CHEMBIOCHEM

Supporting Information

Physicochemical Characterization of Polymer-Stabilized Coacervate Protocells

N. Amy Yewdall[+], Bastiaan C. Buddingh[+], Wiggert J. Altenburg, Suzanne B. P. E. Timmermans, Daan F. M. Vervoort, Loai K. E. A. Abdelmohsen, Alexander F. Mason,* and Jan C. M. van Hest*[a]

cbic_201900195_sm_miscellaneous_information.pdf

Supplementary information

1 Materials

All chemicals were used as received unless otherwise stated. For the synthesis of terpolymer: monomethoxy poly(ethylene glycol) 2 kDa was purchased from Rapp Polymere, trimethylene carbonate was purchased from TCI Europe. For the preparation of modified amylose derivatives: amylose (12-16 kDa) was supplied by Carbosynth and 3-chloro-2-hydroxypropyltrimethyl ammonium chloride (65 wt% in water) was supplied by TCI Europe. All other chemicals and reagents were supplied by Sigma-Aldrich.

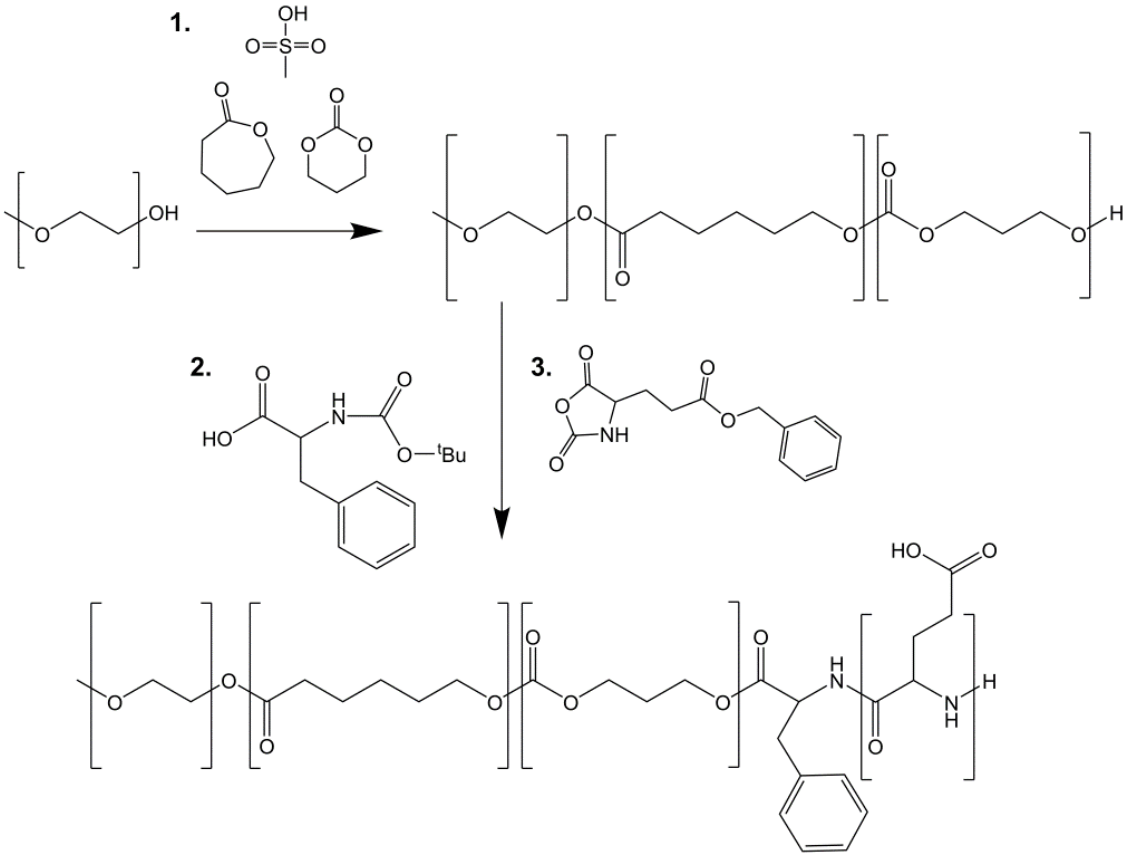
2 Supplementary Methods

PEG-*b*-PCLgPTMC-*b*-PGA terpolymer[1]

Step 1 - Preparation of poly(ethylene glycol)-poly(caprolactone-*gradient*-trimethylene carbonate) (PEG-PCl$_g$TMC). According to a modified literature procedure the organocatalyzed ring-opening polymerization of ε-caprolactone and trimethylene carbonate was performed, aiming for a composition of PEG_{44}-PCl_{50}-*g*-TMC_{50}.[2] Monomethoxy-PEG-OH macroinitiator (2 kDa, 0.2 mmol, 400 mg) was weighed into a round- bottomed flask along ε-caprolactone (ε-CL, 10 mmol, 1108 μL) and trimethylene carbonate (TMC, 10 mmol, 1121 mg) and dried *via* azeotropic evaporation of added toluene (x3). The dried reagents were then re-dissolved in dry toluene (20 mL) and methanesulfonic acid (0.6 mmol, 39 μL) was added, under argon. The reaction mixture was stirred at 30 °C for 6 hrs, after which time it was precipitated into ice cold methanol. This yielded 2.1 g of a waxy solid (85 % yield) and the composition of the resulting copolymer was confirmed by ^{1}H NMR (measured composition was PCL_{51}, $PTMC_{54}$) , comparing the protons of PEG (3.65-3.7 ppm), terminal methyl unit (singlet at 3.40 ppm) to PCL CH_2 (multiplet at 2.40-2.25 ppm) and PTMC CH_2 (multiplet at 2.2-1.8 ppm). GPC analysis (using a PL gel 5 μm mixed D column, with THF and PS standards) yielded a Đ of 1.1.

Step 2 - Chain-end modification with Boc-L-phenylalanine and deprotection. For the incorporation of a terminal amine onto the diblock copolymer, we performed a carbodiimide-mediated esterification between PEG-PCl*g*TMC and Boc-L-phenylalanine (Boc-L-Phe). 1.3 g (*ca*. 0.1 mmol) of PEG-PCl*g*TMC copolymer was dissolved in acetonitrile and to it was added dicyclohexylcarbodiimide (0.2 mmol, 41.9 mg), Boc-L-Phe (0.5 mmol, 23.4 mg), and 4-dimethylaminopyridine (0.02 mmol, 2.5 mg) on ice. The reaction mixture was stirred for 24 hrs at RT and afterwards placed directly in the freezer overnight to facilitate precipitation of dicyclohexylurea by-product. After cold filtration of the reaction mixture it was concentrated and then precipitated into cold methanol. ^{1}H NMR was used to check the product to confirm removal of the terminal TMC signal at 1.92 ppm and emergence of Boc protons at 1.45 ppm before deprotection. The resulting copolymer was then dissolved in 5 mL DCM, to which 5 mL of trifluoro acetic acid (TFA) was added (on ice) and the mixture was allowed to warm to RT and stirred for 2 hr. After 2 hr the solvent was evaporated and the copolymer was washed with $NaHCO_3$, 1 M NaCl and brine before drying on $MgSO_4$, filtration and evaporation of the majority of the solvent so that the copolymer could be precipitated from ice cold methanol. ^{1}H NMR clearly showed that the signal arising from the terminal TMC group had disappeared, due to addition of phenylalanine at the terminus, and aromatic protons were visible at around 7.2-7.3 ppm. GPC analysis before and after deprotection yielded a Đ of 1.1, indicating that TFA treatment did not facilitate copolymer hydrolysis.

Step 3 - Polymerization and deprotection of *N*-carboxyanhydride γ-benzyl L-glutamate (NCA-BLG). Following a published method for the controlled polymerization of NCA-BLG by lowering the temperature and maintaining a constant flow of N_2 in order to remove CO_2 by-product,[3] we aimed to add between 8 and 10 PBLG units onto the amine-terminus of the copolymer chains. 1 g of Phe-terminated copolymer was weighed into a Schlenk flask and dissolved with *ca*. 3 mL of dry DMF and cooled in an ice bath. To the cooled mixture 160 mg of NCA-BLG was added under Ar and the reaction was left under a constant flow of N_2 for 24 hrs. The product was precipitated into cold methanol and analysed by ^{1}H NMR to confirm the overall composition and, in particular, the presence of benzylic and aromatic protons at 5.0-5.2 and 7.1-7.4 ppm, respectively. Benzyl-protected terpolymer was dissolved in 10 mL THF and 10 mL of methanol was added before applying the H-Cube at 60 °C with 30 bar of H_2 pressure and a flow rate of 1 mL/min to facilitate removal of benzyl protecting groups. The product was concentrated then precipitated into cold ether, dissolved in dioxane before lyophilisation to yield a waxy solid, 0.9 g (85 % yield). ^{1}H NMR was used to confirm successful deprotection of the PBLG units and GPC data indicated that the polydispersity didn't increase beyond 1.1 during this process.



