Interactions Governing the Self-Assembly of Globular Protein-Polymer Block Copolymers

by

Christopher N. Lam

B.S. Chemical and Biomolecular Engineering, Rice University, 2010
Minor in Computational and Applied Mathematics, Rice University, 2010

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2016

© 2016 Massachusetts Institute of Technology. All rights reserved.
Interactions Governing the Self-Assembly of Globular Protein-Polymer Block Copolymers

Christopher N. Lam

Submitted to the Department of Chemical Engineering on March 11, 2016
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy in Chemical Engineering

Abstract

Engineering enzymes and other proteins into biocatalysts or bioelectronic devices has the potential to lead to a new generation of energy-generating and energy conversion technologies. Controlling the hierarchical structure of protein materials from the nanoscale single molecule level up to the microscale material morphology is critical to improving their function. Lithographic patterning methods such as electron beam lithography, dip-pen nanolithography, and nanograftin allow proteins to be patterned with nanoscale resolution, but parallelization to increase throughput remains a significant challenge. While templated self-assembly enables patterning in three dimensions, maximizing protein loading and controlling orientation are challenges that remain to be addressed. Self-assembly provides a low-cost method to nanopattern proteins for biofunctional devices with high operational efficiency through control over three-dimensional spatial arrangement and orientation.

Complementary experimental techniques were used to investigate the phase behaviors of globular protein-polymer block copolymers and provide insight into the relevant physics and thermodynamics governing their self-assembly. In particular, methodical permutations were made to the protein block to understand the relationship between protein interactions and protein-polymer block copolymer self-assembly. Order-disorder and order-order transitions were demonstrated for the first time within a rich window of phase space of hexagonal, lamellar, perforated lamellar, and micellar phases that were dependent on coil fraction. Protein-polymer net repulsive interactions were discovered to be important for self-assembly. The type of nanostructures formed at a given coil fraction are different between globular-coil and coil-coil systems due to the anisotropy between protein and coil shape and interactions and minor differences in solvent selectivity. A set of structurally homologous proteins in which the chemical composition and surface interaction potential were varied globally throughout the entire sequence and locally through single point mutations demonstrated highly similar phase behavior, revealing that coarse-grained properties such as the protein shape, size, solubility, surface charge, and virial coefficient can capture the general shape of the phase diagram in nonselective solvents. Engineering greater changes in protein electrostatic interactions and virial coefficient demonstrated that the electrostatic environment of proteins may be designed to tune the morphologies of protein-polymer block copolymers, both enhancing and suppressing formation of nanostructures through attractive and repulsive interactions, respectively.

A combination of small-angle neutron scattering experiments, theory, and coarse-grained modeling and simulation was used to elucidate the shape of protein-polymer block copolymers in dilute solution and quantify their interactions. Modeling protein-polymer interactions using repulsive Weeks-Chandler-Andersen potentials showed that the polymer exists as a relatively unperturbed coil extended away from the protein. The coarse-grained representation additionally provides a simple way to model the conformation of protein-polymer conjugates with strong interactions that result in the polymer wrapping around the protein in a shroud-like configuration.

Thesis supervisor: Bradley D. Olsen, Paul M. Cook Career Development Associate Professor of Chemical Engineering
Acknowledgements

I have been told more than once—if my memory serves me correctly—that if one could start graduate school knowing what one would know at the end, then it would only take 2 years to graduate, or a much shorter amount of time than the typical 5 - 7 year graduate timeline. As I am nearing the graduate school finish line, I reflect upon my experiences and the knowledge that I have gained and feel that, even knowing what I know now, I still wouldn’t be able to have accomplished all this in 2 years. This is because of all the extraordinary help and support that I have received along my journey to this point, and I have much to be thankful for and many people to graciously thank and acknowledge for helping me get to where I am today.

I am very grateful for having the opportunity to join the Olsen Lab and having Brad as a role model and mentor to me during my graduate career as a young and growing scientist. I approached graduate school as a learning opportunity for me to become an independent scientist, but without Brad’s advice and guidance, I would have been completely lost. It has been a difficult journey but one that has made my graduate experiences fulfilling, and I would like to thank Brad very much for helping me achieve the opportunities that I have and for his mentorship, leadership, and friendship.

I would like to thank my committee members, Professor Arup Chakraborty, Professor Paula Hammond, and Professor Bruce Tidor for their support and for asking thoughtful questions not only to help guide me along my research but also to help guide my critical thinking as a scientist. I also thank them for all their career advice as I neared completion of my doctoral studies.

I remember when Wei-Ren visited MIT campus with his son. I didn’t realize back then that that was the small beginning of what would end up being a long-term collaboration throughout the last few years of graduate school. It is rare to find someone else who is as nervous as I am, but I have really enjoyed my discussions with Wei-Ren, both scientific and non-scientific topics, and I look forward to learning much from during my postdoc at Oak Ridge.

When I joined the Olsen Lab, there was a total of seven members—Bokyung, Mitchell, Matt, Carla, Shengchang, Gabo, and myself. Almost everything in the lab was new to me, and I couldn’t have gotten off to such a strong and happy start without their camaraderie as friends inside and outside the lab. As the lab grew quickly, it only meant more labmates and friends and greater teamwork, and I enjoyed working alongside Yan, Vinay, Michelle, Charlotte, Dongsook, Manos, Minkyu, Reggie, Aaron, Guokui, and Xuehui. As the lab continued to grow even more, I would like to thank Allie, Carrie, Brian, Daphne, Cici, Andreia, Sybele, Vivi, Bruno, Yun Jung, and Tom for all their help and kindness.

When I entered my second year and began experimenting with light scattering, Dr. Kooi at the ISN devoted a great deal of his time and effort to help configure the laser setup for my experiments. I would like to thank him for all his help and for continuing to be extremely supportive of my experiments and projects. I would also like to thank Lionel for welcoming me into Professor Swager’s Lab and allowing me to use their equipment and also for his dedication to training students and maintaining equipment at the ISN. I have also frequented the BIF many times and thank Debby for her expertise and help with my data acquisition and analysis.

I would like to thank Amy, Brad’s cousin, from the Department of Civil and Environmental Engineering at MIT, for helping me mount a motor onto our fermentor when I first started and for her recommendation that I enroll in the training course at the MIT Edgerton Student Shop. That was one of my most memorable experiences at MIT. I’ve always wanted to learn to build things, and Mark Belanger, who runs the Edgerton Student Center, has helped me with numerous projects. I also thank Ilan, who helped train me during the training course, and also Ken, Brian, and Hayami for their help with machining needs at the Hobby Shop.

I dedicate this thesis to my family. An only child, I am grateful to have many cousins with whom I am close, and I truly appreciate my cousin Eric, who is like a brother to me, for his moral support during my graduate studies and for traveling all the way to Cambridge from California on such short notice to support me during my thesis defense. A first-generation Vietnamese American, my family left their
families when they first came to America with barely enough to survive, and they worked hard—harder than I have in graduate school—to sacrifice everything so that I could grow up to have everything that they did not growing up. Raised by my mother, father, and two grandmothers, they have been and are always right beside me, supporting my every decision, and they have been and always will be my source of strength, inspiration, and happiness.
# Table of Contents

Acknowledgements ................................................................................................................. 6
List of Schemes .......................................................................................................................... 10
List of Tables .............................................................................................................................. 10
List of Figures ............................................................................................................................ 10

Chapter 1. Introduction ............................................................................................................. 14

1.1 Enzymes .............................................................................................................................. 14
1.2 Protein Patterning Techniques ........................................................................................... 15
1.3 Block Copolymer Self-Assembly ......................................................................................... 20
1.4 Adding Complexity to Block Copolymer Design ................................................................. 21
1.5 Globular Protein-Polymer Block Copolymers .................................................................... 22
1.6 Thesis Overview .................................................................................................................. 27
1.7 References .......................................................................................................................... 28

Chapter 2. Experimental Methods ............................................................................................ 38

2.1 Synthesis of Globular Protein-Polymer Block Copolymers ............................................... 38
  2.1.1 Synthesis of Maleimide-Functionalized Polymers ......................................................... 38
  2.1.2 Synthesis of Globular Proteins ....................................................................................... 50
  2.1.3 Bioconjugation .............................................................................................................. 69
2.2 Characterization Techniques ............................................................................................... 72
2.3 Computer Experiments and Calculations ......................................................................... 86
  2.3.1 Coarse-Grained Molecular Dynamics Simulation ......................................................... 86
  2.3.2 Model Fitting ................................................................................................................ 98
2.4 References .......................................................................................................................... 103

Chapter 3. Phase Transitions in Concentrated Solution Self-Assembly of Globular Protein-
Polymer Block Copolymers ...................................................................................................... 106

3.1 Abstract ............................................................................................................................... 106
3.2 Introduction ........................................................................................................................ 107
3.3 Experimental Methods ...................................................................................................... 110
3.4 Results and Discussion ........................................................................................................ 113
3.5 Conclusions ......................................................................................................................... 127
3.6 References .......................................................................................................................... 129

Chapter 4. The Nature of Protein Interactions Governing Globular Protein-Polymer Block
Copolymer Self-Assembly ......................................................................................................... 132

4.1 Abstract ............................................................................................................................... 132
4.2 Introduction ........................................................................................................................ 133
4.3 Materials and Methods ...................................................................................................... 138
4.4 Results and Discussion ...................................................................................................... 139
4.5 Conclusions ......................................................................................................................... 154
4.6 References .......................................................................................................................... 155

Chapter 5. The Effect of Protein Electrostatic Interactions on Globular Protein-Polymer Block
Copolymer Self-Assembly ......................................................................................................... 161

5.1 Abstract ............................................................................................................................... 161
5.2 Introduction ................................................................................................................. 162
5.3 Materials and Methods .............................................................................................. 165
5.4 Results and Discussion .............................................................................................. 167
5.5 Conclusions ................................................................................................................. 183
5.6 References ................................................................................................................... 184
Chapter 6. The Shape of Protein-Polymer Conjugates in Dilute Solution ....................... 188
   6.1 Abstract ....................................................................................................................... 188
   6.2 Introduction ................................................................................................................. 188
   6.3 Experimental Methods .............................................................................................. 192
   6.4 Results and Discussion .............................................................................................. 197
   6.5 Conclusions ................................................................................................................. 211
   6.6 References ................................................................................................................... 212
Chapter 7. Conclusions ..................................................................................................... 215
   7.1 Summary ..................................................................................................................... 215
   7.2 Outlook ....................................................................................................................... 217
Appendix A. List of Abbreviations .................................................................................... 221
Appendix B. Supporting Information for Chapter 3 .......................................................... 223
Appendix C. Supporting Information for Chapter 4 .......................................................... 229
Appendix D. Supporting Information for Chapter 5 .......................................................... 249
Appendix E. Supporting Information for Chapter 6 .......................................................... 282
List of Schemes

Scheme 4-1. Cartoon structures of mCherry, EGFP, and their electrostatic surface potentials; protein sequence alignment of mCherry and EGFP; electrostatic surface potential of mCherry variants and amino acid sequence

Scheme 5-1. Cartoon structure of superfolder GFP and electric dipole moments and electrostatic surface potentials of neutral, superfolder, and supercharged GFP

Scheme 6-1. Chemical structures of mCherry-polymer conjugates and coarse-grained cylinder-coil bead representation

List of Tables

Table 3-1. Composition of mCherry-PNIPAM block copolymers
Table 3-2. Thermal transitions for mCherry-PNIPAM
Table 4-1. Composition of mCherry-/EGFP-PNIPAM block copolymers
Table 4-2. Thermal transitions mCherry-/EGFP-PNIPAM block copolymers
Table 4-3. Biophysical properties of mCherry and EGFP
Table 5-1. Composition of GFP(O)-, GFP(-8)-, and GFP(-21)-PNIPAM block copolymers
Table 5-2. Thermal transitions for GFP(O)-, GFP(-8)-, and GFP(-21)-PNIPAM
Table 6-1. Molecular properties of mCherry-polymer block copolymers
Table 6-2. DLS and Guinier analysis of mCherry-polymer block copolymers
Table 6-3. Results of model fitting of mCherry-polymer block copolymers results using a theoretical scattering function
Table B-1. Concentration and temperature conditions corresponding to representative SAXS patterns in Figure 3-3
Table B-2. Scattering length densities of mCherry, PNIPAM, H2O, and D2O
Table C-1. mCherry protein sequence and primers for preparation of mCherry variants
Table C-2. Fourier self-deconvolution analysis of FTIR spectra
Table D-1. Values of cluster scores from ClusPro server results for GFP-GFP interactions

List of Figures

Figure 2-1. Column chromatography purification of EMP
Figure 2-2. 1H-NMR spectrum of EMP
Figure 2-3. 1H NMR of exo-3a,4,7,7a-Tetrahydro-2-(3-hydroxypropyl)-4,7-epoxy-14-isoindole-1,3(2H)-dione (imide)
Figure 2-4. 1H NMR of EMP-imide
Figure 2-5. 1H NMR of MEA
Figure 2-6. Time series of PHPA
Figure 2-7. Time series of POEGA
Figure 2-8. Fluorescence spectroscopy of EGFP and variants
Figure 2-9. SDS-PAGE showing selective bioconjugation to EGFP
Figure 2-10. Cartoon of crystal structure of superfolder GFP and mutated residues to achieve a net charge of -25
Figure 2-11. Cartoon of crystal structure of superfolder GFP and mutated residues to achieve a net charge of +15

Figure 2-12. Amino acid sequences for GFP variants

Figure 2-13. Fermentation of mCherryS131C

Figure 2-14. SDS-PAGE of mCherryS131C and EGFP

Figure 2-15. FPLC chromatogram of mCherry

Figure 2-16. SDS-PAGE of FPLC fractions of mCherry

Figure 2-17. Synthesis and purification steps of mCherry- and EGFP-PNIPAM conjugates

Figure 2-18. FPLC chromatogram of mCherry-PEG28k

Figure 2-19. SDS-PAGE of FPLC fractions of mCherry-PEG28k

Figure 2-20. Native and SDS-PAGE of purified protein-polymer bioconjugates

Figure 2-21. Casting setup of conjugates and photo of mCherry- and EGFP-PNIPAM plastics

Figure 2-22. Annealing chamber setup

Figure 2-23. Liquid samples loaded in Banjo cells for SANS

Figure 2-24. SANS and Guinier analysis of mCherry

Figure 2-25. PDB2PQR web server for APBS electrostatics calculations

Figure 2-26. Coarse-grained cylinder model

Figure 2-27. Average mean squared displacement and orientational and end-to-end correlation functions for coarse-grained mCherry simulation

Figure 2-28. Model fitting from coarse-grained cylinder model to mCherry SANS

Figure 3-1. Representative SAXS as a function of temperature for mCherry-PNIPAM

Figure 3-2. Phase diagrams for mCherry-PNIPAM

Figure 3-3. Representative SAXS patterns for mCherry-PNIPAM

Figure 3-4. Contrast matching peak intensities and curve fits for mCherry-PNIPAM

Figure 3-5. Order-disorder transition concentration as a function of coil fraction for mCherry-PNIPAM

Figure 3-6. DSC of mCherry-PNIPAM

Figure 4-1. Phase diagrams comparing EGFP-PNIPAM and mCherry-PNIPAM

Figure 4-2. Representative SAXS patterns for EGFP-PNIPAM and mCherry-PNIPAM

Figure 4-3. Solid-state SAXS and TEM images of EGFP-PNIPAM and mCherry-PNIPAM

Figure 4-4. Cell culture color intensity, CD, and Native PAGE of mCherry variants

Figure 4-5. SAXS of PNIPAM conjugates of mCherry and mCherry variants

Figure 5-1. Phase diagrams of GFP(0)-, GFP(-8)-, and GFP(-21)-PNIPAM

Figure 5-2. Comparing FWHM of GFP(0)-, GFP(-8)-, and GFP(-21)-PNIPAM

Figure 5-3. Cartoon representations of ClusPro server results for GFP-GFP interactions

Figure 5-4. SAXS patterns of GFP(-21)-PNIPAM in the presence of salt

Figure 6-1. Model fitting of mCherry-polymer block copolymers using a theoretical scattering function

Figure 6-2. Model fitting from fitting a coarse-grained cylinder model to mCherry SANS

Figure 6-3. Model fitting from fitting a coarse-grained cylinder-coil model to mCherry-polymer SANS

Figure 6-4. Form factors and visualizations of cylinder-coil model as WCA interactions increase in magnitude and as the cutoff radius changes to approach that of the full Lennard-Jones interaction potential

Figure B-1. Gel permeation chromatography of PNIPAM for Chapter 3

Figure B-2. SDS-PAGE and Native PAGE of mCherry-PNIPAM
Figure B-3. Representative depolarized light scattering power fractions for mCherry-PNIPAM
Figure B-4. SANS contrast matching data of mCherry-PNIPAM
Figure B-5. Zimm plot of mCherry
Figure C-1. DNA and amino acid sequences of EGFP and EGFPC14S
Figure C-2. Gel permeation chromatography of PNIPAM for Chapter 4
Figure C-3. SDS-PAGE of test bioconjugation of EGFP with maleimide-PEG
Figure C-4. Native and SDS-PAGE of mCherry- and EGFP-PNIPAM conjugates in Chapter 4
Figure C-5. Depolarized light scattering and turbidimetry results for mCherry- and EGFP-PNIPAM
Figure C-6. Zimm plot of EGFPC14S
Figure C-7. Cloud point of EGFPC14S/PNIPAM and mCherry/PNIPAM blends
Figure C-8. UV-Vis absorption spectra of mCherry and variants
Figure C-9. SAXS of PNIPAM conjugates of mCherry and mCherry variants in the presence of 200 mM NaCl
Figure C-10. UV-Vis absorption spectra of EGFP-PNIPAM
Figure C-11. Solid-state Fourier transform infrared spectroscopy of solid-state EGFP-PNIPAM
Figure C-12. Circular dichroism spectroscopy of EGFP and EGFP-PNIPAM
Figure C-1. Gel permeation chromatography of PNIPAM for Chapter 5
Figure D-1. DNA and amino acid sequences of GFP(0), GFP(-8), and GFP(-21)
Figure D-2. UV-Vis absorption spectra of GFP variants and conjugates of PNIPAM after dialysis
Figure D-3. UV-Vis absorption spectra of GFP variants and conjugates after self-assembly and rehydration
Figure D-4. Circular dichroism spectroscopy of GFP variants and conjugates of PNIPAM after dialysis
Figure D-5. Radially averaged SAXS patterns for GFP(0)-, GFP(-8)-, and GFP(-21)-PNIPAM21k
Figure D-6. Radially averaged SAXS patterns for GFP(0)-, GFP(-8)-, and GFP(-21)-PNIPAM30k
Figure D-7. Depolarized light scattering and turbidimetry of GFP(0)-PNIPAM21k
Figure D-8. Depolarized light scattering and turbidimetry of GFP(-8)-PNIPAM21k
Figure D-9. Depolarized light scattering and turbidimetry of GFP(-21)-PNIPAM21k
Figure D-10. Depolarized light scattering and turbidimetry of GFP(0)-PNIPAM30k
Figure D-11. Depolarized light scattering and turbidimetry of GFP(-8)-PNIPAM30k
Figure D-12. Depolarized light scattering and turbidimetry of GFP(-21)-PNIPAM30k
Figure D-13. Differential scanning calorimetry of GFP-PNIPAM
Figure D-14. Cartoon representations of ClusPro server results for GFP-GFP interactions
Figure D-15. SAXS patterns of GFP(-21)-PNIPAM in the presence of NaCl at different concentrations
Figure E-1. Gel permeation chromatography of PNIPAM, PHPA, POEGA, and PEG for Chapter 6
Figure E-2. SDS-PAGE and Native PAGE of mCherry-polymer block copolymers
Figure E-3. UV-Vis absorption spectra of mCherry-polymer block copolymers
Figure E-4. Guinier analysis of mCherry, homopolymers, and mCherry-polymer block copolymers
Figure E-5. Model fitting results from fitting homogeneous form factor to mCherry SANS
Figure E-6. Model fitting results from fitting worm-like chain model form factor to homopolymer SANS data
Figure E-7. Determination of contrast match point of mCherry
Figure E-8. Log-log scale plots of model fitting of theoretical scattering function to mCherry-polymer SANS
Figure E-9. Form factor time correlation functions as a function of Q
Figure E-10. Reduced $\chi^2$ values from fitting coarse-grained coil to homopolymer SANS data as a function of bead number
Figure E-11. Plots of SANS data and best fit from using coarse-grained coil model
Figure E-12. Model fitting results from coarse-grained cylinder-coil to mCherry-polymer SANS
Figure E-13. Truncated and shifted Lennard-Jones potential as a function of cutoff distance and second virial coefficient between cylinder-coil beads as a function of cutoff distance
Figure E-14. Form factors of coarse-grained cylinder-coil model with increasing interaction magnitude at different cutoff distances
Chapter 1. Introduction

1.1 Enzymes

Enzymes date back as early as 1833,\textsuperscript{1} from when French chemist Anselme Payen first extracted the enzyme diastase. Since then, countless other enzymes with unique functionalities have been discovered, their crystal structures determined with high resolution (the first being lysozyme in 1965), and their mechanisms studied in great detail, yet we are still perplexed by the extraordinary properties of these molecules. Compared to synthetic catalysts, enzymes have much higher turnover frequencies, \( k_{\text{cat}} \), that range from \( 10^{-1} \) to \( 10^{7} \) s\(^{-1} \), and their large substrate binding affinities (\( \approx 1/K_m \), where \( K_m \) is the Michaelis constant) and substrate specificities (\( k_{\text{cat}}/K_m \)) lead to minimal side reactions.\textsuperscript{2,3} They catalyze reactions using the same physical organic principles as other reactions but are astronomically better at it; the fundamental mechanistic understanding of enzymatic catalysis is still developing, and someday the challenge of synthesizing enzyme mimics may be solved.\textsuperscript{4} Recently, enzymes from nature have attracted much attention for their incorporation directly into materials to address challenges in energy, environment, medicine, national security, and more. To demonstrate the breadth of enzyme applications, a number of examples will be briefly mentioned. Interest in enzyme biofuel cell (BFC) has been revitalized in the past decade by the development of new electrode materials that can eliminate the need for electrolyte membranes due to the enzyme reaction specificity at both electrodes, allowing miniaturization of BFCs for implantable bioelectronics and bionics.\textsuperscript{5,6} The advantages of enzymatic treatment processes has led to interest in the development of enzyme processes for waste treatment\textsuperscript{7} and soil decontamination.\textsuperscript{8,9} Pioneered for glucose monitoring with glucose oxidase in patients, biosensors have been developed for quick and easily accessible diagnostics in biomedical applications as well as environmental monitoring\textsuperscript{10} and neurotoxin
sensing and decontamination for national security.\textsuperscript{11-13} Reaction centers (RCs),\textsuperscript{14} photosynthetic complexes,\textsuperscript{15} and fluorescent proteins\textsuperscript{16} can be utilized in light-harvesting materials for bioelectronics and energy applications. Fluorescent proteins have even been incorporated into lasers as a biological gain medium.\textsuperscript{17,18} Enzymes have also been impactful in the pharmaceutical, food and agriculture, paper, and textile industries.\textsuperscript{19} The discovery of enzymes from extremophilic microorganisms has opened many opportunities for novel enzymatic applications in industrial processes with harsh reaction conditions.\textsuperscript{20,21} While enzymes generally function in solution, the immobilization of enzymes has garnered much interest due to the ability to provide a stabilizing environment for enzymes that can potentially also enhance enzyme efficiency and functionality.\textsuperscript{24} Additionally, enzyme immobilization is essential in commercial applications, allowing facile separation of the enzymes from the product and reuse of the enzymes, which provide cost savings that help reduce the high expenditures often associated with enzyme production and purification.\textsuperscript{22,23}

\textbf{1.2 Protein Patterning Techniques}

Augmenting materials design and function with the functionality of enzymes has immense potential but faces a number of engineering challenges—enzyme stability, control over the spatial position and orientation, and enzyme density—that are addressed through different techniques of protein nanopatterning. Enzymes are intricately folded molecules that are susceptible to loss of activity through unfolding and denaturation when placed in a non-native environment with suboptimal or deleterious pH and temperature.\textsuperscript{25-27} Care must be taken to incorporate enzymes in such a manner that their active sites are easily accessible by controlling their position and orientation, allowing for facile transport of substrate, product, and
cofactors.\textsuperscript{28,29} Providing a pathway that enables charge transport is essential for enzymatic biofuel cell applications.\textsuperscript{30} Moreover, because proteins are relatively large compared to transition metal catalysts, the ability to achieve a high density of enzyme can significantly improve device functionality. To meet specific engineering criteria for designing enzyme-based technologies, the following three broad approaches to nanopatterning are typically employed and will be briefly discussed: lithography, templated assembly, and direct self-assembly.

**Lithography.** Lithography enables top-down patterning of structures through the definition of a template, followed by the transfer of that template to produce a functional protein nanopattern. A wide variety of methods have been developed for fabricating such patterns, including photolithography, electron beam lithography (EBL), dip-pen nanolithography (DPN), nanografting, and nanocontact printing (nCP). Photolithography and EBL are well-developed techniques for processing and patterning integrated circuits in the semiconductor industry and have been adapted to pattern biomolecules. While photolithography is a parallel technique that can achieve high throughput processing, it requires high capital and operational costs; more importantly for device fabrication, the harsh processing conditions and chemical modification of proteins to accommodate photoreactive moieties may reduce protein activity.\textsuperscript{31,32} One of the greatest advantages of EBL is its ability to create precise nanopatterns with the highest resolution of all lithographic techniques, capable of fabricating features with 5 nm resolution reproducibly.\textsuperscript{33} Similar to photolithography, EBL is a costly technique that requires specialized equipment, but it is also a serial process and is slow relative to other printing and stamping techniques. DPN, a scanning probe microscopy (SPM)-based lithographic technique, is one of the most versatile techniques for nanopatterning biomolecules due to its ability to create patterns under conditions of ionic strength, pH, and hydration that are compatible with many different
biomolecules by using an atomic force microscopy (AFM) tip. Although traditionally limited by its serial writing process, the throughput of DPN has been improved by the development of passive probe arrays and active arrays and polymer pen lithography (PPL), which provides DPN with the added advantage of large-area patterning capabilities in printing methods. Nanoshaving and nanografting are two additional SPM techniques that operate by using an AFM tip to remove regions of a self-assembled monolayer (SAM) and allow binding of biomolecules to the fabricated pattern. Nanocontact printing (nCP) is a higher resolution form of microcontact printing (μCP) that allows patterns of biomolecules to be created facilely and rapidly in parallel by binding biomolecules to an elastomeric stamp that is then used to transfer the biomolecules to a patterned substrate. Additional lithographic methodologies for controlling protein nanopatterning include nanoimprint lithography, nanopipettes or nanopens, and the use of paramagnetic metal masks to design a magnetic field onto a substrate to assemble ferromagnetic nanoparticles to create patterns for protein immobilization.

All of these lithography processes rely on forming areas that can selectively bind proteins using one of three approaches: physical adsorption, covalent immobilization, or bioaffinity interactions. Physical adsorption utilizes electrostatic, ionic, hydrophobic/hydrophilic, or hydrogen-bonding interactions to immobilize proteins to the substrate directly or indirectly. Chemisorption approaches to covalently attach proteins to functionalized patterned surfaces focus primarily on functional groups including amines, carboxyls, thiols, epoxies, and photoactive materials. Amine and carboxyl chemistries are generally not site specific, potentially leading to immobilization of proteins in a random orientation that may adversely impact protein function. Site-specific covalent immobilization reactions may enable control of protein orientation, and short linkers between the protein and patterned surface have
been found to be beneficial for protein stability.\textsuperscript{53,61–63} Bioaffinity interactions—between enzyme and substrate, receptor protein and its recognition pair, or antigen–antibody pairs—such as that between biotin and avidin ($K_a \approx 1015 \text{ L mol}^{-1}$),\textsuperscript{64} while not covalent, can provide very strong interactions for forming stable protein patterns.\textsuperscript{65–67} Many of these chemistries have been extensively reviewed in a number of excellent recent articles.\textsuperscript{54,22,68–72}

**Templated Assembly.** Using a three-dimensional template to guide protein assembly is a second major approach for controlling protein nanopatterning. Advantages of this approach include the capability of controlling protein nanostructure in three dimensions, the potential to maintain and improve enzyme stability and activity with a suitable choice of a template that provides a favorable environment for the protein, and the ability to prepare nanostructured scaffolds using well-established processing methods. Templated assembly techniques can be categorized into two groups: first, three-dimensional structures that form interfaces such as silica sol-gel structures\textsuperscript{73,74} or nanoporous gold structures\textsuperscript{75} allow proteins to be physisorbed, chemisorbed, or bound to the surface, and secondly, co-assembly of enzymes with polymer or lipid can form nanostructured materials where interactions between protein and polymer or lipid guide the protein into specific nanodomains. The net charge on the protein surface may be modulated by controlling pH, allowing electrostatic interactions to be used to immobilize the proteins in methods such as layer-by-layer (LbL) deposition and coacervate micelle formation. Hydrophobic interactions are important in the selective adsorption of proteins onto BCP-patterned surfaces. Conjugation of polymers to proteins can direct the protein into nanopatterned regions of chemical similarity,\textsuperscript{76} and membrane proteins may be self-assembled within lipid and polymer templates using hydrophobic/hydrophilic patterns, providing a highly effective method to control their structural organization in a biomimetic environment.\textsuperscript{77,78}
Direct Self-Assembly. Conjugating polymers to proteins and engineering fusion proteins are two methods in which a structure-directing functionality may be afforded to proteins, allowing a bottom-up approach to assemble proteins directly. While these methods necessitate modifications to the protein, the direct self-assembly approach can achieve high protein densities within a three-dimensional nanostructure. A great body of literature on protein-polymer conjugates provides versatility in the multitude of routes toward the synthesis of conjugated globular protein-polymer block copolymers with site-specific control over conjugation of a single polymer.\textsuperscript{79}–\textsuperscript{86} Polymers with controlled molecular weights and low dispersities can be synthesized by living polymerizations, primarily atom transfer radical polymerization (ATRP),\textsuperscript{87}–\textsuperscript{91} reversible addition-fragmentation transfer (RAFT),\textsuperscript{92}–\textsuperscript{95} and nitroxide-mediated radical polymerization (NMRP).\textsuperscript{96,97} Genetic engineering of the protein amino acid sequence is often required to functionalize the protein with an appropriate moiety for bioconjugation; given the many different synthetic strategies for protein-polymer conjugation, it is possible to synthesize protein-polymer conjugates for almost any protein of interest. The protein and polymer are covalently linked under mild reaction conditions to maintain the protein's native conformation using either a grafting to approach, in which the two molecules are reactively coupled, or a grafting from approach, in which a protein is used as a macroinitiator to grow the polymer chain directly through a controlled free-radical polymerization. Fully biosynthetic analogues to hybrid protein-polymer conjugates can be synthesized by designing protein fusions that incorporate elastin-like polypeptides (ELPs) as the coil domain to direct self-assembly. Advantages of protein fusion constructs include their relative ease of large-scale biosynthesis and purification and their "monodispersity" and precise control over the fusion block composition, making fusion protein systems a potentially cost-effective alternative to protein-polymer conjugates.
1.3 Block Copolymer Self-Assembly

Block copolymers belong to a rich area of soft matter that have become ubiquitous in everyday applications, owing to their unique properties arising from their complex structures, and the idea to apply the thermodynamic self-assembly property of block copolymers to protein-polymer block copolymers is one of great scientific and industrial interest.\(^{98,99}\) Block copolymers comprise different polymer chains that are linked together covalently. Advances in synthetic chemistry enable various block copolymer compositions and chain connectivity, including linear, branched, cyclic, star-block, and miktoarm-star copolymers topologies, and a number of review articles thoroughly discuss the vast molecular diversity of block copolymers.\(^{100-102}\) Depending upon their composition and molecular architecture, block copolymers are able to self-assemble into a variety of spatially periodic structures on the mesoscale, thermodynamically governed by an energy minimization between enthalpic and entropic forces. The phase behavior of linear AB block copolymers has been predicted by self-consistent mean field theory to be governed by the Flory-Huggins interaction parameter $\chi$, the degree of polymerization $N$, and the composition, represented as the coil fraction of the A block. The theoretical phase diagram agrees well with the experimental phase diagram of poly(styrene)-b-poly(isoprene).\(^{98}\) Block copolymers have also demonstrated rich phase behavior in solution. In concentrated block copolymer solutions, theoretical studies predict that a “dilution approximation” can be applied to map the neat block copolymer phase behavior onto that of a solution in a neutral solvent.\(^{103,104}\) According to the dilution approximation, a neutral solvent should uniformly screen the unfavorable A-B interactions, leading to the term $\phi \chi N$ governing the phase behavior, where $\phi$ represents the block copolymer volume fraction; in the semidilute regime, however, theory predicts different scaling behavior due to non-negligible effects of chain
Experiments have confirmed the dilution approximation scaling behavior of order-order transitions (OOTs) and that the dilution approximation breaks down at the order-disorder transition (ODT), following instead scaling behavior of the semidilute regime. The preferential swelling of a particular chain due to solvent selectivity leads to more complex phase behavior, and lyotropic and thermotropic OOTs and ODTs resulting from this asymmetric swelling have been observed in organic and aqueous solvents.

The rich phase behavior and properties of block copolymers allow them to be used in numerous applications. For example, they can be used as selectively permeable membranes, their interconnected nanodomains can allow efficient transport of substrate, product, and charge, and their morphologies have also been used to template materials for applications such as integrated circuit fabrication. Coupling the functionality of enzymes with the complex phase behavior of block copolymers can lead to the development of novel biofunctional materials suitable to address many industrial challenges.

1.4 Adding Complexity to Block Copolymer Design

Novel and practical solutions often require block copolymers that incorporate structural components with specific chain topologies, which adds complexity to the theoretical depiction of their phase behavior. Experimental results demonstrate phase behavior that significantly differs from the relatively symmetric phase behavior of coil-coil block copolymers, forming lamellar morphologies over a wide range of coil volume fraction. Rod-coil block copolymers in the weak segregation limit form stable lamellae, nematic, and isotropic phases, and rod-coil systems with greater segregation strength were able to pack onto a hexagonal lattice in which the rod nanodomains formed rectangular aggregates. Theory predicts that the self-assembly
behavior of rod-coil block copolymers, in addition to the Flory-Huggins interaction, $\chi N$, and coil fraction, $\phi$, are governed by the Maier-Saupe interaction, $\mu N$, and the geometric parameter $v$, which represents the coil to rod length ratio, and detailed experimental investigations of the model rod-coil block copolymer poly(alkoxyphenylenevinylene-$b$-isoprene) (PPV-$b$-PI) provide the first universal phase diagram of rod-coil block copolymers from quantitative calculations of the four governing parameters.\textsuperscript{110}

Block copolymers incorporating polypeptides that have secondary structure have also been observed to have rich phase behavior. Rod-like $\alpha$-helical-coil bioconjugates can self-assemble into lamellar phases over a wide range of coil volume fractions,\textsuperscript{122,123} hexagonally packed arrays within lamellar and cylindrical nanostructures,\textsuperscript{124–126} and micelles and vesicles.\textsuperscript{127–132} Conjugates incorporating $\beta$-sheets have been shown to laterally aggregate into fibrous networks,\textsuperscript{133} and the primary $\beta$-helix motif contributes specific interactions that induce formation of one-dimensional nanostructures.\textsuperscript{134} Cyclopeptide-containing conjugates have been demonstrated to form tubelike structures comprising a core of stacked cyclopeptide rings with a polymer shell that laterally aggregate into weak networks.\textsuperscript{135,136}

1.5 Globular Protein-Polymer Block Copolymers

The direct self-assembly of globular protein-polymer block copolymers presents a bottom-up method for structure control that can achieve exceptionally high protein density while enabling efficient transport of substrates/products and stabilizing the protein, filling a technology gap in protein nanopatterning and biofunctional materials. Enzymes and proteins are as diverse in their shape, size, and composition as they are in their functions, and their specific folded shape and anisotropic hydrophobic, ionic, and hydrogen bonding interactions result in a complex set of
interactions that affect the thermodynamics of self-assembly of globular protein-polymer block copolymers. The rich physics underlying the self-assembly of these block copolymers presents many questions of fundamental interest, the solutions to which are important for realizing the potential applications of direct assembly of globular protein-polymer block copolymers.

Self-Assembly. Early syntheses of protein-polymer conjugates by Hoffman and coworkers targeted designs of thermoresponsive conjugates for affinity separations, enzyme recovery and recycling, and control of protein ligand binding affinity. A multitude of chemistry techniques to synthesize singly tethered protein-polymer conjugates have since been developed, attesting to the extremely large interest in these protein-polymer hybrid systems as biofunctional materials. However, the understanding of the self-assembly behavior of these materials is still rudimentary, and experiments characterizing their self-assembly have primarily focused on dilute solution properties. The dilute solution aggregation behavior of giant amphiphiles comprising a hydrophilic protein head conjugated to an apolar polymer tail have been thoroughly characterized, demonstrating a wide variety of different structures. Nolte and coworkers investigated various enzyme-polymer giant amphiphiles, characterizing formation of horseradish peroxidase-b-polystyrene and myoglobin-b-polystyrene into vesicles, bovine serum albumin-b-polystyrene into micelles, and lipase B-b-polystyrene into µm-long fibers resulting from aggregation of micellar rods. More diverse macromolecular structures have been observed in triblock biohybrid systems consisting of a protein conjugated to a diblock copolymer, including micelles, vesicles, singular Y-junctions, and toroids. Proteins conjugated to thermoresponsive polymers exhibiting lower critical solution temperature (LCST) behavior also show self-assembly into stable micelles and spherical aggregates. Protein fusion systems using thermoresponsive elastin-like polypeptides (ELPs) as the coil-like domain
are fully biosynthetic analogues to protein-polymer conjugates, and results by Chilkoti et al. demonstrate self-assembly into micellar platforms for multivalent display of proteins, as shown schematically in Figure 2.

Self-assembly principles have been applied to form solid-state nanostructured plastics and gels from globular protein-polymer block copolymers. The model globular protein-polymer mCherry-b-poly(N-isopropylacrylamide) has been shown to form many nanostructured phases observed in coil-coil block copolymers—lamellae, perforated lamellae, hexagonal cylinders, and disordered micelles—and changes in polymer chemistry have yielded for the first time a cubic phase consistent with the bicontinuous gyroid structure. The concentration-coil fraction phase diagram of mCherry-b-PNIPAM has been thoroughly investigated and mapped, showing lyotropic and thermotropic ODTs and OOTs in the rich and highly asymmetric phase space that is significantly different than that of coil-coil block copolymers due to the complex, patchy interaction potentials of proteins and the difference in shape between proteins and coil polymers.

Coarse-grained modeling and simulations of tethered nanoparticles with different particle anisotropy including shape variation provide some insight into shape effects on self-assembly. Using an explicit particle-based coarse-grained model, Brownian dynamics simulations by Glotzer et al. of mono-tethered spheres, cubes, and rods were conducted to study the effects of shape anisotropy on self-assembly. Simulations demonstrated canonical microphase separated morphologies observed for coil-coil block copolymers, but notable differences include significant shifts in phase boundaries and modifications to local packing and ordering of the nanoparticles within the self-assembled morphologies due to the specific nanoparticle geometry. A recent study by Damasceno et al. addresses the challenge of predicting self-assembly a priori
from knowledge of the building block shape through Monte Carlo simulations of 145 convex polyhedra.\textsuperscript{156} Preliminary results demonstrate correlation of the isoperimetric quotient (IQ) and coordination number of the shapes as a predictor of shape assembly category. Additionally, a majority of the shapes forming disordered glasses lacks centrosymmetry, suggesting that higher shape symmetry accords particles with increased packing and assembly abilities. Jayaraman and coworkers\textsuperscript{157–162} have applied the polymeric reference interaction site model (PRISM) theory and Monte Carlo (MC) simulations to provide insight on the effects of particle size, tether number and length, monomer sequence, packing fraction, particle-particle attraction strength on the real-space structure and self-assembly behavior in concentrated solutions and in the melt phase.

