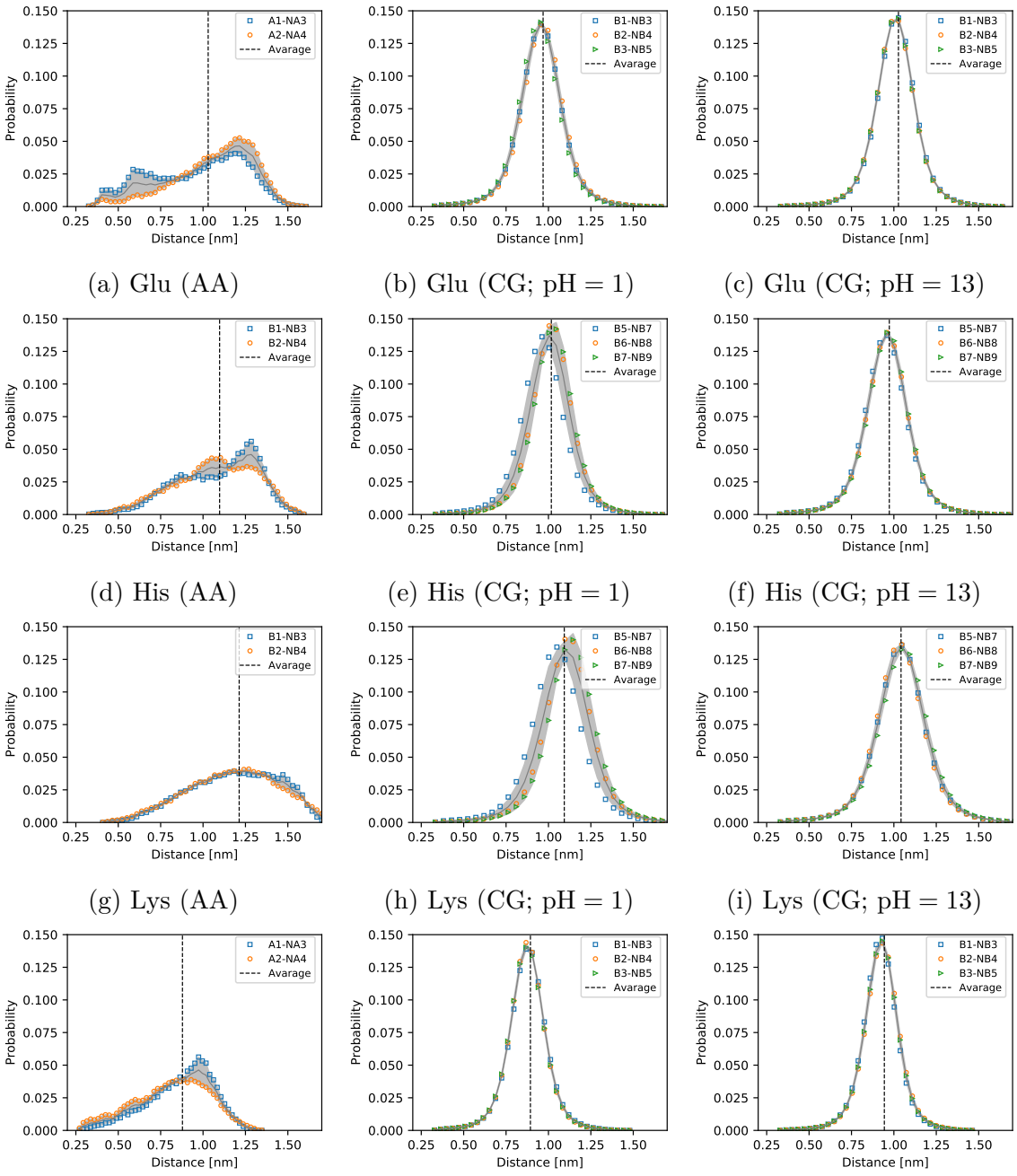
Peptide	Acid Amino acid	distance [nm]	Base Amino acid	distance [nm]
$Glu_5 - His_5$	Glu (AA)	0.92 ± 0.16	His (AA)	0.93 ± 0.18
	Glu (CG; pH = 1)	0.80 ± 0.11	His (CG; pH = 1)	0.86 ± 0.11
	Glu (CG; pH = 13)	0.84 ± 0.11	His (CG; pH = 13)	0.82 ± 0.12
	Glu (Avogadro)	1.04	His (Avogadro)	0.91
$Lys_5 - Asp_5$	Asp (AA)	0.71 ± 0.14	Lys (AA)	1.19 ± 0.18
	Asp (CG; pH = 1)	0.72 ± 0.09	Lys (CG; pH = 1)	0.97 ± 0.14
	Asp (CG; pH = 13)	0.75 ± 0.16	Lys (CG; pH = 13)	0.92 ± 0.27
	Asp (Avogadro)	0.93	Lys (Avogadro)	1.14

in all cases, but the width of this peak depends on the pH and on the type of the amino acid. This difference between the CG and AA distributions can be again attributed to specific details of the atomistic structure, which were not fully included in the CG simulations.

1.3.4 Distances between the next-nearest-neighbour side-chain beads.

Finally, we compared the distances between charged groups on next-nearest-neighbour amino acids, measured in the AA and CG simulations. Table S4 reveals an even greater statistical uncertainty of the average value, approx. 10–20%. Within the given uncertainty, the CG and AA simulations still agree with each other. However, the distances depend on pH: the distances measured in the uncharged state are systematically lower than those measure in the charged state, although this difference is on the verge of the estimated statistical error. The distances between from AA simulations are better matched by CG results at a pH which corresponds to the charged state of the respective group.

This difference can be explained by the electrostatic repulsion between charged groups,



(a) Glu (AA) (b) Glu (CG; pH = 1) (c) Glu (CG; pH = 13)

(d) His (AA) (e) His (CG; pH = 1) (f) His (CG; pH = 13)

(g) Lys (AA) (h) Lys (CG; pH = 1) (i) Lys (CG; pH = 13)

(j) Asp (AA) (k) Asp (CG; pH = 1) (l) Asp (CG; pH = 13)

Figure S5: Distribution of distances between charged group on next-nearest-neighbour amino acids from the AA and CG simulations (the latter at the extreme pH values). The vertical line shows the average distance after averaging over individual pairs. These average values are listed in Table S4.

Table S4: Average distance between the charged group on next-nearest-neighbour amino acids in the AA simulations; Distance between the same atoms determined using energy minimisation in the Avogadro software; Average distance between the A and next-nearest-neighbour A beads, or B and next-nearest-neighbour B bead, in the CG simulations at the extreme pH values.

Peptide	Acid Amino acid	distance [nm]	Base Amino acid	distance [nm]
$Glu_5 - His_5$	Glu (AA)	1.03 ± 0.26	His (AA)	1.10 ± 0.24
	Glu (CG; pH = 1)	0.97 ± 0.14	His (CG; pH = 1)	1.02 ± 0.14
	Glu (CG; pH = 13)	1.03 ± 0.13	His (CG; pH = 13)	$0.97 \pm 0..14$
	Glu (Avogadro)	0.74	His (Avogadro)	0.73
$Lys_5 - Asp_5$	Asp (AA)	0.88 ± 0.21	Lys (AA)	1.21 ± 0.27
	Asp (CG; pH = 1)	0.89 ± 0.12	Lys (CG; pH = 1)	1.09 ± 0.16
	Asp (CG; pH = 13)	0.94 ± 0.12	Lys (CG; pH = 13)	1.04 ± 0.16
	Asp (Avogadro)	0.74	Lys (Avogadro)	0.73

which is absent in the neutral state. Furthermore, we observe that the structures obtained by energy minimisation using the Avogadro software yield significantly lower average distances than any of the simulation models. The distributions from AA simulations in Fig. S5 are all single-peaked and rather broad, while the distributions from CG simulations are narrower and more symmetric. Moreover, the next-nearest-neighbour distances are only slightly greater than the nearest-neighbour distances, demonstrating that the side chains prefer the trans conformation. In turn, the role of chain flexibility is demonstrated by the next-nearest-neighbour distances, which are significantly greater than the distances in all-trans conformations determined from Avogadro. Earlier on, we tested a simpler model in which each amino acid was represented by just one bead. This one-bead model was able to reproduce the nearest-neighbour distances, while yielding next-nearest-neighbour distances approximately twice as large as the nearest-neighbour distances. The two-bead model can approximate well these two distances. They are crucial to account for the electrostatic effect on the ionization of peptides, which suggests that the two-bead model is suited for

quantitative predictions of this effect.

1.4 Computational demands and costs

Various simulations described above dramatically differ in computational demands. Our CG simulations typically required approximately 5 hours of computer run time on a single CPU core for each data point of the titration curve. Thus, one titration curve with 20 data points could be obtained within approximately 100 CPU core-hours. Because all these simulations were independent, they could be run simultaneously on different CPUs, thereby making it possible to obtain the full titration curve within a few hours when run on a small computer cluster.

In contrast, the AA simulations of 100 ns required approximately 5 days on 8 CPU cores, equivalent to approximately 1000 CPU core-hours. However, running one simulation per amino acid was enough to obtain the desired parameters of the CG model. When considering more complicated peptide sequences, it may be necessary to perform such an AA simulation for each pair of amino acids. Thus, running the AA simulations to parametrise the CG model is much more demanding than obtaining the whole titration curve from CG simulations, although the former can be considered a moderate computational demand. In addition, energy minimisation provides a much cheaper and faster way of estimating bond lengths for the CG model. This process is completed within several seconds on a desktop PC, providing parameter values for the amino acids studied here similar to those determined by expensive simulations. Because the energy minimisation could fail for other structures, its predictions should always be verified using an all-atom simulation.

Thus, parametrisation of the model can be completed within one week and, once the model parameters are available, titration curves of various peptide sequences can be obtained within hours. The economic cost of the simulations could be estimated by noting the commercial prices, approximately 0.05 EUR per CPU core-hour. Hence, we estimate the cost of parameterising one amino acid for the CG model as 50 EUR, and the cost of

predicting one titration curve from the CG model as 5 EUR. The above costs include direct and indirect costs of computer time, but they do not include the cost of several person-hours needed for running the simulations and analysing the results.

The cost of performing the simulations should be compared to the cost of purchasing 100 mg of one custom-synthesized peptide, approximately 400 EUR. Thus, merely purchasing the sample to start the experiments is more costly than parametrising the model and obtaining the CG simulation results. Thus, the experimental quantification of the ionisation response is much more expensive than the simulations due to additional instrument time and personnel costs required to perform the experiments.

