Supporting information for

Influence of Charge Block Length on Conformation and Solution Behavior of Polyampholytes

Winnie H. Shi,[†] Rohan S. Adhikari,[†] Dilipkumar N. Asthagiri,[‡] and Amanda B. Marciel[*,†]

[†]*Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, USA*

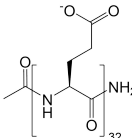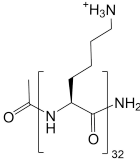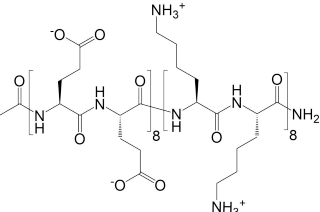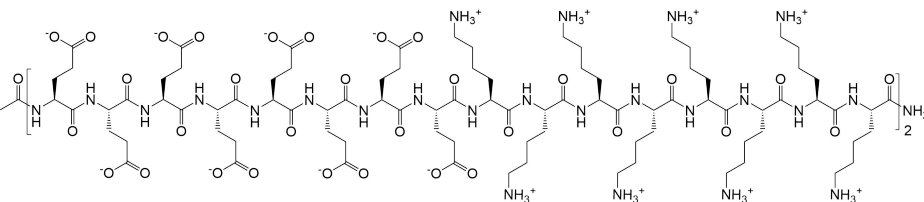[‡]*Oak Ridge National Laboratory, Oak Ridge, TN 37830-6012*

E-mail: am152@rice.edu

Contents

Structures

Table S1: Polyampholytes, chemical structure, and molecular weight (M_w) determined by MALDI-ToF mass spectrometry. The polyampholytes are designated by the single letter amino acid code. The calculated molecular weight is included in parenthesis. Yield after Reversed-Phase High Performance Liquid Chromatography (RP-HPLC) purification is also noted.

Peptide	M_w experiment (calculated)	% Yield
E_{32}	4160 Da (4158.51 Da)	28
K_{32}	4175 Da (4192.91 Da)	36
$E_{16}K_{16}$	4176 Da (4175.71 Da)	19
$(E_8K_8)_2$	4174 Da (4175.71 Da)	25
$(E_4K_4)_4$	4179 Da (4175.71 Da)	28
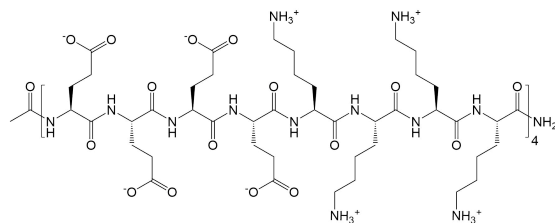
$(E_2K_2)_8$ 4176 Da (4175.71 Da)   17
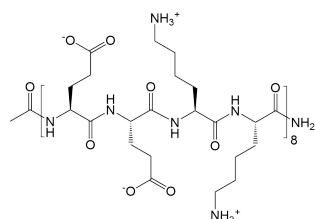
$(EK)_{16}$ 4176 Da (4175.71 Da)   22
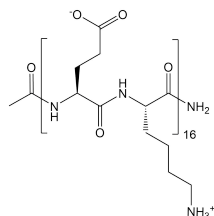
Synthesis

Solid-phase peptide synthesis (SPPS) and cleavage

Sequence specific polyampholytes of increasing block size N_p — $(EK)_{16}$, $(E_2K_2)_8$, $(E_4K_4)_4$, $(E_8K_8)_2$, and $(E_{16}K_{16})$ — as well as polyelectrolytes (E_{32} and K_{32}) were synthesized using Fmoc-based SPPS on a Prelude® X peptide synthesizer from Gyros Protein Technologies. Fmoc-based SPPS at 0.15 mmole scale was carried out on Rink Amide MBHA resin LL (Novabiochem) using Fmoc-L-Lys(Boc)-OH and Fmoc-L-Glu(OtBu)-OH amino acid reagents, O-(6-Chloro- benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HCTU) coupling reagent, and N-Methylmorpholine (NMM) base solution in N,N-Dimethylformamide (DMF) solvent from Gyros Protein Technologies at an excess of 4:3.93:7.86, respectively.[1,2] The Fmoc groups were deprotected using two rounds of mixing with 20 % v/v piperi-

dine/DMF mixture (Gyros Protein Technologies) before each coupling and capped using a 25 % v/v acetic acid/DMF solution after each coupling past the 10th amino acid. For $N_p \leq 4$, the first 11 amino acids were synthesized using single coupling for 20 minutes and every subsequent amino acid was single coupled for 55 minutes. For $N_p > 4$ and homopolymers, the first 11 amino acids were double coupled for 15 minutes each and subsequent amino acids were double coupled for 35 minutes each. Lastly, the polyampholytes were capped at the N-terminus using 25 % v/v acetic acid/DMF solution. A ninhydrin test was performed upon synthesis completion to ensure successful capping. The resin was oven dried overnight to remove excess solvent, and polyampholytes cleaved by shaking in a 95/2.5/2.5 % v/v trifluoroacetic acid (TFA)/H_2O/Triisopropylsilane (TIPS) mixture for 3 hours. Cleaved polyampholytes were separated from resin using a fritted syringe (Torviq). TFA was vaporized by blowing N_2 until a viscous yellow liquid remained. The polyampholytes were precipitated in -30 °C diethyl ether (Sigma Aldrich) and centrifuged. The pellet containing polyampholyte was then dried, dissolved in water, frozen, and lyophilized.

Reversed-Phase High Performance Liquid Chromatography (RP-HPLC), ion exchange, and dialysis

Polyampholytes were purified using Reversed-Phase High Performance Liquid Chromatography (RP-HPLC) with 0.1 % v/v TFA/H_2O for mobile phase A and 0.1 % v/v TFA/acetonitrile for mobile phase B gradient using an Agilent ZORBAX 3000SB-C18 column. Samples were prepared at 20-25 mg lyophilized product per 1 mL mobile phase A. For each round of purification, a 10 minute 0 - 38 % v/v mobile phase B gradient was applied at a steady flow rate of 4 mL/min with 250 μL sample injection volume. The polyampholyte containing peak elutes between 11 - 23 % v/v mobile phase B depending on sequence. Figure S1A illustrates the RP-HPLC mobile phase gradient along with an example of peak collection. Once purified, polyampholytes were characterized and confirmed using matrix assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry using Bruker AutoFlex™ Speed

mass spectrometer (Table S1).

TFA ions were then exchanged for chloride ions using AG® 1-X2 anion exchange resin (Bio-Rad). Purified polyampholytes were mixed in water at 10 mg/mL with 10x resin by weight for 1 hour in a fritted syringe (Torviq) on a rocker at 15° and 40rpm. The resin was filtered out, and ion exchange was confirmed using 19F NMR (Figure S1B).

