Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

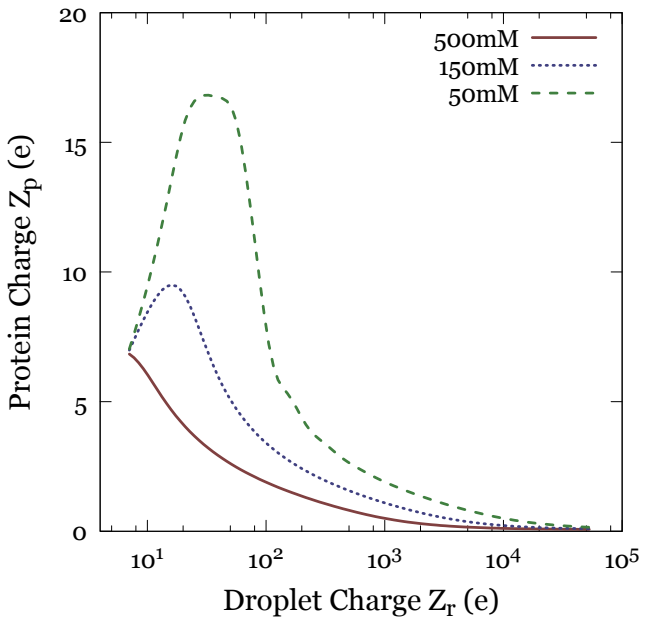
Supporting information for: What Factors Determine the Stability of a Weak Protein-Protein Interaction in a Charged Aqueous Droplet?

Myong In Oh and Styliani Consta*

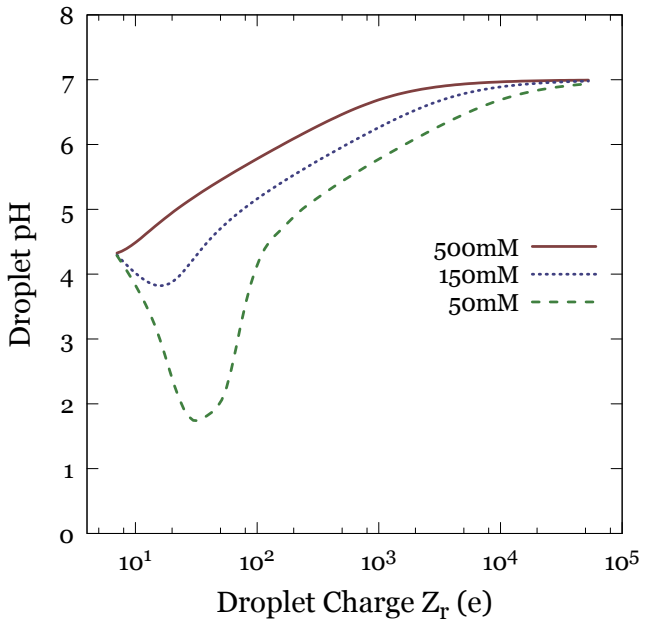
Department of Chemistry, The University of Western Ontario, London, Ontario, Canada N6A 5B7

E-mail: sconstas@uwo.ca

S1 Assignment of the Charge State of a Protein



(a)



(b)

Figure S1: (a) Protein charge Z_p (e) vs. droplet charge Z_r (e) for various concentrations of CH_3COONH_4. (b) Droplet pH vs. droplet charge Z_r (e) for the same concentrations of CH_3COONH_4 as those in (a).

A code written in MAPLE version 2016.0 is attached at the end of the Supporting Information.

S2 Spherical Boundary Condition and Equilibration

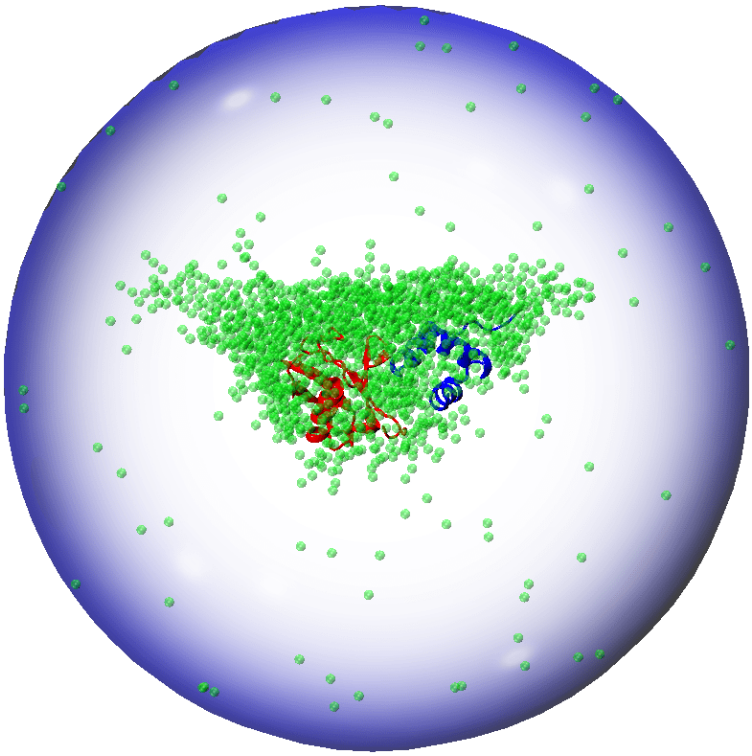


Figure S2: Computational set-up for 2MRO with $+14\ e$ in a droplet of 2000 H_2O molecules. The protein complex is coloured in red (ubiquitin) and blue (ubiquitin-associated domain), and only the oxygen sites of the water molecules are shown in green for clarity. In order to perform equilibrium simulations, the entire droplet system is enclosed in a cavity (coloured in transparent blue). The radius of the cavity used in this study is 20 nm.

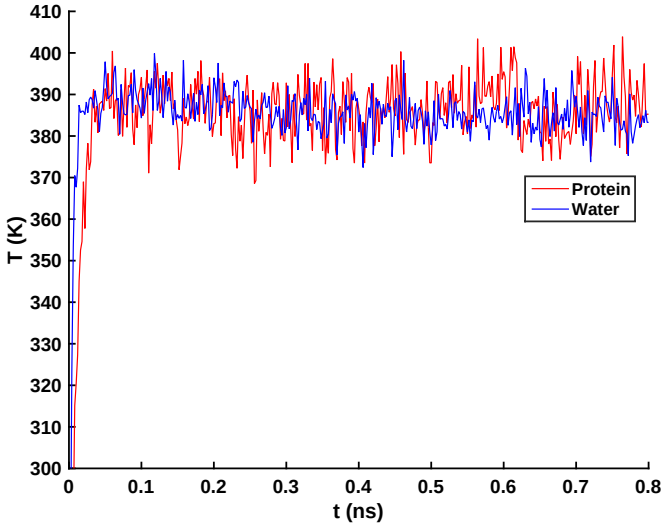


Figure S3: Temperature of the water molecules and 2MRO protein complex during equilibration.

