Article

Cite This: Biomacromolecules 2019, 20, 3641−3647

pubs.acs.org/Biomac

Nature of Amorphous Hydrophilic Block Affects Self-Assembly of an Artificial Viral Coat Polypeptide

Lione Willems,[†] Larissa van Westerveld,[†] Stefan Roberts,[§] Isaac Weitzhandler,[§] Carlos Calcines Cruz,[∥] Armando Hernandez-Garcia,[∥] Ashutosh Chilkoti,[§] Enrico Mastrobattista,[⊥] John van der Oost,[‡] and Renko de Vries*[,†]

[†]Physical Chemistry and Soft Matter and [‡]Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

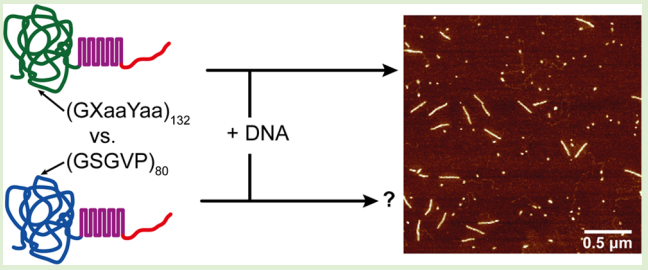
[§]Department of Biomedical Engineering, Duke University, Durham, North Carolina 27708, United States

[∥]Institute of Chemistry, Department of Chemistry of Biomacromolecules, National Autonomous University of Mexico, 04510 Mexico City, Mexico

[⊥]Department of Pharmaceutics, Utrecht Institute of Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands

Supporting Information

ABSTRACT: Consensus motifs for sequences of both crystallizable and amorphous blocks in silks and natural structural analogues of silks vary widely. To design novel silklike polypeptides, an important question is therefore how the nature of either the crystallizable or the amorphous block affects the self-assembly and resulting physical properties of silklike polypeptides. We address herein the influence of the amorphous block on the self-assembly of a silklike polypeptide that was previously designed to encapsulate single DNA molecules into rod-shaped viruslike particles. The polypeptide has a triblock architecture, with a long N-terminal amorphous block, a crystallizable midblock, and a C-terminal DNA-binding block. We compare the self-assembly behavior of a triblock with a very hydrophilic collagen-like amorphous block $(GXaaYaa)_{132}$ to that of a triblock with a less hydrophilic elastin-like amorphous block $(GSGVP)_{80}$. The amorphous blocks have similar lengths and both adopt a random coil structure in solution. Nevertheless, atomic force microscopy revealed significant differences in the self-assembly behavior of the triblocks. If collagen-like amorphous blocks are used, there is a clear distinction between very short polypeptide-only fibrils and much longer fibrils with encapsulated DNA. If elastin-like amorphous blocks are used, DNA is still encapsulated, but the polypeptide-only fibrils are now much longer and their size distribution partially overlaps with that of the encapsulated DNA fibrils. We attribute the difference to the more hydrophilic nature of the collagen-like amorphous block, which more strongly opposes the growth of polypeptide-only fibrils than the elastin-like amorphous blocks. Our work illustrates that differences in the chemical nature of amorphous blocks can strongly influence the self-assembly and hence the functionality of engineered silklike polypeptides.



INTRODUCTION

Self-assembly of proteins is increasingly exploited to design scaffolds for cell growth, nanoparticles for drug delivery, biosensors, and nanoreactors.[1,2] A relatively straightforward approach to the de novo design of self-assembling structural proteins is the use of well-conserved motifs found in natural structural proteins such as elastin, silk, and collagen. These consensus motifs can be repeated and combined in many ways to design sequences for novel self-assembling polypeptides for a wide range of applications.[1,3]

An important class of self-assembling polypeptides is silklike polypeptides that typically consist of both crystallizable and amorphous blocks. Consensus motifs for sequences of both crystallizable and amorphous blocks in natural silks and structural silk analogues, such as spider silks,[4] silkworm silks,[5] and squid sucker ring teeth,[6] vary widely. An obvious question is therefore how the nature of the crystallizable and the amorphous blocks affects the self-assembly and the resulting physical properties of these materials. The same question also arises for de novo design of silklike polypeptides, but so far this issue has not yet been addressed in the literature.

We here consider the influence of the nature of the amorphous block on self-assembly of a triblock polypeptide designed previously to encapsulate single DNA molecules into rod-shaped viruslike particles (VLPs).[7] The triblock polypeptide $C\text{-}S^{Q}_{10}\text{-}K_{12}$ features a dodeca-lysine block (K_{12}) for binding

Received: April 15, 2019
Revised: August 8, 2019
Published: August 16, 2019

to the nucleic acid template, a silklike midblock S^{Q}_{10} = $(GAGAGAGQ)_{10}$ that self-assembles into the rod-shaped core of the VLP via stacking of β-rolls,[8−11] and a collagen-like amorphous block C = $(GXaaYaa)_{132}$ for colloidal stabilization of the VLP particles. Note that the Xaa and Yaa residues in the collagen-like C block are chosen to be highly hydrophilic and mostly uncharged such that the resulting polypeptide adopts a random coil conformation rather than a triple helix.[12]

Apart from a general understanding of the influence of the nature of amorphous blocks on the properties of silk polypeptides, we also have more specific reasons for exploring other kinds of amorphous blocks for the C-S^{Q}_{10}-K_{12} triblock polypeptide. First of all, the nature of the amorphous block strongly impacts the production of the polypeptide in host organisms: while the C-S^{Q}_{10}-K_{12} polypeptide can be obtained at a high yield by secreted expression in the yeast *Pichia pastoris*,[7] expression in *Escherichia coli* is problematic (see SI-1). Next, with respect to applications, since the amorphous block is on the outside of a VLP, it is the amorphous block that will, to a large extent, determine the initial biological response of cells and tissues to the VLP. Therefore, it is of interest to explore amorphous blocks that have been very well characterized with respect to their biological response, such as elastin-like polypeptides. Elastin-like polypeptides (ELPs) with the general sequence E_n^X = $(GXGVP)_n$ are known to be expressed at high yield in *E. coli*,[13] whereas for *P. pastoris*, expression levels for ELPs are relatively low, especially for ELPs with more hydrophobic guest residues X.[14] For studying our previously designed triblock polypeptide with an elastin-like rather than a collagen-like amorphous block, we therefore choose production in *E. coli*.