Polyampholytes were pH adjusted to pH 7 using 6 M NaOH and 1 M HCl before dialysis. 1 wt% pH adjusted polyampholyte solutions were dialyzed using Spectra/Por® 6 Standard RC dialysis membranes with 1 kD MW cut off. 1 L Milli-Q water was used as dialysate and exchanged completely 3 - 4 times per sample with 2 - 4 hours between each exchange. The final salt concentration in solution was determined using ZetaPALS zeta potential analyzer (Brookhaven Instruments) with 0.5 cm SR-560 probe. Conductivities were measured at 0.1 wt% and compared to a calibration curve from 0 to 100 mM NaCl. Final salt concentrations are listed in Table S2.
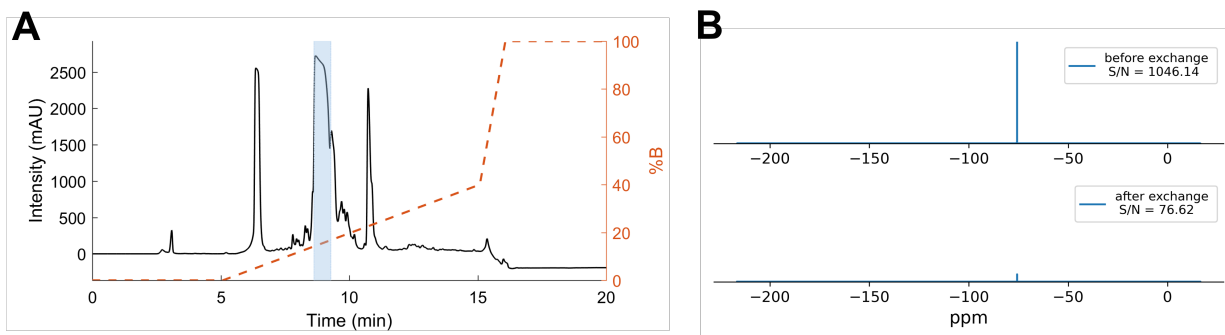
Figure S1: (A) Representative chromatogram of RP-HPLC purification fraction collection and (B) 19F NMR spectra of $(E_2K_2)_8$. The blue region in (A) highlights the collected polyampholyte containing peak. In (B), top and bottom 19F NMR spectra correspond to before and after ion exchange, respectively.

Solution phase behavior characterization

Solution phase behavior was evaluated at pH7 for 0, 100, 500, 1000, and 3000 mM NaCl concentrations. Polyampholyte and homopolymer polyelectrolyte samples at 1 wt% were

Table S2: Salt concentration of dialyzed samples determined from conductivity measurements

Polyelectrolyte/Polyampholyte	NaCl concentration after dialysis (mM)	Molar equivalent NaCl to polyampholyte (mol NaCl/mol polyampholyte)
E_{32}	3.24	13.6
K_{32}	4.75	19.9
$E_{16}K_{16}$	0.301	3.77
$(E_8K_8)_2$	0.137	1.43
$(E_4K_4)_4$	0.178	1.86
$(E_2K_2)_8$	0.173	1.80
$(EK)_{16}$	0.112	1.16

prepared by dissolving lyophilized polypeptide in Milli-Q water at 4 wt% and adjusted to desired salt concentration using 6 M NaCl and Milli-Q water. Polyelectrolyte samples $K_{32}+E_{32}$ and $K_{20}+E_{20}$ at 1 wt% were prepared by dissolving each lyophilized polyelectrolyte powder in Milli-Q water at 4 wt% and combined in the order of poly-L-lysine then 6mM NaCl and Milli-Q water then poly-L-glutamate.[3] K_{20} and E_{20} were purchased from Alamanda ™Polymers and dialyzed and lyophilized according to aforementioned methods. All sample solutions were prepared at 0.2 mL and mixed by vortexing for 30 s and rotating overnight on tube rotator before turbidity measurements(Figure 2 and S2). Turbidity of 1 wt% charged polymer solutions was measured at 500 nm in a 96-well plate using a TECAN Infinite®M1000 Pro microplate reader. The plate was shaken orbitally at 3 mm diameter for 10 seconds before each scan. Transmittance T was calculated from absorbance A using

$$A = -\log_{10}(\frac{I}{I_0}) = -\log_{10}(T)\,, \tag{S.1}$$

where I is light intensity through the sample and I_0 is light intensity through the solvent solution. The resulting values were calculated and plotted with an in-house Jupyter Notebook script.
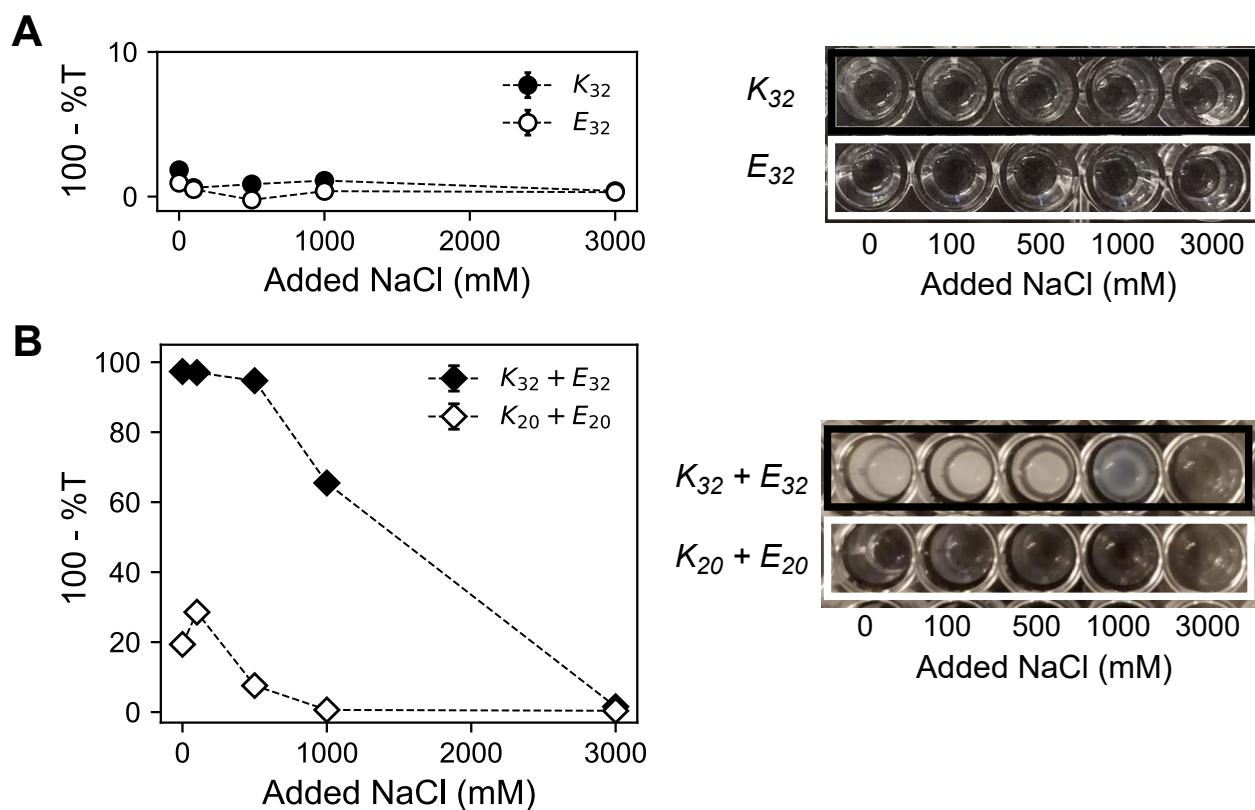
Figure S2: (A) Solution phase behavior of polyelectrolytes K_{32} and E_{32} at 1 wt% at pH 7 for 0, 100, 500, 1000, and 3000 mM NaCl quantified through turbidity measurements. (B) Solution phase behavior of polyelectrolyte complexes $K_{32} + E_{32}$ and $K_{20} + E_{20}$ at 1 wt% at pH 7 for 0, 100, 500, 1000, and 3000 mM NaCl quantified through turbidity measurements. Dashed lines in plots are used to help guide the eye. Photos of solutions in well-plate for visual reference.

Secondary structure characterization

Attenuated total reflection Fourier-transform infrared spectroscopy (ATR FTIR)

ATR FTIR was conducted on a Nicolet™ FTIR Microscope. Samples were prepared at 2 wt% in 100 % D_2O (Cambridge Isotope Laboratories, Inc.) at 0 mM and 3000 mM added NaCl at pH 7. Absorbance measurements were taken with a resolution of 8.000 μm with 64 scans averaging. Air was used as baseline and corresponding D_2O solvent background absorption was subtracted manually. Spectra were normalized to the carbonyl stretching peak at 1564 cm^{-1}. Peaks were identified using local maximum and thresholding. All background subtraction, normalization, and peak identification was performed using an in-house Jupyter notebook script.