S3 Description of the Ub-UbA interface

A protein-protein interface is of high complexity and specificity and may be characterized by such descriptors as its area, shape, and surface complementarity. When the stability of a non-covalently bound protein complex is investigated, there is an important question to be asked: "What are the residues involved in the formation of the protein-protein interface?". As protein-protein interfaces are often hydrophobic, hydrophobicity is a driving force in the recognition of a partner protein.[S1] It is known that the hydrophobic patch centered on isoleucine-44 (ILE44) of ubiquitin serves as the recognition site bound by most ubiquitin-binding domains including those found in a shuttle protein.[S2] UbA domains, which form a compact bundle of three α-helices stabilized by a hydrophobic core, have hydrophobic surface patches, and these patches often act as binding sites for other proteins.[S3] Consequently, UbA domains have been found to play a crucial role in many other protein-protein interactions.[S4] Direct visualization of the protein complex reveals that the initial interface formed by ubiquitin and the UbA domain in 2MRO is indeed mostly hydrophobic. Figure S4 (b) shows what residues are involved in the formation of the interface in 2MRO. The orange and black dots indicate two amino acids (identified by their residue numbers on axes) that are within 8 $\mathrm{\AA}$ and 4 $\mathrm{\AA}$ in distance, respectively. The distance was measured based on the locations of their β-carbons to account for the orientation of their side chains. Also, glycine was omitted due to the lack of a side chain. The analysis of this contact map reveals that 62 $\%$ of the amino acids at the interface are sorted as hydrophobic ones. Interestingly, many of the polar and charged amino acids are located only at the outer verge of the interface with direct access to water. Therefore, the core region of the interface is predominated by the presence of hydrophobic residues, as supported by the computational prediction of hot spots.[S5,S6] Hot spots are a small fraction of residues in the interface that cause destabilization of the bound state, leading to an increase in the binding free energy of at least 2.0 kcal/mol upon alanine mutagenesis. They make large contributions to the stability (or lifetime) of the protein complex, and they are present mostly in a central region of the interface, hidden from solvent.[S7] As shown in Figure S4 (a), the hot spots are leucine-8, isoleucine-44, valine-70, and leucine-73 in Ub, and alanine-423, serine-424, and phenylalanine-

427 in the UbA domain. Except serine-424, the rest of the hot spots are hydrophobic residues.

The buried surface area (BSA) of a protein complex is a hydrophobic surface removed from contact with solvent and buried in the interface. BSA is a descriptor related to the binding affinity of a PPI, whose magnitude is estimated to be approximately 0.025 kcal/mol per 1 Å^2 of the hydrophobic surface. It was calculated that for 2MRO it is ≈ 624 Å^2 which is in the range of a typical transient protein complex.[S8] Based on the estimation, the binding free energy at the interface is estimated to be ≈ 15.6 kcal/mol. The dissociation constant (K_D) of the complex of monomeric Ub and the UbA domain has been measured to be 150 ± 16 mM by analyzing NMR titration curves,[S9] which is again in the typical range of a transient PPI.

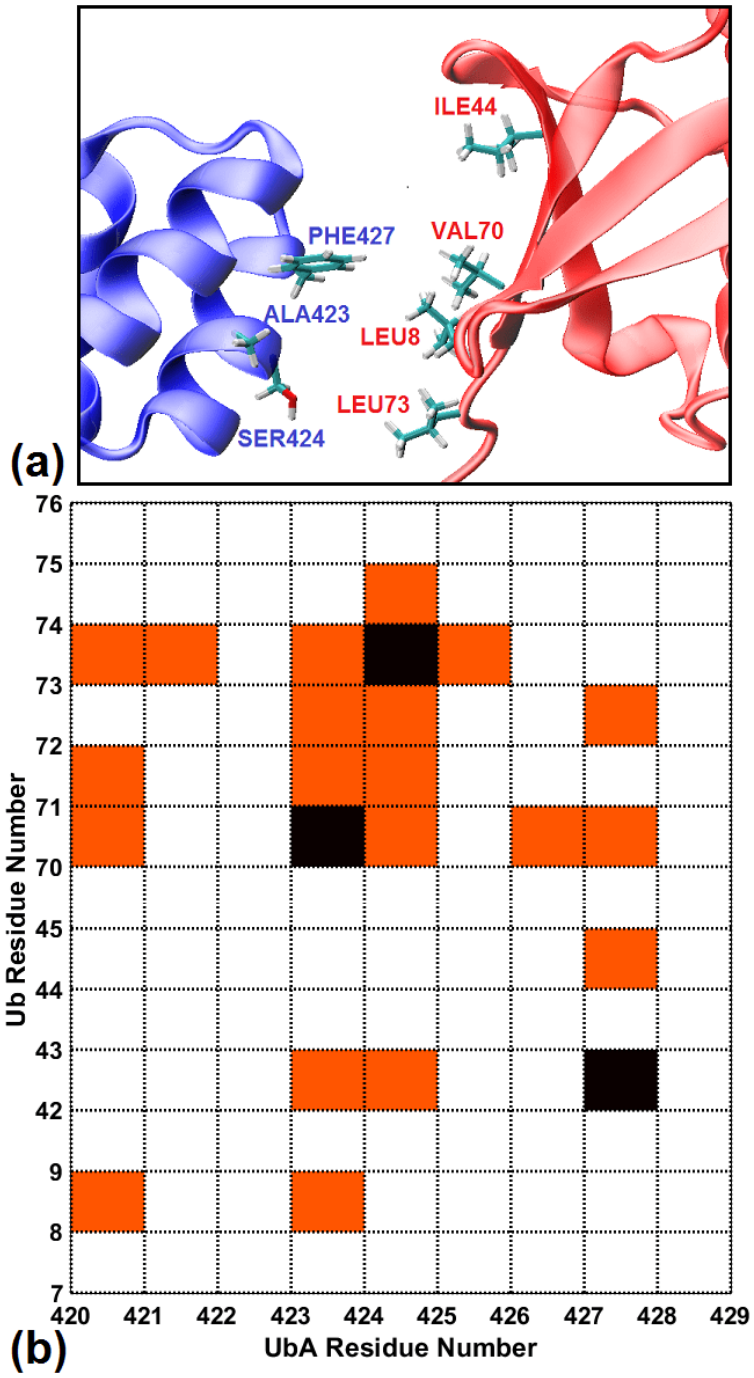


Figure S4: (a) Hot spots at the hydrophobic interface of the protein complex. The ubiquitin molecule is coloured in red, whereas the ubiquitin-associated domain is in blue. (b) Contact map showing a pair of interacting residues around the hot spots at the hydrophobic protein-protein interface. The orange and black blocks indicate the distances between the β-carbons of two residues on the protomers less than 8 Å and 4 Å, respectively.

