SUPPORTING INFORMATION

Noncanonical Self-Assembly of Highly Asymmetric Genetically Encoded Polypeptide Amphiphiles into Cylindrical Micelles

*Jonathan R. McDaniel†, Isaac Weitzhandler†, Sylvain Prevost‡, Kevin B. Vargo§, Marie-Sousai Appavou‖, Daniel A. Hammer§, Michael Gradzielski‡, Ashutosh Chilkoti†**

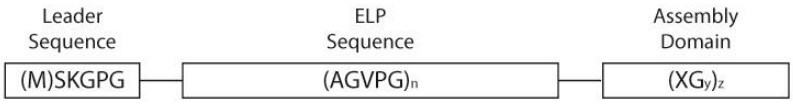
†Department of Biomedical Engineering and Research Triangle MRSEC, Duke University, Durham, North Carolina, 27708, USA.

‡Stranski-Laboratorium fur Physikalische und Theoretische Chemie, Institut fur Chemie, Technische Universitat Berlin, 10623, Berlin, Germany.

§Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA, 19104, USA.

‖Jülich Centre for Neutron Science (JCNS). Forschungszentrum Jülich GmbH. Outstation at MLZ. Lichtenbergstraße 1. 85747 Garching, Germany.

*Professor Ashutosh Chilkoti, chilkoti@duke.edu.



Leader Sequence

ATG AGC AAA GGG CCG GGC
(M) S K G P G

ELP Sequence

(GCC GGA GTG CCT GGT GCA GGT GTG CCA GGC GCG GGT GTT CCA GGA GCA GGC GTT CCA GGT GCG GGT GTT CCT GGC
A G V P G A G V P G A G V P G A G V P G A G V P G

GCC GGA GTG CCT GGT GCA GGT GTG CCA GGC GCG GGT GTT CCA GGA GCA GGC GTT CCA GGT GCG GGT GTT CCT GGC)$_n$
A G V P G A G V P G A G V P G A G V P G A G V P G

$(LGG)_8$

CTG GGC GGT CTG GGT GGC CTG GGT GGT CTG GGT GGC CTG GGC GGT CTG GGT GGT CTG GGC GGC TTG GGT GGC
L G G L G G L G G L G G L G G L G G L G G L G G

TAC TGA TAA ACT
Y . . .

$(IGG)_8$

ATC GGC GGT ATT GGT GGC ATT GGT GGT ATC GGT GGC ATC GGC GGT ATC GGT GGT ATT GGC GGC ATC GGT GGC
I G G I G G I G G I G G I G G I G G I G G I G G

TAC TGA TAA ACT
Y . . .

$(HGG)_8$

CAT GGC GGT CAC GGC GGC CAC GGT GGT CAC GGT GGC CAT GGC GGT CAT GGT GGT CAC GGC GGT CAT GGT GGC
H G G H G G H G G H G G H G G H G G H G G H G G

TAC TGA TAA ACT
Y . . .

$(WGG)_8$

TGG GGC GGT TGG GGC GGC TGG GGT GGT TGG GGT GGC TGG GGC GGT TGG GGT GGT TGG GGC GGC TGG GGT GGC
W G G W G G W G G W G G W G G W G G W G G W G G

TAC TGA TAA ACT
Y . . .

$(FGG)_8$

TTC GGC GGT TTT GGC GGC TTC GGT GGT TTT GGT GGC TTT GGC GGT TTC GGT GGT TTT GGC GGC TTC GGT
F G G F G G F G G F G G F G G F G G F G G F G

TAC TGA TAA ACT
Y . . .

$(FG)_8$

TTC GGC TTT GGC TTC GGT TTC GGT TTT GGT TTC GGT TTC GGC TTT GGC TAC TGA TAA ACT
F G F G F G F G F G F G F G F G Y . . .

F_8

TTC TTT TTC TTC TTT TTC TTC TTT TAC TGA TAA
F F F F F F F F Y . .

$(YGG)_8$

TAT GGC GGT TAC GGC GGC TAC GGT GGT TAT GGT GGC TAT GGC GGT TAC GGT GGT TAT GGC GGC TAC GGT GGC
Y G G Y G G Y G G Y G G Y G G Y G G Y G G Y G G
TAC TGA TAA ACT
Y . . .

