Self-Assembling Micelles Based on an Intrinsically Disordered Protein Domain

Sarah H. Klass,[a] Matthew J. Smith, [a] Tahoe A. Fiala, [a] Jess P. Lee, [a] Anthony O. Omole, [a] Kenneth H. Downing, [b] and Matthew B. Francis[a,b,c,*]

[a]Department of Chemistry, University of California, Berkeley, CA 94720, and [b]Lawrence Berkeley National Laboratory, Molecular Biophysics and Integrated Bioimaging Division, Structural Biology and Imaging Department, and [c]Materials Sciences Division, Berkeley, CA 94720.

KEYWORDS Intrinsically disordered proteins, self-assembly, micelles, surfactants, biologically derived materials

ABSTRACT: The self-assembly of micellar structures from di-block polymers that contain hydrophilic and hydrophobic domains has been of great interest for the encapsulation of drugs and other hydrophobic molecules. While most commercially used surfactants are derived from hydrocarbon sources, there have been recent efforts to replace these with biodegradable, non-toxic, biologically synthesized alternatives. Previous examples have primarily examined naturally occurring self-assembling-proteins, such as silk and elastin-like sequences. Herein, we describe a new series of fusion proteins that have been developed to self-assemble spontaneously into stable micelles that are 27 nm in diameter after enzymatic cleavage of a solubilizing protein tag. The sequences of the proteins are based on a human intrinsically disordered protein, which has been appended with a hydrophobic segment. The micelles were found to form across a broad range of pH, ionic strength, and temperature conditions, with critical micelle concentration (CMC) values below 1 μM being observed in some cases. The reported micelles were found to solubilize hydrophobic metal complexes and organic molecules, suggesting their potential suitability for catalysis and drug delivery applications. Furthermore, the inherent flexibility in the design of these protein sequences enables the encoding of additional functionalities for many future applications. Overall, this work represents a new biomolecular alternative to traditional surfactants that are based on non-renewable and poorly biodegradable hydrocarbon sources.

Introduction

It has been estimated that greater than 30% of eukaryotic proteins are disordered or contain disordered regions of 50 consecutive amino acids or more.[1] These intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDRs) lack the traditional secondary and tertiary structural motifs found in conventional proteins. Algorithms and databases developed for the purpose of identifying IDPs and IDRs have accelerated our understanding of the various roles this class of proteins can play in cellular regulation and disease.[2–5] Furthermore, the disordered nature of certain IDPs have been shown to be vital to the phase separation that is implicated in the formation of membrane-less organelles.[6–9] IDPs based on elastin-like protein (ELP) sequences that exhibit the ability to undergo thermo-responsive phase transitions have been exploited to develop genetically tailorable protein-based polymers for applications in drug delivery and protein-based materials that are naturally biodegradable.[10–12]

In contrast to the abundant examples of synthetically produced self-assembling peptides,[13–17] the recombinant expression of protein-based materials is more amenable to scale up and allows for precise genetic manipulation. To date, successful examples of using this approach to prepare micellular materials have started with naturally occurring motifs that are already known to self-assemble, including the formation of vesicles from the sunflower protein oleosin[18] and the formation of hydrogels by leucine zipper proteins.[19] Most notably, the self-assembly of the naturally phase separating elastin-based IDPs can be controlled by appending these sequences with hydrophobic-rich,[20–23] or post-translationally modified hydrocarbon tails.[11] These proteins have been shown to form a variety of temperature-sensitive structures (including micelles),[21,24] and they have been evaluated for their drug delivery potential *in vivo*.[25,26]

However, given their great abundance and chemical diversity, IDPs remain an underutilized source of biologically derived and biodegradable polymers, with the majority of research being conducted on a small number of sequences that have shown utility in the development of protein-based self-assembled materials.[9,12,23,27] In an effort to expand the sequence space of self-assembling IDPs, we have explored methods to introduce self-assembly, through genetic modification, to a novel IDP sequence that normally shows no phase separation or aggregation behavior.

Herein, we describe the construction and characterization of a new class of self-assembling proteins based on an intrinsically disordered sequence derived from the human neurofilament heavy arm side chain (NF-H) protein.[28] *In vivo* the NF-H domain of interest, referred to as IDP herein, forms extended protrusions that flare out from the core regions of cylindrical fibers, providing mechanical integrity to the axons of neurons and acting as a regulator of neurofilament transport.[29–32] These disordered protein regions involve a characteristic repeat sequence that can be generalized as "SPAE(A/V)K", which has been found to expand and contract in response to pH, phosphorylation, and ionic strength conditions.[33,34] These sequences do not assemble or aggregate on their own, instead behaving more like unstructured organic polymers. As a result, there is significant interest in using this sequence to create "smart biomaterials" in the form of protein-based environmentally

sensitive polymer brushes.[35,24] By designing and adding a hydrophobic domain to the C-terminus of the IDP sequence, we sought to generate an IDP-based di-block copolymer that, through entropic and enthalpic forces, undergoes self-assembly in aqueous solutions, **Figure 1**.

In an effort to develop and characterize a new class of self-assembling proteins, we report an engineered protein construct that allows facile expression and purification, with the potential for scalable production. Upon cleavage of an initial soluble domain, the construct self-assembles into micelles that are highly stable with respect to temperature, pH, and ionic strength conditions. Micelle structures were observed at sub-micromolar concentrations, and the ability of these constructs to solubilize hydrophobic cargo molecules in their core regions was demonstrated. These IDP-based micelles provide tunable, bio-renewable, biodegradable, protein-based surfactants with similar properties to commonly used surfactants found in detergents, drug delivery systems, and encapsulating agents that are traditionally derived from chemically modified hydrocarbons which pose a great health and environmental impact.[36,37] Furthermore, this approach shows promise for the development of new protein-based materials from previously overlooked sequences that may pose new or interesting biomaterials properties.

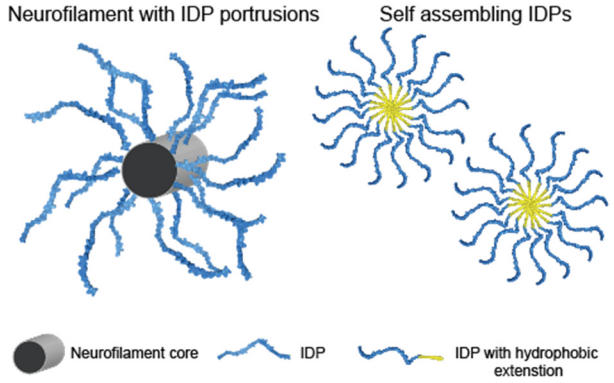

Figure 1. Inspiration for the design of self-assembling IDPs. (Left) IDP proteins derived from neurofilament heavy arm side chains extend as hair-like protrusions from a cylindrical protein core. (Right) by genetically encoding a hydrophobic domain at the C-terminus of the IDP sequence, self-assembly around an artificial hydrophobic core is achieved.