To study the impact of the nature of the amorphous block, we compare self-assembly and co-assembly with DNA of a new polypeptide E^{S}_{80}-S^{Q}_{10}-K_{12} (produced in *E. coli*) with that of the C-S^{Q}_{10}-K_{12} polypeptide (previously produced in *P. pastoris*). To make a good comparison with the amorphous collagen-like C block, we use an elastin-like polypeptide with a length of 80 pentamers (E^{S}_{80}) to match the length of the approximately 400 amino-acid-long C block. As a guest residue in the elastin-like polypeptide, we use serine (X = S) since it is uncharged and hydrophilic. For this guest residue, at the conditions of co-assembly of the polypeptide with DNA, the ELP block will be far below its transition temperature and adopt a random coil configuration.

MATERIALS AND METHODS

Materials. The C-S^{Q}_{10}-K_{12} polypeptide was produced and purified as described previously.[7] For the full amino acid sequence of the C-S^{Q}_{10}-K_{12} polypeptide, in particular, the choice of Xaa and Yaa amino acids in the collagen-like amorphous block C = (GXaaYaa), see ref 7.

A pET24a(+) PRe-RDL cloning vector was constructed as previously described.[15] Custom oligonucleotides were synthesized by Integrated DNA Technologies Inc. All restriction enzymes, the calf-intestinal phosphatase (CIP), and the Quick Ligation kit were purchased from New England Biolabs, and the T4 DNA ligase buffer from Invitrogen. DNA miniprep, gel purification, and PCR purification kits were from Qiagen Inc. Chemically competent *E. coli* cells (strains EB5Alpha and BL21 (DE3)) were ordered from EdgeBioSystems. The NoLimits 2000 bp linear dsDNA, PageBlue protein staining solution, SYBR Safe DNA stain, 6× DNA loading dye, GeneRuler 1 kb DNA ladder, and 50× TAE buffer were purchased from Thermo Fisher Scientific. The 10% mini-protean TGX precast protein gels and Precision Plus Protein All Blue Prestained Protein Standard were ordered from Bio-Rad, and agarose was bought from Brunschwig Chemie.

Preparation of Oligolysine-Encoding Cloning Vector. Two complementary custom oligonucleotides encoding the oligolysine K_{12} were designed: 5′-cAAGAAAAAGAAGAAAAAGAAGAAGAAAAAAGAAGgg3′ and 5′-CTTCTTTTTCTTCTTCTTCTTTTTCTTCTTTTTCTTgcc-3′. To anneal the two oligonucleotides, they were diluted in T4 DNA ligase buffer to a concentration of 2 μM, heated to 95 °C for 2 min, and then slowly cooled to room temperature over 3 h. Linear pET24a(+) PRe-RDL cloning vector was prepared by the digestion of approximately 2 μg of vector for 2 h at 37 °C with 5 U of *BseRI*; the 5′ ends of the linearized vector was dephosphorylated using 10 U of CIP for 30 min at 37 °C. Next, a PCR purification kit was used to remove the enzymes and the annealed oligonucleotides were ligated into the linearized vector by incubating a mixture of both in 1× Quick ligase buffer for 5 min at room temperature in the presence of Quick ligase. The ligation product was then transformed into EB5Alpha chemically competent cells and the cells were plated on TBdry plates supplemented with 45 μg/mL kanamycin. Colonies containing the plasmid with a copy of the oligolysine K_{12} were selected by colony PCR and confirmed by DNA sequencing.

Gene Construction by PRe-RDL. The method of recursive directional ligation by plasmid reconstruction (PRe-RDL)[15] was used to construct a plasmid encoding for the E^{S}_{80}-S^{Q}_{10}-K_{12} polypeptide. To obtain the so-called A-fragment, a previously produced[16] PRe-RDL cloning vector coding for E^{S}_{80}-S^{Q}_{10} was digested with *BglI* and *AcuI* enzymes for 3 h at 37 °C and the *BglI* × *AcuI* fragment containing the E^{S}_{80}-S^{Q}_{10} gene was purified by gel purification. The B-fragment was obtained by digesting the cloning vector coding for K_{12} with *BglI* and *BseRI* for 3 h at 37 °C, followed by gel purification of the *BglI* × *BseRI* fragment containing the K_{12} gene. Ligation of the A- and B-fragments using Quick ligase in 1× Quick ligase buffer for 5 min at room temperature resulted in the formation of the E^{S}_{80}-S^{Q}_{10}-K_{12} gene. The ligation product was transformed into EB5Alpha chemically competent cells, and the cells were plated on TBdry plates supplemented with 45 μg/mL kanamycin. The sequence was confirmed by DNA sequencing. Apart from the methionine start codon, the exact amino acid sequence of the encoded polypeptide is $(GSGVP)_{80}(GAGAGAGQ)_{10}GK_{12}G$, and this corresponds to an expected molecular weight of 39 159.97 Da.

Expression and Purification. To express the triblock polypeptide E^{S}_{80}-S^{Q}_{10}-K_{12}, the plasmid as described above was transformed into *E. coli* BL21(DE3) chemically competent cells, which were then used to inoculate a 10 mL starter culture of TB medium supplemented with 45 μg/mL kanamycin. The starter culture was incubated overnight at 37 °C on a shaker at 200 rpm. Next, the starter culture was used to inoculate a culture of 2 L TB medium supplemented with 45 μg/mL kanamycin. Cells were grown for a total time of 24 h at 37 °C on a shaker at 200 rpm. To induce polypeptide expression, isopropyl β-D-1-thiogalactopyranoside at a final concentration of 1 mM was added to the medium 8 h after the inoculation. After polypeptide expression, cells were pelleted by centrifugation at 2000g for 10 min at 10 °C and resuspended in 25 mL of 20 mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES) (pH 8.0). The cells were lysed by once pressing through a cooled French press at 1.5 bar. The lysate was centrifuged at 29 000g for 12 min at 4 °C to pellet insoluble cellular debris, and the supernatant (soluble lysate) was collected for further purification. To remove nucleic acids, the soluble lysate (~30 mL) was mixed with 4 mL of 10% (w/v) poly(ethylenimine) and centrifuged at 29 000g for 12 min at 4 °C. The polypeptide in the remaining supernatant (soluble cleared lysate) was further purified exploiting both the thermal properties of the elastin-like block (E^{S}_{80}) by using inverse transition cycling (ITC),[13] and the strong positive charge of the oligolysine block (K_{12}) in cation-exchange chromatography.