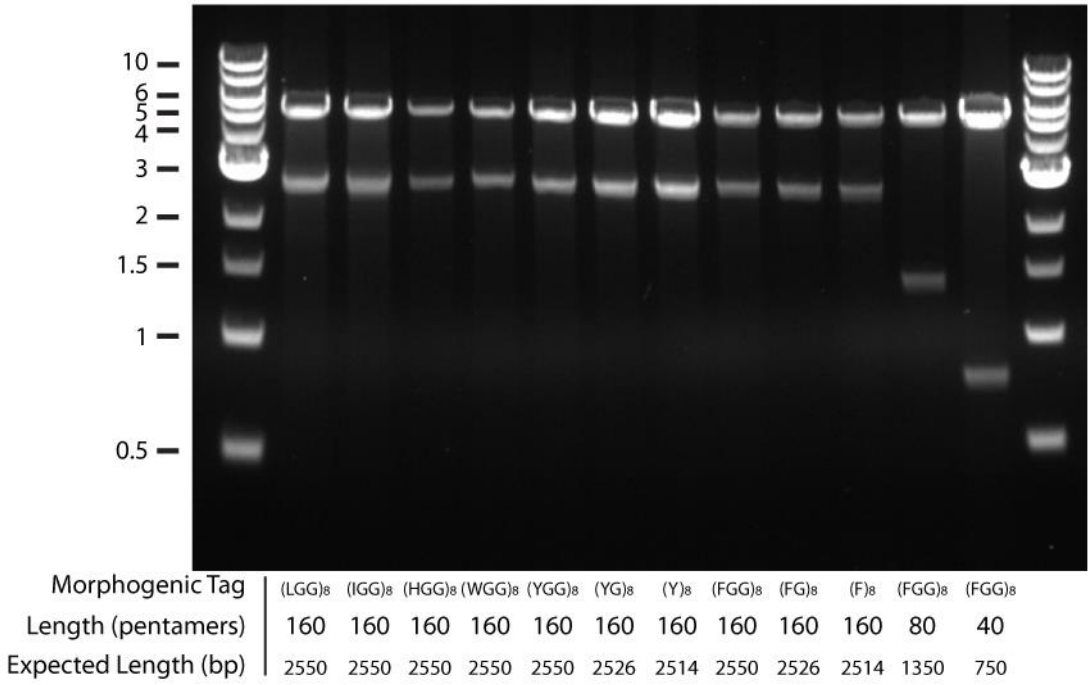
$(YG)_8$

TAT GGC TAC GGC TAC GGT TAT GGT TAT GGT TAC GGT TAT GGC TAC GGC TAC TGA TAA ACT
Y G Y G Y G Y G Y G Y G Y G Y G Y G Y . . .

Y_8

TAT TAC TAT TAT TAT TAC TAC TAT TAC TGA TAA
Y Y Y Y Y Y Y Y Y . .

Figure S1. Gene sequence of genetically encoded amphiphiles. The amphiphiles consist of a leader peptide (MSKGP) to increase the yield of the purified product, followed by the ELP sequence and then by the assembly domain. The methionine is cleaved during expression. The repeat unit n=4, 8, and 16 represent the 40, 80, and 160 pentamer sequences, respectively. Each assembly domain is shown with both the shorthand notation and the actual sequence, followed by a single tyrosine that allows $A_{280\ nm}$ protein quantification (ε = 1290 $mol^{-1}cm^{-1}$).



Morphogenic Tag	$(LGG)_8$	$(IGG)_8$	$(HGG)_8$	$(WGG)_8$	$(YGG)_8$	$(YG)_8$	$(Y)_8$	$(FGG)_8$	$(FG)_8$	$(F)_8$	$(FGG)_8$	$(FGG)_8$
Length (pentamers)	160	160	160	160	160	160	160	160	160	160	80	40
Expected Length (bp)	2550	2550	2550	2550	2550	2526	2514	2550	2526	2514	1350	750

Figure S2. Genetically encoded amphiphile gene libraries. The amphiphile genes were run on a 1% agarose gel and stained with Sybr Safe. The left and right lanes represent a size standard ladder with the length (kb) shown on the left. The remaining lanes represent diagnostic digests of the constructs restricted with BamHI-HF and XbaI (hence appending 66 bp of flanking sequences to each band). The composition of the morphogenic domain and the expected length of the amphiphiles (shown in basepairs and elastin-like polypeptide pentamers) are displayed on the bottom.

Table S1. MALDI-MS of asymmetric amphiphiles.

Amphiphile	Expected MW (Da)[a]	Measured MW (Da)	Error (%)
A160	61636.7	61425.2	-0.3
A160-$(IGG)_8$	63454.8	63495.7	0.1
A160-$(LGG)_8$	63454.8	63434.1	-0.0
A160-$(HGG)_8$	63646.7	64147.3	0.8
A160-$(WGG)_8$	64039.3	n.d.	n.d.
A160-$(YGG)_8$	63855	63852.6	0.0
A160-$(YG)_8$	63398.5	63463.5	0.1
A160-Y_8	62942.1	63027.8	0.1
A160-$(FGG)_8$	63727	63760.9	0.1
A160-$(FG)_8$	63270.6	63319.6	0.1
A160-F_8	62814.1	n.d.	n.d.
A80-$(FGG)_8$	33212.4	33233.3	0.1
A40-$(FGG)_8$	17955.1	17959.9	0.0

a Expected molecular weights were determined by inputting the theoretical amino acid composition of the peptide into the exPASy Proteomics Server, which provides the sum of the isotopically averaged masses of each individual amino acid.
n.d. No data