S4 Contact Map

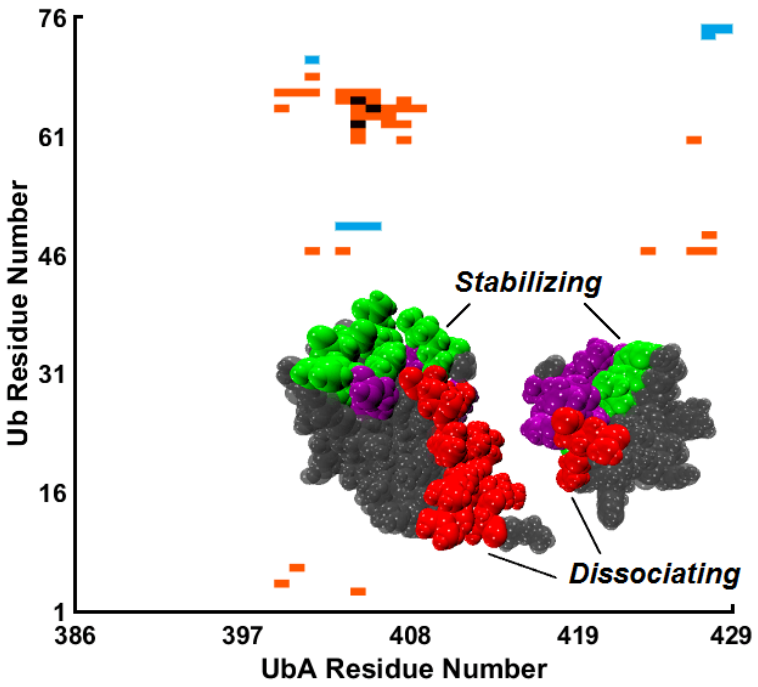


Figure S5: Example of the contact maps showing the type of residues (indicated by their residue numbers) that are involved in the interface of an intermediate complex that leads to dissociation and stabilization after protomer reorientation. The orange/black colouring scheme is the same as in Figure S4 for the dissociating complex, while the blue is for the stabilizing one but corresponds to the black for the dissociating one. The inset shows different interfaces found in the protein complex. These interfaces are the surfaces explored by the protomers' partners. The water molecules are not shown. The interface coloured in red leads to complex dissociation after protomer reorientation, whereas that in green causes stabilization of the intermediate complex. The purple colour indicates the region shared by both red and green patches. No interface formation is found in the gray region.

S5 Statistical Analysis

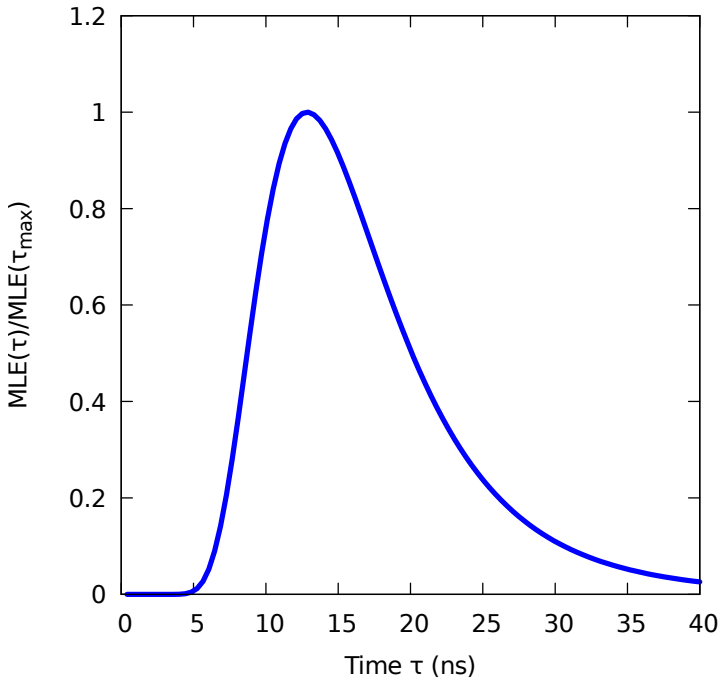


Figure S6: Maximum likelihood estimate (MLE(τ)) normalized by MLE(τ_{max}). $P(\{t_1, \cdots t_N\} \mid \tau)$ as a function of τ.

In our statistical analysis we use the method of maximum likelihood estimate (MLE).[S10] In MLE one assumes the statistical model of the observations and one maximises the probability of the observed events by varying the fitting parameters. In our studies we use the MLE instead of a simple average of the time, because we want to take into account the fragmentation events that occurred beyond the cut-off time. We assume that the distribution of the fragmentation times of the droplet is given by the following normalized exponential probability density

$$P(t) = \lambda e^{-\lambda t} \tag{1}$$

where $\lambda = 1/\tau$. We use the maximum likelihood estimate to find the value of τ. The conditional

probability of the times $\{t_1, \cdots t_N\}$ given τ is

$$P(\{t_1, \cdots t_N\} \mid \tau) = \prod_{i=1}^{K} \frac{1}{\tau} e^{-t_i/\tau} \prod_{i=K+1}^{N} e^{-t_{max}/\tau} = e^{(N-K)t_{max}/\tau} \prod_{i=1}^{K} \frac{1}{\tau} e^{-t_i/\tau} \tag{2}$$

where t_{max} = 10 ns, N = 14, K = 8.

By Bayes's theorem we have that $P(\tau \mid \{t_1, \cdots t_N\}) \approx (constant) P(\{t_1, \cdots t_N\} \mid \tau) P(\tau)$. Assuming that $P(\tau)$ is uniform, the maximum likelihood of $P(\tau \mid \{t_1, \cdots t_N\}) \approx P(\{t_1, \cdots t_N\} \mid \tau)$. By plotting $P(\{t_1, \cdots t_N\} \mid \tau)$ as a function of τ in Fig. S6 or alternatively by maximizing Eq. 2 we find that τ equals 13 ns.

The next step is to test whether the exponential guess of the time distribution and the theoretical distribution are in agreement. We consider the cumulative probability density (CPD) of the fragmentation times provided by the simulations and the theoretical $1 - e^{-t/\tau}$ probability density as a function of time. In Fig. S7 we plot the experimental CPD (blue line) and a diagonal (red line) for the same points in time. If the theoretical and the experimental CPD were the same, then all the points will lie along the diagonal. The blue line shows the difference from the diagonal.

