Electronic Supplementary Material (ESI) for Polymer Chemistry.

Supporting Information

Secondary structure drives self-assembly in weakly segregated globular protein-rod block copolymers

Helen Yao[1†], *Kai Sheng*[2†], *Jialing Sun*[2], *Shupeng Yan*[2], *Yinqing Hou*[2], *Hua Lu*[2], *Bradley D. Olsen*[1*]

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

[2] College of Chemistry and Molecular Engineering, Peking University, Beijing 100871 P.R.China

[†] H.Y. and K.S. contributed equally to this work.

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

Sequence of TEV-CG-eGFP

MENLYFQCGGGGSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNF
KIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFV
TAAGITLGMDELYKGLEHHHHHH

TEV cleavage site
Gly_4Ser linker

Poly(amino acid) (PAA) Gel Permeation Chromatography (GPC) Traces

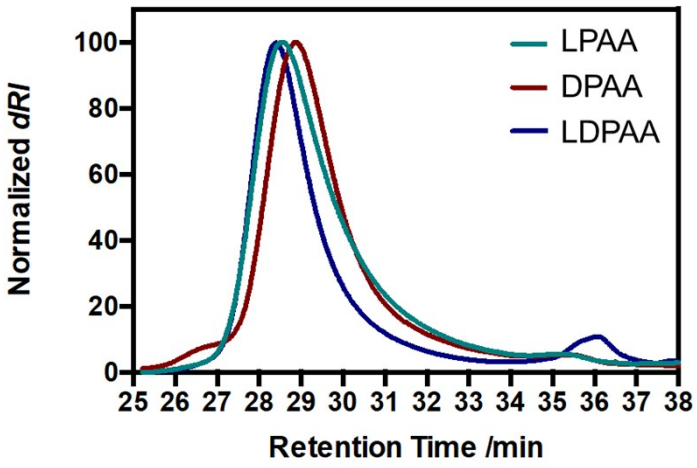

Figure S1. Normalized differential refractive index (*dRI*) signal as a function of retention time as measured from GPC for each PAA.

Table S1. Polymer GPC characterization

Name	Polymer	Molar Mass (kDa)	Degree of Polymerization	Đ
LPAA	$P(EG)_3$-$_L$-Glu_{88}	24.3	88	1.02
DPAA	$P(EG)_3$-$_D$-Glu_{97}	26.7	97	1.02
LDPAA	$P(EG)_3$-$_{L/D}$-Glu_{100}	27.5	100	1.02

Proton Nuclear Magnetic Resonance (NMR) Spectra for PAAs

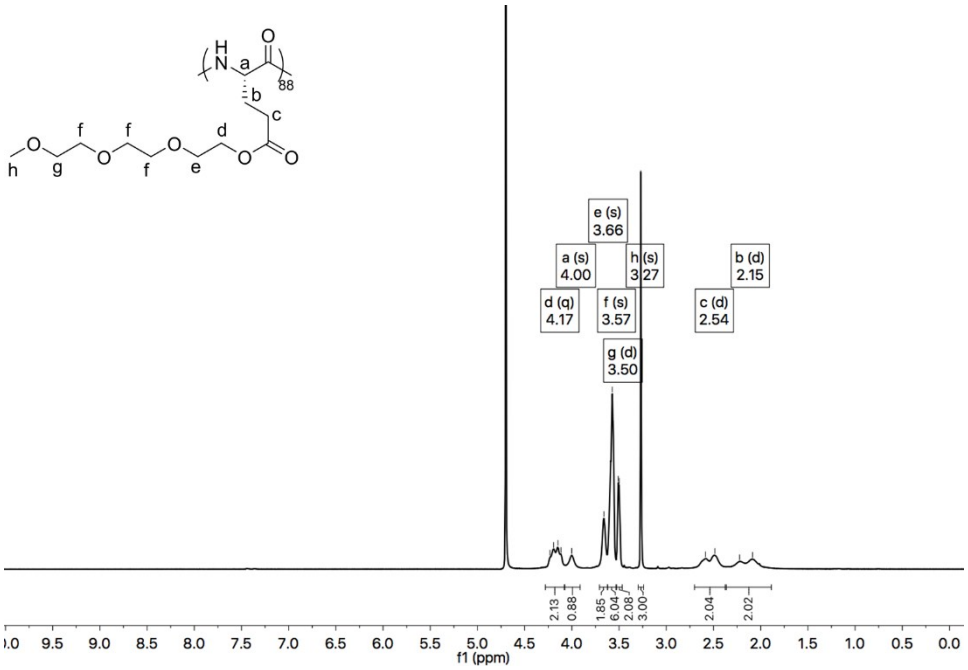

Figure S2. Proton NMR spectrum of $P(EG)_3$-$_L$-Glu_{88} PAA (LPAA) with proton groups labeled on the inset molecule. Peak areas are shown underneath the spectrum and peak chemical shifts are shown next to each peak with the proton group.

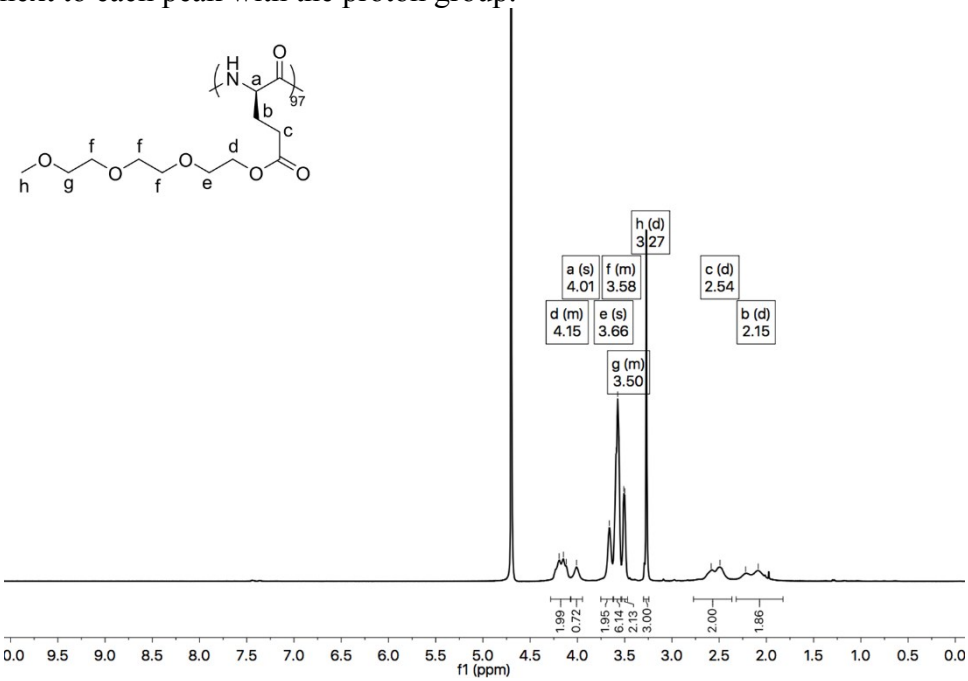

Figure S3. Proton NMR spectrum of $P(EG)_3$-$_D$-Glu_{97} PAA (DPAA) with proton groups labeled on the inset molecule. Peak areas are shown underneath the spectrum and peak chemical shifts are shown next to each peak with the proton group.