Results and Discussion

Design of Sequences. A major challenge in the recombinant expression of new amphiphilic proteins is that, by their very nature, these constructs are prone to assembling into higher ordered systems. Aggregation of the proteins during expression can lead to truncation due to premature ribosome departure and toxicity to host cells. Self-assembly properties can also lead to difficulties during purification, and the highly disordered nature of these proteins can result in increased protease susceptibility.[38] To allow the expression and purification of a fully-soluble protein, our construct consisted of a 161 amino acid segment of the human neurofilament IDP region that was fused to the C-terminus of maltose binding protein (MBP) with an N-terminal 6xHis tag (**Figure 2**). A thrombin cleavage site was positioned between the two protein segments. The MBP region both enhanced the solubility of the protein constructs during expression and purification and increased the overall production yields.[39]

The IDP sequence itself was mostly composed of 25 repeats of the amino acid sequence SPAEAK (see Supporting Information for specific sequences). To the C-terminus of this region was added a YWCA sequence to allow fluorescence detection of the peptides and the potential for chemical labeling with maleimides. To engender self-assembly properties, some constructs were also appended with the hydrophobic segments shown in **Figure 2b,c**:

The self-assembling regions of these peptides are referred to as "2Yx#A" herein, where # = 2,3,4. A higher number indicates a longer hydrophobic sequence.

Construction of Plasmids and Expression of Proteins. The IDP segment was initially prepared as a fusion to a 6xHis-MBP protein and was constructed using two gene blocks due to cloning difficulties arising from its repetitive sequence. The generated MBP-IDP plasmid served as the basis for all additional constructs. The MBP-IDP-2Yx(2-4)A plasmids were constructed by performing overhang PCR on the MBP-IDP (or subsequent) plasmid.

Pure monomeric IDP protein was obtained in a similar manner to IDP-2Yx(2-4)A proteins, as reported in Supporting information. Characterization using DLS revealed an average diameter of 11.3±0.8 nm, which was consistent with literature reports.[35] The MBP-IDP-2Yx(2-4)A constructs were expressed in Rosetta (DE3) pLysS competent cells, as this strain has been shown to stabilize

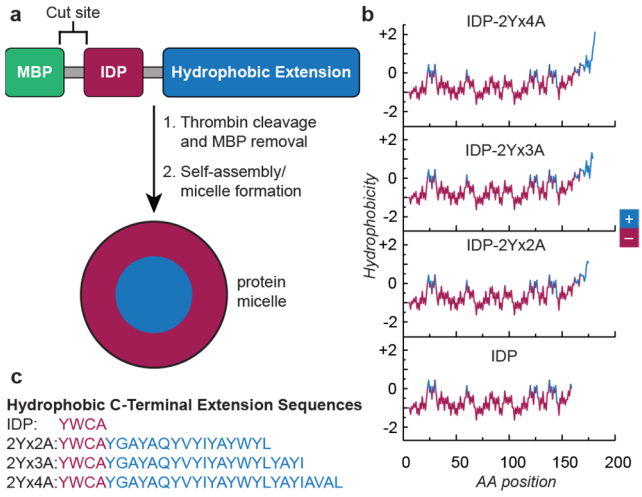

Figure 2. Design of an amphiphilic protein construct. (a) An intrinsically disordered protein (IDP) segment is fused to a hydrophobic sequence. Following cleavage of the MBP protein, the amphiphilic portion self-assembles. (b) Hydrophobicity plots of the designed sequences are shown, following cleavage of the MBP regions. The values are from the Kyte-Doolittle hydrophobicity scale[40] with a window size of 9. Values greater than 0 indicate a hydrophobic region (blue) while those less than zero are hydrophilic (red). The plots were generated using the Expasy ProtScale tool (https://web.expasy.org/protscale). (c) The specific hydrophobic sequence regions are shown for the constructs used in this report. The c-terminal residues of IDP are shown in red while the hydrophobic extensions are in blue.

pET recombinants that encode proteins that can affect cell growth and viability. This cell line also improves the performance of rarely used *E. coli* codons through the additions of non-native tRNA genes on a separate plasmid.[41]

Following cell lysis, the constructs were purified by nickel nitroloacetic acid (Ni-NTA) affinity chromatography owing to the

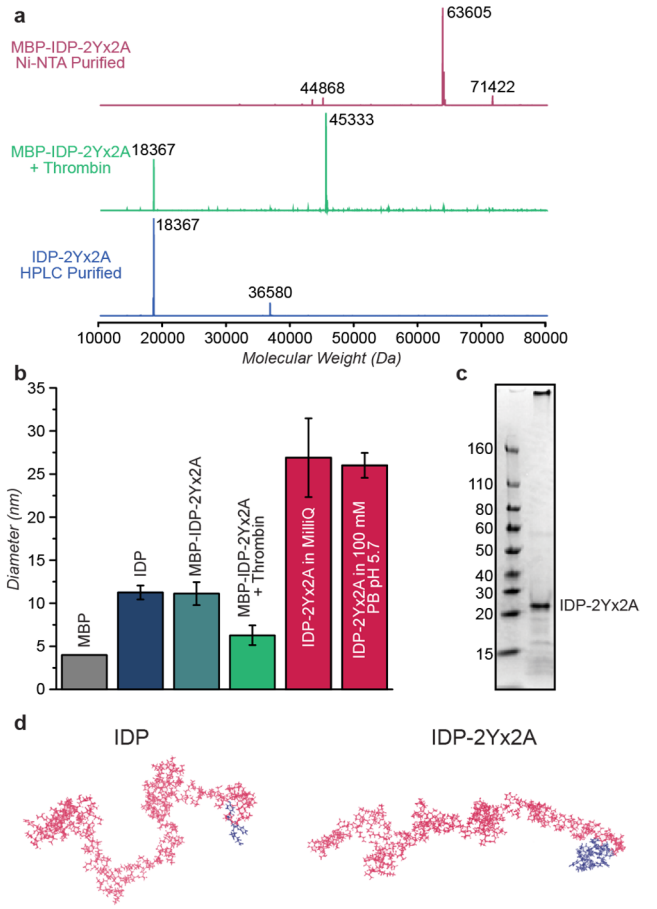

Figure 3. Analysis of the protein constructs. (a) LCMS analyses are shown for MBP-IDP-2Yx2A (red), IDP-2Yx2A after MBP removal with thrombin (green), and IDP-2Yx2A after HPLC purification (blue). The MBP-IDP-2Yx2A expected molecular weight is 63605 Da. In the IDP-2Yx2A spectra, the cysteine residue is capped as a disulfide with beta-mercaptoethanol, corresponding to an expected molecular weight of 18366 Da. The IDP-2Yx2A disulfide dimer appears at 36580 Da in the bottom spectrum. (b) DLS analysis of IDP-2Yx2A throughout the purification process shows changes in the diameters of the constructs. Prior to cleavage, the MBP-IDP-2Yx2A protein has a similar diameter to that of IDP itself, and is presumably monomeric. Following thrombin cleavage, a diameter between that of the MBP monomer and IDP is observed. After removal of the MBP, a sharp increase in the average diameter is observed, indicating self-assembly in both pure water and in buffer (phosphate, pH 5.7). The polydispersity indices (PDIs) of the IDP-2Yx2A samples above the CMC are generally between 0.2-0.5, depending on the buffer conditions (see Supporting Information Figure S4 for full DLS traces for IDP and IDP-2Yx2A). (c) A Coomassie-stained SDS page gel of HPLC purified IDP-2Yx2A is shown, indicating >95% purity. (d) MD simulations of IDP and IDP-2Yx2A at 50 ns. Residues in red indicate the IDP portion of each construct, blue residues indicate the hydrophobic c-terminal residues.

binding affinity of the 6xHis group encoded on the MBP portion of the fusion protein. By LC-MS and SDS-PAGE analysis, the MBP-IDP-2Yx2A constructs were found to express well and with high purity following affinity purification (Supporting Information Figure S3).