In 0 mM added NaCl, the ATR FTIR spectra show the presence of β sheet secondary structure in most polyampholytes except $(E_4K_4)_4$. Secondary structure is more prominent for phase separated sequences $(E_8K_8)_2$ and $(E_{16}K_{16})$ (Figure S3A). In 3000 mM added NaCl, all sequences show two peaks at $\sim 1645\,\text{cm}^{-1}$ and $\sim 1563\,\text{cm}^{-1}$ indicative of a random coil like conformation (Figure S3B).

Circular dichroism (CD)

CD measurements were taken on a Jasco J-1500 circular dichroism spectrophotometer. For each measurement, 800 μL samples of 0.1 wt% polyampholyte were loaded into 1 cm pathlength cuvette (9-F-Q, Starna Cells). Ellipticity was measured from 185-260 nm, with 2 accumulations at a digital integration time (D.I.T) of 2 seconds, scanning speed of 50 nm/min, and 1.00 nm bandwith at a temperature of 20 °C. All collected signals were background subtracted. The exported data was then converted from millidegrees, $m°$, to mean residue
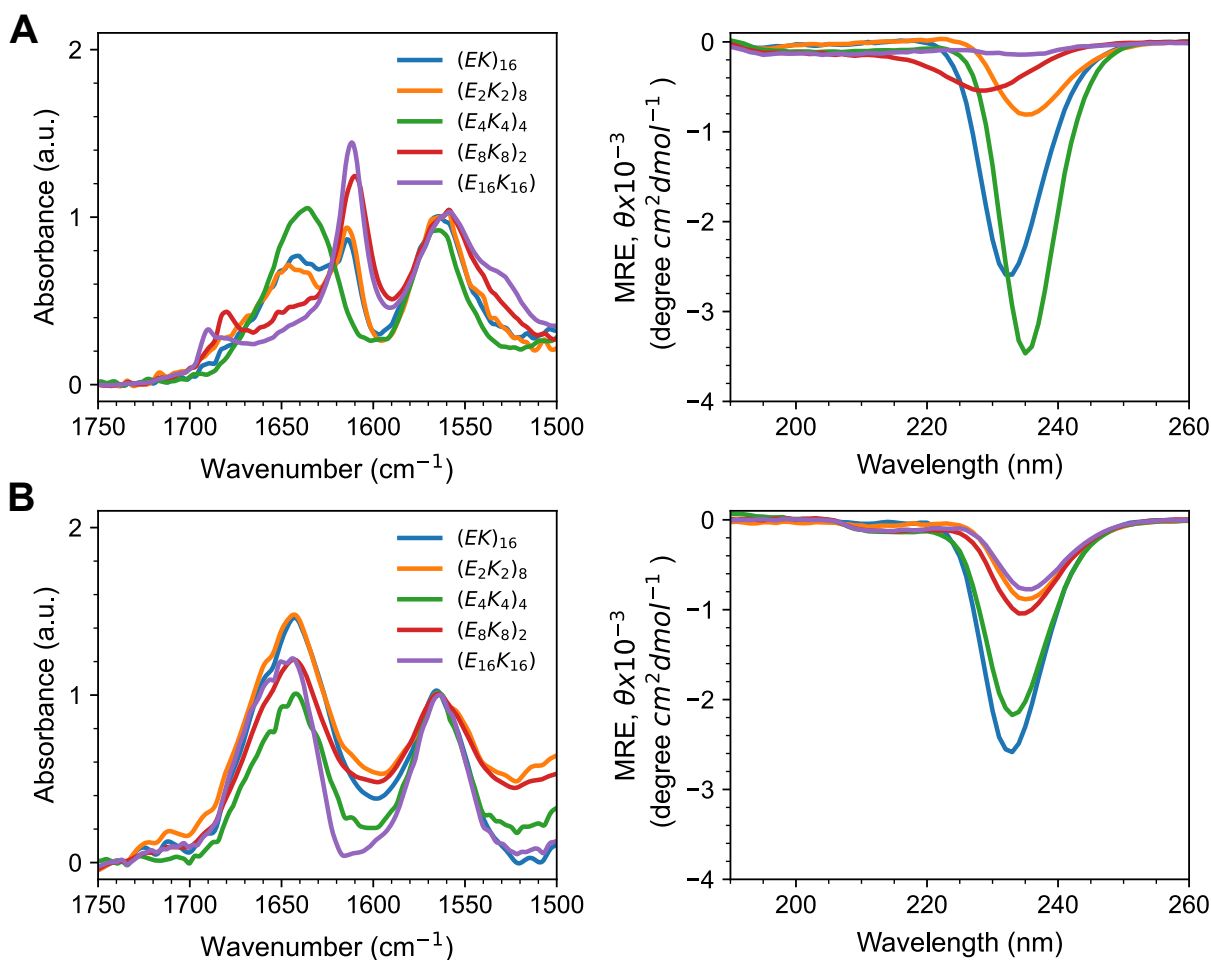
Figure S3: FTIR and CD spectra of polyampholytes at (A) 0 mM added NaCl and (B) 3000 mM added NaCl. Primary peaks in wavenumbers for ATR FTIR are as follows for (A) $(E_{16}K_{16})$: 1560, 1613, 1690 cm^{-1}; $(E_8K_8)_2$: 1561, 1611, 1680 cm^{-1}; $(E_4K_4)_4$: 1565, 1636 cm^{-1}; $(E_2K_2)_8$: 1564, 1614, 1644 cm^{-1}; and $(EK)_{16}$: 1564, 1614, 1641 cm^{-1}; and (B) $(E_{16}K_{16})$: 1564, 1649 cm^{-1}; $(E_8K_8)_2$: 1564, 1645 cm^{-1}; $(E_4K_4)_4$: 1563, 1643 cm^{-1}; $(E_2K_2)_8$: 1563, 1646 cm^{-1}; and $(EK)_{16}$: 1567, 1645 cm^{-1}.

ellipticity (MRE)

$$MRE = \frac{m^{\circ} \cdot M_w}{10 \cdot L \cdot C \cdot N} \,, \tag{S.2}$$

where L is the cell path length (cm), C is the concentration (g/L), and N is the number of amino acids present in the polyampholyte. All data was analyzed with in-house Jupyter Notebook scripts.

A negative peak between 230-240 nm is observed for all polyampholytes. This peak correlates with the relative magnitude of the 1644 cm^{-1} peak observed with ATR FTIR. Negative maxima near 230 nm have been observed for proteins with β-sheet structures.[4]

Viscometry

Dilute concentration regime was determined via intrinsic viscosity $[\eta]$ measurements. The measurements were performed using a Cannon-Ubbelohde Semi-Micro manual viscometer (CANNON Instrument Company) submerged in a water bath of 20 °C to ensure consistent temperature conditions. Polyampholyte solutions were prepared at 5.5 wt% in Milli-Q water and then diluted and mixed in the viscometer until the solution kinematic viscosity was the same as water (1.003 mm^2/s at 20°C).[5] Viscosity was measured in triplicate at each concentration.