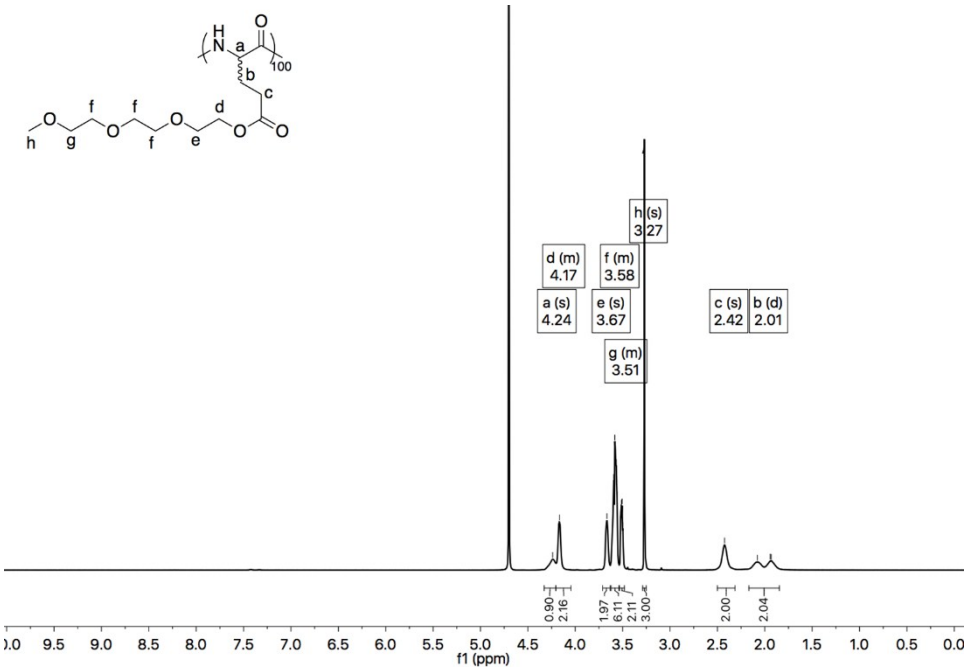

Figure S4. Proton NMR spectrum of $P(EG)_{3\text{-}L/D}\text{-}Glu_{100}$ PAA (LDPAA) with proton groups labeled on the inset molecule. Peak areas are shown underneath the spectrum and peak chemical shifts are shown next to each peak with the proton group.

Mass Spectra for TEV-CG-eGFP and CG-eGFP

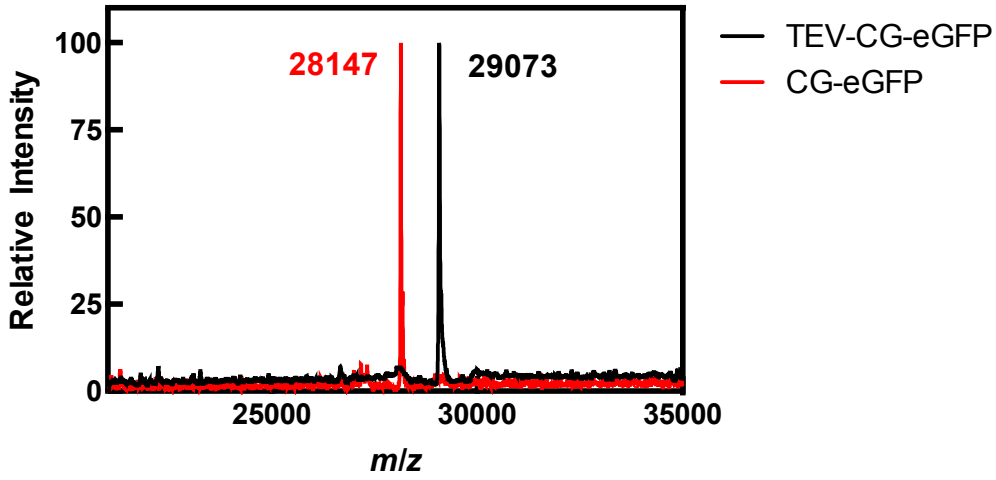

Figure S5. Mass spectra of TEV-CG-eGFP and CG-eGFP.

Table S2. Mass spectrometry results

Species	Expected Mass[a]	Observed Mass
TEV-CG-eGFP	29091	29073
CG-eGFP	28165	28147

[a] Expected mass was calculated from the protein sequence.

Cloud point temperatures for PAAs

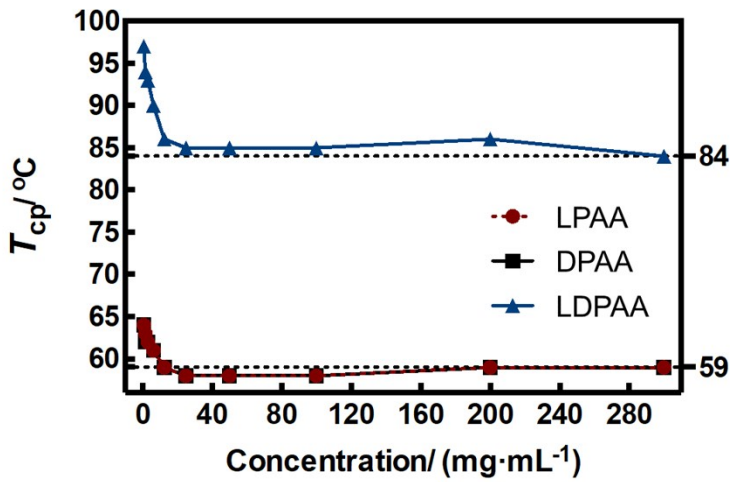

Figure S6. Cloud point temperature (T_{cp}) of all PAAs as a function of concentration in aqueous solution. For all PAAs, T_{cp} was too high to measure during SAXS because eGFP is expected to denature by 50 °C. The right axis is labeled with the estimated cloud point temperature at high concentration (300 mg/mL).