For the cation-exchange chromatography, the soluble cleared lysate was loaded onto a UNO S6 column (Bio-Rad) connected to a BioLogic DuoFlow chromatography system supplied with a QuadTec detector from Bio-Rad. Unbound molecules were removed from the column by washing with four column volumes of 20 mM HEPES buffer (pH 8.0). Bound molecules were eluted by using a salt gradient

DOI: 10.1021/acs.biomac.9b00512
Biomacromolecules 2019, 20, 3641−3647

from 0.0 to 1.0 M of NaCl in a 20 mM HEPES buffer (pH 8.0) applied over 10 column volumes. During elution, 1 mL fractions of the samples were collected from the column. Fractions containing the polypeptide were first dialyzed against Milli-Q water and then subjected to one round of ITC using the following protocol: (1) aggregation of the ELP-containing polypeptide was induced at room temperature by the addition of 2.5−3.0 M ammonium sulfate to the dialyzed sample, followed by incubation at 37 °C for 15 min. (2) Sample was centrifuged at 40 °C for 12 min at 29 000*g* to pellet the aggregated polypeptides. (3) Polypeptides in the pellet were solubilized in 50 mM HEPES buffer (pH 8.0). To facilitate solubilization, samples were vortexed at 1000 rpm for 90 min at room temperature. (4) To remove any remaining insoluble matter, the samples were first cooled on ice and then centrifuged at 4 °C for 12 min at 29 000*g*. Finally, all polypeptides were dialyzed against Milli-Q water, lyophilized, and stored at room temperature.

Characterization. Purity of the polypeptides was assessed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS). SDS-PAGE was carried out using 10% mini-protean TGX precast protein gels, 1× Laemmli running buffer and Precision Plus Protein All Blue Prestained Protein Standard, and gels were stained with PageBlue protein staining solution. For MALDI-TOF MS, a matrix solution was prepared of 15 mg/mL 2.5-dihydroxyacetophenone and 4.5 mg/mL diammonium hydrogen citrate in 75% ethanol. Next, 1 volume of this matrix solution was mixed with 1 volume of 1 mg/mL polypeptide and 1 volume of 3% (v/v) trifluoroacetic acid, and 1 μL of this mixture was dried on an 800 μm spot of an MTP AnchorChip 384 target (Bruker). Analysis was then carried out on an UltrafleXtreme mass spectrometer (Bruker).

Atomic Force Microscopy (AFM). Polypeptide stock solutions of 100 μM were prepared in a 10 mM phosphate buffer (pH 7.4). To promote solubilization of the freeze-dried polypeptides, they were incubated at 65 °C for 1 h. This temperature is high enough to disrupt aggregates and assemblies of the polypeptides, but not so high as to cause irreversible changes to their assembly after cooling down.[17] Next, polypeptides were further diluted in 10 mM phosphate buffer (pH 7.4) to a concentration of 1.8 μM and, if so indicated, mixed with DNA at a final DNA concentration of 1 μg/mL, which corresponds to a polypeptide-to-DNA molar charge ratio N/P of 7. All samples were incubated for 24 h at room temperature to allow formation of complexes. For atomic force microscopy (AFM), 5 μL of a sample that was prepared as described in the previous section was deposited on a clean silicon surface and incubated for 2 min. Next, the surface was rinsed with 1 mL of Milli-Q water to remove salts and nonabsorbed particles and dried slowly under a gentle N_2 stream. Samples were analyzed on a NanoScope MultiMode 8 system (Bruker) in the ScanAsyst (PeakForce Tapping) imaging mode, using ScanAsyst-Air cantilevers (Bruker). Areas of 5 × 5 μm were scanned at 512 × 512 pixels and a line rate of 0.977 Hz. All images were subjected to a second-order flattening using NanoScope Analysis 1.40 software. If so indicated, lengths of fibrils were measured using FiberApp software.[18] Settings used for FiberApp are: α = 0, β = 500, γ = 20, kappa1 = 20, kappa2 = 10, step = 1 pixel, iterations = 100, "Use A* pathfinding algorithm". Weight-averaged fibril lengths were calculated assuming the mass of the fibrils was proportional to their length. The standard error of weighted means (sem_w) was approximated using the method of block averages. Data were subdivided into five subsets, weight-averaged lengths were calculated for the subsets, and the standard error of the weighted means was taken to be the standard deviation of the weight-averaged lengths for the subsets.

Gel Shift Assays. To quantify binding of the E^S_{80}-S^Q_{10}-K_{12} triblock polypeptide to DNA, linear dsDNA (2000 bp) was incubated at room temperature for 2 h with increasing concentrations of the polypeptide. To that purpose, a stock solution of 100 μM polypeptide in 10 mM phosphate buffer (pH 7.4) was prepared in the same way as for AFM. The stock solution was then diluted to the desired polypeptide concentrations and mixed with DNA. The DNA concentration was always 15 ng/μL, and the polypeptide concentrations were 0.00, 1.54, 2.31, 3.08, 3.85, 5.77, 7.69, and 26.92 μM. The corresponding N/P ratios were 0, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, and 7.0. For each mixture, an aliquot containing 52.5 ng of DNA was loaded onto a 1% agarose gel that was supplemented with 1× SYBR Safe DNA stain. After electrophoresis for 90 min at 60 V, the gel was scanned using a Bio-Rad Gel Doc EZ Imager.

RESULTS AND DISCUSSION

Cloning, Expression, and Purification. The plasmid encoding the triblock polypeptide E^S_{80}-S^Q_{10}-K_{12} was constructed by recursive directional ligation by plasmid reconstruction (PRe-RDL), as described before.[15] Figure 1a,b shows a schematic picture of the final E^S_{80}-S^Q_{10}-K_{12} plasmid and a schematic representation of the structure of the corresponding polypeptide when co-assembled with DNA.