Notably, the above cost estimation assumes that the required protocols for running the simulations, performing the experiments and analysing the results are readily available and that all steps can be performed routinely. It does not include the costs and effort needed to establish these protocols.

2 Experiments

2.1 Materials

The following peptides with acetyl and amide terminal groups and trifluoroacetate (TFA) as counterion were purchased from Biomatik LLC, Wilmington, Delaware, USA: Ac-E5-H5-NH2 (Glu_5-His_5; $M = 1390.33\,g/mol$; lot number P180808-DG671108 97.16% HPLC purity for CZE and lot number P190902-LL671108 97.54% HPLC purity for potentiometric titrations and NMR); Ac-K5-D5-NH2 ($Lys_5 - Asp_5$; $M = 1275.36\,g/mol$; lot number P180711-JQ665893 95.83% HPLC purity for CZE and lot number P190816-LC665893 96.99% HPLC purity for titrations and NMR). All peptides were purified, and HPLC and MS spectra were measured for all peptide sequences.

The buffers used for CZE experiments were prepared by mixing of weak acid with strong base (pH less than 7) and by mixing of strong acid with weak base (pH more than 7).

Compositions of the used buffers are specified in Table S5.

Standardised solutions of HCl and NaOH from Carl Roth GmbH (Karslruhe, Germany) were used to prepare 0.1 M stock solutions, which were subsequently diluted to 0.01 M. To prevent contamination by CO_2, the standardised solutions were kept under soda lime at least 24 hours before the measurements.

Deuterium oxide 99.8 % purity with a trace of 3-(trimethylsilyl)-1-propanesulfonic acid sodium salt (DSS) of 97 % purity from Sigma-Aldrich was used for field-frequency lock.

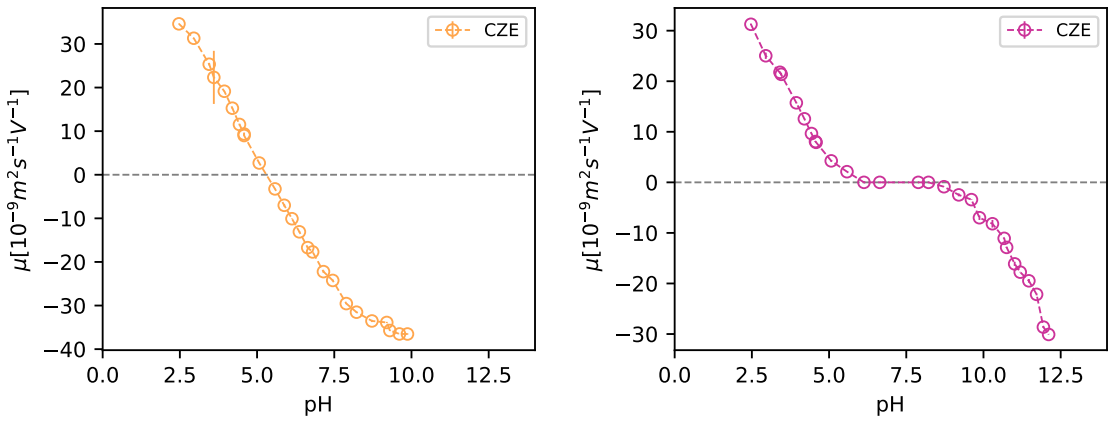
2.2 Capillary Zone Electrophoresis experiments (CZE)

2.2.1 Instrumentation and experimental protocol

All CZE experiments were performed using Agilent 7100 capillary electrophoresis equipment operated under ChemStation software (Agilent Technologies, Waldbronn, Germany). Detection was performed with the built-in diode array detector (DAD). Fused fluorocarbon capillaries (50 μm i.d., 375 μm o.d.) by Agilent Technologies with a total length of 50 cm and effective length to the DAD detector of 41.5 cm were used to perform the experiments. Before the first use, each new capillary was flushed with 0.1 M sodium hydroxide from Agilent Technologies and then with deionised water for 10 min. All CZE measurements were performed in the running background electrolyte (BGE) with an ionic strength of 10 mM. The samples of peptide solutions with a concentration of monomeric units

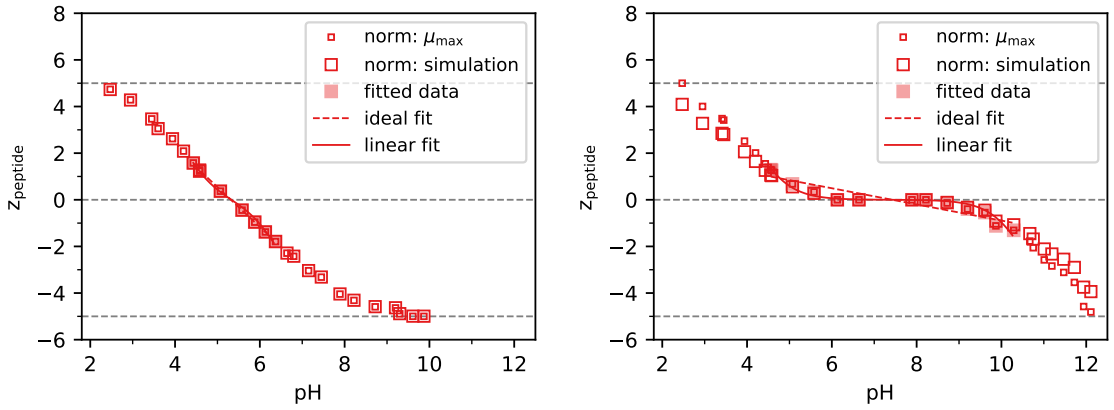
Table S5: The buffers prepared for CZE measurements (ionic strength is always 10 mM).

Acid	Base	pH range
Formic	Lithium hydroxide	2.5 - 3.5
Acetic	Lithium hydroxide	3.9 - 4.6
Cacodylic	Lithium hydroxide	4.6 - 7.2
Hydrochloric	Tris(hydroxymethyl)aminomethane	7.5 - 9.2
Hydrochloric	Ammonium hydroxide	9.3 - 10.7
Hydrochloric	Triethylamine	10.3 - 12.0



(a) $Glu_5 - His_5$ (b) $Glu_5 - His_5$

Figure S6: Electrophoretic mobilities of the peptides determined from CZE experiments.



(a) $Glu_5 - His_5$ (b) $Lys_5 - Asp_5$

Figure S7: Total charge on the peptides calculated from electrophoretic mobilities determined by CZE. Small squares represent the data re-normalised by $|\mu^{max}(pH)|$. Large squares represent the data re-normalised using $\alpha(pH)$ obtained from simulations at $|\mu^{max}(pH)|$. Filled squares show the data points which were fitted by the ideal titration (solid line), and by a line (dashed line). Isoelectric points determined by both fits coincide at least within two significant figures.

[Glu] = [His] = [Asp] = [Lys] =0.5 mM were prepared by dissolving the relevant amount of peptide directly in the running buffer. The PeakMaster 5.3 was used to calculate the properties of all buffers used in CZE experiments.[S6,S7] These buffers are listed in Table S5. We used pure 10 mM HCl and LiOH only at the lowest and highest pH (pH = 2.0 and pH = 12.1, respectively). All running buffers were filtered with 0.45 μm PVDF membranes. The samples were injected hydrodynamically using the pressure of 30 mbar for 5 s; the applied voltage was always ± 10 kV. The solutions were thermostated at 25°C. Capillary was flushed by running buffer for 3 min before each measurement, and each run was repeated three times. DAD detection was performed at a wavelength of 200 nm. The CEVal software[S8] was used to analyse the raw data and to determine the effective mobility.

2.2.2 Determination of charge on the peptide

From CZE, we obtained the absolute electrophoretic mobilities, as shown in Fig. S6. To obtain the charge on the peptide from the mobilities, we first renormalized the absolute mobilities by their maximum values for the given peptide, $\mu^{\max} = \max(|\mu(\mathrm{pH})|)$.

$$z_{\mathrm{cze}}(\mathrm{pH}) = \frac{\mu(\mathrm{pH} - \mathrm{pI}_{\mathrm{cze}})}{|\mu^{\max}|} \tag{4}$$

This renormalisation procedure is based on the assumption that diffusion coefficients of the peptides do not significantly change with pH, as confirmed independently by DOSY NMR (Fig. S20). Ideally, one should observe a plateau in the mobility at a high or low pH value, indicating that the peptide is fully ionised. However, we were not always able to measure the maximum mobility at a pH value, which would ensure that the peptide was fully ionised. Therefore, such a normalisation could not yield a reliable value of the charge on the peptide. To correct this deficiency, we renormalised the $\mu^{\max}(\mathrm{pH}^{\max})$ by the peptide charge determined

from the simulations, $z_{\mathrm{sim}}(\mathrm{pH}^{\mathrm{max}} + \Delta\mathrm{pI})$

$$z_{\mathrm{cze}}(\mathrm{pH}) = \frac{\mu(\mathrm{pH} - \mathrm{pI}_{\mathrm{cze}})}{\mu^{\mathrm{max}}(\mathrm{pH}^{\mathrm{max}})} z_{\mathrm{sim}}(\mathrm{pH}^{\mathrm{max}} + \Delta\mathrm{pI}) \tag{5}$$

where $\Delta\mathrm{pI} = \mathrm{pI}_{\mathrm{cze}} - \mathrm{pI}_{\mathrm{sim}}$. The effect of different renormalisations is shown in Fig. S7. The value of $z_{\mathrm{cze}}(\mathrm{pH})$ of $\mathrm{Glu}_5 - \mathrm{His}_5$ is barely affected by the different normalisation because the $\mu^{\mathrm{max}}(\mathrm{pH}^{\mathrm{max}} = 10)$ is well in the plateau region where the peptide is fully ionised. The value of $z_{\mathrm{cze}}(\mathrm{pH})$ of $\mathrm{Lys}_5 - \mathrm{Asp}_5$ is visibly affected by the different normalisation because the $\mu^{\mathrm{max}}(\mathrm{pH}^{\mathrm{max}} \approx 2.5)$ is still in the region where the peptide should not be fully ionized.

We note that we also attempted to compute the charge on the peptides without renormalisation, using the diffusion coefficients determined from NMR (Fig. S20). However, this attempt yielded a significantly lower peptide charge, which was not consistent with other methods (simulations, NMR, and titrations). This observation is in line with the notion that electrophoretic mobility yields the effective charge, rather than the bare charge of the analyte.[S9–S11] The difference between the effective and bare charge increases with the increase in the charge on the peptide, particularly affecting the result at high and low pH values.

