Electronic Supplementary Material (ESI) for Soft Matter.

Supporting Information

SANS Partial Structure Factor Analysis for Determining Protein-Polymer Interactions in Semidilute Solution

*Aaron Huang,† Helen Yao † and Bradley D. Olsen**

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

† A.H. and H.Y. contributed equally to this work.

*Corresponding Author
Bradley D. Olsen
TEL: (617) 715-4548
Email: bdolsen@mit.edu

mCherry Purification

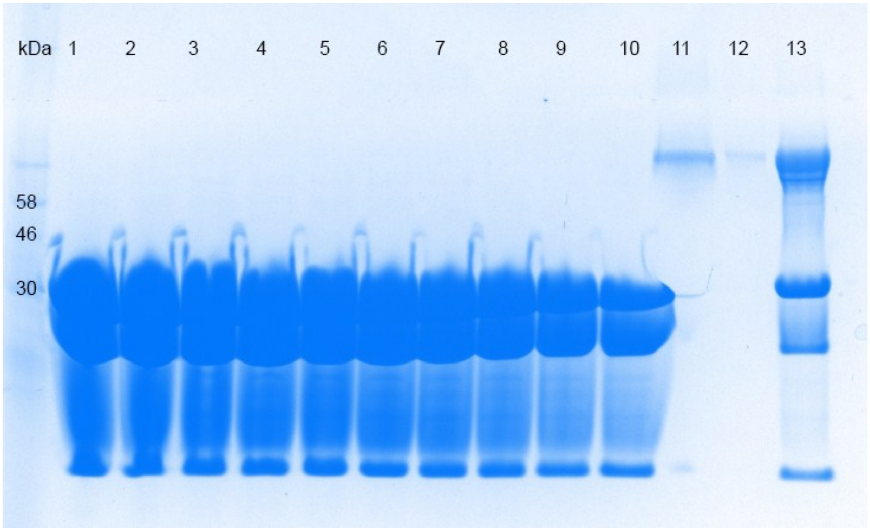

Figure S1. SDS-PAGE of FPLC purified mCherry. Lanes contain the following: 1-12) elutions at 24 mL to 73.5 mL in increments of 4.5 mL. Lane 13 is the column wash with 2M NaCl after elution. Column volume was 10 mL.

Polymer Gel Permeation Chromatography (GPC) Traces

For PNIPAM and POEGA, GPC was run on a Waters system using two ResiPore columns in *N,N*-dimethylformamide (DMF) with 0.02 M lithium bromide (LiBr) as the mobile phase. The dn/dc values of PNIPAM and POEGA in DMF are 0.0761 and 0.042 mL g^{-1}, respectively. For PDMAPS, GPC was run on an Agilent Technologies 1260 Infinity system using two Aquagel columns in 0.5 M NaCl (aq) with 0.02% sodium azide as the mobile phase. The dn/dc value for PDMAPS in 0.5 M NaCl is 0.1423 mL g^{-1}. All GPC signals were collected using Wyatt light scattering and refractive index detectors.

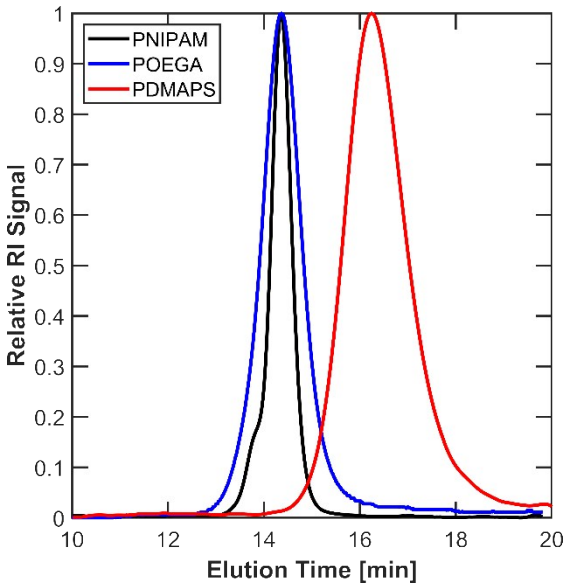

Figure S2. Gel permeation chromatography (GPC) traces for PNIPAM (M_n = 26.3 kDa, Đ = 1.08), POEGA (M_n = 26.4 kDa, Đ = 1.13), and PDMAPS (M_n = 26.7 kDa, Đ = 1.10).

Estimation of Hydration Number (n_H)

SANS samples of PNIPAM and PDMAPS were prepared at a concentration of 13.75 mg/mL in deuterium oxide. POEGA was prepared at a concentration of 8 mg/mL in deuterium oxide. The SANS data for 100% D_2O solutions were fit with a form factor and incoherent background term (Equation S1).

$$I(Q) = Nv_{monomer}\phi(\rho_{solvent} - \rho_{polymer})^2 P(Q) + B \tag{S1}$$

The form factor, *P(Q)*, is the Debye function for Gaussian chains (Equation S2).