Small Angle X-ray Scattering (SAXS) Intensity Curves

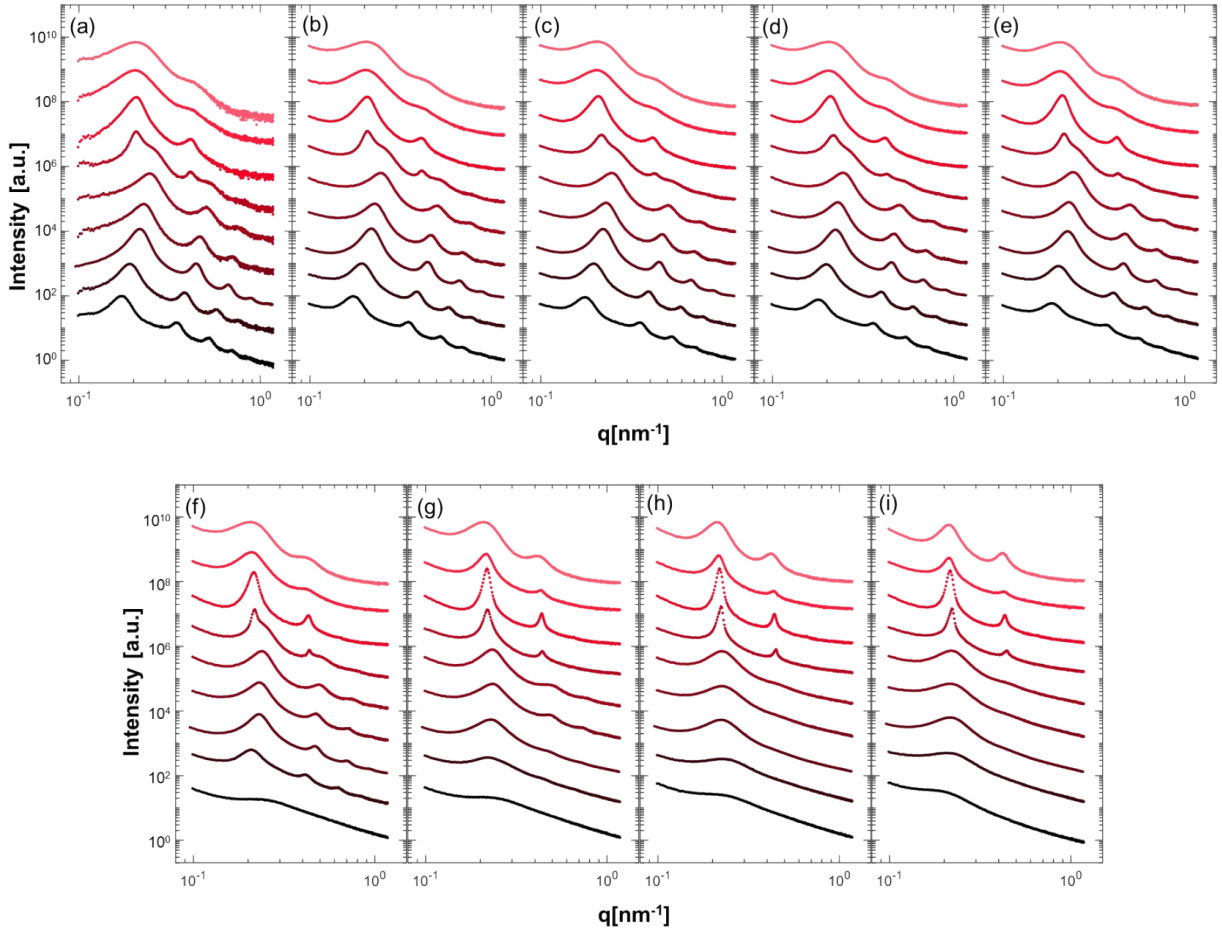

Figure S7. SAXS intensity curves for LPAA-eGFP at (a) 10 °C, (b) 15 °C, (c) 20 °C, (d) 25 °C, (e) 30 °C, (f) 35 °C, (g) 40 °C, (h) 45 °C, and (i) 50 °C. For each plot, from bottom to top, the concentration varies from 20 to 60 wt% in 5 wt% increments. SAXS curves are offset for clarity.

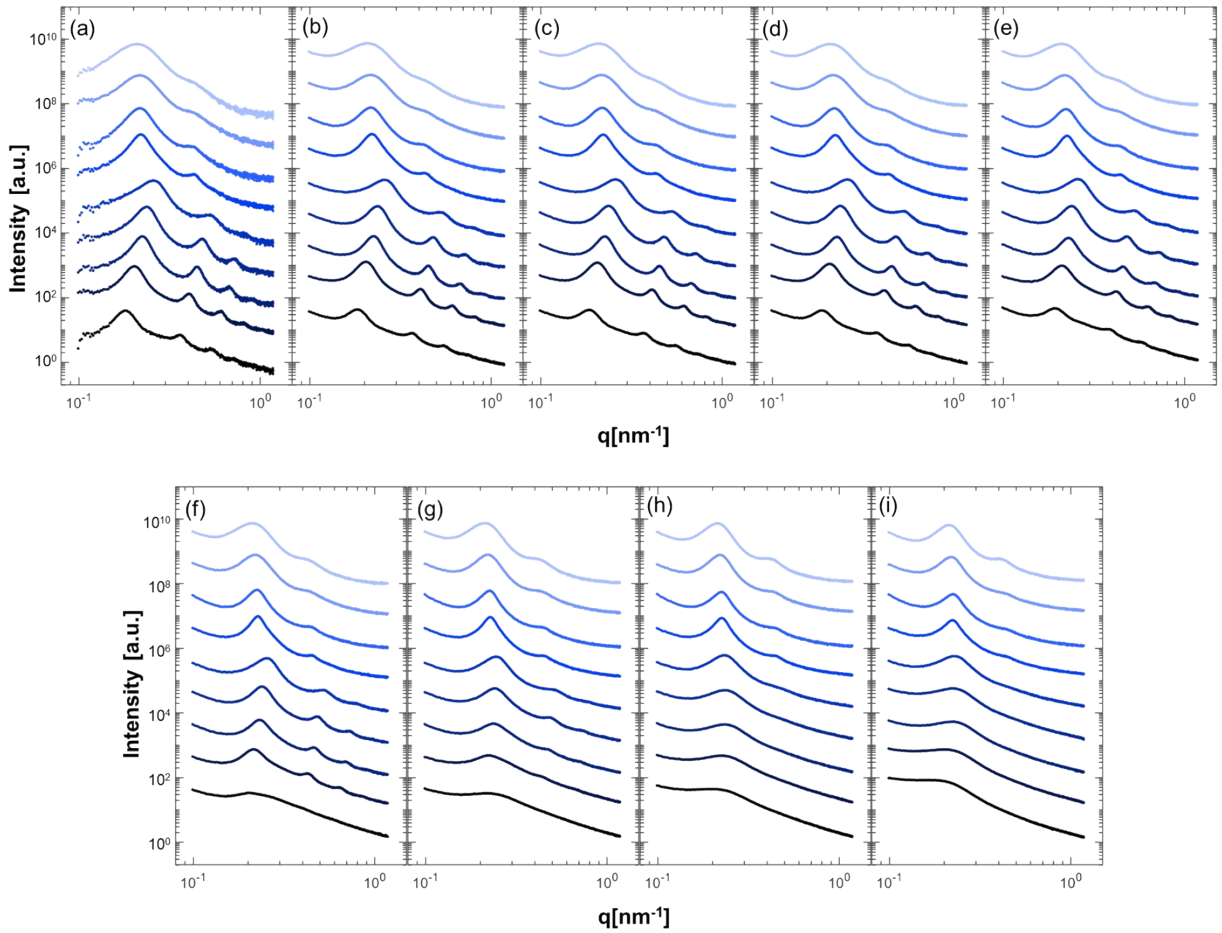

Figure S8. SAXS intensity curves for DPAA-eGFP at (a) 10 °C, (b) 15 °C, (c) 20 °C, (d) 25 °C, (e) 30 °C, (f) 35 °C, (g) 40 °C, (h) 45 °C, and (i) 50 °C. For each plot, from bottom to top, the concentration varies from 20 to 60 wt% in 5 wt% increments. SAXS curves are offset for clarity.