By DLS, MBP-IDP-2Yx2A showed a similar diameter to IDP (11.1±1.3 nm), indicating that it was still monomeric at this stage. Constructs MBP-IDP-2Yx3A and MBP-IDP-2Yx4A, which only differ from MBP-IDP-2Yx2A in the addition of 4 and 8 hydrophobic amino acids, respectively, resulted in lower bacterial titers. Additionally, for MBP-IDP-2Yx3A and MBP-IDP-2Yx4A, smaller cell pellets were observed after expression, soapy foams formed upon sonication, and truncations were observed in the LC-MS traces following NiNTA purification (Supporting Information Figure S1). Additionally, a DLS hydrodynamic diameter of 36±15 nm was observed for MBP-IDP-2Yx3A and a diameter of 115±6 nm was observed for MBP-IDP-2Yx4A (Supporting Information Figure S2). Taken together, it appears that the MBP-IDP-2Yx3A and MBP-IDP-2Yx4A constructs are aggregation-prone prior to MBP cleavage, making them more difficult to express and purify.

MBP cleavage and purification of the IDP-2Yx2A constructs. Following Ni-NTA purification, MBP-IDP-2Yx2A was cleaved using thrombin and purified by preparative scale reversed phase chromatography to ensure high purity for this study; however, ion exchange chromatography could also be used to purify the final construct on a larger scale (albeit at slightly lower purity, see Supporting Information Figure S3). The resulting protein was stored as a solid following lyophilization. For use, the dry powder was resuspended in MilliQ water by vortexing and sonicating prior to the addition of the desired buffer.

Following MBP cleavage and protein purification, a hydrodynamic diameter of 26.9 ± 4.6 nm was observed in MilliQ water (**Figure 3b**). This size correlates with micelle-like structures with an 11 nm hydrophilic IDP portion and a 2.5 nm hydrophobic interior, as would be expected given the apparent size of the IDP segment alone by DLS (11.1 nm) plus a 16 amino acid hydrophobic extension if an alpha-helical configuration is assumed (2.4 nm, 1.5Å per residue). The two constructs were further analyzed using molecular dynamic (MD) simulations. After running the simulations for 50

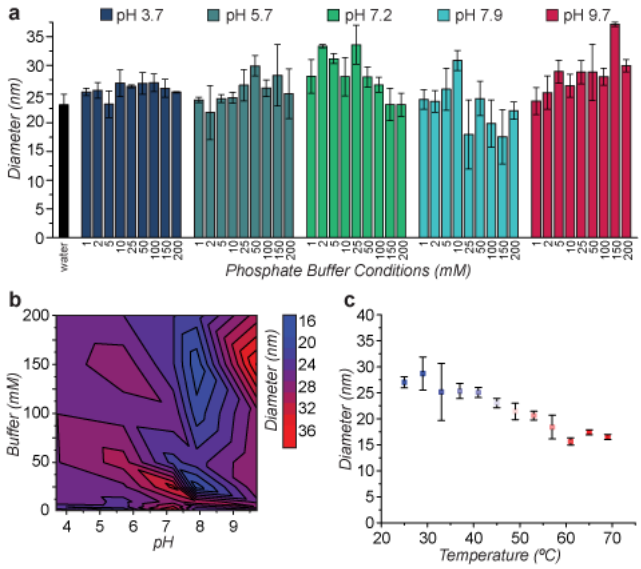

Figure 4. pH and temperature effects on micelle size, as measured by DLS. (a) DLS measurements of IDP-2Yx2A micelles were performed under varying pH and buffer conditions. Over all pH values and salt concentrations examined, the average diameter was 26.2±4.3 nm. (b) The diameter changes occurring upon pH and salt variation are summarized as a topographical plot. The largest variability in diameter is observed as the pH changes from 7 to 8, but disassembly was not observed. (c) The influence of temperature on micelle size is determined by DLS. As the temperature increases, the micelle diameter decreases from 27.0±1.1 nm to 16.5 ± 0.5 nm.

ns, both constructs were similar to the sizes measured by DLS for a single monomer. Between 40 and 50 ns of simulation, the radius of gyration (Rg) fluctuated between 3-4 nm for IDP and 4-5 nm for 2Yx2A (Supporting Information Figure S10). In both constructs, the IDP portion showed no indication of collapse over the 50 ns simulation time (Supporting Information movie 1 and 2).

Effects of salt, pH, and temperature on IDP-2Yx2A micelles. The effects of changes to salt concentration and pH were evaluated for the protein assemblies using DLS. Structures consistent with micelles were observed at pH values ranging from 3.7 -9.7 and salt concentrations of 0-200 mM K_xPO_4 (**Figure 4a**), NaCl, and $CaCl_2$ (Supporting Information Figure S5). At most pH values, the micelle sizes were consistent at ~27 nm with low polydispersity. However, at pH 7.9, significant heterogeneity and polydispersity in the micellar diameters were observed (**Figure 4a,b**). Interestingly, this pH corresponded to the most collapsed state of the hydrophilic IDP portion, as determined in previous studies.[35,33] This may affect the self-assembly behavior of IDP-2Yx2A, and was a property we sought to explore further.[35]

To examine the thermal stability of the micelles, DLS measurements of IDP-2Yx2A were taken upon heating from 25 to 70 °C. As the temperature increased, the average diameter and polydispersity of the sample decreased, **Figure 4c**. No precipitation was observed, and the size change was found to be reversible upon cooling. Similar trends are reported in the literature for conventional SDS micelles and casein micelles, where the micellar volume decreases with increasing temperature.[42,43] This is likely due to factors including the hydrophobic tail packing and the interactions between water and the hydrophilic headgroups.

Determination of the critical micelle concentration (CMC). As polymeric micelles often find use in biomedical[44] and aquatic environments[45], we sought to explore their stability in buffers that reflect the pH of blood (1xPBS, pH 7.4) and rainwater (phosphate buffer, pH 5.7, 100 mM; see Supporting Infomration S1 for specific buffer compositions).[46] Herein these buffers will be referred to as pH 7.4 buffer and pH 5.7 buffer, respectively.

The CMC values of IDP-2Yx2A in these buffers was determined using fluorescence spectroscopy. The emission ratios of the first and third vibronic bands of pyrene were measured at varying concentrations of protein at room temperature. The I1/I3 ratio is known to increase with increasing polarity of the probe environment.[14,47] At high concentrations, an I1/I3 ratio of 0.9-1.0 was observed in the IDP-2Yx2A samples for both buffers. This indicated that the medium surrounding the pyrene was more hydrophobic than when in phosphate buffer alone (I1/I3 = 1.33) In contrast, no change in the I1/I3 band ratio was observed for the IDP protein lacking the 2Yx2A sequence at any concentration in the pH 5.7 buffer (I1/I3 = 1.33-1.36 over a range of 1 μM to 100 μM).