$$P(Q) = \frac{2\left(e^{-x} + x - 1\right)}{x^2},\ x = Q^2 R_g{}^2 \tag{S2}$$

Here, there are three fit parameters: the prefactor $Nv_{monomer}\phi(\rho_{solvent} - \rho_{polymer})^2$, the radius of gyration (R_g), and the incoherent background (B). From the prefactor, the scattering length density (SLD) of the polymer ($\rho_{polymer}$) was calculated. The volume was assumed to be the molecular volume of a monomer, calculated as

$$v_{monomer} = \frac{M_{monomer}}{N_A \delta_{bulk}}, \tag{S3}$$

where $M_{monomer}$ is the molecular weight of the monomer, N_A is Avogadro's number, and δ_{bulk} is the bulk density of the polymer. The rest of the prefactor is comprised of the degree of polymerization (N), the volume fraction of polymer in solution (ϕ), and the SLD of the solvent ($\rho_{solvent}$). From the SLD of the polymer, the hydration number, or the number of water molecules associated with each monomer of the polymer chain, can be calculated from the definition of SLD as the ratio of scattering length (b) to the volume (v):[1]

$$\rho_{polymer,100\%D_2O} = \frac{b_{polymer} + n_H b_{D_2O}}{v_{monomer} + n_H v_{D_2O}} \tag{S4}$$

The scattering lengths of each polymer and D_2O were computed from tabulated values, and the volume of water was assumed to be 30.4 Å^3. All three polymers approximately behave as Gaussian chains, with slight deviations in PNIPAM in the mid-Q region and in PDMAPS in the high-q region, as shown in Figure S3. While the behavior of POEGA chains (Figure S3b) in dilute solution is well-described by the Debye form, PNIPAM (Figure S3a) and PDMAPS (Figure S3c) do not behave exactly like ideal chains in a theta solvent. For PNIPAM, water is a very good solvent, so the polymer chains are more likely to be in a swollen state. PDMAPS is zwitterionic and has additional electrostatic interactions with water molecules.

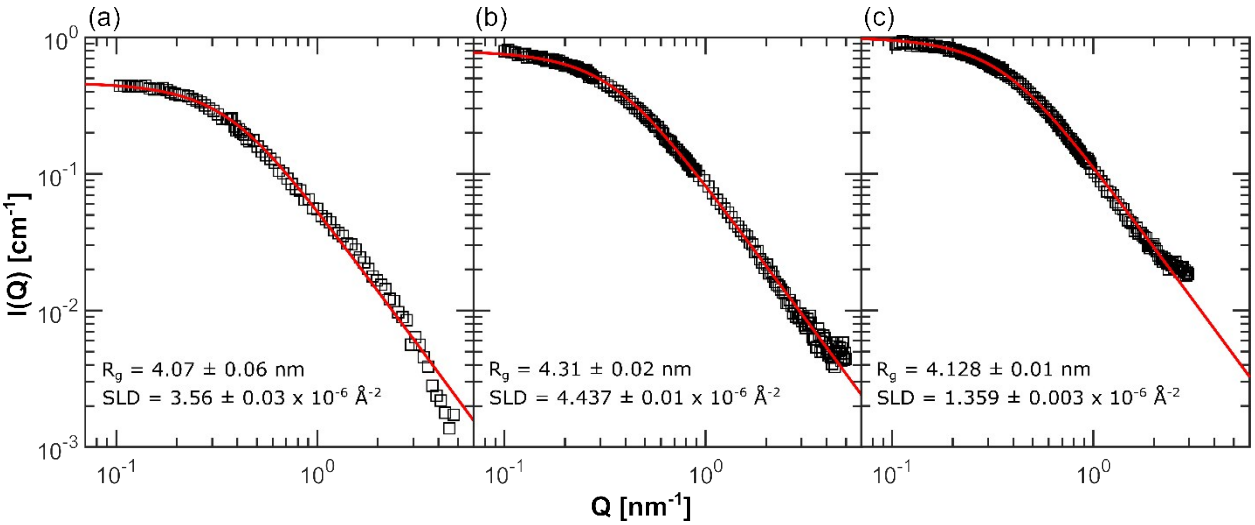

Figure S3. Debye form factor fits for (a) PNIPAM, (b) POEGA, and (c) PDMAPS. Open squares are experimental SANS absolute intensities with incoherent background subtracted. The red line is the Debye form factor fit calculated using the inset radii of gyration and polymer SLD. Uncertainties are 1σ.

The SLD of polymer in 100% D_2O allows for the calculation of the hydration number (Table S1). The hydration number varies inversely with the C_{ODT} (the lowest concentration at which order is observed), which is lowest for PNIPAM and highest for PDMAPS (Figure 1). This supports the idea that hydration water improves the propensity for ordering. Moreover, PHPA and PDMAPS seem to have close to 0 hydration water molecules. To allow for the possibility of variation in hydration number in semidilute solution, sensitivity analysis was performed for the contribution of the hydration number to the polymer SLD (Figure S4).