**Protein-Polymer Interactions.** In analogy to the relationship between polymer blends and block copolymers, the phase behavior of protein-polymer solutions presents an important reference for the development of theories of protein-polymer conjugate thermodynamics. The free energy landscape delineating the self-assembly behavior of globular protein-polymer conjugates is expected to be governed by the same physics and thermodynamics underlying protein-polymer interactions involved in protein aggregation, partitioning, and crystallization.\textsuperscript{163–166} Experimental studies of protein-polymer solutions have revealed features such as the crystallization slot phenomenon for proteins.\textsuperscript{167} Liquid-liquid coexistence between protein-rich and polymer-rich phases have also been studied.\textsuperscript{168–170} Depletion interactions in protein-polymer solutions have been demonstrated to be significant,\textsuperscript{171–175} and a theoretical basis for resolving the protein-solvent, protein-protein, and ideal mixture interactions contributing to the protein second virial coefficient has been developed.\textsuperscript{176}

The interactions between protein and polymer remain a topic of active investigation. Poly(ethylene oxide) (PEO)\textsuperscript{177,178} has been the model polymer of many investigations of protein-
polymer interactions due to its interest for biomedical applications. The main platform for probing protein-polymer interactions has involved studying protein interactions with a PEO-coated surface,\textsuperscript{179} and experiments have been complemented by modeling efforts and single chain mean-field theory.\textsuperscript{180–183} Early studies reported that protein-PEO interactions primarily involved excluded volume interactions. Ingham\textsuperscript{173} studied the effects of pH, PEO size, and protein size in the precipitation of proteins by PEO and showed that attractive or repulsive chemical interactions between PEO and protein are dominated by an entropic mechanism of precipitation such as excluded volume. Densiometry experiments\textsuperscript{184,185} between PEO and a number of proteins—lysozyme, chymotrypsinogen A, bovine serum albumin, β-lactoglobulin, RNase A, tubulin—found the preferential solvent interaction parameter, $(\partial g_{PEO}/\partial g_{protein})_{T,\mu_{water},\mu_{PEO}}$, to be negative and increasing in magnitude with PEO molecular weight, showing that PEO is excluded from the proteins due to a preferential hydration of the proteins. Increases in the protein chemical potential with added PEO further indicate repulsive interactions between protein and polymer. Some different proposed theories for the PEO repulsive and antifouling interactions with proteins included excluded volume, osmotic repulsion, and high mobility of PEO,\textsuperscript{186} a lack of PEO protein-binding sites, a low interfacial free energy of PEO and water, and the highly hydrated surface of PEO.\textsuperscript{187,188} Aqueous systems of PEO homopolymers demonstrated that each ethylene oxide unit is associated with three water molecules, forming a protective hydration shell around the polymer.\textsuperscript{189,190} The inhomogeneous protein surface, however, can lead to local interactions of nonpolar protein patchy regions with PEO through hydrophobic interactions.\textsuperscript{191}

The molecular structure of mono-PEGylated proteins have been described by two competing models—the shroud model,\textsuperscript{192,193} in which the polymer chain enwraps and shields the
protein, and the dumbbell model, in which the polymer chain is extended away from the protein as a hydrated coil, the former being the more commonly accepted model. Small-angle neutron scattering (SANS) successfully isolated the polymer scattering in mono-PEGylated lysozyme and human growth hormone and corroborated the dumbbell model (Figure 4); the model fitting results suggest repulsive behavior between protein and polymer, consistent with studies by other approaches.

1.6 Thesis Overview

The focus of this thesis is the interactions governing the thermodynamics of globular protein-polymer block copolymer self-assembly, and a series of studies will be presented and discussed in the upcoming chapters with the goal of providing some insight into the fundamental understanding of protein-polymer interactions. First, Chapter 2 discusses the core techniques used throughout this thesis, covering both synthetic and computational methods. Chapter 3 presents and discusses the first study of the self-assembly of a globular protein-polymer block copolymer in concentrated solution, using the model globular protein-polymer block copolymer mCherry-b-poly(N-isopropylacrylamide) (PNIPAM). Chapters 4 and 5 involve variations on the globular protein block that allow the effect of altering protein interactions on the self-assembly behavior to be investigated independently of the protein shape. In Chapter 4, the globular proteins mCherry, enhanced green fluorescent protein (EGFP), and single-point mutation variants of mCherry are studied; in Chapter 5, the effect of protein electrostatics is explored by using a supercharged green fluorescent protein (GFP) and comparing to two other GFP sequences, the superfolder GFP, which has a formal net charge of -8, and a variant mutated to have an overall neutral formal net charge. Chapter 6 uses a combination of experimental small-
angle neutron scattering (SANS) experiments and analysis and coarse-grained molecular dynamics (MD) simulation to resolve the single-molecule conformation of a set of globular protein-polymer block copolymers comprising mCherry conjugated to four different polymers and to investigate the sensitivity of the molecular conformation to protein-polymer interactions.

1.7 References

1. Payen, A. & Persoz, J. F. Memoir on diastase, the principal products of its reactions, and their applications to the industrial arts. in Annales de Chimie et de Physique 53, 73–92 (1833).


69. Minteer, S. D. *Enzyme Stabilization and Immobilization*.


Chapter 2. Experimental Methods

2.1 Synthesis of Globular Protein-Polymer Block Copolymers

2.1.1 Synthesis of Maleimide-Functionalized Polymers

Synthesis of 2-ethylsulfanylthiocarbonyl sulfanyl-2-methylpropionic acid (EMP)

\[
\text{SH} + \text{NaOH} + \text{CS}_2 + \text{CHC}_3 + \text{HCl} \rightarrow \text{\begin{tabular}{c}
\text{S} \\
\text{\text{-}} \\
\text{\text{-}} \\
\text{\text{-}} \\
\text{COOH}
\end{tabular}}
\]

The synthesis of the chain transfer agent EMP was adapted from the work of You and Oupický.\(^1\)

7.21 mL of ethanethiol (6.2 g, 0.1 mol) was added to a 1 L round bottom flask, to which 73.3 mL of acetone was added. Then 1.13 mL Aliquat336 (1.0 g, 0.002 mol) was added, and the reaction mixture was cooled in an ice bath under active nitrogen flow. A solution of 50 g NaOH in 100 mL H\(_2\)O was prepared, and 9 mL was added to the reaction mixture dropwise over 10 minutes. A solution of carbon disulfide (6.03 mL, 7.6 g, 0.1 mol) in 12.64 mL of acetone was added using a dropping funnel over 30 minutes. 12 mL of chloroform (17.8 g, 0.15 mol) was then added. 80 mL of the concentrated NaOH solution prepared earlier (40 g, 1.0 mol) was added dropwise over a total of 20 minutes. After allowing the mixture to stir overnight under nitrogen, phase separation was observed, and the reaction turned a very dark red/black color. 200 mL of H\(_2\)O was added, followed by 80 mL of concentrated HCl (36.5 - 38.0%). The product was extracted in the organic phase by washing 3 times using diethyl ether. After removing the ether by rotary evaporation, a deep red black, viscous liquid was obtained. Next, the product was purified by column chromatography using approximately 400 mL of silica gel and 1:1 hexanes:ether as the mobile phase (Figure 2-1). The first fraction was a bright yellow byproduct that was discarded. The second fraction, which contained the desired product, was orange-brown and travelled through the column in broad streaks. After collecting the second fraction, solvent was removed
by rotary evaporation. EMP was then purified by vacuum distillation using a micro distillation apparatus. Distillation was started in an oil bath set at approximately 80 °C, and the temperature was gradually increased under vacuum. At approximately 110 °C and 300 mTorr, the first fraction containing a yellow byproduct began distilling and was collected until the vacuum level decreased to about 150 mTorr. It was difficult to maintain a low vacuum, and a heat gun was used to help facilitate the distillation carefully without overheating and causing the mixture to "bump." The second fraction began to collect at 170 °C and 500 mTorr, and the bright orange red fraction was distilled into a clean collecting flask. The product was transferred to a 20 mL scintillation vial and allowed to cool to room temperature before being transferred to the -20 °C freezer. ¹H NMR (CDCl₃, δ): 1.33 (t, 3H, -CH₂CH₃), 1.73 (s, 6H, -C(CH₃)₂COOH), 3.30 (q, 2H, -CH₂CH₃).
Figure 2-1. Column chromatography purification of 2-ethylsulfanylthiocarbonyl sulfanyl-2-methylpropionic aid (EMP) in 1:1 hexanes:ether. (left) The green highlighted rectangle shows the yellow fraction that elutes first. (right) The green highlighted Erlenmeyer flask shows the first yellow fraction, and the blue highlighted round bottom shows collection of the orange-brown second fraction.
Synthesis of exo-3a,4,7,7a-Tetrahydro-2-(3-hydroxypropyl)-4,7-epoxy-14-isoindole-1,3(2H)-dione (imide)

\[
\begin{align*}
\text{O} & \quad + \quad \text{MeOH} \quad \xrightarrow{56 \degree C \text{ for } 3 \text{ days}} \\
\text{O} & \quad + \quad \text{NH}_2\text{CH}_2\text{CH}_2\text{OH} \quad \rightarrow \\
\text{O} & \quad + \quad \text{H}_2\text{O}
\end{align*}
\]

This reaction was based on the work of Neubert and Snider.\textsuperscript{2} 2.08 mL (2.04 g, 0.027 mol) of 3-amino-1-propanol was added dropwise to a solution of exo-3,6-epoxy-1,2,3,6-tetrahydropthalic anhydride (4.5 g, 0.027 mol) in 250 mL of MeOH in a 1 L round bottom flask. The reaction was stirred at 56 °C for at least 3 days. Methanol was then removed by rotary evaporation to give a clear yellow oil. 100 mL of dichloromethane was added, and the mixture became cloudy. The
mixture was washed 3 times, each with approximately 100 mL of brine. The organic phase was
dried over sodium sulfate, rotovapped, and then dried under vacuum overnight to give a solid
white powder. Yield: 1.3 g (22%). $^1$H NMR (CDCl$_3$, $\delta$): 1.75 (tt, 2H, -CH$_2$(CH$_2$)CH$_2$-), 2.88 (s,
2H, -NC(O)CH-), 3.52 (t, 2H, -CH$_2$O-), 3.65 (t, 2H, -NCH$_2$-), 5.27 (s, 2H, -CH(O)-), 6.52 (s, 2H,
-CHCH-).

![Figure 2-3. $^1$H NMR of exo-3a,4,7,7a-Tetrahydro-2-(3-hydroxypropyl)-4,7-epoxy-14-isoindole-
1,3(2H)-dione in deuterated chloroform. The peak at 7.26 ppm corresponds to chloroform. The
broad peak at 1.57 ppm is residual water.](image)

**Synthesis of EMP-imide RAFT agent**

![Synthesis of EMP-imide RAFT agent](image)

N,N'-dicyclohexylcarbodiimide (DCC) coupling was used to synthesize the maleimide-
functionalized chain transfer agent EMP-imide as detailed in Bays et al.$^3$ EMP (1.846 g, 8.23
mmol), imide (2.02 g, 9.05 mmol), and DCC (2.04 g, 9.87 mmol) were dissolved in 75 mL of dry dichloromethane in a 250 mL round bottom flask and stirred at room temperature under ambient pressure overnight. The solvent was removed by rotary evaporation, and the product was purified by column chromatography using 1:1 hexanes:ethyl acetate as the mobile phase. The first yellow fraction contained unreacted EMP and was discarded. The second yellow fraction was collected, rotovapped, and dried under vacuum overnight. Yield: 2.35 g (66%). $^1$H NMR (CDCl$_3$, $\delta$): 1.34 (t, 3H, -S-CH$_2$CH$_3$), 1.73 (s, 6H, -C(CH$_3$)$_2$-), 1.96 (tt, 2H, -CH$_2$(CH$_2$)CH$_2$-), 2.87 (s, 2H, -NC(O)CH-), 3.30 (q, 2H, -CH$_2$CH$_3$), 3.59 (t, 2H, -NCH$_2$-), 4.10 (t, 2H, -CH$_2$O-), 5.28 (s, 2H, -CH(O)-), 6.53 (s, 2H, -CHCH-).

Figure 2-4. $^1$H NMR of EMP-imide in deuterated chloroform.
Sublimation of N-isopropylacrylamide (NIPAM)

N-isopropylacrylamide (TCI America, part no. TCI0401-500g, stabilized with hydroquinone monomethyl ether (MEHQ)) was loaded into a 50 mL sublimation apparatus (Chemglass Life Sciences, part no. CG-3036) with a circulating water jacket at about 4 - 6 °C and submerged in an oil bath set at 60 °C. Sublimation was performed under vacuum at approximately 100 mTorr. Normally, about 20 g of NIPAM were loaded into the sublimation apparatus, and sublimation was allowed to proceed until a small yellow-white solid remained. The residue was approximately 5% of the original mass, containing some NIPAM and the inhibitor MEHQ, and was discarded. The purified NIPAM was collected from the sides of the sublimation apparatus and transferred to a glass jar, which was sealed with Parafilm and stored at -20 °C.

Purification of Hydroxypropyl Acrylate

2-hydroxypropyl acrylate (HPA) (Sigma-Aldrich, 95%, mixture of isomers, part no. 370932-1L) was passed through a basic alumina column to remove MEHQ inhibitor. Approximately 50 mL of monomer was added to a column with a diameter of 0.625 in. Approximately 1 in. of basic alumina was loaded for 10 mL of monomer. Clear and colorless liquid monomer was collected in a beaker submerged in an ice bath over a period of 8 hours, transferred to a 20 mL scintillation vial, and stored at 4 °C in a chemical refrigerator.

Purification of 2-Methoxyethyl Acrylate

2-methoxyethyl acrylate (MEA) (Sigma-Aldrich, 98%, part no. 408913-250mL) was passed through a basic alumina column to remove MEHQ inhibitor. Approximately 50 mL of monomer was added to a column with a diameter of 0.625 in. Approximately 1 in. of basic alumina was
loaded for 10 mL of monomer. Clear and colorless liquid monomer was collected in a beaker submerged in an ice bath, transferred to a 20 mL scintillation vial and stored at 4 °C in a chemical refrigerator.

**Synthesis and Purification of 2-(2-Methoxyethoxy)Ethyl Acrylate**

The synthesis of 2-(2-methoxyethoxy)ethyl acrylate (MEEA) was referenced from the works of Auvergne et al.\(^4\) and Hua et al.\(^5\) 20 mL of dry dichloromethane were added to a 100 mL 3-neck flask and kept cool in an ice bath. 4.95 mL 2-(2-methoxyethoxy)ethanol and 7.33 mL trimethylamine were added. 10 mL of dry dichloromethane and 3.71 mL acryloyl chloride were added dropwise to the solution while stirring to give an opaque yellow mixture. After mixing all the reagents, the reaction was stirred in an ice bath for 1 hour. The ice bath was removed, and the reaction was allowed to stir for 2 more hours; the mixture turned a dark yellow, forming a salt layer on top. The reaction was filtered to remove Et₃NHCl salt. The mixture was then washed twice with 0.1 M HCl(aq) to remove Et₃N, twice with saturated NaHCO₃(aq) to remove acrylic acid, and twice with 0.5 M NaCl(aq) to remove salts. The resulting mixture was brown orange and opaque and was dried with sodium sulfate. The sodium sulfate was removed by filtration, and the solution was rotovapped to give a liquid orange-brown product. The solution was then passed through a silica gel column using 4:3 hexanes:ethyl acetate as the mobile phase and then rotovapped to give a clear, slightly yellow liquid. Approximately 5 % by weight of hydroquinone inhibitor was added, and the solution was vacuum distilled at approximately 150 mTorr and 90 °C. The product was dissolved in dichloromethane and washed quickly with 0.1 M KOH(aq) 4 times and once with water to remove hydroquinone, resulting in a clear product.
Polymerization of NIPAM

Polymerization of N-isopropylacrylamide (NIPAM) was performed by reversible addition-fragmentation chain transfer (RAFT) polymerization to obtain low dispersity polymer at 55 °C in acetonitrile using Schlenk line techniques. The CTA used was EMP-imide, and azobisisobutyronitrile (AIBN) (recrystallized twice from methanol) was used as the initiator. The initial NIPAM concentration was 2 M, and the monomer to CTA ratio was varied from
300:1 to 1600:1 to target different molecular weights. Three freeze-pump-thaw cycles were performed. The first vacuum pump cycle normally achieved a vacuum level of about 80 - 90 mTorr, and by the final pump cycle, the vacuum level achieved was about 50 - 60 mTorr. Due to coupling through the protected maleimide functionality, polymerizations were limited to a maximum conversion of 30% to minimize the presence of higher molecular weight coupling species. Reactions were quenched in liquid nitrogen and exposed to air. The solvent was removed by rotary evaporation, leaving solid chunks of yellowish polymer and white unreacted monomer. The PNIPAM/NIPAM mixture was dissolved in milliQ water overnight. Vacuum filtration was performed to ensure the removal of any unreacted AIBN. Three rounds of thermal precipitation at 40 °C using a Sorvall RC 6+ centrifuge at 11000 rpm for 1.75 hr were performed to remove unreacted monomer. PNIPAM was then lyophilized for at least 3 days and transferred to a 20 mL scintillation vial. The deprotection of the maleimide group was accomplished through a retro-Diels-Alder reaction by heating the solid polymer at approximately 120 °C under vacuum for 2 hours. PNIPAM was stored at -20 °C in scintillation vials and sealed with Parafilm.

**Polymerization of HPA and Copolymerization of MEA and MEEA**

Experiments were performed to identify synthesis conditions for poly(HPA) (PHPA) and poly(MEA-co-MEEA) by RAFT. Henceforth, poly(MEA-co-MEEA) will be referred to as poly(oligoethylene glycol acrylate) (POEGA). As with the polymerization of PNIPAM, EMP-imide and AIBN were used as the CTA and initiator, respectively. For PHPA, a time series was taken with [HPA]₀ = 4 M and [HPA]:[EMP-imide]:[AIBN] = 200:1:0.2. The polymerization of PHPA was carried out in acetonitrile at 60 °C in a 100 mL Schlenk flask that was degassed by
three freeze-pump-thaw cycles. Time points were taken via cannula transfer under nitrogen, and molecular weights and dispersities were determined by gel permeation chromatography using a Waters Breeze 1525 HPLC system with a series 2414 refractive index detector, calibrated with poly(methyl methacrylate) standards, and \(\text{N},\text{N}-\text{dimethylformamide (DMF)}\) with 0.01 LiBr as the mobile phase. A time series of the polymerization of POEGA was carried out in 1,4-dioxane at 60 °C in a total reaction volume of 6 mL with \([\text{MEA}]_0 = [\text{MEEA}]_0 = 1\ \text{M}\) and \([\text{MEA}]:[\text{MEEA}]:[\text{EMP-imide}]:[\text{AIBN}] = 120:120:1:0.2\). A 100 mL Schlenk flask was used for the reaction and was degassed three times by freeze-pump-thaw before polymerization. Time series of PHPA and POEGA shown in Figures 2-5 and 2-6 showed that RAFT polymerization with a maleimide-functionalized CTA was successful. However, reaction conditions still needed to be optimized to obtain better controlled polymerization with lower dispersities, and challenges of polymer deprotection and bioconjugation needed to be solved to incorporate these polymers into protein-polymer conjugates. The final polymerization conditions for PHPA and POEGA are detailed in the thesis of Dongsook Chang (Ph.D., 2015).
Figure 2-6. Time series of PHPA.

Figure 2-7. Time series of POEGA.
2.1.2 Synthesis of Globular Proteins

Protein Cloning. This thesis uses the fluorescent model globular proteins mCherry, enhanced green fluorescent protein (EGFP), and green fluorescent protein (GFP). The fluorescence properties of these model proteins made assays of their functionality relatively quick and easy using UV-Vis spectroscopy.

mCherry. The pQE80 vector containing the DNA encoding mCherry was originally provided in the DH10B cell line as a gift from the lab of Prof. David Tirrell. The cells were grown on an LB agar plate with ampicillin resistance. After growing a 5 mL culture in LB media overnights, the plasmid DNA was extracted by performing a miniprep and then transformed into the *Escherichia coli* expression cell line SG13009. SG13009 cells contain the repressor gene pREP4, a plasmid with kanamycin resistance that encodes for the *lac* repressor, which limits expression of the protein gene until induction by isopropyl β-D-1-thiogalactopyranoside (IPTG). Because the natural sequence of mCherry does not contain any cysteine residues, it is amenable to single site-specific bioconjugation by Michael addition using thiol-maleimide chemistry, contingent upon introducing a single cysteine residue without disrupting the folded protein structure. The serine residue in the 131st position of mCherry was selected to be changed to cysteine for bioconjugation, as this residue was exposed in a loop region on the end of the β-barrel opposite to the N- and C-termini. The gene encoding this variant, mCherryS131C, with codons optimized for *E. coli* expression, was purchased and placed into the pQE9 plasmid using a double digest and ligation at the BamHI and HindIII restriction sites. The plasmid containing the DNA for mCherryS131C was originally prepared by Carla Thomas and is detailed in her thesis (Ph.D., 2014).
**EGFP.** For studies with EGFP, the gene containing a 6xHis tag at the N-terminus was provided in the plasmid pQE80L as a gift from the lab of Prof. David Tirrell. The EGFP gene sequence showed that there were a total of 3 cysteine residues in the protein, and from referencing the crystal structure (PDB 2Y0G), the three cysteines were located at the N-terminus (residue number 14), near an edge of the cylindrical β-barrel structure (residue number 65), and inside the β-barrel near the chromophore (residue number 87). Site-directed mutagenesis was used to prepare the following series of EGFP variants in which different combinations of cysteines were substituted by serine: EGFPC14S, EGFPC65S, EGFPC14S&C65S, EGFPC87S, EGFPC65S&C87S, and EGFPC14S&C65S&C87S. Fluorescence spectroscopy of the variants after expression and cell lysis (Figure 2-8) showed that only the C14S variant successfully expressed and retained an active chromophore. No combinations in which mutations involving the cysteine near the chromophore (C87S) resulted in an active chromophore, demonstrating the importance and sensitivity of specific interactions near the chromophore to lead to proper chromophore maturation.

A test bioconjugation reaction to 5k maleimide poly(ethylene glycol) (PEG) in 20 mM Tris-Cl, pH = 8.0, showed that only the single cysteine in the loop region of the N-terminus (residue number 14) was solvent-accessible (Figure 2-9). Test bioconjugation results show that EGFPC14S, in which the N-terminal cysteine is mutated to serine, is unable to react with 5k maleimide PEG. The cysteine near the chromophore (residue number 87) is sufficiently buried within the β-barrel and is unreactive toward 5k maleimide PEG.
Figure 2-8. Fluorescence spectroscopy of EGFP and different EGFP variants in which a different combination of cysteines are substituted with serine. Variants in which the cysteine near the chromophore is substituted with serine (C87S) do not retain an active chromophore.

Figure 2-9. SDS-PAGE showing selective bioconjugation to EGFP. Lanes 1-3 represent ladder, crude bioconjugation reaction with EGFP, and crude bioconjugation reaction with EGFP C14S. Test bioconjugations show that only one of the three cysteines in EGFP is solvent-accessible.
GFP. GFP variants with overall net charges of approximately -30, -20, -10, and 0 were designed to explore the effect of protein electrostatics on globular protein-polymer block copolymer self-assembly. As reported by Liu and coworkers, a supercharged GFP gene sequence with a net formal charge of -30 is available to use. The sequence for the starting "superfolder" GFP variant (sfGFP), which has a net charge of -7, reasonably close to -10, was used as the control.

Next, for the variant with a net charge of -20, the sequence reported by Liu and coworkers for the GFP (-25) sequence was used as the base sequence, and some amino acids were "un-mutated" back to the starting "superfolder" GFP to design a gene with a net charge of -20. Figure 2-10 shows the crystal structure for sfGFP with the amino acids that are mutated to reach a net charge of -25 shown as sticks and spheres. The three amino acids shown as spheres are the three that were mutated to return the residue to its original identity in the parent sequence.
superfolder sequence. The mutations involved E52K, D214K, and D230T, giving a total change in the net charge of +5 relative to GFP (-25).

![Image of crystal structure](Image)

**Figure 2-11.** Crystal structure of the superfolder GFP (PDB 2B3P) and the residues that are mutated to reach a net charge of +15, shown as sticks and spheres. The five amino acids shown as spheres are potential candidates for mutation to reach a net charge of approximately 0.

For the variant with a net charge of approximately 0, the sequence for a variant with a net charge of +15, reported in a publication by McNaughton et al., was used as a reference. Figure 2-11 shows the crystal structure for sfGFP with the amino acids that are mutated to reach a net charge of +15 shown as sticks and spheres. The five amino acids shown as spheres are the five that were chosen to be mutated from the starting sfGFP sequence. These mutations correspond to D133K, Q157R, E172K, D197R, and N212K, giving a total change in the net charge of +7 relative to sfGFP, which has a net charge of -7.

The gene sequences for the GFP variants with net charges of -30, -20, and 0 are shown in Figure 2-12, along with the sequence for the starting GFP, sfGFP, and other variants with net charges of -25, +15, and +36 for reference.

Lastly, an N-terminal cysteine that is available for bioconjugation in the EGFP sequence was incorporated into the sequences of the GFP variants. Adding the cysteine at the N-terminus,
the amino acid sequence GGA in the sequences for the GFP variants was replaced with the sequence containing the cysteine that is in the EGFP sequence, GSACELMV. This added an additional glutamic acid, decreasing the net charge of each mutant by -1, so the gene sequences have net charges of -31, -21, -8, and -1. Then for only the sequence with a net charge of -1, the mutation Q157R was made to change the net charge to 0.

**GFP (-30)**
MGH\textcolor{red}{HHHH}HHS\textcolor{blue}{GA}C\textcolor{red}{ELMV}SKGE\textcolor{red}{ELF\textcolor{red}{P}}GV\textcolor{red}{V}PIL\textcolor{red}{V}L\textcolor{red}{E}DGV\textcolor{red}{N}HG\textcolor{red}{E}FSVRGE\textcolor{red}{EG}G

**GFP (-25)**
MGH\textcolor{red}{HHHH}HHS\textcolor{blue}{GA}C\textcolor{red}{ELMV}SKGE\textcolor{red}{ELF\textcolor{red}{P}}GV\textcolor{red}{V}PIL\textcolor{red}{V}L\textcolor{red}{E}DGV\textcolor{red}{N}HG\textcolor{red}{E}FSVRGE\textcolor{red}{EG}G

**GFP (-20)**
MGH\textcolor{red}{HHHH}HHS\textcolor{blue}{GA}C\textcolor{red}{ELMV}SKGE\textcolor{red}{ELF\textcolor{red}{P}}GV\textcolor{red}{V}PIL\textcolor{red}{V}L\textcolor{red}{E}DGV\textcolor{red}{N}HG\textcolor{red}{E}FSVRGE\textcolor{red}{EG}G

**sfGFP (-7)**
MGH\textcolor{red}{HHHH}HHS\textcolor{blue}{GA}C\textcolor{red}{ELMV}SKGE\textcolor{red}{ELF\textcolor{red}{P}}GV\textcolor{red}{V}PIL\textcolor{red}{V}L\textcolor{red}{E}DGV\textcolor{red}{N}HG\textcolor{red}{E}FSVRGE\textcolor{red}{EG}G

**GFP (-30)**
DAT\textcolor{red}{GELT}LFICTT\textcolor{red}{G}LP\textcolor{red}{P}WPTLVTPLL\textcolor{red}{T}YGV\textcolor{red}{Q}FS\textcolor{red}{P\textcolor{red}{D}}HMPK\textcolor{red}{QHDF}FK

**GFP (-25)**
DAT\textcolor{red}{GELT}LFICTT\textcolor{red}{G}LP\textcolor{red}{P}WPTLVTPLL\textcolor{red}{T}YGV\textcolor{red}{Q}FS\textcolor{red}{P\textcolor{red}{D}}HMPK\textcolor{red}{QHDF}FK

**GFP (-20)**
DAT\textcolor{red}{GELT}LFICTT\textcolor{red}{G}LP\textcolor{red}{P}WPTLVTPLL\textcolor{red}{T}YGV\textcolor{red}{Q}FS\textcolor{red}{P\textcolor{red}{D}}HMPK\textcolor{red}{QHDF}FK

**sfGFP (-7)**
DAT\textcolor{red}{GELT}LFICTT\textcolor{red}{G}LP\textcolor{red}{P}WPTLVTPLL\textcolor{red}{T}YGV\textcolor{red}{Q}FS\textcolor{red}{P\textcolor{red}{D}}HMPK\textcolor{red}{QHDF}FK

**GFP (-30)**
SAM\textcolor{red}{PEGY}V\textcolor{red}{Q}ERT\textcolor{red}{ISFKD}DGTYK\textcolor{red}{T}RAE\textcolor{red}{V}FK\textcolor{red}{E}G\textcolor{red}{D}T\textcolor{red}{L}V\textcolor{red}{N}RI\textcolor{red}{L}K\textcolor{red}{IDF}K\textcolor{red}{E}D\textcolor{red}{G}N

**GFP (-25)**
SAM\textcolor{red}{PEGY}V\textcolor{red}{Q}ERT\textcolor{red}{ISFKD}DGTYK\textcolor{red}{T}RAE\textcolor{red}{V}FK\textcolor{red}{E}G\textcolor{red}{D}T\textcolor{red}{L}V\textcolor{red}{N}RI\textcolor{red}{L}K\textcolor{red}{IDF}K\textcolor{red}{E}D\textcolor{red}{G}N

**GFP (-20)**
SAM\textcolor{red}{PEGY}V\textcolor{red}{Q}ERT\textcolor{red}{ISFKD}DGTYK\textcolor{red}{T}RAE\textcolor{red}{V}FK\textcolor{red}{E}G\textcolor{red}{D}T\textcolor{red}{L}V\textcolor{red}{N}RI\textcolor{red}{L}K\textcolor{red}{IDF}K\textcolor{red}{E}D\textcolor{red}{G}N

**sfGFP (-7)**
SAM\textcolor{red}{PEGY}V\textcolor{red}{Q}ERT\textcolor{red}{ISFKD}DGTYK\textcolor{red}{T}RAE\textcolor{red}{V}FK\textcolor{red}{E}G\textcolor{red}{D}T\textcolor{red}{L}V\textcolor{red}{N}RI\textcolor{red}{L}K\textcolor{red}{IDF}K\textcolor{red}{E}D\textcolor{red}{G}N

**GFP (-30)**
ILGH\textcolor{red}{K}\textcolor{red}{L}\textcolor{red}{E}YNF\textcolor{red}{S}\textcolor{red}{H}NV\textcolor{red}{Y}IT\textcolor{red}{A}\textcolor{red}{D}K\textcolor{red}{Q}\textcolor{red}{N}\textcolor{red}{G}\textcolor{red}{I}K\textcolor{red}{A}\textcolor{red}{F}\textcolor{red}{F}\textcolor{red}{I}\textcolor{red}{R}\textcolor{red}{H}\textcolor{red}{N}\textcolor{red}{V}\textcolor{red}{E}\textcolor{red}{D}\textcolor{red}{G}\textcolor{red}{S}\textcolor{red}{V}\textcolor{red}{Q}\textcolor{red}{L}\textcolor{red}{A}\textcolor{red}{D}\textcolor{red}{H}\textcolor{red}{Y}\textcolor{red}{Q}\textcolor{red}{N}

**GFP (-25)**
ILGH\textcolor{red}{K}\textcolor{red}{L}\textcolor{red}{E}YNF\textcolor{red}{S}\textcolor{red}{H}NV\textcolor{red}{Y}IT\textcolor{red}{A}\textcolor{red}{D}K\textcolor{red}{Q}\textcolor{red}{N}\textcolor{red}{G}\textcolor{red}{I}K\textcolor{red}{A}\textcolor{red}{F}\textcolor{red}{F}\textcolor{red}{I}\textcolor{red}{R}\textcolor{red}{H}\textcolor{red}{N}\textcolor{red}{V}\textcolor{red}{E}\textcolor{red}{D}\textcolor{red}{G}\textcolor{red}{S}\textcolor{red}{V}\textcolor{red}{Q}\textcolor{red}{L}\textcolor{red}{A}\textcolor{red}{D}\textcolor{red}{H}\textcolor{red}{Y}\textcolor{red}{Q}\textcolor{red}{N}

**GFP (-20)**
ILGH\textcolor{red}{K}\textcolor{red}{L}\textcolor{red}{E}YNF\textcolor{red}{S}\textcolor{red}{H}NV\textcolor{red}{Y}IT\textcolor{red}{A}\textcolor{red}{D}K\textcolor{red}{Q}\textcolor{red}{N}\textcolor{red}{G}\textcolor{red}{I}K\textcolor{red}{A}\textcolor{red}{F}\textcolor{red}{F}\textcolor{red}{I}\textcolor{red}{R}\textcolor{red}{H}\textcolor{red}{N}\textcolor{red}{V}\textcolor{red}{E}\textcolor{red}{D}\textcolor{red}{G}\textcolor{red}{S}\textcolor{red}{V}\textcolor{red}{Q}\textcolor{red}{L}\textcolor{red}{A}\textcolor{red}{D}\textcolor{red}{H}\textcolor{red}{Y}\textcolor{red}{Q}\textcolor{red}{N}

**sfGFP (-7)**
ILGH\textcolor{red}{K}\textcolor{red}{L}\textcolor{red}{E}YNF\textcolor{red}{S}\textcolor{red}{H}NV\textcolor{red}{Y}IT\textcolor{red}{A}\textcolor{red}{D}K\textcolor{red}{Q}\textcolor{red}{N}\textcolor{red}{G}\textcolor{red}{I}K\textcolor{red}{A}\textcolor{red}{F}\textcolor{red}{F}\textcolor{red}{I}\textcolor{red}{R}\textcolor{red}{H}\textcolor{red}{N}\textcolor{red}{V}\textcolor{red}{E}\textcolor{red}{D}\textcolor{red}{G}\textcolor{red}{S}\textcolor{red}{V}\textcolor{red}{Q}\textcolor{red}{L}\textcolor{red}{A}\textcolor{red}{D}\textcolor{red}{H}\textcolor{red}{Y}\textcolor{red}{Q}\textcolor{red}{N}

**GFP (-30)**
TP\textcolor{red}{IG}DG\textcolor{red}{P}V\textcolor{red}{L}

55
Figure 2-12. Amino acid sequences for GFP variants with different total net formal charge, based upon the superfolder GFP sequence.

Gene Sequences

mCherry

ATGAGAGGATCGCATCACCATCACCATCACCATCACCGATCCATGTGCTCTAAAGGTGAAGAAGATAACATGGCCATCATCAAGGGAGTTCTATGCCATTCAAGGGTGACATGGGAGGCTCCGTGAGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCACCGACGCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCTCTCCTCCGAGCGGATGTACCCCGAGACCGCGCCCTGAGGGCGAGATCAGAGGCTGAAGCTGAAGGACGGCGCCACTACGCCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAG

mCherryS131C

ATGAGAGGATCGCATCACCATCACCATCACCATACCGATCCATGTGCTCTAAAGGTGAAGAAGATAACATGGCCATCATCAAGGGAGTTCTATGCCATTCAAGGGTGACATGGGAGGCTCCGTGAGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCACCGACGCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCTCTCCTCCGAGCGGATGTACCCCGAGACCGCGCCCTGAGGGCGAGATCAGAGGCTGAAGCTGAAGGACGGCGCCACTACGCCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCGGCGCCACTCCACACCAGCGGCGGATCGTACAAG

56
GGATATTCTGTCTCCGCAGTTCATGTATGGTAGTAAAGCGTACGTTAAACATCCGGCA
CGATATCCCAGATTATCTGAAAACGTCTTTTTCCGGAAGGCTTCAATATGGGAACGCTG
GATTTAGCAGAAGAAGATCCTGTCGGAAAGCAGGACATGAGATGCTACCGAGCAGG
AAGATGGTGGCAGTGAAGGCGAATCAAACAGCAGCTGAAACGTGAGAATGAGGAAGG
ACGTAAACCGCCACAAGTTCAGCGGCTTCCCGGAGAGCGAGAGCGAGATCGGACCTA
GGCGAAGCTGCAAGTCAACCCGGCTGGCAAGCGGAGGGGATGCCACCTACGGCAAGC
CAATCGACTACTTCGAGCTGAAGGAGACACCCTGAGACGGCAGGTGCTGAGCAG

**EGFP**

ATGAGAGGATCGCATCACCATCACCATCACCATCACGGATCCGCATGCGAGCTCATGGTGAG
CAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAACTGGACGGCG
ACGTAAACCGCCACAAGTTCAGCGGCTTCCCGGAGAGCGAGAGCGAGATCGGACCTA
GGCGAAGCTGCAAGTCAACCCGGCTGGCAAGCGGAGGGGATGCCACCTACGGCAAGC
CAATCGACTACTTCGAGCTGAAGGAGACACCCTGAGACGGCAGGTGCTGAGCAG

**EGFPC14S**

ATGAGAGGATCGCATCACCATCACCATCACCATCACGGATCCGCATGCGAGCTCATGGTGAG
CAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAACTGGACGGCG
ACGTAAACCGCCACAAGTTCAGCGGCTTCCCGGAGAGCGAGAGCGAGATCGGACCTA
GGCGAAGCTGCAAGTCAACCCGGCTGGCAAGCGGAGGGGATGCCACCTACGGCAAGC
CAATCGACTACTTCGAGCTGAAGGAGACACCCTGAGACGGCAGGTGCTGAGCAG

AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCCAGTTCGTGAACTGGACGGCG

57
CGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGATCCACTATGATGAAAGCTTAA

GFP(0)

ATGGGACATCACATACCATACCATACAGGATCCGCCTGTGTAACGTGTATGTTTTCCAAAAAGGTGAAGAACGTCCCGGATGCGTCTCCGCTGTTGCTGACTGAGCATTATGGGACATCACATCCGTTGAACTGATGGGTATTGACATGTTAACGGTCACAAATTTTCTGTTCGTGGTGAAAGGCGAAGGTGACGCTACTAACGGTAAACTGACTCTGAAATTCATCTGCACCACCGGCAAGCTGCCGGTCCCATGGCCGACCCTGGTGACCACCTGACCTATGGTGTTCAGTGTTTCTCTTGTTATCCGGACCACATGAAACAGCACGATTTCTTTAAGTCCGCTATGCCGGAGGGTTACGTGCAGGAACGTACGATTTTTTAAAGATGACGGCACTTATAAAACTCGTGCCGAAGTTAAATTTGAGGGCGACACTGGTGAACCCGTCATTACAACTCTCACAACGTTTATATCACCGCTGACAAACAGAAAAACGGCATTTAAAATCCGTCATAACGTTGAAGACGGTCTGTTCAACTGGCTGACCATTACCAGCAAAACACCCGTATCGGTGACGGCCCTGTTCTGCTGCCGCTACAACCATTACCTGTCTACGCAGAGCGCGCTGTCTAAAGACCCGAACGAAACGTGACCATATGGTTCTGCTGGAATTTGTCACCGCGGCGGGCATTACCCATGCGATGGACGAACTGTATAG

GFP(-8)

ATGGGACATCACATACCATACCATACAGGATCCGCCTGTGTAACGTGTATGTTTTCCAAAAAGGTGAAGAACGTCCCGGATGCGTCTCCGCTGTTGCTGACTGAGCATTATGGGACATCACATCCGTTGAACTGATGGGTATTGACATGTTAACGGTCACAAATTTTCTGTTCGTGGTGAAAGGCGAAGGTGACGCTACTAACGGTAAACTGACTCTGAAATTCATCTGCACCACCGGCAAGCTGCCGGTCCCATGGCCGACCCTGGTGACCACCTGACCTATGGTGTTCAGTGTTTCTCTTGTTATCCGGACCACATGAAACAGCACGATTTCTTTAAGTCCGCTATGCCGGAGGGTTACGTGCAGGAACGTACGATTTTTTAAAGATGACGGCACTTATAAAACTCGTGCCGAAGTTAAATTTGAGGGCGACACTGGTGAACCCGTCATTACAACTCTCACAACGTTTATATCACCGCTGACAAACAGAAAAACGGCATTTAAAATCCGTCATAACGTTGAAGACGGTCTGTTCAACTGGCTGACCATTACCAGCAAAACACCCGTATCGGTGACGGCCCTGTTCTGCTGCCGCTACAACCATTACCTGTCTACGCAGAGCGCGCTGTCTAAAGACCCGAACGAAACGTGACCATATGGTTCTGCTGGAATTTGTCACCGCGGCGGGCATTACCCATGCGATGGACGAACTGTATAG

GFP(-21)