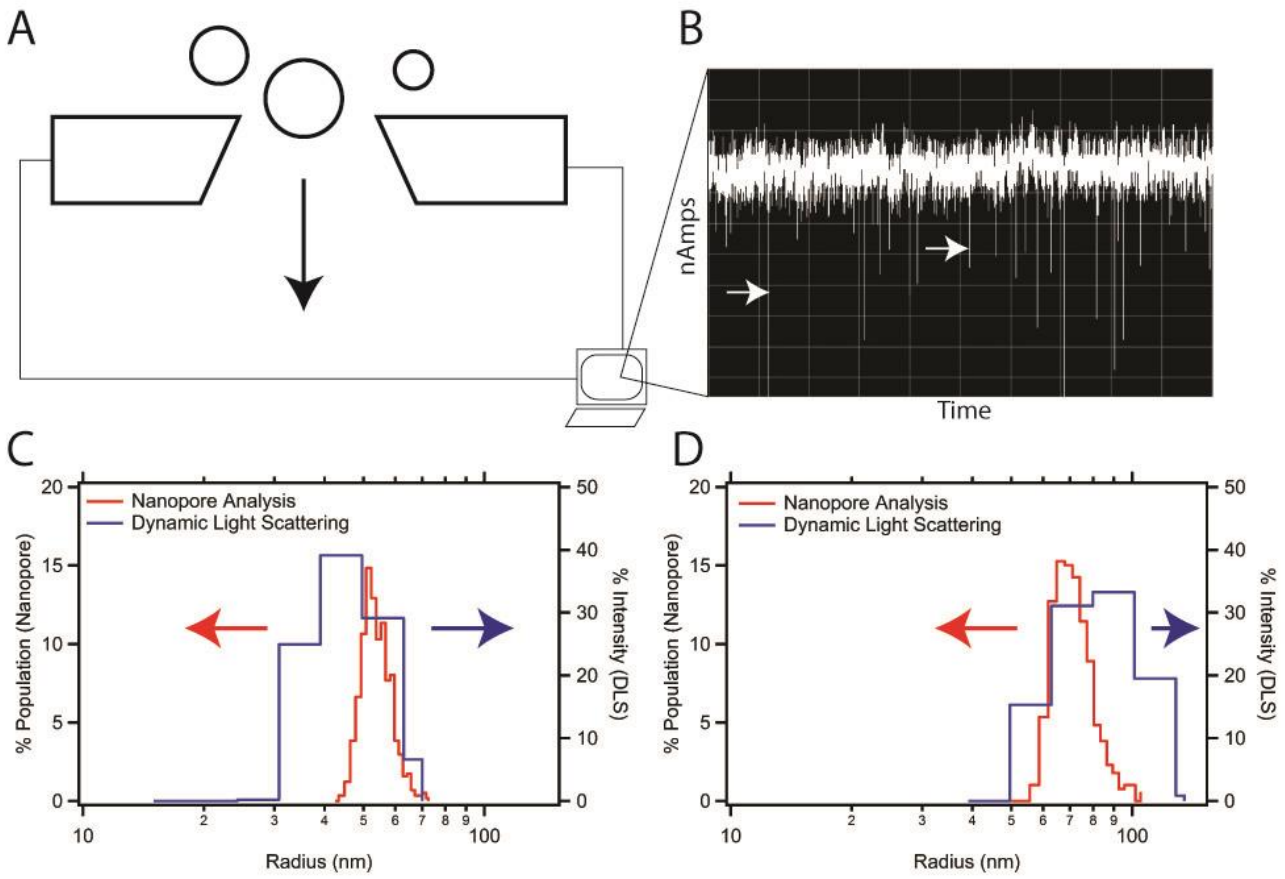


Figure S3. Nanoparticle size measured with tunable resistive pulse sensing (TRPS). (A-B) Pulsatile increases in measured resistance as nanoparticles flow through pores creates blockade events (white arrows) whose amplitude is proportional to the nanoparticle diameter. (C) A160-$(FGG)_8$ and (D) A160-$(YG)_8$ as measured with TRPS (red, left axis) and dynamic light scattering (blue, right axis).

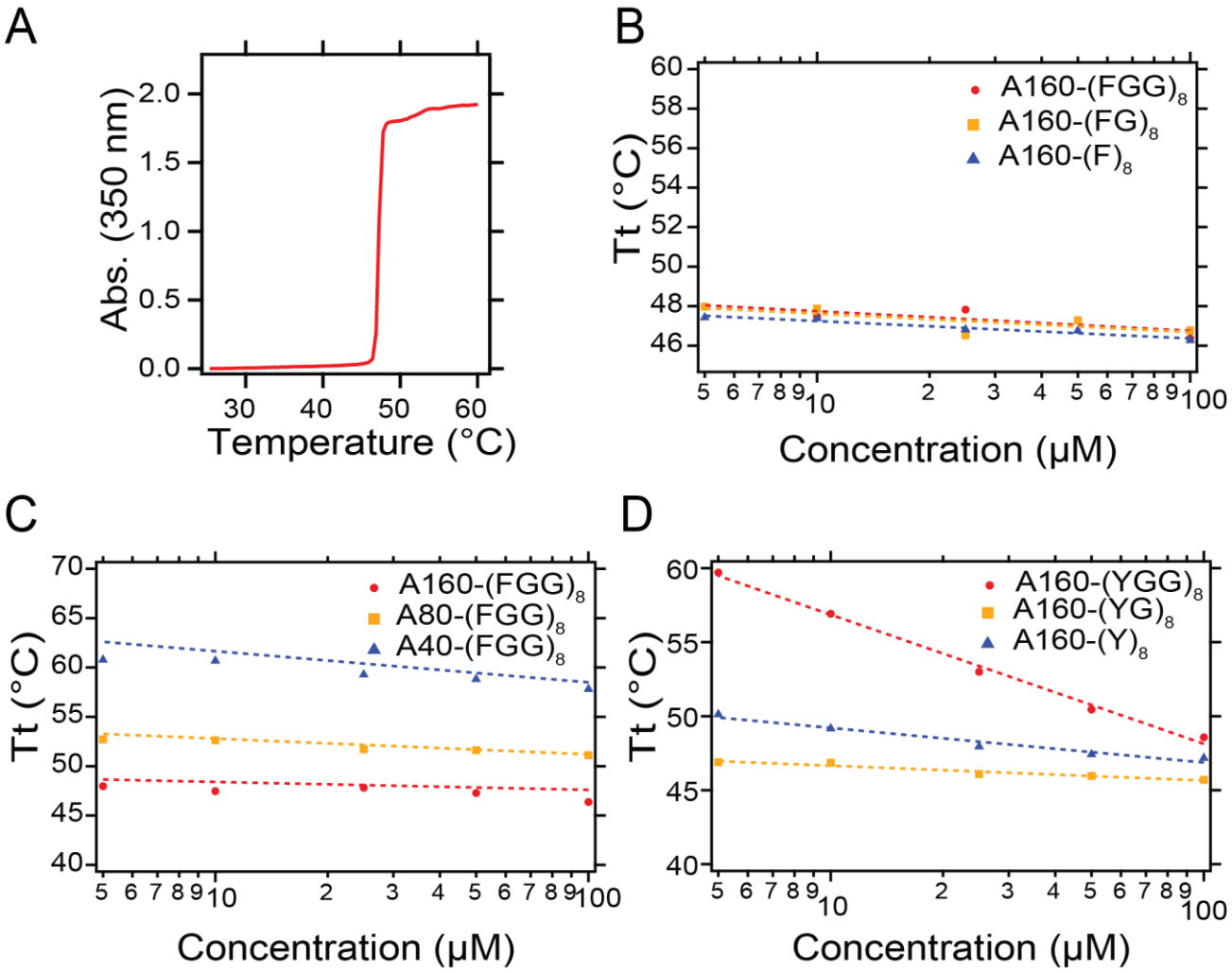


Figure S4. Thermal behavior of asymmetric amphiphiles. (A) A typical thermal turbidimetry profile, in this case for the amphiphile A160-(FGG)$_8$ at 50 μM. The transition temperature is defined as the temperature at which the first derivative reaches its maximum (here, 47.3 °C). (B) Transition temperature of the amphiphiles A160-(FGG)$_8$, A160-(FG)$_8$, and A160-(F)$_8$ plotted against concentration. Because all three amphiphile self-assemble, there is almost no concentration dependence. (C) Transition temperature of the amphiphiles A160-(FGG)$_8$, A80-(FGG)$_8$, and A40-(FGG)$_8$ plotted against concentration. Because the ELP chain length varies considerably between the three, each amphiphile transitions in a different temperature range, with A40-(FGG)$_8$ transitioning at the highest temperature and A160-(FGG)$_8$ transitioning at the lowest. (D) Transition temperature of the amphiphiles A160-(YGG)$_8$, A160-(YG)$_8$, and A160-(Y)$_8$ plotted against concentration. A160-(YGG)$_8$ does not self-assemble and displays a strong concentration dependence, whereas the other two amphiphiles do self-assemble, and transition independently of concentration.

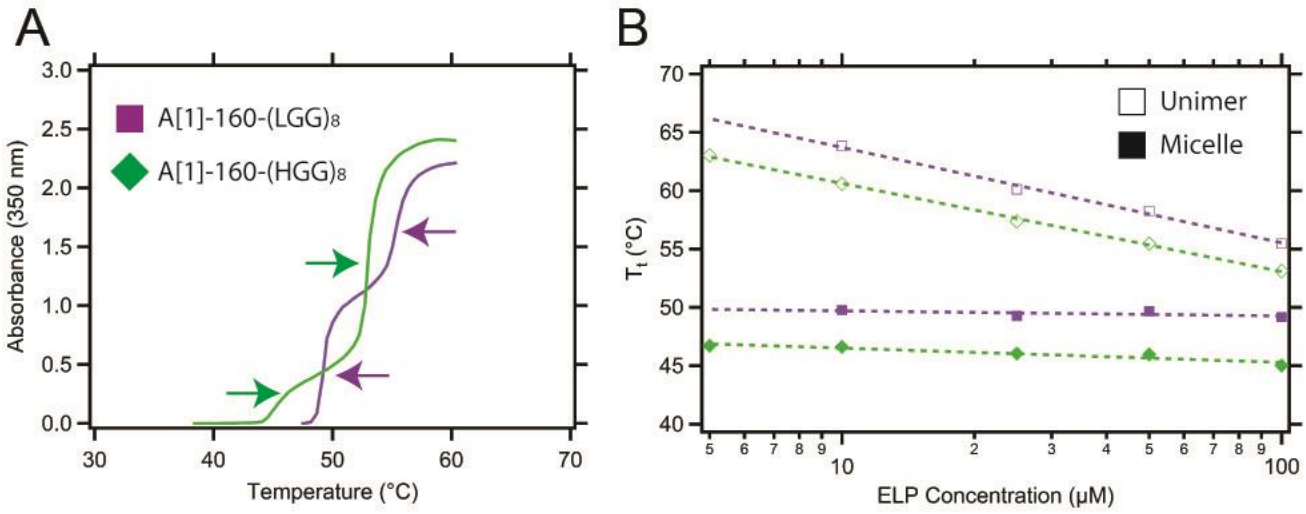


Figure S5. Thermal turbidimetry of asymmetric amphiphiles with minor self-assembling populations. (A) Thermal turbidimetry profile at 100 μM amphiphile concentration of A160-$(LGG)_8$ and A160-$(HGG)_8$, both of which remain primarily unimeric with a minor secondary population of self-assembled nanoparticles. Unlike a typical asymmetric amphiphile (see SI Figure 4A for thermal turbidimetry profile), both constructs display two separate transition temperatures, corresponding to the self-assembled and unimeric states. The first arrow in each curve indicates the transition of the self-assembled population from nanoparticle to collapsed aggregate. The second arrow indicates the transition of the unimeric population to a collapsed aggregate, which occurs at a higher temperature due to lower local concentration in the unimeric population than in the self-assembled population. (B) Transition behavior as a function of concentration for the A160-$(LGG)_8$ and A160-$(HGG)_8$. The first transition of the self-assembled population displays little concentration dependence, but the second transition of the unimeric population does depend on concentration.

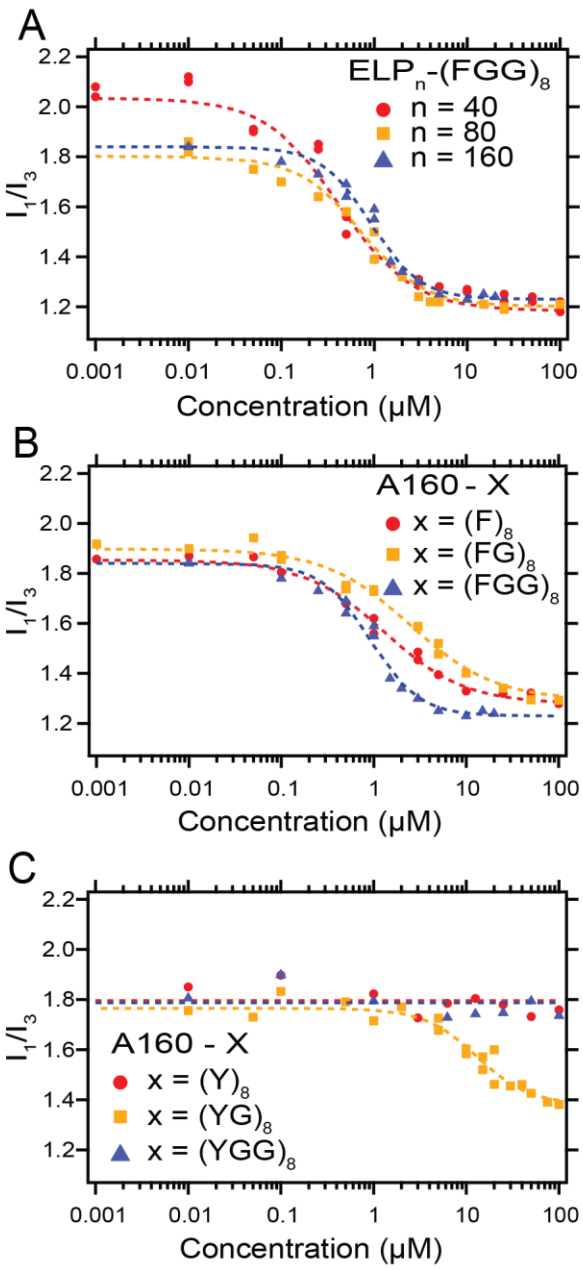


Figure S6. Pyrene analysis of asymmetric amphiphiles. (A) A40-$(FGG)_8$, A80-$(FGG)_8$, A160-$(FGG)_8$ self-assemble at CACs of 1.3 μM, 2.5 μM, and 0.9 μM and have similar minimal I_1/I_3 values of 1.28, 1.31, and 1.26 (Table 3). (B) A160-F_8, A160-$(FG)_8$, and A160-$(FGG)_8$ self-assemble at CACs of 1.3 μM, 2.5 μM, and 0.9 μM (Table 3) and have minimal I_1/I_3 values of 1.28, 1.31, and 1.26 (Table 3). (C) A160-$(YG)_8$ self-assembles at a CAC of 13 μM with a minimal I_1/I_3 value of 1.40 (Table 3). A160-$(YGG)_8$ and surprisingly A160-Y_8 do not provide a well-defined hydrophobic core for pyrene to partition into at any concentration.

Table S2. Amino acid and amphiphile molecular volumes and scattering length densities (SLDs).