The time, t recorded from the viscometer was converted to kinematic viscosity, ν using $\nu = K \cdot t$, where κ is the viscometer constant empirically determined by the manufacturer. Since we were well within the viscosity range of the viscometer, the Hagenback-Couette correction term was not needed.[6] Kinematic viscosity was converted to dynamic viscosity, η with $\nu = \eta/\rho$ where ρ is the density of the solution. Due to the low polyampholyte concentrations, the density of water at 20 °C was used. Intrinsic viscosity $[\eta]$ was extrapolated from Huggins and Kraemer plots (Figure S4A) for $\eta > 1.1 dLg^{-1}$.[7] For Huggins plot, a linear

fit was applied to $\frac{\eta - \eta_s}{\eta_s c}$ versus concentration c according to Huggins equation

$$\frac{\eta - \eta_s}{\eta_s c} = [\eta] + k_H[\eta]^2 c + ... \tag{S.3}$$

where η_s is the solvent viscosity and k_H is the Huggins coefficient. For Kraemer plot, a linear fit was applied to $\frac{ln(\eta/\eta_s)}{c}$ versus c according to Kraemer equation,

$$\frac{\ln(\eta/\eta_s)}{c} = [\eta] + k_K[\eta]^2 c + ... \tag{S.4}$$

where k_K is the Kraemer coefficient.[7]

Intrinsic viscosity was extrapolated from the intercept of the Huggins and Kraemer equations and Huggins and Kraemer coefficients were determined from the slope. Using $[\eta]$, The radius of hydration R_h was estimated from Einstein's viscosity equation for a suspension of spheres in solution[7]

$$[\eta] = \frac{5}{2}\frac{N_A V_h}{M}. \tag{S.5}$$

Here, N_A is Avogadro's number, M is the molecular weight of the solute, and V_h is the hydrodynamic volume which we defined as $V_h \approx \frac{4}{3}\pi R_g^3$ (sphere approximation) for calculating R_g. We further plotted specific viscosity $\eta_{sp} = \frac{\eta - \eta_s}{\eta_s}$ relative to $[\eta]c$, using $[\eta]$ determined from Huggins and Kraemer plots (Figure S4B). Here, specific viscosity scales to concentration with a factor of 1.1, which is close to 1, indicative of a dilute solution. Hence, from Huggins and Kraemer equations as well as scaling factor determination, it is determined the concentrations used for all experiments in this study lie well within dilute solution concentration regime.
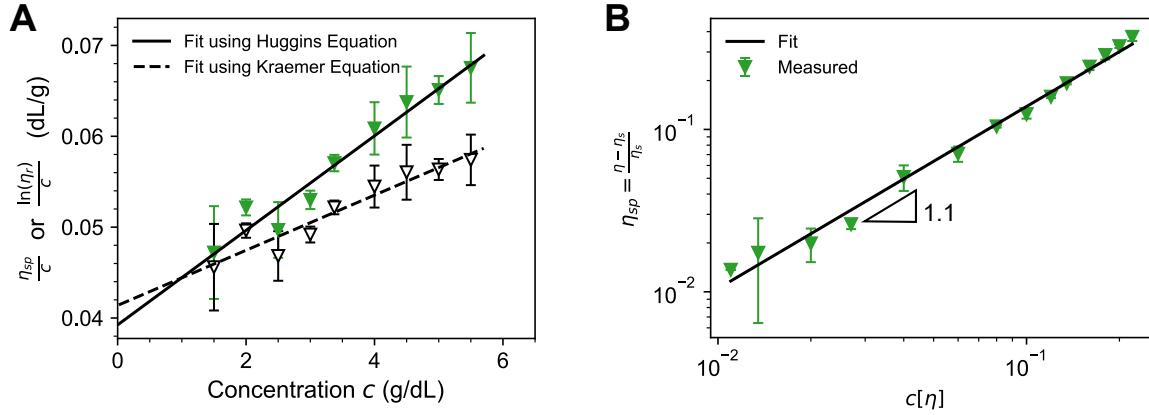
Figure S4: (A) Determination of intrinsic viscosity of $(E_4K_4)_4$ using Huggins plot (solid) and Kraemer plot (dashed). $[\eta]$ was determined to be ~ 4.0 dL g^{-1} which approximates to R_g of 1.3Å, consistent with R_g from SAXS. k_H and k_K were determined to be 3.9 and 2.0 respectively. Both coefficients are positive which indicates intra-chain interactions[8] consistent with observations from SAXS. (B) Specific viscosity η_{sp} of $(E_4K_4)_4$ with respect to dimensionless solution concentration. Slope was determined to be 1.1 indicative of dilute concentration regime.[9]

Small angle X-ray scattering (SAXS)

Sample preparation and experimental methods

Polyampholytes were dissolved in 0, 100, 500, 1000, or 3000 mM NaCl in Milli-Q water at 2 wt%. Each sample was mixed on a tube rotator in oscillation mode for at least 12 hours before loading into 2 mm OD quartz capillaries (Charles Supper). SAXS experiments were performed at beamline 12-ID-B at the Advanced Photon Source at Argonne National Laboratory. Each sample and its corresponding solvent were shot three times at 0.1 second exposure per shot. The samples were moved vertically after each shot to minimize beam damage. Scattering profiles were averaged in triplicate and normalized by transmission at the beamline. Background subtraction of the sample from the solvent was performed manually using in-house Jupyter Notebook scripts with a multiplication factor α on the background solvent such that the spectra converge at high q values to adjust for inconsistencies with capillary glass thickness (Figure S7A).

Guinier analysis

Guinier analysis was used to find radius of gyration (R_g). The background subtracted spectra were fit to

$$I(q) = I_0 \, \exp\left(\frac{-R_g^2 q^2}{3}\right) \tag{S.6}$$

by performing a linear fit to $ln(I(q))$ vs. q^2 plot in the domain $qR_g < 1.3$ for globular particles. The q range for fitting was determined using Guinier analysis tool in BioXTAS RAW.[10] The data was refit using in-house Jupyter Notebook script to find error bars of one standard deviation σ as shown in Figure 3B and 5B. Example linear fit of background subtracted data for Guinier analysis is shown in Figure S7B and resulting R_g is tabulated in Table S3.

Table S3: Radius of gyration R_g from Guinier analysis for all polyampholytes

Polyampholyte	Added NaCl (mM)	R_g (Å)
$E_{16}K_{16}$	3000	16.1±0.4
$(E_8K_8)_2$	1000	16.7±0.2
$(E_8K_8)_2$	3000	16.9± 0.7
$(E_4K_4)_4$	0	14.1± 0.2
$(E_4K_4)_4$	100	15.6± 0.2
$(E_4K_4)_4$	500	15.4± 0.2
$(E_4K_4)_4$	1000	17.8± 0.3
$(E_4K_4)_4$	3000	16.4± 0.5
$(E_2K_2)_8$	0	15.5± 0.2
$(E_2K_2)_8$	100	18.1±0.2
$(E_2K_2)_8$	500	15.4±0.2
$(E_2K_2)_8$	1000	17.8±0.3
$(E_2K_2)_8$	3000	18.2±0.4
$(EK)_{16}$	0	14.8± 0.2
$(EK)_{16}$	100	17.3±0.2
$(EK)_{16}$	500	17.5± 0.4
$(EK)_{16}$	1000	18.5±0.4
$(EK)_{16}$	3000	16.8±0.5

Kratky analysis

Kratky analysis was used to qualitatively evaluate the conformational shape of each polyampholyte. R_g and I_0 from Guinier analysis was used to calculate dimensionless forms of $q^2I(q)$ and q. Peak positions were found from Kratky plots using plot trace. The resulting Kratky plots are shown in Figure S5. All calculations were performed using in-house Jupyter Notebook scripts.

Form factor and scaling factor fitting

Form factors were assigned and fit to $I(q)$ based on initial shape evaluation from Kratky analysis. For polyampholytes behaving as an ideal polymer chain, we applied the Debye form factor for Gaussian chains.

$$P(q) = \frac{2}{q^4 R_g^4}[\exp\left(-q^2 R_g^2\right) - 1 + q^2 R_g^2] \tag{S.7}$$

For $N_p = 1, 2$, and 4 at 0 mM added salt, we observed a downturn at $q > 0.3$ Å^{-1} (Figure S5). Since direct and appropriate analytical form factors were not found, we applied the Gaussian approximation as reference.