2.2.3 Determination of the isoelectric point

To determine the isoelectric point from the CZE data, we determined the intersection of $z_{\mathrm{cze}}(\mathrm{pH})$ with $z = 0$ by fitting the date in the range $z \in \{-2, 2\}$ for he $\mathrm{Glu}_5 - \mathrm{His}_5$ peptide, and $z \in \{-1.5, 1.5\}$ for he $\mathrm{Lys}_5 - \mathrm{Asp}_5$ peptide. To ensure that our result was not affected by the arbitrary choice of the fitting range, we tested various ranges and fitted the date using the ideal titration curve and a straight line. Because both fit functions were symmetric around the pI, and the data was approximately symmetric as well, all fits yielded consistent values of pI within two significant figures. Additionally, the use of symmetric fit functions ensures that the determined isoelectric point is not affected by the renormalisation of the effective mobility. The curves obtained from the fits and the data points used in the fits are shown

in Fig. S7. We avoided extending the range to higher values of z because the fit functions were not appropriate approximations, and the fit results were significantly affected by a few data points that were farther from the target value $z = 0$. The precise determination of pI of $Lys_5 - Asp_5$was particularly tricky because its charge is almost zero in a broad pH range.

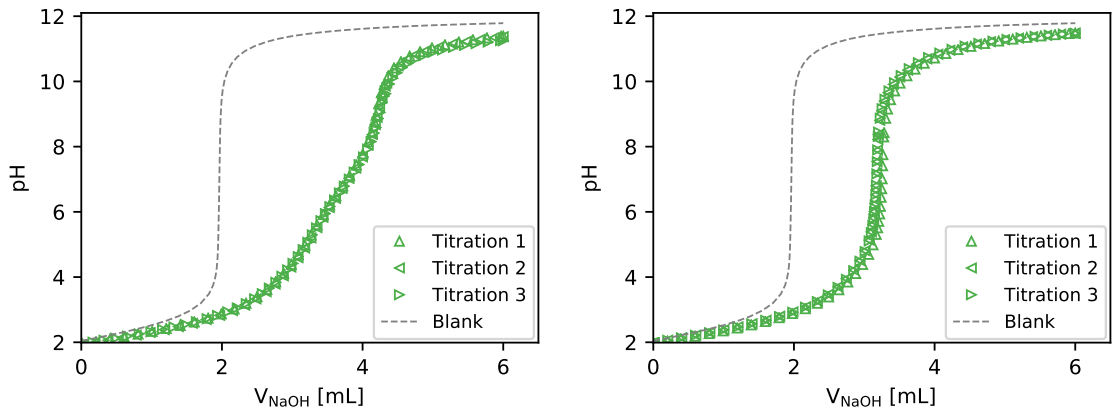
2.3 Potentiometric Titration

2.3.1 Instrumentation and experimental protocol

Potentiometric titrations were performed using a Metrohm 888 Titrando Compact titrator equipped with a Metrohm LL Biotrode 3 mm glass electrode, a Pt1000 temperature sensor, a titration vessel for 1 ml, magnetic stirrer and Titrando Software. In order to prevent the absorption of carbon dioxide from the air, standardised solutions of HCl and NaOH from Carl Roth GmbH (Karslruhe, Germany) were used to prepare 0.1 M stock solutions, which were subsequently diluted to 0.01 M. Moreover, the standardised solutions were kept under soda lime at least 24 hours before the measurements. The solutions of the peptides were prepared at concentrations of monomeric units [Glu] = [His] = [Asp] = [Lys] = 5 mM dissolved in 0.01 M standardised HCl. Sample volumes of approximately 2 ml were weighed to determine the precise amount and then titrated by standardised 0.01 M NaOH using an automated dynamic pH titration method with signal drift 1 mV and waiting time $10 - 50$ s. To prevent excessive contamination by CO_2, the stock solutions were kept under soda lime, and the titration vessels were sealed during the titration, for 15–90 min, depending on the sample. Blank titrations were performed under the same conditions, before and after each peptide titration, to estimate the reproducibility and reliability of the procedure and to estimate the concentration of CO_2 in the stock solution.

2.3.2 Determination of the charge on the peptide

The primary output of titration is the solution pH as a function of the volume of the added NaOH (V_{NaOH}), as shown in Fig. S8. To calculate the charge on the peptide, we used the



(a) $Glu_5 - His_5$ (b) $Lys_5 - Asp_5$

Figure S8: Potentiometric titration of the peptides.

electroneutrality condition

$$\sum_i z_i c_i = 0 \tag{6}$$

where the summation runs over all ionic species in the system.

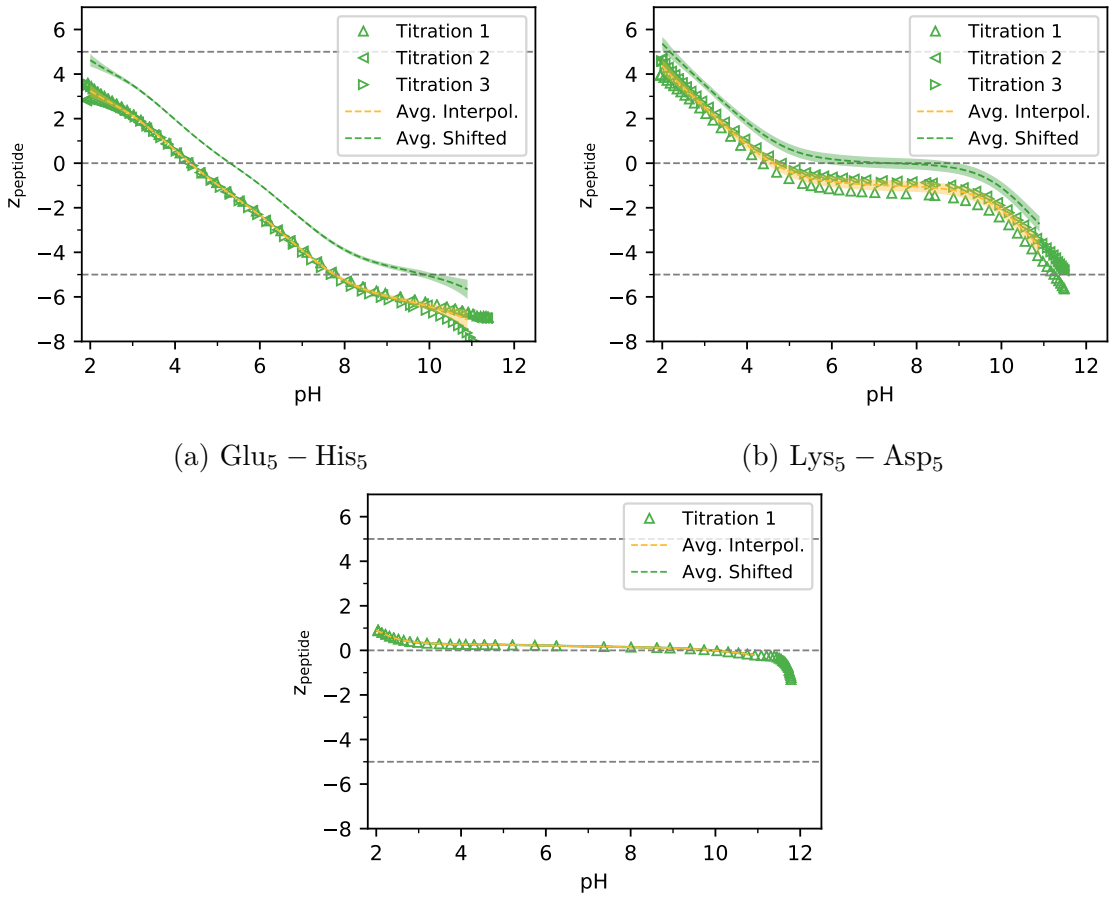
$$z_{\text{titration}} = \frac{(V_{\text{HCl}}((c_{\text{H}} - c_{\text{OH}}) + c_{\text{HCl}}) - V_{\text{NaOH}}(c_{\text{H}} - c_{\text{OH}} + c_{\text{NaOH}}))}{(c_{\text{peptide}} V_{\text{NaOH}})} - x_{\text{TFA}} \tag{7}$$

where c stands for concentration, V for volume, $z_{\text{titration}}$ is the charge on the peptide determined from titration, $x_{\text{TFA}} = n_{\text{TFA}}/n_{\text{base}}$ is the mole fraction of the trifluoroacetate counterions contained in the peptide sample, relative to the number of basic side-chains on the peptide, V_{HCl} is the initial volume of HCl in which the peptide was dissolved. The concentrations c_{H} and c_{OH} were calculated from the measured pH and from the pK_{w}, accounting for the variation of both quantities with temperature and assuming that activity coefficients are equal to one. When assuming that $x_{\text{TFA}} = 1$, this procedure yielded the values of $z_{\text{titration}}(\text{pH})$, which were similar to $z_{\text{cze}}(\text{pH})$, albeit shifted to lower values of z. As shown in Fig. S9, repeated runs of the same titration yielded highly reproducible $z_{\text{titration}}(\text{pH})$, except for pH $\lesssim 3$ and pH $\gtrsim 10$.

We used blank titrations, as shown in Fig. S9c, to assess the reliability and reproducibility

of the calculation of $z_{\text{titration}}(\text{pH})$. These blank titrations were performed before and after each set of titrations with a peptide sample. Ideally, they should yield $z_{\text{titration}}(\text{pH}) = 0$ in the whole range. The example of a typical blank run shown in Fig. S9c highlights that the yielded $|z_{\text{titration}}(\text{pH})| \lesssim 0.1$, except for pH $\lesssim 3$ and pH $\gtrsim 11$. In the high- and low-pH range, the calculation of $z_{\text{titration}}(\text{pH})$ is very sensitive to the precision of the pH measurement and to the value of $\text{p}K_{\text{w}}$, as evidenced by the steep increase or decrease in $z_{\text{titration}}(pH)$ in Fig. S9c. The effect of CO_2 is noticeable at pH $\gtrsim 10$, and this effect was stronger in longer titrations. Therefore, this effect was weak in the blank titration, which was quick, and stronger in the titration of peptides, which were slower.