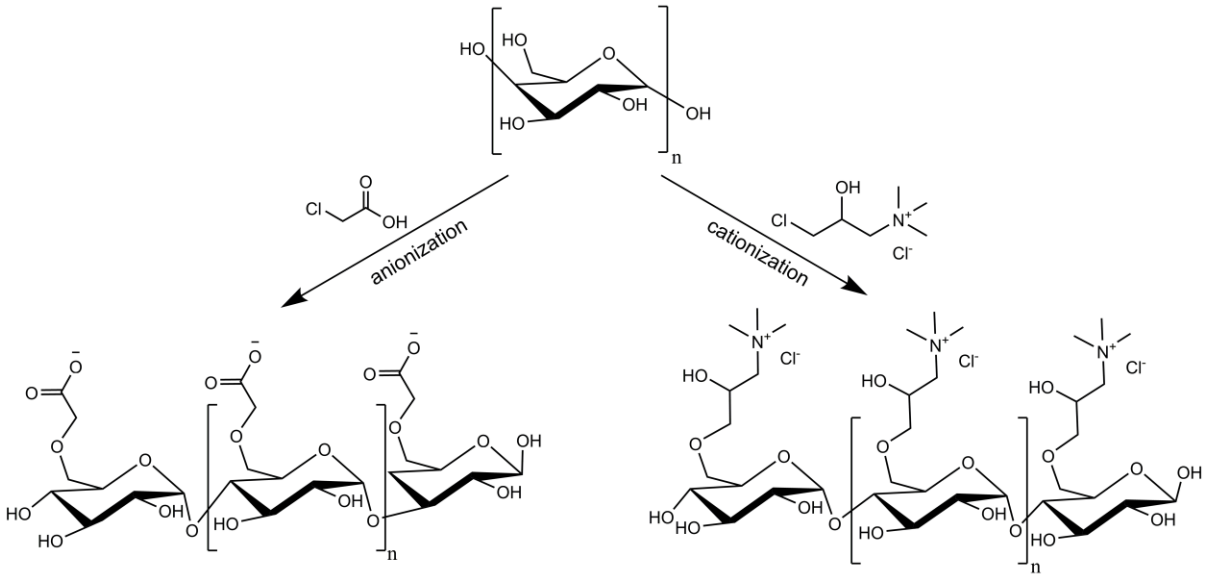
Scheme 1: Synthesis of PEG-*b*-PCL*g*PTMC-*b*-PGA terpolymer. Poly(ethylene glycol) monomethyl ether was used to initiate the ring opening polymerisation of ε-caprolactone and trimethylene carbonate (**step 1**). The terminal alcohol of this polymer was subsequently modified *via* a Steglich esterification with Boc-L-Phe-OH to yield a primary amine after TFA deprotection (**step 2**). The final poly(L-glutamic acid) block was introduced by the ring opening polymerisation of *N*-carboxyanhydride γ-benzyl L-glutamate, followed by hydrogenation (**step 3**).

Modified amylose (Q-Am and CM-Am)

Both quaternized (Q-Am) and carboxymethylated (CM-Am) were prepared in accordance with previously published procedures.[1]

Q-Am was prepared by dissolving 1.5g of amylose and 2.78g of NaOH in 14.25 mL of Milli-Q at 35 °C. After complete dissolution of the amylose, 11.64 mL of 3-chloro-2-hydroxypropyltrimethylammonium chloride solution (60 wt% in water) was added dropwise into the stirring reaction mixture, which was subsequently left to react overnight. After this time, the mixture was neutralized with acetic acid and precipitated into 200 mL of cold ethanol. The resulting precipitate was re-dissolved in Milli-Q water and dialysed extensively against water using regenerated cellulose dialysis tubing (Spectrum Labs, USA) with a 3.5 kDa MWCO before lyophilization. Around 5 g of Q-Am was obtained from this reaction (*ca*. 80 % yield). ^{1}H NMR characterization data are presented in Figure S1a.

CM-Am was prepared by dissolving 1.5g of amylose and 3.6 g of NaOH in 15 mL of Milli-Q at 70 °C. After complete dissolution of the amylose, 2.7 g of chloroacetic acid was added and the reaction mixture was left to stir for 2 hours. After the reaction, the mixture was neutralized with acetic acid and precipitated into 200 mL of cold ethanol. The resulting precipitate was re-dissolved in Milli-Q water and dialysed extensively against water using regenerated cellulose dialysis tubing (Spectrum Labs, USA) with a 3.5 kDa MWCO before lyophilization. Around 3 g of CM-Am was obtained from this reaction (*ca*. 90 % yield). ^{1}H NMR characterization data are presented in Figure S1b.



Scheme 2: Modification of amylose with either chloroacetic acid or 3-chloro-2-hydroxypropyltrimethylammonium chloride solution to form CM-Am and Q-Am, respectively.

DNA Molecular Biology.

All DNA was ordered through Integrated DNA Technologies (IDT). The constructs were optimized using the IDT Codon Optimization Tool for *Escherichia coli (E. coli)* with the amino acid sequence as input and ordered as gBlock Gene Fragments (IDT). The vector pET-28a and the gBlocks were both digested using the restriction enzymes, NcoI and XhoI (New England Biolabs). The digested vector was purified using the QIAquick Gel Extraction Kit (Qiagen). The gBlocks were purified by QIAquick PCR Purification Kit (Qiagen). Both purified components were ligated, 1:3 ratio of vector to insert, using T4 Ligase (New England Biolabs). XL-1 Blue bacterial cells (Agilent Technologies) were transformed with the ligation product for amplification and extracted using the QIAprep Spin Miniprep Kit (Qiagen). The constructs were verified using Sanger sequencing (BaseClear).

Protein Expression and Purification

The expression and purification protocol was based on previous reports.[4] Constructs were transformed into BL21(DE3) bacterial cells (Novagen). Expressions were carried out using lysogeny broth (LB) supplemented with kanamycin (100 μg ml^{-1}). 600 mL 2YT media was inoculated using an overnight culture grown at 37°C, 200 r.p.m. to an optical density (OD_{600}) of 0.6 and then induced with isopropyl β-D-1-thiogalactoyranoside (IPTG) (0.1mM final concentration). The proteins were expressed at 20°C, 150 r.p.m for 18 h. The cells were harvested by centrifugation (2700 *xg*, 15 min, 4°C) and the pellet were either lysed immediately or flash frozen in liquid N_2 and stored at -20 °C. The

cell pellets were resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0; for the GFP(+36) variant, 1 M NaCl was used) and lysed using an EmulsiFlexC3 High-Pressure homogenizer (Avestin) at 15000 psi for three rounds. For GFP(+36), 10 mg mL^{-1} DNaseI (PanReac AppliChem) and 5 mg mL^{-1} RNase A (ThermoFisher) were added 30 min before lysis. Cell debris and insoluble proteins were removed by centrifugation (15.000 xg, 10 min, 4°C). The His-tagged proteins were purified from the soluble lysate using Ni-NTA affinity chromatography (His-Bind Resin, Novagen). The lysates were incubated with the His-bind resin for 1.5 h at 4°C on a shaking table, prior to loading onto an empty gravity flow column (Bio-Rad). The resin was washed twice with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0; again for the GFP(+36) variant, 1 M NaCl was used). The His-tagged proteins were eluted from the resin using elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0, for the GFP(+36) variant, 1 M NaCl was used). Extensive dialysis was performed using a 10 kDa molecular weight cut off (MWCO) membrane (Millipore), into PBS, pH 7.4 (for the superfolder GFP(-8) and GFP(-30) variants) or 50 mM NaPi, 600 mM NaCl, pH: 7.5 (for the GFP(+36) construct. Within the first 2 h, the solvent was replaced twice. The samples were concentrated using 3 kDa MWCO Amicon Ultra Filters (Millipore) and were further purified by size-exclusion chromatography. The samples were passed through a HiLoad Superdex 200 pg preparative column (GE Healthcare) connected to an ÄKTApurifier FPLC (GE Healthcare) at a flow rate of 1 mL min^{-1}, and the absorbance at 280 and 488nm were both monitored. The eluted fractions were analyzed using SDS-PAGE (4-20% Mini-PROTEAN TGX Precast Protein Gel, Bio-Rad) and the most pure fractions were pooled and concentrated. Protein concentrations were calculated using the 280 nm absorbance determined by the ND-1000 spectrophotometer (Thermo Scientific) and theoretical extinction coefficients of 20400 M^{-1} cm^{-1} used as determined by the online ProtParam tool (ExPASy). The samples were aliquoted for single use, flash frozen in liquid N_2 and stored at -80 °C. The purity of the samples were ascertained by using both SDS-PAGE as well as liquid chromatography quadrupole time of flight mass spectrometry (LC-MS Q-TOF) (**Figure S3 and S4**).

LC-MS Q-TOF

The exact molecular weight of the purified GFP variants were determined using a high resolution LC-MS system consisting of an ACQUITY UPLC I-Class system coupled to a Xevo G2 QTOF (Waters). The protein was separated (0.3 mL min^{-1}) on a Polaris C18A reverse phase column 2.0 x 100 mm (Agilent) using a 15 % to 75 % acetonitrile gradient in water supplemented with 0.1 % v/v formic acid before analysis in positive mode in the mass spectrometer. Deconvolution of the m/z spectra was done using the MaxENTI algorithm in the Masslynx v4.1 (SCN862) software.