The I1/I3 ratios of the IDP-2Yx2A samples in both buffers were fit to a logistic function (R^2 for pH 7.4 = 0.97; R^2 for pH 5.7=0.97), with EC_{50} values of 26 μM for pH 7.4 buffer and 13 μM in pH 5.7 buffer. However, because the CMC was so low, and the final concentration of pyrene must be maintained at 2 μM to obtain good quality fluorescence spectra, this assay can only provide an upper limit of the CMC.[48] A more accurate determination of CMC may be simply the emergence of the I3 band, which indicates a CMC of 2.5 μM (the lowest concentration at which a distinct I3 band was observed) (**Figure 5**). This value is reflected by the EC_{20} of the logistic fit. From the EC_{20}, IDP-2Yx2A CMC values were determined to be 17.5 μM in pH 7.4 buffer and 2.6 μM in pH 5.7 buffer. These results again agree with the reported behavior of the IDP sequence at near-neutral pH, where the sequence is known to be in its

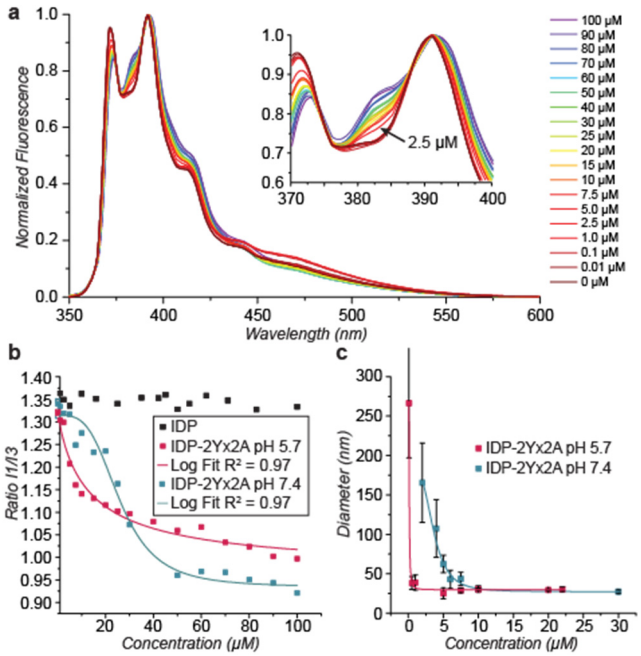

Figure 5. CMC evaluations at pH 5.7 and pH 7.4. (a) Fluorescence traces of pyrene with 0-100 μM IDP-2Yx2A in pH 5.7 buffer are shown. An enhanced *I3* fluorescence band of pyrene is observed at concentrations as low as 2.5 μM. (b) Pyrene fluorescence was similarly evaluated for IDP-2Yx2A in pH 7.4 buffer and for the non-assembling IDP in pH 5.7 buffer. The ratios of the first and third vibronic bands (*I1*/*I3*) are plotted and fit to logistic functions to determine upper limits of the CMC values. Both IDP-2Yx2A samples show enhancement of the pyrene *I3* band, indicating micelle formation. IDP-2Yx2A has a lower CMC value in pH 5.7 buffer than in pH 7.4 buffer, with EC_{20} values of 2.6 μM and 17.5 μM, respectively. No *I3* band enhancement was observed for IDP at any concentration. (c) DLS analysis of IDP-2Yx2A micelles is shown at decreasing concentrations in pH 5.7 buffer and pH 7.4 buffer.

most collapsed state.[35] This can be rationalized by a balancing of the glutamic acid and lysine residue charges, causing an overall decrease in the charge of the hydrophilic domain (and therefore a decrease in hydrophilicity). This results in a less stable assembly. At pH 5.7, side chain protonation results in a net positive charge (increased hydrophilicity) for the hydrophilic domains and thus promotes a more stable and extended assembly.

These results were also compared to a dilution study performed using dynamic light scattering under the same buffer conditions. By DLS, the micelle size (~27 nm) and low polydispersity were maintained at concentrations above 10 μM in both buffers. As the concentration decreased further, the micelles swelled and became less uniform at pH 7.4. In agreement with the fluorescence data, the samples in pH 5.7 buffer showed consistent diameters and low polydispersities down to 0.5 μM. It should be noted that the data point at 0.1 μM in pH 5.7 buffer, where a dramatic increase in size is observed, occurs near the detection limit of the DLS instrument.

Validation of size, shape, and CMC by Cryo-TEM and SAXS. Cryogenic transmission electron microscopy (Cryo-TEM) and Small Angle X-ray Scattering (SAXS) analysis confirmed the size, shape, and CMC of the micelles. SAXS analysis was conducted on IDP-2Yx2A in pH 7.4 buffer and in pH 5.3 buffer, compared to IDP itself as a non-assembling control. The SAXS intensity curves showed a clear difference between the IDP-2Yx2A samples and the

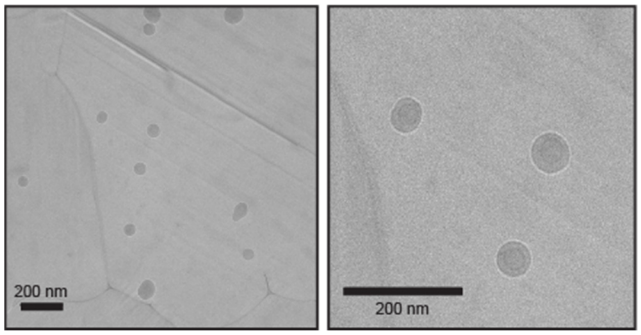

Figure 6. Cryo-TEM analysis of a 0.4 μM suspension of IDP-2Yx2A micelles in pH 5.3 buffer. Image analysis using ImageJ revealed an average diameter of 50.5±12.1 nm. This value is consistent with the hydrodynamic diameter of 48.4±10.6 nm for 0.4 μM samples measured by DLS.

IDP sample, with both IDP-2Yx2A samples showing a dramatic increase in the low q region. This indicated the presence of large particles. Unfortunately, due to the high concentration of protein required for SAXS analysis, a monodisperse core-shell model cannot be applied to the data. We hypothesize that this could be due to inter-micelle interactions, leading to a small fraction of larger and elongated particles in the sample at higher concentrations that are not observed by Cryo-TEM. However, a lower limit (excluding the presence of larger aggregates) radius of gyration (R_g) and radius of the cross-section (R_c) can be calculated. For the IDP-2Yx2A samples, R_g values obtained from the Guinier analysis were greater than 9.0 nm, while the R_c values were between 3.6 and 5.2 nm. For non-assembling IDP the R_g was calculated to be greater than 4.0 nm and an R_c value was not observed. This indicates a clear difference in the assembly states of IDP and IDP-2Yx2A (Supporting Information Figure S6).