ATGGGACATCACATACCATACCATACAGGATCCGCCTGTGTAACGTGTATGTTTTCCAAAAAGGTGAAGAACGTCCCGGATGCGTCTCCGCTGTTGCTGACTGAGCATTATGGGACATCACATCCGTTGAACTGATGGGTATTGACATGTTAACGGTCACAAATTTTCTGTTCGTGGTGAAAGGCGAAGGTGACGCTACTAACGGTAAACTGACTCTGAAATTCATCTGCACCACCGGCAAGCTGCCGGTCCCATGGCCGACCCTGGTGACCACCTGACCTATGGTGTTCAGTGTTTCTCTTGTTATCCGGACCACATGAAACAGCACGATTTCTTTAAGTCCGCTATGCCGGAGGGTTACGTGCAGGAACGTACGATTTTTTAAAGATGACGGCACTTATAAAACTCGTGCCGAAGTTAAATTTGAGGGCGACACTGGTGAACCCGTCATTACAACTCTCACAACGTTTATATCACCGCTGACAAACAGAAAAACGGCATTTAAAATCCGTCATAACGTTGAAGACGGTCTGTTCAACTGGCTGACCATTACCAGCAAAACACCCGTATCGGTGACGGCCCTGTTCTGCTGCCGCTACAACCATTACCTGTCTACGCAGAGCGCGCTGTCTAAAGACCCGAACGAAACGTGACCATATGGTTCTGCTGGAATTTGTCACCGCGGCGGGCATTACCCATGCGATGGACGAACTGTATAG

58
ACGGTCATGAGTTTCTGGCGTGCGAAGGTGAAGGTGATGCTACCGAAGGCGAGCTGACTCTGAAATTTATTTGCACCACCGGTAAACTGCCGGTTCCGTGGCCAACGCTG
GTTACCACTCTGACGTATGGTGTACAGTGCTTTTCCCGCTACCCAGACCACATGAAACAACGACTTCTTCAAGTCTGCCATGCCGGAAGGCTACGTGCAGGAACGTACGAT
CAGCTTCAAGGATGATGGTACTTACAAAACTCGCGCAGAAGTGAATTTCAACCGGCGCTGGTATACACACGGTATGGATGAGCTGTATAAA

GFP(-31)
ATGGGACATCACCATCACCATCACGGATCCGCGTGCGAACTGATGGTGTCCAAAGGC
CGAAGAACTGTTTGACGGCGTTGTGCCGATCCTGGTAGAGCTGGATGGTGATGTGAA
CGGTCACGAGTTAGCTGGCCGATCACTATCAGCAGAATACTCCTATCGGTGATGGCCC
GGTACTGCTGCCGGACGACCACTACCTGTCCACTGAGTCTGCCCTGTCCAAAGATCC
GAACGAAAAACGTGACCATATGCTGCTGTGAATTTGTAACCGGCGCGTGGTATACAC
CCACCGTAGATGGATGAGCTGTATAAG

Expressed Protein Amino Acid Sequences

<table>
<thead>
<tr>
<th>mCherry</th>
<th>mCherryS131C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MRGSHHHHHHHSVMVGFEED</td>
</tr>
<tr>
<td>21</td>
<td>NMAIKKEFMRFKVHMEGSVN</td>
</tr>
<tr>
<td>41</td>
<td>GHEFEIIEEIEGEPRYETGQT</td>
</tr>
<tr>
<td>61</td>
<td>AKLKVTKGGLPFAWDILSP</td>
</tr>
<tr>
<td>81</td>
<td>QFMYGDSKAYKHAPIDPYL</td>
</tr>
<tr>
<td>101</td>
<td>KLSFPEGKFWERVMNFEDGG</td>
</tr>
<tr>
<td>121</td>
<td>VVTVTQDSLQDGEPFIIYKVK</td>
</tr>
<tr>
<td>141</td>
<td>LRGTFNFSDPVMQKKTMTG</td>
</tr>
<tr>
<td>161</td>
<td>EASSRMEFPDGLKGEIKQ</td>
</tr>
<tr>
<td>181</td>
<td>RLKLKDGGHYDAEVKTTYKA</td>
</tr>
<tr>
<td>201</td>
<td>KKPQVLPQAYVNINKLDITS</td>
</tr>
<tr>
<td>221</td>
<td>HNEDYITIVEQY ERAEGRHST</td>
</tr>
<tr>
<td>241</td>
<td>GGMDELYK</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
</tr>
</tbody>
</table>

**EGFP**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>MRGSHHHHHHGSACELMVSK</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>GEEFTGVVIPVLVDGDNV</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>GHSFVGSSGEDEGATYKGLT</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>LFICITGKLVPWPTVTVTT</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>LTYGVCFSRYPDHMKQHDF</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>FKSAMPEGYVQERTIFKDD</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>GNYKTRAEVFKFEGDTLVNRI</td>
<td></td>
</tr>
<tr>
<td>141</td>
<td>ELKGFDFKEDGNILGHKLEY</td>
<td></td>
</tr>
<tr>
<td>161</td>
<td>NNSHNVYMADQKNIGKV</td>
<td></td>
</tr>
<tr>
<td>181</td>
<td>NFKIRHNYEDGSVQLADHYQ</td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>QTNPIDGPVLIDNHYLST</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>QSAJKDFPNEKRDHMVLEF</td>
<td></td>
</tr>
<tr>
<td>241</td>
<td>VTAAGITLMGDDELYKGSTMM</td>
<td></td>
</tr>
<tr>
<td>261</td>
<td>KA</td>
<td></td>
</tr>
</tbody>
</table>

**GFP (O)**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>MGHHHHHHGSACELMVSKGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>ELFSDGVVIPVLDEDGDNV</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>KFSVREGEDEGATNGKTLK</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>FICTTGGKLVPWPTVTVTT</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>YGVCFCSRYPDHMKQHDF</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>SAMPEGYVQERTISFDDGT</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>YKTRAEVFKFEGDTLVNRIEL</td>
<td></td>
</tr>
<tr>
<td>141</td>
<td>KGIDFKEKGNILGHKLEYNF</td>
<td></td>
</tr>
<tr>
<td>161</td>
<td>NSHNVYITADKQNGKIKAF</td>
<td></td>
</tr>
<tr>
<td>181</td>
<td>KIRHNYLDGVSDLADHYQON</td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>TPGDGPVLIDNHYLSTQS</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>ALSKDPNEKRDHMVLEFVT</td>
<td></td>
</tr>
<tr>
<td>241</td>
<td>AAGITHGMDELYK</td>
<td></td>
</tr>
</tbody>
</table>

**GFP (-8)**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>MGHHHHHHGSACELMVSKGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>ELFSDGVVIPVLDEDGDNV</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>KFSVREGEDEGATNGKTLK</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>FICTTGGKLVPWPTVTVTT</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>YGVCFCSRYPDHMKQHDF</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>SAMPEGYVQERTISFDDGT</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>YKTRAEVFKFEGDTLVNRIEL</td>
<td></td>
</tr>
<tr>
<td>141</td>
<td>KGIDFKEKGNILGHKLEYNF</td>
<td></td>
</tr>
<tr>
<td>161</td>
<td>NSHNVYITADKQNGKIKAF</td>
<td></td>
</tr>
<tr>
<td>181</td>
<td>KIRHNYLDGVSDLADHYQON</td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>TPGDGPVLIDNHYLSTQS</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>ALSKDPNEKRDHMVLEFVT</td>
<td></td>
</tr>
</tbody>
</table>

**GFP (-31)**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>MGHHHHHHGSACELMVSKGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>ELFSDGVVIPVLDEDGDNV</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>KFSVREGEDEGATNGKTLK</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>FICTTGGKLVPWPTVTVTT</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>YGVCFCSRYPDHMKQHDF</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>SAMPEGYVQERTISFDDGT</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>YKTRAEVFKFEGDTLVNRIEL</td>
<td></td>
</tr>
<tr>
<td>141</td>
<td>KGIDFKEKGNILGHKLEYNF</td>
<td></td>
</tr>
<tr>
<td>161</td>
<td>NSHNVYITADKQNGKIKAF</td>
<td></td>
</tr>
<tr>
<td>181</td>
<td>KIRHNYLDGVSDLADHYQON</td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>TPGDGPVLIDNHYLSTQS</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>ALSKDPNEKRDHMVLEFVT</td>
<td></td>
</tr>
</tbody>
</table>

| 241 | AAGIDGHGMDELYK |

60
Protein Expression. The proteins mCherry, mCherryS131C, and EGFP, each containing an N-terminal 6xHis tag, were expressed in the *E. coli* strain SG13009 containing the pREP4 repressor plasmid. A freshly grown bacterial colony was inoculated in 5 mL Lysogeny broth (LB) medium supplemented with 0.2 mg/mL ampicillin and 0.05 mg/mL kanamycin at 37 °C overnight for between 12 - 18 hours. At the start of my graduate studies, mCherryS131C was expressed in Terrific broth (TB), but the richer medium did not provide ample time for the protein to fold during expression, which did not consistently result in a high yield of folded mCherryS131C (culture flasks did not become a dark red purple color). Similarly unreliable expression results were found when using the 2xYT medium.

A typical expression consisted of inoculating 1 L of LB medium in a 2.8 L Fernbach flask with 5 mL of overnight culture and incubating until OD$_{600}$ ≈ 0.8 - 1.0 at 37°C and 29°C for mCherry/mCherryS131C and EGFP, respectively. Expressions were induced with 1 mM IPTG, and cells were allowed to continue to grow for 6 hours at 37 °C and for 12 hours at 29 °C for mCherry/mCherryS131C and EGFP, respectively. Successful expressions resulted in cultures that were visibly violet red (mCherry) or pale yellow-green (EGFP). The cells were subsequently harvested by centrifugation, and the pellets were stored at -80 °C until they were purified.

The proteins GFP(0), GFP(-8), and GFP(-21), each containing an N-terminal 6xHis tag, were transformed into Tuner(DE3) cells for overexpression. A typical expression for GFP(0) and GFP(-8) consisted of inoculating 1 L of LB medium with 5 mL of overnight culture and incubating until OD$_{600}$ ≈ 0.8 - 1.0 at 37 °C, after which the cultures were induced with 1 mM IPTG and harvested after growing for 9 - 10 hours at 25 °C. GFP(-21) was expressed for 11 - 12
hours at 37 °C after inoculating and then harvested. Cell pellets were stored at -80°C until they were purified.

Brief experiments near the beginning of my graduate studies were conducted to ferment mCherry S131C in a 7 L single wall autoclavable bioreactor manufactured by Applikon Biotechnology. A typical fermentation consisted of inoculating 5 L of LB medium with 25 mL of overnight culture and allowing the mixture to ferment for approximately 20 - 24 hours. Expression of mCherry S131C was visually confirmed by the entire fermenter turning a deep red purple color (Figure 2-13). Approximately 1.2 - 1.4 g of protein were obtained after purification from a 5 L fermentation, but the yield most likely could have been improved had expression conditions been optimized and the expression been induced with IPTG. For the quantity of protein that was needed for bioconjugation and self-assembly characterization, the yield from multiple expressions in 1 L Fernbach flasks was sufficient, and plans to optimize fermentation conditions were not pursued further.

![Figure 2-13. Fermentation of mCherry S131C.](image-url)
Protein Purification. Frozen cell pellets were resuspended in lysis buffer, incubated with 1 mg/mL lysozyme at 4 °C for 30 min, and sonicated using a Branson Sonifier 250 set at a duty cycle of 50% and an output control of 5. The volume sonicated was limited to 100 - 200 mL and was carefully submerged in an ice bath during the sonication to keep the temperature of the protein mixture from increasing. The sonication time was limited to a maximum of 15 minutes to prevent the mixture from overheating. Generally, 15 minutes is not sufficient to lyse all the cells in a 200 mL volume, as the cell debris after centrifugation was observed to contain some reddish purple or green cell mass (corresponding to the purification of mCherry or EGFP/GFP, respectively). Therefore, two 15 minute cycles were used to lyse the cells in a 200 mL volume, and the mixture was allowed to cool down between sonication cycles. The lysate was clarified by centrifuging in a Fiberlite F13-14 x 50ey fixed angle rotor at 12000 rpm for 20 minutes at 4 °C two times. While the first centrifugation removed most of the cell debris, a small amount of particulate matter was spun down in the second spin. A second spin is not necessary but probably helps to reduce fouling of the Ni-NTA resin. The protein was then purified by Ni-NTA metal affinity chromatography. Ni-NTA was added to the clarified supernatant (approximately 10 mL of Ni-NTA used to bind 100 mg protein) and rocked at 4 °C for at least 2 hours and up to overnight (approximately 8 hours).

To purify the protein from 1 L of bacterial cell culture, the mixture containing the protein solution and Ni-NTA resin was loaded onto a Bio-Rad Econo-Column 20 cm in length with an inner diameter of 2.5 cm. The approximately 100 mL of flowthrough solution was collected in an Erlenmeyer flask. The resin was then washed twice using 50 mL of wash buffer each time. Afterward, elution buffer was used to elute the protein. When the elution fraction began to turn a bright purple or green for mCherry or EGFP/GFP, respectively, a fresh 50 mL conical tube was
used to collect the pure protein fraction until the elution fraction was colorless. During the wash and elution steps, buffers were added slowly to avoid disturbing the settled Ni-NTA resin.

Protein purification was scaled up to purifying 4 L of protein expression at a time using approximately 40 mL of Ni-NTA resin in a larger Bio-Rad Econo-Column 20 cm in length with an inner diameter of 5.0 cm. The wash buffer quantity was scaled up by a factor of 1.5 to ensure that all the resin was being washed thoroughly in a column of larger diameter, so approximately 600 mL of wash buffer was used to wash 40 mL of Ni-NTA resin. As with the smaller scale purification procedure, the protein was eluted until the elution fraction became colorless. Purity was confirmed by SDS-PAGE (Figure 2-14).

![Figure 2-14](image)

**Figure 2-14.** Representative SDS-PAGE of mCherryS131C (lane 1) and EGFP (lane 2). The two lower molecular weight bands in mCherryS131C correspond to smaller fragments resulting from hydrolysis of the mCherry chromophore acylimine bond during SDS-PAGE analysis.
For purification of proteins containing a cysteine residue, 10 mM β-mercaptoethanol (352 μL in 500 mL of buffer) was included to prevent disulfide bond formation. The following are the compositions of the lysis, wash, and elution buffers used:

<table>
<thead>
<tr>
<th>Lysis buffer</th>
<th>Wash buffer</th>
<th>Elution buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.45 g NaH₂PO₄⋅H₂O (50 mM)</td>
<td>3.45 g NaH₂PO₄⋅H₂O (50 mM)</td>
<td>3.45 g NaH₂PO₄⋅H₂O (50 mM)</td>
</tr>
<tr>
<td>8.77 g NaCl (300 mM)</td>
<td>8.77 g NaCl (300 mM)</td>
<td>8.77 g NaCl (300 mM)</td>
</tr>
<tr>
<td>0.34 g imidazole (10 mM)</td>
<td>0.68 g imidazole (20 mM)</td>
<td>8.50 g imidazole (250 mM)</td>
</tr>
<tr>
<td>500 mL milliQ H₂O</td>
<td>500 mL milliQ H₂O</td>
<td>500 mL milliQ H₂O</td>
</tr>
<tr>
<td>pH = 8.0</td>
<td>pH = 8.0</td>
<td>pH = 8.0</td>
</tr>
</tbody>
</table>

Table 1. Table of protein and dialysis buffer composition and molar extinction coefficients for excitation peaks and 280 nm.

<table>
<thead>
<tr>
<th>Protein</th>
<th>MW (kDa)</th>
<th>Yield / L (mg)</th>
<th>Dialysis buffer</th>
<th>Excitation peak(s) (nm)</th>
<th>Excitation peak(s) (M⁻¹ cm⁻¹)</th>
<th>ε₂₈₀ (M⁻¹ cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mCherry</td>
<td>28.13</td>
<td>100</td>
<td>20 mM Tris-Cl, pH = 8.0</td>
<td>586</td>
<td>ε₅₈₆ = 72000</td>
<td>34380</td>
</tr>
<tr>
<td>EGFP</td>
<td>29.46</td>
<td>50</td>
<td>20 mM Tris-Cl, pH = 8.0</td>
<td>488</td>
<td>ε₄₈₈ = 56000</td>
<td>21890</td>
</tr>
<tr>
<td>GFP(0)</td>
<td>28.49</td>
<td>60</td>
<td>20 mM Tris-Cl, 250 mM NaCl, pH = 8.0</td>
<td>485</td>
<td>ε₄₈₅ = 83300</td>
<td>19035</td>
</tr>
<tr>
<td>GFP(-8)</td>
<td>28.40</td>
<td>70</td>
<td>20 mM Tris-Cl, pH = 8.0</td>
<td>485</td>
<td>ε₄₈₅ = 83300</td>
<td>19035</td>
</tr>
<tr>
<td>GFP(-21)</td>
<td>28.43</td>
<td>15</td>
<td>20 mM Tris-Cl, pH = 8.0</td>
<td>490, 395</td>
<td>N/A</td>
<td>19035</td>
</tr>
</tbody>
</table>

The proteins mCherry and EGFP-C14S that do not have a solvent-accessible cysteine residue for thiol-maleimide bioconjugation were purified by Ni-NTA as described above and also by anion-exchange fast protein liquid chromatography (FPLC) with an Äkta Pure 25 to obtain higher purity for static light scattering (SLS) experiments. The binding buffer used was
20 mM Tris-Cl, pH = 8.0, which was filtered and degassed prior to purification. A HiTrap Q Sepharose HP 5 mL column from GE Healthcare was equilibrated with at least 15 column volumes (CVs) of buffer. Protein solution was loaded onto the column using a 150 mL superloop at a flow rate of 2.0 mL/min. The column was subsequently washed with at least 10 CVs of binding buffer. The proteins were eluted using an NaCl linear gradient of 0 mM to 250 mM over 18 CVs, collecting fractions every 1.25 mL.

The purity of the fractions was assessed by both UV-Vis analysis, comparing the ratio of the excitation peak absorbance to $A_{280}$, and by SDS-PAGE. For example, Figure 2-15 shows a representative FPLC chromatogram of mCherry purification. The first eluted peak encompasses roughly fractions A12 through C3, and fractions within this range of interest were assayed by UV-Vis and SDS-PAGE.
Figure 2-15. Representative FPLC chromatogram of mCherry. The first peak in the UV flow monitor reading at 280 nm (light blue trace) contains the highest purity mCherry fractions. The lightly shaded blue region corresponds to the fractions collected after analyses by UV-Vis (Table 2) and SDS-PAGE (Figure 2-16).

Table 2. UV-Vis analysis of FPLC fractions corresponding to first elution peak in Figure 2-15.

<table>
<thead>
<tr>
<th>fraction</th>
<th>Dilution for UV-Vis</th>
<th>A280</th>
<th>A586</th>
<th>A586/A280</th>
</tr>
</thead>
<tbody>
<tr>
<td>A12</td>
<td>10</td>
<td>0.73581934</td>
<td>0.6217311</td>
<td>0.84495075</td>
</tr>
<tr>
<td>B1</td>
<td>20</td>
<td>0.91336745</td>
<td>0.86649007</td>
<td>0.94867631</td>
</tr>
<tr>
<td>B2</td>
<td>40</td>
<td>0.51400328</td>
<td>0.65906966</td>
<td>1.28222851</td>
</tr>
<tr>
<td>B3</td>
<td>40</td>
<td>0.48876673</td>
<td>0.74955845</td>
<td>1.53357093</td>
</tr>
<tr>
<td>B4</td>
<td>40</td>
<td>0.45544437</td>
<td>0.76756579</td>
<td>1.68531185</td>
</tr>
<tr>
<td>B5</td>
<td>40</td>
<td>0.42394114</td>
<td>0.74255878</td>
<td>1.75156104</td>
</tr>
<tr>
<td>B6</td>
<td>40</td>
<td>0.39943194</td>
<td>0.71108234</td>
<td>1.78023403</td>
</tr>
<tr>
<td>B7</td>
<td>40</td>
<td>0.36013731</td>
<td>0.64827245</td>
<td>1.80007022</td>
</tr>
<tr>
<td>B9</td>
<td>40</td>
<td>0.27635589</td>
<td>0.50610054</td>
<td>1.8313615</td>
</tr>
<tr>
<td>B12</td>
<td>40</td>
<td>0.13953935</td>
<td>0.24572402</td>
<td>1.76096584</td>
</tr>
<tr>
<td>C1</td>
<td>20</td>
<td>0.19325659</td>
<td>0.31575584</td>
<td>1.63386003</td>
</tr>
<tr>
<td>C2</td>
<td>20</td>
<td>0.1474546</td>
<td>0.22229096</td>
<td>1.50752134</td>
</tr>
<tr>
<td>C3</td>
<td>10</td>
<td>0.26545668</td>
<td>0.37776324</td>
<td>1.42306928</td>
</tr>
</tbody>
</table>
Figure 2-16. SDS-PAGE of (a) FPLC fractions corresponding to first elution peak in Figure 2-15 and (b) mCherry purified by Ni-NTA affinity chromatography only, showing the presence of a small fraction of high molecular weight impurities.

SDS-PAGE (Figure 2-16) shows that fractions B4 through C1 look quite pure and free of high molecular weight impurities (present from purification by Ni-NTA affinity chromatography alone) that are shown in Figure 2-16b. UV-Vis analysis shows a clear correlation between the first elution peak and the $A_{586}/A_{280}$ ratio of the mCherry fractions. In Table 2, the $A_{586}/A_{280}$ ratio increases and reaches a maximum of about 1.83 at fraction B9 and then decreases. For mCherry and mCherryS131C, fractions with a ratio of $A_{586}/A_{280} \geq 1.6 - 1.7$ were combined and dialyzed against 10 mM Tris-Cl, pH = 7.0 (for SLS), and 20 mM Tris-Cl, pH = 8.0 (for bioconjugation), respectively. Fractions that had a ratio of $A_{586}/A_{280} > 1.7$ that juxtaposed the region of fractions with ratios > 1.7 that looked clean by SDS-PAGE were considered pure enough and included in the final purified fraction. Returning to Figures 2-12 and 2-13 and Table 2, the fractions B4 through C1 were combined. EGFPC14S fractions with a ratio of $A_{488}/A_{280} \geq 2.0$ were combined and dialyzed against 10 mM Tris-Cl, pH = 7.0 for SLS measurements. The ratio
A$_{586}$/A$_{280}$ for mCherry/mCherryS131C purified by Ni-NTA only was approximately 1.5, and the ratio A$_{488}$/A$_{280}$ for EGFP purified by Ni-NTA only was approximately 1.8.

mCherryS131C used for bioconjugation to the polymers PNIPAM, poly(hydroxypropyl acrylate) (PHPA), poly(oligoethylene glycol acrylate) (POEGA), and poly(ethylene glycol) (PEG) for dilute solution SANS characterizations were also purified by both Ni-NTA and anion-exchange chromatography prior to bioconjugation to help improve the purity of the protein for bioconjugation.

2.1.3 Bioconjugation

Thiol-maleimide coupling was used to synthesize globular protein-polymer conjugates of mCherry, EGFP, and GFP used in this thesis (Figure 2-17). Coupling reactions were performed in 20 mM Tris-Cl, pH = 8.0 for all conjugates except for the neutral formal net charge GFP(0), for which bioconjugation was carried out in 20 mM Tris-Cl, pH = 8.0, with the addition of 250 mM NaCl, which was necessary to help keep the protein soluble. Protein solutions at approximately 1 - 2 mg/mL were first incubated with 10x molar excess tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) at 4 °C for approximately 45 - 60 minutes. A 6x molar excess of maleimide-functionalized polymer was then added to the solution and
allowed to dissolve and react at 4 °C overnight for up to 20 - 24 hours. The reaction mixture was then precipitated three times in 1 M ammonium sulfate solution at pH = 8.0 to remove unreacted protein. The excess polymer was removed by metal affinity chromatography using Ni-NTA. The conjugate + excess polymer mixture was allowed to bind to Ni-NTA resin at 4 °C overnight for at least 12 hours before column purification at 4 °C. The resin was washed with at least 4 - 5 column volumes of binding buffer 5 times (20 mM Tris-Cl, pH = 8.0 for all conjugates except GFP(-21), for which 20 mM Tris-Cl, 250 mM NaCl, pH = 8.0, was used) to ensure that all excess polymer is removed. The mCherry-PEG conjugates were purified by anion exchange chromatography using 10 mM Tris-Cl, pH = 7.0 as the binding buffer and eluting with NaCl (Figures 2-15 and 2-16).

---

**Figure 2-18.** FPLC chromatogram of mCherry-PEG28k.
Figure 2-19. SDS-PAGE of fractions sampled from D3 through G11 corresponding to the FPLC chromatogram in Figure 2-18. The final elution fraction during the final column elution with high salt concentration to clean and wash the column is also included, labeled “final elute,” showing a larger percentage of mCherry present.

Purified bioconjugate was then dialyzed against milliQ water. At least 8 dialysis changes were performed in intervals of at least 3 hours. Purified bioconjugate was then stored at 4 °C. Figure 2-20 shows representative native and SDS-PAGE gels of mCherry-PNIPAM, -PHPA, -POEGA, and -PEG conjugates and EGFP-PNIPAM conjugate.

Figure 2-20. Native and SDS-PAGE of purified protein-polymer bioconjugates. (a) Native PAGE and (b) SDS-PAGE of mCherry conjugates. Lanes 1 - 5 represent (1) mCherry, (2) mChPEG28k, (3) mChPN26k, (4) mChPOE26k, and (5) mChPHPA30k. (c) Native PAGE and (d) SDS-PAGE of EGFP-PNIPAM13k, which is labeled as lane 6.
2.2 Characterization Techniques

Concentrated Solution Sample Preparation. Conjugate solutions were concentrated to approximately 100 mg/mL using Millipore Ultra-15 centrifugal filters with a molecular weight cutoff of 10 kDa. Solutions were centrifuged at 5410 rpm (5000 g) using a Fiberlite F13-14 x 50cy fixed angle rotor in the Sorvall RC 6 Plus centrifuge or at 4000 rpm (3488 g) in a TX-750 4 x 750 mL swinging bucket rotor in the Sorvall Legend XTR benchtop centrifuge. Bulk solid samples were prepared by drop-casting aliquots of concentrated bioconjugate solution onto Teflon sheets and drying under vacuum overnight at room temperature. Usually, each aliquot was about 15 - 20 µL. A J-KEM Scientific Digital Vacuum Regulator Model 200 was used to control the vacuum ramp rate at 50 mTorr/hr until full vacuum was reached (≈ 25 Torr), after which the samples were allowed to dry further under full vacuum overnight for at least 8 hours. The pellets were then harvested and stored in an Eppendorf tube sealed with Parafilm and at 4 °C until needed. Figure 2-21 shows the process of casting and drying pellets under vacuum and the resulting dried pellets.

![Figure 2-21](image)

**Figure 2-21.** (Left) GFP-PNIPAM conjugate pellets being casted and dried in a vacuum oven. (Right) Photo of dried EGFP- and mCherry-PNIPAM pellets.

Solid-State Sample Preparation. Solid-state samples of mCherry-PNIPAM and EGFP-PNIPAM were prepared and characterized. Samples were prepared at room temperature by
casting from concentrated aqueous solutions at approximately 100 mg/mL into 1 mm thick galvanized aluminum washers with an inner diameter of 5 mm with one side sealed by Kapton tape. Using a J-KEM Scientific Digital Vacuum Regulator, the water evaporation rate was controlled by using a ramp rate of 300 Torr/hr to a final set point of 50 Torr. The samples were then allowed to dry overnight for at least 8 hours. The self-assembly of samples directly after casting was measured by small-angle X-ray scattering (SAXS), and additional samples were annealed in water for 8 hours at room temperature before their self-assembly was characterized by SAXS. The annealing chamber consisted of a glass jar, in which an inverted small beaker was placed (Figure 2-22). MilliQ water was added to the jar to a height of approximately 1 cm. The sample was then placed carefully on top of the inverted beaker, and the jar was sealed and annealed under ambient temperature conditions in a cabinet to shield the fluorescent proteins from light exposure.

![Figure 2-22. Setup of an annealing chamber.](image)

**Small-Angle X-Ray Scattering.** Concentrated solution samples were prepared in 1 mm thick washers and sealed on both sides with Kapton tape. Washers with an inner diameter of 5 mm were used for studies with mCherry- and EGFP-PNIPAM conjugates, but in later studies
involving supercharged GFP-PNIPAM conjugates, washers with an inner diameter of 3 or 3.5 mm were used to reduce the amount of sample volume needed. The X-ray beam spot size is on the order of 1 mm, and the smaller inner diameter dimension was sufficiently large to be able to characterize the samples. Concentrated solution samples were measured between 10 and 40 °C in 5 °C increments, and samples were equilibrated for 10 minutes at each temperature prior to data acquisition. Because the beam energy and flux will vary between different beamlines and also between different experimental trips to the same beamline, an additional test sample was measured before starting actual sample measurements to assess the minimum exposure time needed to obtain good signal-to-noise while minimizing beam damage.

For SAXS experiments conducted at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory, Beamline X27C, the beam current, incident intensity, transmitted intensity, and acquisition time were recorded for every measurement. The following standards were collected:

**Open:** No sample, shutter open, beam on. This is recorded to obtain the incident and transmitted intensities, and the 2D scattering pattern does not need to be recorded. The beam at NSLS operated in decay mode, and this was measured once per injection.

**Background:** A 2D scattering pattern is collected with the beam off for a small amount of time (typically 0.1 seconds). This is used to correct the measured intensities and is measured once per injection.

**Dark:** A 2D scattering pattern with the beam off is collected for the same duration as the sample measurement, ideally, but with time corrections, this measurement
does not have to be collected for the same amount of time as the sample collection time. This is also measured once per beam injection.

**Empty:** This standard refers to an aluminum washer with only Kapton (either one or both sides sealed with Kapton for solid-state and concentrated solution samples, respectively). The 2D scattering pattern is acquired for the same amount of time as the sample is, but with time corrections, it does not have to be collected for the same amount of time as the sample collection time. This measurement is also recommended to be acquired once per injection.

The following equation was used to reduce the sample scattering data:

\[
I(Q) = \frac{1}{T_{\text{sample}} \cdot BC_{\text{sample}}} \left[ \frac{(I_{\text{sample}} - I_{\text{background}})}{t_{\text{sample}}} - \frac{(I_{\text{dark}} - I_{\text{background}})}{t_{\text{dark}}} \right] 
- \frac{1}{T_{\text{empty}} \cdot BC_{\text{empty}}} \left[ \frac{(I_{\text{empty}} - I_{\text{background}})}{t_{\text{empty}}} - \frac{(I_{\text{data}} - I_{\text{background}})}{t_{\text{dark}}} \right] 
\]

(2.1)

where \( I \) is intensity, \( T \) is transmittance, \( BC \) is beam current, and \( t \) is time. Transmittance is calculated as

\[
T_{\text{sample}} = \frac{I_{t, \text{sample}}}{I_{t, \text{open}}} \frac{I_{t, \text{open}}}{I_{t, \text{sample}}} 
\]

(2.2)

Where \( I_t \) is the transmitted intensity and \( I_i \) is the incident intensity. The subscript “sample” corresponds to the values taken when the sample is positioned in the beam path, and the subscript “open” corresponds to the values taken in the “open” configuration.

For SAXS experiments conducted at the Advanced Photon Source (APS) at Argonne National Lab, Beamlines 12-ID-B and 12-ID-C,D, the spec software automatically corrects for
the dark field, distortion from the fiber optics, and the nonuniformity of the beam flux, reducing
the data as it is collected. Only sample and empty measurements were needed at APS.

For SAXS experiments conducted at the Advanced Light Source (ALS) at Lawrence
Berkeley National Lab, Beamline 7.3.3, dark field measurements were not needed, and only
samples and empty Kapton were measured. The Nika Igor software package was used for data
reduction and analysis according to the following equation:

\[ I(Q) = \frac{1}{T_{\text{sample}}} I_{\text{sample}} - \frac{1}{T_{\text{empty}}} I_{\text{empty}} \]  

(2.3)

**Small-Angle Neutron Scattering.** Small-angle neutron scattering (SANS) was performed at the
ORNL HFIR General Purpose Small-Angle Neutron Scattering (GP-SANS) Diffractometer and
the NIST NCNR NG-7 SANS instrument. Measured scattering intensities were corrected for
detector sensitivity and the background from the empty cell and placed on an absolute scale
using a calibrated standard. For form factor measurements, dilute solutions were measured to
avoid the possible coherent scattering contribution from intermolecular spatial correlations, and
solutions were filtered using 0.1 \( \mu \)m Whatman Anotop 10 syringe filters. Protein and
bioconjugate concentration measurements by absorbance at 280 nm were performed after sample
filtration to quantify the exact concentration measured by SANS. Homopolymer solutions were
not filtered directly to avoid precipitation during filtration; instead, the polymer solutions were
prepared by dissolving in solutions filtered with 0.02 \( \mu \)m Whatman Anotop 10 syringe filters.
Contrast matching of mCherry was performed at the GP-SANS beamline instrument using a
wavelength of 4.72 Å at two sample-to-detector distances of 0.3 and 8 m, covering a Q-range of
0.00658 Å\(^{-1}\) to 0.933 Å\(^{-1}\); samples were loaded in quartz Banjo cells (Figure 2-23) with a 1 m
path length. A 1 m\(^2\) area detector with an array of 192 1 m long \( ^3 \)He Linear Position-Sensitive
Detectors with 5.2 mm × 4 mm pixel resolution was used. Dilute solution form factor
measurements were also performed at the NG-7 SANS instrument at the National Institute of Standards and Technology using a wavelength of 6 Å with wavelength spreads of 15% and three sample-to-detector distances of 1, 4, and 13 m, covering a Q-range of 0.00348 Å⁻¹ to 0.518 Å⁻¹. Solution samples were loaded into titanium Hellma cells with quartz windows and a path length of 1 mm. The scattered neutrons were detected by a 640 mm x 640 mm ³He position-sensitive proportional counter with a 5.08 mm x 5.08 mm resolution. Concentrated bioconjugate solutions were measured at the Lujan Center at the Los Alamos Neutron Science Center on the Low Q Diffractometer using a 2-dimensional position-sensitive proportional counter detector measuring 59 cm in diameter. The concentrated solutions were prepared in the same manner as they were for characterization by SAXS.

**Figure 2-23.** Photo showing liquid samples loaded in 1 mm thick Banjo cells at Beamline 6, the Extended Q-Range Small-Angle Neutron Scattering Diffractometer (EQ-SANS) at the Spallation Neutron Source (SNS) at Oak Ridge.

**Guinier Analysis.** For a dilute solution of particles, the scattering intensity per unit of volume containing $N$ identical particles in a volume $V$ is given by

$$I(Q) = \frac{N}{V} \frac{1}{V_{\text{part}}} P(Q) = \Phi \frac{1}{V_{\text{part}}} P(Q)$$

(2.4)

where $V_{\text{part}}$ is the volume of the particle and $P(Q)$ is the form factor of the particle. When $Q$ tends to 0, $P(Q)$ has a general form that depends only on the particle radius of gyration. The intensity for a dilute solution can be written as
\[ I(Q) \approx \Phi \int_{Q_{\text{min}}}^{Q_{\text{max}}} \left[ 1 - \frac{(Q R_G)^2}{3} - \ldots \right] e^{-(Q R_G)^2/3} \approx \Phi \int_{Q_{\text{min}}}^{Q_{\text{max}}} e^{-(Q R_G)^2/3} \]  

which is known as the Guinier approximation. This approximation is only valid in the Guinier regime, which is defined as \( Q R_G < 1 \). Guinier analysis was performed by plotting \( \ln[I(Q)] \) vs. \( Q^2 \), from which \( R_G \) and zero-angle scattering intensity, \( I(0) \), can be determined from the slope and intercept of the best-fit line, respectively. The analysis was performed such that the approximation was satisfied by maintaining \( Q_{\text{max}} R_G < 0.8 \). Figure 2-24 shows an example of SANS data of a dilute solution of mCherry and the Q-range over which Guinier analysis was performed (region demarcated by the vertical dashed lines) and a plot of linear regression analysis of \( \ln[I(Q)] \) vs. \( Q^2 \). The value of \( I(0) \) obtained from Guinier analysis is shown as a solid horizontal line, and the dashed horizontal lines represent the 95% confidence interval. The radius of gyration, \( R_G = 1.56 \pm 0.15 \) nm, agrees reasonably well with the dimensions of mCherry estimated from the protein crystal structure, which has a diameter and length of about 2.5 and 4.4 nm, respectively.

Figure 2-24. (a) SANS data of a dilute solution of mCherry. The dashed vertical lines represent the Q-range over which Guinier analysis was performed. The solid horizontal line and dashed lines represent the zero-angle scattering intensity, \( I(0) \), calculated from Guinier analysis and the 95% confidence interval. (b) Guinier plot and linear regression to determine the radius of gyration and \( I(0) \) from the slope and intercept, respectively.
Depolarized Light Scattering. Depolarized light scattering (DPLS)\(^8,9\) was performed on samples loaded into a 1 mm thick Teflon mold with an inner diameter of 3 mm and sealed between two quartz disks in a brass sample holder. A Coherent OBIS LX660 laser with wavelength \(\lambda = 662\) nm (chosen to ensure that this wavelength is beyond the absorption cutoff of the fluorescent proteins mCherry, EGFP, and GFP studied in this thesis) and continuous wave output power 20 mW was used. Samples were equilibrated at 10 °C for 20 minutes and then heated at 0.5 °C/min. to 40 °C, allowed to equilibrate for 5 minutes, and then cooled at 0.5 °C to 10 °C. A NESLAB recirculating water chiller was programmed in LabVIEW to control the temperature of the sample stage. The static birefringence was corrected for transmission and dark field background; necessary background detector readings and the intensity of the incident laser beam needed for data analysis were measured at least once at the beginning of each day of DPLS experiments.

1. \(I_{\text{dark}}\) is the detector voltage when the laser is turned off or the shutter is closed.

2. \(I_{\text{cross}}\) is the detector voltage when the front and rear polarizers are at perpendicular orientation to one another.

3. \(I_{\text{open}}\) is the detector voltage reading measured with the laser operating at 20 mW with the neutral density (ND) filter set to at least 2 (to prevent saturation and damage of the photodiode detector) and without any sample loaded.

DPLS measurements of protein-polymer block copolymer samples that were highly ordered and birefringent were measured with the ND filter set to 2, and samples that were poorly ordered or not birefringent were measured with the ND filter set to 1. The maximum voltage reading of the detector was on the order of 10, and in general, the ND filter was configured such that the detector voltage reading was on the order of 1. Sample transmission measurements were
performed with the ND filter set to 2. In case the birefringence intensity, $I_B$, the sample transmission intensity, $I_T$, $I_{open}$, and $I_{cross}$ were measured at different ND settings, the measurements were converted to a common ND setting before data analysis. The normalized static birefringence signal, also referred to as the power fraction, $I_{PF}$, was calculated according to the following equation:

$$I_{PF} = \frac{1}{T_s} \left( \frac{I_B - I_{dark}}{I_{open}} - \frac{I_{cross} - I_{dark}}{I_{open}} \right)$$

(2.6)

$T_s$ refers to the sample transmission, calculated as the sample transmission intensity normalized by the incident laser intensity, $I_T/I_{open}$. Transitions were reversible, and data analysis was performed on the first heating cycle.

**Turbidimetry.** For turbidimetry measurements, the same apparatus as that used for DPLS was used. However, the rear polarizer that is perpendicular to the incident laser polarization was removed to enable measurements of sample transmission. Macrophase separation transitions were defined as the temperature corresponding to a 10% reduction in the initial sample transmittance, according to previous methods. Transitions were reversible, and data analysis was performed on the first heating cycle.

**Static Light Scattering.** Static light scattering (SLS) was performed on mCherry using a Wyatt Technology DAWN Heleos II in offline-batch mode equipped with a 120 mW solid-state GaAs laser operating at 658 nm at 10 °C. While mCherry has very little absorbance at 658 nm, the low level of excitation resulted in excess scattering signal from mCherry fluorescence, yielding a weight average molar mass calculated from SLS that was approximately an order of magnitude
larger than that of mCherry that had been purified by Ni-NTA followed by anion exchange chromatography. Therefore, 658 nm bandpass filters were installed at three angle detectors—50, 81, and 134°—that that were used for light scattering measurements. SLS was performed on EGFPC14S using a Wyatt Technology DAWN 8 in offline-batch mode with a 120 mW solid-state GaAs laser operating at 658 nm at 25 °C. SLS of mCherry and EGFPC14S was performed in 10 mM Tris-Cl, pH = 7.0. At least 4 different concentrations were prepared and filtered using 0.1 μm Whatman Anotop 10 syringe filters. Protein sample concentrations were measured after being filtered and injected into the light detector using a syringe pump for measurements. The weight-average molar mass and second-virial coefficient were determined from Zimm analysis using the Astra software.