Succinylated HRP uptake into the coacervate phase

HRP proteins, succinylated and non-succinylated, (5 µL of 5.6 µM solutions) were either mixed with 45 µL of PBS, or 30 µL Q-amylose (1 mg/mL) and 15 µL Cm-amylose (1 mg/mL) for the coacervate containing sample. Both samples with (+, Figure S9) and without coacervates (-, Figure S9) were hard spun at 21000 xg for 10 mins, and the top 7 µL (“polymer poor phase’’) was gently pipetted off and mixed with 2X gel loading dye. Samples were boiled at 100 °C for 10 min and loaded, with a Precision Plus Protein™ molecular weight ladder (BioRad), onto a 10 well 4–20% Mini-PROTEAN® TGX™ Precast Protein Gel (BioRad). The gel was run according to manufacturer’s instructions. Proteins were stained with Bio-Safe™ Coomassie G-250 stain according to manufacturer’s protocol.

3 Supplementary Tables and Figures

Table S1 Coacervate formation

Formation of coacervates at varying Q-Am:CM-Am ratios, using $DS_Q = DS_{CM} = 0.8$

Q-Am:CM-Am	1:10	1:3	1:2	1:1	2:1	3:1	10:1
Coacervate formation	no	no	no	yes	yes	Yes, smaller	no

Figure S1 Amylose characterisation

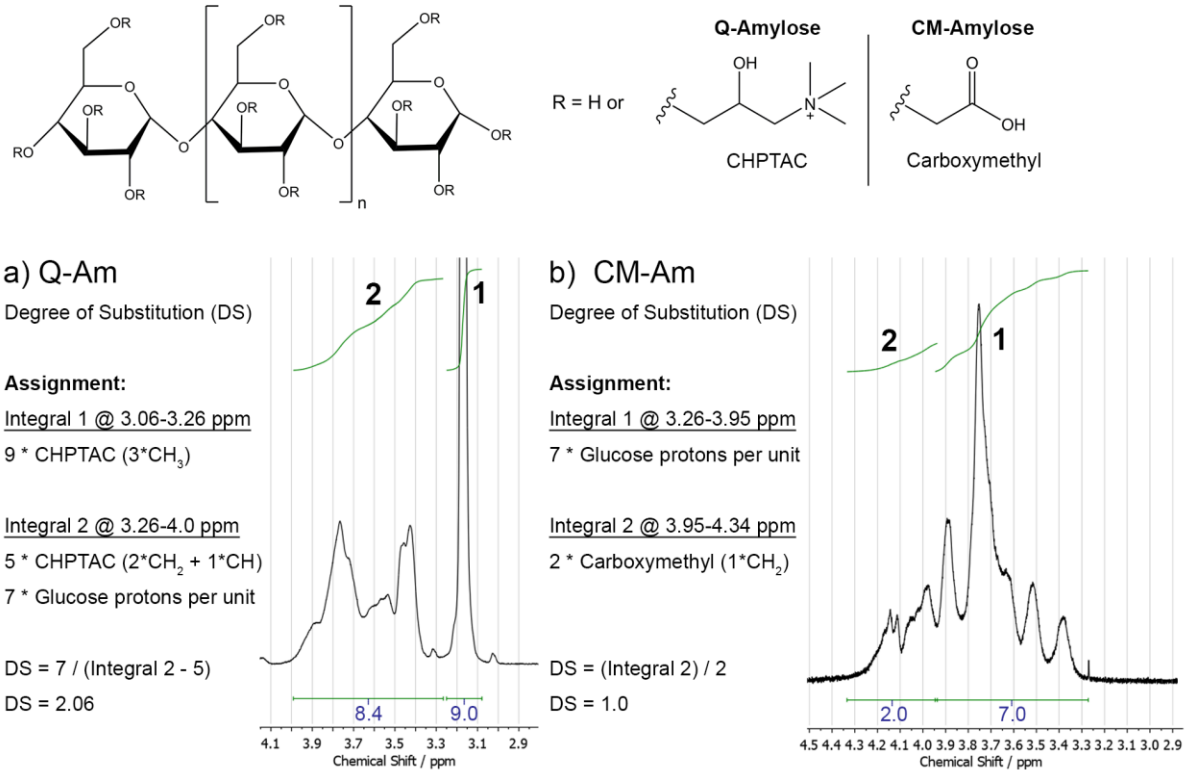


Figure S1. ^{1}H NMR processing of highly charged amylose derivatives, including calculation of the degree of substitution (DS = number of modifying groups per glucose unit), of both quaternized (Q-Am) and carboxymethylated (CM-Am) products.

Figure S2 GFP sequence alignment

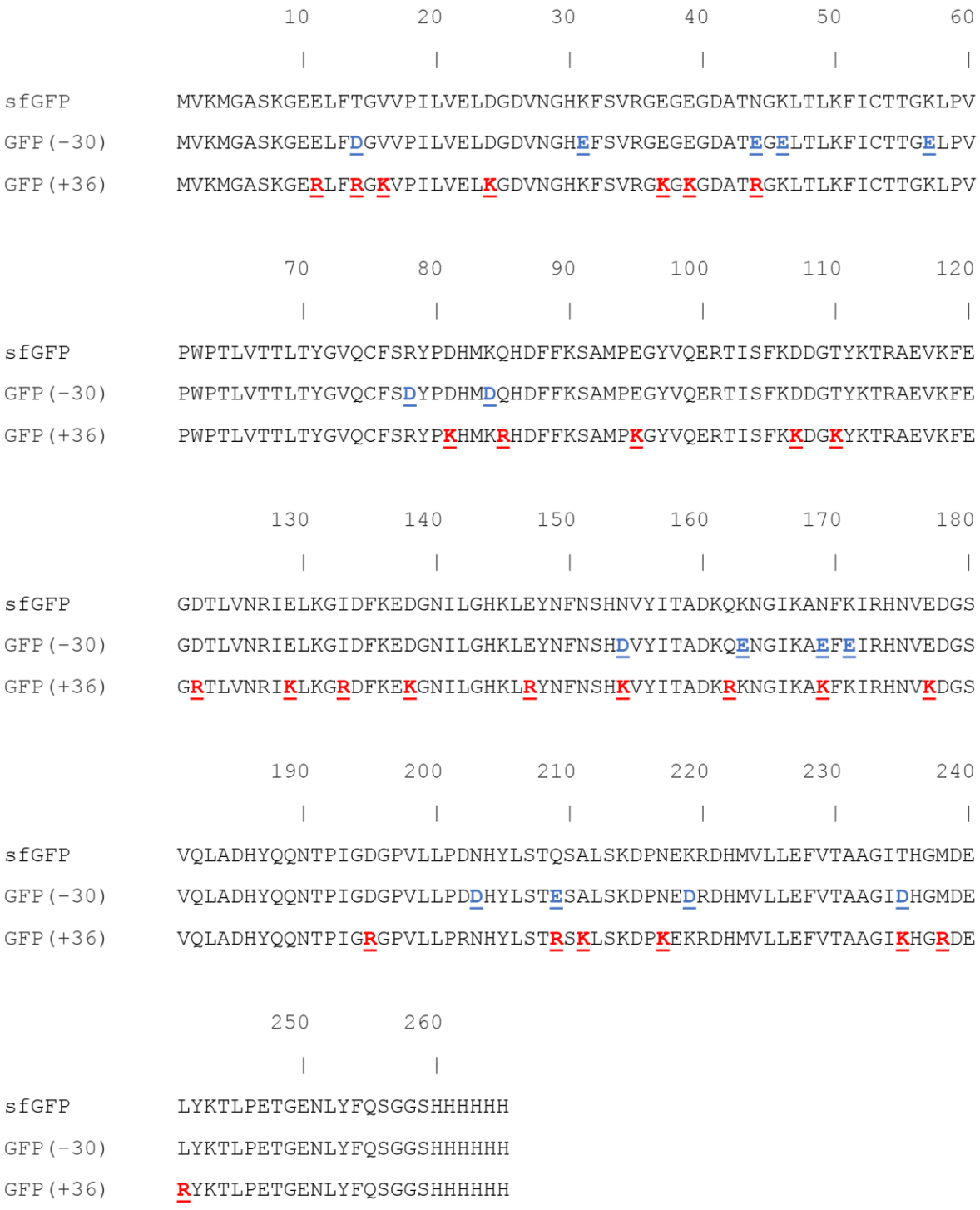


Figure S2. Sequence alignment of the GFP variants with superfolder GFP(-8) as reference was performed using the online tool CLUSTALW (Rhone-Alpes Bioinformatic Pole Gerland). Negatively charged amino acids substitutions are highlighted in **blue**, positively charged amino acids substitutions in **red.**