The spherical shapes of the micelles ware confirmed by Cryo-TEM. At a concentration of 0.4 μM IDP-2Yx2A, spherical micellar structures were observed. The low concentration of particles resulted in slightly larger diameters and more polydispersity (see Figure 5). This was confirmed by DLS prior to Cryo-TEM analysis, where the IDP-2Yx2A samples used for Cryo-TEM showed an average diameter of 48.4±10.6 nm (Supporting Information Figure S7). Cryo-TEM images were obtained by pre-exposing the grid to photons prior to image acquisition to remove ice crystals from the samples. Embedded in the remaining vitrified ice, spherical micelles were observed (**Figure 6**). Image analysis using ImageJ revealed an average diameter of 50.5±12.1 nm. Additionally, in some particles, a core-shell like structure could be observed. We hypothesize that this is due to the hydration differences between the micelle core and shell. This should be especially prominent at low concentrations, at which the micelles swell.

Solubilization of hydrophobic molecules and chemical modification of the micelle exterior. A key application of surfactants and micelles stems from their ability to solubilize hydrophobic molecules in aqueous solution. To evaluate the ability of IDP-2Yx2A micelles to do this, three molecules with unique functions were chosen. The antifungal agent pyraclostrobin, with a solubility of 1.9 mg/L in water, was first loaded into the micelle structure by combining both pyraclostrobin and lyophilized IDP-2Yx2A in a small volume of THF, followed by dilution with pH 5.7 buffer. Any remaining insoluble pyraclostrobin was removed by centrifugation, and the THF was removed under vacuum. Following an additional round of resuspension in water and subsequent centrifugation, the supernatant was diluted 1:1 with acetonitrile before HPLC analysis to quantify the amount of pyraclostrobin that had been solubilized (Supporting Information Figure S8). In the presence of IDP-2Yx2A, substantially more pyraclostrobin was solubilized than occurred in buffer alone (Figure 7a). Based on UV measurements, 15.2±8 molecules of pyraclostrobin were loaded per protein monomer. Assuming each IDP-2Yx2A molecule is a cone of with a length of 13.5 nm and a base radius of 1.25 nm, it can be estimated that each 27 nm micelle sphere contains about 470 protein molecules. This suggests that thousands of pyraclostrobin molecules are contained within each micelle.

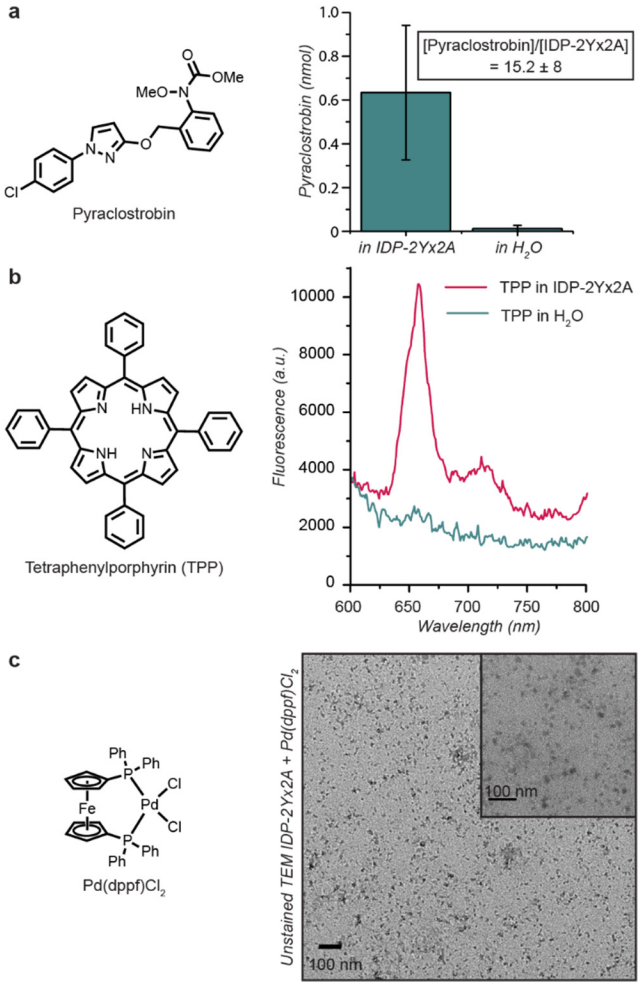

Figure 7. Evaluation of small molecule loading into IDP-2Yx2A protein micelles. (a) Pyraclostrobin shows substantially greater solubility (0.6±0.3 nmol) in 3 μM IDP-2Yx2A solution vs. water alone (0.01± 0.01 nmol). The molar ratio of pyraclostrobin to IDP-2Yx2A protein monomers is 15.2±8. (b) The loading of tetraphenyl-porphyrin (TPP) for singlet oxygen generation was evaluated by fluorescence emission. When solubilized in the hydrophobic environment of IDP-2Yx2A, the characteristic emission spectra of TPP was observed. No material dissolved in water alone. (c) Unstained TEM images of hydrophobic Suzuki coupling reagent $Pd(dppf)Cl_2$ show loading into IDP-2Yx2A micelles. The heavy metals Pd and Fe provide sufficient contrast to image the micelle interior, which have a measured diameter averaging 15±8 nm (over 4000 particles analyzed).

As a second cargo type, a singlet oxygen-generating tetraphenyl-porphyrin (TPP) molecule used for photodynamic therapy (PDT) was evaluated similarly. TPP is completely insoluble in water and

thus is non-fluorescent. However, upon incubation with IDP-2Yx2A, characteristic emission peaks around 650 and 715 nm were observed (Figure 7b). Finally, a Suzuki coupling catalyst ($Pd(dppf)Cl_2$) with poor water solubility was added to IDP-2Yx2A micelles. Loading was then analyzed by TEM without additional staining. Consistent with the catalyst being concentrated in the interior of the micelle structures, small, spherical concentrations of metal complexes were observed at the expected sizes of 15±8 nm (Figure 7c). This result indicates that upon the loading of this hydrophobic catalyst precursor, the spherical micelle-like shape of the particles is maintained and does not undergo aggregation. Furthermore, this demonstrates potential applications in the micellar catalyst and the templating of metal nanoparticles.

Conclusion

Developing biologically-derived and biodegradable surfactants could help to alleviate the environmental effects of detergents that have been traditionally hydrocarbon-based. The move to replace these compounds with synthetic peptide-based amphiphiles has resulted in many highly successful self-assembling systems; however, their syntheses typically involve the use of organic solvents and the costs associated with their production can limit scale-up potential. The new IDP-based proteins described herein offer high yields using inexpensive bacterial expression systems, and exhibit CMC values that are lower than many other types of surfactants. They offer excellent thermostability and remain assembled across a broad range of pH, ionic strength, and temperature conditions. These features suggest that, even in their current form, they can find use in a number of industrial applications. The ability to house hydrophobic cargo molecules, along with the overall nanoscale dimensions, open the door to applications in targeted drug delivery. Perhaps the greatest opportunity is the potential for tuning the sequences to obtain desired solubilizing, assembling, and response properties, as every position along the polypeptide chain can be specified on the genetic level. We are currently exploring the potential of this promising class of surfactants for use as paint emulsifiers, antifungal dispersants, and bioremediation materials.