**Dynamic Light Scattering.** Dynamic light scattering (DLS) was performed on a Wyatt DynaPro Nanostar using a wavelength of 658 nm and a scattering angle of 90°. Data were acquired from ten 10-second acquisitions. Concentrations were prepared at 1 mg/mL for measurements and filtered using 0.1 μm Whatman Anotop 10 syringe filters prior to measurement. Samples were loaded into plastic Eppendorf UVette cuvettes for measurements. DYNAMICS software was used for data analysis.

**Differential Scanning Calorimetry.** Approximately 4 - 8 mg of solution samples were loaded into a hermetically sealed aluminum pan for differential scanning calorimetry (DSC). Data for mCherry-PNIPAM and EGFP-PNIPAM conjugates were acquired using a Texas Instruments Q-10 differential scanning calorimeter in Professor Swager’s lab. Data for the supercharged GFP-PNIPAM conjugates were acquired using a TA Instruments Discovery differential scanning calorimeter at the Institute for Soldier Nanotechnologies (ISN). Samples were equilibrated at 5 °C for 10 minutes,
followed by two cycles of ramping to 45 °C at 5 °C/min., holding isothermally for 2 minutes, cooling at 5 °/min. to 5 °C, and holding isothermally for 2 minutes. Temperature transition values were extracted from the peak point in the initial heating cycle using the instrument software.

**Zeta Potential.** Zeta potential measurements of mCherry, EGFPC14S, mCherryE26S, mCherryE26K, mCherryK123S, and mCherryK123E were made using a Zetasizer Nano ZS (Malvern Instruments Ltd.) at the Institute for Soldier Nanotechnologies (ISN). Protein samples were prepared at a concentration of 0.25 mg/mL in 20 mM Tris-Cl, pH = 8.0. The proteins mCherryE26S, mCherryE26K, mCherryK123S, and mCherryK123E containing a single cysteine were measured with 5 mM TCEP. The results were analyzed using the Smoluchowski model.

**Circular Dichroism.** Circular dichroism (CD) spectroscopy was performed using an Aviv model 202 CD spectrometer at the Biophysical Instrumentation Facility. Proteins are measured in 20 mM Tris-Cl, pH = 8.0, while protein-polymer conjugates are measured in water; samples were prepared at a concentration of 0.2 mg/mL and filtered using 0.1 μm Whatman Anotop 10 syringe filters prior to measurement. Measurements were acquired in a 0.1 cm path length cuvette at a scan rate of 6 nm/min.

**Fourier Transform Infrared Spectroscopy.** Silicon wafers of test grade, P type with dopant, (100) orientation, 500 μm thick, 1 - 50 Ω·cm resistivity) were used as the substrate for FTIR sample preparation. Measurements were made using a Thermo Nexus 870 spectrometer available in the Gleason Lab. For sample measurements, 64 scans at a 2 cm⁻¹ resolution were
made. Data analysis was performed using OMNIC software by measuring silicon wafer as the background spectrum, which was subtracted from the sample spectrum, followed by baseline correction. Fourier Self-Deconvolution (FSD) curves were calculated using OMNIC in the amide I region between 1600 and 1700 cm\(^{-1}\). An enhancement factor of 2.5 and a bandwidth around 30 cm\(^{-1}\) were chosen for general secondary structure identification. Each peak from the FSD spectra was assigned to a secondary structure,\(^{11-18}\) and the relative areas under each peak were used to estimate secondary structure content in the solid-state samples.

**Macrophase Separation Behavior of Protein/Polymer Blends.** The macrophase separation behavior of blends of mCherry and EGFPCC14S with PNIPAM with a molar mass of 27 kDa was studied in 10 mM Tris-Cl, pH = 7.0. Protein solutions in 10 mM Tris-Cl, pH = 7.0, were concentrated using Millipore Ultra-15 centrifugal filters with a molecular weight cutoff of 10 kDa. The concentrated protein solutions were then diluted to 250 mg/mL, and a 250 mg/mL solution of PNIPAM27k in 10 mM Tris-Cl, pH = 7.0, was also prepared. Protein and polymer solutions were mixed in a 1:1 ratio volumetrically and allowed to equilibrate at 4 °C. Blend compositions from mixing protein and polymer solutions each at 200, 150, 100, and 50 mg/mL were prepared. The cloud point was measured by UV-Vis using a temperature ramp of 0.1 °C/min. from 4 to 35 °C. The cloud point was determined as the temperature at which the sample was 50% transmissive.

**Protein Electrostatics Calculations.** The Poisson-Boltzmann equation (PBE) is a continuum model used to describe electrostatic interactions between solutes in aqueous salt solutions. The software package Adaptive Poisson-Boltzmann Solver (APBS) has been developed to solve the
Poisson-Boltzmann equation to model biomolecular solvation. There is a plugin to use APBS with PyMOL, but I was not able to get that to work successfully and instead opted to calculate protein electrostatics using the APBS web solver. Instructions for using the web version of APBS can be found at the following URL: www.poissonboltzmann.org/docs/calculating/.

Briefly, to use APBS, the protein structure must be prepared in the PQR format before APBS can proceed with continuum electrostatics calculations. The PDB2PQR web server (www.poissonboltzmann.org/docs/structures-ready/) can be used to accomplish this, and the screenshot below in Figure 2-25 shows the web version where a user can provide the protein PDB ID or a PDB file for conversion into a PQR file. A number of different force fields are available. For mCherry and EGFP electrostatics calculations, the PARSE forcefield was used, and PROPKA was used to assign protonation states at pH = 7.

After generating the PQR file, the user must select the link to solve using the APBS web solver. There are advanced options to change settings such as grid spacing, but using the default values was sufficient for this analysis. After the calculations have been completed, the web server java jmol plugin is the quickest way to visualize the results. Different views of the protein electrostatics surface potentials were exported as a Persistence of Vision Raytracer (POV-Ray) image with the extension .pov and rendered at high resolution using POV-Ray.
Calculation of Protein Solvation Energy. Following the method of Eisenberg and McLachlan,\textsuperscript{19} the solvation free energies for the globular proteins mCherry (PDB 2H5Q) and EGFP (PDB 2Y0G) were calculated using the software PyMOL.\textsuperscript{20–22} Modeller\textsuperscript{23} was used to add the missing residues and atoms to the PDB files for mCherry and EGFP, and PROPKA\textsuperscript{24–27} was used to assign charge states at pH = 7.0. The residues glutamic acid and aspartic acid with pKa values less than 5.5 were assumed to be in the conjugate base $\text{O}^-$ state, and the residues lysine, arginine, and histidine with pKa values greater than 8.5 were assumed to be prononated as $\text{N}^+$. 

Figure 2-25. PDB2PQR web server for use to prepare protein structure files in the PQR format for APBS electrostatics calculations.
Using PyMOL, the charge was assumed to be concentrated on the most solvent exposed O or N atom. In PyMOL, hydrogen atoms were added to the molecule, a solvent radius of 1.4 Å was specified to determine the solvent-accessible surface, and the dot density was set to be the maximum value of 4 for the highest achievable surface area calculations. The procedure consists of calculating the solvent-accessible atomic areas of the proteins, weighting each atomic area by its atomic solvation parameter (ASP), followed by summation.

2.3 Computer Experiments and Calculations

2.3.1 Coarse-Grained Molecular Dynamics Simulation

Molecular dynamics simulation based upon the Kremer-Grest model was used to simulate the coarse-grained conformations of protein cylinders, homopolymer coils, and mCherry-polymer cylinder-coil particles. MD simulations of single coil and single cylinder-coil particles were used to calculate their form factors to fit to SANS data and to explore the dependence of the cylinder-coil particle form factor on cylinder-coil bead interactions. The theoretical background, the computational implementation, and details of the model fitting will be discussed in the following sections. The development of the simulation code was done in C++ and adapted from Muzhou Wang’s original MD code for simulating the diffusion of rod-coil block copolymers. The work in this thesis is restricted to very low density systems with implicit solvent, and trajectories were simulated using a 64-bit Linux machine using an Intel Core i7-4700MQ processor.

Theoretical Background. A coarse-grained model based upon the Kremer-Grest model was used to model mCherry, homopolymer, and mCherry-polymer conjugates. The coarse-grained
representation of an mCherry-polymer conjugate was constructed by attaching a polymer coil of \( N \) monomers of mass \( m_{\text{coil}} \) and size \( \sigma_{\text{coil}} \) attached to a 28-bead cylinder (4 layers of 7 beads packed into a hexagonal geometry), each of mass \( m_{\text{cyl}} \) and size \( \sigma_{\text{cyl}} \). Nearest-neighbor beads are connected by finite-extensible nonlinear springs with the potential

\[
U_{s,i} = \begin{cases} 
-\frac{1}{2}kR_0^2 \ln \left( 1 - \frac{r_i^2}{R_0^2} \right), & r_i \leq R_0, \\
\infty, & r_i > R_0,
\end{cases}
\]

(2.7)

where \( l \) refers to either cylinder or coil beads, \( r \) is the separation between monomers, \( R_0 \) is a cutoff distance, and \( k \) is the interaction strength. The protein and polymers are well solvated in water under SANS measurement conditions, so cylinder-cylinder bead and coil-coil bead excluded volume interactions were modeled as solvophilic and purely repulsive using the Weeks-Chandler-Andersen (WCA) potential,

\[
U_{WCA,i} = \begin{cases} 
4\varepsilon_i \left[ \left( \frac{\sigma_i}{r} \right)^6 - \left( \frac{\sigma_i}{r} \right)^{12} + \frac{1}{4} \right], & r \leq 2^{1/6}\sigma_i, \\
0, & r > 2^{1/6}\sigma_i,
\end{cases}
\]

(2.8)

where the subscript \( k \) in the interaction term refers to the three different possible pairwise interactions—cylinder-cylinder beads, coil-coil beads, and cylinder-coil beads—and \( l \) refers to either cylinder or coil beads. WCA potentials were also applied for cylinder-coil bead interactions. The cylinder and coil bead sizes are not highly asymmetric, so the average of the two was used as an effective size for evaluating the WCA interaction potential following the Lorentz-Berthelot combining rule. 30 The sensitivity of cylinder-coil interactions was explored by changing the interaction strength \( \varepsilon_{\text{cyl-coil}} \) and also by adding attractive interactions between cylinder-coil beads. To gradually introduce a long-range attractive potential, a truncated and shifted Lennard-Jones (LJ) potential was used:
where the Lennard-Jones potential is given as

$$U_{LJ}(r) = \begin{cases} U_{LJ}(r) - U_{LJ}(r_c), & r \leq r_c \\ 0, & r > r_c \end{cases}$$

(2.9)

The parameters can be expressed as $R_0 = 1.5\sigma$, $k = 30\varepsilon/\sigma^2$, and $k_BT = \varepsilon$, where $\varepsilon$ and $\sigma_{cyl}$ are the energy and length scales of the simulation, respectively. The beads are separated by an average distance of $0.97\sigma$, as the probability distribution of $r$ for cylinder and coil beads, $P(r)$, is peaked around its mean of $f_0 P_i(r)rdr = 0.97\sigma$. Stiff three-bead bending potentials were applied along the length of the cylinder and around each hexagonal bend and are given by Eqn. 2.11,

$$U_{\theta,1} = 1000\varepsilon (1 - \cos \theta)$$

$$U_{\theta,2} = 1000\varepsilon (1 - \cos (\theta - 60^\circ))$$

(2.11)

where $\theta$ is the angle between the segments connecting adjacent monomer beads, with $\theta = 0$ corresponding to a fully extended configuration. These potentials were used in molecular dynamics simulations where the equation of motion for each monomer is given by

$$m \ddot{R} = -\nabla_j (U_{WCA} + U_S + U_\phi) - \xi \dot{R} + W$$

(2.12)

$R$ is the position of each monomer, $m$ the mass, and $t$ is time with a characteristic unit of $\tau = \sigma_{cyl}(m_{cyl}/\varepsilon)^{1/2}$. The last two terms correspond to a drag force and a Brownian force that regulates the temperature of the simulation using the Langevin thermostat. $w$ is a random Brownian force that obeys the fluctuation-dissipation relation,

$$\langle w_i(t) \rangle = 0$$

$$\langle w_i(t)w_j(t') \rangle = 2k_BT \delta_{ij} \delta(t-t')$$

(2.13)
is the identity tensor, and the drag coefficient is given by \( \xi = 0.5 \tau^{-1} \), maintaining the temperature at \( T = \varepsilon/k_B \). For form factor calculations, a single particle was simulated in a cubic box of nondimensionalized length 23 with periodic boundary conditions, and the form factor can be determined from the positions of the monomer beads that are stored at specified time steps.

**Implementation.** The coordinates of simulated particles were randomly initialized and equilibrated, and their trajectories were determined by numerically integrating Eqn. 2-12. The positions and velocities of the particles were written to an output file that can be analyzed, and for the work discussed in this thesis, only the Cartesian coordinate positions of the monomer beads were used to numerically evaluate the form factor of the particles. The positions and velocities are determined by numerically integrating the equation of motion for each bead. Writing the second-order differential equation in Eqn. 2-12 as two first-order differential equations in terms of the particle position \( R \) and velocity \( \nu \),

\[
\begin{align*}
\frac{dR}{dt} &= \nu \\
\frac{d\nu}{dt} &= \frac{\xi}{m} \nu - \frac{F + W}{m}
\end{align*}
\]

where \( F = -\nabla \left( U_{WC} + U_{s} + U_{\delta} \right) - \xi \frac{dR}{dt} \), the sum of all the non-Brownian forces. Numerical integration was performed using the leapfrog algorithm, a second-order method with a relatively simple integration. The time dimension was discretized into equal steps of \( \Delta t \), and at each time step, the first-order differential equations for the particle position and velocity were integrated.
forward, assuming that the current particle position and velocity are constant at each time step.

The procedure is outlined as follows:

1. Positions are updated to half time steps by integrating 2.13 over $\Delta t/2$,

$$
\vec{R}\left(t + \frac{\Delta t}{2}\right) = \vec{R}(t) + \frac{\Delta t}{2} \vec{V}\left(t - \frac{\Delta t}{2}\right)
$$

(2.15)

2. Velocities are updated over a full time step by solving 2.13. The solution to this first-order inhomogeneous differential equation uses the integrating factor $e^{\xi t/m}$,

$$
\frac{d}{dt}(V e^{\xi t/m}) = \frac{F + W e^{\xi t/m}}{m}
$$

$$
\vec{V}(\Delta t) = \vec{V}(0) e^{-\xi \Delta t/m} + \frac{F}{\xi} \left(1 - e^{-\xi \Delta t/m}\right) + \int_{0}^{\Delta t} \exp\left(-\frac{\xi}{m}(\Delta t - t)\right) \frac{W}{m} dt
$$

(2.16)

Denoting $s$ as a random vector whose Cartesian components are normally distributed with zero mean and unit variance, the velocities are updated over $\Delta t$ according to

$$
\vec{V}\left(t + \frac{\Delta t}{2}\right) = A \vec{V}\left(t - \frac{\Delta t}{2}\right) + B F\left(\vec{R}(t)\right) + C \vec{s}
$$

(2.17)

$$
A = e^{-\frac{\xi \Delta t}{m}}
$$

$$
B = \frac{1}{\xi} \left(1 - e^{-\frac{\xi \Delta t}{m}}\right)
$$

$$
C = \sqrt{\frac{k_B T}{m} \left(1 - e^{-\frac{\xi \Delta t}{m}}\right)}
$$

3. The positions are then updated another half time step using the updated velocities,

$$
\vec{R}\left(t + \frac{\Delta t}{2}\right) = \vec{R}(t) + \frac{\Delta t}{2} \vec{V}\left(t + \frac{\Delta t}{2}\right)
$$

(2.18)

Uniformly distributed random numbers between 0 and 1 were generated using the Mersenne twister algorithm MT19937. Normally distributed random numbers can then be generated from these uniform numbers using the Ziggurat algorithm.
Originally developed to simulate tens of thousands of monomer beads for rod-coil systems at a very high number density, computationally efficient algorithms are used in the code, and while the full extent of the time savings may not necessarily have been realized in the single particle simulations described in this thesis, these algorithms are necessary for dense systems and will be described briefly. The majority of the computational time required in updating the particle positions and velocities is the evaluation of the interactions from either WCA or Lennard-Jones potentials between all possible monomer-monomer pairs. Evaluating all the interactions straightforwardly would be quadratic in algorithmic complexity, and, while this would be acceptable for simulations of a single particle consisting of an order of $10^2$ monomer beads, it quickly becomes too computationally demanding as the number of beads grows. A more efficient algorithm can be used by taking advantage of the fact that for particles interacting through purely repulsive WCA interactions, the excluded volume interactions are nonzero only if the inter-monomer distance is less than the cutoff distance of $r_c = 2^{1/6} \sigma$. Most pairs are not close enough to experience this short-range interaction. The code implements a cell list algorithm that does not evaluate pairwise interactions that are greater than the cutoff distance $r_c$. This is accomplished by discretizing the simulation box into cubic cells of length $l_c > r_c$ and assigning every monomer to one of these cells. Monomers can only be within the cutoff distance $r_c$ if they belong to adjacent cells. The algorithm proceeds by evaluating the interactions of a monomer with other monomers in the same cell and then with all other monomers in only the neighboring 26 cells; because interactions are pairwise, only 13 neighboring cells are selected. The cell list algorithm has an algorithmic complexity of $O(N)$. The storage efficiency is optimized by using a linked list to store the monomer indices of each cell. Using a linked list also optimizes the
sorting of monomers into cells, as insertion into linked lists is $O(1)$, whereas insertion into dynamic arrays is $O(N)$.

The potentials used in this model are quite stiff and contain many singularities, and initialization and equilibration of the particles within the simulation box was important to maintain numerical stability and accuracy. The random initialization of particles may lead to monomer beads that overlap, and the MD code uses the slow push-off method developed by Auhl et al.\textsuperscript{33} While the initialization of single particles has a significantly higher probability of avoiding monomer overlap and is much simpler compared to the initialization of a dense system of tens of thousands of particles, care was still required to make sure that for a single conjugate particle with on the order of $10^2$ monomers that monomer overlap and numerical instabilities are avoided prior to MD simulation. This was accomplished by using an algorithm that gradually increases the strength of the pairwise WCA or Lennard-Jones interactions from zero to their full value by using a time-varying functional form of these forces. These potentials and forces are given by the following:

\begin{equation}
U_{ij, \text{push}} = \begin{cases} 
  U_{ij}, & r > r_e \\
  (r - r_e) \frac{\partial U_{ij}}{\partial r}(r_e) + U_{ij}(r_e), & r \leq r_e 
\end{cases}
\tag{2.19}
\end{equation}

\begin{equation}
F_{ij, \text{push}} = \begin{cases} 
  \frac{\partial U_{ij}}{\partial r}, & r > r_e \\
  \frac{\partial U_{ij}}{\partial r}(r_e), & r \leq r_e 
\end{cases}
\tag{2.20}
\end{equation}

Here, $r_e$ is a cutoff distance that varies with time. When $r_e = r_c = 2^{1/6} \sigma$, the push-off potential is 0, and the equilibrium potential is fully recovered as $r_e$ decreases to 0. The push-off forces were applied to pairs of monomers that are not adjacent to one another on the same particle; for pairs of monomers that are separated by a distance less than or equal to $r_e$, a push-off force
corresponding to a separation of \( r_e \) was applied. Push-off forces were not evaluated between pairs of beads within the same coarse-grained cylinder particle.

In this thesis, simulations of 28-bead mCherry particles were done to verify the diffusion and correlation functions of the model. Single-particle MD simulations were computed for homopolymer coils and for cylinder-coil bioconjugate particles, and details of these implementations along with the coarse-grained models used to represent mCherry, homopolymers, and mCherry-polymer conjugates are discussed next.

**mCherry**

A coarse-grained model was developed for the red fluorescent globular protein mCherry. mCherry resembles a cylinder with approximate dimensions of 4.4 nm in length and 2.5 nm in diameter based upon its crystal structure (PDB 2H5Q).\(^3\) Therefore, 28 beads were positioned as 4 layers of 7 beads packed into a hexagonal geometry to represent the cylindrical \( \beta \)-barrel shape of mCherry (Figure 2-26).

![Figure 2-26](image)

**Figure 2-26.** Coarse-grained representation of cylindrical mCherry shape. (a) Bead numbering refers to the connectivity of the beads used in the coarse-grained MD code. (b) Lines show the hexagonal arrangement of the 4 layers of beads.
The coarse-grained model for mCherry was verified and validated by performing some quick MD simulations and analyzing the mCherry diffusion, orientational correlation functions, and end-to-end vector correlation functions. The coarse-grained cylinder form factor was also fit to experimental mCherry form factor measurements by SANS. Five 28-bead mCherry particles were initialized in a cubic box of length 23 and simulated to $\tau = 2 \times 10^4$ with a time step of $\Delta t = 0.01\tau$. Discarding the initial time steps to $\tau = 100$, the trajectories were analyzed to determine the mean squared displacement (MSD) of the center-of-mass, the orientational correlation function, and the end-to-end vector correlation function. The orientational vector of mCherry was defined to be the vector pointing from bead 28 to 25 (Figure 2-11) along the cylindrical axis. The end-to-end vector was defined to be the diagonal vector pointing from bead 16 to 1 (Figure 2-11).

According to the Rouse model, the diffusion coefficient is given by $D = \frac{kT}{N\xi}$. For a 28-bead mCherry particle and using a friction coefficient of $\xi = 0.5\tau^{-1}$, $D = 1/14 = 0.0714$. Thus, the MSD of a single mCherry should have a slope of $6D = 3/7 = 0.429$, which was observed in the simulation results (Figure 2-27). As shown in Figure 2-27b and 2-23c, the correlation functions of the orientational vector and the end-to-end vector of mCherry quickly decay to 0. The correlation functions were calculated using a rolling average.
Figure 2-27. (a) Average mCherry mean squared displacement, (b) orientational correlation function, and (c) end-to-end vector correlation function of a simulation of five 28-bead mCherry particles in a cubic box with a length of 23. In the Rouse model, the diffusivity $1/14$, and the MSD agrees with the expected slope of $3/7$ (red solid line in (a)). Insets in (b) and (c) show that the orientational and end-to-end vector correlation functions quickly decay to 0.

To determine the size of each mCherry bead $\sigma_{\text{mCh}}$, the numerical form factor of the coarse-grained representation was fit to SANS data of mCherry at approximately 8 mg/mL in 10 mM Tris-Cl, pH = 8.0 at $T = 5\,^\circ\text{C}$. For a particle comprising $P$ spherical subunits, the analytical expression of the form factor is given by the Debye formula:

$$P_{\text{sys}}(Q) = \frac{1}{M^2} \sum_{i,j=1}^{P} M_i M_j F_i(Q) F_j(Q) \frac{\sin(Qr_{ij})}{Qr_{ij}}$$

(2.21)

where $F_i(Q)$ and $F_j(Q)$ represent the form factor amplitude of the $i^{th}$ and $j^{th}$ particle, respectively, and $r_{ij}$ is the distance between the centers of the $i^{th}$ and $j^{th}$ particles. The scattering lengths $M$ of all the spheres in the protein are taken to be equal and are normalized to 1. A multiplicative factor that accounts for the concentration and scattering contrast are additional fitting parameters, resulting in the fitting equation

$$I_{\text{sys}}(Q) = C T \times P_{\text{sys}}(Q) + B$$

(2.22)
Results from fitting the coarse-grained cylinder form factor to the protein SANS data using nonlinear least squares regression shows that a good fit is achieved with a value of $\sigma_{\text{cyl}} = 11.23$ (Figure 2-28).

![Graph showing experimental data and coarse-grained model fit](image)

**Figure 2-28.** Model fitting results from fitting the form factor of the coarse-grained 28-bead cylinder representation to mCherry SANS form factor data. The dashed vertical line denotes data points at low Q that deviate from 95% confidence interval bounds on the zero-angle scattering value from Guinier analysis by 10% that are omitted in the model fitting.

**Polymer**

Homopolymer coils in a good solvent are coarse-grained as beads connected by springs with excluded volume interactions. MD simulations of coils with $N = 20$ to 80 beads in which the coil bead size is $\sigma_{\text{coil}} = 1$ are computed. The time correlation of the form factors was calculated, and because the correlation time was found to be less than 1000, form factors of the coils are calculated by averaging the form factors of 1000 independent simulation configurations (one configuration is sampled every 1000$\tau$.) Single coil chains are initialized randomly in a cubic box of length 23, equilibrated using the slow push-off algorithm over a total time of $\tau = 10^4$, and then simulated to a final time of $\tau = 1.01 \times 10^6$ using a time step of $\Delta t = 0.01\tau$. The coil conformations from MD simulations as a function of $N_{\text{coil}}$ monomer beads are used to fit
experimental SANS form factors of the four polymers to determine the best-fit number of coarse-grained beads to represent each polymer. In the nonlinear least squares regression algorithm, when calculating the form factor of the single coils, the scaling parameter representing the size of the coil beads, $a_{\text{coil}}$, is calculated as a ratio relative to $a_{\text{cyl}}$ such that the coarse-grained polymer volume fractions agree with those of the experimental values.

$m\text{Cherry-Polymer}$

Coarse-grained models of mCherry-polymer conjugates are constructed by attaching coils of $N_{\text{coil}}$ monomer beads to the 28-bead mCherry model at bead number 28, the center bead at the “bottom” of the coarse-grained mCherry structure (Figure 2-26). In the bioconjugate models, in addition to a scaling parameter representing the coil bead size, $a_{\text{coil}}$, that is scaled by the cylinder bead size, to match the simulation polymer volume fractions with those from experiment, a mass $m_{\text{coil}}$ is associated with each coil bead that is scaled by the cylinder bead mass $m_{\text{cyl}}$ such that the ratio of the density of the polymer to that of the protein matches the experimental value. Single cylinder-coil conjugate molecules are initialized in a cubic box of length 23, equilibrated using the slow push-off algorithm over a total time of $\tau = 10^4$, and then simulated to a final time of $\tau = 1.01 \times 10^6$ using a time step of $\Delta t = 0.005\tau$. Because of the larger number of total beads simulated in the cylinder-coil conjugate particles and different masses accounted for the cylinder and coil monomers, a smaller time step was needed to maintain numerical stability.

**Data Visualization Using Visual Molecular Dynamics.** Molecular dynamics trajectories are visualized using the software Visual Molecular Dynamics, developed by the Theoretical and Computational Biophysics Group at the University of Illinois at Urbana-Champaign and the
Beckman Institute. More information about the software, including a link to download it, can be found at the following URL: http://www.ks.uiuc.edu/Research/vmd/. The XYZ file format is used for loading and visualizing MD trajectories in VMD. It should be noted that while the XYZ file format is a chemical file format and is used to represent the atomic positions of a molecule. It was suitable for plotting the coordinates of beads from the coarse-grained MD simulations.

In the .xyz files used to visualize only the coarse-grained mCherry cylinder beads or the coarse-grained polymer coil beads, the beads were arbitrarily assigned to correspond to the element carbon. In the .xyz files used to visualize coarse-grained mCherry-polymer MD trajectories, the cylinder beads were assigned to be the element carbon, while the coil beads were assigned to be the element hydrogen; this was arbitrarily chosen for visualization purposes, and the van der Waals sizes of the beads can be modified in the visualization, so that they do not necessarily have to correspond to those of carbon and hydrogen.

2.3.2 Model Fitting

Results from form factor SANS measurements were fit to form factor theoretical and numerical expressions using least-squares optimization. The chi-squared ($\chi^2$) functional is used as a measure for the deviation between experimental data and the model, and is given by

$$\chi^2 = \sum_{i=1}^{N} \left( \frac{I^{\text{exp}}(q_i) - I^{\text{mod}}(q_i)}{\sigma_i} \right)^2$$

(2.23)

where $I^{\text{exp}}(q_i), i = 1,\ldots,N$ represents the data points measured for the independent variable $q_i$, $I^{\text{mod}}(q_i)$ is the model intensities that depend on the parameters $a_i, i = 1,\ldots,M$, and $\sigma_i$ represents the statistical uncertainties for the data points. To account for the finite instrumental resolution,
the model intensities are smeared using the instrumental resolution function when fit to the experimental data, $R(\langle q \rangle, q)$:

$$I_{\text{model}}(\langle q \rangle) = \int R(\langle q \rangle, q) \frac{d\sigma(q)}{d\Omega} dq$$

(2.24)

where $\frac{d\sigma(q)}{d\Omega}$ is the scattering cross section. The resolution function contains contributions from the finite collimation, the spatial resolution of the detector, and the wavelength distribution. For typical SANS experiments, it has been shown that the contributions to the resolution function can approximated well with Gaussian functions. This allows the instrumental resolution function to be calculated analytically, and the smeared intensities can be calculated straightforwardly with a single numerical integral. For dilute solution form factor scattering, smearing from the instrumental resolution results in an imperceptible change to the scattering, but it was included in the model fitting procedure for completeness. Nonlinear least squares regression was performed using Matlab’s lsqnonlin.m function. 95% confidence intervals were calculated using Matlab’s nlparci.m function.

**mCherry**

The scattering function of a homogeneous cylinder with radius $R$ and length $L$ was used to model the form factor of mCherry from SANS, and the results were found to agree reasonably well with measurements from the protein crystal structure. The scattering function of a homogeneous cylinder was given by Fournet and is as follows:

$$P_{\text{cyl}}(Q) = \frac{\pi^{1/2}}{2J_1(Q R \sin \alpha) \sin \left(\frac{Q L \cos \alpha}{2}\right)} \left[ \frac{Q R \sin \alpha}{Q L \cos \alpha / 2} \right]^{1/2} \sin \alpha \, d\alpha$$

(2.25)
where \( J_1(x) \) is the first-order Bessel function. In the model fitting analysis, a contrast factor \((CT)\) parameter that accounts for the sample concentration and particle volume and a fitting parameter for the incoherent background \((B)\) are included, so the equation used to fit mCherry SANS dilute solution scattering is

\[
I_{mC}(Q) = CT \times P_{\text{sys}}(Q, R, L) + B \tag{2.26}
\]

where \( R, L, CT, \) and \( B \) are the fitting parameters.

**Homopolymers**

Dilute solutions of the homopolymers PNIPAM, PHPA, POEGA, and PEG were measured by SANS in D\(_2\)O at \( T = 5 \) °C, and at this low temperature, water is a good solvent for these four polymers. The single particle scattering of these homopolymers swollen in a good solvent is well described the scattering function of a worm-like chain (WLC) with excluded volume interactions, \( P_{\text{WLC}}(Q, L, a_K, R_{CS}) \),\(^{37,38}\) where \( L \) represents the chain contour length, \( a_K \) is the Kuhn length, and \( R_{CS} \) is the chain cross-sectional radius. The sample concentration and particle volume and incoherent background are accounted for in the SANS model fitting, so the equation used to fit homopolymer SANS dilute solution scattering is

\[
I_{\text{polymer}}(Q) = CT \times P_{\text{WLC}}(Q, L, a_K, R_{CS}) + B \tag{2.27}
\]

where \( a_K, CT, \) and \( B \) are the unknown fitting parameters. Because the contour length of the polymer is the same regardless of whether it is unconjugated or conjugated to mCherry, the contour length \( L \) was fixed in the fitting procedure. The contour length of a homopolymer is estimated by using the number average molecular weight of the polymer measured by GPC and the molecular weight of the repeat unit to determine the number of repeat units in the polymer, \( N_{\text{repeat}} \), rounded to the nearest integer, and then multiplying that value by the length of the repeat
unit, estimated from the bond lengths. The repeat unit for PNIPAM, PHPA, and POEGA includes two sp$^3$ hybridized carbon-carbon bonds. Using a value of 1.54 Å for the carbon-carbon bond length, the contour lengths of PNIPAM, PHPA, and POEGA are estimated by

$$L = \sin \left( \frac{109.5^\circ}{180^\circ} \pi \right) \times 1.54 \ \text{Å} \times 2 \times N_{\text{repeat}}$$

(2.28)

The repeat unit for PEG has two carbon-carbon single bonds and a carbon-oxygen single bond. Using a value of 1.43 Å for the carbon-oxygen single bond, the contour length for PEG is thus estimated as

$$L = \left[ \sin \left( \frac{109.5^\circ}{180^\circ} \pi \right) \times 1.54 \ \text{Å} + \sin \left( \frac{112^\circ}{180^\circ} \pi \right) \times 1.43 \ \text{Å} \times 2 \right] \times N_{\text{repeat}}$$

(2.29)

The molecular volume of the polymer also remains the same regardless of whether the polymer is conjugated or not, and the molecular volume can be calculated from the polymer density; knowing the polymer contour length and volume, the chain cross-sectional radius can be calculated from $V_{\text{polymer}} = \pi R_{cs}^2 L$.

mCherry-Polymer Conjugates

In a publication by Li et al., the scattering function of a colloid-polymer conjugate with polymer excluded volume interactions was derived from the two-point spatial correlation function. The form factor of an mCherry-polymer conjugate is given by

$$P_{\text{conjugate}}(Q) = P_{\text{colloid}}(Q) + P_{\text{polymer}}(Q) + P_{\text{colloid-polymer}}(Q)$$

(2.30)

where the colloid and polymer scattering contributions are given by
\[ P_{mcCh}(Q) = (\Delta \rho_{mcCh} V_{mcCh})^2 P_{norm, cyl}(Q, R_{mcCh}, L_{mcCh}) \]

\[ P_{polymer}(Q) = \left( \frac{b L}{a_K} \right)^2 P_{\pi, L}(Q, L, a_K, R_{CS}) \]  

(2.31)

where \( P_{norm, cyl}(Q, R_{mcCh}, L_{mcCh}) \) refers to the normalized expression of the form factor of a homogeneous cylinder given in Eqn. 2.25. \( \Delta \rho_{mcCh} \) is the scattering length difference between mCherry and the solvent D\textsubscript{2}O, and \( V_{mcCh} \) is the protein volume; \( b \) refers to the total bound coherent scattering length of a Kuhn segment.

The Debye formula (Eqn. 2.21) was also used to calculate the form factor of the coarse-grained model for mCherry-polymer conjugates and fit the form factor scattering. In the coarse-grained model, there are two different types of beads—"mCherry" beads and "polymer" beads. The scattering lengths of all beads in the protein are taken to be \( M_{mcCh} = 1 \), and the scattering length of polymer beads are allowed to be a fitting parameter. As with the model fitting equations for mCherry and the homopolymers, contrast factor (CT) and incoherent background (B) terms are included, so that there are three fitting parameters in the coarse-grained model fitting.

\[ I_{cyl-coil}(Q) = CT \times P_{cyl-coil}(Q) + B \]  

(2.32)

The form factor of a coarse-grained cylinder-coil structure of \( P \) spherical coarse-grained beads, in which the first 28 beads correspond to the 28-bead cylinder unit, is calculated by
\begin{align*}
P_{\text{cyl-coil}}(Q) &= \frac{1}{M^2} \sum_{j=1}^{r} M_j F_i(Q) F_j(Q) \sin(Qr_{ij}) Qr_0 \\
&\quad + \sum_{l=1}^{28} M_{cyl} F_{cyl}(Q, \sigma_{cyl}) \frac{\sin(Qr_{il})}{Qr_{il}} F_{cyl}(Q, \sigma_{cyl}) \frac{\sin(Qr_{ij})}{Qr_{ij}}
&+ \sum_{l=1}^{28} M_{\sigma_{cyl}} F_{cyl}(Q, \sigma_{cyl}) \frac{\sin(Qr_{il})}{Qr_{il}} F_{cyl}(Q, \sigma_{cyl}) \frac{\sin(Qr_{ij})}{Qr_{ij}}

P_{\text{cyl-coil}}(Q) &= \frac{1}{M^2} \left[ \sum_{l=-1}^{28} M_{\text{cyl}} F_{\text{cyl}}(Q, \sigma_{\text{cyl}}) \frac{\sin(Qr_{il})}{Qr_{il}} F_{\text{cyl}}(Q, \sigma_{\text{cyl}}) \frac{\sin(Qr_{ij})}{Qr_{ij}} \\
&+ \sum_{l=1}^{28} M_{\text{coil}} F_{\text{coil}}(Q, \sigma_{\text{coil}}) \frac{\sin(Qr_{il})}{Qr_{il}} F_{\text{coil}}(Q, \sigma_{\text{coil}}) \frac{\sin(Qr_{ij})}{Qr_{ij}} \right] \\
M &= \sum_{j=1}^{r} M_j
\end{align*}

\textbf{2.4 References}

Chapter 3. Phase Transitions in Concentrated Solution Self-Assembly of Globular Protein-Polymer Block Copolymers

3.1 Abstract

The phase behavior of mCherry-b-PNIPAM (mChP) block copolymers with four different PNIPAM coil fractions is investigated in concentrated aqueous solution as a function of both concentration and temperature, demonstrating both order-order transitions (OOTs) and order-disorder transitions (ODTs) in globular protein-polymer block copolymers. Independent of coil volume fraction from 0.25 to 0.70, the temperature-concentration phase diagrams share several common features. At low concentrations, mCherry-b-PNIPAM forms a homogeneous disordered phase, and macrophase separation into an ordered conjugate-rich phase and a solvent-rich phase is observed at temperatures above the PNIPAM thermoresponsive transition temperature. mChP solutions are also observed to undergo a low-temperature ODT driven by increasing concentration. The order-disorder transition concentration (ODTC) behavior of mChP is minimized for symmetric conjugates, suggesting that repulsive solvent-mediated protein-polymer interactions provide a driving force for self-assembly. Both coil fraction and solvent selectivity have large effects on the morphologies formed—disordered micelles, hexagonally packed cylinders, lamellae, and perforated lamellae are identified with the combination of small-angle X-ray scattering (SAXS), depolarized light scattering (DPLS), turbidimetry, and differential scanning calorimetry (DSC). An OOT is observed upon increasing temperature for three of the studied coil fractions at concentrations of 40-50 wt.% due to changing solvent selectivity. SANS contrast-matching experiments show that water is weakly selective for PNIPAM at low temperatures and strongly selective for mCherry at high temperatures.
3.2 Introduction

Biofunctional and biomimetic materials have received a great deal of recent attention, with a focus on understanding the design principles used in nature and incorporating these evolutionarily optimized designs into novel synthetic materials. In particular, globular proteins have been incorporated into a multitude of novel devices—bioelectronics, biocatalysis, photovoltaic devices, and fuel cells. To achieve optimally functional devices, numerous design parameters need to be considered and controlled—protein stability and lifetime within a material, protein orientation, and protein nanostructure to optimize enzyme packing density and minimize transport limitations. Protein-polymer bioconjugates formed by covalently linking proteins and synthetic polymers have been shown to improve the properties of proteins in materials, finding potential application in biotechnology, nanotechnology, drug delivery, protein separation and purification, and biosensor and switch development. Microphase separation between synthetic polymers and proteins also affords these bioconjugate materials with novel self-assembly properties that can be controlled with rational design of the molecules.