Because the blank titrations did not show the shift observed in the peptide titrations, we attributed this shift to the unknown excess of TFA anions contained in the peptide samples. Indeed, the excess TFA, commonly found in peptide samples after deprotection from the BOC groups during solid-state synthesis, results in $x_{\text{TFA}} > 1$. This assumption was supported by the results from the titration of a different batch of the same peptide, which yielded slightly shifted titration curves (not shown). Unfortunately, the amount of peptide samples was too low to quantitatively analyse the TFA content. To correct for the unknown amount of TFA, we first interpolated titration data from different runs and then used the interpolated data to compute the average $z_{\text{titration}}(\text{pH})$ and to estimate the accuracy using the standard deviation of different runs. Then, we used the isoelectric point determined from CZE to adjust the value of x_{TFA}, so that $z_{\text{titration}}(\text{pH} = \text{pI}_{\text{cze}}) = 0$. The results of all intermediate steps of the titration data processing are shown in Fig. S9.



(a) $Glu_5 - His_5$

(b) $Lys_5 - Asp_5$

(c) Blank

Figure S9: Total charge on the peptides and on the blank control, calculated from potentiometric titrations. The Yellow line shows the interpolated average over different runs. The green line show the average shifted by adjusting x_{TFA} so that the curves matched the isoelectric point determined from CZE.

2.4 NMR

2.4.1 NMR measurements and instrumentation

All NMR data were recorded using a Bruker AVANCE III spectrometer operating at the proton Larmor frequency of 600 MHz equipped with a cryogenically cooled probe and stabilising the temperature at 25 °C. The samples were prepared by dissolving the peptides in 10 mM HCl to obtain 15 gL^{-1} peptide concentration. The pH was adjusted by adding NaOH. A capillary insert containing deuterium oxide with a trace of sodium trimethylsilylpropanesulfonate (DSS) was used for field-frequency lock and chemical shift referencing. ^{1}H spectra were acquired with water suppression using the excitation sculpting method.[S12] Measurements of translational diffusion coefficients were performed with the double stimulated echo experiment with bipolar pulse field gradients described by,[S13] combined with water suppression. The gradients were 1.5 ms long with 24 linearly spaced amplitudes spanning the range $0 - 60\,Gcm^{-1}$, and the diffusion time was 300 ms. The calibration was performed using a standard sample of 1% H_2O in D_2O (doped with $GdCl_3$), for which the value of the HDO diffusion coefficient at 25 °C is $1.9 \times 10^{-9}\,m^2s^{-1}$. All data processing and fitting of the diffusion coefficients has been done using the MestReNova and GNAT software.[S14] The chemical shifts were determined by referencing to the signal of DSS for ^{1}H NMR and of TFA for ^{13}C NMR. At a very low pH the signals of TFA were affected by its ionization. Therefore, the chemical shifts of peptides at low pH were referenced to the chemical shifts of CH_3 terminal groups of the peptides. The MestReNova Software was used to analyse both 1D and 2D spectra, including the determination centers of mass of multiplets and the ranges of the peaks.

2.4.2 Degree of ionisation from the NMR spectra

Fig. S10 and S13 show ^{1}H and ^{13}C spectra for Glu-His and Lys-Asp, respectively, at pH ranging from 1 to 13, with the chemical structure of the amino acids and the assignment of

all peaks. 2D NMR spectra, COSY and ^{1}H-^{13}C HSQC, (Fig. S16 and S17) at pH 2 were used for peak assignment. Specific atoms have been identified in literature as "good reporters" of ionisation, that is, their chemical shifts predominantly reflect ionisation changes on the nearby ionisable group.[S15] Typically, good reporters are located far from the backbone and as close as possible to the ionisable group. Following Ref.,[S15] we used CB for Asp, CD and CG for Glu, CG and CE1 for His, and CD and CE for Lys. We were able to identify two good reporters for each amino acid except for Asp, for which we were able to identify only one good reporter because the signal of CG overlaps with carbonyl signals from the backbone. Details of the spectra, highlighting how the peaks shift with the pH, are shown in Fig.S11, Fig.S12, Fig.S14 and Fig.S15.

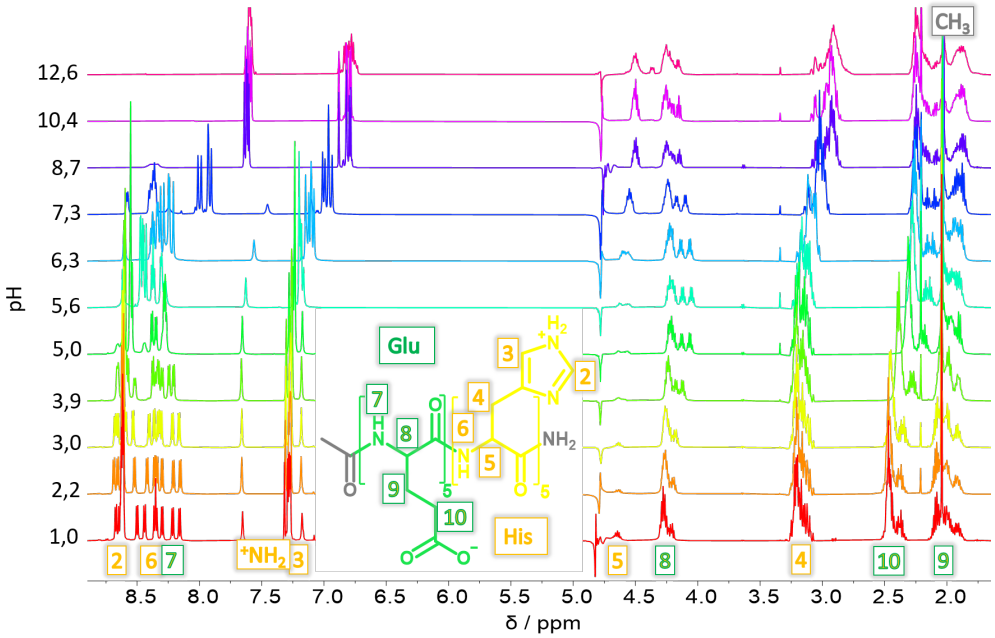
The chemical shifts of the good reporters, which could be unambiguously identified in the spectra, were used to calculate the degree of ionisation of each amino acid in the oligopeptide. These peaks typically consisted of multiple sub-peaks, reflecting the fact that same amino acids in different positions in the peptide chain were not equivalent. To determine the average degree of ionisation of each type of amino acid, we use the centre of mass of the corresponding peak, which should be equivalent to averaging the degree of ionisations of 5 amino acids of the same type. We calculated the degree of ionisation by normalising the chemical shifts as follows:

$$\alpha_{base} = \frac{\delta_{max} - \delta}{\delta_{max} - \delta_{min}} \tag{8}$$

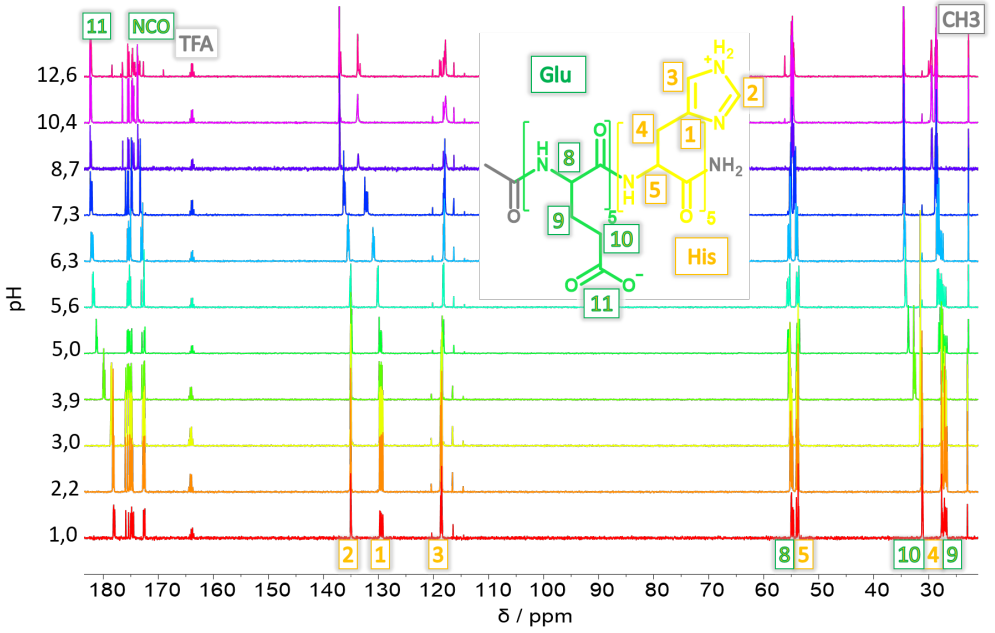
for bases (Lys and His) and

$$\alpha_{acid} = \frac{\delta - \delta_{min}}{\delta_{max} - \delta_{min}} \tag{9}$$

for acids (Asp and Glu).



(a) ^{1}H



(b) ^{13}C

Figure S10: NMR spectra of $Glu_5 - His_5$ at various pH.