Table S1. Hydration number for dilute solution polymers

Polymer	**n_H**
PNIPAM	5.98 ± 0.08*
POEGA	0.765 ± 0.005*
PDMAPS	0.236 ± 0.007*

*1σ

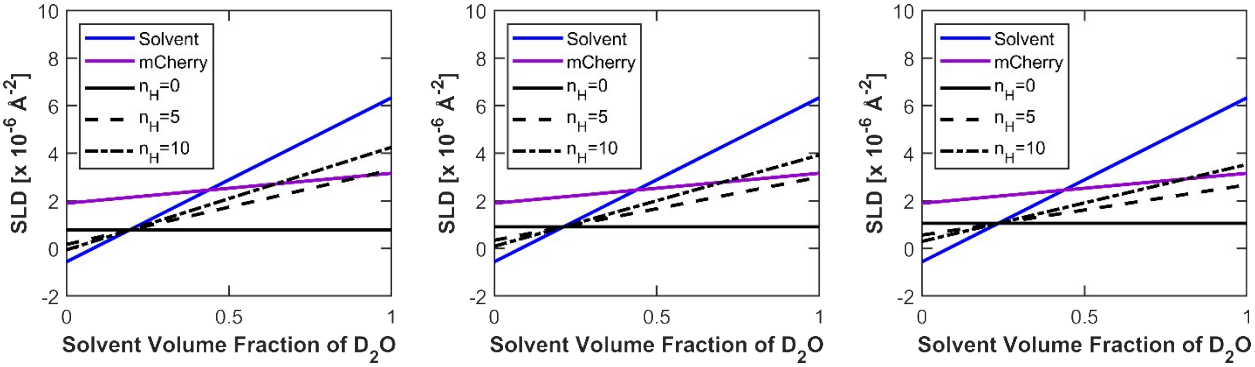

Figure S4. Variation in scattering length density (SLD) with hydration number (n_H) and solvent volume fraction of D_2O for (a) PNIPAM, (b) POEGA, and (c) PDMAPS.

Calculation of Overlap Concentration, Contour Length, and Correlation Length

The overlap concentration, defined as the concentration at which polymer chains begin to interact with each other in solution, demarcates the boundary between dilute and semi-dilute solutions. The overlap volume fraction was estimated for PNIPAM, POEGA, and PDMAPS using the following equation:[2]

$$\phi^* = \frac{\text{molecular volume of polymer}}{\text{volume of polymer coil}} = \frac{\frac{M}{N_A \rho}}{v_{coil}} \tag{S5}$$

Here, M is the molar mass of the polymer, N_A is Avogadro's number, ρ is the polymer density, and v_{coil} is the volume of the polymer coil, which can be estimated as the cube of the end-to-end distance $\left\langle r^2 \right\rangle^{1/2}$. Given that the end-to-end distance is difficult to obtain from SANS, we can estimate it using the radius of gyration, namely that $\left\langle r^2 \right\rangle^{1/2} = \frac{1}{6} R_g^2$. Table S2 provides the calculated overlap concentrations of each polymer. The contrast-variation SANS experiments were performed at 5 vol%, which is slightly above the overlap concentration for each polymer, confirming that they were done in the semidilute regime.

The correlation lengths, ξ, of the polymers were calculated using the following relation, which is valid for polymers with a coil volume of v_{coil} and a Kuhn length of b in semidilute solution with a solvent parameter of υ:[2]

$$\xi = b \left(\frac{b^3}{v_{coil}} \right)^{2\upsilon - 1 / 3\upsilon - 1} \phi^{-\upsilon / 3\upsilon - 1} \tag{S6}$$

The correlation length was estimated using the dilute solution radius of gyration assuming a theta solvent ($\upsilon = 0.5$). The Kuhn length was obtained from the radius of gyration and degree of polymerization (N):

$$b = R_g \sqrt{6 / N} \tag{S7}$$

Table S2. Overlap Concentration and Characteristic Lengths

Polymer	ρ [g/cm^3]	R_g [nm]	ϕ^*	ξ [nm]	L [nm]
PNIPAM	1.05	4.56	0.03	14.7	20.7
POEGA	1.05	4.24	0.04	15.7	15.6
PDMAPS	1.37	4.12	0.03	21.0	8.18

The correlation lengths were used to predict a range for the depletion thickness of each polymer, if there are depletion interactions between proteins. Colloid-polymer depletion literature is divided over the length scale of the depletion thickness in the protein limit ($R_{protein} << \xi$), with predictions ranging from the length scale of the protein itself to the correlation length of the polymer above the overlap concentration.[3]

The contour length, L, was calculated by approximating the polymer backbone as a chain of tetrahedrally oriented C–C atoms (bond length, l = 0.154 nm; bond angle, α = 109.5°C)

$$L = Nl \cos(\alpha / 2) \tag{S8}$$

SANS Intensity Background Subtraction and Correction for Incoherent Scattering

SANS intensity curves were corrected for background scattering and incoherent scattering by taking an average of the absolute intensities measured at high Q (Q > 4.054 nm^{-1}), where the scattering intensities have decayed into a flat line. Figure S5 depicts the range over which the flat background was obtained and the flat background that was subtracted from all intensity curves.

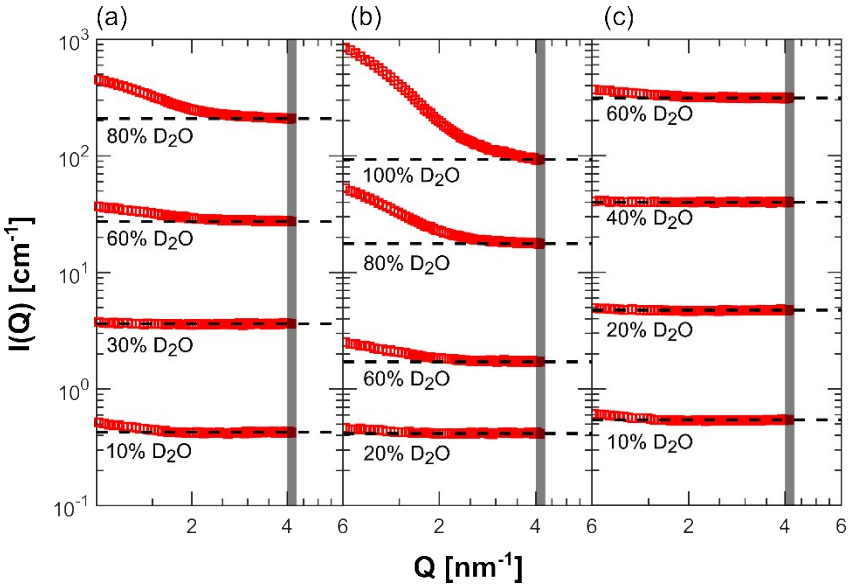

Figure S5. Absolute SANS intensity curves for (a) PNIPAM, (b) POEGA, and (c) PDMAPS showing the estimated background and incoherent scattering component that was subtracted prior to partial structure factor decomposition. Data for Q < 1 nm^{-1} have been truncated for clarity. The gray box illustrates the range over which the intensity was averaged.