Experimental Procedures

Expression and purification MBP-IDP. MBP-IDP plasmids were prepared as described in the Supporting Information and transformed into *E. coli* BL21(DE3) competent cells. Starter cultures (20 ml Lauria Broth (LB, VWR Life Sciences), including 50 mg/L kanamycin) were grown from single colonies, grown overnight at 37 °C, and used to inoculate 1 L of Terrific Broth (TB. Sigma) media that contained 50 mg/L kanamycin. Cultures were grown to an OD ~ 0.5, cooled for 20 min at 16 °C, induced with 0.5 mM IPTG, and expressed (~ 18 h) at 25 °C. Cells were harvested by centrifugation for 15 min at 4,000*x*g at 4 °C. The resulting pellet was lysed in 30 mL of lysis buffer (20 mM HEPES, pH = 7.5, 300 mM NaCl, 10 mM β-mercaptoethanol, 10 mM imidazole = buffer A) supplemented with one tablet of EDTA-free SigmaFast Protease Inhibitor (Sigma Aldrich), 2 mM PMSF, and 10 mg of lysozyme. The resuspended sample was lysed by sonication (amplitude 50%, 2:4 seconds on:off for 10 min), followed by 20 min of centrifugation at 24,000*x*g at 4 °C. The supernatant was filtered through a 40 μm Steriflip filter (Millipore), and loaded onto a 5 ml NiNTA column (Protino, Machery Nagel) connected to an Akta purifier that was pre-equilibrated with buffer A. The column was washed with 50 mL (10 column volumes) of 20 mM HEPES (pH = 7.5), 300 mM NaCl, 10 mM βMe, and 10 mM imidazole. The protein was eluted with 20 mM HEPES (pH = 7.5), 300 mM NaCl, 10 mM β-mercaptoethanol, and 250 mM of imidazole. Imidazole was removed by spin concentration with 20 mM HEPES (pH=7.5) buffer containing 100 mM NaCl.

Expression and Purification IDP-2Yx(2-4)A. Plasmids were transformed into *E. coli* Rosetta 2 pLys competent cells. Starter cultures in 20 mL of LB with 50 mg/L kanamycin were grown overnight at 37 °C from single colonies and used to inoculate 1 L of TB media containing 50 mg/L kanamycin. Cultures were grown to an OD ~ 0.5, cooled for 20 min at 16 °C, induced with 0.1 mM IPTG, and expressed for about 6 h at 16 °C. Cells were harvested by centrifugation for 15 min at 4,000xg at 4 °C. The pellet was transferred to a 50 ml Falcon tube in PBS buffer, and spun down for 10 min at 4000 rcf (g). The resulting pellet (~5 g) was lysed in 30 ml of lysis buffer (20 mM HEPES, pH = 7.5, 300 mM NaCl, 10 mM imidazole = buffer A) supplemented with one tablet of EDTA-free SigmaFast Protease Inhibitor (Sigma Aldrich), 2 mM PMSF, and 10 mg lysozyme. The resuspended sample was lysed by sonication (amplitude 50%, 2:4 seconds on:off for 10 minutes) followed by 20 min of centrifugation at 24,000 rcf (g) at 4 °C. The supernatant was filtered through a 40 μm Steriflip filter (Millipore), and loaded onto a 5 ml NiNTA column (Protino, Machery Nagel) connected to an Akta purifier that was pre-equilibrated with buffer A. The column was washed with 50 ml (10 CV) of 20 mM HEPES (pH = 7.5), 300 mM NaCl, 10 mM imidazole. The protein was eluted with 20 mM HEPES (pH = 7.5), 300 mM NaCl, 250 mM imidazole. Purity and identity of the recovered protein were analyzed by gel and LC/MS. Imidazole was removed by spin concentration with 20 mM HEPES (pH=7.5), 100mM NaCl.

MBP cleavage procedure. MBP-IDP and MBP-IDP-2Yx(2-4)A samples were subsequently digested with 1 mg of thrombin protease (high purity from Bovine, MP Biomedicals). Complete digestion was achieved at room temperature after 1 h, as confirmed by ESI-TOF LCMS.

Molecular dynamic simulations. MD simulations were carried out using the Desmond software package and run with the OPLS_2005 force field[49] available through Maestro. Both 2Yx2A and IDP were modeled in an alpha helical conformation at time zero. Simulations were carried out at 300 K and a constant pressure of 1.01325 bar. Each structure was fully solvated with SPC water in a cube with an edge length of 20Å. Each simulation was run for a total of 50 ns.

Protein purification. The IDP sample was diluted to a final volume of 50 mL with 20 mM HEPES (pH = 7.5) containing 10 mM β-mercaptoethanol and loaded onto a 1 mL HiTrap HP cation exchange column connected to an Akta purifier (GE Healthcare). The column was washed with 10 column volumes of 20 mM HEPES (pH = 7.5) buffer containing 10 mM β-mercaptoethanol and eluted with 20 mM HEPES (pH = 7.5) buffer containing 10 mM β-mercaptoethanol and 1 M NaCl. The protein was >95% pure by SDS-PAGE gel electrophoresis and LCMS.The IDP-2Yx(2-4)A samples were purified using reversed phase purification on a Biotage chromatography system. To these samples were added 10% acetonitrile (ACN), and the resulting solutions were loaded onto a 10 g C18 Biotage SNAP Bio 300A reversed phase column that had been equilibrated with 10% ACN in H_2O + 0.1% TFA. The column eluted using a gradient of 10-100% ACN, with the desired product eluting around 40% ACN. The fractions containing IDP-2Yx2A were analyzed by ESI/TOF LC/MS for purity. Pure fractions were collected and lyophilized to dryness, yielding the protein product as a white powder. SDS-PAGE analysis with Coomassie staining indicated >95% purity.

DLS Analysis. DLS analysis was performed on a Malvern Instruments Zetasizer Nano ZS. Data plots and standard deviations were calculated from an average of three measurements, each of which consisted of 13 runs. Measurement data are presented as a diameter determined by the %Number distribution. Unless otherwise stated,

all measurements were taken at 25 °C after a 2 min period of temperature equilibration.

CMC determination by pyrene fluorescence. To each sample, a 2 μM solution of pyrene in the appropriate buffer was added, and the solutions were allowed to equilibrate for 5 min. Each protein solution was then diluted with an additional portion of a 2 μM pyrene solution in the appropriate buffer to keep the pyrene and salt concentrations constant, but decrease the protein concentration. Fluorescence emission spectra were collected on a Fluoromax-4 Spectrofluorometer (HORIBA Scientific) exciting at 335 nm with a 5 nm window and monitoring emission from 350-600 nm. The florescence emission intensities of the first and third vibronic bands of pyrene were recorded.

Cryogenic Transmission Electron Microscopy (Cryo-TEM). These experiments were run on a JEM 2100F electron microscope (JEOL, Tokyo, Japan) that was equipped with a field emission gun and a F224HD CCD camera (TVIPS, Gauting, Germany). The microscope was operated at a 200 kV accelerating voltage, and the specimen was cooled to −180 °C using a liquid-nitrogen-cooled cryoholder.