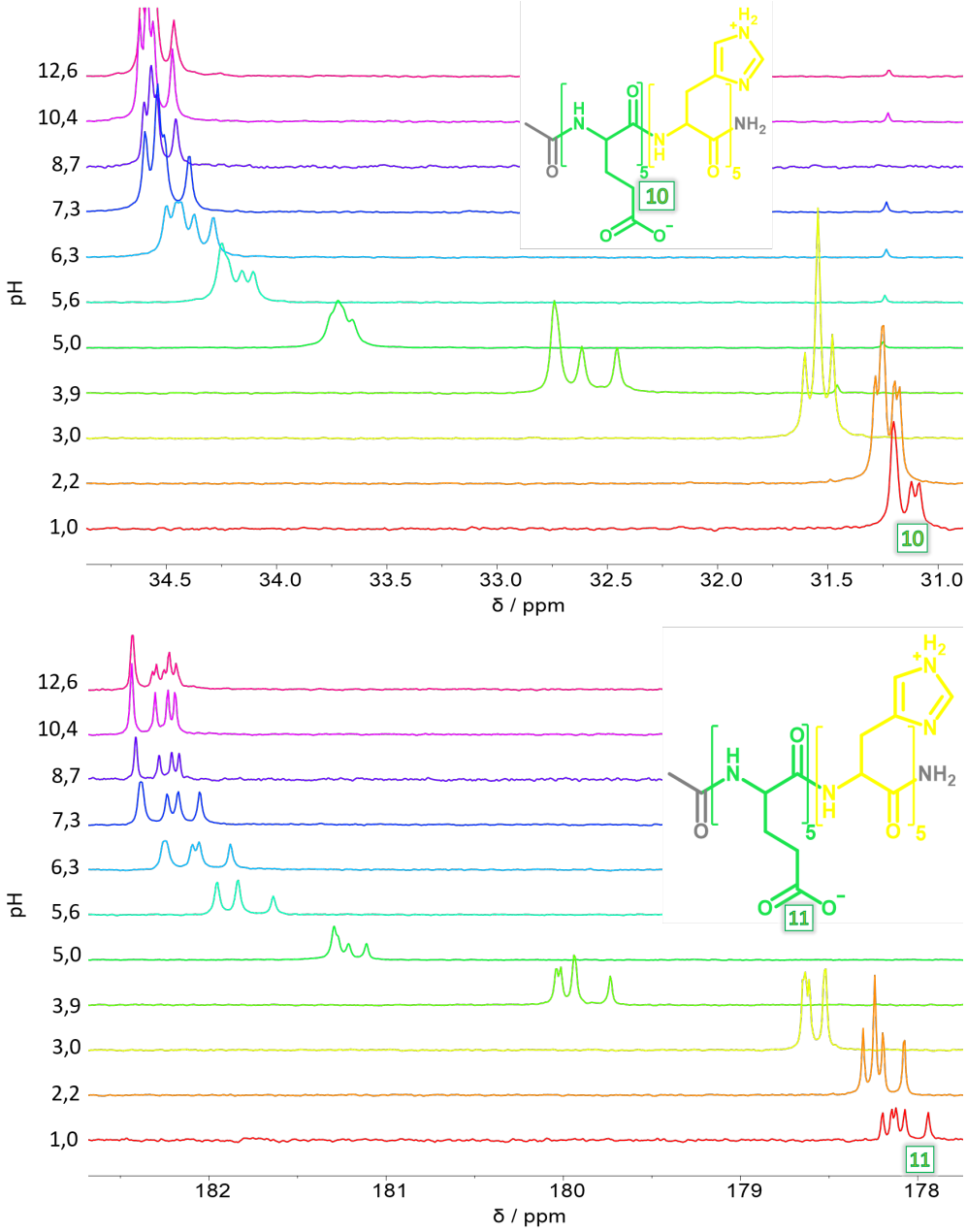


Figure S11: Details of NMR spectra of good reporters for Glu in $Glu_5 - His_5$ at various pH.

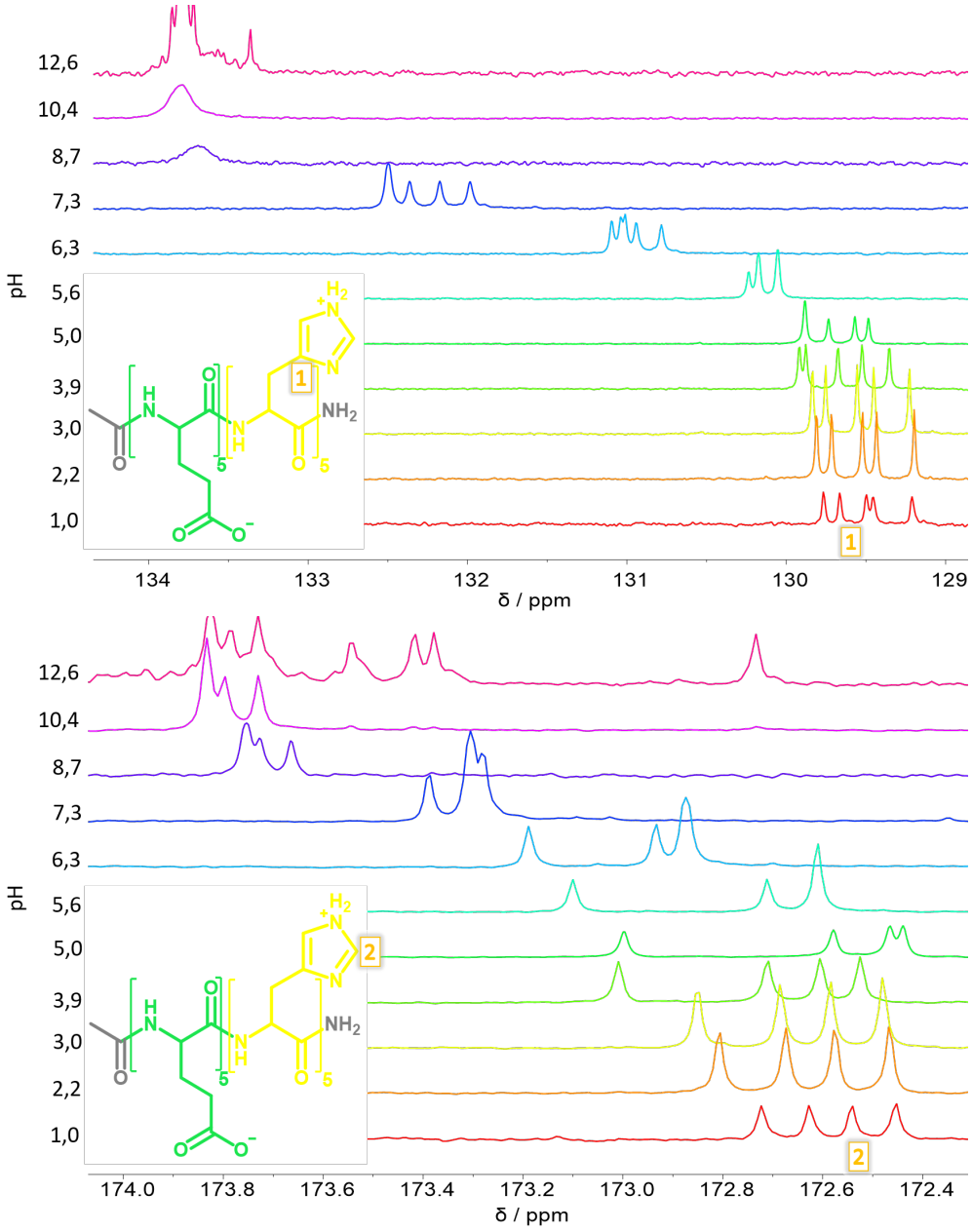
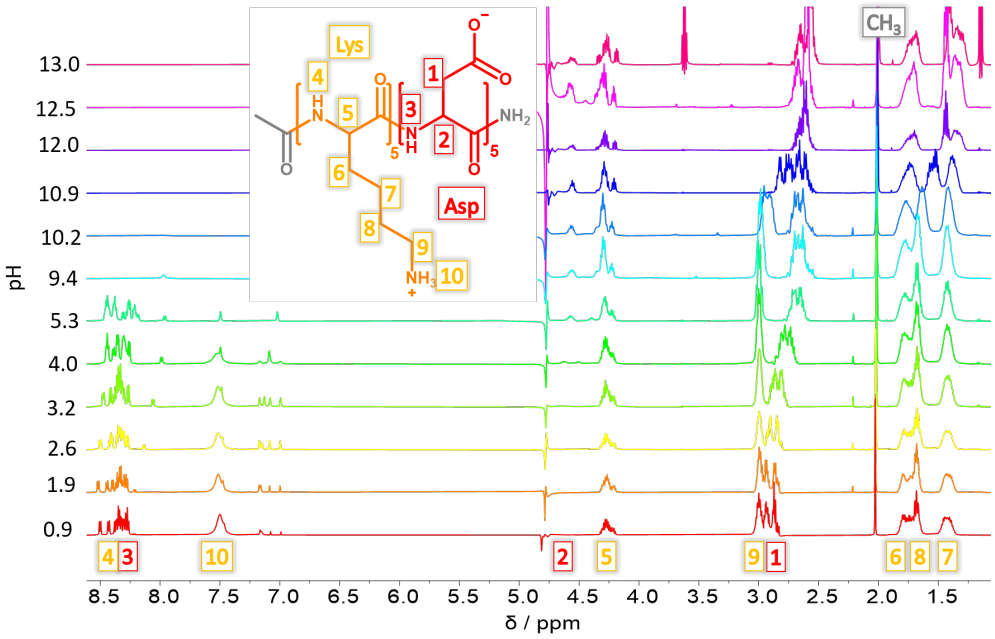
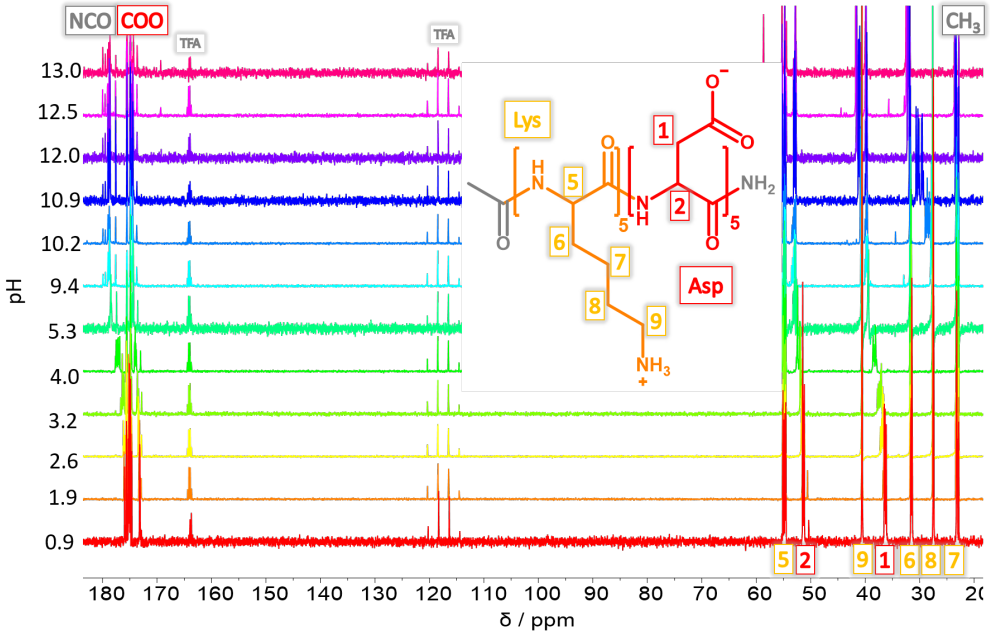


Figure S12: Details of NMR spectra of good reporters for His in $Glu_5 - His_5$ at various pH.



(a) ^{1}H



(b) ^{13}C

Figure S13: NMR spectra of $Lys_5 - Asp_5$ at various pH.