The complexity of incorporating a protein block with a specific folded geometry and anisotropic hydrophobic, ionic, or hydrogen bonding interactions is expected to significantly alter the physics of block copolymer self-assembly. Even in the comparatively simple case of rodlike polypeptides, molecular shape has a large impact on phase behavior. For example, \(\alpha\)-helix-coil bioconjugates have been demonstrated to microphase separate into lamellar phases for a wide range of coil volume fractions or hexagonally packed arrays within lamellar and cylindrical nanostructures. In addition, conjugates incorporating \(\beta\)-sheets have been shown to laterally aggregate into fibrous networks, and the primary \(\beta\)-helix motif contributes
interactions that induce one-dimensional nanostructure formation.\textsuperscript{27} Bioconjugates incorporating cyclopeptides have also been synthesized, forming tubelike structures comprising a core of stacked cyclopeptide rings with a polymer shell which laterally aggregate into weak networks.\textsuperscript{28,29}

Studies of globular protein-polymer bioconjugates have primarily focused on dilute solution self-assembly. For example, Lavigne\textsuperscript{e}ur et. al\textsuperscript{30} show that enhanced green fluorescent protein (EGFP) bioconjugates with random copolymers of ethylene glycol methyl ether acrylate (EGMEA) and methoxy ethoxy ethyl acrylate (MEEA) form spherical aggregates at high temperatures due to the thermoresponsive behavior of the acrylate copolymer. Protein-polymer conjugates have also been designed to form micelles as drug delivery vehicles.\textsuperscript{31,32} Studies of horse-radish peroxidase and myoglobin conjugated to a hydrophobic synthetic polymer tail show aqueous solution self-organization into vesicles.\textsuperscript{33} A triblock biohybrid system consisting of a globular protein and a diblock polymer has demonstrated a richer set of macromolecular structures in aqueous solution including micelles and vesicles and singular Y-junctions and toroids.\textsuperscript{34}

The self-assembly of coil-coil block copolymers in solution provides a comparison to self-assembly in more complex bioconjugate block copolymers. Coil-coil block copolymer/solvent blends exhibit similar morphologies in concentrated solutions or gels and the bulk due to similar repulsive interactions between the different blocks; however, solvent selectivity and concentration both affect the final morphology.\textsuperscript{35} Self-consistent field theory (SCFT) simulations\textsuperscript{36} and experimental order-order transition (OOT) measurements\textsuperscript{37} have shown that the phase behavior of block copolymers in a nonselective solvent follows the dilution approximation, where $\chi_{ABN}$ in melts is substituted with $\phi_{\chi_{ABN}}$ in the nonselective solvent. $\chi_{AB}$ is
the Flory-Huggins interaction parameter, \( N \) is the degree of polymerization, and \( \phi \) represents the block copolymer concentration. The dilution approximation fails as solvent concentration increases, because chain swelling and excluded volume interactions become significant; therefore, in the semidilute concentration regime, treatment of the system as a melt of blobs results in a different scaling.\(^{38}\) Experiments have shown that the order-disorder transition (ODT) follows the semidilute regime scaling.\(^{37,39,40}\) Asymmetric swelling in selective solvents leads to lyotropic and thermotropic OOTs and ODTs and has been observed experimentally in aqueous and organic solvents.\(^{41,42}\)

Recently, it has been demonstrated that globular protein-polymer block copolymers may be self-assembled into solid-state nanostructured materials similar to neat block copolymers.\(^{14}\) To preserve the folded structure of the protein, these materials must be self-assembled from solution, with solid materials produced by the evaporation of water or other solvent. The nanostructures reflect the solution self-assembled structure that formed at a concentration where the polymer lost mobility to rearrange its nanostructure. Therefore, understanding the equilibrium phase behavior of conjugates in concentrated solution is critical to describing the self-assembly of these systems. Furthermore, despite significant interest in polypeptide-polymer and protein-polymer conjugate self-assembly, the interactions that govern nanostructure formation during this type of process remain poorly understood.

In this work, the concentrated solution self-assembly of the model globular protein-polymer diblock copolymer mCherryS131C-b-PNIPAM is studied, demonstrating for the first time the ability to measure order-disorder transitions (ODTs) and order-order transitions (OOTs) in these complex block copolymers. Neutron scattering is used to probe the relationship between block hydration within the self-assembled nanostructures and ordering transitions. Finally, the
dependence of ordering on molecular design is used to shed light on relevant molecular interactions that govern self-assembly in concentrated solutions.

3.3 Experimental Methods

Synthesis

N-isopropylacrylamide (NIPAM) was polymerized by reversible addition-fragmentation chain transfer (RAFT) polymerization as previously described.\textsuperscript{14} The monomer to chain transfer agent (CTA) ratio was varied between 300:1 and 1600:1, and the reaction time was controlled to achieve the desired polymer molecular weight. The absolute molecular weight and polydispersity were characterized by gel permeation chromatography using an Agilent Technologies 1260 Infinity system, a Wyatt miniDAWN TREOS MALS light scattering detector, and a Wyatt Optilab T-rEX refractometer. N,N-dimethylformamide with 0.02 M LiBr was employed as the mobile phase. (Figure B-1).

The gene encoding for the red fluorescent protein variant mCherryS131C containing an N-terminal His tag\textsuperscript{43} was expressed in the \textit{E. coli} strain SG13009 containing the pREP4 repressor plasmid, grown in Terrific Broth at 37 °C in a 5 L fermentor for 24 hours without induction, and harvested. The cells were resuspended in lysis buffer (50 mM NaH\textsubscript{2}PO\textsubscript{4}, 300 mM sodium chloride, 10 mM imidazole, 10 mM β-mercaptoethanol (BME), pH 8.0), incubated with 1 mg/mL lysozyme at 4 °C for 30 min, and sonicated. The lysate was clarified, and the protein was purified using Ni–NTA metal affinity chromatography. Throughout the purification, 10 mM BME was used in all buffers. Elution fractions containing purified protein were dialyzed into 20 mM Tris-Cl buffer, pH = 8. The yield in the elution fractions was determined spectrophotometrically using the absorbance peak at 586 nm (extinction coefficient of 72,000 M\textsuperscript{-1}cm\textsuperscript{-1})
The purity of the protein was confirmed by denaturing gel electrophoresis (SDS-PAGE).

The maleimide-thiol coupling reaction between mCherryS131C and maleimide end-functionalized PNIPAM was performed in 20 mM Tris buffer, pH 8.0, as described previously. Following the maleimide-thiol coupling reaction, the conjugate was precipitated twice in 1 M (NH₄)₂SO₄. After the final precipitation, a clear supernatant indicated the removal of all soluble, unconjugated mCherry. The bioconjugate was then purified from the excess polymer by metal affinity chromatography using Ni-NTA; wash fractions to remove excess PNIPAM were monitored by checking for precipitation above the LCST to ensure complete removal of excess polymer. Purity was analysed using SDS-PAGE and native PAGE (Figure B-2), demonstrating the preparation of pure bioconjugates. A set of five bioconjugates with different PNIPAM coil fractions was synthesized, as detailed in Table 3-1. Polymer volume fractions were calculated as described previously.

Table 3-1. Composition of mCherry-PNIPAM Block Copolymers

<table>
<thead>
<tr>
<th>Copolymer</th>
<th>PNIPAM Mₙ (kg/mol)</th>
<th>PDIₜₙₚₐₚ</th>
<th>Bioconjugate Mₙ (kg/mol)</th>
<th>ϕₜₙₚₐₚ</th>
</tr>
</thead>
<tbody>
<tr>
<td>mChP8</td>
<td>8.2</td>
<td>1.18</td>
<td>36.3</td>
<td>0.25</td>
</tr>
<tr>
<td>mChP17</td>
<td>17.2</td>
<td>1.12</td>
<td>45.3</td>
<td>0.41</td>
</tr>
<tr>
<td>mChP27</td>
<td>27.0</td>
<td>1.09</td>
<td>55.1</td>
<td>0.52</td>
</tr>
<tr>
<td>mChP30</td>
<td>29.7</td>
<td>1.14</td>
<td>57.8</td>
<td>0.54</td>
</tr>
<tr>
<td>mChP57</td>
<td>57.1</td>
<td>1.06</td>
<td>85.2</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Sample Preparation and Characterization

Conjugate solution was concentrated to approximately 100 mg/mL using Millipore Ultra-15 centrifugal filters with a molecular weight cutoff of 10 kDa. Bulk solid samples were prepared by drop-casting aliquots of the concentrated bioconjugate solution onto Teflon sheets and drying under vacuum overnight at room temperature. The solid material was then rehydrated to the desired concentration.
Small-angle X-ray scattering (SAXS) samples were loaded into a 1 mm thick washer and sealed with Kapton tape. The bioconjugates with PNIPAM molecular weights of 17k and 30k were measured at Brookhaven NSLS X27C, while the bioconjugates with PNIPAM molecular weights of 8k and 57k were collected at the Argonne National Laboratory Advanced Photon Source at Beamline 12-ID-C,D. Samples were equilibrated at 10 °C for 20 minutes and for 10 minutes at all other temperatures prior to data collection. SAXS data were collected and corrected for empty cell and dark field scattering. Acquisition times were minimized to prevent beam damage. All observed transitions were reversible with temperature, indicating that minimal beam damage is observed.

Small-angle neutron scattering (SANS) of mChP27 was used to analyse the hydration levels of the mCherry and PNIPAM domains at 10 and 40 °C. Samples were loaded into a 1 mm thick washer and sealed between two quartz discs. Data acquisition was performed at Los Alamos National Laboratory at the Lujan Neutron Scattering Center on the Low Q Diffractometer. Samples were equilibrated at 10 °C for 20 minutes prior to data collection. Samples were then heated to 40 °C and allowed to equilibrate for 10 minutes. SANS data were collected and corrected for empty cell and dark field scattering.

Turbidimetry and depolarized light scattering\textsuperscript{46,47} were performed on samples loaded into a 1 mm thick Teflon mold and sealed between two quartz disks. A Coherent OBIS LX660 laser with wavelength $\lambda = 662$ nm (beyond the absorption cutoff of mCherry) and continuous wave output power 20 mW was used. Samples were equilibrated at 10 °C for 20 minutes and then heated at 0.5 °C/min. to 40 °C, allowed to equilibrate for 5 minutes, and then cooled at 0.5 °C/min. to 10 °C. The static birefringence was corrected for transmission and dark field background. For turbidimetry measurements, the same apparatus was used without the rear
polarizer to enable measurement of sample transmission. Macrophase separation transitions were defined as the temperatures corresponding to a 10 % reduction in the initial sample transmittance, according to previous methods.\textsuperscript{48} Transitions were reproducible upon repeated cycling, and data analysis was performed on the first heating cycle.

For differential scanning calorimetry (DSC) measurements, bioconjugate solution was loaded into a hermetically sealed aluminium pan. Data was acquired using a Texas Instruments Q-10 differential scanning calorimeter. The sample was equilibrated at 5 °C for 10 minutes, followed by two cycles of ramping to 45 °C at 5 °C/min, holding isothermally for 2 minutes, cooling at 5 °C/min to 5 °C, and holding isothermally for 2 minutes. Temperature transition values were extracted from the initial heating cycle.

Static light scattering was performed on mCherry without the S131C mutation using a Wyatt Technology DAWN Heleos II equipped with a 120 mW solid-state GaAs laser at 658 nm configured in offline-batch mode. mCherry protein without the S131C mutation was additionally purified using ion exchange chromatography to improve the accuracy of measurements. 658 nm bandpass filters were installed at three angle detectors—50, 81, and 134°—for measurements to prevent excess scattering signal from mCherry fluorescence. Five different concentrations at 1.97, 4.02, 6.10, 8.16, and 10.2 mg/mL were measured in 10 mM Tris Cl pH = 7.0. Measurements were made at 10 °C, and $dn/dc = 0.185$ mL/g for proteins was used for data analysis.\textsuperscript{49,50}

3.4 Results and Discussion

Phase Behavior in Concentrated Solution

The temperature-concentration phase diagrams of globular protein-polymer conjugates show common behavior independent of coil fraction, but the type of nanostructures formed in
solutions or gels shows a strong dependence on the polymer block volume fraction. SAXS patterns, shown in Figure 3-1, allow both order-disorder transitions (ODTs) and order-order transitions (OOTs) to be identified as a function of concentration and temperature. For a 20 wt.% solution of mChP17, a disordered phase is observed at low temperature, as evidenced by a single broad peak due to the correlation hole effect in block copolymers. Heating this solution results in a transition to a microphase separated phase between 30 and 35 °C, as indicated by narrowing of the primary peak and the appearance of a broad higher order reflection. Increasing concentration to 40 wt.% at a temperature below 20 °C causes the system to cross an order-disorder transition concentration, leading to the observance of a sharp primary peak and a weak 2q* peak corresponding to a weakly ordered lamellar phase (Figure 3-1b). Upon heating to 25 °C, the weak 2q* peak broadens, and the material undergoes an order-order transition to a hexagonally packed conformation by 30 °C. The hexagonal phase has clearly resolved q*, √3q*, and 2q* peaks. Further heating results in a small increase in domain spacing and a decrease in the sharpness of the higher order reflections, suggesting a second transition. As concentration is increased again to 50 wt.% (Figure 3-1c), an order-order transition from lamellae to a hexagonal phase is observed below 20 °C. At 50 wt.%, the sample displays q*, √3q*, and 2q* peaks throughout the observed temperature range.

DSC, turbidimetry, and DPLS (Table 3-2) provide insight into the nature of the structural transitions observed by SAXS. For concentrations ranging from 20 wt.% to 40 wt.%, DSC shows an endothermic transition between 34 and 30 °C due to desolvation of the PNIPAM block of the block copolymer. Higher transition temperatures are observed at lower block copolymer concentrations, consistent with DSC measurements of aqueous homopolymer PNIPAM solutions. This transition coincides with the ordering transition observed between
30 and 35 °C in the 20 wt.% solution of mChP17 and the OOT observed between 25 and 30 °C in the 40 wt.% solution. In the 50 wt.% solution a clear transition is no longer observed by DSC, consistent with the lack of structural transition observed in SAXS.

![Graph](image_url)

**Figure 3-1.** SAXS as a function of temperature for mChP17 at (a) 20 wt. %, (b) 40 wt. %, and (c) 50 wt. % showing both ODT and OOT behavior.

For the 20 wt.% solution of mChP17, the transition observed in DSC corresponds to a macrophase separation transition, as evidenced by a sharp drop in the transmission of the sample (Table 3-2). DPLS shows that neither the low temperature disordered phase nor the high temperature ordered phase is birefringent, suggesting that the high temperature solution is a macrophase separated mix of a water-rich phase and a block copolymer-rich disordered micellar phase, consistent with previous observations on dilute solutions and solid-state samples. Increasing concentration results in an increase in the macrophase separation transition temperature: in the 40 wt.% solution, this transition corresponds to the second transition observed in SAXS between 35 and 40 °C. At 40 wt.% DPLS indicates that none of the three phases are birefringent, suggesting that there is relatively weak ordering and small grain sizes in both the hexagonal phase and the lamellar phase. At the highest concentration of 50 wt.%,
macrophase separation is no longer observed. However, a transition from a birefringent hexagonal phase to a non-birefringent hexagonal phase is observed at 32.5 °C. While the peak positions in the SAXS pattern do not change upon heating, there is a pronounced shift in the relative peak intensities around 30 °C that suggests a change in the form factor of the polymer structure. Form factor calculations for hexagonally packed cylinders suggest shrinking of the PNIPAM cylinders due to water rejection at elevated temperatures can suppress the $\sqrt{3}q^*$ peak, relative to the $2q^*$ reflection.

The compilation of these different transitions provides the phase diagram for mChP17 (coil fraction $\phi = 0.41$) as a function of temperature and concentration, as illustrated in Figure 3-2. At temperatures below the thermal transition temperature of the PNIPAM block, increasing concentration results in a transition from a disordered phase to a lamellar phase to a hexagonally packed cylinder phase. Up to 33 wt.% conjugate the material is disordered, with the lamellar phase observed until 40 wt.% concentration and the hexagonal cylinder phase observed above 40 wt.% For concentrations of 40 wt.% and below, heating results in a transition from a homogeneous disordered phase to a macrophase separated state with an ordered block copolymer-rich phase and a water-rich phase. A PNIPAM collapse-induced OOT is observed at concentrations of 37 and 40 wt.%, where a weakly ordered lamellar phase transitions to a hexagonal phase due to an effectively decreased PNIPAM coil fraction, eventually becoming macrophase separated at high temperatures.

Changing the polymer fraction in the block copolymer has a large impact on the observed phase behavior (Figure 3-2). The bioconjugate with the smallest polymer block, mChP8 (coil fraction $\phi = 0.25$) shows a large disordered region at low temperatures and low concentration, with ordered nanostructures only observed below the PNIPAM block LCST at concentrations
greater than 37 wt.%. At these higher concentrations, the primary peak shows a significant
decrease in full width half maximum, consistent with a disorder to order transition.
Representative SAXS patterns for each phase are shown in Figure 3-3. With increasing
concentration to 40 and 50 wt.%, self-assembly is observed at low temperatures, and no
macrophase separation is observed upon heating. Unexpectedly, given the coil fraction of the
polymer, a nonbirefringent and weakly ordered lamellar phase is suggested by peaks at q* and
2q* at low temperature.

At solution concentrations of 37 wt.% and below, heating above the PNIPAM block
LCST triggers macrophase separation as confirmed by DSC and turbidimetry (Table 3-2) due to
attractive interactions between the PNIPAM chains. SAXS curves show a decrease in q* of the
primary peak as well as a transition from a broad to a sharp narrow peak with a shoulder near
1.8-1.9q*. This macrophase separated structure is not birefringent, consistent with a disordered
micellar ordering within the block copolymer- rich phase. At concentrations above 37 wt.% and
temperatures above 20-25 °C, a gradual PNIPAM collapse induces an order-order transition
(OOT) from a weakly lamellar to a hexagonally packed structure, as evidenced by peaks at q*,
\( \sqrt{3} q^* \), and 2q*. This higher temperature phase remains transmissive throughout the temperature
range studied, indicating that at high enough concentrations the system may remain macrophase
homogeneous even at elevated temperatures. Neither the low temperature nor the high
temperature phase is birefringent (Table 3-2), suggesting that both phases contain significant
disorder.

In a nearly symmetrical diblock copolymer with \( \phi_{PNIPAM} = 0.54 \) (mChP30), lamellar
nanostructures are observed at low temperature and sufficiently high concentration. For this coil
fraction the ODT concentration is 30 wt.%. Figure 3-3 shows representative scattering profiles
for both weak and strong ordering, where the degree of order improves with increasing block copolymer concentration and results in the observation of birefringent nanostructures at 40 wt.% and above. As for lower coil fractions, macrophase separation is observed at high temperatures (Table 3-2) for concentrations up to 50 wt.% At concentrations up to 33 wt.% macrophase separation corresponds with the PNIPAM desolvation transition observed by DSC, while at 37, 40, and 50 wt.% macrophase separation occurs at a temperature above this transition.

At 40 and 50 wt%, the PNIPAM desolvation transition identified by DSC corresponds closely to an order-order transition between the birefringent lamellar phase and a non-birefringent phase at 28.1 and 29.3 °C, respectively. As illustrated in Figure 3-3, the SAXS pattern for this phase exhibits an additional low-q 100 shoulder on the 001 primary peak and a set of higher order reflections that can be indexed to 110, 200, and 002. This type of scattering pattern has been previously observed in neat mCherry-PNIPAM diblock copolymers due to the formation of perforated lamellar domains. Therefore, at 30 °C and above, it is hypothesized that the thermal transition behavior of PNIPAM causes collapse of the PNIPAM-rich lamellar nanodomains, resulting in a structural transition from lamellae to perforated lamellae.

For the largest PNIPAM block studied (φPNIPAM = 0.70, mChP57) increasing concentration at low temperature also results in a transition from a disordered to a lamellar structure. Weak ordering is first observed at 35 wt.%, and the quality of ordering increases with increasing concentration as evidenced by birefringence and the appearance of a visible peak at 3q* for the 50 wt.% sample. Turbidimetry measurements show that the solutions macrophase separate upon heating at all concentrations studied, similar to mChP30. However, at concentrations from 25 to 35 wt.%, a homogeneous disordered micellar phase is observed in
between the low temperature ordered or disordered phase and the high temperature macrophase separated region. In this intermediate region, the SAXS scattering pattern (Figure 3-3) shows a decrease in intensity and increase in full width half maximum of the primary peak and broad higher reflections between 1.5 and 1.8$q^*$.

Despite undergoing macrophase separation above 36 $^\circ$C, the materials remain in a birefringent lamellar phase at 50 wt.% (Figure B-3).

---

**Figure 3-2,** Phase diagrams of (a) mChP8 (b) mChP17 (c) mChP30 (d) mChP57 as a function of temperature and concentration. The various phases are denoted as disordered (Dis), disordered micellar (DM), nonbirefringent lamellar (NB Lam), lamellar (Lam), nonbirefringent hexagonal (Hex), hexagonally packed cylinders (HPC), and perforated lamellar (PL). Open symbols represent regions where macrophase separation between a conjugate rich ordered phase and a water-rich phase is observed.
Table 3-2. Thermal transitions for mChP systems as measured by DPLS, turbidimetry, and DSC.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Conc. (wt.%)</th>
<th>TDPLS(^a) (°C)</th>
<th>Tt (°C)</th>
<th>TDSC (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mChP8</td>
<td>20</td>
<td>○ 29.3</td>
<td>33.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>○ 29.0</td>
<td>31.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>○ 29.1</td>
<td>31.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>○ 39.3</td>
<td>30.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>○</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>○</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>mChP17</td>
<td>20</td>
<td>○ 28.2</td>
<td>33.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>○ 28.8</td>
<td>31.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>○ 28.4</td>
<td>32.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>○ 30.9</td>
<td>30.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>○ 36.2</td>
<td>30.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>32.5</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>mChP30</td>
<td>20</td>
<td>○ 31.2</td>
<td>33.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>○ 31.0</td>
<td>30.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>○ 31.3</td>
<td>31.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>○ 34.4</td>
<td>31.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>28.1</td>
<td>36.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>29.3</td>
<td>39.6</td>
<td></td>
</tr>
<tr>
<td>mChP57</td>
<td>20</td>
<td>○ 32.8</td>
<td>32.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>○ 32.2</td>
<td>32.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>○ 35.0</td>
<td>30.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>○ 35.7</td>
<td>30.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>○ 35.7</td>
<td>28.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>● 36.3</td>
<td>28.4</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) A temperature listed for a transition in DPLS means that the sample initially begins birefringent at 10 °C and loses birefringence at the listed transition temperature. ○ denotes samples that never display birefringent behavior at all temperatures, ● denotes samples that remain birefringent at all temperatures, and – denotes samples in which thermal transitions are not observed.

Solvent Distribution

The distribution of water within the nanostructures is critical to determining phase transitions and the type of nanostructure formed. To determine the hydration levels within mCherry and PNIPAM domains as a function of temperature during self-assembly, a nearly symmetric mChP with a polymer block molar mass of 27 kDa was analysed by small-angle neutron scattering. A bioconjugate nearly symmetric in composition (\(\phi_{\text{PNIPAM}} = 0.52\), mChP27) was chosen because it provides a balanced matrix for studying water segregation. 30 wt.% solutions were prepared in four different H\(_2\)O/D\(_2\)O blend compositions—90/10, 80/20, 70/30, and 60/40—to provide variable contrast with both PNIPAM and mCherry blocks (Figure B-4).
Using calculated scattering length densities for each molecular component (Table B-1), the scattering contrast was modeled as a function of water partitioning in the block copolymer. The scattering intensity $I \sim \Delta SLD^2 = C\Delta SLD^2$, where $C$ is a constant and $\Delta SLD = f(\epsilon, H_2O/D_2O)$, was fit to the experimental peak intensities at the four $H_2O/D_2O$ compositions to determine $C$ and $\epsilon$, where $\epsilon$ is the volume fraction of water hydrating the PNIPAM block.

**Figure 3-3.** Representative SAXS patterns of (a) mChP8 (b) mChP17 (c) mChP30 (d) mChP57. The various phases are denoted as disordered (Dis), disordered micellar (DM), nonbirefringent lamellar (NB Lam), lamellar (Lam), hexagonal (Hex), hexagonally packed cylinders (HPC), and perforated lamellar (PL). The concentration and temperature conditions corresponding to each representative SAXS pattern can be found in Table B-1.

Theoretically, the pure PNIPAM and mCherryS131C components are contrast matched by an $H_2O/D_2O$ blend volume fraction composition of 81/19 and 68/32, respectively. Experimentally, at $T = 10 \degree C$, the peak scattering intensity decreases with increasing $D_2O$ composition (Figure 3-4), and fitting the scattered intensity of the primary peak for the four blend compositions indicates that under these conditions, contrast matching is achieved at a blend composition of 14% volume $H_2O$ and 86% volume $D_2O$. In order to achieve contrast matching at this composition, 59% of the water is located within the PNIPAM nanodomains, while 41% of the water is associated with the mCherry nanodomains. Therefore, at low
temperatures, the PNIPAM coils are hydrated to a greater extent than mCherry, and the water is a slightly polymer-selective solvent.

At 40 °C, where thermal collapse of the PNIPAM is expected to lead to selective hydration of the protein nanodomains, fitting the scattering intensity leads to a contrast matching condition of 85.5/14.5 H2O/D2O blend as seen in Figure 3-4b. This is nearly identical to the contrast matching condition for neat PNIPAM, corresponding to only 12% volume of the water hydrating the PNIPAM block and 88% of the water residing within the mCherry nanodomains. This result clearly confirms the strong protein-selective nature of the solvent above the thermal transition temperature of mChP.

Changes in hydration with increasing temperature are responsible for both macrophase separation and order-order transitions. In mChP8 and mChP17 at low temperatures, lamellar morphologies are stable due to the preferential hydration and swelling of PNIPAM. With increasing temperature, however, PNIPAM chain collapse effectively decreases the PNIPAM coil fraction, favouring hexagonally packed structures with greater interfacial curvature. A similar change in PNIPAM hydration is also responsible for the lamellar to perforated lamellar transition in mChP30 and the lamellar to disordered micellar transition in mChP57.

Water partitioning also accounts for the observation of lyotropic transitions from weakly ordered lamellar phases to hexagonal phases in mChP17. In lower concentration solutions, the water swells the PNIPAM domains, increasing the coil fraction and pushing the conjugates towards the lamellar phase. However, as the concentration of block copolymer is increased, the effective polymer fraction decreases, resulting in a transition to a hexagonal phase. Although water is slightly selective for the PNIPAM block at low temperature, both phases remain highly
hydrated such that selective hydration is unlikely to be a large factor in the low temperature order-disorder transition concentration.

![Figure 3-4](image.png)

**Figure 3-4.** Small-angle neutron scattering intensities (data points) and curve fits (solid line) for mChP27 in different H₂O/D₂O blend compositions at (a) T = 10 °C and (b) T = 40 °C.

Knowledge of solvent selectivity allows comparison between the behavior of mCherry-b-PNIPAM and the well-studied polystyrene-polyisoprene (PS-PI) diblock copolymers in dialkylphthalate solvents. mCherry-b-PNIPAM bioconjugates at 20 °C and below are observed to have lower ODTCs than those observed for the PI-PS systems studied in a neutral dioctyl phthalate (DOP) solvent but greater than those observed in the more selective solvents dibutyl phthalate (DBP) and diethyl phthalate (DEP). This is consistent with the weakly selective nature of water as a solvent for mChP block copolymers at low temperature. The type of nanostructures formed at a given coil fraction are different between the globular protein-coil and coil-coil systems due to the difference in chain shape and minor differences in solvent selectivity.

Studies of polystyrene-polyisoprene block copolymers (PS-PI) with polystyrene coil fraction $\phi_S = 0.23$, in the neutral solvent dioctyl phthalate (DOP), show disordered and cylindrical morphologies. In contrast, mCherry-PNIPAM bioconjugate with $\phi_{PNIPAM} = 0.25$
shows weakly ordered lamellar structures by 40 wt.%. The phase diagram for PS-PI with $\phi_S = 0.40$ shows a disordered phase transitioning to lamellae near a polymer volume fraction of 0.50. For the bioconjugate with $\phi_{PNIPAM} = 0.41$, a weakly ordered lamellar phase is first seen at 37 wt.%, while a cylindrical morphology is observed at 50 wt.%. A symmetric PS-PI diblock copolymer shows similar behavior to that for $\phi = 0.40$ and has strong similarity to the mCherry-PNIPAM bioconjugate with $\phi_{PNIPAM} = 0.54$, where a disorder to lamellar ordering transition occurs by 30 wt.%.

Lastly, the phase behavior of PS-PI with $\phi_S = 0.70$ shows only disordered and cylindrical phases. Unexpectedly, mCherry-PNIPAM near this PNIPAM coil fraction is observed to form weakly ordered lamellar nanostructures at 35 wt.%, and lamellar phases at 40 and 50 wt.%. The preference for lamellar domains at high coil fractions is consistent with self-assembly in rod-coil block copolymers, where the presence of a rigid block such as the globular protein may introduce an energy penalty for curved interfaces required to form cylindrical or spherical nanodomains.

The phase behavior of PS-PI in a series of increasingly polystyrene-selective solvents—di-$n$-butyl phthalate (DBP), diethyl phthalate (DEP), and dimethyl phthalate (DMP)—show thermotropic OOTs similar to those observed in the mChP block copolymer system. Similar to coil-coil diblock copolymers in selective solvents, the phase boundaries between ordered phases for mChP reflect the temperature dependence of the solvent selectivity: the OOTs correspond to a reduction in PNIPAM coil fraction with increasing solvent selectivity. Concentrated mChP solutions with increasing temperature and thus increasing protein-selectivity are observed to pack into hexagonal morphologies, but the polymers do not order into FCC or BCC spherical domains as has been observed for PS-PI solutions. Both experimental and SCFT results have
demonstrated lamellae and cylinder coexistence regions for coil-coil diblock copolymers in selective solvents,\textsuperscript{52,55} whereas mChP30 appears to adopt a perforated lamellar morphology in solution near the thermal transition temperature.

**ODTC and Interactions**

At low concentrations, all coil fractions of conjugate are well-solvated in a homogeneous, disordered state; beyond a minimum density of molecules an order-disorder transition concentration (ODTC) is observed. Because self-assembly is observed at low temperatures where both protein and polymer domains are highly hydrated, solvent-mediated interactions must drive microphase separation in this regime. In a binary system such as protein-polymer diblocks, only solvent mediated protein-protein, polymer-polymer, and polymer-protein interactions can influence self-assembly.

In the hybrid protein-polymer system, quantitative measures of solvent-mediated protein-protein and polymer-polymer interactions in the dilute solution limit are accessible through measurement of the second virial coefficient, $A_2$, using static light scattering. In 10 mM Tris Cl pH = 7.0 at $T = 25 \degree C$, the second virial coefficient of mCherry (without thiol mutation) was found to be $1.1e-4 \text{ mol mL}^{-2}$, illustrating that the proteins have slightly repulsive self-interactions (Figure B-5). For polymers, the repulsive interaction regime for $A_2 > 0$ corresponds to a Flory-Huggins parameter $\chi < 0.5$ where the polymers are soluble. Second virial coefficients of PNIPAM reported in the literature range from $0.79e-4$ to $1.5e-4 \text{ mL mol}^{-2}$ at $20 \degree C$ in water,\textsuperscript{56} with only a weak molecular weight dependence. This demonstrates repulsive interactions between PNIPAM molecules in dilute solution. These interactions in the dilute solution limit suggest that both protein-protein and polymer-polymer repulsive interactions will not act to promote microphase separation.
Trends in the ODTC as a function of coil fraction provide a basis for determining the relative role of protein-protein, protein-polymer, and polymer-polymer interactions in governing the self-assembly behavior in concentrated systems. The ODTC (Figure 3-5) is defined as the concentration at which ordering is first observed at temperatures less than 20 °C, where both protein and polymer blocks are known to be well solvated. The ODTC has a minimum for the symmetric diblock copolymer, with an increasing ODTC as the coil fraction deviates from 0.50. In the concentrated regime, if protein-protein interactions were most important, one would expect that the ODTC would be lowest at low coil fraction due to the fact that protein concentration increases faster with increasing block copolymer concentration for a small coil fraction. However, the ODTC is minimized when the concentration of PNIPAM and mCherry is balanced. Therefore, it is hypothesized that repulsive mCherry-PNIPAM interactions account for the largest contribution to the net protein-polymer repulsion leading to self-assembly.

![Figure 3-5. Order-disorder transition concentration as a function of PNIPAM coil fraction, showing that a minimum is reached near the symmetric composition diblock copolymer.](image)

Macrophase Separation

Although macrophase separation observed for all four coil fractions is driven by the thermal collapse of PNIPAM blocks, polymer molecular weight and conjugate concentration also
have a large impact on this transition. For all coil fractions studied, the PNIPAM thermal transition temperature decreases with increasing solution concentration, as measured by DSC (Figure 3-6). The PNIPAM homopolymer thermal transition also decreases with increasing solution concentration and PNIPAM molecular weight,$^{57,58}$ consistent with DSC results for the conjugates. In contrast, the macrophase separation transition temperatures increase with increasing solution concentration. Therefore, PNIPAM collapse only corresponds directly to macrophase separation at low concentrations. Instead, decreasing PNIPAM solubility results in microphase separated structures that will then form either a homogeneous phase or a macrophase separated structure depending upon the ability of the mCherry nanodomains to accommodate all of the water rejected by the PNIPAM domains. Increasing conjugate concentration enables more water to be accommodated, resulting in macrophase separation at a higher degree of PNIPAM collapse and an increase in the macrophase separation transition temperature above the PNIPAM thermal transition temperature.

An increase in PNIPAM coil fraction reduces the mCherry to water ratio, resulting in a higher macrophase separation transition concentration. For mChP8 at 40 and 50 wt.% and mChP17 at 50 wt.%, the mCherry to water ratio is sufficiently high such that the system remains homogeneous. In mChP30 and mChP57, however, macrophase separation is observed even at 40 and 50 wt.% due to the inability of the mCherry domains to accommodate the water from PNIPAM desolvation.

3.5 Conclusions

The self-assembly of mCherry-b-PNIPAM (mChP) diblock copolymers in concentrated aqueous solutions was explored, demonstrating for the first time OOTs and ODTs in globular
protein-containing diblock copolymers. The resulting temperature-concentration phase diagrams for mCherry-PNIPAM bioconjugates with four coil fractions from 0.25 to 0.70 show several common features independent of coil fraction: at low concentrations (20-30 wt.%), the block copolymers form homogeneous, disordered solutions at low temperature with macrophase separation between an ordered, conjugate-rich phase and a water-rich phase occurring at elevated temperatures due to PNIPAM chain collapse. As concentration is increased, conjugates undergo a low temperature ODT, and the samples become homogeneous at elevated temperature. SANS contrast matching experiments show that water is a slightly selective solvent for the PNIPAM block at low temperatures, but it becomes strongly selective for the mCherry block above the thermal desolvation transition of the PNIPAM block.

The coil fraction of the conjugates has a large effect on the type of ordered phases formed, the position of the ODT concentration, and the concentration at which macrophase separation ends. Lamellar and hexagonal phases ($\phi_{\text{PNIPAM}} = 0.25$), hexagonally packed cylinders and hexagonal phases ($\phi_{\text{PNIPAM}} = 0.41$), lamellae and hexagonally perforated lamellae ($\phi_{\text{PNIPAM}} = 0.69$).
and lamellae (\(\phi_{\text{PNIPAM}} = 0.70\)) were observed with increasing coil fraction, with OOTs between high and low temperature ordered structures observed at 40-50 wt.% for the three lower coil fractions. The ODTC reaches a minimum for the symmetric conjugate, suggesting that protein-polymer repulsive interactions are important for governing self-assembly. Increasing block copolymer concentration increases the mCherry to water ratio, leading to higher macrophase separation temperatures. Increasing coil fraction reduces the mCherry to water ratio; for low PNIPAM coil fraction, solutions are able to remain homogeneous at high concentration, whereas macrophase separation is observed up to 50 wt.% with increasing PNIPAM coil fraction.

### 3.6 References

Chapter 4. The Nature of Protein Interactions Governing Globular Protein-Polymer Block Copolymer Self-Assembly

4.1 Abstract

The effects of protein surface potential on the self-assembly of protein-polymer block copolymers are investigated in globular proteins with controlled shape through two approaches: comparison of self-assembly of mCherry-poly(N-isopropylacrylamide) (PNIPAM) bioconjugates with structurally homologous enhanced green fluorescent protein (EGFP)-PNIPAM bioconjugates, and mutants of mCherry with altered electrostatic patchiness. Despite large changes in amino acid sequence, the temperature-concentration phase diagrams of EGFP-PNIPAM and mCherry-PNIPAM conjugates have similar phase transition concentrations. Both materials form identical phases at two different coil fractions below the PNIPAM thermal transition temperature and in the bulk. However, at temperatures above the thermoresponsive transition, mCherry conjugates form hexagonal phases at high concentrations while EGFP conjugates form a disordered micellar phase. At lower concentration, mCherry shows a two-phase region while EGFP forms homogeneous disordered micellar structures, reflecting the effect of changes in micellar stability. Conjugates of four mCherry variants with changes to their electrostatic surface patchiness also showed minimal change in phase behavior, suggesting that surface patchiness has only a small effect on the self-assembly process. Measurements of protein/polymer miscibility, second virial coefficients, and zeta potential show that these coarse-grained interactions are similar between mCherry and EGFP, indicating that coarse-grained interactions largely capture the relevant physics for soluble, monomeric globular protein-polymer conjugate self-assembly.
4.2 Introduction

Globular proteins and enzymes have great potential for incorporation into novel biofunctional nanomaterials to address challenges in medicine, catalysis, defense, and energy. For example, the substrate specificity and high catalytic activity of proteins render them attractive for use in glucose sensors\textsuperscript{1,2} and in neurotoxin sensing and detoxification.\textsuperscript{3-5} Globular proteins have also garnered interest for bioelectronics and other energy-related applications including photovoltaic devices\textsuperscript{6} and biofuel cells.\textsuperscript{7-9} For many of these applications, an optimized protein-based material must have mechanical integrity, maintain and improve protein stability and longevity, control three-dimensional packing and orientation within a material to achieve a high areal density of active sites, and provide structured channels for transport of reactants/products.\textsuperscript{8} Self-assembly is being actively investigated as a method to control the nanostructure of protein-based materials, potentially enabling all of these design criteria to be achieved. One route to protein self-assembly is templating within a nanostructure formed by materials such as block copolymers or lipids. Selective segregation into a single domain of the nanostructure may be driven by bioconjugation (i.e. PEGylation),\textsuperscript{10} ionic interactions,\textsuperscript{11} or hydrophobic interactions with specific regions of membrane proteins.\textsuperscript{12} In addition to these three-dimensional templating methods, selective adsorption onto a nanopatterned block copolymer surface has been shown to lead to two-dimensional protein arrays.\textsuperscript{13,14}

An alternative to templated self-assembly is the incorporation of a protein of interest into an amphiphilic or block copolymer-like molecule, providing a route to direct assembly of the biofunctional component. In contrast to traditional block copolymers composed of Gaussian coil polymers,\textsuperscript{15,16} the self-assembly of fully folded and functional protein blocks in block copolymers has the potential to introduce significant complexity due to the highly specific
protein shape\cite{17, 18} and anisotropic hydrophobic, ionic, and hydrogen bonding interactions between different protein molecules. The protein surface heterogeneity or patchiness and charge fluctuation due to the acidic and basic groups have been shown to reduce the net repulsive electrostatic interactions between proteins in solution.\cite{19} Charge fluctuation may induce attractive interactions between proteins, which is now understood to be related to an intrinsic molecular property of a protein known as the charge capacitance.\cite{20} Recent experimental studies,\textsuperscript{21-23} theory, and simulations\textsuperscript{24, 25} of mixtures of charged colloids and neutral nonadsorbing polymers have also demonstrated that electrostatic interactions can stabilize the mixture against depletion-induced phase separation. In addition to electrostatic effects, the coarse-grained shape of the globular protein may lead to different interactions and phase behavior. Bioconjugates incorporating structurally simple peptide motifs—\(\alpha\)-helix,\textsuperscript{26, 27} \(\beta\)-sheets,\textsuperscript{28} \(\beta\)-helix,\textsuperscript{29} and cyclopeptide rings\textsuperscript{30, 31}—have been observed to display peptide structure-dependent phase behavior. For example, even relatively simple \(\alpha\)-helix-coil bioconjugates have a phase diagram markedly different from that of traditional coil-coil block copolymers.\textsuperscript{32-34}

Due to their large technological interest, significant effort has also been directed towards understanding the self-assembly of globular protein-polymer conjugate block copolymers. Most work has investigated dilute solution self-assembly into spherical aggregates,\textsuperscript{35} micelles for drug delivery vehicles,\textsuperscript{36, 37} and vesicles.\textsuperscript{38} Triblock biohybrid systems consisting of a globular protein conjugated to a diblock copolymer have demonstrated a richer set of macromolecular structures in aqueous solution including \(Y\)-junctions and toroids.\textsuperscript{39} Triblock bioconjugates with both linear and \(Y\)-shaped molecular architectures have also been explored and shown to form stable thermoresponsive spherical micelles.\textsuperscript{40, 41} Recently, our group has demonstrated the ability to apply these same self-assembly principles to form solid-state nanostructured plastics\textsuperscript{42, 43} and
gels\textsuperscript{44} from globular protein-polymer block copolymers. Using the model globular protein-polymer mCherry-\textit{b}-poly(\textit{N}-isopropylacrylamide), several nanostructured phases have been observed, and phase diagrams were developed that show lyotropic and thermotropic order-disorder and order-order transitions (ODTs and OOTs) in solution.