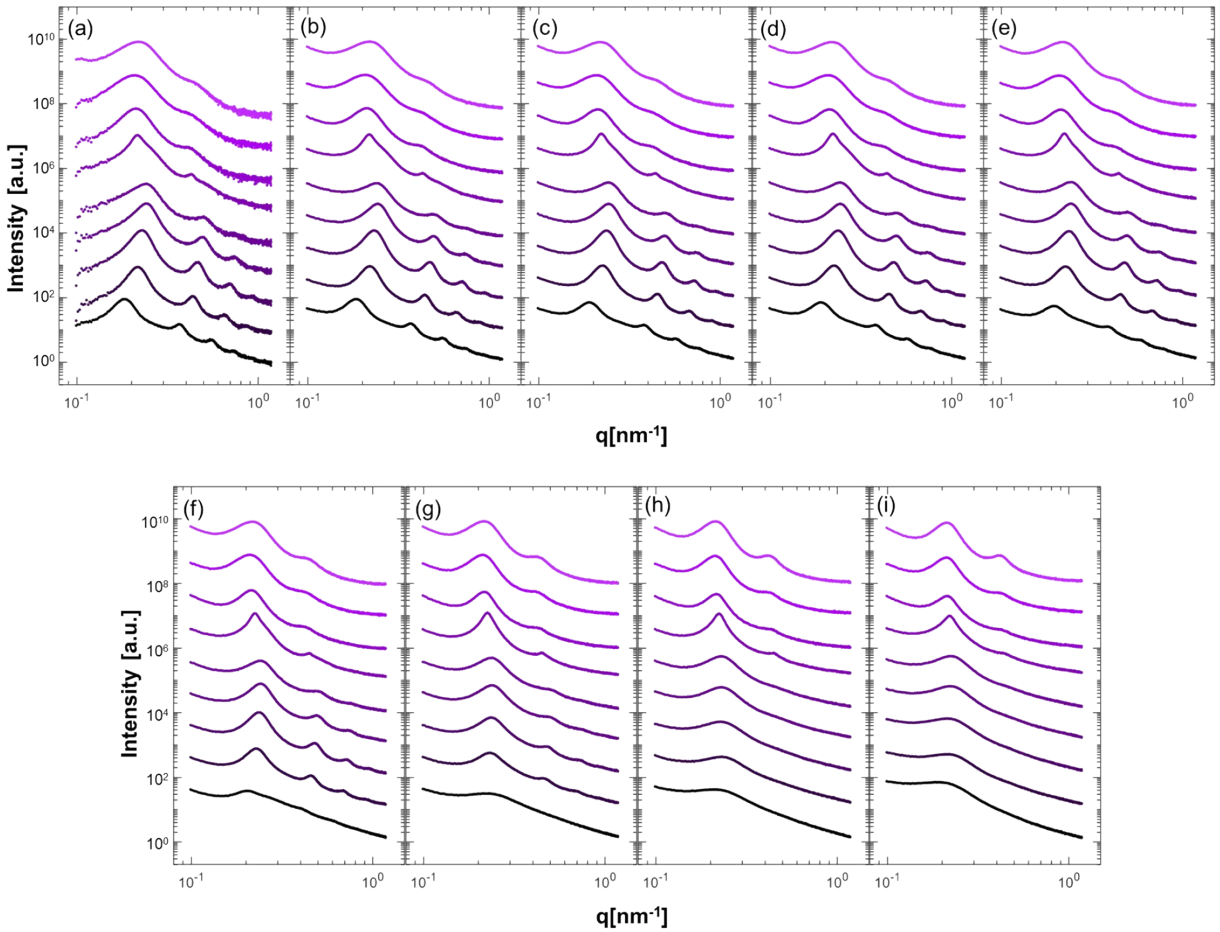

Figure S9. SAXS intensity curves for L:D1:1-eGFP at (a) 10 °C, (b) 15 °C, (c) 20 °C, (d) 25 °C, (e) 30 °C, (f) 35 °C, (g) 40 °C, (h) 45 °C, and (i) 50 °C. For each plot, from bottom to top, the concentration varies from 20 to 60 wt% in 5 wt% increments. SAXS curves are offset for clarity.

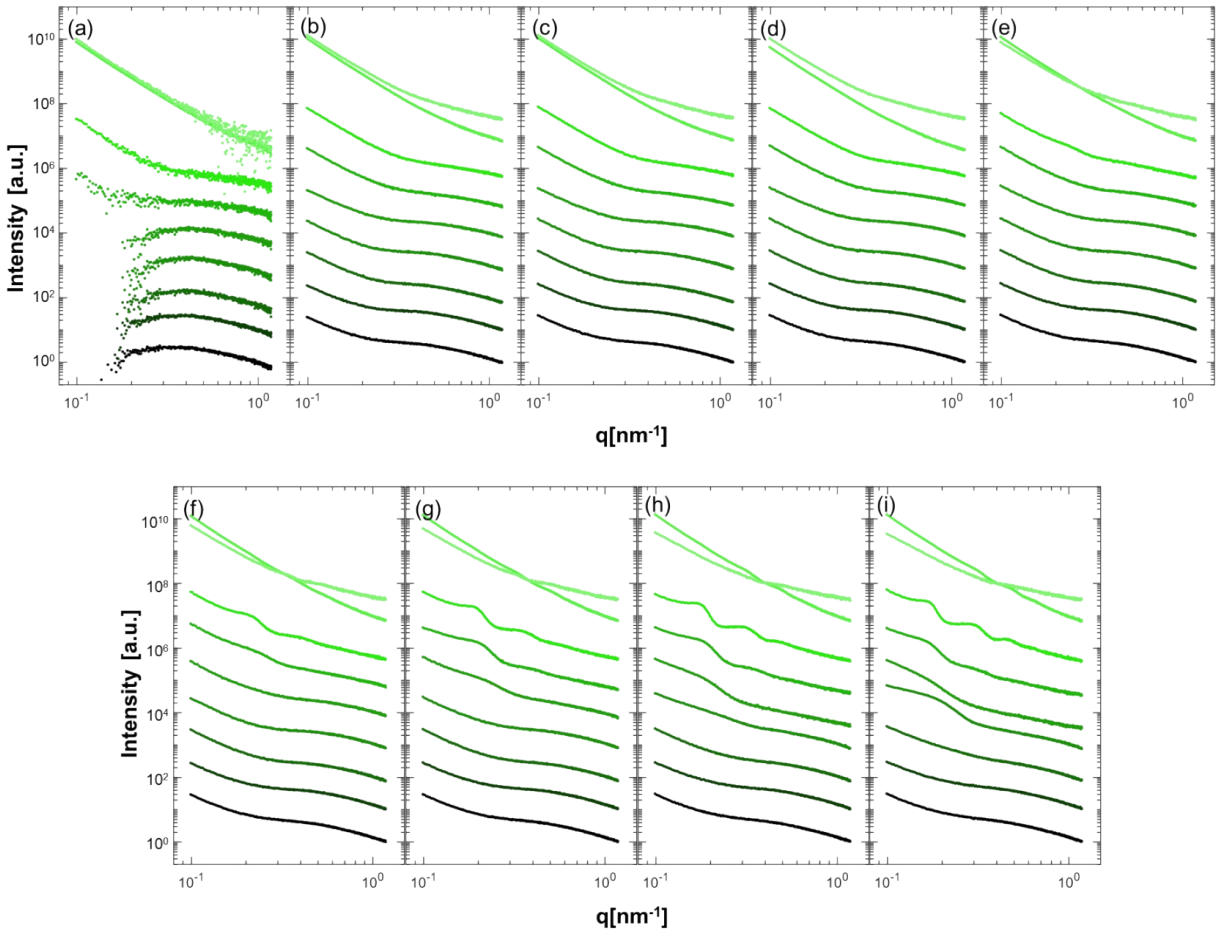

Figure S10. SAXS intensity curves for LDPAA-eGFP at (a) 10 °C, (b) 15 °C, (c) 20 °C, (d) 25 °C, (e) 30 °C, (f) 35 °C, (g) 40 °C, (h) 45 °C, and (i) 50 °C. For each plot, from bottom to top, the concentration varies from 20 to 60 wt% in 5 wt% increments. SAXS curves are offset for clarity.

Birefringence and Transmission Data (from Depolarized Light Scattering)

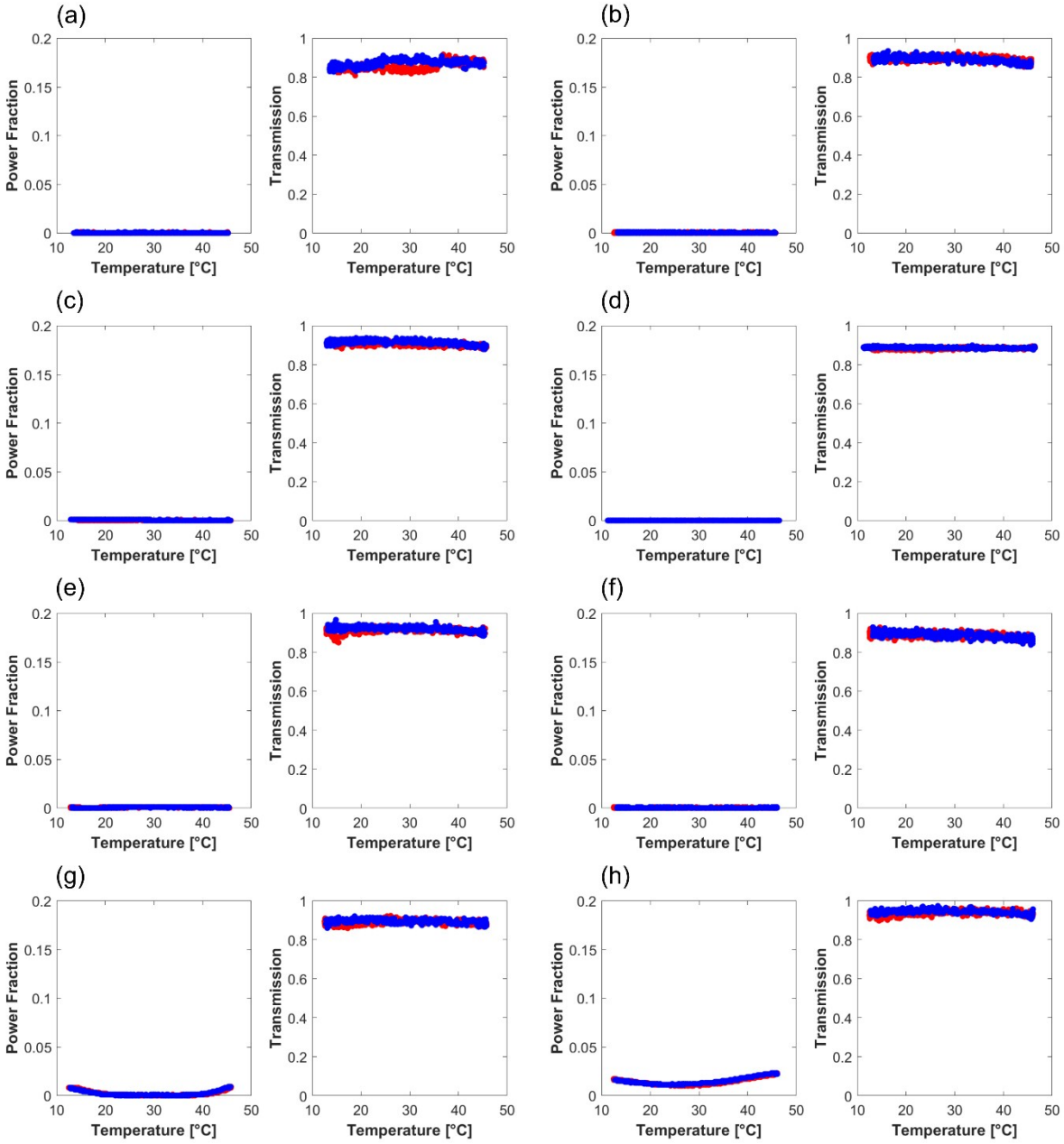

Figure S11. LPAA-eGFP birefringence (power fraction) and transmission curves as a function of temperature as measured by depolarized light scattering (DPLS). Heating ramps are depicted in red, and cooling ramps are depicted in blue. Each double panel corresponds to a different wt%: (a) 25%, (b) 30%, (c) 35%, (d) 40%, (e) 45%, (f) 50%, (g) 55%, (h) 60%. There is no plot for 20% because this sample was too liquid-like to remain in the sample holder. Based on the behavior of the more dilute samples, 20% is not expected to exhibit any birefringence.

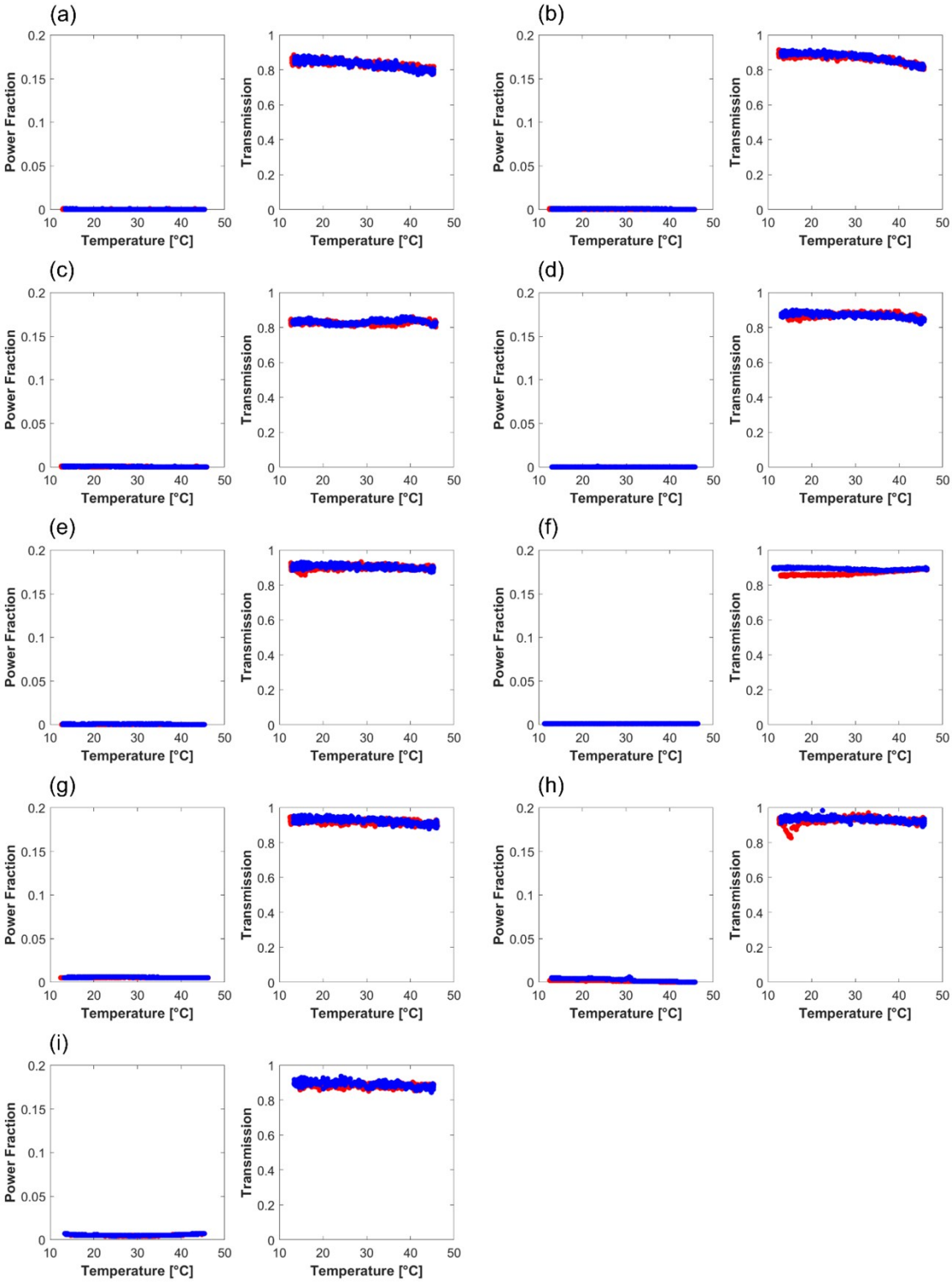

Figure S12. DPAA-eGFP birefringence (power fraction) and transmission curves as a function of temperature as measured by depolarized light scattering (DPLS). Heating ramps are depicted in red, and cooling ramps are depicted in blue. Each double panel corresponds to a different wt%: (a) 20%, (b) 25%, (c) 30%, (d) 35%, (e) 40%, (f) 45%, (g) 50%, (h) 55%, (i) 60%.

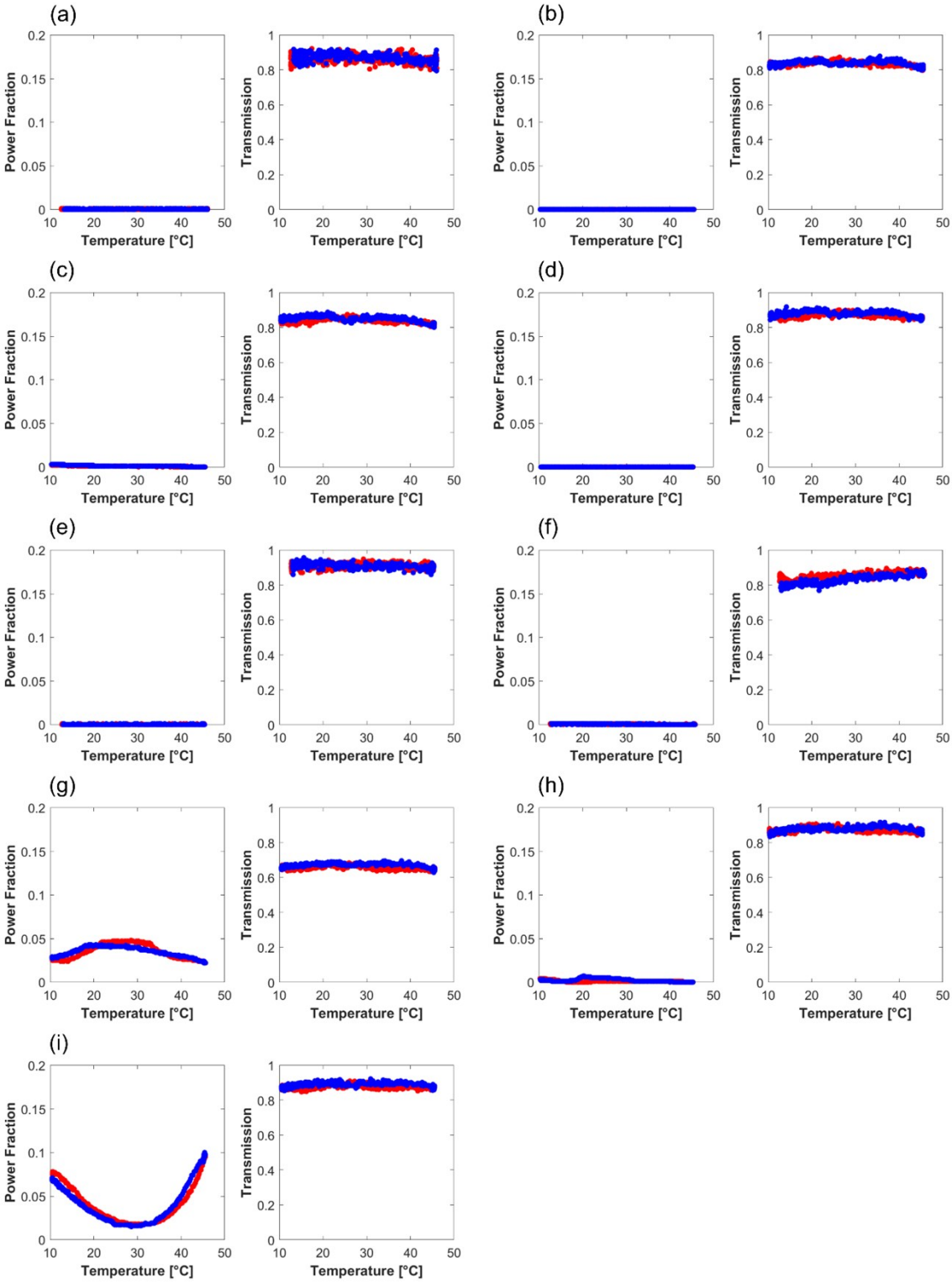

Figure S13. L:D1:1-eGFP birefringence (power fraction) and transmission curves as a function of temperature as measured by depolarized light scattering (DPLS). Heating ramps are depicted in red, and cooling ramps are depicted in blue. Each double panel corresponds to a different wt%: (a) 20%, (b) 25%, (c) 30%, (d) 35%, (e) 40%, (f) 45%, (g) 50%, (h) 55%, (i) 60%.

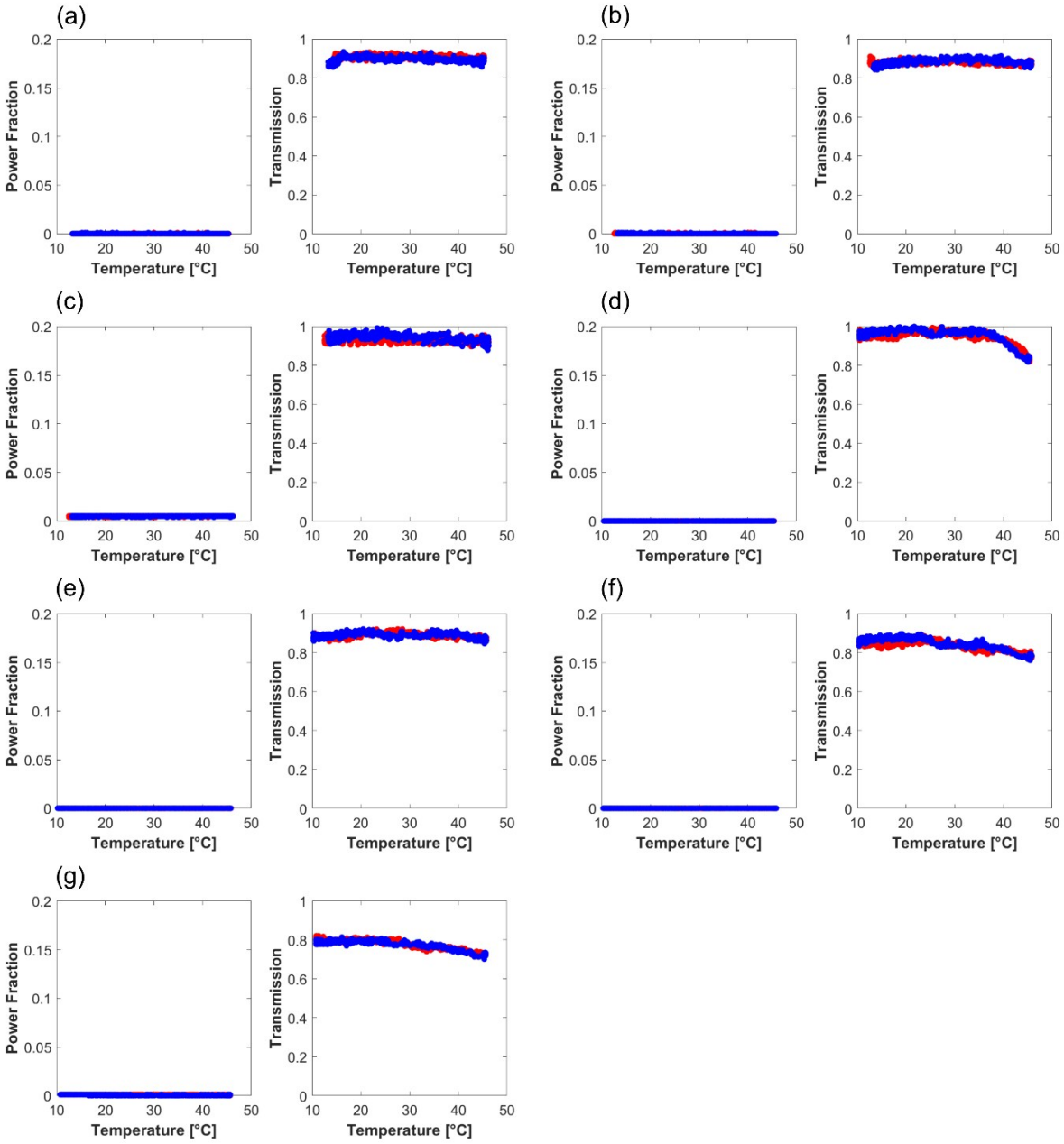

Figure S14. LDPAA-eGFP birefringence (power fraction) and transmission curves as a function of temperature as measured by depolarized light scattering (DPLS). Heating ramps are depicted in red, and cooling ramps are depicted in blue. Each double panel corresponds to a different wt%: (a) 20%, (b) 25%, (c) 30%, (d) 35%, (e) 40%, (f) 45%, (g) 50%. There are no data for 55 and 60 wt% because these were too turbid to measure.

Domain Spacing

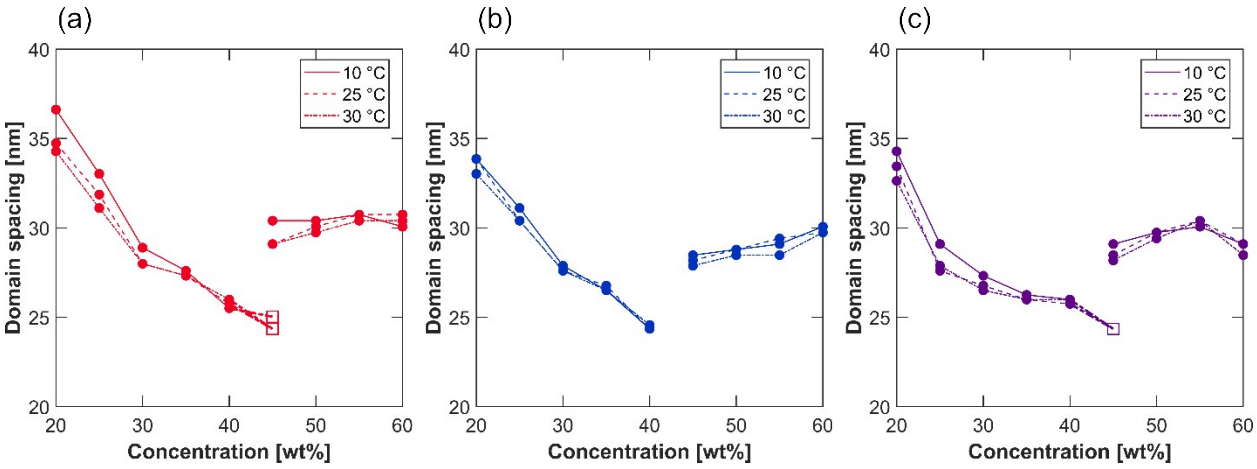

Figure S15. Domain spacing as a function of concentration at different temperatures for (a) LPAA-eGFP, (b) DPAA-eGFP, and (c) L:D1:1-eGFP. There is a non-monotonic dependence of domain spacing on concentration due to the presence of two lamellar phases but no significant dependence on temperature. The open square is the domain spacing calculated for the second set of peaks for 45 wt%, which exhibited two lamellar phases in samples containing LPAA-eGFP.

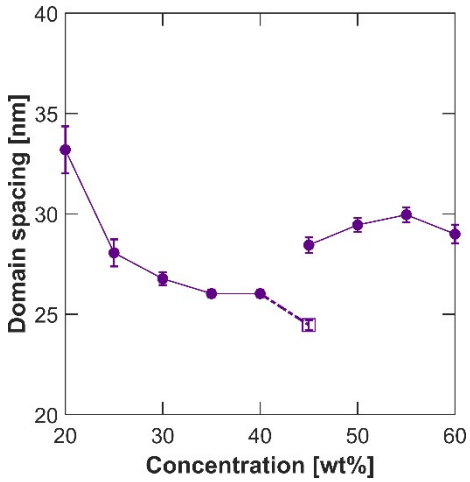

Figure S16. Domain spacing as a function of concentration averaged over all temperatures for L:D1:1-eGFP. There is a non-monotonic dependence of domain spacing on concentration. The closed circles represent domain spacing calculated from the main SAXS peak. The open square is the domain spacing calculated for the second set of peaks for 45 wt.%, which exhibited two lamellar phases.

Circular Dichroism (CD) Spectroscopy for the L:D1:1-eGFP Blend

CD spectroscopy was used to confirm that the fold of the 1:1 blend (by mass of bioconjugate) of LPAA-eGFP and DPAA-eGFP retained its fold after blending and lyophilization. A pre-lyophilization 1:1 blend was prepared by mixing LPAA-eGFP and DPAA-eGFP and dissolving at 0.5 mg/mL in water. This was vortexed and rocked at 4 °C overnight to allow for equilibration. The post-lyophilization 1:1 blend was prepared by dissolving the blended sample in water at 0.5 mg/mL, vortexing, and rocking at 4 °C overnight. The CD experiments were conducted on at the MIT Biophysical Instrumentation Facility (BIF) on a Jasco J-1500 CD spectrometer with a data pitch of 0.5 nm, a scan speed of 50 nm/min, a CD scale of 200 mdeg/1.0 dOD, and a digital integration time of 4 s. The data were baseline subtracted using a water background and averaged over three replicates. The CD spectra shown in Figure S17 were taken at three temperatures: 10, 25, and 50 °C, which span the range of the SAXS experiments. The lyophilized product has a slightly smaller CD signal from the raw material but the same features. This is indicative of a concentration difference between the two samples, due to slight differences in weighing the sample.

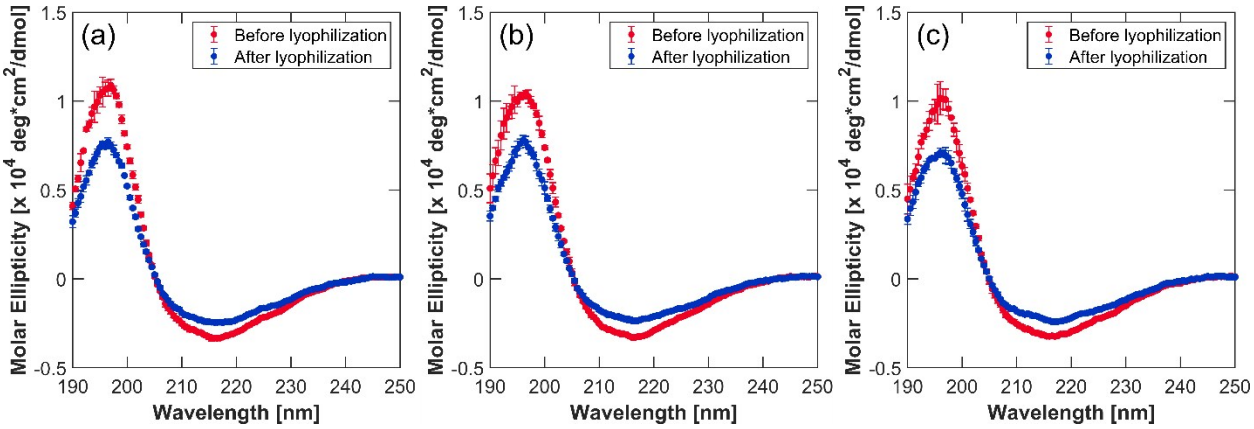

Figure S17. CD spectra for 1:1 blend of LPAA-eGFP and DPAA-eGFP before and after lyophilization at (a) 10 °C, (b) 25 °C, and (c) 50 °C.