We use the following criterion

$$w^2 = \int_0^{0.5365} (F_{\exp}(x) - x)^2 dx \tag{3}$$

in the same spirit as the von Mises measure in order to compare the theoretical with the experimental CPD. The lower and upper limit of integration is estimated from $1 - e^{-t/\tau}$ at $t = 0$ and $t = 10$, respectively. The value of $t = 10$ is used because this is the maximum time for which we have data. Equation 3 gives the von Mises integral to be I = 0.010855.

In Fig. S8 the blue line is the probability density of the Mises measure of 10^6 sets of 14 points that were computer-generated. Each of these sets were generated from the distribution shown in Eq. 1 (with $\lambda = 1/\tau = 1/13\ ns$). The red line is the integral of the distribution. For the 14 MD times, the measure is 0.01, which indicates that the 80 % of the measures are less than 0.01.

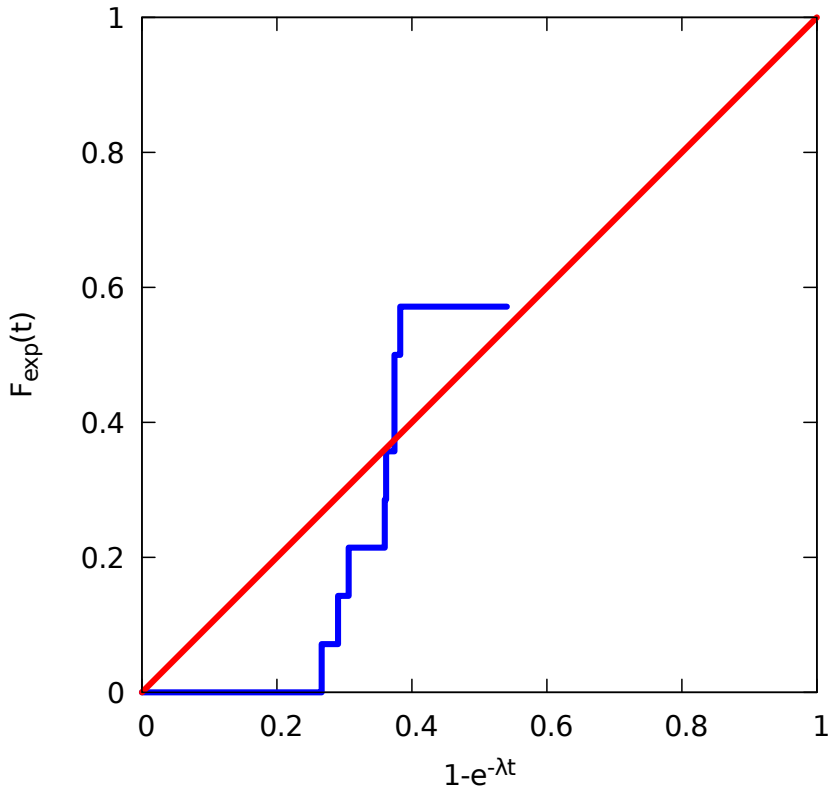


Figure S7: Cumulative probability density (CPD) of the fragmentation times taken from the simulations (blue line) and the theoretical $1 - e^{-t/\tau}$ (where $\tau = 1/\lambda$) probability (red line).

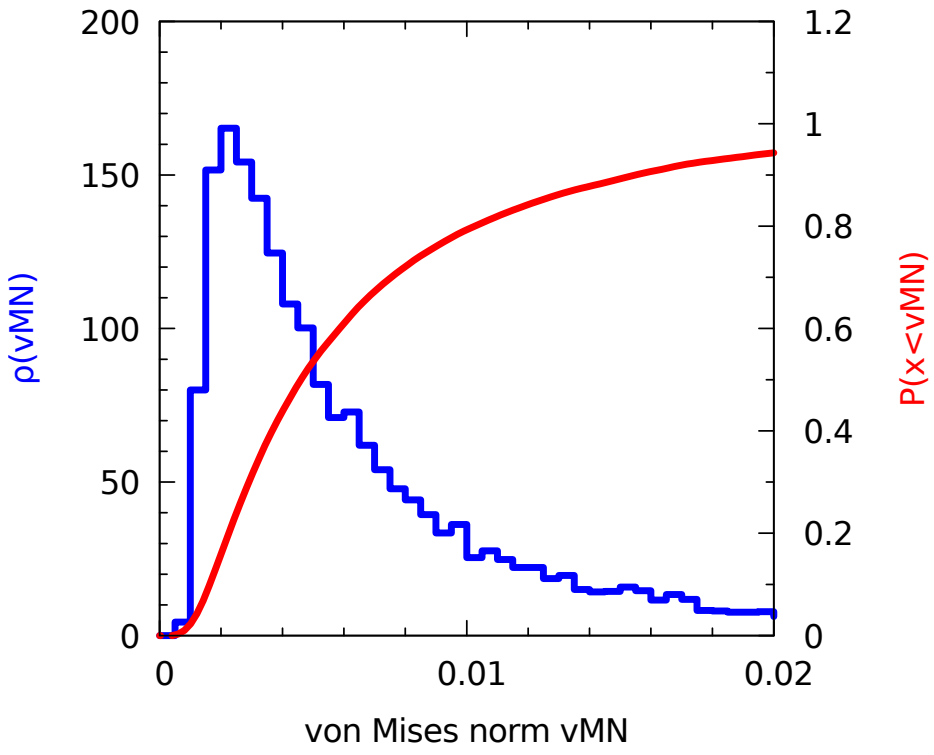


Figure S8: Histogram of the Mises measures of 10^6 points shown by the blue line and the integral of the histogram shown by the red line.