A central question in the development of self-assembly as a technique to control globular protein and enzyme nanostructure is whether studies of a specific protein can provide predictive information or design rules for materials built from a different protein. In particular, are there universal rules for self-assembly analogous to those for traditional block copolymers that can be used to guide material design, or are such rules precluded by the functional diversity of proteins? Although the complexity of protein sequence and structure lead to a wide variety of shapes and specific interactions, relatively simple theories of colloidal interactions have been able to capture a great deal of the solution thermodynamics in globular protein solutions,\textsuperscript{45-47} suggesting that there may also exist universal relationships that govern the thermodynamics of globular protein-containing block copolymers. A critical step in developing an understanding of such structure-property relationships is to identify the relevant level of coarse-graining at which the system can be considered.

In this study, the role of protein surface potential and patchiness in the self-assembly process is investigated through direct comparisons of the self-assembly of conjugates containing structurally similar proteins with different surface amino acids using two approaches (Scheme 4-1). First, the structurally homologous but phylogenetically distant fluorescent proteins mCherry, an optimized variant of the mRFP1 monomer derived from the tetrameric Discosoma sp. fluorescent protein “DsRed”,\textsuperscript{48} and enhanced green fluorescent protein (EGFP), a variant of the green fluorescent protein from the jellyfish Aequorea victoria,\textsuperscript{49, 50} are used to show that
chemically distant protein sequences can undergo the same self-assembly when the protein fold and coarse-grained properties—protein shape, size, surface charge, surface hydrophobicity, and second virial coefficient—are similar. Second, mCherry variants with the point mutations K123E, K123S, E26S, and E26K are designed to study the effects of electrostatic patchy interactions on the surface of mCherry-PNIPAM conjugates. The similarity in phase behavior between these mutants indicates the patchiness of the protein surface plays a minimal role in the self-assembly process. Studies of phase behavior in concentrated solution and the solid state specifically highlight similarities and differences between the materials that suggest that coarse-grained surface interactions and biophysical properties are sufficient for understanding different aspects of the self-assembly behavior of globular protein-polymer block copolymers.
**Scheme 4-1.** a) Cartoon structures of mCherry (PDB file 2H5Q) and EGFP (PDB file 2Y0G) showing similar β-barrel structure. Different views of the electrostatic surface potential (±5 kT/e) at the solvent-accessible surface of EGFP and mCherry rendered from solutions of the linearized Poisson-Boltzmann equation using the Adaptive Poisson-Boltzmann Solver (APBS). Red represents negative values, and blue represents positive values. b) Protein sequence alignments of mCherry and EGFP showing the low identity (29%) in chemical composition (highlighted in gray). Sequence alignment results are generated from William Pearson’s program implementing the algorithm of Huang and Miller. c) Electrostatic surface potentials (±5 kT/e) at the solvent-accessible surface of the different mCherry mutants showing the perturbation in protein surface patchiness due to the point mutations E26K, E26S, K123E, and K123S. d) mCherry protein sequence for preparation of mCherry variants in this study. Mutation locations are highlighted in blue at the mCherry protein sequence.
4.3 Materials and Methods

Synthesis

Poly(N-isopropylacrylamide) (PNIPAM) was synthesized by reversible addition-fragmentation chain transfer (RAFT) polymerization as previously described. The proteins mCherry, all mCherry variants, and enhanced green fluorescent protein (EGFP), each containing an N-terminal 6xHis tag, were expressed in the E. coli strain SG13009 containing the pREP4 repressor plasmid. The protein mCherry referred to in this paper contains a single-site mutation to the original sequence as reported by Tsien, where residue 131 is mutated from serine to cysteine to yield a unique reactive thiol, as reported previously. The coupling reactions between EGFP or all mCherry and maleimide end-functionalized PNIPAM were performed in 20 mM Tris buffer, pH 8.0, as described previously. Details of syntheses and DNA cloning are included in the Supplementary Information.

Two pairs of bioconjugates with different PNIPAM coil fractions (Mn = 12.8 kDa, ϕPNIPAM = 0.37, and Mn = 26.3 kDa, ϕPNIPAM = 0.55) (Supporting Information) were synthesized using mCherry and EGFP, as detailed in Table 4-1, and five bioconjugates of mCherry mutants with the same PNIPAM coil fraction (Mn = 26.0 kDa, ϕPNIPAM = 0.54) were prepared. Abbreviations used to refer to mCherry and EGFP conjugates UV-Vis spectrophotometry, FTIR, and CD spectroscopy of the EGFP conjugates show retention of protein folding in the self-assembled state (Figure C-10, Figure C-11, Figure C-12, and Table C-2). Protein folding and activity retention for mCherry conjugates have been previously demonstrated and reported.

Sample Preparation and Characterization

Concentrated solution samples were prepared by rehydrating the solid globular protein-polymer block copolymer material to the desired concentration. Solid-state samples were cast
from concentrated aqueous solutions under processing conditions that have previously been shown to maximize ordering and the preservation of protein function in mCherry materials.\textsuperscript{42,43} The concentrated solution phase behavior of mCherry- and EGFP-PNIPAM conjugates is characterized using a combination of small-angle X-ray scattering (SAXS), depolarized light scattering (DPLS) and turbidimetry, and differential scanning calorimetry (DSC). Solid-state self-assembly observed by SAXS is confirmed by transmission electron microscopy (TEM). Static light scattering (SLS) and protein/polymer macrophase blend studies were used to quantify protein-protein and protein-polymer demixing behavior. The alteration of the surface potentials was measured using zeta potential, and the retention of secondary structures of the mCherry mutants was characterized by CD spectroscopy. CD spectroscopy was performed using an Aviv model 202 CD spectrometer to measure far UV circular dichroism spectroscopy of mCherry and mutants in 20 mM Tris-Cl, pH = 8.0, at concentrations of 0.2 mg/mL in a 0.1 cm path length cuvette at a scan rate of 12 nm/min.; EGFP was measured in 20 mM Tris-Cl, pH = 8.0, and EGFP-PNIPAM conjugates were measured in water. Details of the aforementioned characterization and sample preparation techniques are included in the Supplementary Information.

4.4 Results and Discussion

Self-Assembly of Structurally Homologous Proteins mCherry and EGFP

Comparison of EGFP and mCherry block copolymer self-assembly in concentrated solution shows that the general shape of the phase diagram is common to both globular protein-polymer conjugates. However, small shifts in transition temperatures and changes in the type of structure formed at high temperature are specific details of the phase behavior that depend upon the protein’s surface potential. Both similarities and differences are illustrated in phase diagrams.
of EGFPP12.8 and mChP12.8 ($\phi_{\text{PNIPAM}} = 0.37$) and EGFPP26.3 and mChP26.3 ($\phi_{\text{PNIPAM}} = 0.55$) (Figure 4-1) in which the proteins mCherry and EGFP are conjugated to the same two polymers to enable a direct comparison. The characterization details of the block copolymers used in this study may be found in Table 4-1. Phase identifications are performed using a combination of SAXS to identify nanodomain symmetry (Figure 4-2), DSC to measure the PNIPAM thermal transition, DPLS to measure the optical anisotropy of nanostructures, and turbidimetry to identify macrophase separation (Table 4-2).

Table 4-1. Composition of mCherry/EGFP-PNIPAM Block Copolymers

<table>
<thead>
<tr>
<th>Compound</th>
<th>PNIPAM $M_n$ [kg/mol]</th>
<th>PDI$_{\text{PNIPAM}}$</th>
<th>Bioconjugate $M_n$ [kg/mol]</th>
<th>$\phi_{\text{PNIPAM}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFPP12.8</td>
<td>12.8</td>
<td>1.11</td>
<td>42.3</td>
<td>0.37</td>
</tr>
<tr>
<td>mChP12.8</td>
<td>12.8</td>
<td>1.11</td>
<td>40.9</td>
<td>0.37</td>
</tr>
<tr>
<td>EGFPP26.3</td>
<td>26.3</td>
<td>1.10</td>
<td>55.8</td>
<td>0.55</td>
</tr>
<tr>
<td>mChP26.3</td>
<td>26.3</td>
<td>1.10</td>
<td>54.4</td>
<td>0.55</td>
</tr>
<tr>
<td>mChP26.0</td>
<td>26.0</td>
<td>1.05</td>
<td>54.2</td>
<td>0.54</td>
</tr>
<tr>
<td>mChK123EP26.0</td>
<td>26.0</td>
<td>1.05</td>
<td>54.2</td>
<td>0.54</td>
</tr>
<tr>
<td>mChK123SP26.0</td>
<td>26.0</td>
<td>1.05</td>
<td>54.2</td>
<td>0.54</td>
</tr>
<tr>
<td>mChE26SP26.0</td>
<td>26.0</td>
<td>1.05</td>
<td>54.2</td>
<td>0.54</td>
</tr>
<tr>
<td>mChE26KP26.0</td>
<td>26.0</td>
<td>1.05</td>
<td>54.2</td>
<td>0.54</td>
</tr>
</tbody>
</table>
Figure 4-1. Phase diagrams of (a) EGFP12.8 (b) mChP12.8 (c) EGFP26.3 (d) mChP26.3 as a function of temperature and concentration. Phases are identified as disordered (Dis), disordered micellar (DM), nonbirefringent lamellar (NB Lam), lamellar (Lam), and hexagonal (Hex). Open symbols represent regions where macrophase separation between a conjugate-rich ordered phase and a water-rich phase is observed.
Figure 4-2. Representative SAXS intensities are shown for (a) EGFPP12.8 (b) mChP12.8 (c) EGFPP26.3 and (d) mChP26.3. The various phases are denoted as disordered (Dis), disordered micellar (DM), nonbirefringent lamellar (NB Lam), lamellar (Lam), and hexagonal (Hex).
Table 4-2. Thermal transitions for EGFPP and mChP block copolymers.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Conc. [wt.%]</th>
<th>$T_{DPLS}^{a)}$ [°C]</th>
<th>$T^{b)}$ [°C]</th>
<th>$T_{DSC}$ [°C]</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>o</td>
<td>34.6 - 38.1</td>
<td>34.7</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>o</td>
<td>34.1 - 35.8</td>
<td>34.2</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>o</td>
<td>33.2 - 36.1</td>
<td>34.1</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>o</td>
<td>–</td>
<td>34.3</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>o</td>
<td>–</td>
<td>31.3</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>o</td>
<td>–</td>
<td>25.9</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>o</td>
<td>&gt;27.4</td>
<td>34.5</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>o</td>
<td>&gt;28.2</td>
<td>31.3</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>o</td>
<td>&gt;27.6</td>
<td>32.1</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>o</td>
<td>–</td>
<td>30.2</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>o</td>
<td>–</td>
<td>26.3</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>o</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>o</td>
<td>33.0 - 36.4</td>
<td>33.2</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>o</td>
<td>32.1 - 35.7</td>
<td>32.8</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>o</td>
<td>31.7 - 35.5</td>
<td>32.1</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>32.3</td>
<td>–</td>
<td>33.0</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>32.2</td>
<td>–</td>
<td>31.7</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>35.1</td>
<td>–</td>
<td>31.8</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>*</td>
<td>–</td>
<td>29.9</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>o</td>
<td>&gt;31.8</td>
<td>33.3</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>o</td>
<td>&gt;32.2</td>
<td>32.3</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>o</td>
<td>&gt;32.5</td>
<td>32.1</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>35.4</td>
<td>&gt;34.3</td>
<td>31.2</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>35.7</td>
<td>&gt;35.7</td>
<td>29.5</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>37.6</td>
<td>&gt;37.4</td>
<td>29.8</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>*</td>
<td>–</td>
<td>27.4</td>
<td></td>
</tr>
</tbody>
</table>

a) A temperature listed for a transition in DPLS means that the sample is initially birefringent at 10 °C and loses birefringence at the listed transition temperature. o denotes samples that never display birefringent behavior at all temperatures, • denotes samples that remain birefringent at all temperatures; b) – denotes samples in which thermal transitions are not observed.

Under conditions where water is a good solvent for both PNIPAM and mCherry/EGFP, the phase behavior of the two conjugates is surprisingly similar given the large differences in surface potential between the two proteins. The pair of bioconjugates with PNIPAM coil fraction around 0.37 show a disordered phase at temperatures below the PNIPAM
thermoreponsive transition temperature at concentrations up to 40 wt.%. The disordered region is followed by the appearance of a nonbirefringent lamellar phase at 50 wt.%, as evidenced by SAXS peaks at q* and 2q*. Copolymers with a coil volume fraction of approximately 0.55 show a wider concentration window for microphase separation and a larger variety of ordered phases. At 20 wt.% concentration, both EGFPP26.3 and mChP26.3 show a disordered phase at temperatures below the PNIPAM block thermoresponsive transition temperature. The C_{ODT} depends weakly upon the identity of the protein block, increasing from 30 wt.% to 33 wt.% upon changing the protein from EGFP to mCherry. For all concentrations above C_{ODT}, conjugates of both proteins self-assemble into lamellar phases at 10 °C. The ordering improves with increasing concentration, as the lamellar morphology of both conjugates becomes birefringent at 37 wt.%.

With increasing temperature, both sets of conjugates show disordered micellar structures for concentrations up to 40-50 wt.%. However, the observation of disordered micelles as a homogeneous phase depends significantly on the protein interactions. Consistent with previous observations, the disordered micelles of mChP macrophase separate into conjugate-rich and conjugate-poor phases at 40 °C and low concentrations, remaining macrophase homogeneous at sufficiently high concentrations. Because the PNIPAM is dehydrated and segregated from water at elevated temperatures, it is likely that these high temperature phases are kinetically arrested morphologies. In contrast, EGFP-PNIPAM conjugates are observed to undergo a moderate increase in turbidity around 30 °C, but the transmission recovers upon further heating, suggesting the formation of a narrow two-phase region accompanying the transition from the low temperature disordered to the high temperature disordered micellar nanostructure. These observations are consistent with previously studied thermoresponsive protein-polymer conjugates containing EGFP, where micelles were found to be stable in solution above the
polymer thermal transition. In contrast, mChP micelles were found to be unstable, even at concentrations as low as 1%. Other authors have similarly found that the type of globular protein in a micellar corona can potentially have a large impact on micelle stability. The results here show that this effect is manifest in the concentrated solution self-assembly behavior under poor solvent conditions for the polymer block.

At sufficiently high concentrations, the conjugate solutions remain homogeneous as temperature is increased, but the deswelling of the PNIPAM domain at high temperatures causes an effective reduction in polymer coil fraction. This leads to order-order transitions (OOTs) from lamellar to disordered micellar and lamellar to hexagonal phases for EGFPP12.8 and mChP12.8, respectively. At higher coil fractions, transitions from lamellar to hexagonal to disordered micellar and lamellar to hexagonal phases are observed for EGFPP26.3 and mChP26.3, respectively. Therefore, phase diagrams for all four materials (Figure 4-1) illustrate that the identity of the protein has a significant impact on self-assembly primarily at high concentration and high temperature. The higher OOT temperatures in EGFPP12.8 relative to mChP12.8 and for the hexagonal phase observed at 45 wt.% in EGFPP26.3 relative to mChP26.3 indicate that a smaller effective PNIPAM volume fraction (greater extent of desolvation) is required in the EGFP conjugates to form phases with higher interfacial curvature, leading to shifts in the phase transition lines between mCherry-PNIPAM and EGFP-PNIPAM conjugates. The high thermal and folding stabilities of mCherry and EGFP, the thermal reversibility of the phase transitions (Figure C-5), and circular dichroism experiments performed before and after self-assembly (Figure C-12 and prior results) suggest that differences in the phase behaviors between mCherry and EGFP conjugates are not the result of irreversible effects of protein folding and/or aggregation but rather due to differences in protein interactions.
The domain spacings of the low temperature lamellar phases are only slightly larger in EGFP conjugates than in mCherry conjugates, consistent with the highly similar size and molar mass of both pairs of conjugate molecules. For a coil fraction of 0.37, the domain spacings are relatively close: \( d = 20.4 \pm 0.5 \text{ nm} \) and \( 19.9 \pm 0.5 \text{ nm} \) for EGFP12.8 and mChP12.8, respectively, at \( T = 10 \text{ °C} \). However, at a coil fraction of 0.55, the lamellae have a domain spacing that ranges from \( d = 28.6 \pm 1.4 \text{ nm} \) to \( 29.9 \pm 1.5 \text{ nm} \) with changing concentration for EGFP26.3 and a domain spacing of \( d = 27.3 \pm 1.2 \text{ nm} \) to \( 28.6 \pm 1.4 \text{ nm} \) for mChP26.3. The domain spacings for both EGFP26.3 and mChP26.3 in the concentrated solution phase are consistent with a bilayer lamellar morphology (Supporting Information).

EGFP and mCherry conjugates also display small differences in PNIPAM desolvation temperatures due to the effect of the protein block. DSC measurements show that the PNIPAM thermal transition temperature for all four conjugates decreases with increasing concentration, consistent with observed behavior of PNIPAM homopolymers. However, the PNIPAM collapse transition for mChP is lower than EGFP at the same PNIPAM coil fraction and bioconjugate concentration, and the difference increases with increasing concentration. This indicates that the presence of mCherry slightly favors the desolvation of PNIPAM in comparison to EGFP, suggesting that mCherry more favorably competes for water in the hydrated structures.

As expected from solution self-assembly studies, both EGFP- and mCherry-based block copolymers self-assemble into similar solid-state structures when formed by solution casting from a nonselective solvent, further suggesting that coarse-grained effects play a dominant role in governing self-assembly from neutral solvents. The solid-state SAXS patterns for EGFP12.8 show a disordered phase both before and after solvent annealing. TEM (Figure 4-3) shows poorly ordered structures, consistent with the SAXS results. mChP12.8 SAXS patterns show
slightly better ordering as evidenced by a narrower primary peak, but TEM images continue to show features lacking significant long range order.

Figure 4-3. (a) Solid-state SAXS of EGFPP26.3, mChP26.3, EGFPP12.8, and mChP12.8 after annealing in water at room temperature and the corresponding TEM images of (b) EGFPP12.8 (c) mChP12.8 (d) EGFPP26.3 and (e) mChP26.3. TEM images of EGFPP26.3 and mCh26.3 show lamellar morphologies, while the lower coil fraction materials have little

EGFPP26.3 and mChP26.3 samples both show lamellar morphologies in the solid state by SAXS and TEM (Figure 4-3). Although the nanostructures appear qualitatively similar by TEM, differences are observed in SAXS patterns. In the as-cast sample, EGFPP26.3 shows no higher-order reflections, while mChP26.3 shows a second order reflection. Following an 8 hr anneal in water at room temperature, both materials show improved ordering as evidenced by a slight sharpening of higher order reflections. However, the peak intensities in the two materials differ, suggesting differences in the form factor between the two materials. It is known that perfectly symmetric lamellae will have a forbidden \(2q^*\) reflection,\(^6\) which may account for its suppression in EGFPP26.3 and suggest that mChP26.3 packs into slightly asymmetric lamellae.
The domain spacings of the lamellar structures are also observed to be slightly different: 21.3 ± 0.3 nm and 23.2 ± 0.4 nm for EGFPP26.3 and mChP26.3, respectively. From scaling relationships for domain spacing (Supporting Information), both EGFPP26.3 and mChP26.3 are hypothesized to form a bilayer lamellar morphology. Given the protein dimensions, this slightly smaller domain spacing for EGFPP26.3 is consistent with the formation of a symmetric bilayer lamellar nanostructure. It is interesting to note that while EGFPP26.3 leads to a smaller domain spacing in the neat block copolymer, in the concentrated solution phase, the lamellar domain spacing of EGFPP26.3 is greater than that of mChP26.3.

Although the two proteins have very different amino acid sequences, the coarse-grained properties of mCherry and EGFP show significant similarity. Therefore, the similarities in self-assembly behavior between mCherry and EGFP suggest that the chemical details of the solvent-accessible protein surface are less important than coarse-grained interactions or protein shape in determining the thermodynamics of self-assembly. Measurements of the protein-protein second virial coefficient and protein solubility in PNIPAM solutions show that EGFP and mCherry are similar, consistent with the similarity of the self-assembly phase behavior for conjugates of these proteins. In 10 mM Tris-Cl, pH = 7.0 at T = 25 °C, the second virial coefficient of EGFPC14S was measured to be 1.0 x 10^{-4} ± 0.2 mol mL g^{-2} (Figure C-6), comparable to that measured for mCherry (1.1 x 10^{-4} ± 0.1 mol mL g^{-2}) in the same buffer. Although the virial coefficients are similar, EGFP is more highly charged and has a lower free energy, ΔGs, than mCherry. EGFP displays a relatively large region of negative electrostatic potential and has a more negative surface charge than mCherry at pH = 8.0 as measured by zeta potential, consistent with the fact that EGFP has a lower pI than mCherry (Table 4-3). The solvation free energies for mCherry and EGFP were calculated to be -12.5 kcal/mol and -9.9 kcal/mol (Supplementary Information),
respectively, using the method developed by Eisenberg and McLachlan.\textsuperscript{62} mCherry has a larger solvation free energy, suggesting that EGFP has a slightly more hydrophobic solvent-accessible surface than does mCherry, and the hydrophobic and electrostatic effects appear to offset, resulting in similar virial coefficients for the two proteins.

**Table 4-3. Biophysical properties of mCherry and EGFP.**

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Theor. M&lt;sub&gt;W&lt;/sub&gt; [kDa]</th>
<th>M&lt;sub&gt;W&lt;/sub&gt; from SLS [kDa] *</th>
<th>A&lt;sub&gt;2&lt;/sub&gt; [mL mol g&lt;sup&gt;-2&lt;/sup&gt;] (x 10&lt;sup&gt;4&lt;/sup&gt;)\textsuperscript{b}</th>
<th>Zeta Potential [mV]</th>
<th>Measured pI</th>
<th>( \Delta G_s ) (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mCherry</td>
<td>28.1</td>
<td>38.5 ± 0.4</td>
<td>1.1 ± 0.1</td>
<td>-21.9 ± 5.4</td>
<td>6.2\textsuperscript{63,b)}</td>
<td>-12.5</td>
</tr>
<tr>
<td>EGFP14S</td>
<td>29.5</td>
<td>39.1 ± 0.1</td>
<td>1.0 ± 0.2</td>
<td>-26.8 ± 6.2</td>
<td>4.8-5.0\textsuperscript{64,5.6\textsuperscript{63}}</td>
<td>-9.9</td>
</tr>
</tbody>
</table>

\( a) \) Measurements made in 10 mM Tris-Cl, pH = 7.0, \( T = 25 \) °C; \( b) \) Value of pI reported here corresponds to DsRed2, a dimeric RFP similar in amino acid sequence to the mCherry mutant.

Protein/polymer miscibility is assessed through macrophase separation experiments on blends of concentrated protein and polymer solutions (Figure C-7), suggesting that the interactions between protein and polymer are similar for both proteins studied. Blend compositions with equal mass of protein and PNIPAM were prepared up to a total concentration of 250 mg/mL; at this concentration, both EGFP/PNIPAM and mCherry/PNIPAM solutions macrophase separate into protein-rich and polymer-rich phases at 4 °C. Both proteins are soluble to greater than 300 mg/mL concentration, so all experiments are performed within the range of protein solubility. Macrophase separation originates from the same thermodynamic driving force as molecular self-assembly, and as in coil-like polymers it occurs at a lower concentration than block copolymer self-assembly.\textsuperscript{65,66} As concentration is decreased, the behavior for mCherry and EGFP solutions is largely similar, suggesting that the majority of the physics driving phase separation is invariant with respect to the detailed changes in surface potential. However, mCherry/PNIPAM solutions show a slightly higher demixing temperature than EGFP14S/PNIPAM solutions. This illustrates that EGFP14S/PNIPAM blends macrophase separate at a slightly lower concentration than mCherry/PNIPAM blends at a given temperature,
which is consistent with the slightly lower C_{ODT} observed in the EGFP conjugates. These similarities in phase behavior, despite large differences in the protein’s specific surface amino acid residues, suggest that the coarse-grained protein-polymer conjugate properties have the largest impact on the miscibility behavior driving protein-polymer self-assembly in these systems, while the specific surface potential of the proteins plays a secondary role. The idea that protein-polymer self-assembly thermodynamics may be governed by coarse-grained interactions is consistent with theories of protein solutions which are able to capture phase behavior and crystallization by treating complex proteins as simple colloids with coarse-grained interactions measured by parameters such as the second virial coefficient.\(^{67-69}\)

**Comparison of mCherry Mutants with Modified Patchy Interactions**

In order to better understand the generality of surface patchiness effects on the self-assembly of globular protein-polymer block copolymers, single point mutations on multiple residues of mCherry were performed and the self-assembly of a few mCherry variants with different total charges was characterized. Based on an electrostatic potential map analysis of the mCherry crystal structure, seven potential mutation sites—\(E26, E30, E34, E100, K123, E123\) and \(D200\)—were identified that could have a large impact on the electrostatic surface potential of the protein without significantly affecting the protein fold due to minimal involvement in salt bridging (Scheme 4-1). To either neutralize or invert the charge of a given residue, negatively charged residues were mutated to serine (S) or lysine (K) while positively charged residues were replaced with S or glutamic acid (E). Small scale cultures of all variants showed different color intensities (Figure 4-4a), although the cell densities in each culture were similar (\(\sim 10\text{mg} / 1.5\text{mL}\)), indicating either different expression levels among the mCherry variants or incorrect folding of mCherry mutants. Activity and secondary structure analyses of the purified mCherry variants
showed that they were not functionally or structurally different than the original mCherry (Figure 4-4b and C-8). Therefore, the different color intensities of mCherry variants during protein expression were mainly caused by their expression levels, not by incorrect folding of mCherry.

An electrophoretic mobility shift assay of most variants on a native gel was performed to confirm the charge changes; the different electrostatic potentials of each mutant compared to mCherry clearly resulted in varying magnitude shifts in the gel (Figure 4-4c). From the results, the variant pair E26S and E26K were selected for conjugate synthesis because they make changes to the surface potentials at the same location that are of different magnitude (both more negative relative to mCherry) and have higher protein expression levels compared to other mutation pairs. Although their expression levels are lower than other mutants, K123S and K123E were also selected since these are the only mutants where positively charged residues are replaced by neutral and negatively charged residues, respectively. The effect of these mutations on the electrostatic surface potential of mCherry is shown in Scheme 4-1.

Consistent with the electrophoretic mobility shift assay of the mutants, zeta potential measurements of the selected variants show differences in the overall surface charge that follow the same order as changes in the formal charge of the protein. Zeta potential values of -29.3±6.4 (K123E), -22.8±7.9 (K123S), -21.9±5.4 (mCherry), -15.8±9.4 (E26S), and -5.1±10.1 mV (E26K) were obtained for the most negatively charged to the least negatively charged mCherry mutants in pH 8 buffer. Therefore, tuning the electrostatic surface patchiness of mCherry was accomplished with no significant structural and functional changes on its original folded structure, leading to a set of mutants where patchiness effects can be directly assessed.

Despite changes in the electrostatic patchiness of the mCherry mutants, globular protein-polymer self-assembly was not significantly affected. A PNIPAM block of molar mass 26.0
Figure 4-4. (a) Cell culture color intensities showing the different expression levels among mCherry and the mCherry variants. (b) Far UV circular dichroism spectroscopy showing that the mutations do not adversely affect the secondary structure. (c) Native PAGE showing monomer forms of the mCherry variants at 50 mM BME.

kg/mol was conjugated to four selected mCherry variants and mCherry (Table 4-1) to produce a set of mCherry-PNIPAM bioconjugates in which the polymer block was controlled. The concentrated solution self-assembly behavior of the resulting conjugates was investigated at $T = 15 \, ^{\circ}C$ to measure the surface charge effects on the order-disorder transition concentration. Irrespective of the mCherry point mutation, lamellar structures of mCherry-$b$-PNIPAM26k appeared at at $C_{ODT}$s of 33-37 wt. % in water (Figure 4-5). Well-ordered structures, as evidenced by sharp higher order peaks in the SAXS curves, always appeared between 37 and 40 wt.%. The domain spacing change at 50 wt.% between the most negative mCherryK123E-$b$-PNIPAM26k and least negative mCherryE26K-26k PNIPAM (a difference in surface charge by 25 mV as measured by zeta potential) was 26.2 nm and 26.6 nm, respectively, indicating no significant electrostatic surface potential effect on globular protein-polymer self-assembly for this
magnitude of electrostatic potential change. As was observed in the similar phase behavior between mCherry- and EGFP-PNIPAM in which the electrostatic surface potentials are significantly different, local changes to the protein surface potential have only small effects on the self-assembly behavior of globular protein-polymer block copolymers.

Figure 4-5. Concentrated solution SAXS of mCherry and its variants conjugated to 26k PNIPAM ($\phi_{\text{PNIPAM}} = 0.54$) in H$_2$O, showing that the order-disorder transition concentration (ODTC) is minimally affected by the surface residue point mutations.

Varying the salt concentration to screen electrostatic interactions indicates that electrostatic interactions between proteins are effectively attractive in these systems, promoting ordering. Compared to the self-assembly behavior in water, the $C_{\text{ODT}}$ in the presence of 200 mM NaCl were measured to be approximately 5 wt.% higher (Figure C-9). In presence of salt, the protein-protein electrostatic interactions are screened; the higher observed $C_{\text{ODT}}$ suggests that these interactions are effectively attractive under the self-assembly conditions. This effect has been observed by several groups in synthetic polymers, where the presence of charged groups can promote phase separation of ionic domains in ionomers. Proteins may be likened to macroions, and while Derjaguin-Landau-Verwey-Overbeek (DLVO) theory predicts repulsive
interactions between two similarly electrically charged colloidal particles, this is in the infinite dilution limit where the particles may exist without their counterions; theory and computational results support the conclusions that the effects of counterions in experimentally accessible conditions cannot be ignored, leading to counterion-mediated interactions. \textsuperscript{72,73} Despite extensive dialysis of protein-polymer conjugates to ultrapure deionized water, counterions strongly bound to the protein surface are expected to remain, and at the relatively high concentrations of self-assembly, the attractive counterion-mediated interactions between protein molecules can be important. Comparison of domain spacings (50 wt.%) in the absence or presence of NaCl only showed subnanometer increases (0.1-0.6nm) due to increasing salt concentration. The fact that a large increase in salt concentration yields only small changes in the domain spacing and protein-protein packing in the self-assembly suggests that the interactions between proteins may already be highly screened in the concentrated state. However, these results do not preclude significant ionic strength effects effect on protein-polymer self-assembly when larger changes in protein charge are investigated.

4.5 Conclusions

The self-assembly of PNIPAM conjugated to mCherry and EGFP, two structurally similar proteins with significantly different chemical composition and surface potentials, and the self-assembly of PNIPAM-conjugated mCherry variants differing in point mutations were compared to provide insight into the impact of changes in protein surface potential on self-assembly in a system where the protein structure is controlled. This comparison suggests that coarse-grained properties such as protein shape, solubility, total charge, and virial coefficient are largely responsible for the general shape of the phase diagram in nonselective solvents. However, desolvation of the PNIPAM block with increasing temperature leads to different
phases depending upon the protein block, with mCherry conjugates exhibiting a wide two-phase region and EGFP conjugates forming homogeneous disordered micellar solutions. The presence or absence of macrophase separation in the high temperature disordered micellar phase is correlated with micelle stability determined in dilute solution of conjugates. 

Despite the different surface potentials of mCherry and EGFP, the phase behavior of mCherry and EGFP conjugates is highly similar in neutral solvent conditions, consistent with highly similar protein-polymer miscibility and protein-protein second virial coefficients for the two proteins. The $C_{ODT}$ with increasing concentration is similar for both proteins, and both form similar nanostructures in solution and the solid state. 

Point mutations to mCherry that alter its electrostatic surface patchiness also show little effect on $C_{ODT}$ or the type of nanostructure formed, further indicating that electrostatic patchiness is less important than colloidal properties of the protein. The presence of high NaCl concentration to screen ionic interactions yields an increase in $C_{ODT}$, suggesting that the ionic interactions are effectively attractive in these systems. Therefore, for all structurally similar proteins studied, local changes to the protein surface patchiness led to minimal effects on self-assembly, suggesting that coarse-grained models from polymer and colloidal physics may be applied to effectively capture a large portion of the self-assembly behavior in globular protein-polymer conjugate diblock copolymers. 

4.6 References 


Chapter 5. The Effect of Protein Electrostatic Interactions on Globular Protein-Polymer Block Copolymer Self-Assembly

5.1 Abstract

Mutation of a superfolder GFP was used to design GFP variants with formal net charges of 0, -8, and -21, providing a set of three proteins in which the total charge is varied to tune protein-protein interactions while controlling for the protein size and tertiary structure. After conjugating each of these three GFP variants to poly(N-isopropylacrylamide) (PNIPAM), the concentrated solution phase behavior of these three block copolymers is studied using a combination of small-angle X-ray scattering, depolarized light scattering (DPLS), and turbidimetry to characterize their morphologies. The electrostatic repulsion between supercharged GFP suppresses ordering, increasing the order-disorder transition concentration (CODT) and decreasing the quality of the ordered nanostructures as measured by the full width at half maximum of the primary scattering peak. In contrast, the charge distribution of the neutrally charged GFP results in its largest dipole moment, calculated about the protein’s center of mass, among the three GFP variants and a self-complementary Janus-like electrostatic surface potential that enhances nanostructure formation. The different electrostatic properties result in different protein-protein interactions that affect the high temperature morphologies, including the formation of macrophase separated or homogeneous micellar phases and the smaller hexagonal ordering window of the supercharged GFP. Small improvements in the quality of the ordered nanostructures of GFP(-21)-PNIPAM can be achieved through protein-divalent cation interactions. Therefore, supercharging is demonstrated as a method of tuning the magnitude and directionality of protein-protein interactions to control self-assembly.
5.2 Introduction

The biodegradability and renewability of enzymes, their ability to function under mild reaction conditions, and their high degree of specificity toward substrates make them very desirable materials to address a diversity of challenges. For energy applications, enzymatic fuel cells (EFCs) utilize the ability of enzymes to catalyze fuel oxidation at the anode and oxidant reduction at the cathode, enabling miniaturization of EFCs.1,2 Light-harvesting materials incorporating reaction centers (RCs)3 and photosynthetic complexes4 have been developed for bioelectronics and energy applications. Recently, fluorescent proteins have been used as a biolasing gain medium in solid-state and optofluidic lasers.5,6 The specificity and biodegradability of enzymes have garnered much interest for enzyme technology in environmental applications such as waste and soil treatment through the removal of specific pollutants.7 The unparalleled catalytic capability of enzymes has been utilized to advance sensor technologies, with glucose oxidase having been used to pioneer the development of biosensors for biomedical applications.8 Other sensing technologies have been developed for environmental monitoring9 and neurotoxin detection and decontamination to address challenges in national security.10–12 The pharmaceutical, chemical, food and agriculture, paper, and textile industries have also embraced enzyme technology.13,14

Augmenting materials design with the functionality of enzymes requires overcoming a number of engineering challenges—enzyme stability, control over the spatial arrangement and orientation of the protein, and enzyme density—that can be addressed through different protein nanopatterning techniques. Enzymes are susceptible to loss of activity through unfolding and denaturation when placed in a non-native environment with deleterious pH, solvent, salinity, and
temperature conditions.\textsuperscript{15–17} Care must be utilized to incorporate enzymes in such a manner that their active sites are easily accessible by controlling their position and orientation, allowing for facile transport of substrate, product, and cofactors.\textsuperscript{18,19} Finally, providing a pathway that enables charge transport is also essential for biocatalytic applications.\textsuperscript{20}

Direct self-assembly of proteins as blocks in block copolymers provides an elegant solution to achieving the necessary nanostructural control within a multicomponent material.\textsuperscript{21,22} Inspiration for direct self-assembly as an approach for patterning enzymes is drawn from the rich phase behavior of coil-coil block copolymers.\textsuperscript{23,24} These self-assembly principles have been applied to form nanostructured materials from globular protein-polymer block copolymers, demonstrating the ability to make highly active biocatalytic materials.\textsuperscript{25} In addition to a number of studies of self-assembly in dilute solution\textsuperscript{26–31} and in thin films,\textsuperscript{25} self-assembly principles have been applied to form solid-state nanostructured plastics and gels from globular protein-polymer block copolymers, forming many nanostructured phases observed in coil-coil block copolymers—lamellae, perforated lamellae, hexagonal cylinders, and disordered micelles.\textsuperscript{32–36}

Differences in the phase behavior of coil-coil block copolymers and globular protein-polymer block copolymers reflect differences in the underlying physics governing their self-assembly. Self-consistent mean field theory predicts that the Flory-Huggins interaction parameter $\chi$, the degree of polymerization $N$, and the volume fraction $f_A$, representing the fraction of block $A$, govern the phase behavior. In concentrated block copolymer solutions, according to the “dilution approximation” a neutral solvent will uniformly screen unfavorable A-B interactions, leading to the term $\phi\chi N$ governing the phase behavior, where $\phi$ represents the block copolymer volume fraction in solution.\textsuperscript{37,38} In contrast to a random coil polymer, the diverse shape, size, and composition of enzymes result in a complex set of anisotropic hydrophobic, ionic, and hydrogen bonding interactions. The governing thermodynamics of globular protein-polymer block
copolymers has been shown to undergo lyotropic and thermotropic ODTs and OOTs in a rich a highly asymmetric phase space that is significantly different than that of coil-coil block copolymers due to the interaction potentials of proteins and the difference in shape between proteins and coil polymers.

Detailed studies have elucidated many of the relevant effects governing the self-assembly process in the complex protein-polymer hybrid systems. Changes in polymer chemistry have led to significant changes in phase transition lines and ordered morphologies, yielding for the first time a cubic phase consistent with a bicontinuous gyroid structure; change in polymer topology has also been observed to have a significant effect on phase behavior and micellar stability. In contrast, comparative studies between mCherry and EGFP, two proteins with nearly identical tertiary structure but very different surface potentials, demonstrated very little change in the phase behavior. This suggests that coarse-grained properties such as the second-virial coefficient, protein size, and colloidal shape of the protein are critical to understanding protein-polymer block copolymer phase behavior. However, among these coarse grained parameters, the role of protein electrostatic interactions has not been explored.

Herein, the effect of protein electrostatic interactions is investigated by comparing the phase behavior of globular protein-polymer conjugates containing structurally homologous proteins with different values of formal net charge using a supercharged series of GFP mutants. Using the superfolder GFP as the reference globular protein, a series of mutants is prepared with variable charge, electrostatic surface potential, and electric dipole moment. While conjugates of all three GFP variants to PNIPAM are able to form long-range ordered nanostructures, the electrostatic repulsion of the supercharged protein is shown to affect the phase behavior significantly, suppressing nanostructure formation and reducing the quality of ordering in the self-assembled nanostructures.
Scheme 5-1. (a) Cartoon of the superfolder GFP crystal structure (PDB 2B3P).\textsuperscript{43} (b) The dipole moments are calculated using the Protein Dipole Moments Server\textsuperscript{44} and are illustrated for GFP(0) (left), GFP(-8) (middle), and GFP(-21) (right), which have magnitudes of $1.6 \times 10^{-27}$ (C·m), $5.3 \times 10^{-28}$ (C·m), and $3.0 \times 10^{-28}$ (C·m), respectively. The dipole moments are calculated as $D = \sum r_i q_i$ about the proteins' centers of mass, and the lengths of the arrows are drawn to scale in atomic units. (c) Different views of the electrostatic surface potential ($\pm 10$ kT/e) at the solvent-accessible surface of GFP(0), GFP(-8) (PDB 2B3P), and GFP(-21) rendered from solutions of the linearized Poisson-Boltzmann equation using the Adaptive Poisson-Boltzmann Solver (APBS).\textsuperscript{45-47} Red and blue represent negative and positive values, respectively.