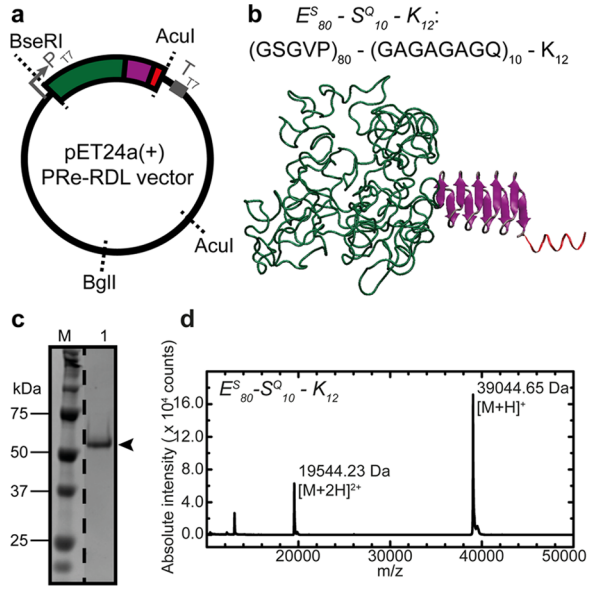


Figure 1. Expression and characterization of the E^S_{80}-S^Q_{10}-K_{12} triblock polypeptide. (a) PRe-RDL plasmid for the expression of the E^S_{80}-S^Q_{10}-K_{12} triblock polypeptide. The elastin-like polypeptide block E^S_{80} is shown in green, the silklike S^Q_{10} midblock in purple, and the oligolysine K_{12} binding block in red. (b) Cartoon of the structure of the triblock polypeptide when co-assembled with DNA. Colors as in (a). Silklike midblock is shown in β-solenoid configuration, which it supposedly adopts after fibrillar assembly onto DNA.[10,11] Also given is the amino acid sequence of the triblock polypeptide. (c) SDS-PAGE of purified E^S_{80}-S^Q_{10}-K_{12}; the arrowhead indicates the polypeptide. Lane M is a molecular weight marker. Gel was stained using PageBlue. (d) MALDI-TOF MS analysis of the polypeptide.

The E^S_{80}-S^Q_{10}-K_{12} triblocks were purified using cation-exchange chromatography followed by inverse transition cycling (ITC).[13] Purity and molar mass of the newly purified polypeptides were confirmed by SDS-PAGE and MALDI-TOF MS (Figure 1c,d). Note that, due to the hydrophilic random coil nature of the large amorphous blocks, the triblocks have an anomalously low mobility in SDS-PAGE compared to the SDS-PAGE molecular weight standards, exactly as has been observed before for the *C*-S^Q_{10}-K_{12} triblocks.[7]

The experimental mass, as determined from MALDI-TOF MS, is approximately 100 Da lower than expected if we assume cleavage of the N-terminal methionine.[19] Possibly, also the subsequent glycine (~75 Da) has been cleaved off, but this or

DOI: 10.1021/acs.biomac.9b00512
Biomacromolecules 2019, 20, 3641−3647

other small modifications are not expected to strongly affect the overall behavior of the polypeptides.

Influence of the Nature of the Amorphous Block on DNA-Binding Properties. To assess the impact of the nature of the amorphous block on the DNA-binding properties of the artificial viral capsid polypeptides, an electrophoretic mobility gel shift assay (Figure 2) was performed and compared to the earlier results for the $C\text{-}S^{Q}_{10}\text{-}K_{12}$ triblocks.[7] The gel shift assay shown in Figure 2 clearly demonstrates that binding of $E^{S}_{80}\text{-}S^{Q}_{10}\text{-}K_{12}$ is highly cooperative with high-mobility complexes (naked DNA; linear dsDNA, 2000 bp) coexisting, with low-mobility complexes (encapsulated DNA). The shift to low-mobility (encapsulated) DNA was observed for polypeptide-to-DNA charge ratios N/P (#nitrogens in polypeptides per #phosphates in DNA) larger than ~0.6. At high N/P ratio, the complexes are observed to run slightly toward the anode. While the observation of highly cooperative binding is similar to that for the $C\text{-}S^{Q}_{10}\text{-}K_{12}$ triblocks, for the latter (under similar conditions), encapsulated DNA was only observed for N/P larger than ~3.[7] Knowing that the amorphous block itself is not directly involved in binding to the DNA, this hints at a significant indirect influence of the nature of the amorphous block on complex formation.

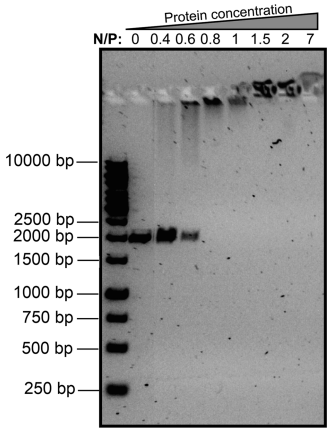


Figure 2. Gel shift assay of $E^{S}_{80}\text{-}S^{Q}_{10}\text{-}K_{12}$. The DNA (linear dsDNA, 2000 bp) concentration was kept constant at 15 ng/μL, and the DNA was incubated with increasing concentrations of the $E^{S}_{80}\text{-}S^{Q}_{10}\text{-}K_{12}$ triblock. The polypeptide-to-DNA molar charge ratios (N/P) are indicated above each lane, and the sizes of the DNA marker lane are labeled on the left. The total amount of DNA loaded into each well is 52.5 ng.

Influence of the Nature of the Amorphous Block on VLP Size Distributions. We hypothesized that the indirect influence of the nature of the amorphous block on the DNA-binding properties of the artificial viral capsid polypeptides occurs through influencing the assembly of the silklike midblocks into fibrils: it is the fibril formation of the silklike midblock that mediates binding cooperativity.[7] Therefore, we next investigated the co-assembly of $E^{S}_{80}\text{-}S^{Q}_{10}\text{-}K_{12}$ with DNA by AFM and compared it with the corresponding data for $C\text{-}S^{Q}_{10}\text{-}K_{12}$. Note that in earlier papers we have shown that VLP lengths obtained using AFM on dried complexes are consistent with those obtained using other methods, in particular, cryo-electron microscopy and static light scattering,[7] as well as AFM in liquid.[20]

Assembly and AFM imaging were performed at a polypeptide concentration of 1.8 μM and an excess of polypeptide over DNA (charge ratio N/P = 7). The AFM images of $E^{S}_{80}\text{-}S^{Q}_{10}\text{-}K_{12}$ and $C\text{-}S^{Q}_{10}\text{-}K_{12}$ in the absence and presence of a 2000 bp linear DNA template are shown in Figure 3.