S6 Root-Mean-Square Deviation of 2MRO at Lower Charge and Temperature

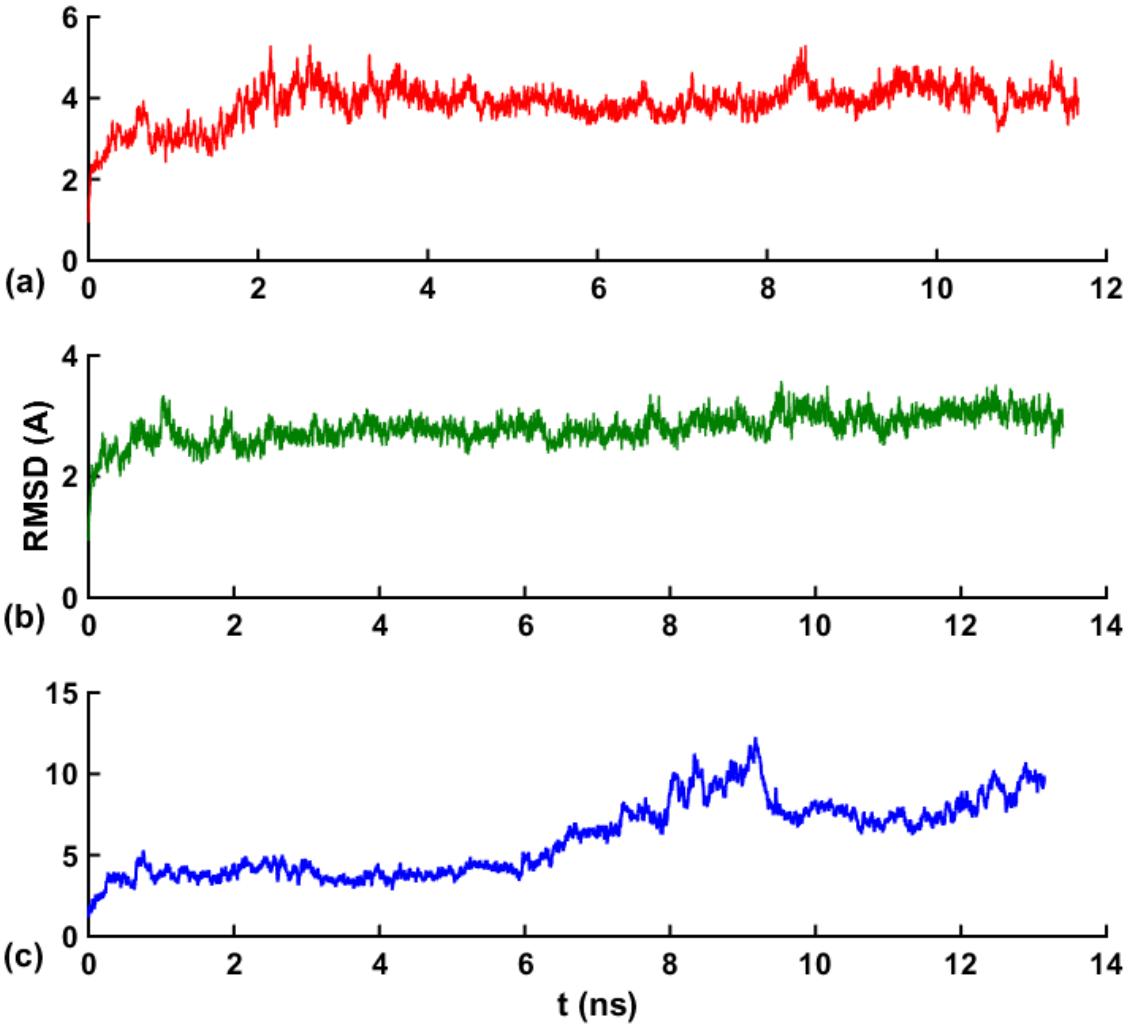


Figure S9: Root-mean-square deviation (RMSD) as a function of time t for (a) $2\mathrm{MRO}^{10+}$ in 2000 $\mathrm{H_2O}$ molecules at 390 K, (b) $2\mathrm{MRO}^{14+}$ with the stabilizing protein-protein interface in 1400 $\mathrm{H_2O}$ molecules at 300 K, and (c) the same 2MRO but with the destabilizing protein-protein interface.

S7 Different ”Star” Morphologies Due to Protomer Reorientation

We observed that different numbers of spikes are developed on the droplet surface during the evolution of the droplet, as shown in Figure S10. The lifespan and the location of those protrusions can vary. High temperature enhances surface fluctuations of the droplet, generating short-living spikes as indicated by a dashed black circle in Figure S10 (c). They may appear in different locations on the droplet surface. On the contrary, the two spikes pointing outward from each protomer (indicated by dashed squares in the same figure) are long-living and rather fixed in location.

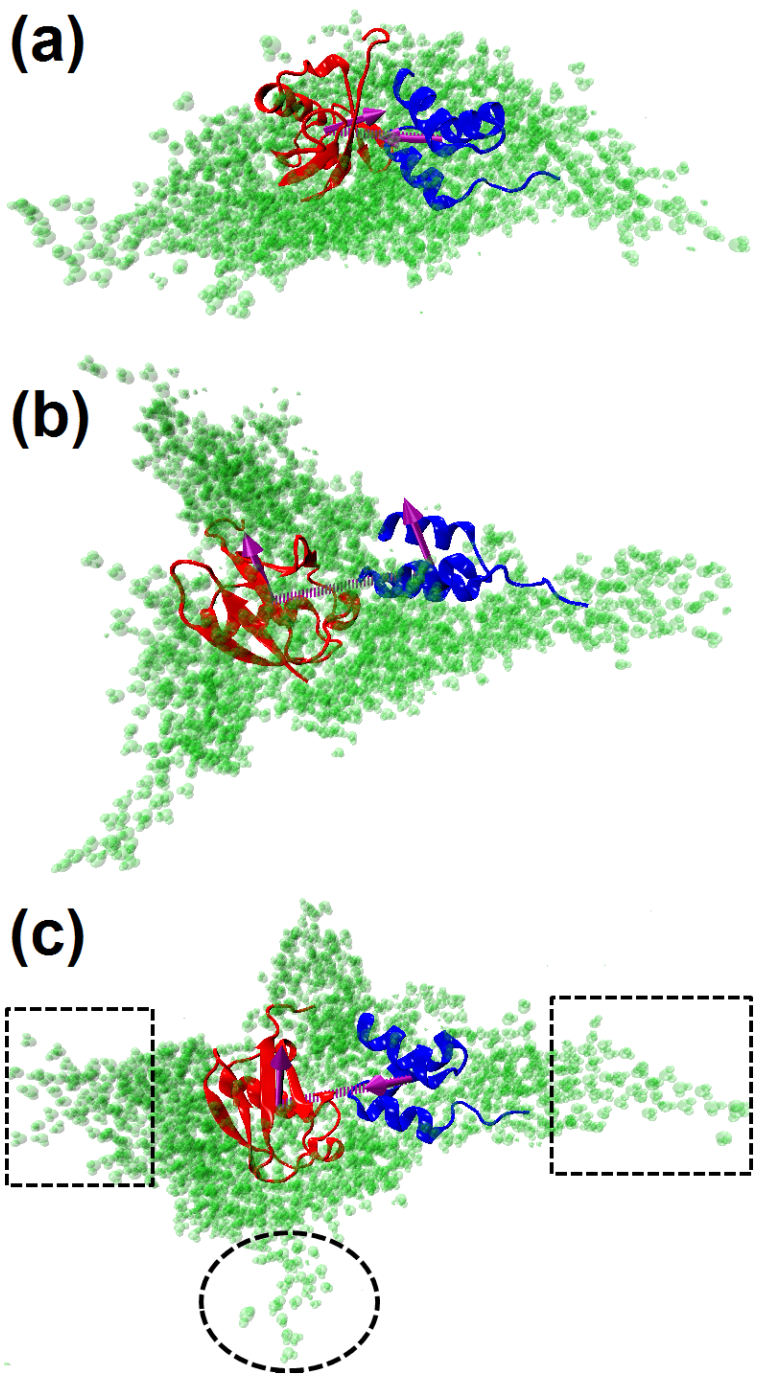


Figure S10: Different star morphologies of the aqueous droplet that contains 2MRO with +14 e and $\approx$ 1400 H_2O molecules. The same colouring scheme for the protein complex and the arrows was used as in Figure 2 in the main tex, and the water molecules are coloured in transparent green. The transient spikes and the long-living ones are indicated by a dashed black circle and a dashed black square, respectively.