5.3 Materials and Methods

Synthesis

Poly(N-isopropylacrylamide) (PNIPAM) was synthesized by reversible addition-fragmentation chain transfer (RAFT) polymerization as previously described.\textsuperscript{36} The molar mass and dispersity were analyzed by gel permeation chromatography using an Agilent Technologies 1260 Infinity system equipped with two ResiPore, 7.5 x 300 mm columns (Agilent Technologies), each with a molecular weight range up to 500,000 g/mol, in N,N-dimethylformamide (DMF) with 0.02 M lithium bromide (LiBr) as the mobile phase (Figure D-1). The superfolder green fluorescent protein (GFP) (overall formal net charge of -8) and two variants with overall formal net charges of -21 and 0 (DNA and amino acid sequences shown in Figure D-2), each containing an N-terminal 6xHis tag, were expressed in the Escherichia coli strain Tuner(DE3). Henceforth, the GFP variants will be referred to by their charge as GFP(0),
GFP(-8), and GFP(-21). GFP(0) and GFP(-8) were expressed by inoculating 1 L of LB medium with 5 mL of overnight culture and incubating at 37 °C until OD$_{600}$ ≈ 0.8 - 1.0, after which the cultures were induced with 1 mM IPTG and harvested after growing for 9 - 10 hours at 25 °C. The supercharged variant, GFP(-21), was expressed for 11 - 12 hours at 37 °C after inoculating and then harvested. Cell pellets were resuspended in lysis buffer, incubated with 1 mg/mL lysozyme at 4 °C for 30 minutes, and sonicated. After clarifying the lysate by centrifugation, the proteins were purified by Ni-NTA metal affinity chromatography (Figure D-3). GFP(-8) and GFP(-21) were dialyzed to 20 mM Tris-Cl, pH = 8.0, and GFP(0) was dialyzed to 20 mM Tris-Cl, 250 mM NaCl, pH = 8.0 due to its low solubility in low ionic strength buffers.

**Bioconjugation**

The coupling reactions between the GFP variants and maleimide end-functionalized PNIPAM were performed in 20 mM Tris-Cl, pH = 8.0 for GFP(-8) and GFP(-21) and in 20 mM Tris-Cl, 250 mM NaCl, pH = 8.0 for GFP(0). Protein solutions at approximately 1 - 2 mg/mL were first incubated with a 10x molar excess of tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) at 4 °C for approximately 45 - 60 minutes. A 6x molar excess of maleimide-functionalized polymer was then added to the solution and allowed to dissolve and react at 4 °C overnight for up to 20 - 24 hours. After completion, the conjugation reaction was precipitated three times in 1 M ammonium sulfate solution at pH = 8.0 to remove unreacted protein. Unreacted polymer was subsequently removed by Ni-NTA metal affinity chromatography. Purified bioconjugate was then dialyzed against milliQ water. Purity was confirmed by both SDS-PAGE and Native PAGE (Figure D-4). Two sets of bioconjugates with different PNIPAM coil fractions (M$_n$ = 21.6 kDa, $\phi_{PNIPAM} = 0.52$, and M$_n$ = 30.7 kDa, $\phi_{PNIPAM} = 0.60$) (Supporting Information) were synthesized using GFP(0), GFP(-8), and GFP(-21), as detailed in Table 5-1.
UV-Vis spectrophotometry and circular dichroism (CD) spectroscopy show retention of protein optical activity and secondary structure after conjugation and after self-assembly and rehydration (Figures D-5 – D-8).

<table>
<thead>
<tr>
<th>Molecule</th>
<th>PNIPAM $M_n$ (kg/mol)</th>
<th>$D_{PNIPAM}$</th>
<th>Bioconjugate $M_n$ (kg/mol)</th>
<th>$\phi_{PNIPAM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP(0)PN21</td>
<td>21.6</td>
<td>1.06</td>
<td>50.1</td>
<td>0.52</td>
</tr>
<tr>
<td>GFP(-8)PN21</td>
<td>21.6</td>
<td>1.06</td>
<td>50.0</td>
<td>0.52</td>
</tr>
<tr>
<td>GFP(-21)PN21</td>
<td>21.6</td>
<td>1.06</td>
<td>50.1</td>
<td>0.52</td>
</tr>
<tr>
<td>GFP(0)PN30</td>
<td>30.7</td>
<td>1.09</td>
<td>59.2</td>
<td>0.60</td>
</tr>
<tr>
<td>GFP(-8)PN30</td>
<td>30.7</td>
<td>1.09</td>
<td>59.1</td>
<td>0.60</td>
</tr>
<tr>
<td>GFP(-21)PN30</td>
<td>30.7</td>
<td>1.09</td>
<td>59.2</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Sample Preparation and Characterization

Concentrated solution samples were prepared by rehydrating the solid globular protein-polymer block copolymer to the desired concentration. The concentrated solution phase behavior of GFP-PNIPAM conjugates was characterized using a combination of small-angle X-ray scattering (SAXS) (Figures D-9 and D-10), depolarized light scattering (DPLS) and turbidimetry (Figures D-11 – D-16), and differential scanning calorimetry (DSC) (Figure D-17). Further details of these characterization techniques are provided in the Supporting Information.

5.4 Results and Discussion

Electrostatic Effects on Self-Assembly of Structurally Homologous Proteins

A series of three proteins was synthesized with a range of formal charge: 0, -8, and -21, referred to as GFP(0), GFP(-8), and GFP(-21) (Scheme 1). The different total net charge and charge distribution of the three GFPs result in their different electric dipole moments. Because the dipole moment of a protein with a nonzero monopole will depend upon the placement of the protein within the coordinate system, the dipole moments are calculated as $D = \sum r_i q_i$, with the proteins centered at their center of mass, enabling a meaningful comparison of the dipole
moments of the three GFP variants. The electric dipole moments are calculated using the Protein Dipole Moments Server. The magnitudes of the electric dipole moments are found to decrease in the order GFP(0) > GFP(-8) > GFP(-21); they are calculated to be $1.6 \times 10^{-27}$ C·m, $5.3 \times 10^{-28}$ C·m, and $3.0 \times 10^{-28}$ C·m for GFP(0), GFP(-8), and GFP(-21), respectively. The dipole reflects the greater polarization of GFP(0) which results in a Janus-like electrostatic surface potential, promoting complementary electrostatic patchy directional interactions that can potentially improve self-assembly.

Comparison of the self-assembly in concentrated solution of GFP variants with different total formal charges of 0, -8, and -21 shows common morphologies formed by all three variants but also reveals the impact that total net charge and electrostatic surface potential have on self-assembly. These similarities and differences are illustrated in the phase diagrams for $\phi_{\text{PNIPAM}} = 0.52$ and $\phi_{\text{PNIPAM}} = 0.60$ (Figure 5-1) in which the proteins GFP(0), GFP(-8), and GFP(-21) are conjugated to the same molecular weight PNIPAM to enable a direct comparison. Phase identifications are performed using a combination of SAXS to identify nanodomain symmetry, DPLS to measure the optical anisotropy of nanostructures, turbidimetry to identify macrophase separation, and DSC to measure the PNIPAM thermal transition (Table 5-2).
Figure 5-1. Phase diagrams of (a) GFP(0)PN21, (b) GFP(-8)PN21, (c) GFP(-21)PN21, (d) GFP(0)PN30, (e) GFP(-8)PN30, (f) GFP(-21)PN30 as a function of temperature and concentration. Phases are identified as disordered (Dis), disordered micellar (DM), lamellar (Lam), and hexagonal (Hex). Open symbols represent regions where macrophase separation between a conjugate-rich ordered phase and a water-rich phase is observed. Semi-transparent symbols demarcated by the dashed lines denote ordered morphologies that are nonbirefringent.

At low temperature where water is a good solvent for both protein and polymer, the different charge of the GFP variants leads to a significant shift in the disorder-to-order concentration (C_{ODT}) at the larger coil fraction of $\phi_{\text{PNIPAM}} = 0.60$. All three GFPs show a lyotropic C_{ODT} from an isotropic disordered phase to a lamellar morphology, and the C_{ODTs} are similar at $\phi_{\text{PNIPAM}} = 0.52$: 35 wt.% for GFP(0) and GFP(-8) and 33 wt.% for GFP(-21). The quality of the lamellar ordering is observed to improve with increasing concentration, becoming birefringent at 45 wt.% for all three GFP variants at a PNIPAM coil fraction of 0.52. However, the birefringence of the supercharged GFP(-21) is more than an order of magnitude less than that of GFP(0) and GFP(-8) (Figures D-11 – D-13). At $\phi_{\text{PNIPAM}} = 0.60$, a significant difference in the C_{ODTs} is observed; the C_{ODT} of GFP(0) is 33 wt.%, while the C_{ODT} of GFP(-8) and GFP(-21) is significantly greater at 47 wt.%. The lamellar nanostructures of all three GFP variants at $\phi_{\text{PNIPAM}}$
= 0.60 are birefringent at their respective C_{ODT}s. For GFP(-21), SAXS shows a single peak at 43 wt.%, but at 25 and 30 °C, a reflection at 2q* is observed; it is likely that at low temperatures at 43 wt.% that the 2q* peak may be suppressed due to symmetry between the protein and polymer nanodomains.48

The slopes of the C_{ODT} transition lines are different between the sets of conjugates at the near-symmetric coil fraction reflects the differences in monopole-monopole interactions of the GFP variants. Within the resolution of the measured temperatures and concentrations, the slope C_{ODT} lines for \( \phi_{\text{PNIPAM}} = 0.52 \) (Figures D-1a, D-1b, and D-1c) are approximately 7.5 °C/wt.% for GFP(0) (over the concentration range of 35 - 37 wt.%), 3.0 °C/wt.% for GFP(-8) (over the concentration range of 35 - 40 wt.%), and 2.1 °C/wt.% for GFP(-21) (over the concentration range of 33 - 40 wt.%). The gradual tilting of the C_{ODT} transition line with increasing protein net charge shows that there is a greater propensity toward disordering due to the increasing monopole-monopole interactions with increasing protein charge. However, in these lower coil fraction conjugates, this is manifest most strongly at slightly elevated temperatures as the PNIPAM hydration starts to decrease, decreasing the unperturbed size of the coil block. At the larger coil fraction of \( \phi_{\text{PNIPAM}} = 0.60 \), the C_{ODT} occurs at a high packing density for GFP(-8) and GFP(-21) than for GFP(0), reflecting the stronger propensity toward self-assembly in the neutral net charge GFP. However, the transition lines are nearly vertical for all three of these higher coil fraction conjugates.

Increasing temperature is observed to lead to a disordered micellar phase for all three GFP variants above the PNIPAM desolvation transition, and the micellar stability shows trends due to the difference in net charge and the polymer coil fraction. At the smaller coil fraction of \( \phi_{\text{PNIPAM}} = 0.52 \), GFP(0)-PNIPAM forms stable, homogeneous micellar phases. In contrast,
GFP(-8) and GFP(-21) show a decrease in transmission below 90% of the initial transmission, the criterion chosen to determine the macrophase separation temperature, and only remain macrophase homogeneous at sufficiently high concentrations. For GFP(-8), at concentrations of 35 - 40 wt.%, the transmission recovers to greater than 90% of the initial transmission (Table 5-2 shows the temperature range in which the two-phase region is observed), and no decrease in transmission is observed at 45 and 50 wt.% (Figure D-12). The formation of a narrow two-phase region during the transition to the high temperature, macrophase homogeneous, disordered micellar nanostructure was previously observed in PNIPAM conjugates of the similar protein enhanced green fluorescent protein (EGFP).35 Similarly to GFP(-8), GFP(-21) shows a drop in transmission that recovers with further heating up to 40 °C, but the transmission does not recover to at least 90% of the initial transmission. Hence, the phase diagram for GFP(-21) in Figure 5-1c denotes the formation of a two-phase region for all concentrations less than 50 wt.%.

At the larger coil fraction of $\phi_{PNIPAM} = 0.60$, the propensity toward macrophase separation is opposite to that at the smaller coil fraction. For GFP(0), in which solutions remained homogeneous at all concentrations studied at the smaller coil fraction, macrophase separation is now observed at all concentrations studied at the larger coil fraction. GFP(-8) and GFP(-21) display greater regions of micellar phases that remain homogeneous at the larger coil fraction. GFP(-8) remains homogeneous at all concentrations, and GFP(-21) forms a narrow two-phase region at intermediate concentrations of 30 - 40 wt.%, the transmission recovering above 90% above approximately 36 °C, and shows no decrease in transmission at concentrations greater than 40 wt.%. In contrast to the other two GFPS, GFP(-21) forms a homogeneous disordered phase at high concentration that persists even up to 40 °C at 43 and 47 wt.%.
The net protein charge is observed to affect micellar phase stability at high temperature, with the intermediate charge of GFP(-8) showing the greatest micellar stability overall. When the GFP is overall net neutral, at the larger coil fraction, the PNIPAM desolvation driving force at high temperature is sufficiently high to lead to macrophase separation at all concentrations. While GFP(-8) and GFP(-21) both form micellar phases at both coil fractions, the greater charge and monopole-monopole interactions in GFP(-21) leads to a slightly greater tendency toward micelle aggregation, resulting in GFP(-21) having a larger two-phase region than GFP(-8).

The PNIPAM desolvation temperatures of the GFP conjugates are also measured by DSC, which shows that the desolvation temperature decreases with increasing concentration for conjugates of all three GFP variants (Figure D-17). This is consistent with the behavior of PNIPAM homopolymer and of previously studied mCherry-PNIPAM and EGFP-PNIPAM conjugates. The macrophase separation transition temperatures, in contrast, increase with increasing solution concentration. More water is accommodated through hydration of the protein domains with increasing conjugate concentration, which results in an increase in the macrophase separation transition temperature above the PNIPAM desolvation temperature needed to achieve a higher degree of PNIPAM collapse for macrophase separation to occur.
**Table 5-2.** Thermal transitions for GFP(0)-, GFP(-8)-, and GFP(-21)-PNIPAM block copolymers.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Conc. (wt.%)</th>
<th>T&lt;sub&gt;DPLS&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt; (°C)</th>
<th>T&lt;sub&gt;t&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt; (°C)</th>
<th>T&lt;sub&gt;DSC&lt;/sub&gt; (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP(0)PN21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>o</td>
<td>o</td>
<td>31.3</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>o</td>
<td>o</td>
<td>30.9</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>o</td>
<td>o</td>
<td>30.4</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>o</td>
<td>o</td>
<td>29.4</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>o</td>
<td>o</td>
<td>29.6</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>o</td>
<td>o</td>
<td>28.8</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>o</td>
<td>o</td>
<td>28.7</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>32.9</td>
<td>o</td>
<td>26.9</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>o</td>
<td>o</td>
<td>26.2</td>
<td></td>
</tr>
<tr>
<td>GFP(-8)PN21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>o</td>
<td>32.2</td>
<td>31.8</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>o</td>
<td>32.2</td>
<td>31.7</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>o</td>
<td>32.2</td>
<td>31.4</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>o</td>
<td>31.5</td>
<td>29.9</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>o</td>
<td>31.6 - 39.3</td>
<td>28.4</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>o</td>
<td>32.7 - 39.6</td>
<td>28.0</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>o</td>
<td>32.6 - 38.5</td>
<td>27.3</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>o</td>
<td>34.1</td>
<td>27.6</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>34.9</td>
<td>o</td>
<td>26.5</td>
<td></td>
</tr>
<tr>
<td>GFP(-21)PN21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>o</td>
<td>33.1</td>
<td>29.8</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>o</td>
<td>33.2</td>
<td>29.8</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>o</td>
<td>32.9</td>
<td>29.3</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>o</td>
<td>32.9</td>
<td>29.1</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>o</td>
<td>33.0</td>
<td>28.7</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>o</td>
<td>33.0</td>
<td>28.8</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>o</td>
<td>33.3</td>
<td>28.8</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>o</td>
<td>33.4</td>
<td>27.2</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>o</td>
<td>o</td>
<td>25.4</td>
<td></td>
</tr>
<tr>
<td>GFP(0)PN30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>o</td>
<td>o</td>
<td>32.8</td>
<td>31.7</td>
</tr>
<tr>
<td>25</td>
<td>o</td>
<td>o</td>
<td>32.9</td>
<td>31.3</td>
</tr>
<tr>
<td>27</td>
<td>o</td>
<td>o</td>
<td>33.2</td>
<td>31.3</td>
</tr>
<tr>
<td>30</td>
<td>o</td>
<td>o</td>
<td>32.7</td>
<td>30.8</td>
</tr>
<tr>
<td>33</td>
<td>o</td>
<td>o</td>
<td>32.8</td>
<td>30.2</td>
</tr>
<tr>
<td>37</td>
<td>32.8</td>
<td>32.8</td>
<td>29.8</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>33.0</td>
<td>33.0</td>
<td>29.5</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>35.8</td>
<td>35.8</td>
<td>28.9</td>
<td></td>
</tr>
<tr>
<td>GFP(-8)PN30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>o</td>
<td>o</td>
<td>30.4</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>o</td>
<td>o</td>
<td>30.1</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>o</td>
<td>o</td>
<td>29.1</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>o</td>
<td>o</td>
<td>27.7</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>o</td>
<td>o</td>
<td>25.7</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>30.4</td>
<td>o</td>
<td>27.0</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>o</td>
<td>o</td>
<td>26.5</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>o</td>
<td>o</td>
<td>26.3</td>
<td></td>
</tr>
<tr>
<td>GFP(-21)PN30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>o</td>
<td>32.6</td>
<td>31.7</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>o</td>
<td>33.2 - 36.6</td>
<td>31.1</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>o</td>
<td>33.2 - 36.3</td>
<td>30.3</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>o</td>
<td>o</td>
<td>29.7</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>o</td>
<td>o</td>
<td>28.7</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>o</td>
<td>o</td>
<td>27.3</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>o</td>
<td>o</td>
<td>26.4</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>o</td>
<td>o</td>
<td>25.8</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>A temperature listed for a transition in DPLS denotes that the sample is initially birefringent at 10 °C and loses birefringence at the listed temperature. <sup>b</sup> denotes samples that never display birefringent behavior within the studied temperature range; <sup>b</sup> denotes samples that remain birefringent throughout the entire studied temperature range. <sup>b</sup>The symbol - signifies that no thermal transition is observed.
Differences in the phase behavior among the GFP variants at high concentration and high temperature are also observed, where the deswelling of the PNIPAM domain results in an effective reduction in polymer coil fraction. At $\phi_{\text{PNIPAM}} = 0.52$, an order-order transition (OOT) from lamellar to hexagonal is observed at 45 and 50 wt.% at 30 °C for both GFP(0) and GFP(-8). GFP(-21) has a smaller region of hexagonal morphologies, with the transition occurring at 50 wt.%. At the larger coil fraction of $\phi_{\text{PNIPAM}} = 0.60$, GFP(0) has an OOT from lamellar to hexagonal at 40 and 50 wt.% from 30 °C to 40 °C. GFP(-8) similarly transitions from lamellar to hexagonal at 30 °C between 40 and 50 wt.%, but between 40 and 43 wt.%, the hexagonal phase further transitions to a disordered micellar phase at 35 °C. The conjugate of GFP(-21) transitions from lamellar to a disordered phase above 35 °C at 43 and 47 wt.%, but at higher concentrations of 50 and 55 wt.%, only lamellar morphologies are observed throughout the entire temperature range.

The superfolder GFP and EGFP share a high identity (96%) in their sequence alignment, and previously studied EGFP-PNIPAM at a coil fraction of $\phi_{\text{PNIPAM}} = 0.55$ shows good agreement with the phase behavior of GFP(-8)-PNIPAM in this work with small shifts in transition lines due to differences in coil fraction. Similar to GFP(-8)-PNIPAM, EGFP-PNIPAM undergoes a lyotropic transition from a disordered to a lamellar morphology, forms a narrow two-phase region and formed homogeneous disordered micelles at high temperatures, and undergoes an OOT from lamellar to hexagonal at high concentration and temperature. However, the hexagonal window for EGFP-PNIPAM at $\phi_{\text{PNIPAM}} = 0.55$ is smaller than those for GFP(-8) at the studied coil fractions of 0.52 and 0.60. Previous comparative studies of EGFP, mCherry, and mCherry variants with single amino acid substitutions demonstrated very similar phase behavior, suggesting that large changes in total charge are required to see a noticeable effect.
Despite undergoing a similar lyotropic transition from a disordered to a lamellar phase, the different supercharged GFP variants show significant differences in the quality of the lamellar morphology at both coil fractions studied. The quality of the ordering can be assessed by the full width at half maximum (FWHM) of the primary scattering peak. Figures 5-2a and 5-2c show a comparison of the FWHM of the primary peaks for birefringent lamellar phases formed by GFP(0), GFP(-8), and GFP(-21) at $\phi_{\text{PNIPAM}} = 0.52$ and 0.60, respectively. The background of the SAXS data is fit using the functional form $Aq^{-4} + Bq^{-2} + C$ and then subtracted from the SAXS data, after which a three-parameter Lorentzian function is used to fit the primary peak: $f(q, q_0, \gamma, I_0) = I_0 \left[\frac{\gamma^2}{(q-q_0)^2+\gamma^2}\right]$. Here, $q_0$ refers to the position of the primary peak, $I_0$ denotes the height of the peak, and $2\gamma$ gives the value of the FWHM. As Figure 5-2a shows, for $\phi_{\text{PNIPAM}} = 0.52$, while the FWHM values of lamellar morphologies of GFP(0) and GFP(-8) at 45 wt.% are similar (0.028 and 0.022 nm$^{-1}$, respectively), the FWHM of GFP(-21) at 50 wt.% is greater by approximately a factor of two, 0.048 nm$^{-1}$. Figure 5-2c shows similar results for $\phi_{\text{PNIPAM}} = 0.60$, where the FWHM values of lamellar morphologies of GFP(0) at 37 wt.% and GFP(-8) at 47 wt.% are 0.018 and 0.024 nm$^{-1}$, respectively, and the FWHM of GFP(-21) at 50 wt.% is again greater by approximately a factor of two, 0.043 nm$^{-1}$. These results clearly indicate a preference for better ordered phases in the bioconjugates with the net neutral GFP with its charge distribution and stronger protein dipole moment.

The weak tendency toward microphase separation and the decreased quality of ordering in the supercharged GFP(-21) conjugates is observed across the entire concentration range explored. Figures 5-2b and 5-2d show the evolution of the FWHM of the primary peak as a function of concentration at 10 °C for both coil fractions. The concentrations plotted along the x-axis have been normalized to the $C_{\text{ODT}}$. Even at low concentrations within the disordered
regime of the bioconjugates, the sharpness of the single peak of the solutions follows the trend
GFP(0) > GFP(-8) > GFP(-21), indicating stronger concentration inhomogeneity in the samples
with near neutral charge. While Figures 5-2a and 5-2c indeed show better ordering of lamellar
nanodomains formed by GFP(0) and GFP(-8) than that of GFP(-21) at high concentrations
within the well-ordered regime, the eventual convergence of the FWHM of the primary peaks of
GFP(-21) conjugates to those of GFP(0) and GFP(-8) suggests that at sufficiently high
concentration, molecular crowding leads to contributions from short-range excluded volume and
van der Waals interactions and also electrostatic interactions that govern self-assembly.51 A
similar decrease in protein-protein electrostatic repulsion owing to increasing molecular
crowding effects has been observed with only a moderate increase in concentration beyond the
dilute regime (< 40 mg/mL) for solutions of α-chymotrypsinogen, with a more pronounced
decrease in protein-protein repulsion at low salt concentrations.52
Figure 5-2. (a) SAXS data for birefringent lamellar morphologies of GFP(0), GFP(-8), and GFP(-21) conjugates at \( \phi_{PNIPAM} = 0.52 \). Fitting a three-parameter Lorentzian function to the primary peak shows that GFP(0) and GFP(-8) form more well-ordered nanostructures than the supercharged GFP(-21), denoted by smaller FWHM values of their primary peaks. (b) Evolution of the FWHM of the primary peak of GFP(0), GFP(-8), and GFP(-21) as a function of concentration at 10 °C normalized to the C_{ODT}. Results for \( \phi_{PNIPAM} = 0.60 \) are shown in (c) and (d).

Despite the large range of formal net charge among these three GFP variants, the total number of charged residues among these three proteins is approximately the same—62 for GFP(0) (31 positive and 31 negative), 60 for GFP(-8) (26 positive and 34 negative), and 65 for GFP(-21) (22 positive and 43 negative). Therefore, it is likely that the greater electrostatic repulsive interactions of the supercharged GFP dominate any effects of changing total charge.
density in the system. This leads to the suppression of nanostructure formation with increasing net charge. The fact that nanostructured morphologies of GFP(0) have the best ordering among those formed by the different GFP variants also suggests that the electrostatic complementarity and dipolar interactions are important coarse-grained properties for understanding self-assembly of globular protein-polymer block copolymers. As illustrated in Scheme 1, the Janus-like electrostatic surface potential results in a dipole moment with the greatest magnitude among the three GFP variants, promoting favorable directional interactions that can help enhance self-assembly. McCoy et al. have quantified both the charge complementarity and electrostatic surface complementarity to investigate protein-protein interfaces, finding that the electrostatic complementarity is a more appropriate and better predictor of intermolecular interactions.\(^{53}\)

**Protein Interactions**

Algorithms developed to compute protein-protein interactions for solving protein docking configurations are applied to the series of GFP variants, providing insight into the most favorable orientations for packing of the GFPs as a function of charge and how this packing might influence the propensity to self-assemble into nanostructured phases. The protein docking server ClusPro 2.0 by Vajda and coworkers\(^{54,55}\) uses PIPER, an FFT-based docking program that calculates the pairwise interaction potential

\[ E = E_{\text{attr}} + w_1 E_{\text{rep}} + w_2 E_{\text{elec}} + w_3 E_{\text{pair}}, \]

where \(E_{\text{attr}}\) and \(E_{\text{rep}}\) represent attractive and repulsive contributions to the van der Waals interaction energy, respectively (shape complementarity), \(E_{\text{elec}}\) represents the electrostatic energy, and \(E_{\text{pair}}\) represents the desolvation energy. The coefficients \(w_1, w_2,\) and \(w_3\) represent the weights of the corresponding terms that are optimally set for different types of docking problems.\(^{54}\) Without a priori knowledge of which interactions are most important, the balanced model is used to enable...
calculations in which all of the interaction terms in the pairwise interaction potential are considered. By rotating and translating one protein relative to another fixed protein, approximately $10^9$ configurations are sampled, from which the 1000 most energetically favorable configurations are selected and grouped using a root mean square deviation (RMSD)-based clustering algorithm. Among the 1000 most favorable configurations from the $10^9$ configurations sampled returned by ClusPro, the scoring function using the pairwise potential cannot meaningfully distinguish between the 1000 configurations, and instead the cluster size is a more reliable measure of model prediction. Figure 5-3 shows cartoon representations of the top 5 clusters (groups of configurations that are similar to within a root-mean-squared deviation radius of 9 Å) of the most favorable configurations, ranked by size, for GFP(0), GFP(-8), and GFP(-21).

The most favorable configurations show qualitatively that GFP(0) and GFP(-8) are better able to pack laterally in a lamellar morphology than the supercharged GFP(-21) (Figure 5-3). The principal axis that is aligned with the beta barrel of the GFP variants (represented as black arrows in Figure 5-3) is used to quantify the relative orientation between the two proteins in each interacting configuration. While most of the configurations have orientation angles that do not correspond to nearly parallel alignment of the beta barrels, most of the configurations for GFP(0) and GFP(-8) results from interactions along the side of the beta barrel that could facilitate lateral packing. Four of these five configurations for GFP(-21), on the other hand, favor interactions between the N-termini in a head-to-head manner (which is obstructed by the polymer conjugation) or between the beta barrel end of one protein and the side of the other. The Supporting Information (Figure D-18, Table D-1) includes additional cartoon representations of the clusters ranked 6 through 10 for the different GFPs, as well as tabulated data of the cluster
sizes and their energetics as represented by the scoring function. The top 10 clusters for GFP(0) and GFP(-8) shown in Figure D-18 are highly similar to each other, differing in ranking by the number of members in each group. While the scoring function is not able to distinguish reliably between the top 1000 configurations for each computation, it is still worthwhile to compare the energies of the configurations for the three different GFPs. Calculating a weighted average of the score of the “center” configuration (Table D-1), weighted by cluster size, shows that the weighted average scores of GFP(0) and GFP(-8) (-530.8 and -553.5, respectively) are greater in magnitude than that of GFP(-21), reflecting that the supercharged GFP has less favorable interactions. The center of a cluster is the configuration with the greatest number of neighbors within a root-mean-squared deviation radius of 9 Å.

Figure 5-3. Cartoon schematic representations of results from ClusPro 2.054,55 of the five largest clusters representing the most probable configurations for interactions between two GFP molecules as a function of total net charge. The orientation angle between proteins is displayed underneath each configuration and is calculated as the angle between the vectors in the directions of the principal axes aligned with the beta barrel of the two proteins, represented as the black arrows.
Modulating Protein Interactions with Divalent Cations

Divalent cations are well-known to form ionic crosslinks between polyanions, providing a method to increase attractive interactions between charged polymers\cite{57,58} that may be particularly effective for highly supercharged proteins in counterbalancing the repulsive effects of high net charge. When Ca$^{2+}$ and Mg$^{2+}$ are added to GFP(-21) in the form of CaCl$_2$ and MgCl$_2$, at 20 or 50 mM, there are minor improvements in the quality of the ordering at the CODT of 47 wt.% without a change in the nanostructure of the self-assembled material. Quantifying the quality of order by fitting a Lorentzian to the primary peaks (Figure 5-4) shows that at a concentration of 20 mM salt, the FWHM only decreases from 0.037 nm$^{-1}$ in H$_2$O to 0.027 and 0.024 nm$^{-1}$ with the addition of Ca$^{2+}$ and Mg$^{2+}$, respectively. At 50 mM salt concentration, the values of the FWHM of the primary peak in the presence of Ca$^{2+}$ and Mg$^{2+}$ are 0.030 and 0.039 nm$^{-1}$, respectively. These values are greater than those at 20 mM, suggesting that at 50 mM, screening dominates the ability of the ions to promote more attractive interactions. As a control, NaCl at 20 and 50 mM (Figure D-19) is added to GFP(-21) and is seen to decrease the quality of ordering, having FWHM values of 0.041 and 0.047, respectively. This has been observed in the previously studied system of the red fluorescent globular protein mCherry conjugated to PNIPAM, in which the addition of 200 mM NaCl increased the CODT by approximately 5 wt.%\cite{35}.

The relatively weak role of counterion valency in tuning interactions even between highly charged proteins can be rationalized by estimating the number of counterions present in the protein solution even after concentration from milli-Q water. Assuming conservatively that GFP(-21) maintains charge neutrality with one monovalent counterion per charged residue, a 47 wt.% solution of GFP(-21)-PNIPAM, $\phi_{\text{PNIPAM}} = 0.60$, can have a maximum counterion concentration on the order of 1000 mM. While the true counterion concentration is likely to be
somewhat lower due to charge pairing within the protein, it is nevertheless difficult to achieve charge clustering in such a highly screened environment.

The addition of salt is observed to increase the domain spacing of the lamellar phase relative to the domain spacing in water, reflecting changes in the thermodynamics of the protein-polymer conjugates due to the presence of salt. The domain spacing in water is 26.4 nm. Upon adding NaCl to 20 and 50 mM, the domain spacing increases to 27.1 and 27.4 nm, respectively. At concentrations of 20 and 50 mM CaCl₂, the domain spacing increases to 28.0 and 27.6 nm, respectively; MgCl₂ increases the domain spacing to 27.6 nm at 20 mM and 27.5 nm at 50 mM. The increase in domain spacing in the presence of salt is attributed to the preferential solvation of the ions by the protein, increasing the effective repulsion between the protein and polymer domains, and, subsequently, the domain spacing. For the divalent cation salts CaCl₂ and MgCl₂, preferential solvation of the cations by the protein is observed at 20 mM, resulting in an improvement in the ordering quality of the lamellar phase discussed previously and an increase in domain spacing. Increasing the divalent cation salt concentration to 50 mM leads to a slight decrease in domain spacing relative to that at 20 mM for both CaCl₂ and MgCl₂.
Figure 5-4. SAXS patterns of GFP(-21), $\phi_{\text{PNIPAM}} = 0.60$, at 47 wt.% show lamellar morphologies in water and in the presence of NaCl, CaCl$_2$, and MgCl$_2$ at (a) 20 mM and (b) 50 mM. FWHM values of the primary peak from fitting a Lorentzian show that Ca$^{2+}$ and Mg$^{2+}$ show an improvement in the quality of ordering at 20 mM; the FWHM values increase at 50 mM, suggesting that electrostatic screening dominates at this concentration. NaCl is not observed to improve the ordering relative to the self-assembly in H$_2$O.

5.5 Conclusions

Using the superfolder GFP with a total formal net charge of -8 as the reference protein, solvent-accessible amino acids were chosen to synthesize a GFP variant with a neutral net charge and another variant with a total net charge of -21. These three GFP variants were conjugated to two PNIPAM molecular weights at coil fractions of $\phi_{\text{PNIPAM}} = 0.52$ and 0.60, providing a set of protein-polymer block copolymers in which the effects of the large perturbations in the protein electrostatic interactions on the self-assembly of these block copolymers may be investigated without changes to the protein tertiary structure.
The formation of long-range ordered nanostructures in the supercharged GFP variant is suppressed relative to conjugates to GFP(0) and GFP(-8), as shown by the larger disordered region in phase space of the supercharged GFP compared to the other two GFP variants. The stronger repulsive electrostatic interactions of GFP(-21) lead to the formation of more poorly ordered nanostructures than those formed by GFP(0) and GFP(-8). On the other hand, enhanced nanostructure formation is observed in GFP(0) conjugates, owing to the magnitude of its dipole moment as well as the self-complementarity of its Janus-like electrostatic surface potential. The divalent cations Ca$^{2+}$ and Mg$^{2+}$ are observed to have a small improvement on the quality of the self-assembled nanostructures of GFP(-21) at $\phi_{\text{PNIPAM}} = 0.60$, as the preferential solvation of the cations by the supercharged GFP promotes both protein-protein interactions and protein-polymer repulsion. Differences in formal net charge and surface charge distribution are observed to affect the micellar stability and macrophase separation behavior of these GFP-PNIPAM conjugates. These results show that electrostatic interactions may be used to tune the morphologies of protein-polymer block copolymers, enhancing nanostructure formation through electrostatic surface patches that may promote protein-protein interactions in complementary packing geometries/orientations and suppressing ordered nanostructures due to repulsive protein-protein interactions.

5.6 References


Chapter 6. The Shape of Protein-Polymer Conjugates in Dilute Solution

6.1 Abstract

Protein-polymer conjugation can significantly affect many different aspects of protein behavior, ranging from their solution properties to their ability to form solution and bulk nanostructured materials. An underlying fundamental question is how the molecular design affects the shape of the conjugate and, consequently, its properties. This work measures the molecular configuration of model protein-polymer conjugates in dilute solution using small-angle neutron scattering (SANS) and uses quantitative model fitting to understand the shape of the molecules. Form factor measurements of four model bioconjugates of the red fluorescent protein mCherry and the polymers poly(N-isopropylacrylamide) (PNIPAM), poly(hydroxypropyl acrylate) (PHPA), poly(oligoethylene glycol acrylate) (POEGA), and poly(ethylene glycol) (PEG) show that these protein-polymer conjugates are well described by a recently developed scattering function for colloid-polymer conjugates that explicitly incorporates excluded volume interactions in the polymer configuration. In the regime where the protein does not exhibit strong interactions with the polymer, modelling the protein-polymer interactions using a purely repulsive Weeks-Chandler-Andersen potential also leads to a coarse-grained depiction of the conjugate that agrees well with its scattering behavior. The coarse-grained model can additionally be used for systems with varying protein-polymer interactions, ranging from purely repulsive to strongly attractive, which may be useful for conjugates with strong electrostatic or hydrophobic attractive interactions.

6.2 Introduction

Enzymes and globular proteins are unique polymers with specific functions and desirable properties, including high catalytic activity, large substrate binding affinity, and substrate
specificity, that are difficult to achieve synthetically. Owing to their functional diversity, enzymes and globular proteins are promising candidates for myriad applications to industrial challenges. For example, glucose sensors\cite{1,2} have been developed for medical diagnostics, and sensors to detect and detoxify neurotoxins have been demonstrated to be effective measures for countering against threats of chemical and biological warfare agents.\cite{3-5} Enzymes have been used to synthesize drug compounds in the pharmaceutical industry.\cite{6} They have also been incorporated into devices for applications in bioelectronics and energy.\cite{7-14} The native function of proteins has been augmented by combining proteins with synthetic polymers, enabling these hybrid materials to be used as therapeutics or agents for drug delivery.\cite{15,16}

Protein design within a materials context is important to expand the scope of technological applications and overcome the physical limitations of proteins that make it challenging for them to be utilized effectively under conditions that differ from their natural environment. Synthetic polymers are chemically diverse, durable, and have mechanical integrity and programmable self-assembly behavior that, when combined with the functional diversity of proteins, can realize properties that neither material can individually. The effect of PEGylation (the modification of proteins with poly(ethylene glycol) (PEG)) on the biological properties of protein-containing block copolymers, has been extensively studied. The largest application of PEGylated proteins is as therapeutic agents.\cite{15} Conjugation to PEG has been demonstrated to increase the circulation half-life of proteins in vivo\cite{17-19} and to increase their stability and immunogenicity.\cite{20-22} The protective efficacy of the polymer shield has been shown to depend upon the number of PEG modifications, the polymer size, and polymer architecture.\cite{18} Applications in affinity separations,\cite{23-26} enzymatic reactions,\cite{27-29} and biosensors\cite{30} of protein-polymer conjugates have also been demonstrated by conjugation of stimuli-responsive polymers
such as PNIPAM to proteins, allowing the conjugates to inherit the “smart” attributes of the polymers.

The polymer block can also guide the self-assembly of protein-polymer conjugates into nanostructured materials in solution and in solid materials. The dilute solution aggregation behavior of giant amphiphiles comprising a hydrophilic protein head conjugated to an apolar polymer tail has been shown to form a wide variety of different structures. Nolte and coworkers have investigated various enzyme-polymer giant amphiphiles and characterized their assembly into vesicles, micelles, and μm-long fibers resulting from aggregation of micellar rods. Dilute solution self-assembled nanostructures enable protein encapsulation for drug delivery. Fully biosynthetic protein fusion systems using thermoresponsive elastin-like polypeptides (ELPs) as the coil-like domain also form micellar platforms for multivalent display of proteins. Recently, the ability to form solid-state nanostructured plastics and gels from globular protein-polymer conjugates has also been demonstrated using the globular proteins mCherry, enhanced green fluorescent protein (EGFP), and catalytically active myoglobin.

For application both as pharmaceutically active compounds and as building blocks for self-assembled nanostructures, it is essential to understand the molecular shape of protein-polymer conjugates. In particular, the therapeutic efficacy of a protein and the type of self-assembled structures formed will be affected by whether the polymer molecule wraps around the protein (shroud confirmation) or is repelled, forming a dumbbell structure. A small number of studies have examined the shape of the polymer chain in protein- or peptide-PEG conjugates. Lu et al. used small-angle X-ray scattering (SAXS) to study the shape of PEGylated antibody fragments, showing that the PEG takes a more extended conformation relative to that of the unconjugated antibody fragments. The structure of coiled-coil peptide-PEG conjugates has
also been modeled by assuming that the polymer follows Gaussian statistics with its center of mass located at a specified distance from the peptide. The PEG chain radius of gyration $R_G$ was found to be very similar to that of PEG in free solution; however, slight compression of the polymer chain was observed when PEG was conjugated to the side of the peptide compared to the end. SANS contrast variation has also allowed the scattering contribution from the PEG in mono-PEGylated lysozyme and human growth hormone to be isolated; fitting of the polymer form factor to the Debye model suggests a dumbbell conformation in which the polymer exists as a random coil adjacent to the protein. However, these previous studies do not account for important nonidealities in the polymer structure: they neglect the potential for interactions between the protein and polymer, the effect of polymer solvation on chain conformation, and the effect of the polymer type on the bioconjugate shape.

Herein, the molecular configuration and intramolecular interactions of four protein-polymer conjugates in dilute aqueous solution are explored. The structures of the globular protein-polymer conjugates comprising the red fluorescent protein mCherry conjugated to poly(N-isopropylacrylamide) (PNIPAM), poly(hydroxypropyl acrylate) (PHPA), poly(oligoethylene glycol acrylate) (POEGA), and poly(ethylene glycol) (PEG) in a good solvent for both protein and polymer are measured by small-angle neutron scattering (SANS). The scattering function recently published by Li et al. for protein-polymer conjugates that explicitly incorporates excluded volume effects within the polymer chain is applied to describe the global conjugate conformation in terms of the microstructure of its constituent components, and the data for all four polymer conjugates are compared to this model without the need for contrast matching to isolate the polymer contribution. Analyses using coarse-grained molecular dynamics (MD) simulations further enable the incorporation of protein-polymer interactions into
the form factor fitting, providing insight into the relevant coarse-grained interactions governing the structure of protein-polymer conjugates with excluded volume interactions.