For polyampholytes that exhibit an upturn at high q, we applied the excluded volume form factor for worm-like chains (WLCs) defined by Pedersen and Schurtenberger with corrections by Chen, Butler, and Magid.[11,12] For this model, the contour length L was fixed to 143 Å, calculated from C-C and C-N bond lengths.

Lastly, scaling of $I(q)$ as $q > 0.2$ Å^{-1} was found by performing a linear fit on $\ln I(q)$ vs. $\ln q$ and finding the slope (Figure S7). Multiple intervals of q was fitted, and the resulting average was reported with standard σ error (Figure 3A and 5A). The resulting $-1/\nu$ factor for polyampholytes with increasing salt concentration is plotted in Figure S6.

All fits were performed using non-linear least squares algorithm from SciPy with consideration of data uncertainties in an in-house Jupyter Notebook python script. The results for the Debye, excluded volume WLC, and thin rod fits are plotted in Figure S5.
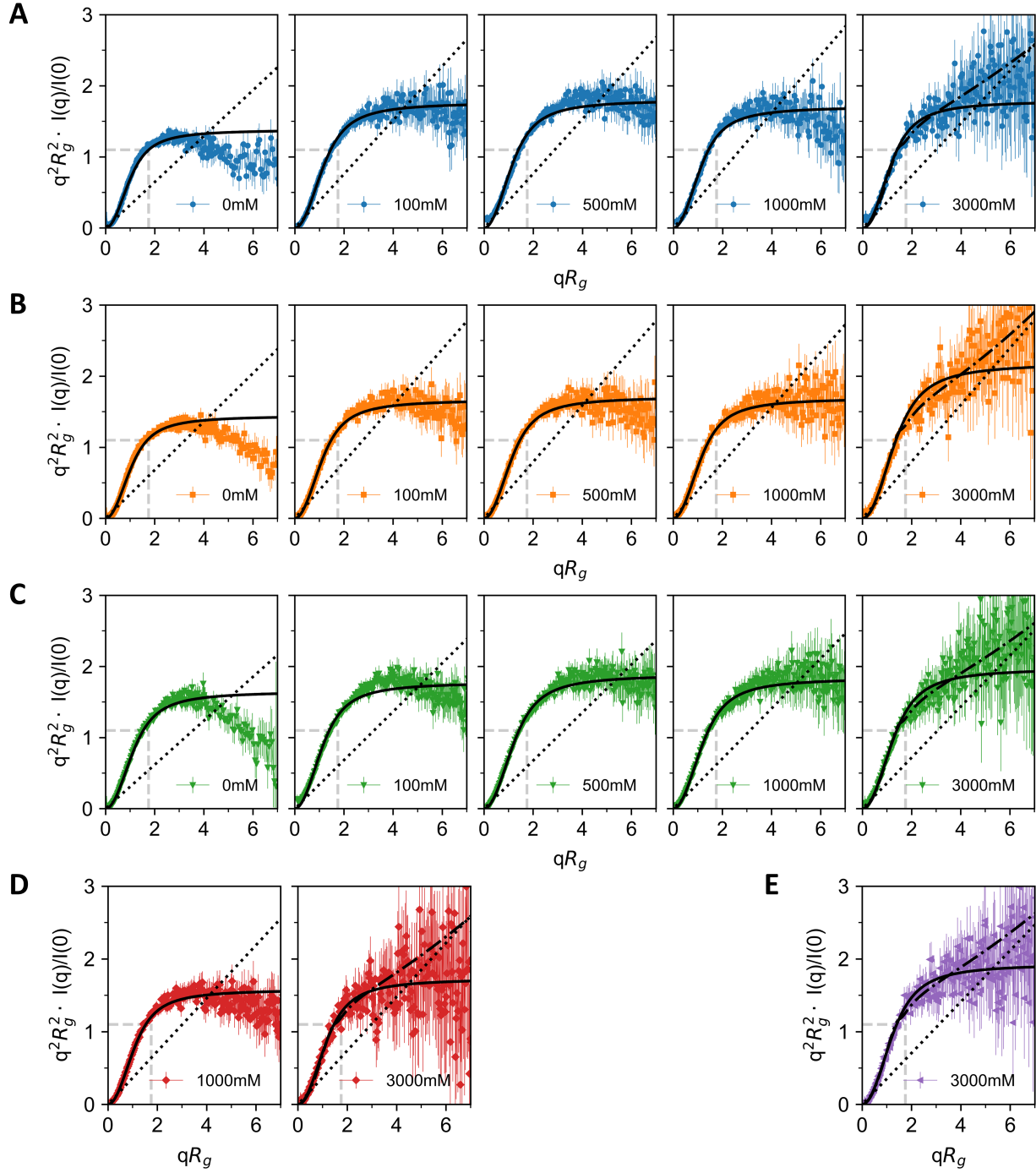
Figure S5: Dimensionless Kratky plots of polyampholytes (A)$(EK)_{16}$, (B)$(E_2K_2)_8$, (C)$(E_4K_4)_4$, (D)$(E_8K_8)_2$, and (E)$(E_{16}K_{16})$ at increasing added NaCl concentrations. Gray dashed lines mark where $q^2 R_g^2 \cdot I(q)/I(0) = 1.1$ and $qR_g = \sqrt{3}$ for globule conformation. Black solid lines, black dot dashed lines, and black dotted lines denote Debye form factor for a Gaussian chain, Pedersen Schurtenberger form factor for semiflexible chain with excluded volume, and form factor for infinitely thin rod, respectively.
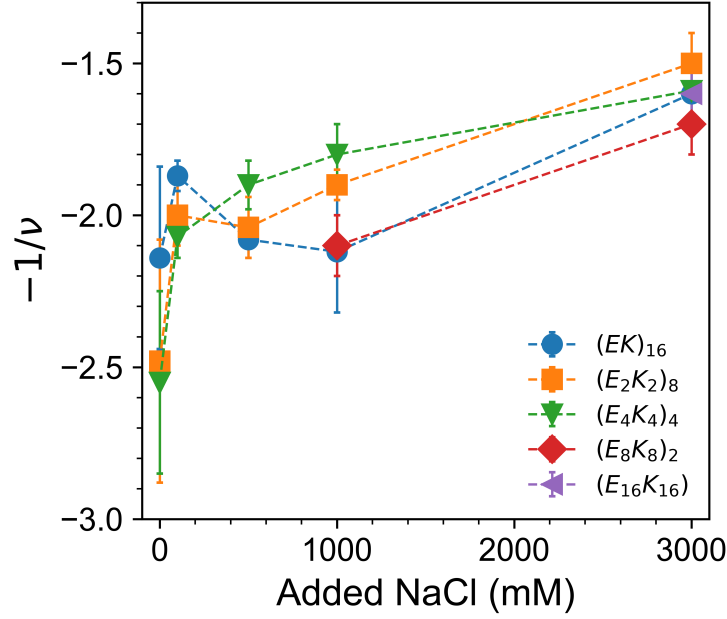
Figure S6: Scaling factor $-1/\nu$ at high q of polyampholyte sequences at increasing NaCl concentrations. We observe a general increase in the slope of scattering intensity of each charge sequence a high q, indicative of salt induced transition from poor solvent to good solvent environment.