Previously, we found[7] that for $C\text{-}S^{Q}_{10}\text{-}K_{12}$, the length of fibrils co-assembled with DNA was equal to the contour length of the DNA template, divided by a factor of approximately 3 and which was called the "packing factor". For a 2000 bp template, the DNA contour length divided by a packing factor of 3 equals approximately 230 nm. For the $C\text{-}S^{Q}_{10}\text{-}K_{12}$ polypeptide, fibrils with this length are only found in the

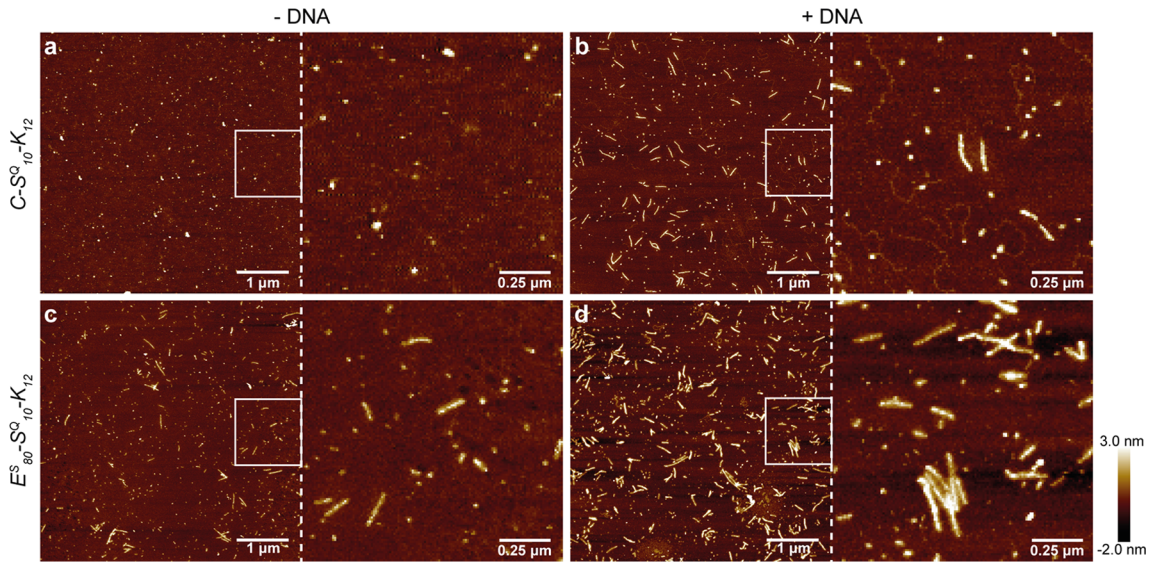


Figure 3. Self-assembly and co-assembly of $C\text{-}S^{Q}_{10}\text{-}K_{12}$ and $E^{S}_{80}\text{-}S^{Q}_{10}\text{-}K_{12}$ triblocks with and without a DNA template. Triblocks were solubilized at 65 °C, diluted to a concentration of 1.8 μM, and incubated at room temperature for 24 h in the absence or presence of linear dsDNA of 2000 bp. (a) $C\text{-}S^{Q}_{10}\text{-}K_{12}$ polypeptide incubated in the absence of DNA. (b) $C\text{-}S^{Q}_{10}\text{-}K_{12}$ polypeptide incubated in the presence of DNA. (c) $E^{S}_{80}\text{-}S^{Q}_{10}\text{-}K_{12}$ polypeptide incubated in the absence of DNA. (d) $E^{S}_{80}\text{-}S^{Q}_{10}\text{-}K_{12}$ polypeptide incubated in the presence of DNA. In each of the four conditions, the right panel is a digital magnification to highlight the formed fibrils.

DOI: 10.1021/acs.biomac.9b00512
Biomacromolecules 2019, 20, 3641−3647

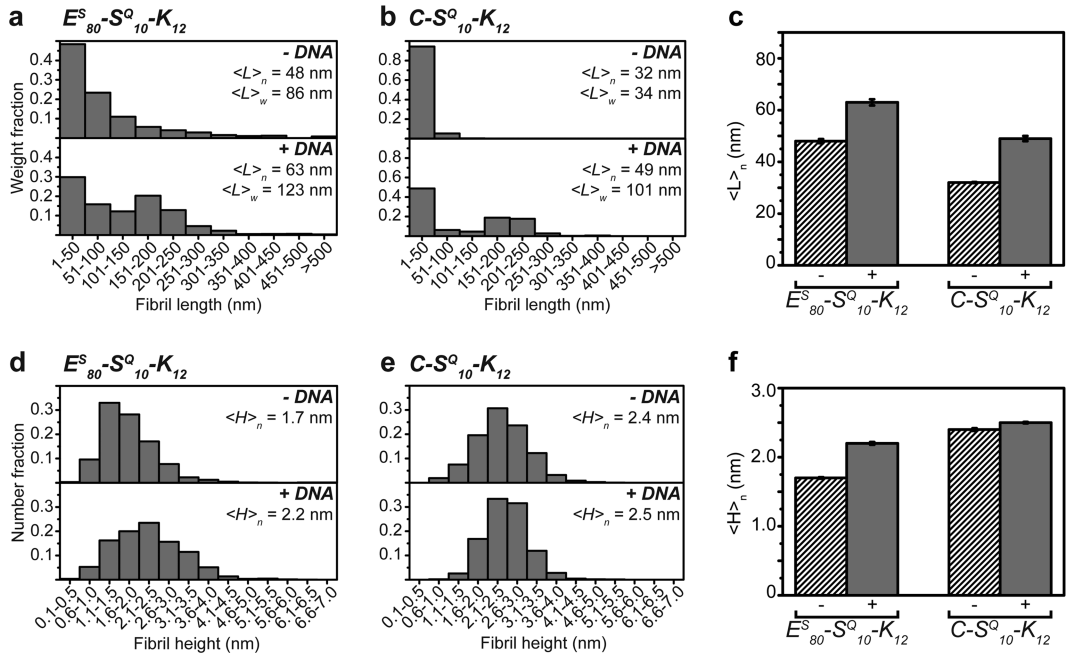


Figure 4. Length and height distributions of fibrils formed by C-S^Q_{10}-K_{12} and E^S_{80}-S^Q_{10}-K_{12} triblocks in the presence and absence of linear dsDNA (2000 bp), as derived from the AFM images. Representative AFM images are shown in Figure 3. Length distributions are shown as the weight fraction of polypeptide incorporated into fibrils within a given length interval. Height distributions are shown as the number fraction of polypeptide within a given interval. Length distributions for (a) E^S_{80}-S^Q_{10}-K_{12} and (b) C-S^Q_{10}-K_{12}, and height distributions for (d) E^S_{80}-S^Q_{10}-K_{12} and (e) C-S^Q_{10}-K_{12} are all split in two graphs, with the top graphs showing the distribution in the absence of DNA and the bottom graphs in the presence of DNA. Note that in (b) the *y*-axis scale differs from that in (a). (c) Number-averaged fibril lengths of both polypeptides summarized in a bar diagram and (f) number-averaged fibril heights of both polypeptides summarized in a bar diagram. The striped bars represent the samples without DNA, and the filled bars represent the samples with DNA. The error bars in (c) and (f) are ±sem (see values in Table 1).