Fourier Transforms of Structure Factors

To perform the inverse Fourier transformation (FT) described in Equation 9, the background-subtracted scattering intensities, *I(Q)*, were extrapolated to $Q = 0$ nm^{-1} and $Q = 50$ nm^{-1}. For the high Q region, the scattering intensities were set to 0 for a range of $Q > Q_{max}$ from experiment to 50 nm^{-1}. This was chosen through a sensitivity analysis. For any $Q > 50$ nm^{-1}, there was no further effect on the inverse FT. For the low Q region, the extrapolation was done using a Guinier-like analysis. Since the concentrations of all protein/polymer blends were designed to be above the overlap concentration (reported in Table S2), this analysis was not intended to derive quantitative results such as radius of gyration or scattering length density but rather to provide a mathematical function for extrapolation. The general form for the small Q region is

$$I(Q) = A\exp(-CQ^2) \tag{S9}$$

where A and C are fitting parameters. The region of scattering intensity that would fulfill equation S9 satisfies the inequality $Q < 1/R_g$. Using the radii of gyration obtained from Debye fits (shown in Figure S3), the data for which $Q < 1/R_g$ were fit to equation S9 in linearized form and concatenated to the actual data and the high Q region (see Figure S6 for small Q fits and S7 for concatenated data).

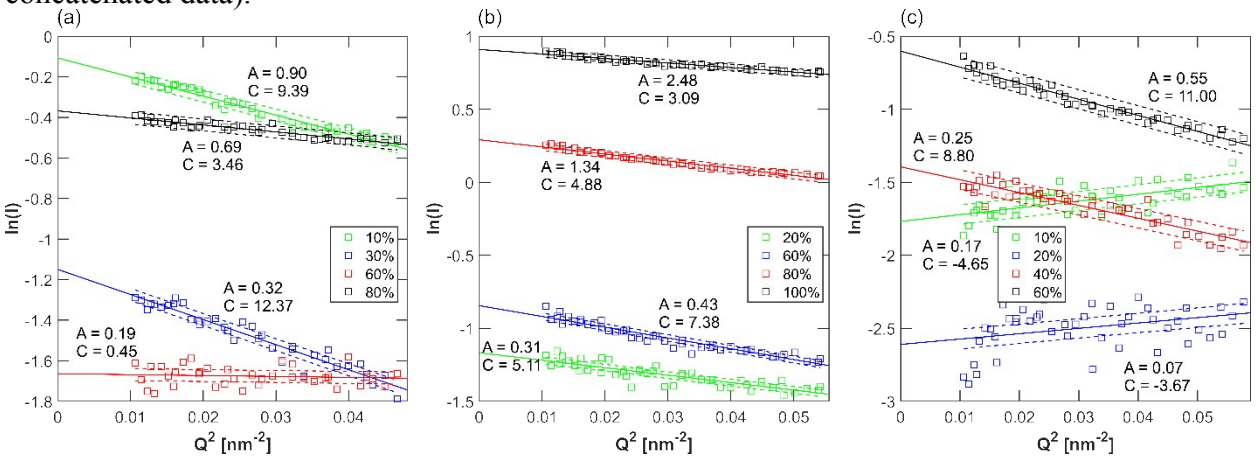

Figure S6. Linearized Guinier-like fits for scattering intensities of blends of mCherry and (a) PNIPAM, (b) POEGA, and (c) PDMAPS. Fitting parameters for equation S9 are displayed next to the corresponding line of fit. Dashed lines indicate 95% confidence intervals for each fit.

The extrapolated scattering intensity was then decomposed into partial structure factors via Equation 2, smoothed, and inverse Fourier transformed into Γ^*_{ij}. Smoothing parameters are shown in Table S3. Optimal smoothing was determined by Fourier transforming Γ^*_{ij} back into partial structure factors and ensuring that these matched with the experimentally obtained partial structure factors. The result of the extrapolated and decomposed data after the inverse FT is compared to the result of the raw, decomposed data after the inverse FT in Figure S8. As shown in Figure S8, the most significant effect of extrapolation (S8d – f) is to change the amplitude of some of the correlation functions (all PNIPAM correlation functions, POEGA/POEGA, and POEGA/mCherry). Some of the peaks in the cross structure factor of PNIPAM/mCherry are also suppressed after extrapolation.