S8 Fission Rate of a Small Ion from Charged Aqueous Droplets

Table S1: Time required to observe a fission event in charged aqueous droplets with separable small ions at T = 300 K. τ_{pos} and τ_{neg} represent the time when fission occurs after equilibration in positively and negatively charged aqueous droplets, respectively. The last row shows the mean value and the standard deviation of the five τ values listed in each column.

System	τ_{pos} (ps)	τ_{neg} (ps)
1	130	70
2	210	160
3	70	340
4	50	120
5	160	380
τ_{avg}	124 ± 65.4	214 ± 137.8

S9 Effect of Counterions: Iodide (I^-) Ions

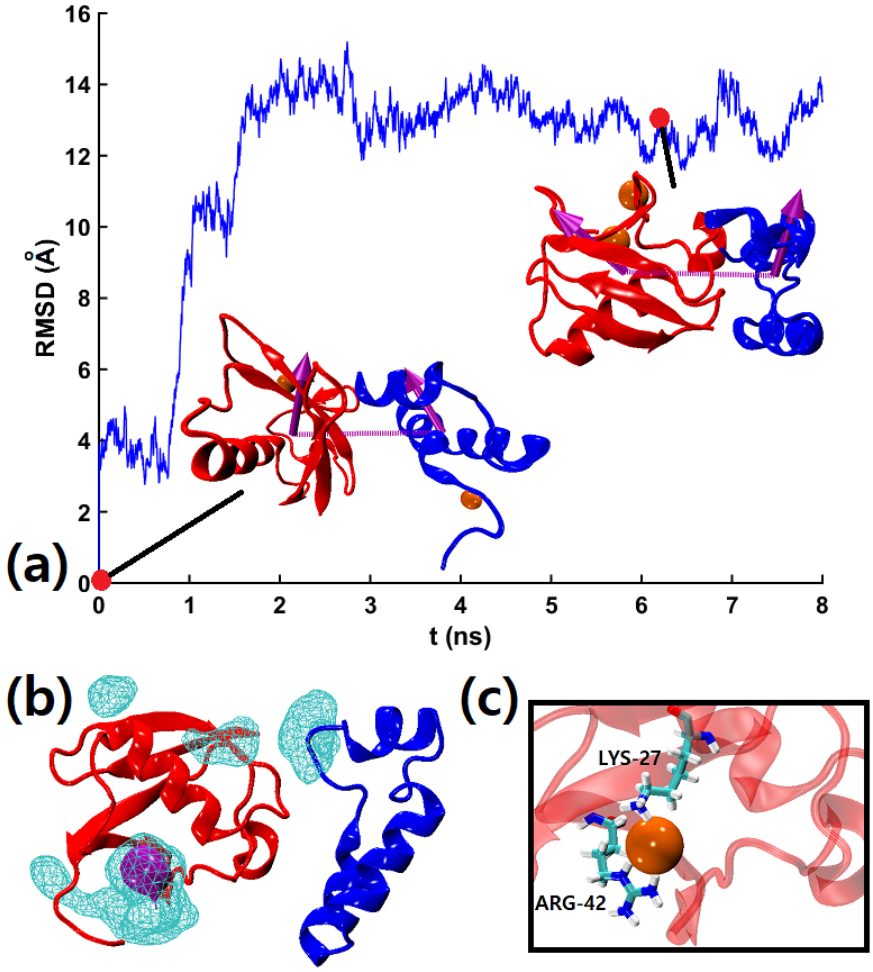


Figure S11: (a) Root-mean-square deviation (RMSD) of the protein complex ion in an aqueous droplet with two I^- ions. The snapshots in (a) show the protein complex conformations at 0 ns and 6.25 ns. (b) Example of the density map showing the location of the two I^- ions around the protein complex. The density map was obtained by averaging the entire MD trajectory ($\approx$ 10 ns). (c) The snapshot magnifies the two positively charged residues, lysine-27 and arginine-42, located at the purple region at the bottom of the ubiquitin molecule shown in (b). The same scheme is used for colouring as in Fig. 8 and 10 (in the main text), except that the orange spheres represent the I^- ions. Water molecules were omitted for better visualization.

S10 Solvent Evaporation Rate of the Aqueous Droplet with the $2PEA^{14+}$ Ion

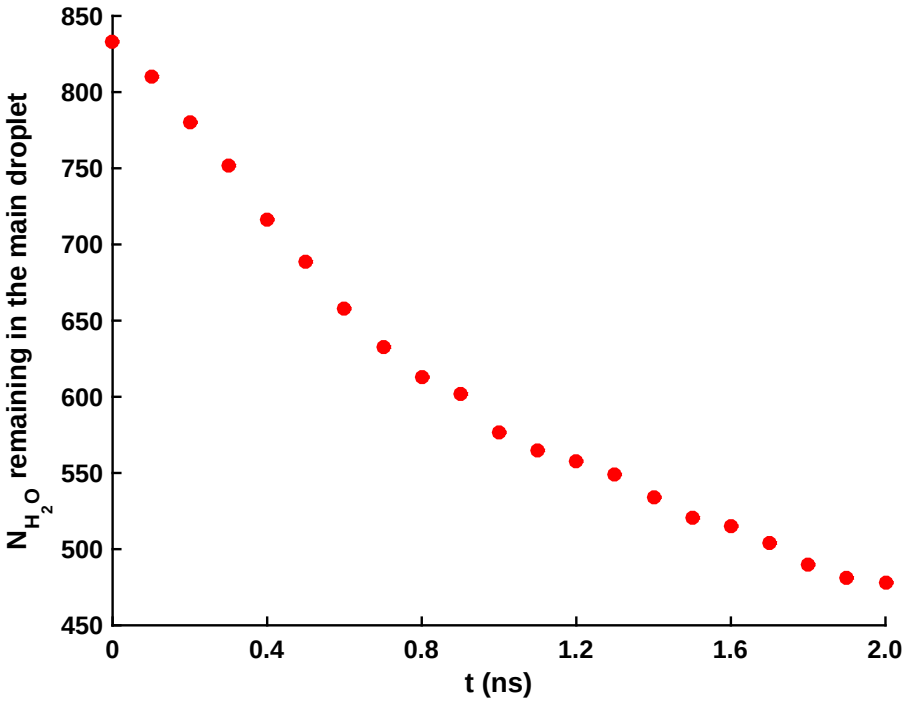


Figure S12: Number of water molecules (N_{H_2O}) remaining in the main droplet that initially contains 833 H_2O molecules and the $2PEA^{14+}$ complex ion as a function of time (t) at 390 K.