Table 1. Average Fibril Lengths and Heights for E^S_{80}-S^Q_{10}-K_{12} and C-S^Q_{10}-K_{12} Polypeptides Assembled with or without 2000 bp dsDNA, Obtained by Analyzing AFM Images[a]

		N	$\langle L\rangle_n$ (nm)	$\langle L\rangle_w$ (nm)	$\langle H\rangle_n$ (nm)
E^S_{80}-S^Q_{10}-K_{12}	1.8 μM	2551	48 (±0.8)	86 (±18.9)	1.7 (±0.01)
	1.8 μM + DNA	2443	63 (±1.2)	123 (±9.4)	2.2 (±0.02)
C-S^Q_{10}-K_{12}	1.8 μM	1478	32 (±0.2)	34 (±1.2)	2.4 (±0.02)
	1.8 μM + DNA	2708	49 (±1.0)	101 (±7.9)	2.5 (±0.01)

[a]All averages (±sem) were calculated using the data obtained from AFM images.

presence of the DNA template (Figure 3a,b). Indeed, for C-S^Q_{10}-K_{12}, in the absence of DNA, fibrils are mostly shorter than 50 nm. In contrast, for E^S_{80}-S^Q_{10}-K_{12}, in the absence of DNA, fibrils with lengths higher than 100 nm are also formed (Figure 3c). In the presence of DNA, the number of fibrils longer than 100 nm drastically increases (Figure 3d), indicating DNA encapsulation, but the encapsulation is somewhat obscured by the presence of nontemplated fibrils as we will show in a more quantitative analysis below.

We performed a quantitative analysis of the AFM images by measuring both the lengths and heights of the fibrils. To that purpose, all fibrils in three or four AFM images of 5 × 5 μm surfaces were measured per condition. Histograms of fibril lengths and heights are shown in Figure 4. Figure 4 also shows bar diagrams of the number-averaged lengths and heights for the fibrils. Average values of fibril lengths and heights, and the number of fibrils analyzed per condition are provided in Table 1. The histograms of fibril lengths are reported as the weight fraction of polypeptides incorporated in fibrils within a certain length range, viz., the weight of the incorporated polypeptides versus the total weight of the polypeptides in the condition. The fibrils have a very wide distribution of lengths; therefore, both the number and weight-averaged lengths are reported in Table 1 ($\langle L\rangle_n$ and $\langle L\rangle_w$).

Confirming our qualitative analysis of the AFM images, the quantitative analysis shows that C-S^Q_{10}-K_{12} indeed forms only short fibrils (with length <50 nm) in the absence of DNA (Figure 4b, top) and that in the presence of the DNA template, longer fibrils with a length of 151−250 nm are enriched (Figure 4b, bottom; also see Figure 4c). For the E^S_{80}-S^Q_{10}-K_{12} polypeptides, the length distribution is much broader in the absence of the template (Figure 4a, top); nevertheless, enrichment in longer fibrils is still observed in the presence of the DNA template (Figure 4a, bottom; also see Figure 4c). The fact that an enrichment in 151−250 nm fibrils is only observed in the presence of the DNA template indicates encapsulation of this template. For the E^S_{80}-S^Q_{10}-K_{12} polypeptide, the peak corresponding to encapsulated DNA occurs at somewhat smaller lengths (151−200 nm) than the expected length of around 230 nm, corresponding to a packing factor of around 3. This could indicate that the packing factor for this polypeptide is somewhat higher than that for the C-S^Q_{10}-K_{12} polypeptide. Further evidence of successful DNA encapsulation by the E^S_{80}-S^Q_{10}-K_{12} polypeptide is provided by the height

DOI: 10.1021/acs.biomac.9b00512
Biomacromolecules 2019, 20, 3641−3647

distributions, which show a shift to thicker fibrils in the presence of the DNA template (Figure 4d–f, Table 1). Replacing the collagen-like hydrophilic block with the elastin-like block thus preserves the fundamental features of VLP formation with DNA.

However, the fibril length distributions for the two polypeptides (Figure 4a–c and Table 1) also highlight an important difference. Very clearly, without a template, the E^S_{80}-S^Q_{10}-K_{12} fibrils formed in the absence of DNA have a length distribution that tails off at much larger lengths than the *C*-S^Q_{10}-K_{12} fibrils. Indeed, at 1.8 μM, the weight-averaged length in the absence of a template is $\langle L \rangle_w$ = 86 ± 19 nm for E^S_{80}-S^Q_{10}-K_{12}, whereas it is $\langle L \rangle_w$ = 34 ± 1 nm for *C*-S^Q_{10}-K_{12}. As a consequence, when co-assembling with DNA, the separation in length between fibrils encapsulating DNA and fibrils not encapsulating DNA is much more pronounced for the *C*-S^Q_{10}-K_{12} polypeptides (compare Figure 4a,b, bottom panels).

Origin of Differences in Self-Assembly is Caused by Differences in Amorphous Blocks. We here find that the amorphous block plays a crucial role in preventing non-templated fibril formation at low concentrations by opposing fibril formation at low concentrations. We also find that the efficiency by which the amorphous block prevents (non-templated) fibril formation at low concentrations critically depends on their chemical nature.