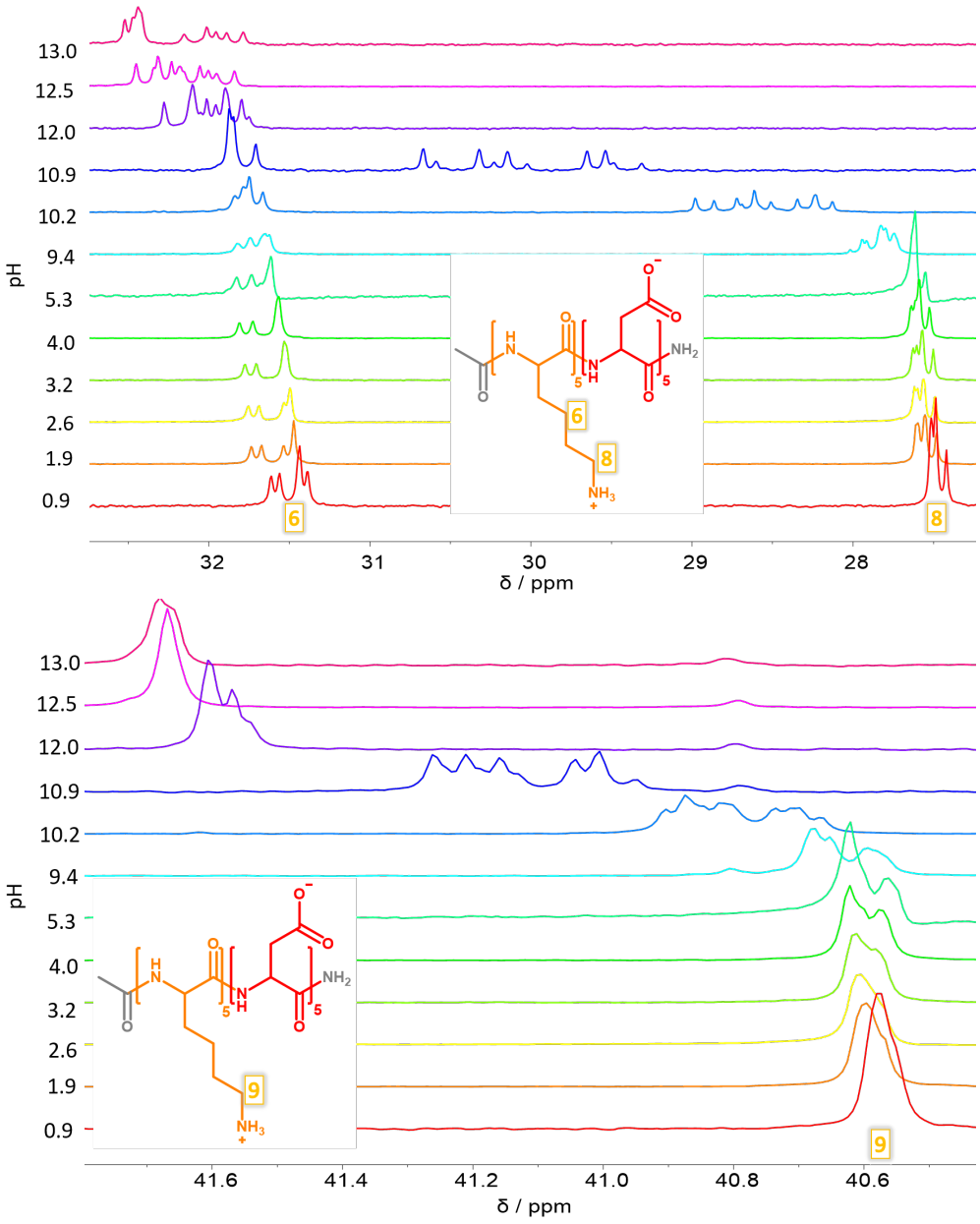


Figure S14: Details of NMR spectra of good reporters for Lys in $Lys_5 - Asp_5$ at various pH.

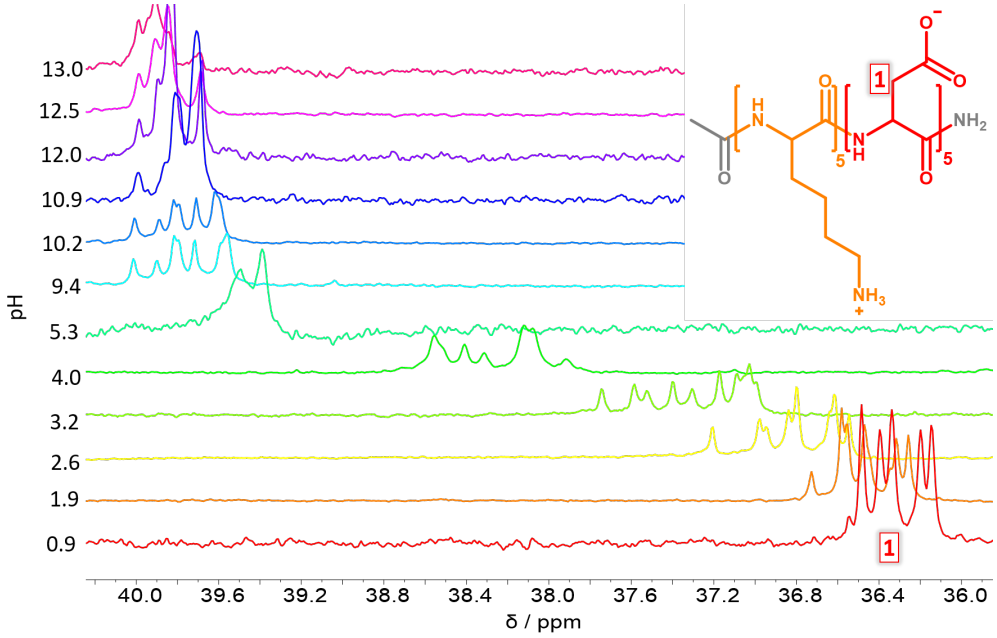
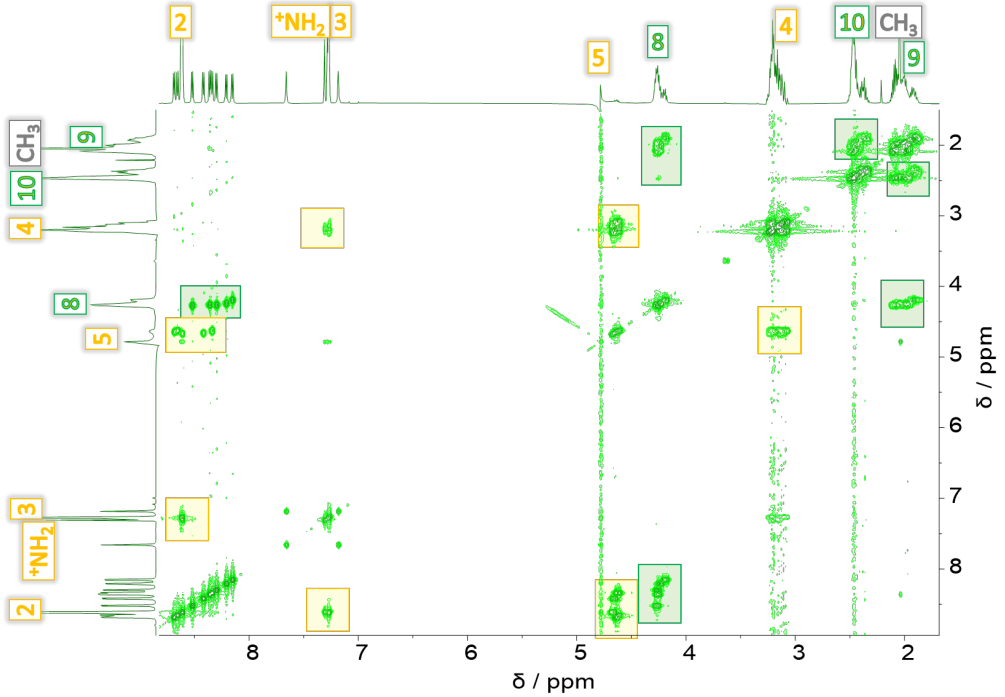
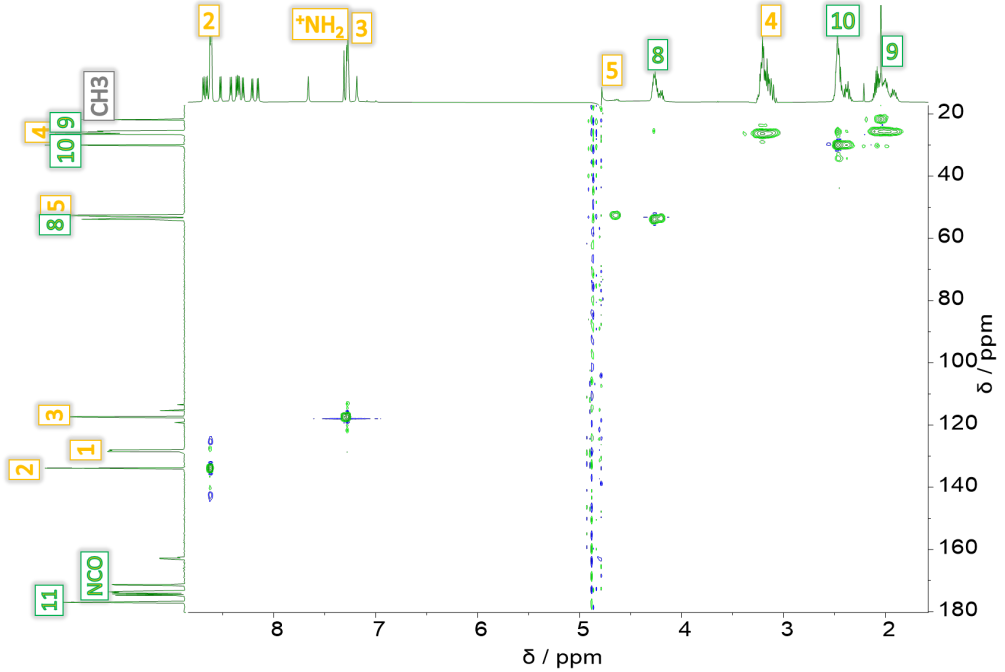


Figure S15: Details of NMR spectra of good reporters for Asp in $Lys_5 - Asp_5$ at various pH.