Figure S7: Example SAXS analysis workflow using $(E_4K_4)_4$ at pH 7 and 1000 mM added NaCl from (A) background solvent subtraction to (B) Guinier analysis to (C) fitting the SAXS profiles for scaling factor (upper) and Debye form factor (lower).

pH titrations

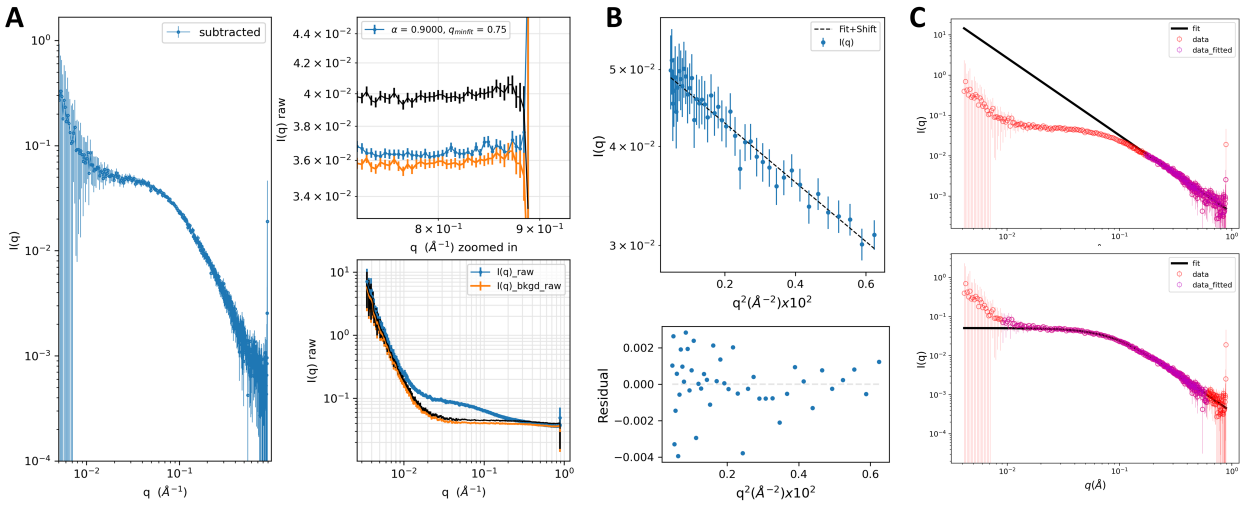pH titrations were conducted using a Thermo Scientific ™ pH meter. Titration samples were prepared at 0.1 wt% polyampholyte in Milli-Q water in 3 mL volumes. Polyampholyte solutions were either titrated to acidic conditions using 0.1 M HCl or titrated to basic conditions using 0.1 M NaOH from pH 7 to minimize counterion concentration which interacts with the charged groups and interferes with solubility and ionization determinations. Titrant was added initially in 1 μL volumes and increased up to 100 μL at extreme pH. Between each addition of HCl or NaOH, the solution was allowed to sit with interspersed vortexing for at least 2 minutes until pH readings between each mixing and sitting cycle are within 0.1 of each other. Molar equivalence was calculated using

$$mol\,eq. = \frac{mol\,HCl}{32 \cdot mol\,PA} \ or \ \frac{mol\,NaOH}{32 \cdot mol\,PA} \tag{S.8}$$

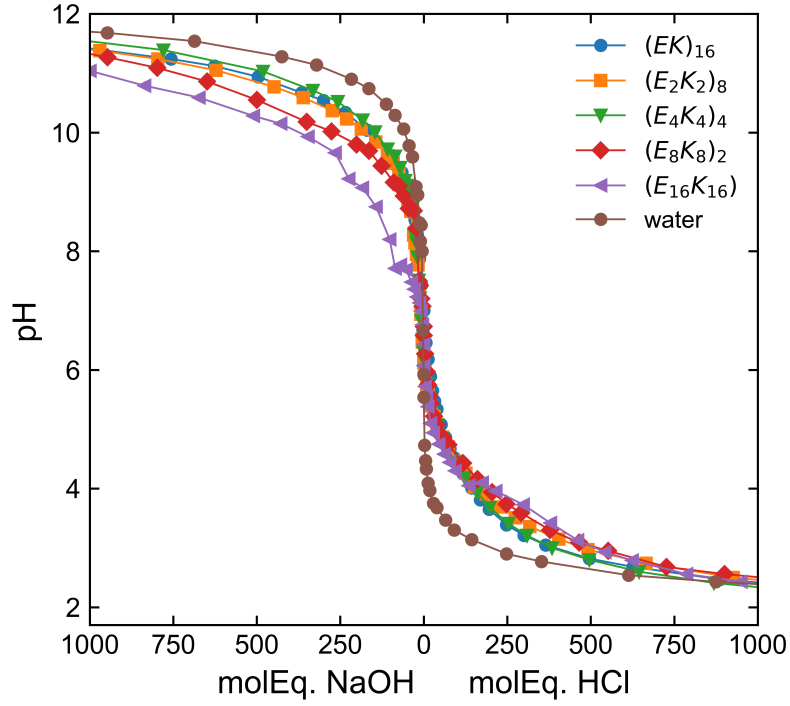depending on whether HCl or NaOH was used for the titration.

Figure S8: pH titration curves of 0.1 wt% polyampholyte solutions in water. The brown titration curve is for only Milli-Q water. $(E_8K_8)_2$ and $E_{16}K_{16}$ phase separate in $\approx$ pH 3.7-10.6 and $\approx$ pH 3.9-10.8, respectively; we continue to measure a titration curve by stirring the phase separated solution.

Simulations

GB/SA

Molecular dynamics simulations with Generalized Born/Surface Accessible (GB/SA)[13] were run on the NAMD[14] platform. In NAMD, the implementation of GB follows the Onufriev-Bashford-Case (OBC)[15] formulation. In the GB/SA approach, the contribution due to attractive solute-solvent electrostatic interactions is modeled by treating the solvent as a continuum dielectric and by using the so-called Coulomb field approximation. The unfavorable free energy of creating a cavity to accommodate the solute and the effects of non-polar interactions are lumped together in a surface-area dependent term. The intra-molecular (vacuum) potential energy comprises the usual bond, angle, dihedral, and non-bonded van der Waals and electrostatic interactions derived from version 36[16] of the CHARMM forcefield[17] with CMAP-corrections for dihedrals.[18]

Linear structures of polyampholytes were N-terminus acetylated and C-terminus amidated using psfgen. Each system was equilibrated for 30 ns with a 2 fs time step. Structures for the R_g calculation were saved every 250 time steps in a 30 ns production run. This procedure was repeated three times to obtain structures from independent trials. The ensemble averaged R_g for each polyampholyte was calculated in a python program using the MDTraj library.[19] GB/SA calculations were performed at 0, 100, 500, and 1000 mM salt. Note that in GB/SA simulations the effect of ions is through the Debye length parameter, i.e. the ions are point particles in a continuous dielectric medium. Table S4 shows the R_g from each independent simulation.

Hydrogen bonding in GB/SA simulations

Analogous to the FTIR study, simulated structures of polyampholytes were tested for the presence of backbone hydrogen bonds. For backbone hydrogen bonding, only the carboxyl oxygen and amide nitrogen of each residue need to be analysed. If the carboxyl oxygen and

Table S4: Mean and standard deviation of the radius of gyration R_g from each of the three GB/SA simulations for 0, 100, 500, and 1000 mM salt.