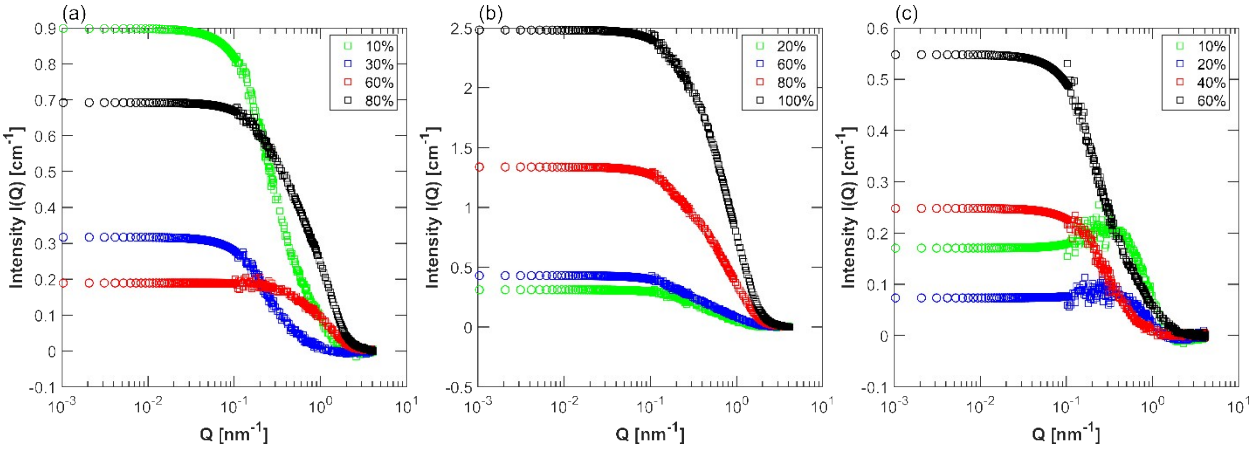

Figure S7. Results of Guinier-like fit (Equation S9) used to extrapolate to $Q = 0$ nm^{-1} for absolute scattering intensities after background subtraction of blends of mCherry and (a) PNIPAM, (b) POEGA, and (c) PDMAPS. For clarity, extrapolation to 50 nm^{-1} is not shown.

Table S3. Smoothing window sizes

Polymer	S_{11}	S_{22}	S_{12}
PNIPAM	40	40	40
POEGA	6	35	20
PDMAPS	5	5	5

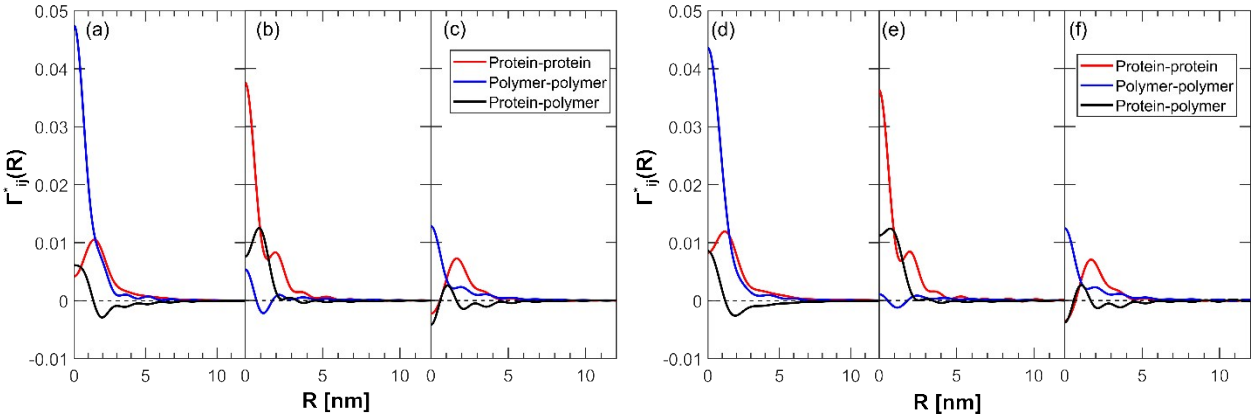

Figure S8. Dimensionless concentration correlation function Γ_{ij}^* between components *i* and *j* for (a, d) PNIPAM/mCherry, (b, e) POEGA/mCherry blend, and (c, f) PDMAPS/mCherry. The set of figures shown in (a – c) were derived from the experimental data with only high *Q* extrapolation. The set of figures shown in (d – f) were derived from experimental data extrapolated on both the low and high *Q* ends.

Non-dimensionalization of the Structure Factor

The partial structure factors were non-dimensionalized using the scattering volumes of the contributing components. For the protein-protein structure factor $S_{11}(Q)$, the scattering volume was the volume of a molecule of mCherry ($V_{mCherry}$) scaled by the volume fraction of mCherry in solution (ϕ_1).

$$S_{11}^{*}(Q) = \frac{S_{11}(Q)}{V_{mCherry}\phi_1} \tag{S10}$$

The molecular volume of mCherry was previously measured to be 33.724 nm^3.[4] The volume fraction was 0.05.

For the polymer-polymer structure factor, the scattering volume was the molecular volume of a single polymer chain scaled by the volume fraction of polymer in solution.

$$S_{22}^{*}(Q) = \frac{S_{22}(Q)}{V_{polymer}\phi_2} \tag{S11}$$

The volume of a polymer chain was calculated as the product of the degree of polymerization (N) and the molecular volume of a monomer unit (v). Table S4 lists the volume per chain and degree of polymerization for each polymer; the volume per monomer unit can be found in Table 1.

$$S_{22}^{*}(Q) = \frac{S_{22}(Q)}{V_{polymer}\phi_2} \tag{S12}$$

Table S4. Polymer Chain Volume

Polymer	N	$V_{polymer}$ **[nm^3]**
PNIPAM	232	9666
POEGA	175	7307
PDMAPS	92	2866

For the protein-polymer cross structure factor, the scattering volume was scaled by the geometric mean of the volumes of each contributing component.

$$S_{12}^{*}(Q) = \frac{S_{12}(Q)}{\left(\sqrt{V_{mCherry}\phi_1}\right)\left(\sqrt{V_{polymer}\phi_2}\right)} \tag{S13}$$