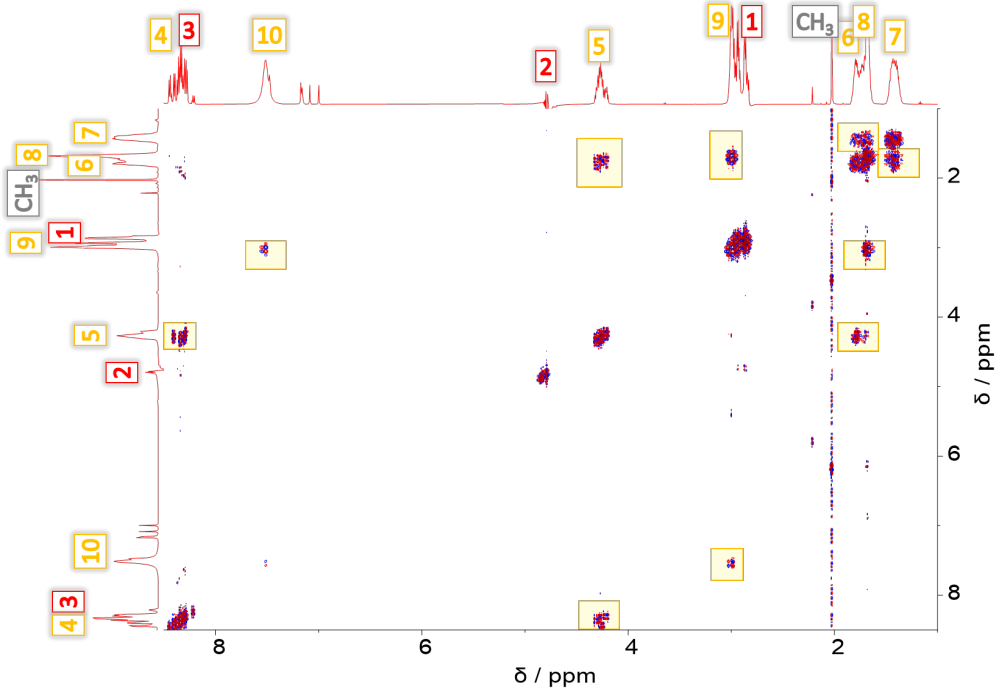


(a) COSY

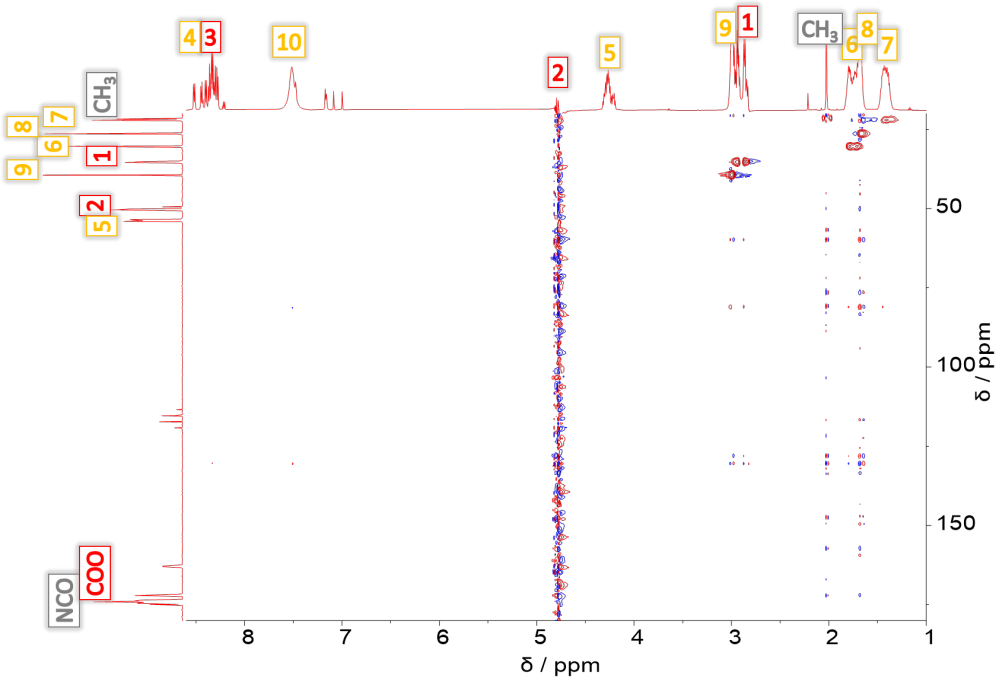


(b) ^{1}H-^{13}C HSQC

Figure S16: 2D NMR spectra of $Glu_5 - His_5$ at pH=2.

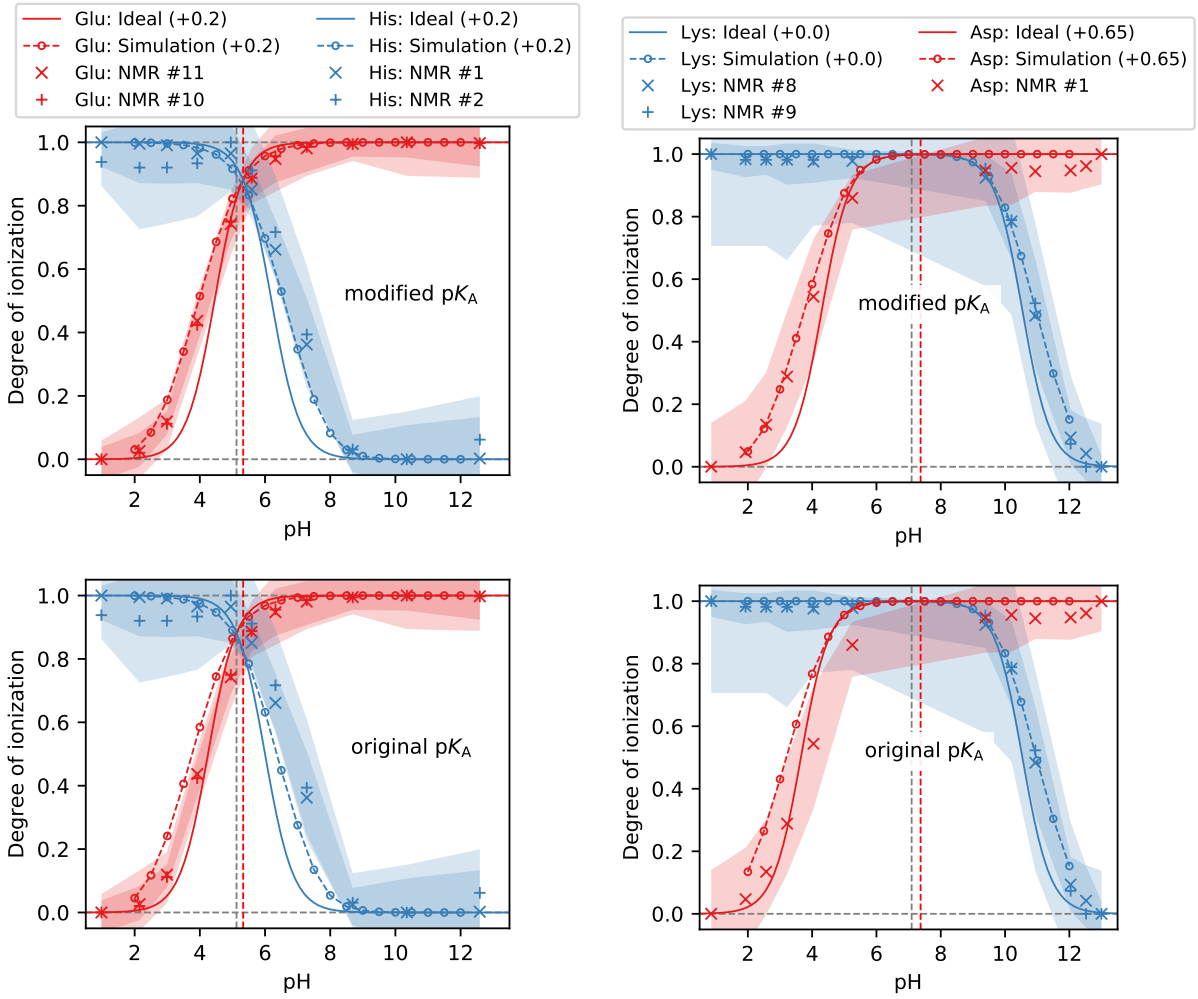


(a) COSY



(b) ^{1}H-^{13}C HSQC

Figure S17: 2D NMR spectra of $Lys_5 - Asp_5$ at pH=2.



(a) $Glu_5 - His_5$ (b) $Lys_5 - Asp_5$

Figure S18: The degree of ionisation of acid and base groups on the peptides predicted from simulations, determined from NMR measurements and from the ideal titration curves. Individual reporter atoms are indicated in the legend. Shaded areas indicate the spread of peaks in the NMR spectra. Red and grey vertical lines represent the isoelectric point determined from CZE and the ideal isoelectric point. Panels in the top row show the ideal and simulation data using the modified pK_A values to correct for the effect of incorporating amino acids into the peptide (see Table S6). Numbers in the legend indicate the reporter atom id in NMR (see Fig. S10 and Fig. S13) and the amount by which the pK_A values were modified in the simulations. The bottom row shows the original uncorrected data for comparison (denoted as "Lit1" in Table S6).

Fig.S18 shows that the degree of ionisations determined using different good reporters agree with each other. However, individual reporters may considerably differ in the spread of the peaks due to the non-equivalence of amino acids of the same type. The panels in the bottom row of Fig.S18a show that the degree of ionisations of both Glu and His exhibit the same trend as that assessed in simulations, albeit shifted to higher pH values, in line with the shift of the isoelectric point between the ideal and CZE results ΔpI. In contrast, Fig.S18b shows that the degree of ionisation of Asp is shifted with respect to the simulation results approximately twice as much as the difference in isoelectric points, while the degree of ionisation of Lys matches the simulations. We attributed these differences to the uncertainty in the pK_A values caused by different substituents. To assess this hypothesis, we performed a new set of CG simulations using a modified set of pK_A values: In $Glu_5 - His_5$, we increased the pK_A of both Glu and His by ΔpI; In $Lys_5 - Asp_5$, we increased the pK_A of Lys by 2ΔpI (see Table S6). Consequently, the isoelectric point from simulations using the modified pK_A values matches pI_{cze}. With the modified pK_A, we obtained an almost perfect match between the simulation and the NMR results, as shown in the top row of the panels in Fig.S18b. This observation supports our hypothesis that the previously observed differences between the pI from simulations and experiments could be attributed to the uncertainty in choosing the right pK_A values for the simulation.