Amino acid / sequence	Formula	Volume (nm^3)	SLD (10^{-4} nm^{-2})
Alanine (A)	C_3H_4DNO	0.0915	2.93
Serine (S)	$C_3H_3D_2NO_2$	0.0991	4.34
Lysine (K)	$C_6H_9D_4N_2O$	0.1762	3.26
Glycine (G)	C_2H_2DNO	0.0664	4.16
Proline (P)	C_5H_7NO	0.1293	1.72
Valine (V)	C_5H_8DNO	0.1417	1.78
Phenylalanine (F)	C_9H_8DNO	0.2034	2.54
Tyrosine (Y)	$C_9H_7D_2NO_2$	0.2036	3.34
SKGPG-$(AGVPG)_{160}$ (A160)	-	79.7854	2.62
$(FGG)_8$-Y	-	2.8932	3.19
$(YG)_8$-Y	-	2.3636	3.52

Analytical model for SANS

For a population of micelles composed of polymers, with isotropic orientation and interaction, the intensity can be approximated as:

$$I(q) = I_{inc} + {}^1N \int pdf(Z)\, P_{mic}(q,Z)dZ\ S(q,\ {}^1N) + I_{chains}(q) \qquad (1)$$

where I_{inc} is a flat background due to incoherent scattering, ${}^1N$ is the number density of micelles (which was determined from the composition of the ELP and the distribution of aggregation numbers Z), *pdf* is the log-normal probability density function applied to the aggregation number Z, P_{mic} is the unnormalized form factor describing the shape, size and scattering length density (SLD) of micelles, *S* is the structure factor accounting for interactions between micelles – here applied in the monodisperse spherical approximation. I_{chains} is the intensity arising from the scattering by polymer chains, a contribution which can be considered independent from the scattering of the micelles due to the very different length scales. The log-normal probability density function is given as:

$$pdf(Z) = \frac{1}{Z\,\sigma\,\sqrt{2\,\pi}} \exp\left[-\frac{(\ln x-\mu)^2}{2\,\sigma^2}\right],\ \mu = \ln\left(\frac{N_{agg}}{\sqrt{1+s^2}}\right), \sigma = \sqrt{\ln(1+s^2)}, \qquad (2)$$

The micelle form factor at q=0 is proportional to the average volume and contrast of the micelles: $P_{mic}(0) = (\overline{V}\ \overline{\Delta SLD})^2$. The length *L* of the cylinder is fitted, and the radius *R* is distributed according to *pdf(Z)*: $R = \sqrt{\frac{Z\,v_{molec}}{(1-h)\pi L}}$, where v_{molec} is the molecular volume of one amphiphile and h the extent of hydration in this cylinder

(volume fraction of water). Based on cryo-TEM data, we chose a model of isotropically oriented homogeneous cylindrical micelles, the form factor of which is expressed as [1]:

$$P_{cyl}(q,Z) = P_{mic}(0)\ \int_0^1 \frac{J_1\left(qR\sqrt{1-x^2}\right)}{qR\sqrt{1-x^2}}\, 2\, j_0\left(\frac{qLx}{2}\right) dx \qquad (3)$$

where R depends on Z, J_1 is the cylindrical Bessel function of the first order and j_0 the spherical Bessel function of zeroth order. The scattering by polymer chains was assumed Gaussian as this model leads to a q^{-2} power law, as observed in the data. The corresponding equation derived by Debye [2] is:

$$I_{chains}(q) = 2\, I_{chains}(0)\ \left(\frac{e^{-q\,R_g}}{(q\,R_g)^2} + \frac{1}{q\,R_g} - \frac{1}{(q\,R_g)^2}\right) \qquad (4)$$

Repulsions between micelles are accounted for with the hard sphere structure factor solution of the Ornstein-Zernike equation with the Percus-Yevick closure relation [3].

$$\alpha = \left[\frac{1+2\,\varphi_{HS}}{(1-\varphi_{HS})^2}\right]^2, \beta = -6\,\varphi_{HS}\ \left[\frac{1+\varphi_{HS}/2}{(1-\varphi_{HS})^2}\right]^2, \gamma = \frac{\alpha\,\varphi_{HS}}{2}, x = 2\,q\,\sqrt[3]{\frac{3}{4\pi}\frac{\varphi_{HS}}{{}^1N}},$$

$$g = \alpha\, j_1(x) + \frac{\beta}{x^3}\left[2\,x\sin x + (2-x^2)\cos x - 2\right] + \frac{\gamma}{x^5}\left[-x^4\cos x + 4\left((3\,x^2-6)\cos x + (x^3-6\,x)\sin x + 6\right)\right],$$

$$S\left(q,\ {}^1N\right) = \left[1 + 24\,\varphi_{HS}\frac{g}{q}\right]^{-1} \qquad (5)$$

As we know the concentration of amphiphiles, their molecular volumes and contrast, $I_{chains}(0)$ can be fixed, while the mean aggregation number provides ${}^1N$ (used for both the form factor and structure factor) and $P_{mic}(0)$. To get a robust set of output parameters, we fitted simultaneously all spectra from a given amphiphile, with a common mean aggregation number and standard deviation, cylinder length, hydration of the micelles and gyration radius of the polymer chains. The only free parameters are the incoherent backgrounds and the hard sphere volume fractions.

The gyration radius of polymer chains is 11 nm. The hydration of micelles is 0.94, and the aggregation numbers are respectively 231 for A160-$(FGG)_8$ and 139 for A160-$(YG)_8$. The cylinders lengths are 174 and 164 nm with mean radii of 23.5 and 21.2 nm, respectively.