Resuspension of protein micelles with hydrophobic compounds. To load hydrophobic molecules into the IDP-2Yx2A micelles, the lyophilized protein power and a hydrophobic molecule of interest were first suspended in THF. The sample was then diluted with the appropriate aqueous buffer to the desired protein concentration. To remove hydrophobic molecules that wee not solubilized, the samples were centrifuged at 13.1*x*g for 1 min. The supernatants were then recovered, ransferred to clean 1.5 mL Eppendorf tubes, and centrifuged again. The supernatants were recovered and transferred to another set of clean 1.5 mL tubes, and the process was repeated for a total of three rounds of centrifugation. The sample was then lyophilized. The lyophilized samples were then reconstituted in MilliQ water, resulting in clear solutions. Even through hydrophobic molecules were observed to precipitate from solution, three rounds of centrifugation and supernatant removal were again performed to ensure that all insoluble material was removed from each solution. The amount of solubilized pyraclostrobin was then determined using HPLC, as detailed in the Supporting Information.

ASSOCIATED CONTENT

Supporting Information. Experimental details and further characterization experiments are provided as Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*mbfrancis@berkeley.edu

ACKNOWLEDGMENT

This work was supported by the BASF CARA program. The SAXS data analysis benefited from the use of the SasView application, originally developed under NSF award DMR-0520547. SasView contains code developed with funding from the European Union's Horizon 2020 research and innovation programmed under the SINE2020 project, grant agreement No 654000. http://www.sasview.org/.

A portion of this work was conducted at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional support comes from the National Institute of Health project MINOS (R01GM105404) and a High-End Instrumentation Grant S10OD018483.

ABBREVIATIONS

LB = Lauria Broth

TB = Terrific Broth

DOTA =1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid

2PCA = 2-pyridine carboxaldehyde

DLS = Dynamic Light Scattering

CMC = Critical micelle Concentration

IDP = Intrinsically Disordered Protein

MBP = Maltose Binding Protein

SAXS = Small Angle X-Ray Scattering

TPP = Tetraphenyl porphyrin

Ni-NTA = Nickel-nitrilotriacetic acid

REFERENCES

(1) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; et al. Intrinsically Disordered Protein. *J. Mol. Graph. Model.* **2001**, *19* (1), 26–59.

(2) Vucetic, S.; Obradovic, Z.; Vacic, V.; Radivojac, P.; Peng, K.; Iakoucheva, L. M.; Cortese, M. S.; Lawson, J. D.; Brown, C. J.; Sikes, J. G.; et al. DisProt: A Database of Protein Disorder. *Bioinformatics* **2005**, *21* (1), 137–140.

(3) Varadi, M.; Vranken, W.; Guharoy, M.; Tompa, P. Computational Approaches for Inferring the Functions of Intrinsically Disordered Proteins. *Front. Mol. Biosci.* **2015**, *2*, 45.

(4) Tompa, P. Intrinsically Disordered Proteins: A 10-Year Recap. *Trends Biochem. Sci.* **2012**, *37* (12), 509–516.

(5) Monastyrskyy, B.; Fidelis, K.; Moult, J.; Tramontano, A.; Kryshtafovych, A. Evaluation of Disorder Predictions in CASP9. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (S10), 107–118.

(6) Wright, P. E.; Dyson, H. J. Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (1), 18–29.

(7) Uversky, V. N. Intrinsically Disordered Proteins in Overcrowded Milieu: Membrane-Less Organelles, Phase Separation, and Intrinsic Disorder. *Curr. Opin. Struct. Biol.* **2017**, *44*, 18–30.

(8) Uversky, V. N.; Kuznetsova, I. M.; Turoverov, K. K.; Zaslavsky, B. Intrinsically Disordered Proteins as Crucial Constituents of Cellular Aqueous Two Phase Systems and Coacervates. *FEBS Lett.* **2015**, *589* (1), 15–22.

(9) Dzuricky, M.; Roberts, S.; Chilkoti, A. Convergence of Artificial Protein Polymers and Intrinsically Disordered Proteins. *Biochemistry* **2018**, *57* (17), 2405–2414.

(10) MacEwan, S. R.; Weitzhandler, I.; Hoffmann, I.; Genzer, J.; Gradzielski, M.; Chilkoti, A. Phase Behavior and Self-Assembly of Perfectly Sequence-Defined and Monodisperse Multiblock Copolypeptides. *Biomacromolecules* **2017**, *18* (2), 599–609.

(11) Mozhdehi, D.; Luginbuhl, K. M.; Simon, J. R.; Dzuricky, M.; Berger, R.; Varol, H. S.; Huang, F. C.; Buehne, K. L.; Mayne, N. R.; Weitzhandler, I.; et al. Genetically Encoded Lipid–polypeptide Hybrid Biomaterials That Exhibit Temperature-Triggered Hierarchical Self-Assembly. *Nat. Chem.* **2018**, *10* (5), 496–505.

(12) Simon, J. R.; Carroll, N. J.; Rubinstein, M.; Chilkoti, A.; López, G. P. Programming Molecular Self-Assembly of Intrinsically Disordered Proteins Containing Sequences of Low Complexity. *Nat. Chem.* **2017**, *9* (6), 509–515.

(13) Vauthey, S.; Santoso, S.; Gong, H.; Watson, N.; Zhang, S. Molecular Self-Assembly of Surfactant-like Peptides to Form Nanotubes and Nanovesicles. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (8), 5355–5360.

(14) Guler, M. O.; Claussen, R. C.; Stupp, S. I. Encapsulation of Pyrene within Self-Assembled Peptide Amphiphile Nanofibers. *J. Mater. Chem.* **2005**, *15* (42), 4507.

(15) Hartgerink, J. D.; Beniash, E.; Stupp, S. I. Self-Assembly and Mineralization of Peptide-Amphiphile Nanofibers. *Science* **2001**, *294* (5547), 1684–1688.

(16) Yuri Zimenkov, †; Steven N. Dublin, †; Rong Ni, †; Raymond S. Tu, ‡; Victor Breedveld, ‡; Robert P. Apkarian, † and; Vincent P. Conticello*, †. Rational Design of a Reversible pH-Responsive Switch for Peptide Self-Assembly. **2006**.

(17) Magnotti, E. L.; Hughes, S. A.; Dillard, R. S.; Wang, S.; Hough, L.; Karumbamkandathil, A.; Lian, T.; Wall, J. S.; Zuo, X.; Wright, E. R.; et al. Self-Assembly of an α-Helical Peptide into a Crystalline Two-Dimensional Nanoporous Framework. *J. Am. Chem. Soc.* **2016**, *138* (50), 16274–16282.

(18) Vargo, K. B.; Parthasarathy, R.; Hammer, D. A. Self-Assembly of Tunable Protein Suprastructures from Recombinant Oleosin. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (29), 11657–11662.

(19) Petka, W. A.; Harden, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A. Reversible Hydrogels from Self-Assembling Artificial Proteins. *Science (80-.).* **1998**, *281* (5375), 389 LP-392.

(20) Hassouneh, W.; Zhulina, E. B.; Chilkoti, A.; Rubinstein, M. Elastin-like Polypeptide Diblock Copolymers Self-Assemble into Weak Micelles. *Macromolecules* **2015**, *48* (12), 4183–4195.