Structure	Salt (mM)	Sim 1, R_g (Å)	Sim 2, R_g (Å)	Sim 3, R_g (Å)	Average, R_g (Å)
$(EK)_{16}$	0	19.7 ± 1.6	18.5 ± 3.2	15.0 ± 1.5	17.7 ± 3.0
$(E_2K_2)_8$	0	17.9 ± 1.8	16.2 ± 1.6	19.7 ± 2.0	17.9 ± 2.3
$(E_4K_4)_4$	0	17.5 ± 1.4	16.5 ± 1.2	17.9 ± 1.3	17.3 ± 1.4
$(E_8K_8)_2$	0	13.0 ± 1.7	16.6 ± 1.4	14.2 ± 0.9	14.6 ± 2.0
$(E_{16}K_{16})$	0	12.7 ± 1.2	13.1 ± 0.5	11.6 ± 0.4	12.5 ± 1.0
$(EK)_{16}$	100	19.4 ± 1.8	17.8 ± 1.5	16.1 ± 1.6	17.8 ± 2.1
$(E_2K_2)_8$	100	16.9 ± 1.4	17.1 ± 1.6	18.4 ± 2.3	17.5 ± 1.9
$(E_4K_4)_4$	100	16.8 ± 2.1	16.2 ± 1.3	19.9 ± 2.4	17.6 ± 2.5
$(E_8K_8)_2$	100	18.2 ± 1.4	17.0 ± 1.9	17.4 ± 1.5	17.5 ± 1.7
$(E_{16}K_{16})$	100	17.2 ± 1.7	13.8 ± 3.7	18.7 ± 1.7	16.6 ± 3.3
$(EK)_{16}$	500	15.3 ± 1.6	19.0 ± 1.3	20.1 ± 1.6	18.1 ± 2.6
$(E_2K_2)_8$	500	18.5 ± 2.3	17.3 ± 1.9	17.3 ± 1.6	17.7 ± 2.0
$(E_4K_4)_4$	500	14.1 ± 1.3	19.1 ± 1.6	16.7 ± 1.7	16.6 ± 2.6
$(E_8K_8)_2$	500	18.6 ± 2.2	18.0 ± 1.7	18.4 ± 2.3	18.4 ± 2.1
$(E_{16}K_{16})$	500	21.6 ± 2.6	20.2 ± 1.4	16.8 ± 1.4	19.5 ± 2.8
$(EK)_{16}$	1000	20.0 ± 2.4	20.5 ± 1.3	14.7 ± 1.7	18.4 ± 3.2
$(E_2K_2)_8$	1000	19.7 ± 2.1	19.1 ± 2.3	17.6 ± 1.3	18.8 ± 2.1
$(E_4K_4)_4$	1000	20.3 ± 1.9	17.4 ± 2.0	20.3 ± 1.8	19.3 ± 2.4
$(E_8K_8)_2$	1000	19.4 ± 1.4	18.1 ± 2.4	18.6 ± 2.2	18.7 ± 2.1
$(E_{16}K_{16})$	1000	18.4 ± 2.0	19.7 ± 1.3	18.6 ± 1.5	18.9 ± 1.7

amide nitrogen are less than 3 Å apart, and the angle between the vectors $\overrightarrow{HN}$ and $\overrightarrow{HO}$ ($\angle NHO$) is greater than 150°, a hydrogen bond is counted. Similar to the R_g calculation, hydrogen bonding was aggregated from 3 independent simulations. For each structure the system was equilibrated for 30 ns with a 2 fs time step. The number of hydrogen bonds in each structure was collected every 500 time steps in a 30 ns production run. The semi-log plot in Figure S9 shows evidence of transient hydrogen bonding for all N_p. However, the trends captured in FTIR for different blockiness are not mirrored here.
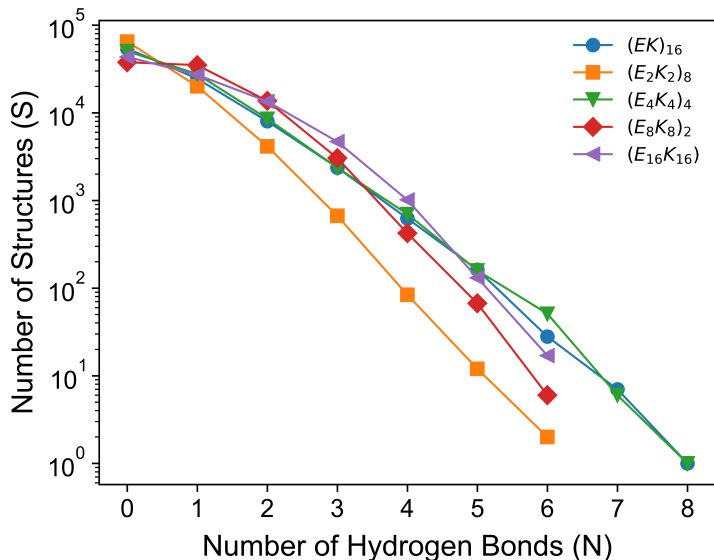

Figure S9: A semi-log plot of number of structures (S) with N hydrogen bonds vs. number of hydrogen bonds (N) for each polyampholyte (varying N_p) from GB/SA.

ABSINTH

Monte Carlo simulations with ABSINTH[20] were run on the CAMPARI simulation engine.[21] In the ABSINTH[20] model, one decomposes the polyampholyte into smaller groups with known hydration free energies. The model additionally includes a model for solvent exclusion drawing upon the earlier work of Lazaridis and Karplus[22] but with a different functional form. Additionally, ABSINTH also incorporates the role of intra-molecular charge-charge interactions to the electrostatic energy of a solute. On this basis, the hydration free energy

of the model is constructed. This together with the usual bonding, angle, dihedral, and non-bonded contributions defines the ABSINTH Hamiltonian. As in GB/SA, the effect of added ions is treated at a continuum level.

Linear polyampholytes were equilibrated for 50 million sweeps in a spherical droplet, 1000 Å in radius. Trajectories for R_g calculation were saved every 1000 sweeps in a 100 million sweep production run. Analogous to GB/SA simulations, calculations for each system were repeated three times. Ensemble averages R_g (from 3 independent trials) were calculated in a python program utilizing the MDTraj library.[19] Table S5 shows the results of independent ABSINTH trials for polyampholytes with varied blockiness N_p. It can be seen from tables S4 and S5, the variation between independent trials of the GB/SA MD simulations are much more noticeable than in the ABSINTH MC simulations, which needs further investigation.

Table S5: Mean and standard deviation of the radius of gyration (R_g) from each of the three ABSINTH simulations.

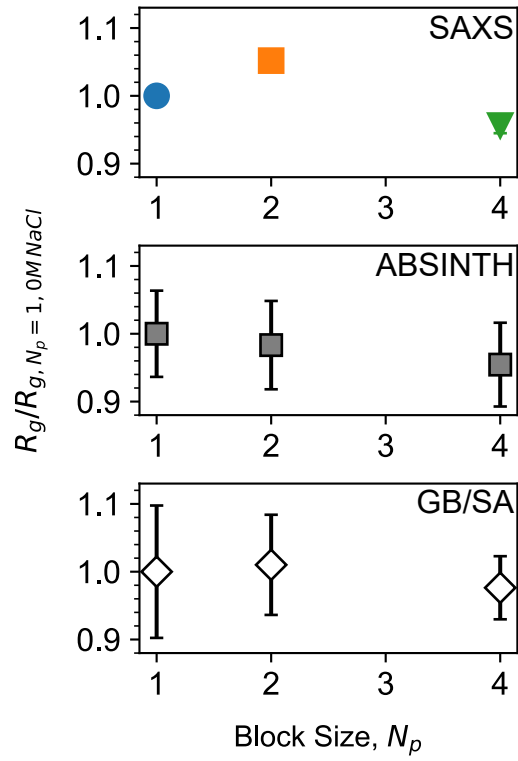| Structure	Sim 1, R_g (Å)	Sim 2, R_g (Å)	Sim 3, R_g (Å)	Average, R_g (Å)
$(EK)_{16}$	22.2 ± 2.4	22.3 ± 2.4	22.2 ± 2.5	22.2 ± 2.5
$(E_2K_2)_8$	21.9 ± 2.5	21.9 ± 2.5	21.9 ± 2.5	21.9 ± 2.5
$(E_4K_4)_4$	21.2 ± 2.4	21.2 ± 2.4	21.3 ± 2.4	21.2 ± 2.4