Both amorphous blocks studied here adopt a random coil conformation and have a length of about 400 amino acids, but the collagen-like block *C* much more strongly opposes fibril formation at low concentrations than the elastin-like block E^S_{80}. We speculate that the key difference between the two polypeptides that gives rise to this different behavior is their relative hydrophilicity. Increased hydrophilicity translates into larger random coil sizes of the amorphous blocks and stronger intermolecular repulsions between these more hydrated random coils that are densely packed along the assembled silk fibrils. This may be expected to lead to the observed prevention of nontemplated fibril formation at low concentrations.

Indeed, the collagen-like and elastin-like blocks have grand average of hydropathy (GRAVY) values[21] that are significantly different from each other. The GRAVY for E^S_{80} is 0.200, indicating that this amorphous block has an intermediate hydrophilicity and is close to its theta point, the transition point between a polymer in good solvent and a polymer in a poor solvent, consistent with the well-known aqueous demixing phase behavior exhibited by elastin-like polypeptides.[22] On the other hand, the collagen-like *C* polypeptide is extremely hydrophilic with a GRAVY value of −1.717.

In future work, it would therefore be interesting to change the nature of the guest residue X in the elastin-like polypeptide consensus motif GXGVP to obtain a polypeptide with a hydrophilicity approaching that of the collagen-like *C* block. The most hydrophilic amino acids, however, are charged residues, which may interfere with the electrostatic interactions that drive template binding. Therefore, the most promising candidates would be zwitterionic elastin-like polypeptides, for example, with sequences as having recently been investigated.[23]

CONCLUDING REMARKS

We have shown that the assembly- and co-assembly behaviors of the *C*-S^Q_{10}-K_{12} and E^S_{80}-S^Q_{10}-K_{12} triblock polypeptides are significantly different. We expect that differences between the chemical nature of amorphous blocks may similarly influence the assembly of a broad range of natural silk polypeptides and engineered silklike polypeptides.

As the much more hydrophilic collagen-like polypeptide *C* is better able to prevent self-assembly into fibrils at low concentrations than the much less hydrophilic elastin-like polypeptide E^S_{80}, we hypothesize that hydrophilicity of the amorphous block is a key variable in determining the assembly of natural silk polypeptides and engineered silklike polypeptides. Indeed, a critical role for hydrophilicity in controlling self-assembly is reported for many other polypeptide block copolymers, with increased hydrophilicity invariably reducing self-assembly.[24−26]

Such a role for the hydrophilicity of amorphous blocks in silklike polypeptides is also consistent with the observation that long hydrophilic blocks surrounding silklike self-assembly blocks may completely prevent fibril formation.[27]

An important conclusion is therefore that the chemical nature of the amorphous blocks, in particular their hydrophilicity, plays an important role in determining the assembly of both natural silks and engineered silklike polypeptides. This also means that it would be very interesting to more systematically study a broader range of amorphous structural polypeptides as amorphous blocks in the context of silklike polypeptides, to better understand the mutual influence of crystallizable and amorphous blocks on self-assembly. These amorphous blocks could include either natural ones such as resilin and abductin[28,29] or designed sequences such as PAS and XTEN.[30,31]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biomac.9b00512.

Expression of triblock polypeptide *C*-S^Q_{10}-K_{12} in *E. coli*: SDS-PAGE and Western blot (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: renko.devries@wur.nl.

ORCID

Enrico Mastrobattista: 0000-0002-6745-2015

Renko de Vries: 0000-0001-8664-3135

Author Contributions

L.W. designed, biosynthesized, and characterized the polypeptides, and performed AFM experiments. S.R. and I.W. contributed to biosynthesis of the polypeptides. L.v.W. contributed to AFM experiments. C.C.C. and A.H.-G. performed the work described in the Supporting Information. A.C. advised on polypeptide design, E.M. and J.v.d.O. contributed to the design of the project, R.d.V. designed the project, L.W. and R.d.V. wrote the manuscript, and all authors have edited the manuscript and have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

L.W. and R.d.V. acknowledge funding by the Dutch Science Organisation (NWO), grant 712.014.003.

DOI: 10.1021/acs.biomac.9b00512
Biomacromolecules 2019, 20, 3641−3647

REFERENCES

(1) Chow, D.; Nunalee, M. L.; Lim, D. W.; Simnick, A. J.; Chilkoti, A. Peptide-Based Biopolymers in Biomedicine and Biotechnology. *Mater. Sci. Eng., R* **2008**, *62*, 125−155.

(2) Carter, N. A.; Geng, X.; Grove, T. Z. Design of Self-Assembling Protein-Polymer Conjugates. In *Protein-based Engineered Nanostructures*; Cortajarena, A.; Grove, T., Eds.; Springer, 2016; Advances in Experimental Medicine and Biology, Vol. *940*; pp 179−214.

(3) Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L. Protein-Based Block Copolymers. *Biomacromolecules* **2011**, *12*, 269−289.

(4) Hardy, J. G.; Romer, L. M.; Scheibel, T. R. Polymeric Materials Based on Silk Proteins. *Polymer* **2008**, *49*, 4309−4327.

(5) Vepari, C.; Kaplan, D. L. Silk as a Biomaterial. *Prog. Polym. Sci.* **2007**, *32*, 991−1007.

(6) Hiew, S. H.; Miserez, A. Squid Sucker Ring Teeth: Multiscale Structure-Property Relationships, Sequencing, and Protein Engineering of a Thermoplastic Biopolymer. *ACS Biomater. Sci. Eng.* **2017**, *3*, 680−693.

(7) Hernandez-Garcia, A.; Kraft, D. J.; Janssen, A. F. J.; Bomans, P. H. H.; Sommerdijk, N. A. J. M.; Thies-Weesie, D. M. E.; Favretto, M. E.; Brock, R.; de Wolf, F. A.; Werten, M. W. T.; van der Schoot, P.; Cohen Stuart, M.; de Vries, R. Design and Self-Assembly of Simple Coat Proteins for Artificial Viruses. *Nat. Nanotechnol.* **2014**, *9*, 698−702.

(8) Schor, M.; Martens, A.; DeWolf, F. A.; Cohen Stuart, M.; Bolhuis, P. G. Prediction of Solvent Dependent β-Roll Formation of a Self-Assembling Silk-like Protein Domain. *Soft Matter* **2009**, *5*, 2658−2665.

(9) Schor, M.; Bolhuis, P. G. The Self-Assembly Mechanism of Fibril-Forming Silk-Based Block Copolymers. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10457−10467.