Figure S3 SDS-PAGE of purified GFP variants

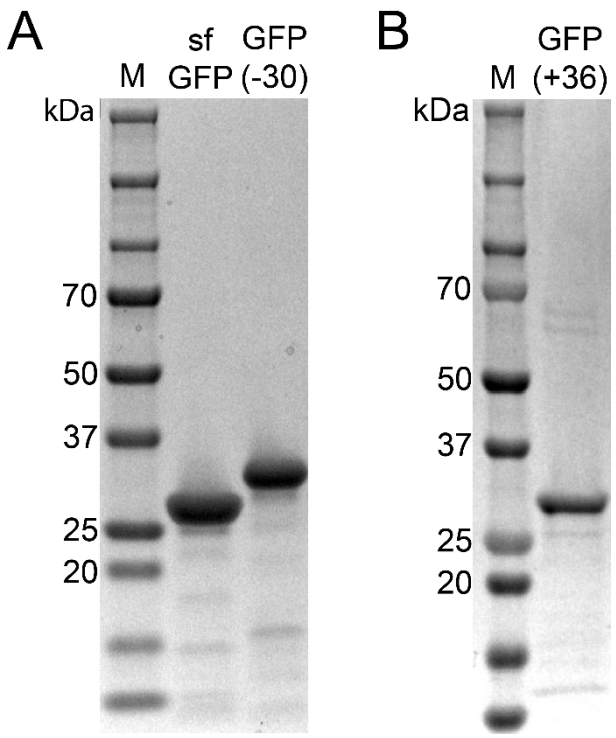


Figure S3. Analysis of the purified GFP variants by SDS-PAGE gel electrophoresis (4-20% Mini-PROTEAN TGX Precast Protein Gel, Bio-Rad). A) Left to right: Marker (Precision Plus Protein™ Dual Color Standards, Bio-Rad), superfolder GFP (sfGFP, -8), GFP(-30). B) Left to right: Marker, GFP(+36). For all three GFP variants only one prominent band is observed at an anticipated height.

Figure S4 LC-MS analysis of purified GFP mutants

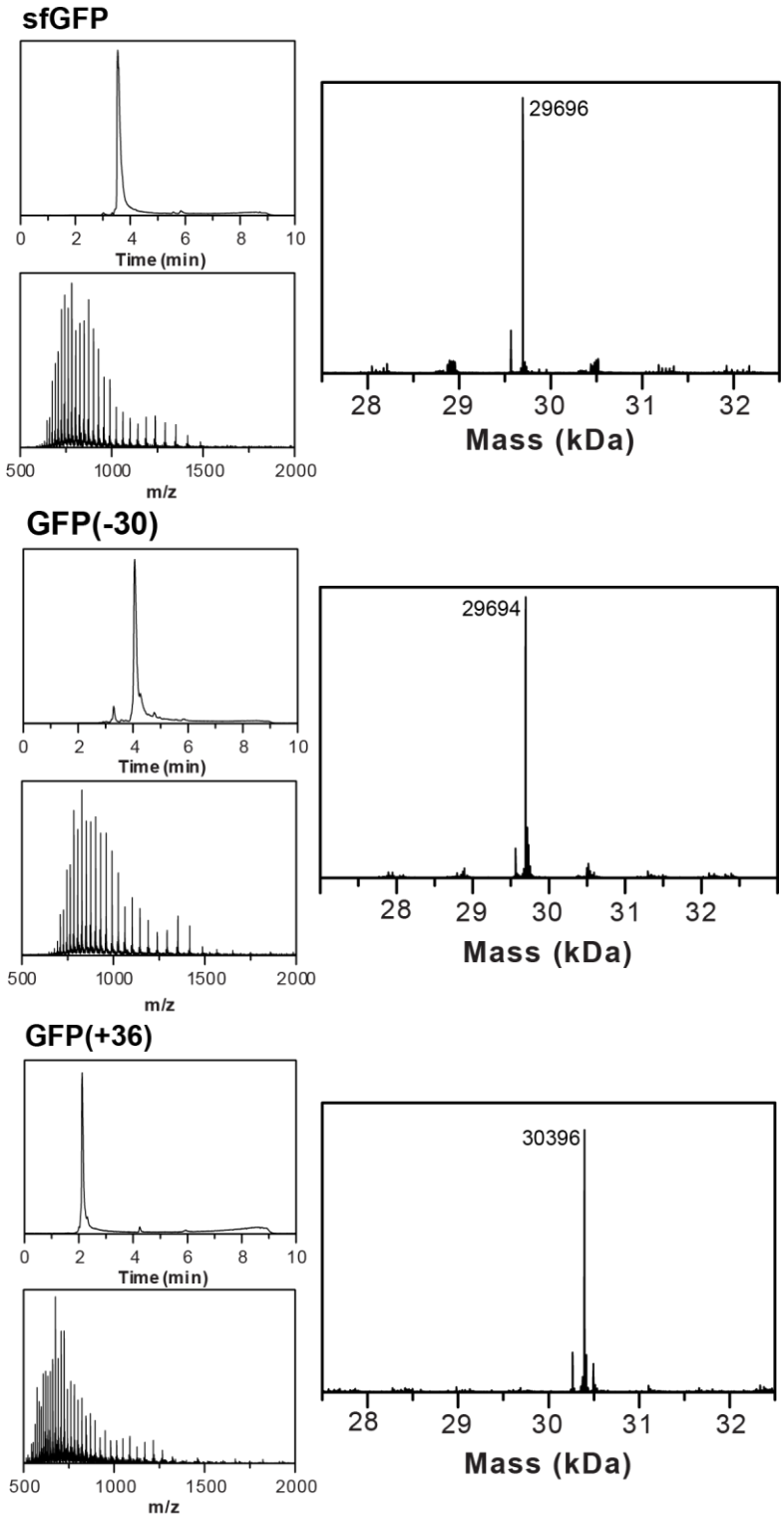


Figure S4. LC-MS Q-TOF analysis of each GFP. For each GFP, the total ion count chromatogram (top left), the m/z spectrum (bottom left), and the resulting mass spectrum (right) are shown. The theoretical mass, as calculated using the protein sequence on the online tool ProtParam (ExPASy), are 29716 Da for sfGFP(-8), 29714 Da for GFP(-30) , and 30416 Da for GFP(+36). All the experimental molecular weights are 20 Da lower, which is expected due to chromophore maturation.

Figure S5 Protein uptake into coacervates

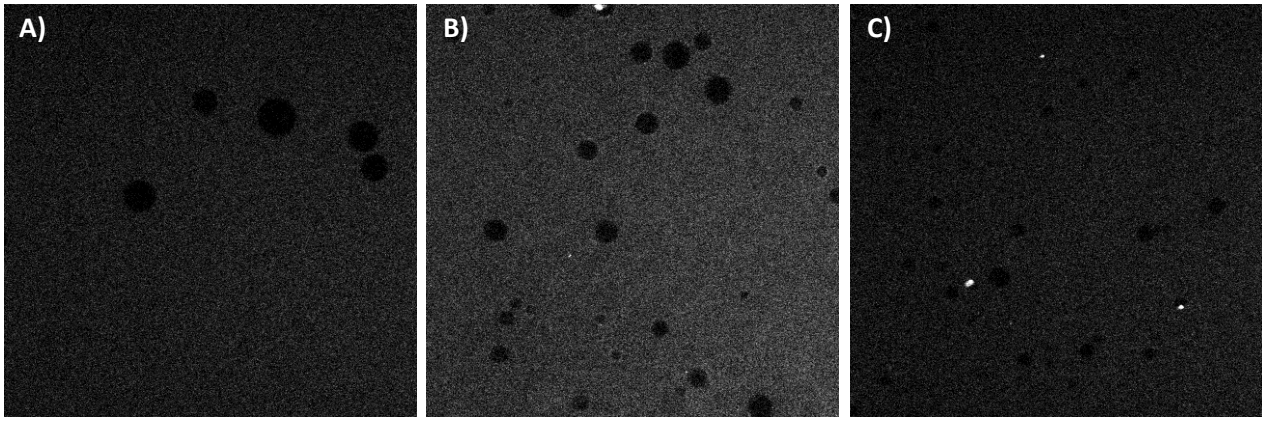


Figure S5. The uptake of GFP(-8) proteins at various charge balances. Q-Am : CM-Am ratios are A) 1:1, B) 2:1, C) 3:1.

Figure S6 Effect of PGA group on terpolymer assembly

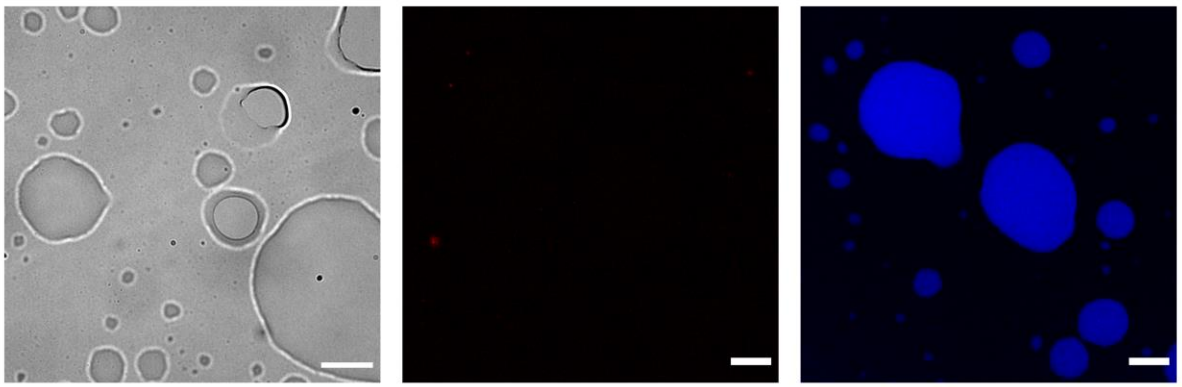


Figure S6. The poly(glutamic acid) moiety of the terpolymer is essential for interfacial self-assembly and stabilisation of coacervate droplets. These microscopy images show that when the diblock copolymer PEG-b-PCLgPTMC is used instead of the terpolymer, no evidence of membrane organisation around the droplets is observed, and no stabilisation occurs. Cy5 emission is from encapsulated Cy5-labelled, succinylated BSA. Scale bar represents 20 µm.