This scale factor is based on the prefactor of the decomposition of the cross structure factor into its intermolecular form factor, $\sqrt{\phi_1\phi_2}P_{12}^{I}(Q)$.[5]

$$S_{12}(Q) = \sqrt{V_{mCherry}\phi_1}\sqrt{V_{polymer}\phi_2}\left(\sqrt{\phi_1\phi_2}P_{12}^{I}(Q)\right) \tag{S14}$$

Partial Structure Factors on Different Scales

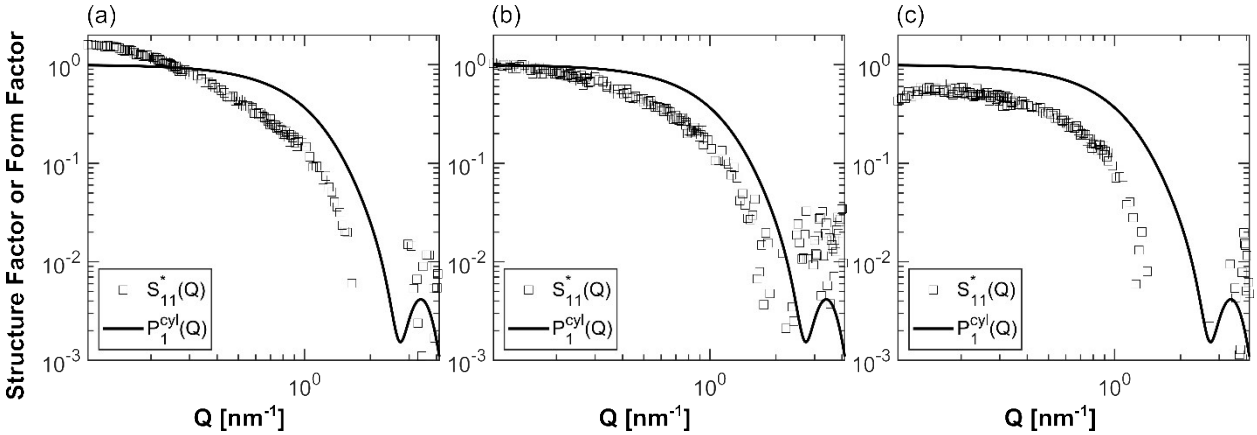

Figure S9. Single molecule and intermolecular components of the non-dimensionalized protein-protein structure factor S_{11}^* plotted on a log-log scale in the presence of (a) PNIPAM, (b) POEGA, and (c) PDMAPS. The single molecule component P_1^{cyl} is the cylinder form factor for mCherry.

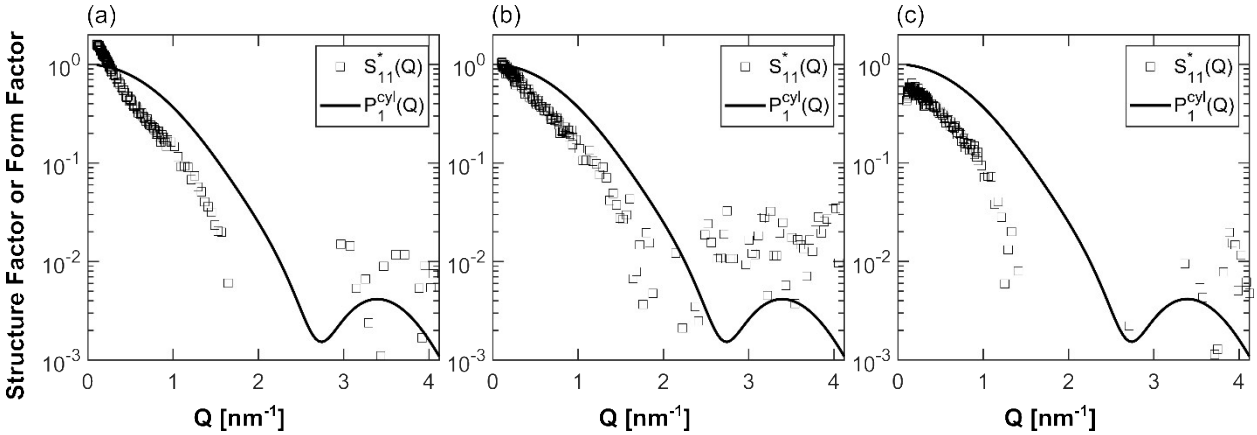

Figure S10. Single molecule and intermolecular components of the non-dimensionalized protein-protein structure factor S_{11}^* plotted on a linear-log scale in the presence of (a) PNIPAM, (b) POEGA, and (c) PDMAPS. The single molecule component P_1^{cyl} is the cylinder form factor for mCherry.

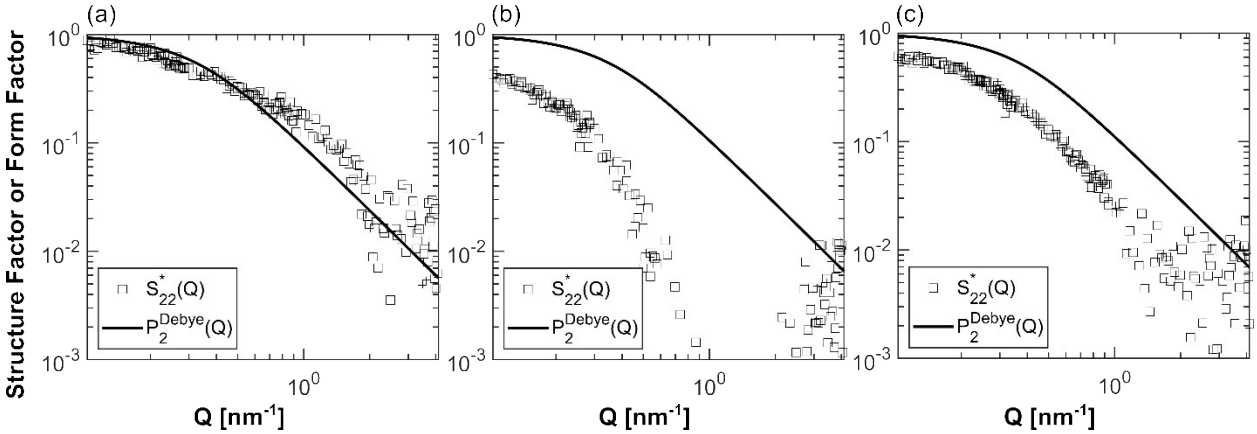

Figure S11. Single molecule and intermolecular components of the non-dimensionalized polymer-polymer structure factor S_{22}^* plotted on a log-log scale for (a) PNIPAM, (b) POEGA, and (c) PDMAPS. The single molecule component P_2^{Debye} is the Debye form factor for Gaussian chains.

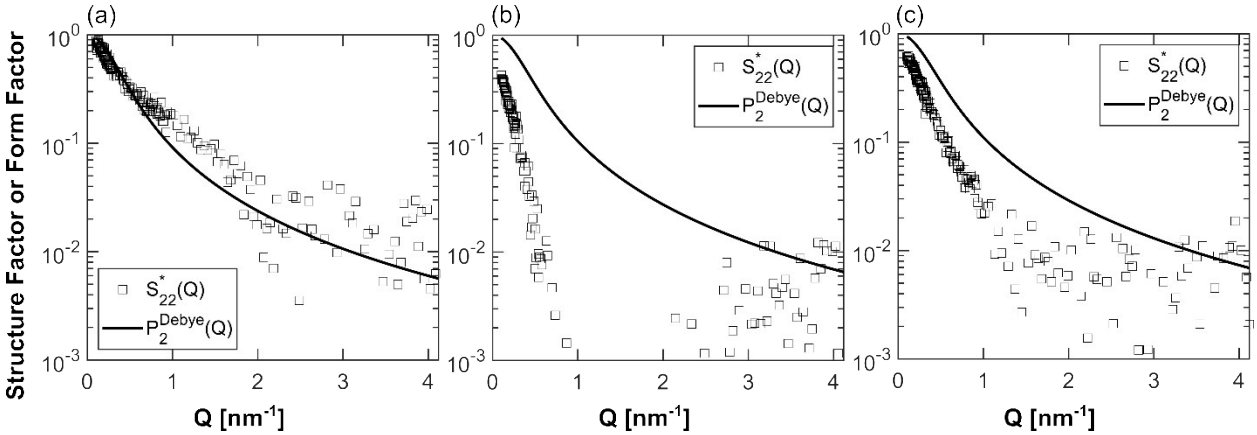

Figure S12. Single molecule and intermolecular components of the non-dimensionalized polymer-polymer structure factor S_{22}^* plotted on a linear-log scale for (a) PNIPAM, (b) POEGA, and (c) PDMAPS. The single molecule component P_2^{Debye} is the Debye form factor for Gaussian chains.

Sensitivity of Structure Factor and Concentration Correlation Function to Hydration Number

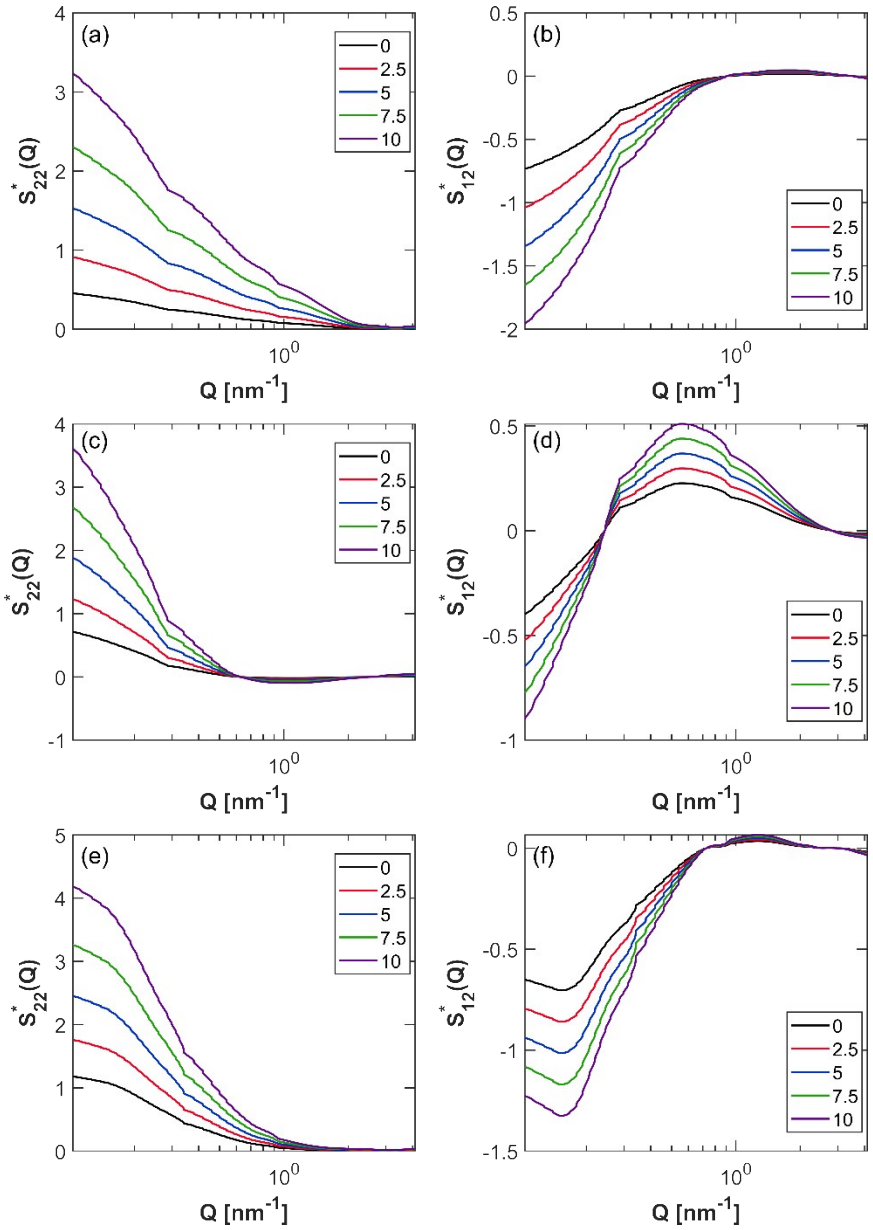

Figure S13. Variation in dimensionless partial structure factor S_{ij}^* with respect to hydration water for (a) PNIPAM/PNIPAM interactions, (b) PNIPAM/mCherry interactions, (c) POEGA/POEGA interactions, (d) POEGA/mCherry interactions, (e) PDMAPS/PDMAPS interactions, and (f) PDMAPS/mCherry interactions demonstrating that hydration number tunes the strength but not nature of interactions.

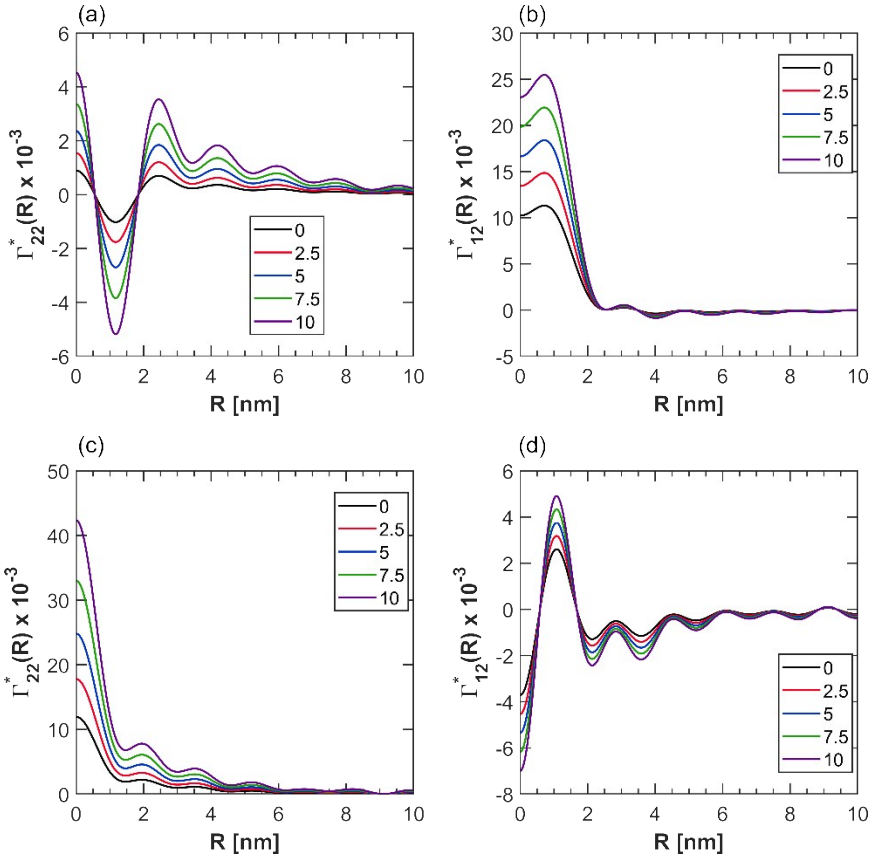

Figure S14. Variation in dimensionless concentration correlation function Γ_{ij}^* with respect to hydration water for (a) POEGA/POEGA interactions, (b) POEGA/mCherry interactions, (c) PDMAPS/PDMAPS interactions, and (d) PDMAPS/mCherry interactions demonstrating that hydration number tunes the strength but not nature of interactions.

References

1. J. D. Nickels, J. Atkinson, E. Papp-Szabo, C. Stanley, S. O. Diallo, S. Perticaroli, B. Baylis, P. Mahon, G. Ehlers, J. Katsaras and J. R. Dutcher, *Biomacromolecules*, 2016, **17**, 735-743.
2. M. Rubinstein and R. H. Colby, *Polymer Physics*, Oxford University Press, New York, 2017.
3. M. Surve, V. Pryamitsyn and V. Ganesan, *The Journal of Chemical Physics*, 2005, **122**, 154901.
4. C. N. Lam, D. Chang, M. Wang, W.-R. Chen and B. D. Olsen, *Journal of Polymer Science Part A: Polymer Chemistry*, 2016, **54**, 292-302.
5. B. Hammouda, *Probing Nanoscale Structures: The SANS Toolbox*, NIST, Gaithersburg, MD, 2010.