(10) Zhao, B.; Cohen Stuart, M. A.; Hall, C. K. Dock 'n Roll: Folding of a Silk-Inspired Polypeptide into an Amyloid-like Beta Solenoid. *Soft Matter* **2016**, *12*, 3721−3729.

(11) Zhao, B.; Cohen Stuart, M. A.; Hall, C. K. Navigating in Foldonia: Using Accelerated Molecular Dynamics to Explore Stability, Unfolding and Self-Healing of the β-Solenoid Structure Formed by a Silk-like Polypeptide. *PLoS Comput. Biol.* **2017**, *13*, No. e1005446.

(12) Werten, M. W. T.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. Secreted Production of a Custom-Designed, Highly Hydrophilic Gelatin in *Pichia Pastoris*. *Protein Eng., Des. Sel.* **2001**, *14*, 447−454.

(13) Hassouneh, W.; Christensen, T.; Chilkoti, A. Elastin-like Polypeptides as a Purification Tag for Recombinant Proteins. In *Current Protocols in Protein Science*; John Wiley & Sons Inc.: Hoboken, NJ, 2010; pp 6.11.1−6.11.16.

(14) Schipperus, R.; Eggink, G.; De Wolf, F. A. Secretion of Elastin-like Polypeptides with Different Transition Temperatures by *Pichia Pastoris*. *Biotechnol. Prog.* **2012**, *28*, 242−247.

(15) McDaniel, J. R.; MacKay, J. A.; Quiroz, F. G.; Chilkoti, A. Recursive Directional Ligation by Plasmid Reconstruction Allows Rapid and Seamless Cloning of Oligomeric Genes. *Biomacromolecules* **2010**, *11*, 944−952.

(16) Willems, L.; Roberts, S.; Weitzhandler, I.; Chilkoti, A.; Mastrobattista, E.; van der Oost, J.; de Vries, R. Inducible Fibril Formation of Silk−Elastin Diblocks. *ACS Omega* **2019**, *4*, 9135−9143.

(17) Cazares-Vargas, E.; Cohen Stuart, M. A.; de Vries, R.; Hernandez-Garcia, A. Template-Free Self-Assembly of Artificial De Novo Viral Coat Proteins into Nanorods: Effects of Sequence, Concentration, and Temperature. *Chem. - Eur. J.* **2019**, *25*, 11058−11065.

(18) Usov, I.; Mezzenga, R. FiberApp: An Open-Source Software for Tracking and Analyzing Polymers, Filaments, Biomacromolecules, and Fibrous Objects. *Macromolecules* **2015**, *48*, 1269−1280.

(19) Hirel, P. H.; Schmitter, M. J.; Dessen, P.; Fayat, G.; Blanquet, S. Extent of N-Terminal Methionine Excision from *Escherichia Coli* Proteins Is Governed by the Side-Chain Length of the Penultimate Amino Acid. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8247−8251.

(20) Marchetti, M.; Kamsma, D.; Cazares-Vargas, E.; Hernandez-García, A.; van der Schoot, P.; de Vries, R.; Wuite, G. J. L.; Roos, W. H. Real-time assembly of virus-like nucleocapsids elucidated at the single-particle level. *Nano Lett.* **2019**, 5746.

(21) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105−132.

(22) Hassouneh, W.; Zhulina, E. B.; Chilkoti, A.; Rubinstein, M. Elastin-like Polypeptide Diblock Copolymers Self-Assemble into Weak Micelles. *Macromolecules* **2015**, *48*, 4183−4195.

(23) Banskota, S.; Yousefpour, P.; Kirmani, N.; Li, X.; Chilkoti, A. Long Circulating Genetically Encoded Intrinsically Disordered Zwitterionic Polypeptides for Drug Delivery. *Biomaterials* **2019**, *192*, 475−485.

(24) Trabbic-Carlson, K.; Meyer, D. E.; Liu, L.; Piervincenzi, R.; Nath, N.; LaBean, T.; Chilkoti, A. Effect of Protein Fusion on the Transition Temperature of an Environmentally Responsive Elastin-like Polypeptide: A Role for Surface Hydrophobicity? *Protein Eng., Des. Sel.* **2004**, *17*, 57−66.

(25) Christensen, T.; Hassouneh, W.; Trabbic-Carlson, K.; Chilkoti, A. Predicting Transition Temperatures of Elastin-like Polypeptide Fusion Proteins. *Biomacromolecules* **2013**, *14*, 1514−1519.

(26) Weitzhandler, I.; Dzuricky, M.; Hoffmann, I.; Garcia Quiroz, F.; Gradzielski, M.; Chilkoti, A. Micellar Self-Assembly of Recombinant Resilin-/Elastin-Like Block Copolypeptides. *Biomacromolecules* **2017**, *18*, 2419−2426.

(27) Beun, L. H.; Storm, I. M.; Werten, M. W. T.; De Wolf, F. A.; Cohen Stuart, M. A.; De Vries, R. From Micelles to Fibers: Balancing Self-Assembling and Random Coiling Domains in pH-Responsive Silk-Collagen-like Protein-Based Polymers. *Biomacromolecules* **2014**, *15*, 3349−3357.

(28) Roberts, S.; Dzuricky, M.; Chilkoti, A. Elastin-like Polypeptides as Models of Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589*, 2477−2486.

(29) Rauscher, S.; Pomès, R. Structural Disorder and Protein Elasticity. In *Fuzziness: Structural Disorder in Protein Complexes*; Fuxreiter, M.; Tompa, P., Eds.; Springer: New York, NY, 2012; Vol. *1*, pp 59−183.

(30) Schlapschy, M.; Binder, U.; Börger, C.; Theobald, I.; Wachinger, K.; Kisling, S.; Haller, D.; Skerra, A. PASylation: A Biological Alternative to PEGylation for Extending the Plasma Half-Life of Pharmaceutically Active Proteins. *Protein Eng., Des. Sel.* **2013**, *26*, 489−501.

(31) Schellenberger, V.; Wang, C. W.; Geething, N. C.; Spink, B. J.; Campbell, A.; To, W.; Scholle, M. D.; Yin, Y.; Yao, Y.; Bogin, O.; et al. A Recombinant Polypeptide Extends the in Vivo Half-Life of Peptides and Proteins in a Tunable Manner. *Nat. Biotechnol.* **2009**, *27*, 1186−1190.

DOI: 10.1021/acs.biomac.9b00512
Biomacromolecules 2019, 20, 3641−3647