Hard Sphere Structure Factor with Percus-Yevick Closure

High concentration LDPAA-eGFP bioconjugates (40 – 50 wt%) at high temperature (> 35 °C) exhibited SAXS curves with wide correlation peaks indicative of aggregation (Figure S10). These can be fit to a hard sphere structure factor with Percus-Yevick closure.[1] The SAXS intensity curves are generally described by Equation S1, where I is the intensity, A is a scale factor, P is the form factor, S is the structure factor, and B is the background intensity.

$$I(q) = AP(q)S(q) + B(q) \tag{S1}$$

The background intensity can be described by the following power law relation (Equation S2), where B_1, B_2, and B_3 are empirical fitting parameters.

$$B(q) = \frac{B_1}{q^4} + \frac{B_2}{q^2} + B_3 \tag{S2}$$

SAXS curves were first masked to remove the correlation peaks and subsequently fit to Equation S2 using non-linear least-squares fitting in MATLAB with the χ^2 parameter as the objective function (**lsqnonlin**). The background scattering was subtracted from the unmasked SAXS curves and fit to Equation S1 in SasView.[2] The sphere form factor (Equation S3) was selected to pair with the hard sphere structure factor.

$$P(q) = P_1\left[\frac{3\left[\sin(qr) - qr\cos(qr)\right]}{(qr)^3}\right]^2 \tag{S3}$$

In Equation S3, P_1 is a scaling factor (which was included in A when fitting) and r is the radius of a sphere. The structure factor is of the form

$$S(q) = \frac{1}{1 - NC(q)} \tag{S4}$$

where $NC(q)$ is derived from the hard sphere potential with Percus-Yevick closure (Equation S5).

$$NC(q) = -24\phi\left(\lambda_1 S_1 - 6\phi\lambda_2 S_2 - \phi\frac{\lambda_1}{2}(S_3 + S_4) \right)$$

$$S_1(q) = \frac{\sin(2qr) - 2qr\cos(2qr)}{(2qr)^3}$$

$$S_2(q) = \frac{(2qr)^2\cos(2qr) - 4qr\sin(2qr) - 4\cos(2qr) + 2}{(2qr)^4} \quad \text{(S4)}$$

$$S_3(q) = \frac{(2qr)^4\cos(2qr) - 4(2qr)^3\sin(2qr) - 12(2qr)^2\cos(2qr)}{(2qr)^6}$$

$$S_4(q) = \frac{24(2qr)\sin(2qr) + 24\cos(2qr) - 24}{(2qr)^6}$$

In Equation S5, ϕ is the volume fraction, r is the radius of the spheres, $\lambda_1 = \frac{(1+2\phi)^2}{(1-\phi)^4}$, and $\lambda_2 = -\frac{(1-\phi/2)^2}{(1-\phi)^4}$.

The resulting fits are shown in Figure S18. The scaling factors could not be extracted since the SAXS curves were not calibrated to an absolute standard. The aggregates sizes and volume fractions are reported in Table S3. Sphere radius from the form factor and the structure factor were confirmed to be within error of one another.

Table S3. Sphere radius and volume fraction for LDPAA-eGFP at select conditions

Concentration [wt%]	**Temperature [°C]**	**Radius, r (form factor) [nm]**	**Radius, r (structure factor) [nm]**	**Volume fraction [ϕ]**
45	45	160 ± 60	180 ± 20	0.36 ± 0.1
45	50	170 ± 90	210 ± 30	0.39 ± 0.1
50	35	160 ± 100	160 ± 40	0.4 ± 0.3
50	40	160 ± 50	180 ± 20	0.45 ± 0.1
50	45	190 ± 40	190 ± 10	0.45 ± 0.08
50	50	200 ± 30	200 ± 10	0.46 ± 0.08

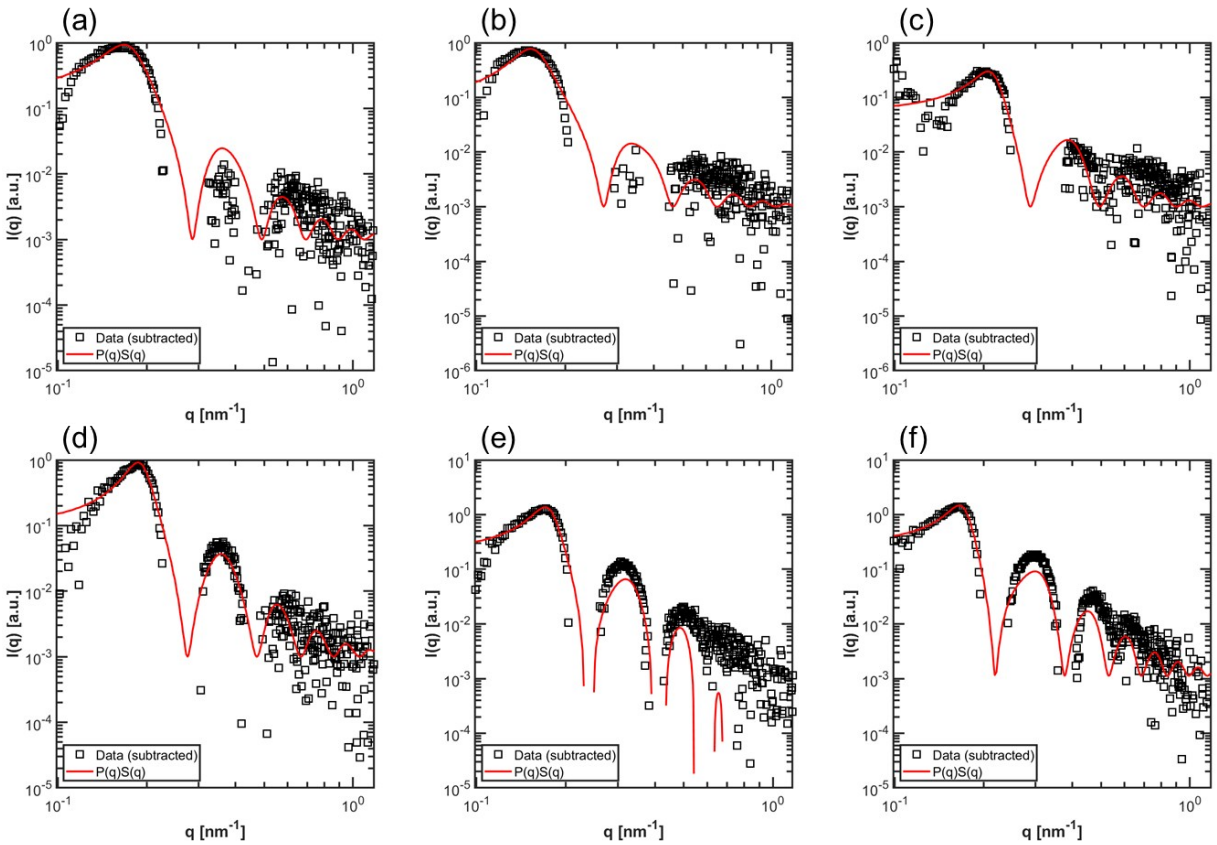

Figure S18. SAXS subtracted intensity curves for LDPAA-eGFP with a fit to the hard sphere structure factor (Equations S1, S3, and S4) for (a) 45 wt%, 45 °C, (b) 45 wt%, 50 °C, (c) 50 wt%, 35 °C, (d) 50 wt%, 40 °C, (e) 50 wt%, 45 °C, and (f) 50 wt%, 50 °C.

References

1. J. K. Percus and G. J. Yevick, *Phys Rev*, 1958, **110**, 1-13.
2. SasView, http://www.sasview.org/, (accessed 1/13/2020).