Figure S10: Relative R_g of polyampholytes at 2 wt% at pH 7 in 0 mM added NaCl solution shown in Figure 3. R_g's of each method are normalized to corresponding R_g of $(EK)_{16}$ at pH 7 in 0 mM added NaCl.
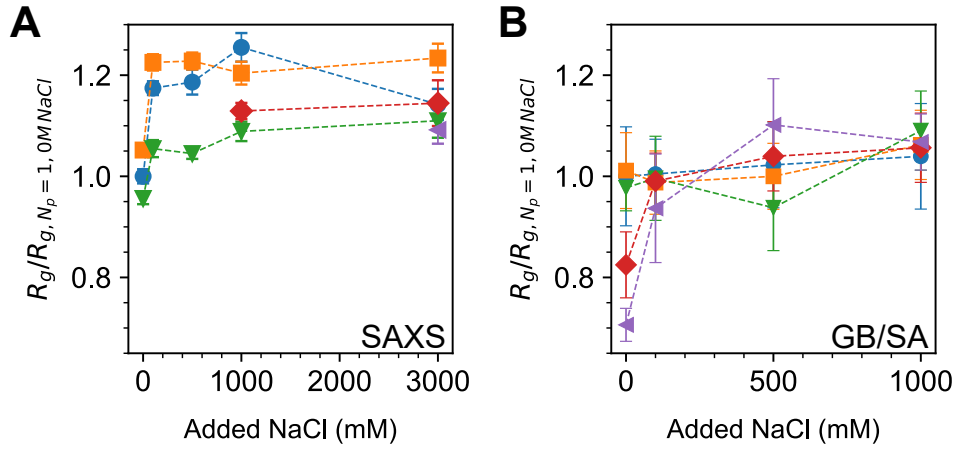

Figure S11: Relative R_g of polyampholytes at 2 wt% at pH 7 in 0, 100, 500, 1000, and 3000 mM added NaCl solution from (A) SAXS experiments and (B) GB/SA simulations. The original R_g's are shown of in Figure 5B and 5C. R_g's of SAXS experiments are normalized to SAXS determined R_g of $(EK)_{16}$ at pH 7 in 0 mM added NaCl, and R_g's of GB/SA simulations are normalized to R_g of $(EK)_{16}$ at pH 7 in 0 mM NaCl from the GB/SA simulations.
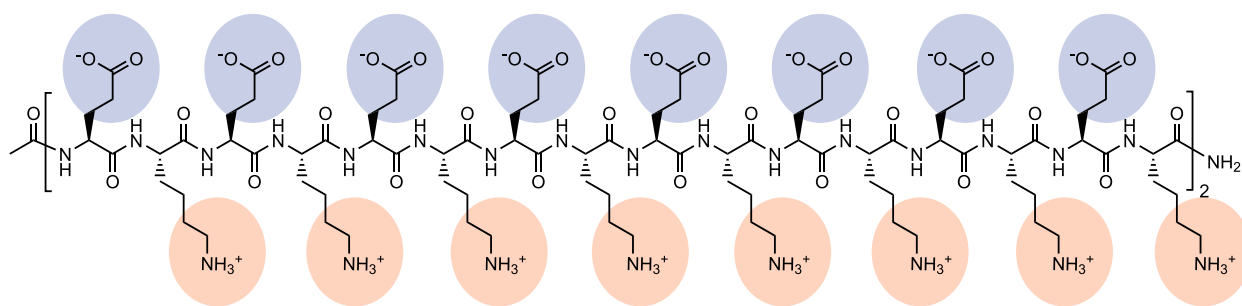
Figure S12: Stereochemistry schematic of $(EK)_{16}$ where anionic residues are highlighted in blue and cationic residues are highlighted in orange.

References

(1) Fields, G. B.; Noble, R. L. Solid Phase Peptide Synthesis Utilizing 9-Fluorenylmethoxycarbonyl Amino Acids. *Int. J. Pept. Protein Res.* **1990**, *35*, 161–214.

(2) Hood, C. A.; Fuentes, G.; Patel, H.; Page, K.; Menakuru, M.; Park, J. H. Fast Conventional Fmoc Solid-Phase Peptide Synthesis With HCTU. *J. Pept. Sci.* **2008**, *14*, 97–101.

(3) Marciel, A. B.; Srivastava, S.; Tirrell, M. V. Structure and Rheology of Polyelectrolyte Complex Coacervates. *Soft Matter* **2018**, *14*, 2454–2464.

(4) Sen, P.; Fatima, S.; Khan, J. M.; Khan, R. H. How methyl cyanide induces aggregation in all-alpha proteins: a case study in four albumins. *Int. J. Biol. Macromol.* **2009**, *44*, 163–160.

(5) Kestin, J.; Sokolov, M.; Wakeham, W. A. Viscosity of Liquid Water in the Range 8°C to 150°C. *J. Phys. Chem. Ref. Data* **1978**, *7*, 941–948.

(6) Wilke, J.; Kryk, H.; Hartmann, J.; Wagner, D.; Eich, A. Visco Handbook: Theory and Application of Viscometry with Glass Capillary Viscome-

ters. 2015; `https://www.xylemanalytics.com/File%20Library/Downloads/SIA_Visco-handbook_English.pdf`.

(7) Hiemenz, P. C.; Lodge, T. P. *Polymer Chemistry*, 2nd ed.; CRC Press: Boca Raton, 2007; pp 334–354.

(8) Davis, R. M.; Russel, W. B. Intrinsic Viscosity and Huggins Coefficients for Potassium Poly(styrenesulfonate) Solutions. *Macromolecules* **1987**, *20*, 518–525.

(9) Dinic, J.; Martínez Narváez, C. D. V.; Sharma, V. *Macromolecular Engineering*; John Wiley & Sons, Ltd, 2022; pp 1–36.

(10) Hopkins, J. B.; Gillilan, R. E.; Skou, S. BioXTAS RAW: Improvements to a Free Open-Source Program for Small-Angle X-Ray Scattering Data Reduction and Analysis. *J. Appl. Crystallogr.* **2017**, *50*, 1545–1553.

(11) Pedersen, J. S.; Schurtenberger, P. Scattering Functions of Semiflexible Polymers With and Without Excluded Volume Effects. *Macromolecules* **1996**, *29*, 7602–7612.

(12) Chen, W.-R.; Butler, P. D.; Magid, L. J. Incorporating Intermicellar Interactions in the Fitting of SANS Data From Cationic Wormlike Micelles. *Langmuir* **2006**, *22*, 6539–6548.

(13) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.

(14) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics With Namd. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(15) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

(16) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, Jr., A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ 1and χ Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

(17) MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack, Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; *et al.*, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(18) MacKerell Jr., A. D.; Feig, M.; Brooks III, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comp. Chem.* **2004**, *25*, 1400–1415.

(19) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDtraj: A Modern Open Library For the Analysis of Molecular Dynamics Trajectories. *BiophysJ* **2015**, *109*, 1528 – 1532.

(20) Vitalis, A.; Pappu, R. V. ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions. *J. Comput. Chem.* **2009**, *30*, 673–699.

(21) Vitalis, A.; Pappu, R. V. Methods for Monte Carlo Simulations of Biomacromolecules. *Annu. Rep. Comput. Chem.* **2009**, *5*, 49–76.

(22) Lazaridis, T.; Karplus, M. Effective Energy Function For Proteins in Solution. *Proteins: Struct. Funct. Genet.* **1999**, *35*, 133–152.