S11 Movies of MD Simulations

S11.1 Dissociation of 2MRO in a Charged Aqueous Droplet

A movie of the molecular dynamics simulation that shows the dissociation mechanism of the protein complex is included. (File: 2MROdiss.avi)

S11.2 Star Formation of a Charged Aqueous Droplet with 2PEA

A movie of the molecular dynamics simulation that shows the star formation of the droplet system is included. (File: StarForm.avi)

References

(S1) Young, L.; Jernigan, R.; Covell, D. *Protein Sci.* **1994**, *3*, 717–729.

(S2) Dikic, I.; Wakatsuki, S.; Walters, K. J. *Nat. Rev. Mol. Cell. Biol.* **2009**, *10*, 659–671.

(S3) Mueller, T. D.; Feigon, J. *J. Mol. Biol.* **2002**, *319*, 1243–1255.

(S4) Hurley, J. H.; Lee, S.; Prag, G. *Biochem. J* **2006**, *399*, 361–372.

(S5) Darnell, S. J.; Page, D.; Mitchell, J. C. *Proteins: Structure, Function, and Bioinformatics* **2007**, *68*, 813–823.

(S6) Zhu, X.; Mitchell, J. C. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79*, 2671–2683.

(S7) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. *Proteins: Structure, Function, and Bioinformatics* **2007**, *68*, 803–812.

(S8) Acuner Ozbabacan, S. E.; Engin, H. B.; Gursoy, A.; Keskin, O. *Protein Eng. Des. Sel.* **2011**, *24*, 635–648.

(S9) Nowicka, U.; Zhang, D.; Walker, O.; Krutauz, D.; Castañeda, C. A.; Chaturvedi, A.; Chen, T. Y.; Reis, N.; Glickman, M. H.; Fushman, D. *Structure* **2015**, *23*, 542–557.

(S10) MacKay, D. J. C. *Information Theory, Inference and Learning Algorithms*; Cambrdige University Press, 2003; `http://www.inference.phy.cam.ac.uk/mackay/itprnn/book.html`.

```
>
>
> pKa := 7.7; Cysteine pKa experimental
> Dpka := 0.5; Variance of the pKa (see ref)
> Q := (x, y) → 1/(1 + 10^(x − y)); Population of the protonated states
> DP := x → 1/(sqrt(2·Pi)·Dpka) · exp(−0.5·((x − pKa)/Dpka)^2);
> A := x → int(Q(x, y)·DP(y), y = −infinity..infinity); Expectation of the population
  of a cycteine sampled from the normal distribution above
>
>
>
> PL := [seq([0.25·t, evalf(A(0.25·t))], t = 0..50)]; Numerical integration
>
> X := PL[1..51, 1];
> Y := PL[1..51, 2];
```

$$pKa := 7.7$$

$$Dpka := 0.5$$

$$Q := (x, y) \to \frac{1}{1 + 10^{x - y}}$$

$$DP := x \to \frac{e^{(-1)\cdot 0.5\,\frac{(x - pKa)^2}{Dpka^2}}}{\sqrt{2\,\pi}\;Dpka}$$

$$A := x \to \int_{-\infty}^{\infty} Q(x, y)\; DP(y)\; dy$$

$PL := [[0., 0.9999999612], [0.25, 0.9999999311], [0.50, 0.9999998775], [0.75, 0.9999997822], [1.00, 0.9999996128], [1.25, 0.9999993115], [1.50, 0.9999987758], [1.75, 0.9999978231], [2.00, 0.9999961290], [2.25, 0.9999931164], [2.50, 0.9999877593], [2.75, 0.9999782335], [3.00, 0.9999612956], [3.25, 0.9999311805], [3.50, 0.9998776445], [3.75, 0.9997824954], [4.00, 0.9996134609], [4.25, 0.9993133948], [4.50, 0.9987814301], [4.75, 0.9978405116], [5.00, 0.9961827391], [5.25, 0.9932808623], [5.50, 0.9882539842], [5.75, 0.9796852503], [6.00, 0.9654214143], [6.25, 0.9424498536], [6.50, 0.9070360768], [6.75, 0.8553475790], [7.00, 0.7846683027], [7.25, 0.6949476145], [7.50, 0.5899728374], [7.75, 0.4773054186], [8.00, 0.3666184590], [8.25, 0.2670288020], [8.50, 0.1847005272], [8.75, 0.1218074474], [9.00, 0.07705073897], [9.25, 0.04708925399], [9.50, 0.02801624565], [9.75, 0.01634306355], [10.00, 0.009403984584], [10.25, 0.005362561446], [10.50, 0.003040626207], [10.75, 0.001718113776], [11.00, 0.0009688392715], [11.25, 0.0005456763623], [11.50, 0.0003071300903], [11.75, 0.0001727989416], [12.00, 0.00009719957767], [12.25, 0.00005466807904], [12.50, 0.00003074488635]]`

$X := [0., 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 3.25, 3.50, 3.75, 4.00, 4.25, 4.50, 4.75, 5.00, 5.25, 5.50, 5.75, 6.00, 6.25, 6.50, 6.75, 7.00, 7.25, 7.50, 7.75,`

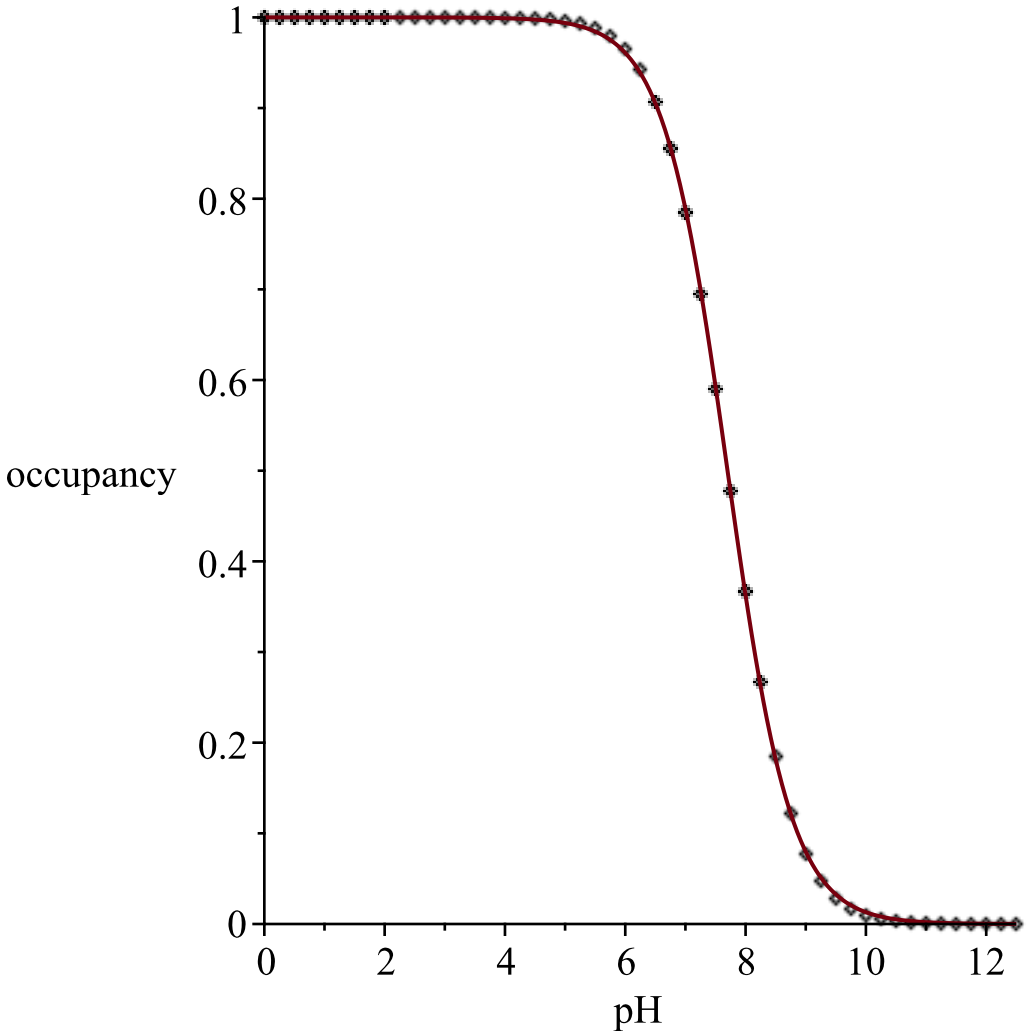
8.00, 8.25, 8.50, 8.75, 9.00, 9.25, 9.50, 9.75, 10.00, 10.25, 10.50, 10.75, 11.00, 11.25, 11.50, 11.75, 12.00, 12.25, 12.50]

$Y := [0.9999999612, 0.9999999311, 0.9999998775, 0.9999997822, 0.9999996128, 0.9999993115, 0.9999987758, 0.9999978231, 0.9999961290, 0.9999931164, 0.9999877593, 0.9999782335, 0.9999612956, 0.9999311805, 0.9998776445, 0.9997824954, 0.9996134609, 0.9993133948, 0.9987814301, 0.9978405116, 0.9961827391, 0.9932808623, 0.9882539842, 0.9796852503, 0.9654214143, 0.9424498536, 0.9070360768, 0.8553475790, 0.7846683027, 0.6949476145, 0.5899728374, 0.4773054186, 0.3666184590, 0.2670288020, 0.1847005272, 0.1218074474, 0.07705073897, 0.04708925399, 0.02801624565, 0.01634306355, 0.009403984584, 0.005362561446, 0.003040626207, 0.001718113776, 0.0009688392715, 0.0005456763623, 0.0003071300903, 0.0001727989416, 0.00009719957767, 0.00005466807904, 0.00003074488635]$ **(1)**

```
> with(plots):
>
> Func := x -> 1/(1+exp(a*(x-b))); Fitting function
> SS := Statistics[NonlinearFit](Func(x), X, Y, x, initialvalues=[a=3, b=pKa],
      parameterranges=[a=1..3, b=0..20], output=[parametervalues, leastsquaresfunction,
      residualstandarddeviation]);
  The best fit
>
>
>
> P1 := pointplot(PL):
> P2 := plot(subs(SS[1], 1/(1+exp(a*(x-b)))), x=0..12.5):
>
>
> display([P1, P2], labels=["pH", "occupancy"]); Plot of the best fit and the
  exact solution
>
> ne := evalf(subs(SS[1], exp(a)));
```

$$Func := x \to \frac{1}{1+e^{a\,(x-b)}}$$

$$SS := \left[[a = 1.88250462852574,\ b = 7.70000001050413],\ \frac{1}{1+e^{1.88250462852574\,x - 14.4952856594223}},\ 0.002447684195 \right]$$



$ne := 6.569939530$ **(2)**

> $N[HIS] := 1$; **Number of ionazable groups of a specific type**
> $N[ASP] := 7$;
> $N[GLU] := 8$;
> $N[LYS] := 10$;
> $N[ARG] := 5$;
>
>

> $Z\_p := x \to \frac{N[HIS]}{1+\exp(1.35\cdot(x-6.6))} - \frac{N[ASP]}{1+\exp(1.19\cdot(3.5-x))} - \frac{N[GLU]}{1+\exp(1.44\cdot(4.2-x))} + \frac{N[ARG]}{1+\exp(1.25\cdot(x-12.0))} + \frac{N[LYS]}{1+\exp(1.25\cdot(x-10.5))} + \frac{2}{1+\exp(1.88\cdot(x-7.7))} - \frac{2}{1+\exp(1.54\cdot(3.3-x))}$;

Expectation value of the charge for 2MRO fot a given value of pH

> $plot(Z\_p(x), x=0..14, labels=["pH", "Charge"])$;

$N_{HIS} := 1$

$N_{ASP} := 7$

$N_{GLU} := 8$

$$N_{LYS} := 10$$

$$N_{ARG} := 5$$

$$Z\_p := x \rightarrow \frac{N_{HIS}}{1 + e^{1.35\,(x - 6.6)}} - \frac{N_{ASP}}{1 + e^{1.19\,(3.5 - x)}} - \frac{N_{GLU}}{1 + e^{1.44\,(4.2 - x)}} + \frac{N_{ARG}}{1 + e^{1.25\,(x - 12.0)}} + \frac{N_{LYS}}{1 + e^{1.25\,(x - 10.5)}} + \frac{2}{1 + e^{1.88\,(x - 7.7)}} - \frac{2}{1 + e^{1.54\,(3.3 - x)}}$$