Figure S7 Line profile analysis of confocal image for thickness measurement

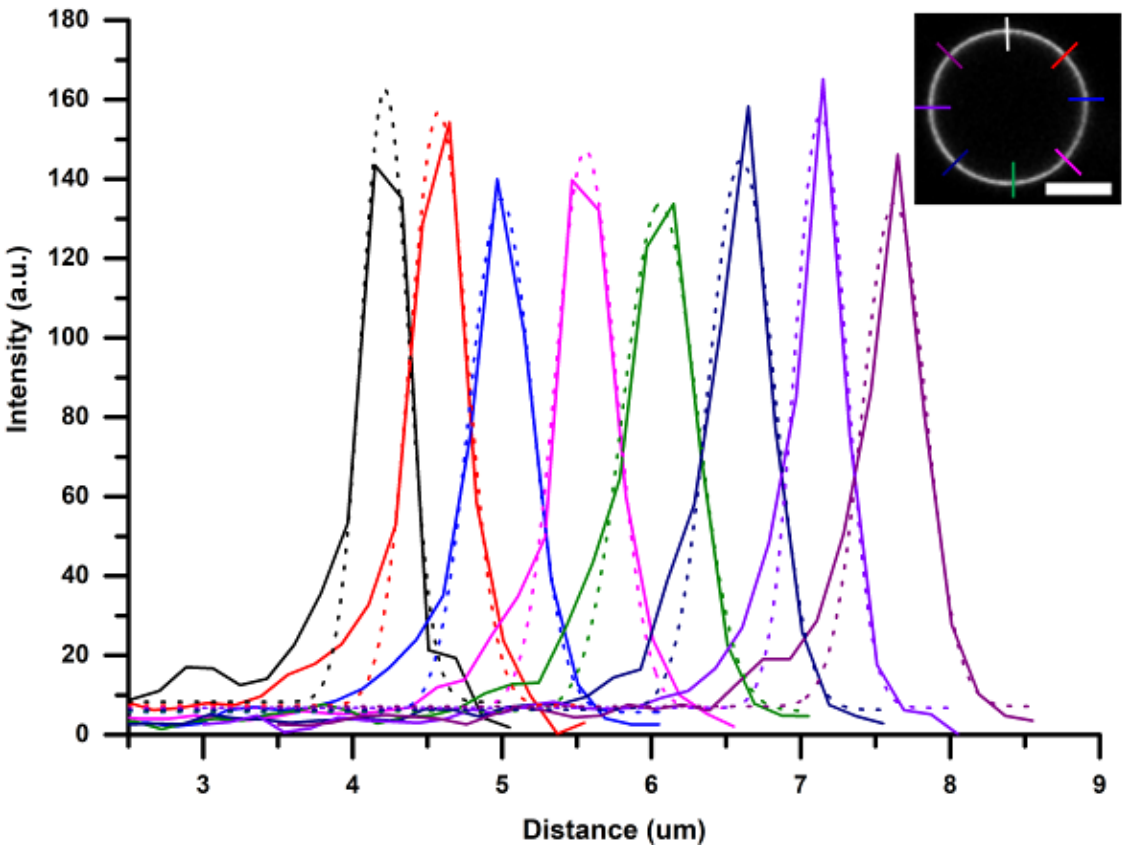
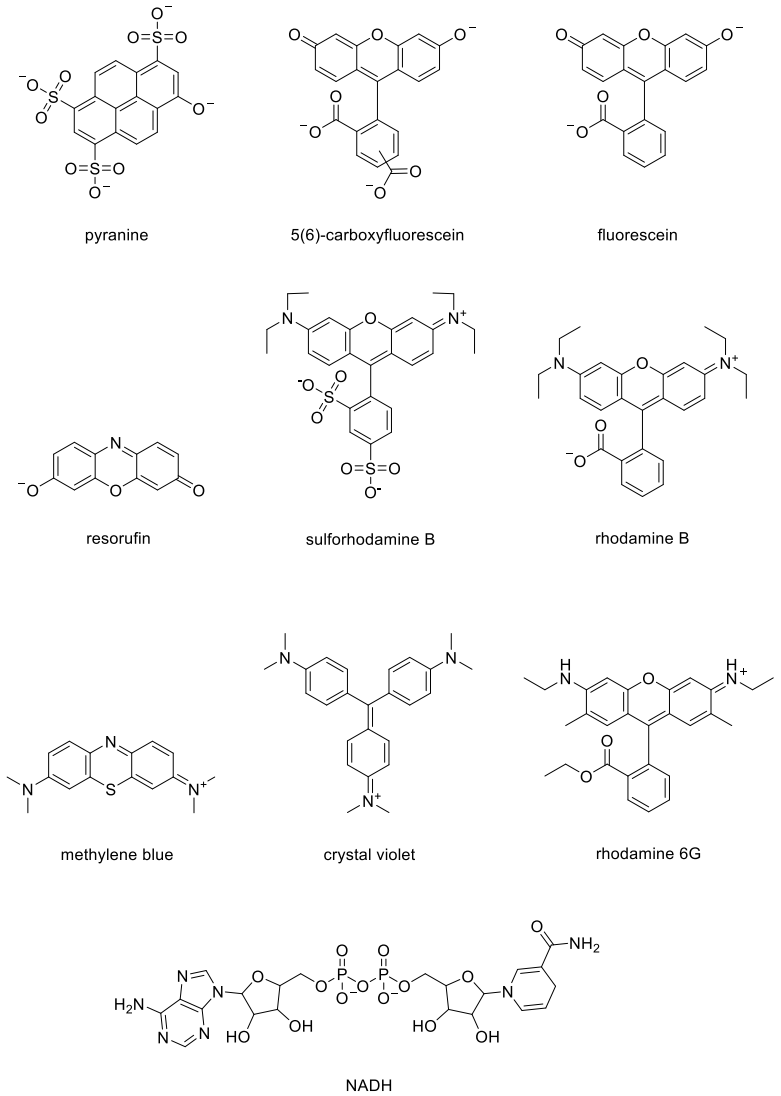


Figure S7. Line profile analysis of 8 radially equidistant regions on an image of a terpolymer-stabilized coacervate protocell (inset, scale represents 2 µm) labelled with Nile red. Overlaid are Gaussian fits to these data (dashed lines), from which Full Width at Half Maximum (FWHM) values were obtained for Figure 4C.

Figure S8 Structures of low molecular weight fluorophores, their estimated charges at pH 7.4, and partitioning coefficients



Fluorophore	Partitioning coefficient
Pyranine	52.1
Carboxyfluorescein	13.1
Fluorescein	6.1
NADH	1.6
Resorufin	3.5
Sulforhodamine B	2.9
Methylene blue	1.4

Figure S9 Structures of different classes of fluorophores

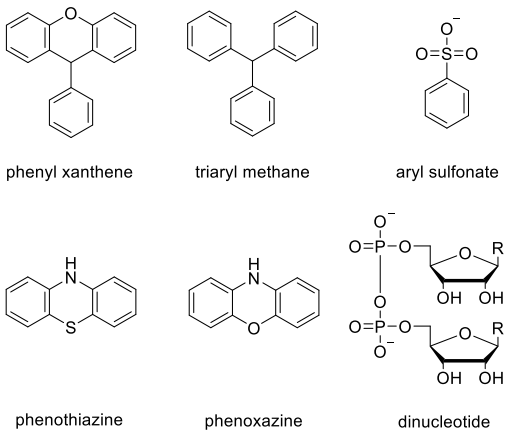


Figure S10 Uptake of small molecule fluorophores by stabilized or unstabilized coacervates

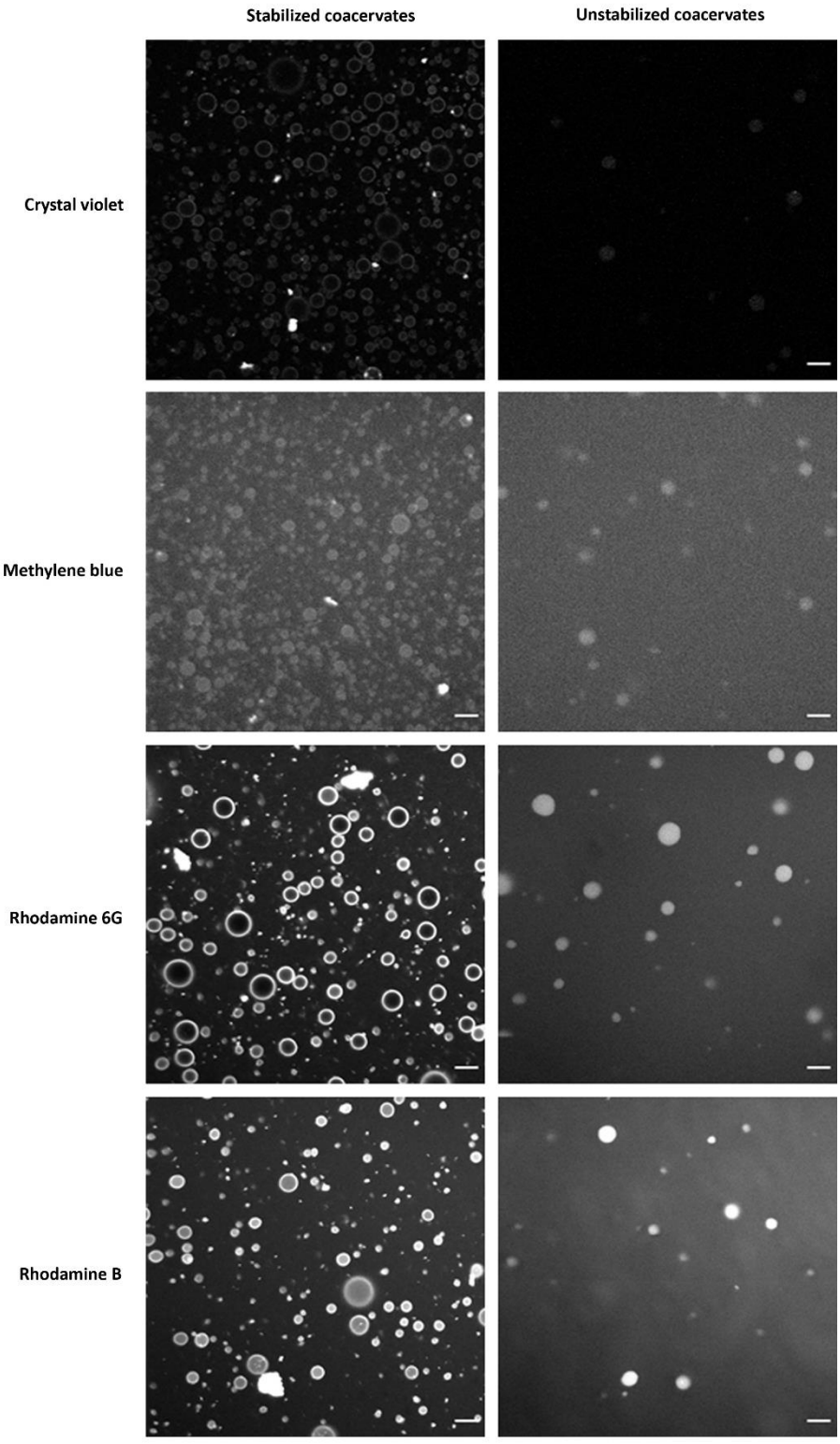


Figure S10. Uptake of several fluorophores into coacervate phase is affected by the terpolymer membrane. Left: uptake by terpolymer-stabilized coacervates. Right: uptake by unstabilized coacervates.

Figure S11 Effect of excess terpolymer on BSA permeability

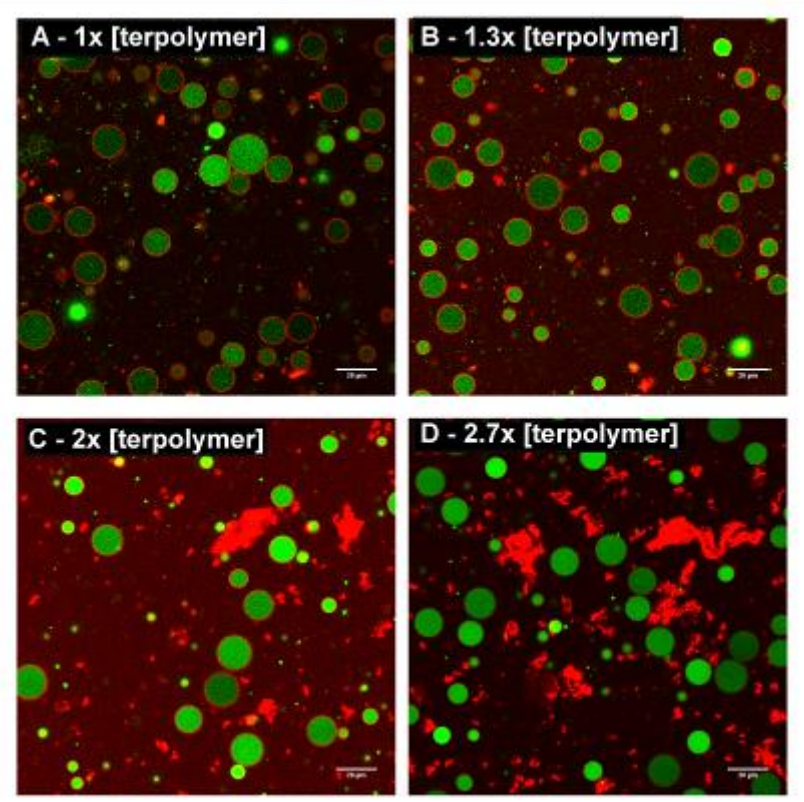


Figure S11. Confocal micrographs demonstrating the effect of excess terpolymer (red, stained with Nile red) compared to normal formation conditions on the permeability of succinylated BSA added after protocell formation (green, FITC). BSA is taken up despite a large excess of terpolymer, which only results in increasing amounts of polymer aggregation. Scale bar represents 20 µm.

Figure S12 Uptake of succinylated HRP into the coacervate phase

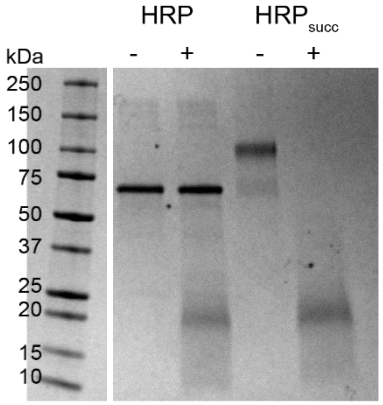
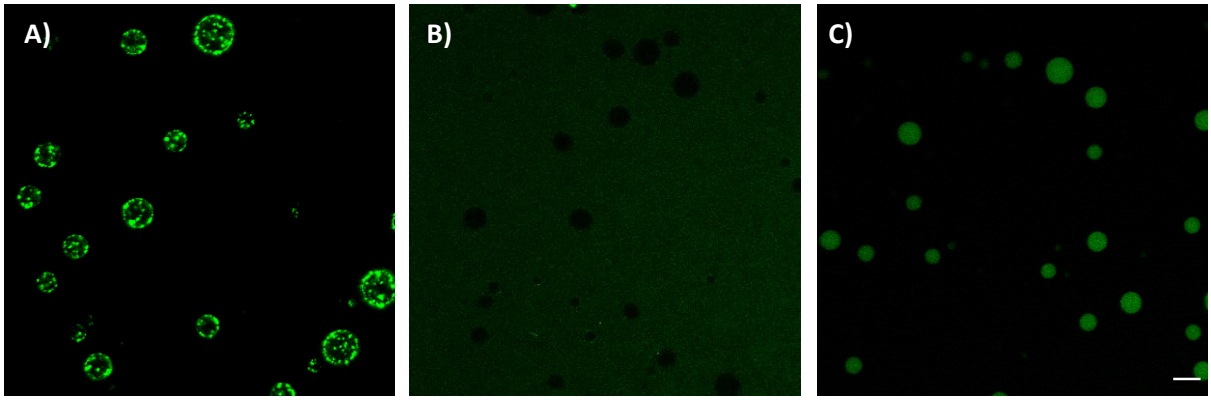


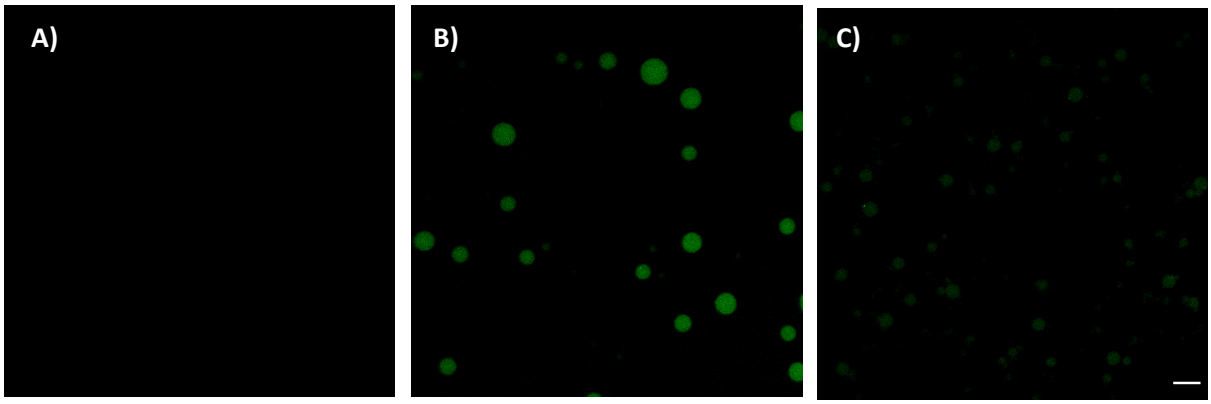
Figure S12. An SDS-PAGE gel showing proteins in buffer without coacervates (-) and the polymer-poor phase of proteins mixed with coacervates (+). Non-succinylated proteins do not readily enter the coacervate phase as shown by presence of protein bands in the (+) samples, whereas succinylated HRP (HRP_{succ}) is readily sequestered into the coacervate phase as indicated by the absence of a corresponding protein band. The diffuse 22 kDa band is present in all (+) samples and is likely to be low molecular weight amylose which interacts with Coomassie dye.