(21) Park, W. M.; Champion, J. A. Thermally Triggered Self-Assembly of Folded Proteins into Vesicles. *J. Am. Chem. Soc.* **2014**, *136* (52), 17906–17909.

(22) 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* (8), 2419–2426.

(23) Wright, E. R.; Conticello, V. P. Self-Assembly of Block Copolymers Derived from Elastin-Mimetic Polypeptide Sequences. *Adv. Drug Deliv. Rev.* **2002**, *54* (8), 1057–1073.

(24) Srinivasan, N.; Kumar, S. Ordered and Disordered Proteins as Nanomaterial Building Blocks. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology* **2012**, *4* (2), 204–218.

(25) MacEwan, S. R.; Chilkoti, A. Applications of Elastin-like Polypeptides in Drug Delivery. *J. Control. Release* **2014**, *190*, 314–330.

(26) Koria, P.; Yagi, H.; Kitagawa, Y.; Megeed, Z.; Nahmias, Y.; Sheridan, R.; Yarmush, M. L. Self-Assembling Elastin-like Peptides Growth Factor Chimeric Nanoparticles for the Treatment of Chronic Wounds. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (3), 1034–1039.

(27) Keeley, F. W.; Bellingham, C. M.; Woodhouse, K. A. Elastin as a Self-Organizing Biomaterial: Use of Recombinantly Expressed Human Elastin Polypeptides as a Model for Investigations of Structure and Self-Assembly of Elastin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2002**, *357* (1418), 185–189.

(28) Lees, J. F.; Shneidman, P. S.; Skuntz, S. F.; Carden, M. J.; Lazzarini, R. A. The Structure and Organization of the Human Heavy Neurofilament Subunit (NF-H) and the Gene Encoding It. *EMBO J.* **1988**, *7* (7), 1947–1955.

(29) Adiga, S. P.; Brenner, D. W. Molecular Basis for Neurofilament Heavy Chain Side Arm Structure Modulation by Phosphorylation. *J. Phys. Chem. C* **2010**, *114* (12), 5410–5416.

(30) Chang, R.; Kwak, Y.; Gebremichael, Y. Structural Properties of Neurofilament Sidearms: Sequence-Based Modeling of Neurofilament Architecture. *J. Mol. Biol.* **2009**, *391* (3), 648–660.

(31) Bhagawati, M.; Rubashkin, M. G.; Lee, J. P.; Ananthanarayanan, B.; Weaver, V. M.; Kumar, S. Site-Specific Modulation of Charge Controls the Structure and Stimulus Responsiveness of Intrinsically Disordered Peptide Brushes. *Langmuir* **2016**, *32* (23), 5990–5996.

(32) Ackerley, S.; Thornhill, P.; Grierson, A. J.; Brownlees, J.; Anderton, B. H.; Leigh, P. N.; Shaw, C. E.; Miller, C. C. J. Neurofilament Heavy Chain Side Arm Phosphorylation Regulates Axonal Transport of Neurofilaments. *J. Cell Biol.* **2003**, *161* (3), 489–495.

(33) Lei, R.; Lee, J. P.; Francis, M. B.; Kumar, S. Structural Regulation of a Neurofilament-Inspired Intrinsically Disordered Protein Brush by Multisite Phosphorylation. *Biochemistry* **2018**, *57* (27), 4019–4028.

(34) Lee, H.; Akers, W.; Bhushan, K.; Bloch, S.; Sudlow, G.; Tang, R.; Achilefu, S. Near-Infrared pH-Activatable Fluorescent Probes for Imaging Primary and Metastatic Breast Tumors. *Bioconjug. Chem.* **2011**, *22* (4), 777–784.

(35) Srinivasan, N.; Bhagawati, M.; Ananthanarayanan, B.; Kumar, S. Stimuli-Sensitive Intrinsically Disordered Protein Brushes. *Nat. Commun.* **2014**, *5* (1), 5145.

(36) Lechuga, M.; Fernández-Serrano, M.; Jurado, E.; Núñez-Olea, J.; Ríos, F. Acute Toxicity of Anionic and Non-Ionic Surfactants to Aquatic Organisms. *Ecotoxicol. Environ. Saf.* **2016**, *125*, 1–8.

(37) Cowan-Ellsberry, C.; Belanger, S.; Dorn, P.; Dyer, S.; McAvoy, D.; Sanderson, H.; Versteeg, D.; Ferrer, D.; Stanton, K. Environmental Safety of the Use of Major Surfactant Classes in North America. *Crit. Rev. Environ. Sci. Technol.* **2014**, *44* (17), 1893–1993.

(38) Baneyx, F. Recombinant Protein Expression in Escherichia Coli. *Current Opinion in Biotechnology*. Elsevier Current Trends October 1, 1999, pp 411–421.

(39) Raran-Kurussi, S.; Waugh, D. S. Unrelated Solubility-Enhancing Fusion Partners MBP and NusA Utilize a Similar Mode of Action. *Biotechnol. Bioeng.* **2014**, *111* (12), 2407–2411.

(40) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157* (1), 105–132.

(41) Tegel, H.; Tourle, S.; Ottosson, J.; Persson, A. Increased Levels of Recombinant Human Proteins with the Escherichia Coli Strain Rosetta(DE3). *Protein Expr. Purif.* **2010**, *69* (2), 159–167.

(42) Hammouda, B. Temperature Effect on the Nanostructure of SDS Micelles in Water. *J. Res. Natl. Inst. Stand. Technol.* **2013**, *118*, 151–167.

(43) Beliciu, C. M.; Moraru, C. I. Effect of Solvent and Temperature on the Size Distribution of Casein Micelles Measured by Dynamic Light Scattering. *J. Dairy Sci.* **2009**, *92* (5), 1829–1839.

(44) Kwon, G.; Naito, M.; Yokoyama, M.; Okano, T.; Sakurai, Y.; Kataoka, K. Block Copolymer Micelles for Drug Delivery: Loading and Release of Doxorubicin. *J. Control. Release* **1997**, *48* (2), 195–201.

(45) Kaczerewska, O.; Brycki, B.; Ribosa, I.; Comelles, F.; Garcia, M. T. Cationic Gemini Surfactants Containing an O-Substituted Spacer and Hydroxyethyl Moiety in the Polar Heads: Self-Assembly, Biodegradability and Aquatic Toxicity. *J. Ind. Eng. Chem.* **2018**, *59*, 141–148.

(46) Charlson, R. J.; Rodhe, H. Factors Controlling the Acidity of Natural Rainwater. *Nature* **1982**, *295* (5851), 683–685.

(47) Piñeiro, L.; Novo, M.; Al-Soufi, W. Fluorescence Emission of Pyrene in Surfactant Solutions. *Adv. Colloid Interface Sci.* **2015**, *215*, 1–12.

(48) Wen, Y.; Li, J. Ultrastable Micelles Boost Chemotherapy. *Nat. Biomed. Eng.* **2018**, *2* (5), 273–274.

(49) Association for Computing Machinery., C.; Sigarch.; IEEE Computer Society. *Proceedings : SC 06 : Powerful beyond Imagination : Proceedings : November 11-17, 2006, Tampa Convention Center, Tampa Florida*; [Association for Computing Machinery], 2006.

SYNOPSIS TOC: