Supporting Information

Library of Random Copolypeptides by Solid Phase Synthesis

Vladimir Dmitrović[‡,a,b], Jos J.M. Lenders[‡,a,c], Harshal R. Zope[‡,d], Gijsbertus de With[a], Alexander Kros[d,*] and Nico A.J.M. Sommerdijk[a,c,*]

[a] Laboratory of Materials and Interface Chemistry and Soft Matter CryoTEM Research Unit, Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

[b] Dutch Polymer Institute, P.O. Box 902, 5600 AX Eindhoven, The Netherlands

[c] Institute for Complex Molecular Systems, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

[d] Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands

[‡] These authors contributed equally to this work.

* Corresponding authors: n.sommerdijk@tue.nl; a.kros@chem.leidenuniv.nl

Materials

Aqueous solutions were prepared in ultrapure water (resistivity 18.2 MΩ·cm at 20 °C). Fmoc-Glu(Otbu)-OH, Fmoc-Lys(Boc)-OH and Fmoc-Ala-OH amino acid derivatives as well as *N*,*N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), piperidine, *N*,*N*-diisopropylethylamine (DIEA), 1H-benzotriazolium,1-[bis(dimethylamino)methylene]-5-chloro-hexafluorophosphate(1-),3-oxide (HCTU) and acetonitrile (ACN) (all peptide synthesis grade) were purchased from Biosolve. Trifluoroacetic acid (TFA), triisopropylsilane (TIS), diethyl ether, α-cyano-4-hydroxycinnamic acid (ACH), lithium chloride (LiCl) and calcium chloride dihydrate ($CaCl_2 \cdot 2H_2O$, ≥98%) were purchased from Sigma-Aldrich. All the reagents required for acid hydrolysis were purchased from Waters.

Synthesis of random polypeptides

The desired random copolypeptides (with a degree of polymerization of 24) were prepared using a solid phase synthesis method. The scale of the synthesis was 0.1 mmol, following the fluorenylmethoxycarbonyl (Fmoc) strategy and using standard Fmoc-derivatized amino acids. Briefly, the synthesis was performed on a fully automated parallel peptide synthesizer, type Syro I (Multisyntech, Witten, Germany), using Rink Amide resin (Iris Biotech, substitution 0.53 mmol/g) as solid support. The desired 4 equivalent mix of amino acid solutions in *N*,*N*-dimethylformamide (DMF) (LiCl, 1 g/L) was prepared and activation of the amino acids was achieved using a 1H-benzotriazolium,1-[bis(dimethylamino)methylene]-5-chloro-hexafluorophosphate(1-),3-oxide (HCTU, in DMF) / *N*,*N*-diisopropylethylamine (DIEA, in

NMP) mixture in a HCTU:DIEA molar ratio of 1:2. Fmoc deprotection was carried out using a 40% (v/v) piperidine solution in DMF. All couplings were performed for 45 min at room temperature, while deprotections were carried out by exposing the resin for 3 min to 40% piperidine, followed by a 12 min exposure to 20% piperidine. A gentle flow of N_2 was maintained throughout the synthesis.

After completion of the solid phase-assisted polymerization the polypeptides were cleaved from the resin using a trifluoroacetic acid (TFA) / triisopropylsilane (TIS) / H_2O (95:2.5:2.5, v/v/v) mixture for 3 hrs. The resulting polymers were precipitated by quenching the solutions with cold diethyl ether (dry) and centrifugation at 400 rpm for 10 min. The pellets were thoroughly washed 3-4 times in cold diethyl ether (dry) and were vacuum-dried to obtain powdered copolypeptides.

MALDI-TOF mass spectrometry

The random copolypeptides were characterized using MALDI-TOF mass spectrometry. Samples of random copolypeptides were prepared using α-cyano-4-hydroxycinnamic acid (ACH) and analyzed on a Bruker Daltonics-Microflex using the linear mode.

Subsequently, the experimental MALDI-TOF patterns were fitted by normal distributions according to Equation 1, where μ is the mean molar mass, σ is the standard deviation, A is a scaling factor and B is the apparent noise level.

Equation 1: $$I_{ND}(M) = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{(M-\mu)^2}{2\sigma^2}} + B$$

The positions and intensities of the peaks of the major distributions in the experimental MALDI-TOF patterns were taken as input to a fitting routine that minimizes the residuals between theory and experiment (non-linear least-squares regression). Figure S1 shows the results after normalization (= after subtracting B and dividing by A). In almost all cases the fit is in good agreement with the experiment and pseudo-R^2 values larger than 0.95 were found, except for a few cases where the signal-to-noise ratio of the MALDI-TOF measurement (S/N = I_{max} / B) was lower than 6.

The as such determined mean molar mass μ was taken as the best approximation for the molecular weight M_n of each copolymer batch, while the corresponding polydispersity index PDI could be calculated from the determined standard deviation σ according to Equation 2. Further, the found M_n was combined with the compositional information from the ^{1}H NMR analysis to derive the degree of polymerization DP. See also Table 1 in the main text for the determined M_n, PDI and DP of each copolymer batch.

Equation 2: $\sigma = M_N\sqrt{PDI-1} = \mu\sqrt{PDI-1} \rightarrow PDI = \frac{M_W}{M_N} = 1+\left(\frac{\sigma}{\mu}\right)^2$

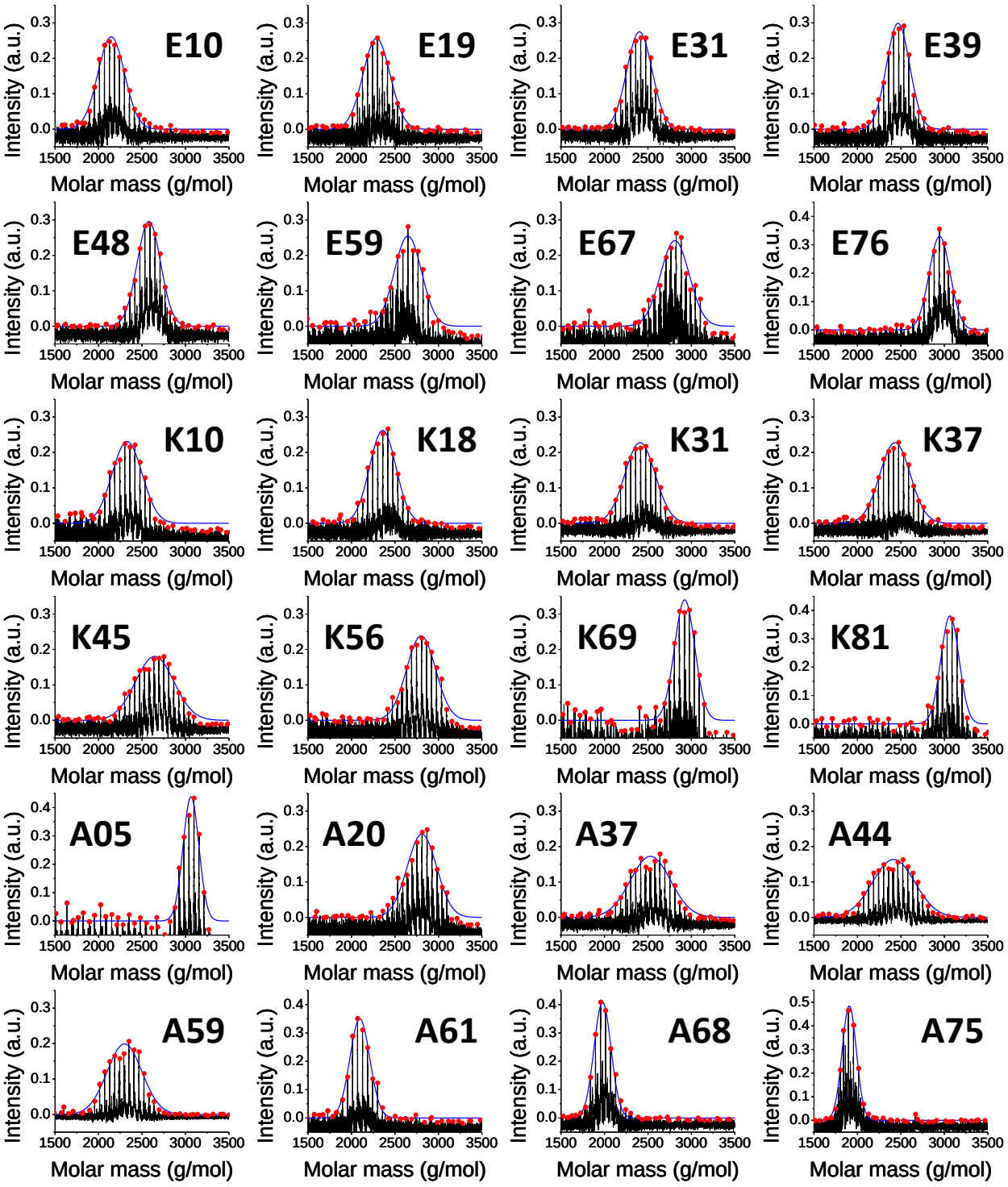
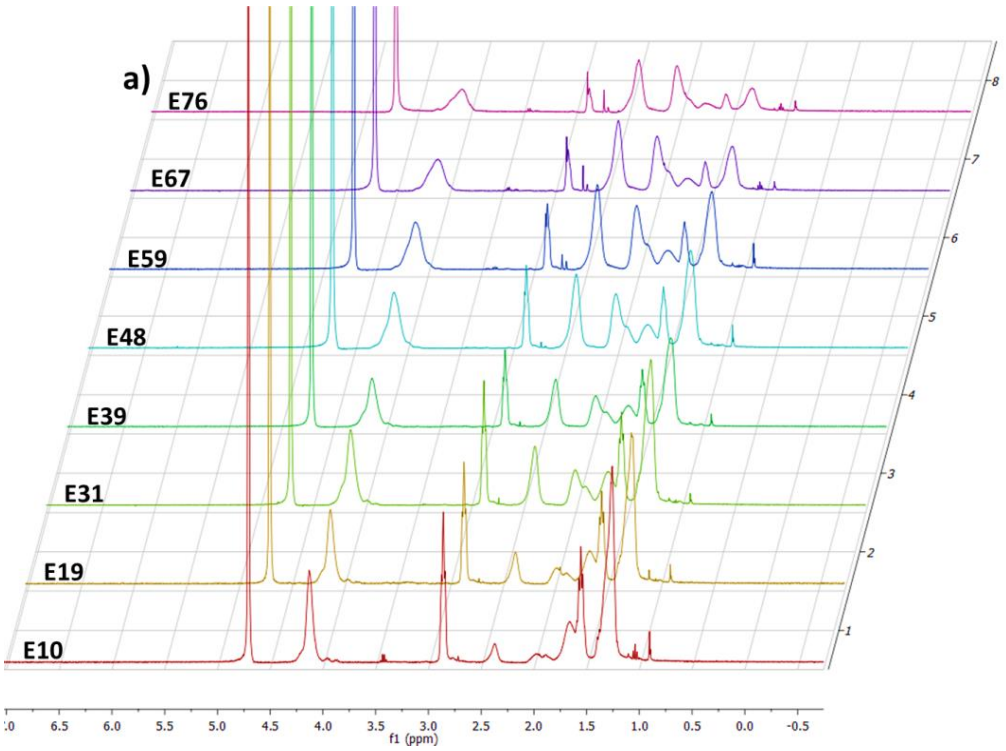


Figure S1. Experimental MALDI-TOF patterns (black) of all copolymer batches and corresponding normal distributions (blue lines) fitted to the positions and intensities (red dots) of the peaks of the observed major distributions.

^{1}H NMR spectroscopy

^{1}H NMR analyses were performed on a Bruker ^{1}H NMR spectrometer, type Mercury 400. For measuring the polypeptide spectra deuterated water (D_2O) was used. The amino acid composition of the polypeptides was determined by comparing the peak intensities corresponding to the signals characteristic for *L*-glutamate (2.48 ppm, 2H) and *L*-lysine (2.98 ppm, 2H) (see Figure S2).

^{1}H NMR (400 MHz, D_2O): Lysine: 1.40 (2H, CH_2), 1.70 (4H, CH_2CH_2), 2.98 (2H, CH_2), 4.2 (1H, CH); Glutamic acid: 2.00 (2H, CH_2), 2.48 (2H, CH_2), 4.2 (1H, CH); Alanine: 1.35 (3H, CH_3), 4.2 (1H, CH).



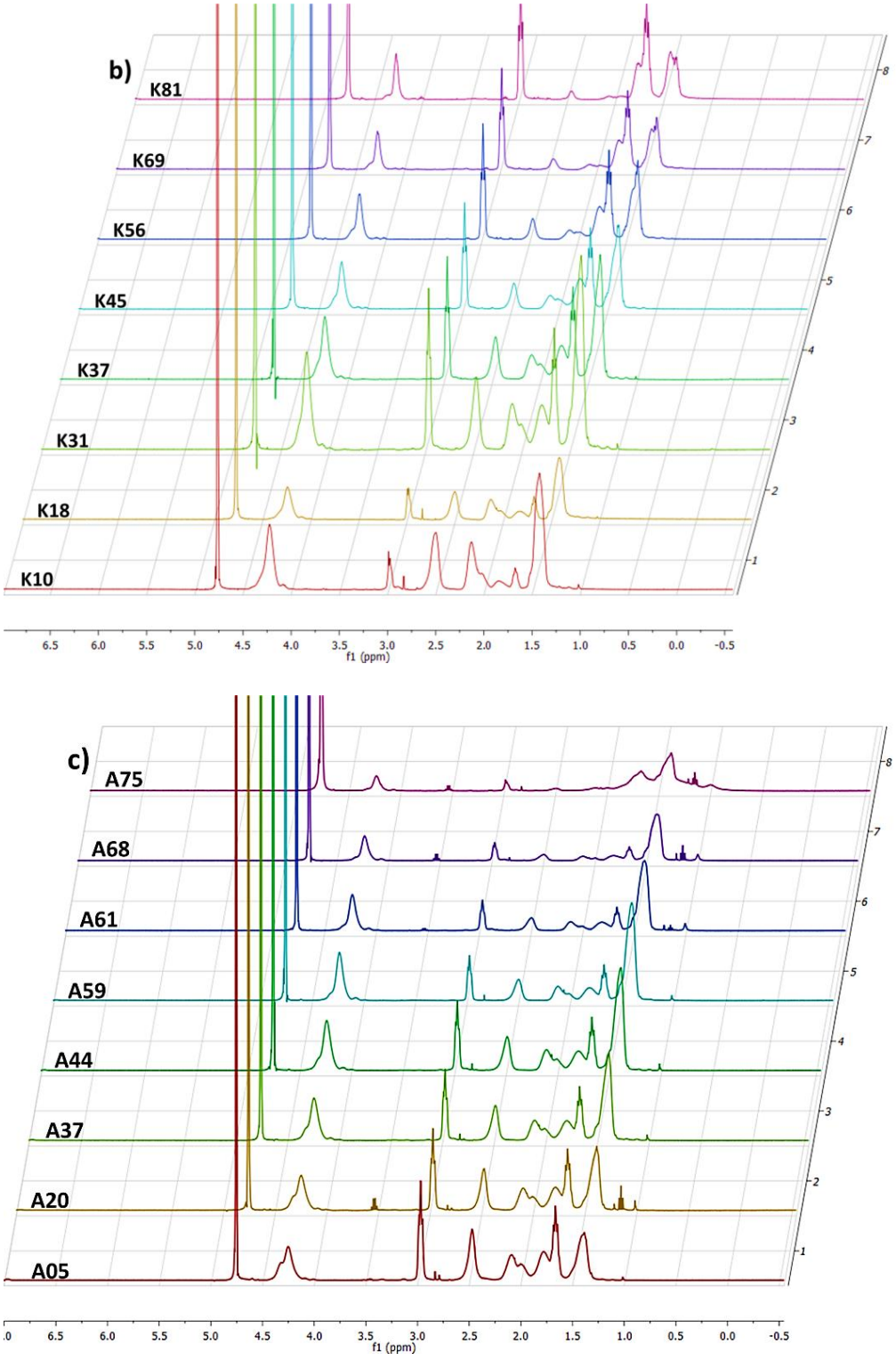


Figure S2. ^{1}H NMR of a) $\mathbf{E_{10}}$-$\mathbf{E_{76}}$, b) $\mathbf{K_{10}}$-$\mathbf{K_{81}}$ and c) $\mathbf{A_{05}}$-$\mathbf{A_{75}}$, where a systematic increase in the peaks characteristic for Glu, Lys and Ala is observed, respectively.

Acid hydrolysis experiments

Briefly, 10 µL of IS (internal standards) ($C^{13}N^{15}$ amino acids) were added to 10 µL of the synthetic peptide solution (500 times) followed by the addition of 70 µL borate buffer (pH 8.5), After 10 s vortexing, 20 µL of AQC reagent was added and the mixture was vortexed again. The sample was heated for 10 min at 55 ºC. After cooling down a 1 µL sample of the reaction mixture was injected into the ultra-pressure liquid chromatography – mass spectrometry/mass spectrometry (UPLC-MS/MS system).[1]

An ACQUITY UPLC™ system with autosampler (Waters Chromatography B.V., Etten-Leur, The Netherlands) was coupled online with a Quattro Premier Xe Tandem quadrupole mass spectrometer (Waters Corporation) and was used in the positive ion electrospray mode. The instrument was operated under Masslynx data acquisition software (Waters, version 4.1). The samples were analyzed by UPLC-MS/MS using a AccQ-Tag™ Ultra 2.1 x 100 mm column (Waters, 1.7 µm particle size). A binary gradient system of water – eluent A (10:1, v/v) (AccQ-Tag, Waters) and 100% eluent B (AccQ-Tag, Waters) was used. Elution of the analytes was achieved by ramping the percentage of eluent B from 0.1% to 90.0% in approximately 9.5 minutes, using a combination of both linear and convex profiles. The flow rate was 0.7 mL/min. The column temperature was maintained at 60 ºC and the temperature of the autosampler tray was set to 10 ºC. After each injection the injection needle was washed with 200 µL strong wash solvent (95% ACN) and 600 µL weak wash solvent (5% ACN).

The Xevo triple quadrupole was used in the positive ion electrospray mode and all analytes were monitored in Selective Reaction Monitoring (SRM) using nominal mass resolution (FWHM 0.7 amu). Next to the derivatization reagent all amino acids were selectively monitored via the transition from the protonated molecule of the AccQ-Tag derivative to the common fragment at m/z 171. Collision energy and collision gas (Ar) pressure were 22 eV and 2.5 mbar, respectively. The complete chromatogram was divided into 4 time windows, restricting the number of SRM transitions to follow and allowing quantitative information to be gathered in each segment. Acquired data was evaluated using Quantlynx software (Waters).

[1] S.M. Rutherfurd, G.S. Gilani, *Current protocols in protein science* **2009**, *11*, Unit 11.9.

pH-net charge relationships from acid-base titrations

Acid-base titrations were carried out using a titration setup from Metrohm Applikon, consisting of three 2 mL Dosino dosing units, a glass pH microelectrode and a magnetic stirring plate, connected to two 809 Titrando titration units and controlled by a computer running Tiamo 2.2 software. One dosing unit was connected to a bottle containing HCl solution (~1.2 M) and one to a bottle containing NaOH solution (~0.5 M).

Aqueous solutions (V_0 = 25 mL) were prepared by dissolving polypeptide (~1 mg/mL) in ultrapure water (resistivity 18.2 MΩ·cm at 25 °C).

Acid-base titrations were carried out as follows. First, HCl solution (V_{HCl}) was titrated into the polypeptide solution while stirring magnetically until a pH of 2 was reached. Second, NaOH solution was titrated at a constant rate (v_{NaOH} = 10 µL/min) while recording the pH over time (t).

The exact concentrations of the HCl and NaOH solutions [HCl] and [NaOH] and the water dissociation constant were determined from the time-pH data of an acid-base titration of ultrapure water (Figure S3a) by solving the set of equations below for all time points while taking into account charge neutrality:

$$[H^+]+[Na^+]=[OH^-]+[Cl^-]$$

with $[H^+]=10^{-pH}$, $[OH^-]=10^{-(pK_w-pH)}$, $[Na^+]=[NaOH]\cdot\frac{V_{NaOH}}{V}$, $[Cl^-]=[HCl]\cdot\frac{V_{HCl}}{V}$,

$$V=V_0+V_{HCl}+V_{NaOH}=V_0+V_{HCl}+v_{NaOH}\cdot t$$

and V_0 = 25 mL, V_{HCl} = 0.2126 mL, v_{NaOH} = 10 µL/min,

resulting in [HCl] = 1.2481 M, [NaOH] = 0.5253 M and pKw = 14.12.

As can be seen in Figure S3b, charge neutrality is maintained over the complete pH range.