4 Supplementary Images

Figure 1:



The uptake of GFP(+36) (A), GFP(-8) (B) or GFP(-30) (C) into coacervates composed of 2:1 Q-Am : CM-Am ($DS_Q = DS_{CM} = 0.8$). Scale bar represents 10 µm.



The uptake of GFP(-30) into coacervates composed of 1:1 (A), 2:1 (B) or 3:1 (C) Q-Am : CM-Am ($DS_Q = DS_{CM} = 0.8$). Scale bar represents 10 µm.

Figure 2C, 2E:

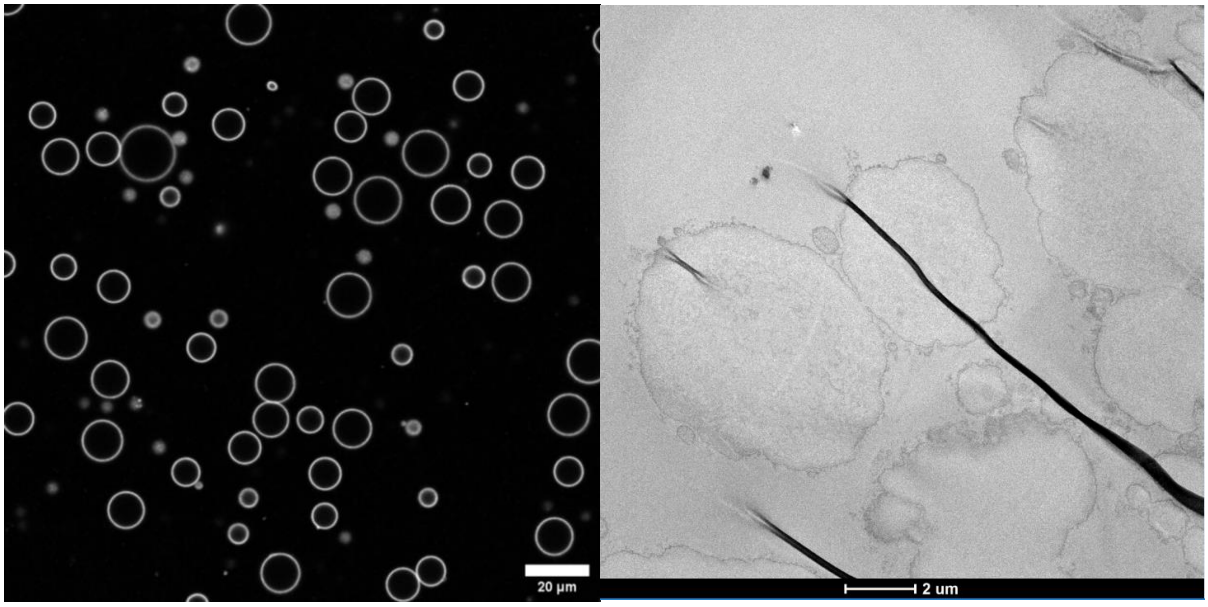
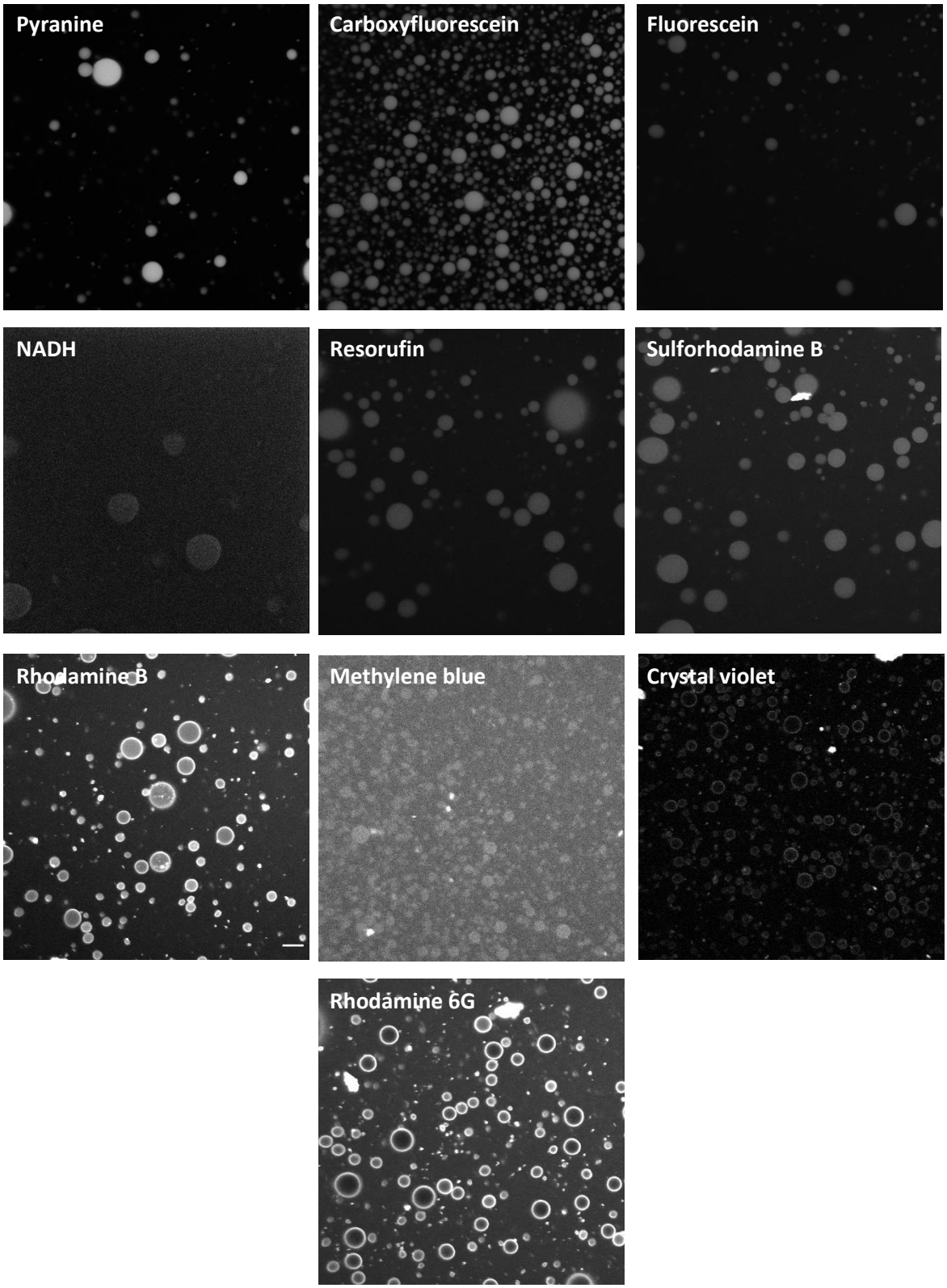
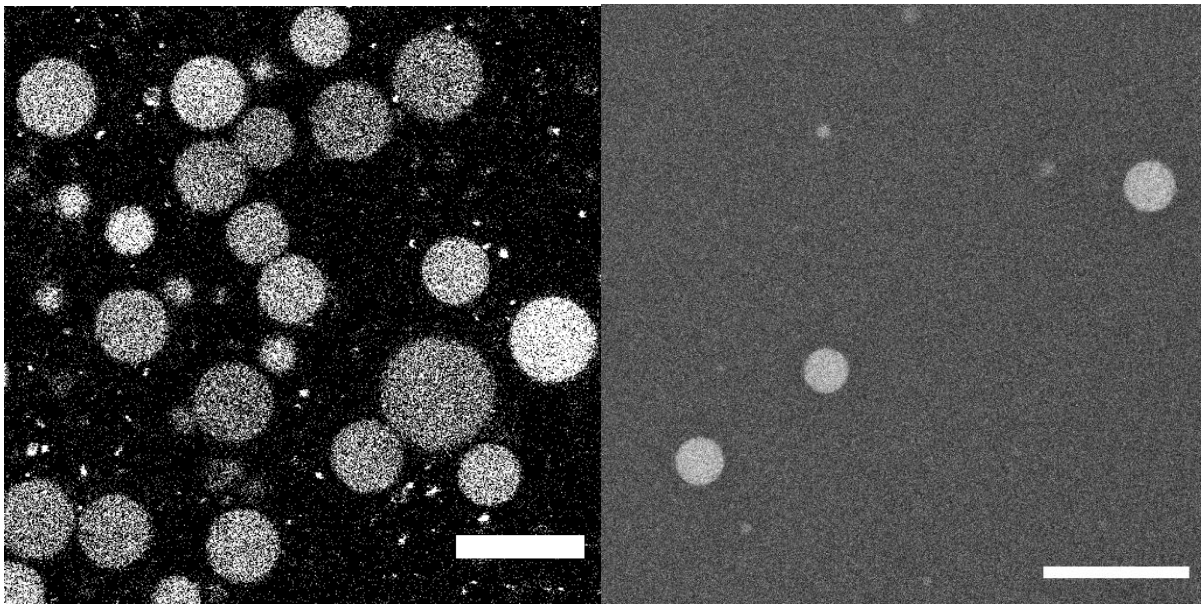