The obtained concentration-dependent data were also used to determine the experimental effective structure factor. Spectra from samples at the three highest concentrations were divided with the scattering function of the sample with lowest concentration (53.1 µM and 48.3 µm for A160-$(FGG)_8$ and A160-$(YG)_8$, respectively). Given the approximation that at the lowest concentration for each amphiphile inter-particle interactions are weak and the structure factor has no effect, then only the form factor is seen at that concentration. Thus dividing the other spectra by the lowest concentration function and normalizing for concentration yields the experimental structure factor. The obtained data for the three higher concentrations are given in fig. S8. These experimental structure factors were fitted with the hard sphere structure factor given by eq. 5. The relatively strong agreement between the experimental structure factor and the theoretical hard sphere structure factor is both verification that this is the appropriate structure factor as well as an independent indication that the morphology (radius and length) of the micelles remains constant upon variation of the concentration.

Table S3. SANS fit parameters for simultaneous fits

Amphiphile	Conc. (μM)	N_{agg}	L (nm)	h	V_{cyl} (nm^3)	R_{cyl} (nm)	ϕ_{HS}	R_g (nm)
A160-$(FGG)_8$	79.0						11.3 %	
A160-$(FGG)_8$	53.1						8.5 %	
A160-$(FGG)_8$	221	231	174	0.94	302 669	23.5	19.9 %	10.7
A160-$(FGG)_8$	300						24.3 %	
A160-(FGG)8	523						31.2 %	
A160-$(YG)_8$	79.0						1.9 %	
A160-$(YG)_8$	48.3						11.1 %	
A160-$(YG)_8$	205	139	164	0.95	231 233	21.2	21.9 %	11.0
A160-$(YG)_8$	296						25.4 %	
A160-$(YG)_8$	488						32.0 %	

h. Hydration by volume fraction.
ϕ_{HS}. Excluded volume fraction.

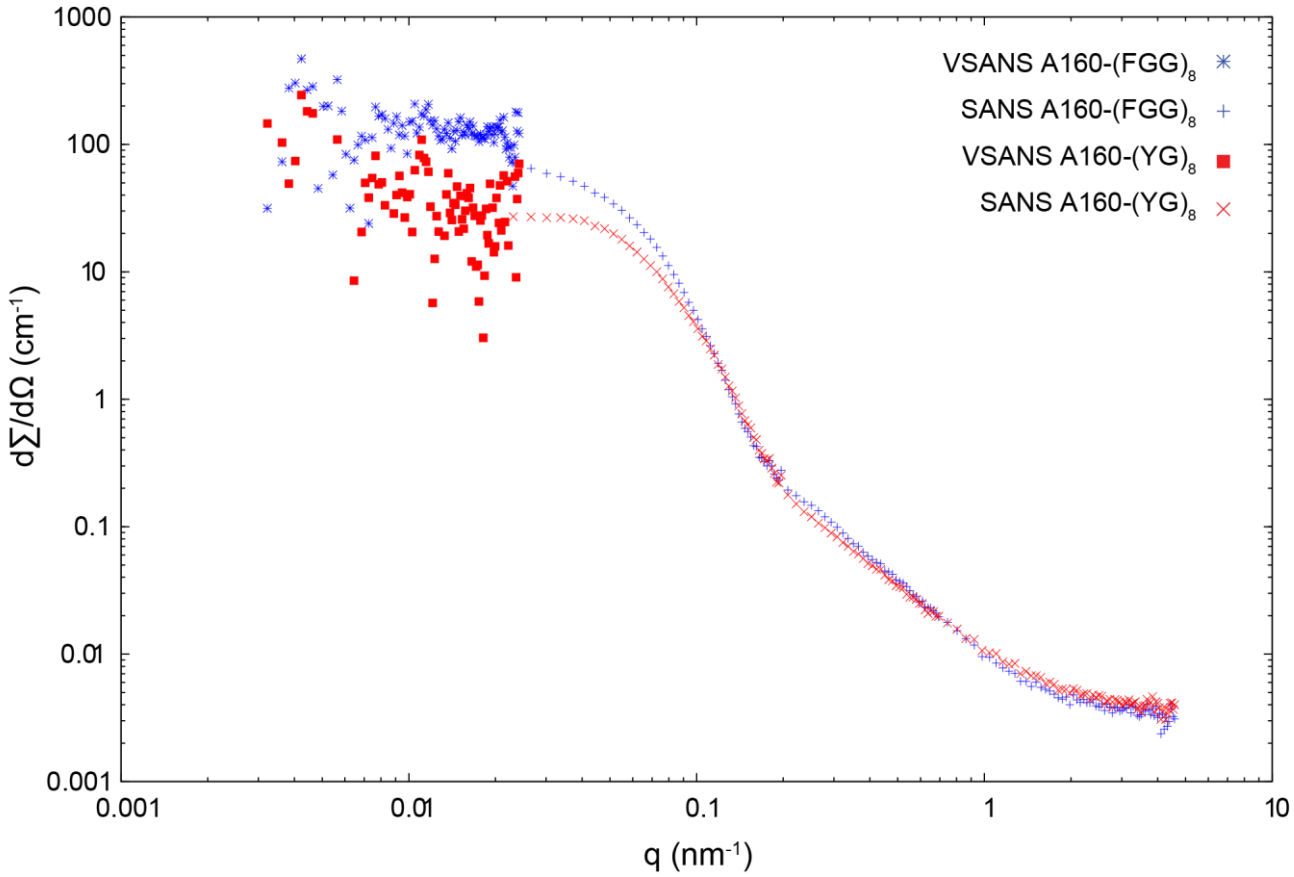


Figure S7. VSANS (very small angle neutron scattering) spectra were acquired for both amphiphiles at 0.5 wt% to confirm that the longest dimension is not significantly longer than 150 nm. Low scattering intensity resulted in a very high signal to noise ratio, making the data not suitable for analytical modeling. However, the absence of a peak in the VSANS region and the constancy of the intensity of either spectrum indicates that neither sample contains aggregates with a dimension significantly larger than 150 nm.