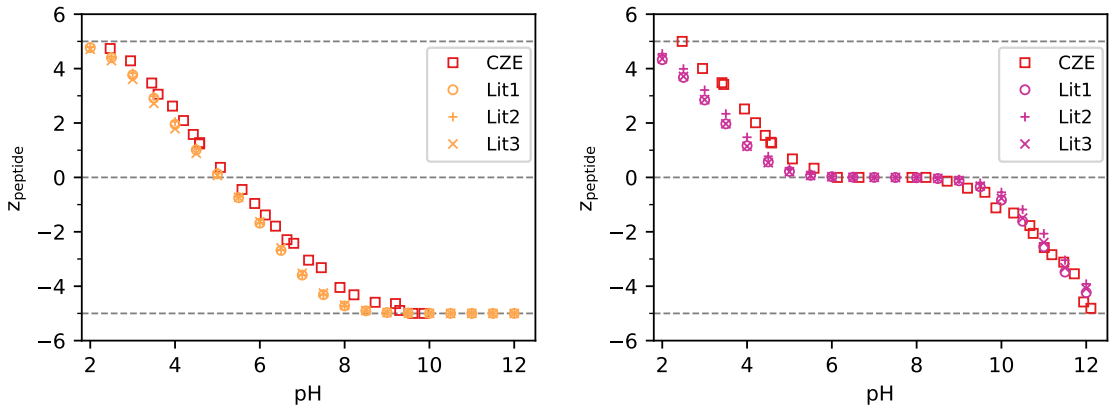
As an alternative hypothesis, one could claim that using different literature sources of

Table S6: Difference between the pK_A values of free amino acids reported in the literature and our estimates of the pK_A values of the same amino acids incorporated in the peptides $Glu_5 - His_5$ and $Lys_5 - Asp_5$.

Abbreviation	Glu	His	Asp	Lys	Source
Original pK_A (Lit1)	4.25	6.00	3.65	10.53	CRC Handbook 1991[S16]
Lit2	4.30	6.00	3.90	10.80	Concepts in Biochemistry 1988[S17]
Lit3	4.15	6.04	3.71	10.67	CRC Handbook 2015[S18]
Modified pK_A	4.45	6.20	4.21	10.53	Our estimate from NMR, Fig. S18

pK_A values of free amino acids might have the same effect as the modifications proposed above because the values reported in the literature are not entirely consistent. Table S6 outlines the pK_A values from several literature sources and the modified pK_A values obtained as described in the previous paragraph. The source labeled "Lit1" was used for the original pK_A values in the manuscript, while the other sources were included only for comparison. The reported pK_A values for His and Glu are quite consistent among different sources and do not vary by more than 0.1, while our estimation suggests that the modified pK_A values should be higher than the original values by $\Delta\mathrm{pI}(\mathrm{Glu}_5 - \mathrm{His}_5) \approx 0.2$. The reported p$K_A$ values for Lys and Asp are less consistent among different sources and do not vary by more than 0.1. Our estimation suggests that the modified pK_A values of Asp should be higher than the original value by $2\Delta\mathrm{pI}(\mathrm{Glu}_5 - \mathrm{His}_5) \approx 0.56$, while the p$K_A$ of Lys should remain unchanged. Thus, the differences between the original and modified pK_A values are greater than the inconsistencies in pK_A values reported in various literature sources.

To assess how different pK_A values from the literature might affect our simulation results, we performed a set of CG simulations using each literature source listed in Table S6. Fig. S19 shows that simulations performed with these sets of pK_A values differ only marginally, with much larger systematic differences between all simulation and CZE results. Thus, we conclude that the differences between simulations and NMR or CZE results cannot be attributed to the uncertainty in choosing a literature source of pK_A values of free amino acids. Instead, they should be attributed to a systematic shift in pK_A caused by the replacement of some substituents on the amino acids upon their incorporation into the peptide.



(a) $Glu_5 - His_5$ (b) $Lys_5 - Asp_5$

Figure S19: Simulation predictions of the total charge of the peptides, compared with experimental data from capillary zone electrophoresis (CZE). The pK_A values used in individual simulations are listed in Table S6.

2.4.3 Diffusion coefficients from DOSY NMR

The DOSY experiment made it possible to calculate the diffusion coefficients, D, of the oligopeptides as function of pH using GNAT software by integrating the peaks. Fig. S20 shows that the values of diffusion coefficients, D(pH), do not exhibit any visible trend, and they do not deviate from the average value beyond the range of the estimated error. The error bar was determined as a standard deviation of D determined from each individual peak of the spectrum. Thus, we conclude that the diffusion coefficients are approximately constant. In the whole pH range, they do not deviate from the average values by more than 10%.

References

(S1) Weik, F.; Weeber, R.; Szuttor, K.; Breitsprecher, K.; de Graaf, J.; Kuron, M.; Landsgesell, J.; Menke, H.; Sean, D.; Holm, C. ESPResSo 4.0 – an extensible software package for simulating soft matter systems. *The European Physical Journal Special Topics* **2019**, *227*, 1789–1816, DOI: 10.1140/epjst/e2019-800186-9.

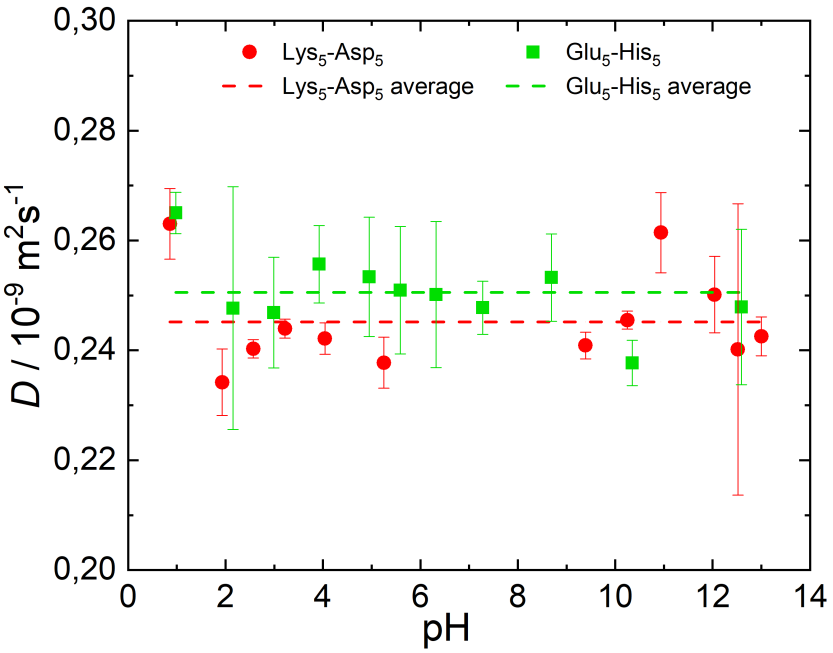


Figure S20: Diffusion coefficients for peptides as a function of pH, determined from DOSY NMR. Horizontal lines represent the diffusion coefficient averaged over all pH values.

(S2) Janke, W. Statistical Analysis of Simulations: Data Correlations and Error Estimation. *Quantum Simulations of Complex Many-Body Systems: from Theory to Algorithms* **2002**, 423–445.

(S3) `http://www.gromacs.org/`.

(S4) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25, DOI: `10.1016/j.softx.2015.06.001`.

(S5) `https://avogadro.cc`.

(S6) Hruška, V.; Svobodová, J.; Beneš, M.; Gaš, B. A nonlinear electrophoretic model for PeakMaster: Part III. Electromigration dispersion in systems that contain a neutral complex-forming agent and a fully charged analyte. Theory. *Journal of Chromatography A* **2012**, *1267*, 102–108, DOI: `10.1016/j.chroma.2012.06.086`.

(S7) Beneš, M.; Svobodová, J.; Hruška, V.; Dvořák, M.; Zusková, I.; Gaš, B. A nonlinear electrophoretic model for PeakMaster: Part IV. Electromigration dispersion in systems that contain a neutral complex-forming agent and a fully charged analyte. Experimental verification. *Journal of Chromatography A* **2012**, *1267*, 109–115, DOI: `10.1016/j.chroma.2012.06.053`.

(S8) Dubský, P.; Ördögová, M.; Malý, M.; Riesová, M. CEval: All-in-one software for data processing and statistical evaluations in affinity capillary electrophoresis. *Journal of Chromatography A* **2016**, *1445*, 158–165, DOI: `10.1016/j.chroma.2016.04.004`.

(S9) Lobaskin, V.; Dünweg, B.; Holm, C. Electrophoretic mobility of a charged colloidal particle: a computer simulation study. *Journal of Physics: Condensed Matter* **2004**, *16*, S4063–S4073, DOI: `10.1088/0953-8984/16/38/021`.

(S10) Hill, R. J.; Saville, D.; Russel, W. Electrophoresis of spherical polymer-coated colloidal particles. *Journal of Colloid and Interface Science* **2003**, *258*, 56–74, DOI: 10.1016/s0021-9797(02)00043-7.

(S11) Griffiths, P. C.; Paul, A.; Stilbs, P.; Petterson, E. Charge on Poly(ethylene imine): Comparing Electrophoretic NMR Measurements and pH Titrations. *Macromolecules* **2005**, *38*, 3539–3542, DOI: 10.1021/ma0478409.

(S12) Hwang, T.; Shaka, A. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A* **1995**, *112*, 275–279, DOI: 10.1006/jmra.1995.1047.

(S13) Jerschow, A.; Müller, N. Suppression of Convection Artifacts in Stimulated-Echo Diffusion Experiments. Double-Stimulated-Echo Experiments. *Journal of Magnetic Resonance* **1997**, *125*, 372–375, DOI: 10.1006/jmre.1997.1123.

(S14) Castañar, L.; Poggetto, G. D.; Colbourne, A. A.; Morris, G. A.; Nilsson, M. The GNAT: A new tool for processing NMR data. *Magnetic Resonance in Chemistry* **2018**, *56*, 546–558, DOI: 10.1002/mrc.4717.

(S15) Hass, M. A.; Mulder, F. A. Contemporary NMR Studies of Protein Electrostatics. *Annual Review of Biophysics* **2015**, *44*, 53–75, DOI: 10.1146/annurev-biophys-083012-130351, tex.ids: hass2015a.

(S16) Lide, D. R. *CRC Handbook of Chemistry and Physics*, 72nd ed.; CRC Press: New York, 1991.

(S17) Stephenson, W. K. *Concepts in Biochemistry*, 3rd ed.; Wiley: New York, 1988.

(S18) Haynes, W. *CRC Handbook of Chemistry and Physics*, 96th ed.; CRC Press: New York, 2015.