Figure 3:

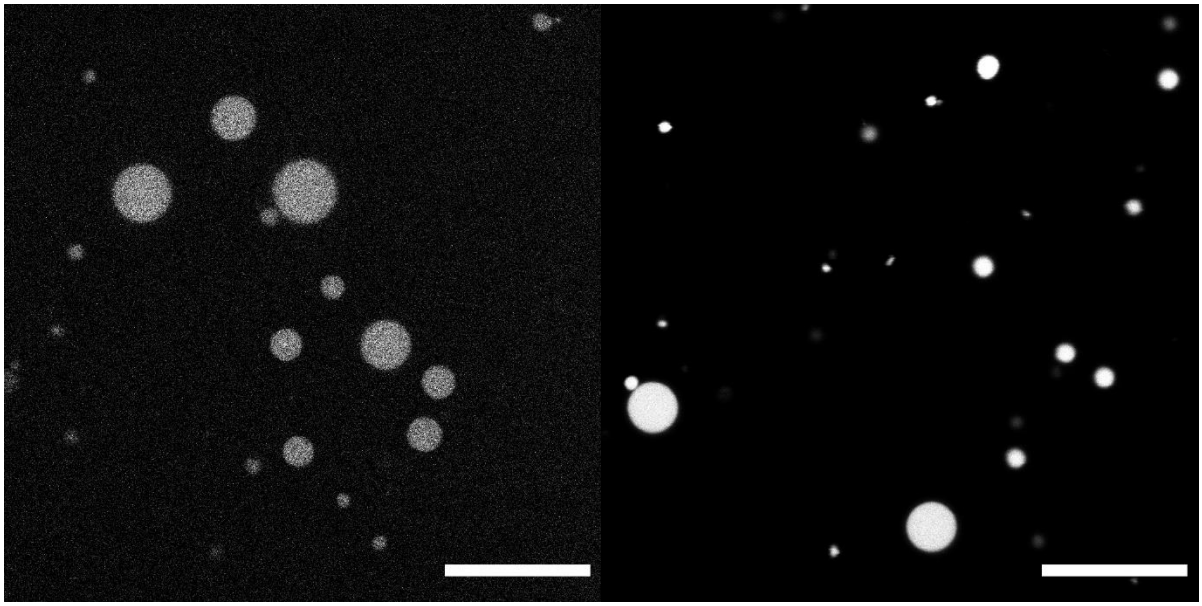


Scale bar represents 10 µm.

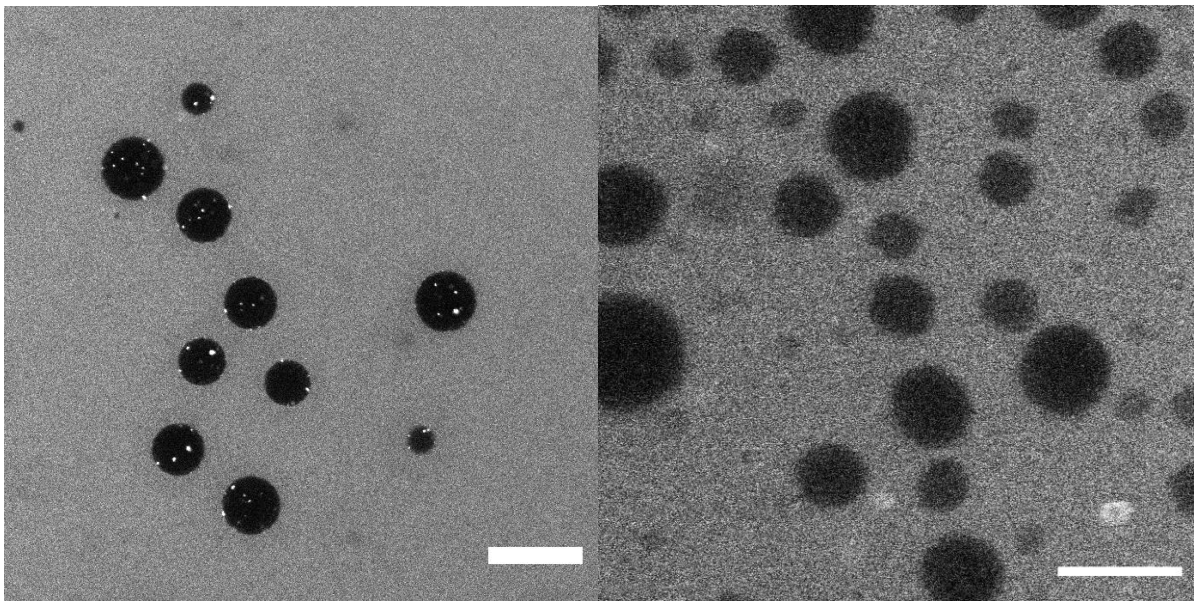
Figure 4:



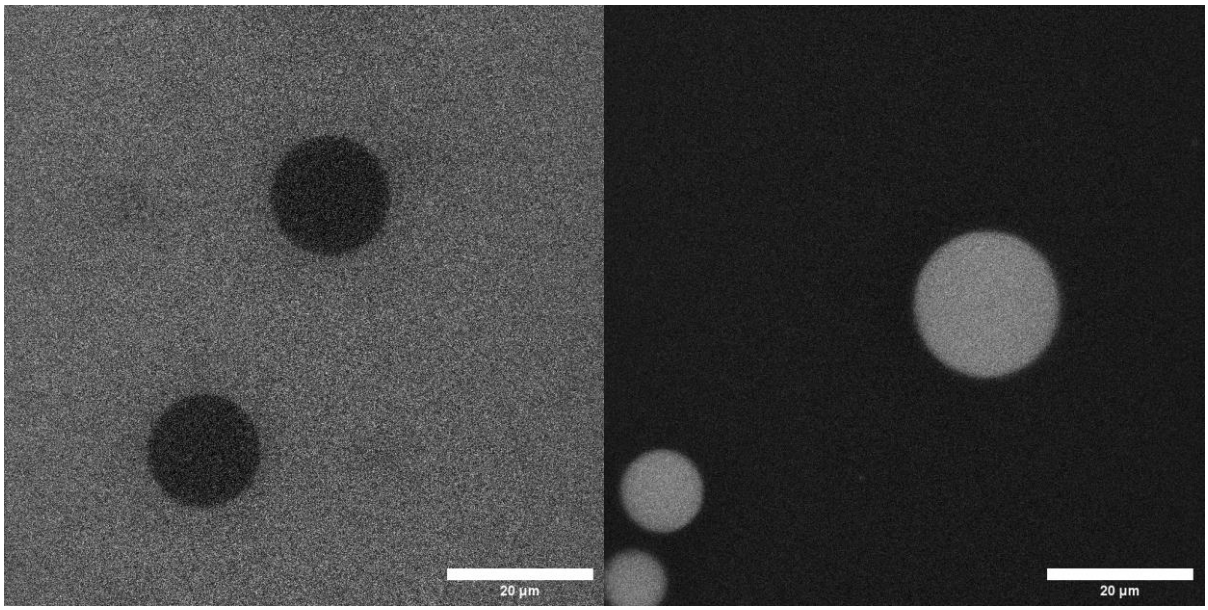
scBSA, 4 kDa Dextran (scale = 20 µm)



DNA 13bp, DNA 27bp (scale = 20 µm)



BSA, sfGFP (scale = 20 µm)



70 kDa Dextran t=0, 70 kDa Dextran t = 18h (scale = 20 µm)

Supplementary References

[1] A. F. Mason, B. C. Buddingh', D. S. Williams, J. C. M. van Hest, *J. Am. Chem. Soc.* **2017**, *139*, 17309–17312.

[2] A. Couffin, D. Delcroix, B. Martín-Vaca, D. Bourissou, C. Navarro, *Macromolecules* **2013**, *46*, 4354–4360.

[3] J. Zou, J. Fan, X. He, S. Zhang, H. Wang, K. L. Wooley, *Macromolecules* **2013**, *46*, 4223–4226.

[4] a) M. S. Lawrence, K. J. Phillips, D. R. Liu, *J. Am. Chem. Soc.* **2007**, *129*, 10110-10112; b) Y. Azuma, R. Zschoche, M. Tinzl, D. Hilvert, *Angewandte Chemie International Edition* **2016**, *55*, 1531-1534.