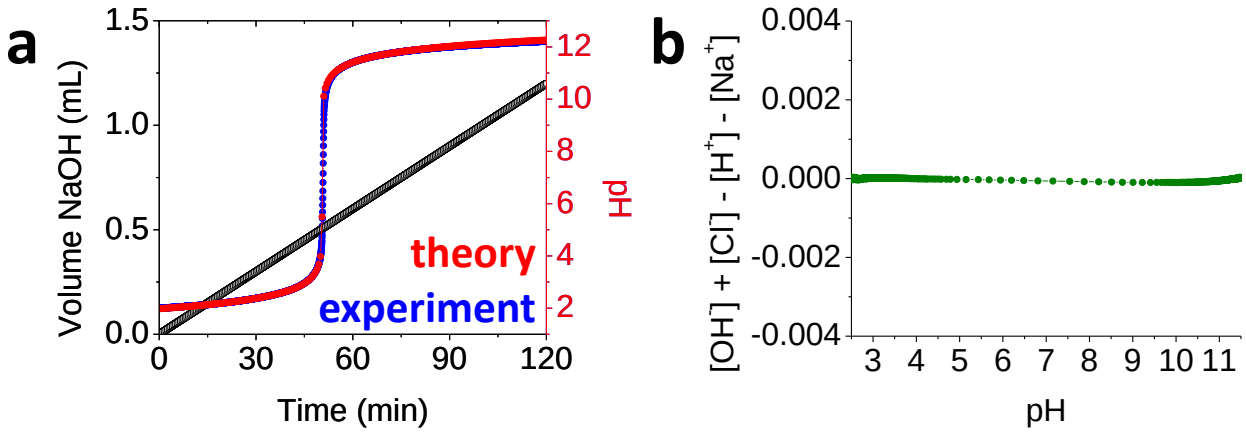


Figure S3. a) Time-pH data from the acid-base titration of ultrapure water. The experimental data is in excellent agreement with the theoretical prediction. b) For [HCl] = 1.2481 M, [NaOH] = 0.5253 M and pK_w = 14.12, charge neutrality is maintained over the complete pH range.

Then, theoretical predictions for the time-pH data of acid-base titrations of various polypeptides (Figure S4) were calculated by assuming the following dissociation equilibria for water and the glutamic acid (E), lysine (K) and terminal amine (N) residues:

$$K_w = [H^+]\cdot[OH^-] = 10^{-14.12} \qquad [H^+] = 10^{-pH},\ [OH^-] = 10^{-(pK_w - pH)}$$

$$K_E = \frac{[H^+]\cdot[E^-]}{[E^0]} = 10^{-4.15} \qquad [E] = [E^0] + [E^-] \qquad \rightarrow [E^-] = \frac{[E]\cdot K_E}{[H^+] + K_E}$$

$$K_K = \frac{[H^+]\cdot[K^0]}{[K^+]} = 10^{-10.67} \qquad [K] = [K^+] + [K^0] \qquad \rightarrow [K^0] = \frac{[K]\cdot K_K}{[H^+] + K_K}$$

$$K_N = \frac{[H^+]\cdot[N^0]}{[N^+]} = 10^{-9.48} \qquad [N] = [N^+] + [N^0] \qquad \rightarrow [N^0] = \frac{[N]\cdot K_N}{[H^+] + K_N}$$

The pH = –log([H^+]) was calculated for all time points by solving this set of equations while taking into account charge neutrality:

$$[H^+] + [Na^+] + [K^+] + [N^+] = [OH^-] + [Cl^-] + [E^-] + [TFA^-]$$

The exact copolymer and trifluoroacetic acid concentrations (because the lysine and terminal amine moieties can be TFA salts) could be determined by fitting the inflection point (steepest part of the curve) of the theoretical prediction to the experimental data (Figure S4). The E, K and A monomer concentrations could be calculated from the as such determined copolymer concentrations, using the molecular weight and degree of polymerization DP as determined by ^{1}H NMR and MALDI-TOF (see Table 1 in the main text).

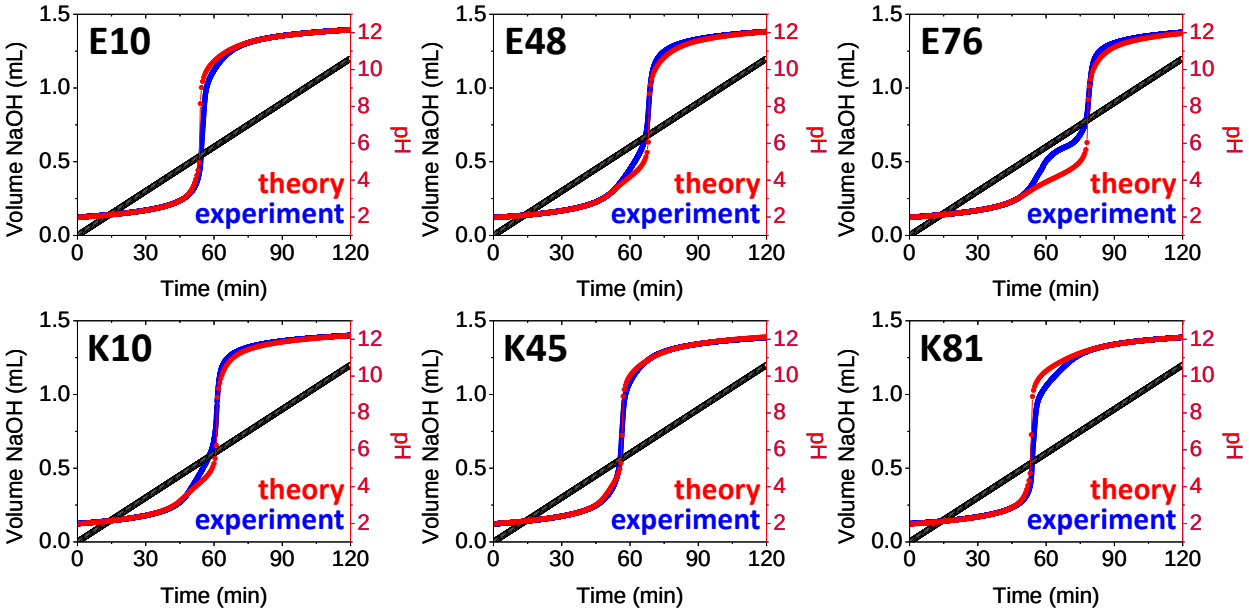


Figure S4. Time-pH data from acid-base titrations of $\mathbf{E_{10}}$, $\mathbf{E_{48}}$, $\mathbf{E_{76}}$, $\mathbf{K_{10}}$, $\mathbf{K_{45}}$ and $\mathbf{K_{81}}$. Comparison of theoretical predictions with experimental data.

Further, theoretical and experimental pH-net charge relationships as shown in Figure 7 were extracted from the theoretical and experimental time-pH data of acid-base titrations (see Figure S4) by defining net charge as the difference between protonated lysine (K^+) and terminal amine (N^+) residues and deprotonated glutamic acid (E^-) residues per copolymer chain and taking into account charge neutrality:

$$\text{net charge} = \frac{[K^+]+[N^+]-[E^-]}{[EKA]} = \frac{[OH^-]+[Cl^-]+[TFA^-]-[H^+]-[Na^+]}{[EKA]}$$

with $[EKA]=\frac{[E]+[K]+[A]}{DP}$ and all the other concentrations as defined earlier.

Circular dichroism (CD) spectroscopy

Circular dichroism (CD) spectra were acquired at room temperature on a Jasco J-815 CD spectrometer between 190 nm and 270 nm in 1 nm increments using a quartz glass cuvette with a path length of 1 cm. Scanning speed and response time were chosen appropriately. The CD spectra were corrected using Milli-Q water or 3 mM $CaCl_2$ solution as a background reference, respectively. The polypeptide concentration was 1 mM residues for all samples. The per residue molar ellipticities were calculated under the assumption that all lysine and terminal amine moieties are trifluoroacetic acid salts.

Cryogenic transmission electron microscopy (cryoTEM)

Copolymer solutions were prepared by dissolving 1 mg/mL copolymer and adjusting the pH to 7 using NaOH solution. Sample preparation was performed using an automated vitrification robot (FEI Vitrobot Mark III) for plunging in liquid ethane.[1] For cryoTEM, 200 mesh Cu grids with Quantifoil R 2/2 holey carbon films (Quantifoil Micro Tools GmbH) were used. All TEM grids were surface plasma treated for 40 seconds using a Cressington 208 carbon coater prior to use. CryoTEM samples were studied on the TU/e cryoTITAN (FEI, www.cryotem.nl) operated at 300 kV, equipped with a field emission gun (FEG), a post-column Gatan Energy Filter (GIF) and a post-GIF 2k x 2k Gatan CCD camera. Gatan DigitalMicrograph and ImageJ were used for TEM image analysis.

[1] M.R. Vos, P.H.H. Bomans, P.M. Frederik, N.A.J.M. Sommerdijk, *Ultramicroscopy* **2008**, *108*, 1478-1483.

Dynamic light scattering (DLS)

DLS measurements were performed on a Zetasizer Nano series instrument (Malvern Instruments Ltd., UK). Copolymer solutions were prepared by dissolving 1 mg/mL copolymer and adjusting the pH to 7 using NaOH solution. For DLS, a standard measuring angle of 173° and a 4 mW He-Ne laser with a wavelength of 633 nm were used. All measurements are presented as number-weighted size distributions and are an average of 10 runs over 10 seconds each. Multiple distributions are presented per sample to show the variation between consecutive measurements due to the typically low count rates. Measurements showing size maxima at 0.6-0.7 nm, which is the detection limit of the DLS instrument, were considered artifacts and left out for clarity, as well as some measurements with size maxima >100 nm due to trace sample inhomogeneities.

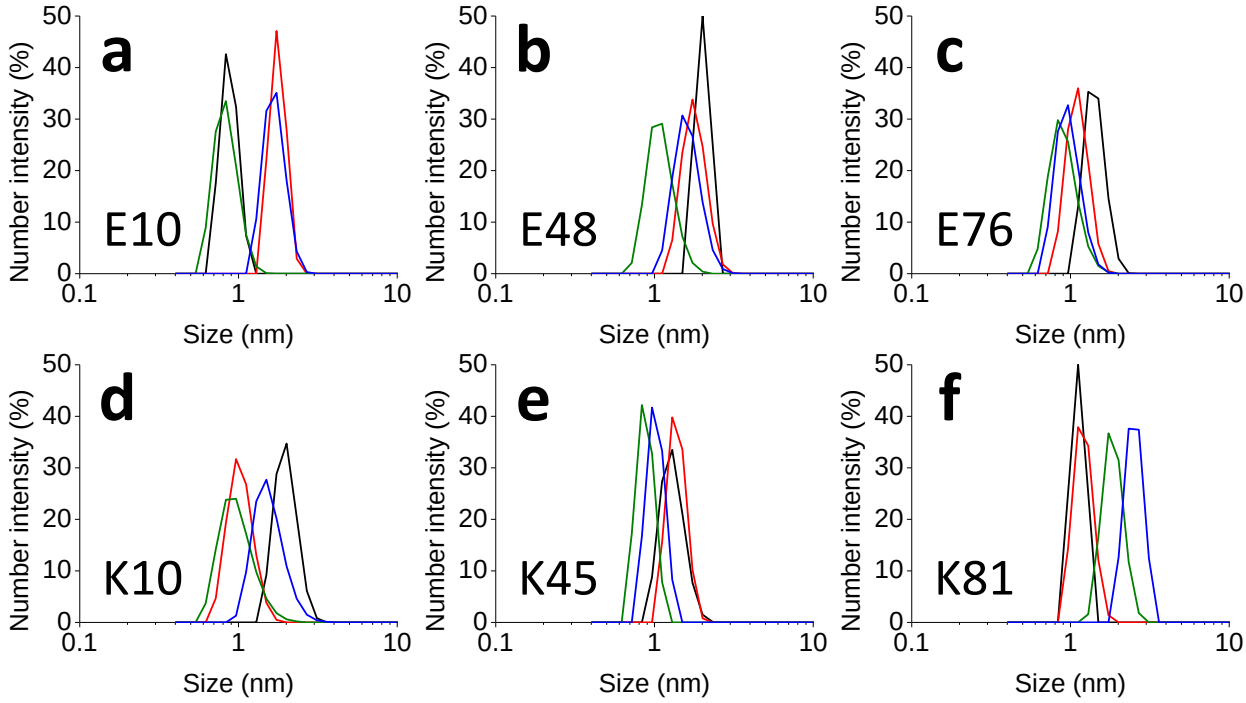


Figure S5. DLS size distributions (plotted as number-weighted intensity) for a) $\mathbf{E_{10}}$, b) $\mathbf{E_{48}}$, c) $\mathbf{E_{76}}$, d) $\mathbf{K_{10}}$, e) $\mathbf{K_{45}}$ and f) $\mathbf{K_{81}}$ in aqueous solution (1 mg/mL, pH 7). Sizes <4 nm are observed.