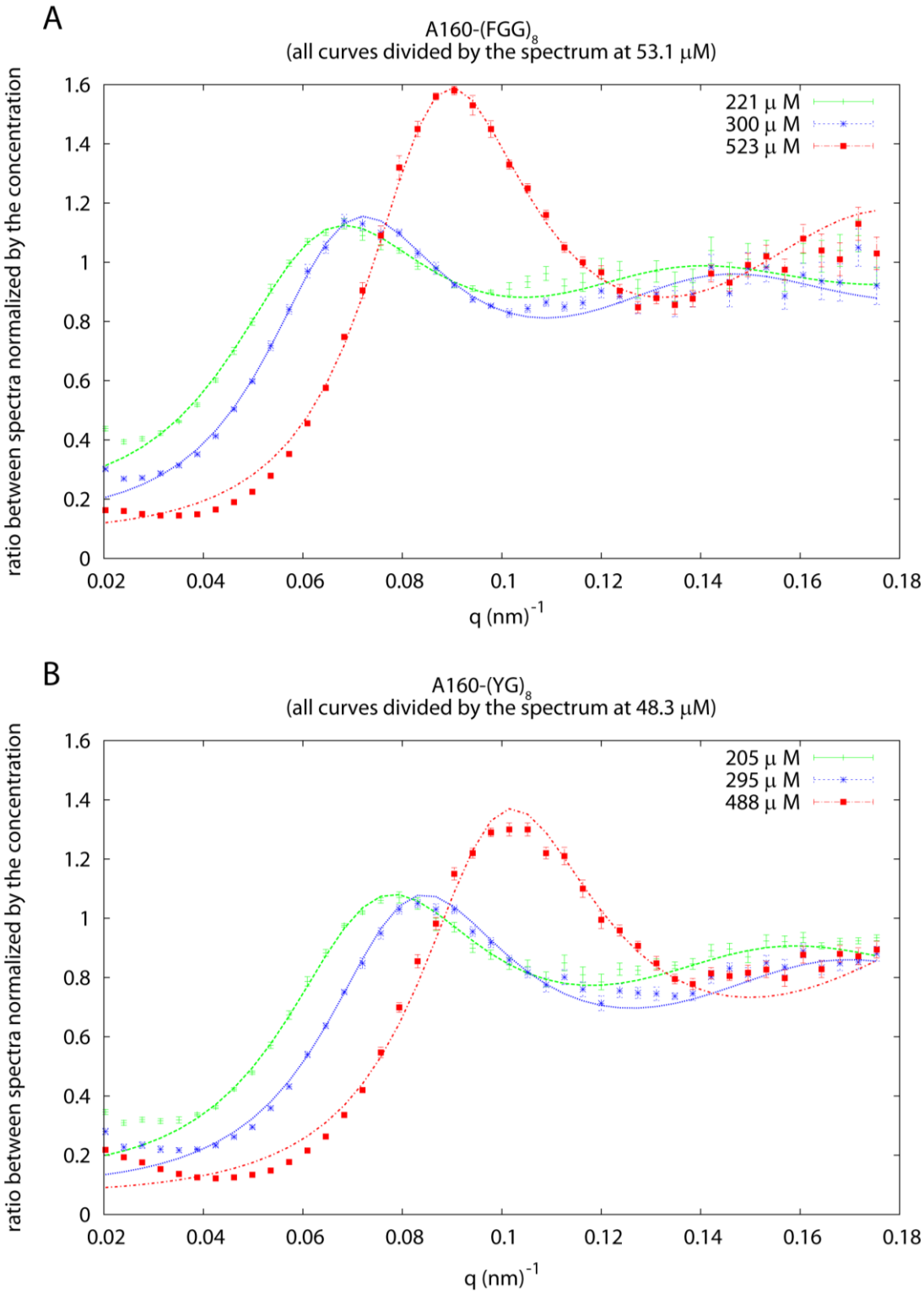


Figure S8. Effective structure factor obtained by dividing the scattering curves of the differently concentrated samples by dividing with the scattering of the most dilute sample and correcting for the concentration difference between the two samples and after subtracting the corresponding incoherent background. a) A_{160}-$(FGG)_8$; b) A_{160}-$(YG)_8$

References

1. Fournet, A.G. J. Polym. Sci., 1956, 19: p. 594-594.
2. Debye, P. J. Phys. Colloid Chem., 1947. 51: p. 18-32.
3. Wertheim, M.S. Phys. Rev. Lett., 1963. 10: p. 321-323.