6.3 Experimental Methods

Synthesis and Characterization of mCherry-b-Polymer Block Copolymers

The chemical structures of the polymers used in this study are illustrated in Scheme 6-1. Syntheses of poly(N-isopropylacrylamide) (PNIPAM), poly(hydroxypropyl acrylate) (HPHA), and poly(oligoethylene glycol acrylate) (POEGA) were performed by reversible addition-fragmentation chain-transfer (RAFT) polymerization using a maleimide-functionalized chain-transfer agent, as reported previously. Poly(ethylene glycol) (PEG) with a single maleimide functional group was purchased from Creative PEGWorks (product number PSB-230, mPEG-MAL, MW 30k). The molar masses and dispersities of the polymers were analyzed by gel permeation chromatography using an Agilent Technologies 1260 Infinity system using two ResiPore, 7.5 x 300 mm columns (Agilent Technologies), each with a molecular weight range up to 500,000 g/mol, in N,N-dimethylformamide (DMF) with 0.02 M lithium bromide (LiBr) as the mobile phase (Figure E-1). Signals were collected from a Wyatt DAWN HELEOS II multi-angle light scattering detector and a Wyatt Optilab T-rEX refractometer.

Protein Expression and Bioconjugation

mCherryS131C protein expression, purification, and thiol-maleimide coupling reactions were carried out as previously described. mCherryS131C was further purified using anion exchange fast protein liquid chromatography (FPLC) with an Äkta Pure 25, using 20 mM Tris-Cl, pH = 8.0 as the binding buffer and eluting with NaCl, to improve the purity of the protein for bioconjugation. Briefly, a 6-fold molar excess of polymer was mixed with mCherryS131C in 20 mM Tris buffer at pH = 8.0 after pre-incubation of the protein with 10-fold molar excess of
tris(2-carboxyethyl)phosphine hydrochloride) (TCEP HCl) for 1 hour to reduce disulfide bonds. Reactions with PNIPAM and PEG were carried out for 1 day; reactions with PHPA and POEGA were carried out for 2 days at 4°C. For mCherry-b-PNIPAM, -PHPA, and -POEGA, the reaction mixtures were precipitated in 1 M ammonium sulfate at pH = 8.0 three times to remove unreacted protein. The resuspended mixture was purified by nickel affinity chromatography to remove unreacted polymers and dialyzed into Milli-Q water. For mCherry-b-PEG, the pure bioconjugate was separated from excess polymer and protein by anion exchange chromatography using 10 mM Tris-Cl pH = 7.0 as the binding buffer and eluting with NaCl. SDS-PAGE and Native PAGE confirm the purity of the final products (Figure E-2). UV-Vis spectroscopy demonstrates that the optical activity of the proteins is preserved after bioconjugation (Figure E-3). Scheme 6-1 shows the chemical structures of the protein-polymer conjugates used in this study, and Table 1 summarizes their compositions.

![Scheme 6-1](image-url)

**Scheme 6-1.** Chemical structure of (a) mCherry-b-PNIPAM, (b) mCherry-b-PHPA, (c) mCherry-b-POEGA, and (d) mCherry-b-PEG. (e) Coarse-grained cylinder-coil bead representation of mCherry-polymer conjugates.
Table 1. Molecular properties of protein-polymer conjugates.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Copolymer</th>
<th>Polymer $M_n$ (kg mol$^{-1}$)</th>
<th>Dispersity (D)</th>
<th>Bioconjugate $M_n$ (kg mol$^{-1}$)</th>
<th>$\phi_{\text{polymer}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mChPN26</td>
<td>mCherry-$b$-PNIPAM</td>
<td>26.3</td>
<td>1.08</td>
<td>54.4</td>
<td>0.55</td>
</tr>
<tr>
<td>mChPH30</td>
<td>mCherry-$b$-PHPA</td>
<td>29.9</td>
<td>1.21</td>
<td>58.0</td>
<td>0.55</td>
</tr>
<tr>
<td>mChPOE26</td>
<td>mCherry-$b$-POEGA</td>
<td>26.4</td>
<td>1.13</td>
<td>54.5</td>
<td>0.55</td>
</tr>
<tr>
<td>mChPEG28</td>
<td>mCherry-$b$-PEG</td>
<td>28.0</td>
<td>1.04</td>
<td>56.0</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Characterization

Small-angle neutron scattering (SANS) was performed at the ORNL HFIR General Purpose Small-Angle Neutron Scattering (GP-SANS) Diffractometer and the NIST NCNR NG-7 SANS instrument. Dilute solutions were measured to avoid the possible coherent scattering contribution from intermolecular spatial correlations, and solutions were filtered using 0.1 μm Whatman Anotop 10 syringe filters. Protein and bioconjugate concentration measurements by absorbance at 280 nm were performed after sample filtration to quantify the exact concentration measured by SANS. Contrast matching of mCherry was performed at the GP-SANS beamline instrument using a wavelength of 4.72 Å at two sample-to-detector distances of 0.3 and 8m, covering a Q-range of 0.00658 Å$^{-1}$ to 0.933 Å$^{-1}$. A 1 m$^2$ area detector with an array of 192 1 m long $^3$He Linear Position-Sensitive Detectors with 5.2 mm x 4 mm pixel resolution was used. Solutions of mCherry without the S131C mutation at 10 mg/mL and $T = 10 \degree C$ were measured in $D_2O/H_2O$ blends at compositions of 100, 90, 80, 70, and 10% $D_2O$ using Banjo cells with a 1 mm path length. The contrast match point and the percentage of exchangeable protons were determined experimentally (SI). Dilute solution form factor measurements of mCherry, the polymers PNIPAM; PHPA; POEGA; and PEG, and the mCherry-polymer bioconjugates were performed at the NG-7 SANS instrument using a wavelength of 6 Å with wavelength spreads of 15% and three sample-to-detector distances of 1, 4, and 13 m, covering a Q-range of 0.00348 Å$^{-1}$.
to 0.518 Å⁻¹. Solution samples were loaded into titanium Hellma cells with quartz windows and a path length of 1 mm. The scattered neutrons were detected by a 640 mm x 640 mm ³He position-sensitive proportional counter with a 5.08 mm x 5.08 mm resolution. Solutions of the homopolymers were prepared in D₂O at 8 mg/mL using filtered D₂O; the solutions were not filtered directly to avoid precipitation during filtration. The mCherry solution was prepared at 8 mg/mL in 10 mM Tris-Cl, pD = 8.0 and conjugate solutions were prepared at a protein concentration of 8 mg/mL in D₂O (determined from the protein and bioconjugate absorption at 280 nm, ε₂₈₀ = 34,380 M⁻¹ cm⁻¹). The scattering patterns of mCherry, the homopolymers, and the bioconjugates were measured at 5 °C. The measured scattering intensity was corrected for detector sensitivity and the background from the empty cell and placed on an absolute scale using a calibrated standard. DLS measurements were performed at a concentration of 1 mg/mL at T = 5 °C in H₂O using a Wyatt DynaPro NanoStar with a laser wavelength of 658 nm.

Model and Simulation Methodology

Molecular dynamics was used to determine the form factor of a coarse-grained representation of mCherry-polymer conjugates with excluded volume interactions and to fit to experimental neutron scattering data. A coarse-grained model based upon the Kremer-Grest model was used to model polymer coils of N monomers of mass mcoil attached to a 28-bead cylinder (4 layers of 7 beads packed into a hexagonal geometry), each of mass mcyₐ representing the β-barrel shape of the red fluorescent globular protein mCherry (Scheme 6-1).

Since mCherry has approximate dimensions of 4.4 nm in length and 2.5 nm in diameter based upon the crystal structure (PDB 2H5Q), this bead model closely resembles the coarse-grained shape of the protein. Nearest-neighbor beads are connected by finite-extensible nonlinear springs with the potential...
where \( l \) refers to either cylinder or coil beads, \( r \) is the separation between monomers, \( R_0 \) is a cutoff distance, and \( k \) is the interaction strength. The protein and polymers are well solvated in water under SANS measurement conditions, so cylinder-cylinder bead and coil-coil bead excluded volume interactions are modeled as solvophilic and purely repulsive using the Weeks-Chandler-Andersen (WCA) potential,

\[
U_{\text{wca}, l} = \begin{cases} 
4 \varepsilon_k \left( \frac{\sigma_{l}}{r} \right)^{12} - \left( \frac{\sigma_{l}}{r} \right)^{6} + \frac{1}{4} \varepsilon_k, & r \leq 2^{1/6} \sigma_{l}, \\
0, & r > 2^{1/6} \sigma_{l},
\end{cases}
\]

where the subscript \( k \) in the interaction term refers to the three different possible pairwise interactions—cylinder-cylinder beads, coil-coil beads, and cylinder-coil beads—and \( l \) refers to either cylinder or coil beads. WCA potentials are also applied for cylinder-coil bead interactions. The cylinder and coil bead sizes are not highly asymmetric, so the average of the two is used as an effective size for evaluating the WCA interaction potential. The parameters can be expressed as \( R_{0,l} = 1.5 \sigma_l \), \( k = 30 \varepsilon/\sigma_l^2 \), and \( k_B T = \varepsilon \), where \( \varepsilon \) and \( \sigma_{cyl} \) are the energy and length scales of the simulation, respectively. The beads are separated by an average distance of \( 0.97 \sigma_l \), as the probability distribution of \( r \) for cylinder and coil beads, \( P_l(r) \), is peaked around its mean of \( \int_0^{\infty} P_l(r) r dr = 0.97 \sigma_l \). Stiff three-bead bending potentials are applied along the length of the cylinder and around each hexagonal bend and are given by Eqn. 3 and Eqn. 4, respectively,

\[
U_{\theta, 3} = 1000 \varepsilon \left( 1 - \cos \theta \right)
\]

\[
U_{\theta, 2} = 1000 \varepsilon \left( 1 - \cos \left( \theta - 60^\circ \right) \right)
\]
where $\theta$ is the angle between the segments connecting adjacent monomer beads, with $\theta = 0$ corresponding to a fully extended configuration. These potentials are used in a molecular dynamics simulation where the equation of motion for each monomer is given by

$$m \frac{d^2 R}{dt^2} = -\nabla_i(U_{wCA} + U_S + U_\theta) - \zeta \frac{dR_i}{dt} + W$$

(5)

$R$ is the position of each monomer, $m$ the mass, and $t$ is time with a characteristic unit of $\tau = \sigma_{cyl}(m_{cyl}/\varepsilon)^{1/2}$. The last two terms correspond to a drag force and a Brownian force that regulates the temperature of the simulation using the Langevin thermostat. $W$ is a random Brownian force that obeys the fluctuation-dissipation relation,

$$\langle W_i(t) \rangle = 0$$

(6)

$$\langle W_i(t)W_j(t') \rangle = 2k_B T \delta_{ij} \delta(t-t')$$

(7)

$I$ is the identity tensor, and the drag coefficient is given by $\zeta = 0.5\tau^{-1}$, maintaining the temperature at $T = \varepsilon/k_B$. Simulations were equilibrated by slow push-off$^{58}$ and integrated using a leapfrog algorithm with a time step $\Delta t = 0.005\tau$, which is within the stability limit of the stiff interaction potentials.

6.4 Results and Discussion

Guinier Analysis

Both SANS and dynamic light scattering show that there is some change in the size of the protein-polymer conjugate in solution as a function of polymer type, even though the molar mass of all the polymers is nearly identical. The hydrodynamic radius, $R_H$, from DLS measurements, and the radius of gyration, $R_G$, determined by model-independent Guinier analysis (Figure E-4), confirm the presence of single conjugate molecules at dilute concentrations (Table 2). From Guinier analysis, $R_G$ for mCherry is calculated to be 1.56 nm, and values of $R_G$ for the polymers
are calculated to be quite similar to one another—4.56, 4.22, 4.24, and 4.39 nm for PNIPAM, PHPA, POEGA, and PEG, respectively. The conjugates are found to have $R_G$ values of 3.46, 4.77, 5.95, and 5.63 nm for mCherry conjugated to PNIPAM, PHPA, POEGA, and PEG, respectively. Three of the conjugates show an increase in radius of gyration relative to mCherry and the corresponding homopolymer; however, both $R_G$ and $R_H$ for mCherry-PNIPAM are found to be somewhat smaller than the values for PNIPAM. It should be noted that for mChPOE26, the low Q upturn in Figure 6-1c signifies aggregation, and, because the Guinier regime is not as well established, this leads to an overestimation of $R_G$ for mChPOE26. Previous SANS contrast variation experiments of mCherry-POEGA have shown that mCherry-POEGA tends to macrophase separate in solvent compositions comprising a majority percentage of D$_2$O, which may lead to a small percentage of stable aggregate species at dilute solution. The shape-factor or asymmetry factor, given by $R_G/R_H$ ($\rho$-ratio), provides additional insight into the molecular shape. $R_G/R_H$ for a globular protein is at times assumed to be equivalent to that of a homogeneous sphere, 0.775.\textsuperscript{59} The ratio for mCherry is 0.54, slightly smaller than that for a globular protein, which could result from the protein having a slightly larger hydrodynamic radius due to the bound water layer. For polymers in solution, the Kirkwood-Riseman theory predicts that $R_H = 0.665R_G$,\textsuperscript{60} and theoretical calculations by Akcasu and Han\textsuperscript{61} predict that $R_H = 0.664R_G$ for theta-solvent conditions and $R_H = 0.537R_G$ for good solvent conditions. The relationship between $R_G$ and $R_H$ for the homopolymers differ from these theoretical calculations, but the ratio of $R_G/R_H$ for PEG is found to be similar to that from a previously reported study.\textsuperscript{50} After conjugation of the polymers to mCherry, the ratios of $R_G/R_H$ for the conjugates are 1.02, 1.26, 1.49, and 0.99 for mCherry-PNIPAM, mCherry-PHPA, mCherry-POEGA, and mCherry-PEG, respectively. For all conjugates, the $R_G/R_H$ ratios are greater than that of mCherry and the corresponding
homopolymers, which is consistent with a molecular structure in which the polymer coil remains in an extended coil configuration instead of wrapping around the protein. For reference, values of $R_G/R_H$ vary from 0.775 for a solid homogeneous sphere to values greater than 2 for extended coils and prolate ellipsoids; $R_G/R_H$ for a hollow sphere is 1.00 and for a thin rod geometry the ratio is approximated by $1.22 \ln(L/D)$, where $L/D$ is the rod aspect ratio. While results comparing the radius of gyration and the ratio $R_G/R_H$ for mCherry, the homopolymers, and the mCherry-polymer conjugates suggest that the polymer adopts a coil configuration in a dumbbell-like structure, the errors associated with these calculations do not preclude the possibility of other structures. Form factor model fitting using both theoretical and computational models must be used to provide additional quantitative information on the conjugate shape in dilute solution.

Table 2. DLS and Guinier analysis results of mCherry-polymer bioconjugates

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$R_H$ (nm)</th>
<th>$R_G$ (nm)$^a$</th>
<th>$R_G/R_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mCherry</td>
<td>2.9 ± 0.9</td>
<td>1.56 ± 0.15</td>
<td>0.54 ± 0.17</td>
</tr>
<tr>
<td>PN26</td>
<td>5.2 ± 1.6</td>
<td>4.56 ± 0.74</td>
<td>0.88 ± 0.32</td>
</tr>
<tr>
<td>PH30</td>
<td>3.8 ± 0.8</td>
<td>4.22 ± 0.50</td>
<td>1.12 ± 0.28</td>
</tr>
<tr>
<td>POE26</td>
<td>4.2 ± 1.6</td>
<td>4.24 ± 0.36</td>
<td>1.02 ± 0.40</td>
</tr>
<tr>
<td>PEG28</td>
<td>6.2 ± 2.0</td>
<td>4.39 ± 0.66</td>
<td>0.70 ± 0.25</td>
</tr>
<tr>
<td>mChPN26</td>
<td>3.4 ± 1.0</td>
<td>3.46 ± 0.32</td>
<td>1.02 ± 0.31</td>
</tr>
<tr>
<td>mChPH30</td>
<td>3.8 ± 0.6</td>
<td>4.77 ± 0.39</td>
<td>1.26 ± 0.22</td>
</tr>
<tr>
<td>mChPOE26</td>
<td>4.0 ± 1.4</td>
<td>5.95 ± 0.38</td>
<td>1.49 ± 0.53</td>
</tr>
<tr>
<td>mChPEG28</td>
<td>5.7 ± 0.7</td>
<td>5.63 ± 0.55</td>
<td>0.99 ± 0.18</td>
</tr>
</tbody>
</table>

$^a$Statistical error estimates correspond to 95% confidence intervals.

Form Factor Analysis: Theoretical Model

The scattering model for colloid-polymer conjugates developed by Li et al. provides a direct relationship between the global conformation of the conjugate and structure of its constituent components, namely through the scattering contributions of the intracolloid
correlation, the intrapolymer correlation, and the colloid-polymer cross correlation to the overall
scattering signature. This model can be tested using scattering measurements for the series of
mCherry-polymer conjugates. First, a dilute solution SANS measurement of mCherry is
independently fit using the model for the form factor of a homogeneous cylinder to model the
protein structure in terms of a simplified geometric object.\textsuperscript{64} The values of the cylinder radius
and length determined from model fitting are 1.51 ± 0.02 nm and 4.73 ± 0.11 nm, respectively,
agreeing reasonably well with the values determined from the protein crystal structure ($R_{mCh} =
1.2$ nm, $L_{mCh} = 4.4$ nm) (Figure E-5). mCherry has been demonstrated to retain its structure and
optical functionality after conjugation,\textsuperscript{39,42,52} and the form factor of mCherry is therefore fixed in
subsequent model fitting calculations of the mCherry-polymer conjugates, yielding a constant
intracolloidal contribution.

Dilute solution SANS measurements of the homopolymers PNIPAM, PHPA, POEGA,
and PEG are also independently fit using the scattering function for a worm-like chain (WLC)
with excluded volume effects, $P_{WL}(Q,L,a_{K},R_{CS})$ (Figure E-6),\textsuperscript{65,66} where $L$ represents the chain
contour length, $a_{K}$ is the Kuhn length, and $R_{CS}$ is the chain cross-sectional radius. From values of
the molecular weights of the homopolymers by GPC and the number of polymer repeat units, the
contour lengths of the polymers are calculated using estimates of bond angles and bond lengths
(SI). Additionally, from the experimental polymer volume fractions of the mCherry-polymer
conjugates and the volume of a cylinder with radius 15.1 Å and length 47.3 Å that best
approximates the form factor of mCherry, the volumes of the polymer chains are estimated,
allowing the cross-sectional radius $R_{CS}$ to be determined. Thus, the SANS form factor
measurements of the homopolymers are fit using the WLC scattering function with only the
Kuhn segment length $a_{K}$, a multiplicative factor that accounts for the concentration and scattering
contrast of the polymer in solution, and the incoherent background as fitting parameters. Having the Kuhn segment length as the single parameter from the WLC model to be used as a fitting parameter enables a direct comparison of the parameter of the polymer when it is isolated in solution compared to when one end is conjugated to the protein. The Kuhn segment lengths of the free polymers are summarized in Table 3; as expected, low temperature water is a good solvent for the polymers, and Kuhn segment lengths are measured ranging from 1.5 to 1.6 nm.

In all cases, the scattering observed for the conjugates shows a clear low $Q$ Guinier region; however, for mCherry-POEGMA, there is an upturn at low $Q$ due to slight aggregation. Good fits of the experimental data with Li’s scattering model suggest that the polymer chain adopts a self-avoiding random coil conformation adjacent to the protein for all polymer chemistries explored. The conformation of the conjugate is given by

$$P_{\text{conjugate}}(Q) = P_{\text{colloid}}(Q) + P_{\text{polymer}}(Q) + P_{\text{colloid-polymer}}(Q)$$  

$$P_{\text{colloid}}(Q) = \left(\Delta \rho_{m\text{Ch}} V_{m\text{Ch}}\right)^2 P_{\text{norm,mCh}}(Q, R_{m\text{Ch}}, L_{m\text{Ch}})$$

$$P_{\text{polymer}}(Q) = \left(\frac{b - L}{a_k}\right)^2 P_{\text{WLC}}(Q, L, a_k, R_{CS})$$

$\Delta \rho_{m\text{Ch}}$ is the scattering length density difference between mCherry and the solvent D$_2$O, $V_{m\text{Ch}}$ the protein volume, and $P_{\text{norm,mCh}}(Q, R_{m\text{Ch}}, L_{m\text{Ch}})$ the normalized protein form factor with radius $R_{m\text{Ch}}$ and length $L_{m\text{Ch}}$ as discussed previously. $L$ is the contour length of the polymer, $a_k$ the Kuhn segment length, $R_{CS}$ the radius of the Kuhn segment cross section, and $b$ is the total bound coherent scattering length of a Kuhn segment, calculated by $b = \Delta \rho_{\text{polymer}}(4\pi R_{CS}^2 a_k)$, where $\Delta \rho_{\text{polymer}}$ is the scattering length density difference between the polymer Kuhn segment and D$_2$O. The cross correlation between the protein and polymer is evaluated numerically.$^{51}$ In the scattering model for colloid-polymer conjugates, the radius and length of the cylinder, as well as
the polymer contour lengths and cross-sectional radii, are fixed, leaving the conjugated polymer Kuhn segment length and the scattering length densities of the protein and polymer Kuhn segment as the only unknown parameters; a multiplicative factor that accounts for the concentration and scattering contrast of the conjugate and the incoherent background are additional parameters, resulting in the fitting equation $P_{\text{conjugate}}(Q) = CT \times P_{\text{conjugate}}(Q) + B$. A good estimate for $\rho_{\text{mCh}}$ for model fitting calculations is determined from a contrast matching experiment to determine the number of exchangeable protons of mCherry (SI). Model fitting results are shown in Figure 6-1 and summarized in Table 3. The results are shown on a log-log axis in Figure E-8.

For all four conjugates, the Kuhn segment length of the polymer in free solution is comparable to that when it is conjugated to mCherry, suggesting that conjugation does not significantly affect the solvation of the polymer chains. The Kuhn segment lengths of PNIPAM and PHPA are observed to decrease slightly and those of POEGA and PEG increase slightly when conjugated to mCherry, respectively. The fitting results also show that the scattering length density of each polymer is greater than the scattering length density of mCherry (Table 3). Surprisingly, the polymer scattering length density values are found to be significantly greater than the values provided by the NIST online scattering length density calculator. The difference is approximately 3.7 times greater for PHPA and POEGA, whereas for PNIPAM and PEG, the difference is approximately a factor of 6 and 7 times greater, respectively.

To understand the origin of this result, the scattering length densities of the polymers are also estimated from Guinier analysis of their dilute solution SANS scattering. The volume of mCherry is estimated as that of the homogeneous cylinder that best fits the mCherry form factor measurement (Figure E5), and then the polymer volume fractions of the mCherry-polymer
conjugates are used to estimate the homopolymer volumes, $V_{\text{polymer}}$. The scattering length density of the polymer is then calculated according to the following equation:

$$I(Q = 0) = \left( \frac{N}{V} \right) \times V_{\text{e, polymer}} \times \left( \rho_{\text{polymer}} - \rho_{D_2O} \right)^2$$  \hspace{1cm} (11)$$

where $(N/V)$ is the number concentration of the polymer solution. Estimated values for the scattering length densities of PHPA and POEGA are found to be similar to those provided by the NIST online calculator, and those of PNIPAM and PEG are about 3 to 4 times greater (Table 3). Details of the calculations are included in the Supporting Information and shown in Table S2. The larger scattering length density values determined from model fitting could result from the packing of D$_2$O molecules within the swollen polymer chains and from D$_2$O molecules that may be associating with the polymers through hydrogen bonding, effects that the scattering model does not incorporate. For example, studies of water-PEG interactions in solution have shown that each ethylene oxide unit is well hydrated, associating with 2-3 water molecules.$^{68}$
Figure 6-1. Nonlinear least-squares regression of cylinder-polymer scattering function (black lines) to experimental mCherry-polymer bioconjugate SANS data (open circles) for (a) mChPN26, (b) mChPHPA30, (c) mChPOE26, and (d) mChPEG28. The dashed vertical lines denote data points at low Q that deviate from 95% confidence interval bounds on the zero-angle scattering value from Guinier analysis by 10% that are omitted in the model fitting. Slight aggregation is observed in mChPOE26, represented by a slight upturn in intensity at low Q.
Table 3. Results from model fitting using a colloid-polymer conjugate scattering function with excluded volume interactions

<table>
<thead>
<tr>
<th>Polymers</th>
<th>$M_c$ (kg mol$^{-1}$)</th>
<th>$L$ (nm)</th>
<th>$\rho_{\text{cyl}}$ (x 10$^4$ Å$^3$)</th>
<th>$\rho_b$ (x 10$^4$ Å$^2$)</th>
<th>$\rho_b$ (Guinier) (x 10$^4$ Å$^2$)</th>
<th>$a_c$ (free polymer) (nm)</th>
<th>$a_c$ (conjugate) (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNIPAM</td>
<td>26.3</td>
<td>67.4</td>
<td>3.16</td>
<td>4.94 ± 0.02</td>
<td>0.777</td>
<td>2.32</td>
<td>1.53 ± 0.05</td>
</tr>
<tr>
<td>PHPA</td>
<td>29.9</td>
<td>66.8</td>
<td>3.16</td>
<td>3.96 ± 0.04</td>
<td>1.07</td>
<td>0.90</td>
<td>1.59 ± 0.04</td>
</tr>
<tr>
<td>POEGA</td>
<td>26.4</td>
<td>50.8</td>
<td>3.16</td>
<td>3.94 ± 0.04</td>
<td>1.05</td>
<td>1.03</td>
<td>1.57 ± 0.03</td>
</tr>
<tr>
<td>PEG</td>
<td>28.0</td>
<td>260.6</td>
<td>3.16</td>
<td>4.79 ± 0.03</td>
<td>0.679</td>
<td>2.08</td>
<td>0.29 ± 0.01</td>
</tr>
</tbody>
</table>

$^a$ Statistical error bounds correspond to 95% confidence intervals. $^b$ Density values for PNIPAM,$^{39}$ PHPA,$^{44}$ and POEGA$^{44}$ are reported previously. The density used for PEG is 1.2 g cm$^{-3}$.67

Form Factor Analysis: Coarse-Grained Model

Coarse-grained molecular dynamics simulations provide the opportunity to additionally introduce variable protein-polymer and polymer-polymer interactions into the form factor model, exploring how a conjugate may transition between “dumbbell” and “shroud” states as a function of these parameters. In this model, the protein is constructed from a rigid body of linked spheres for ease of computation, and the polymer is modeled as a flexible chain of spheres linked to the protein at a defined conjugation site. The relative volume fractions of the blocks and the flexibility of the polymer chain may be controlled to match the experimental system by tuning the size and number of beads in the polymer chain. In order to model conjugate scattering, the protein and homopolymer are first characterized and parameterized separately, and these parameters are then used to model the final conjugate structure.

To model the protein, the form factor of the 28-bead ensemble of homogeneous spheres to approximate the coarse-grained cylindrical shape of the protein mCherry is first fit to SANS form factor data to determine the value of the length scale parameter $\sigma_{\text{cyl}}$. The form factor amplitude of a homogeneous sphere of radius $R$ is given by the following equation.$^{64}$


\[ F_{\text{sphere}}(Q) = \frac{3[\sin(QR) - QR \cos(QR)]}{(QR)^3} \]  

(12)

For a particle comprising \( P \) spherical subunits, the analytical expression for the form factor is given by the Debye formula:\(^{64}\)

\[
P(Q) = \frac{1}{M^2} \sum_{i,j=1}^{P} M_i M_j F_i(Q) F_j(Q) \frac{\sin(Qr_{ij})}{Qr_{ij}}
\]

\[ M = \sum_{i=1}^{P} M_i \]  

(13)

The scattering lengths \( M \) of all spheres in the protein are taken to be equal and are normalized to 1. Results from fitting the coarse-grained cylinder form factor to the protein SANS data using nonlinear least squares regression finds the value of \( \sigma_{\text{cyl}} \) to be 11.23 (Figure 6-2).

**Figure 6-2.** Model fitting results from fitting the form factor of the coarse-grained 28-bead cylinder representation to mCherry SANS form factor data. The bead diameter \( \sigma_{\text{cyl}} \) is determined to be 11.23. Similar to the model fitting described above for fitting to mCherry and polymer SANS data, a contrast factor \( (CT) \) and an incoherent background \( (B) \) are also incorporated into the fitting analysis. The dashed vertical lines denote data points at low \( Q \) that deviate from 95% confidence interval bounds on the zero-angle scattering value from Guinier analysis by 10% that are omitted in the model fitting.

MD simulations of a coarse-grained model as a function of \( N_{\text{coil}} \) monomer beads connected by springs with excluded volume interactions are used to fit the experimental SANS form factors of the four polymers to determine the best-fit number of coarse-grained beads to represent each polymer. The form factor is calculated by applying Eqn. 12 to 1000 independent
simulation configurations (one configuration is sampled every 1000\(r\)) and averaging them. The
time correlation function of the form factors are calculated to ensure that 1000\(r\) exceeds the
correlation time for each simulation. In the nonlinear least squares regression algorithm, when
calculating the form factor of the single coils, the scaling parameter \(\sigma_{\text{coil}}\) is calculated as a ratio
relative to \(\sigma_{\text{cyl}}\) such that the coarse-grained polymer volume fractions agree with those of the
experimental values. Results of the coarse-grained coil model fitting show that 45, 46, 28, and
64 beads give the best fit to the SANS data for PNIPAM, PHPA, POEGA, and PEG, respectively
(Figures E-10 and E-11). Because the form factor fitting was performed such that the bead size
\(\sigma_{\text{coi}l}\) is a function of the number of beads \(N_{\text{coi}l}\) to maintain the correct polymer volume fraction,
the contour lengths of the coarse-grained polymer bead representation do not match the
calculated values shown in Table 3. However, the trend for the optimal number of beads needed
to represent each homopolymer does agree with that of the calculated polymer contour lengths.
The trend in size of each polymer bead also agrees with that of the size of the monomers, with
POEGA having the largest bead size and PEG the smallest.

Coarse-grained cylinder-coil model representations of the mCherry-polymer
bioconjugates are then constructed by attaching the 28-bead cylinder model to coils of length
\(N_{\text{coi}l}\) for each polymer. Average form factors of the four coarse-grained models are calculated
using Eqn. 12 by averaging 1000 independent simulation configurations (one every 1000\(r\), time
correlation functions shown in Figure E-9) to fit to the SANS form factor data of the
corresponding mCherry-polymer bioconjugates. The scattering lengths \(M_i\) represent the
scattering contrast relative to the solvent scattering length, \((\rho_{\text{part}} - \rho_s)V_{\text{part}}\), where \(\rho_{\text{part}}\) refers to the
scattering length density of the particle, and \(V_{\text{part}}\) the volume of the particle; \(\rho_s\) refers to the
scattering length density of D$_2$O, the solvent used in this study. In the fitting procedure, the
scattering length contrast $M_{mCherry}$ is normalized to 1, with the scattering length contrast of the
coil beads used as a fitting parameter along with a multiplicative factor (that accounts for the
conjugate concentration and scattering contrast) and incoherent background. Results in Figure 6-
3 show that the form factors of the coarse-grained representations of the protein-polymer
conjugates agree relatively well using purely repulsive WCA interaction potentials for all
pairwise bead interactions.

Instead of allowing $M_{coil}$ to be a fitting parameter, the ratio $M_{coil}/M_{mCherry}$ can be
estimated, fixing the form factor of the coarse-grained conjugate and allowing the contrast factor
and incoherent background to be the only two fitting parameters. From contrast matching
experiments, the scattering length density of mCherry is determined to be $\rho_{mCh} = 3.16e-6$ Å$^2$
(Supporting Information), from which $M_{cyl} = (\rho_{cyl} - \rho_{D2O})V_{cyl}$ can be determined. The zero-
angle scattering intensity is calculated from Guinier analysis (Figure E-4) and is given by
$I(Q = 0) = \frac{N}{V} (\rho_{part} - \rho_{s})^2 V_{part}^2$, where $N/V$ is the particle number density concentration. The
zero-angle scattering intensity of the homopolymer SANS data is used to estimate $M_{coil} = 
(\rho_{pol} - \rho_{D2O})V_{pol}$. The ratios $M_{coil}/M_{mCherry}$ for the four different polymers PNIPAM, PHPA,
POEGA, and PEG are estimated to be 1.52, 2.10, 2.00, and 1.50, respectively (SI). From the
fitting results shown in Figure 6-3 where $M_{mCherry}$ is normalized to 1, the fit ratios $M_{coil}/M_{mCherry}$
were determined to be 0.61, 1.35, 1.80, and 0.45, respectively. Holding the estimated scattering
lengths constant in the fitting procedure overestimates the ratio of the polymer scattering length
contrast to that of the protein, which is reflected in a qualitative difference in the shape of the of
the coarse-grained form factor compared to the SANS data (Figure E-12), particularly for
mCherry-PNIPAM and mCherry-PEG. The ratios $M_{coil}/M_{mCherry}$ from fitting the coarse-grained
model to the data are smaller than those predicted from Guinier analysis; these results agree
qualitatively with those from fitting the data using Li et al.'s theoretical scattering function, which also finds that the ratio of the polymer scattering length contrast to that of the protein is smaller than the expected value from Guinier analysis. Varying the interaction potential between protein and polymer beads (Figure E-14) causes the polymer beads to enshroud the protein beads more strongly as the interaction length scale increases but does not improve the quality of the fits.

![Graphs showing fit results](image)

**Figure 6-3.** Nonlinear least-squares regression of coarse-grained cylinder-coil form factors (black lines) to experimental mCherry-polymer bioconjugate SANS data (open circles) for (a) mChPN26, (b) mChPHPA30, (c) mChPOE26, and (d) mChPEG28. The dashed vertical lines denote data points at low Q that are omitted in the model fitting, corresponding to the same data omitted in Figure 6-1.

Although all of the experimentally explored conjugates have a similar molecular shape captured by repulsive interactions between protein and polymer, the coarse-grained model is highly sensitive to protein-polymer interactions in the transitional regime between repulsive and
attractive. The sensitivity of the coarse-grained model to colloid-coil bead interactions is explored by attaching 46 coil beads to the 28-bead cylinder model and calculating the form factors as the interaction potential between cylinder-coil beads is varied. In this coarse-grained model, the polymer volume fraction is set to be 0.5, and the ratio of the coil/cylinder bead density is taken to be $1/1.35 = 0.741$ (which represents a polymer density of 1.0 g cm$^{-3}$ relative to the density of mCherry, 1.35 g cm$^{-3}$). The scattering lengths of the cylinder and coil beads are set to be equal to 1 for these calculations. Figure 6-4a shows that the coarse-grained model form factor is insensitive to increases in the repulsive interaction strength, $\varepsilon_{\text{cyl-coil}}$, using a purely repulsive potential. However, when the potential transitions from repulsive to attractive by gradually increasing the cutoff distance in the truncated and shifted Lennard-Jones interaction potential (Figure E-13a) between cylinder-coil beads to transition from the WCA to a traditional Lennard-Jones potential (Figure 6-4b), the form factor shifts to resemble that of a homogeneous sphere, indicating that the coil wraps around the surface of the cylinder (Figure 6-4c). The second virial coefficient $B_2$ between a cylinder-coil bead pair experiencing a truncated and shifted Lennard-Jones interaction potential for this particular 74-bead coarse-grained model is calculated as a function of the cutoff radius. When the cutoff distance is greater than $2^{1/6}\sigma_{\text{cyl-coil}}$, the attractive interaction between cylinder-coil beads dominates, represented by a negative second virial coefficient (Figure E-13b). Increasing the interaction strength $\varepsilon_{\text{cyl-coil}}$ at cutoff distances greater than $2^{1/6}\sigma_{\text{cyl-coil}}$ results in the coil beads wrapping around the cylinder (Figure E-14). Therefore, this coarse-grained approach can capture both dumbbell (where the colloid and polymer exhibit weak or repulsive interactions) and shroud-like configurations (where strong enthalpic interactions exist or the polymer size greatly exceeds that of the colloid), and also those intermediate between these two configurations. This potentially makes this approach to
modeling very general, applying to conjugates where electrostatic or hydrophobic forces between protein and polymer may introduce attractive interactions.

6.5 Conclusions

The form factors of globular protein-polymer conjugates in a good solvent for both protein and polymer are investigated by using a combination of model fitting and coarse-grained MD simulations to understand the structure of four model globular protein-polymer conjugates consisting of the red fluorescent protein mCherry conjugated to four different polymers—PNIPAM, PHPA, POEGA, and PEG—at an approximate volume fraction of 0.5. Model fitting using a scattering function for colloid-polymer conjugates with excluded volume solvent
interactions within the polymer provides evidence that the polymer coil exists as a relatively unperturbed flexible chain attached to the globular protein. A coarse-grained representation of the protein-polymer conjugates is developed, and simulation results show that the conformation of protein-polymer conjugates in a good solvent is captured by a model incorporating purely repulsive interaction potentials. The coarse-grained representation provides a simple way of modeling the conformation of protein-polymer conjugates that may exhibit strong protein-polymer interactions that will be useful for systems with strong hydrophobic and electrostatic interactions between protein and polymer.

6.6 References

16. Ge, J.; Neofytou, E.; Lei, J.; Beygui, R. E.; Zare, R. N. *small* 2012, 8, 3573.
42. Lam, C. N.; Olsen, B. D. Soft Matter 2013, 9, 2393.
61. Akcasu, A. Z.; Han, C. C. Macromolecules 1979, 12, 276.
64. Pedersen, J. S. Advances in Colloid and Interface Science 1997, 70, 171.
Chapter 7 Conclusions

7.1 Summary

This thesis has investigated the phase behavior of globular protein-polymer block copolymers in concentrated aqueous solution. Making permutations to the protein composition while controlling for the protein shape and tertiary structure, the primary goal was to explore the effect of protein interactions on the phase behavior of globular protein-polymer block copolymers and provide insight into the underlying thermodynamics governing the self-assembly of this class of biohybrid block copolymers. Computer experiments in the form of coarse-grained MD simulations were later combined with form factor measurements to study the relationship between protein-polymer interactions and the shape of protein-polymer block copolymers.

The concentrated solution phase behavior of a globular protein-polymer block copolymer was first explored using the model system comprising the red fluorescent globular protein mCherry and the thermoresponsive polymer poly(N-isopropylacrylamide). A rich window of phase space was discovered, and order-disorder transitions and order-order transitions were demonstrated for the first time in globular protein-containing diblock copolymers. Independent of coil fraction, at low concentrations, mCherry-PNIPAM bioconjugates formed homogeneous, disordered solutions at low temperature and macrophase separated between a conjugate-rich phase and a water-rich phase at high temperature. Coil fraction was observed to have a large effect on the order-disorder transition concentration and the type of ordered phases formed. The location of the minimum in CODT at a nearly symmetric coil fraction was a key finding that suggested that protein-polymer repulsive interactions are important for self-assembly.
As a step toward addressing the central question whether there may exist universal rules that can provide a predictive theoretical framework for protein-containing block copolymers incorporating different proteins, a series of globular proteins with nearly identical structure but different surface potential and patchiness was studied. The proteins mCherry and enhanced green fluorescent protein (EGFP) are structurally homologous but have significantly different surface potentials owing to their different chemical composition; additionally, a series of mCherry variants with single point mutations were identified to study the effects of local, electrostatic patchy interactions. Highly similar results in phase behavior of PNIPAM conjugates containing these different proteins demonstrate that coarse-grained properties such as the protein shape, size, solubility, surface charge, and virial coefficient can capture the general shape of the phase diagram in nonselective solvents. However, protein surface potential and protein-protein interactions were observed to have a strong effect on the stability of micellar phases.

Next, greater changes in protein electrostatic interactions were introduced into a series of globular protein-polymer block copolymers by engineering two GFPs using the superfolder GFP as the reference protein to arrive at three GFP variants with variable charge (0, -8, and -21), electric dipole, and virial coefficient. The electrostatic repulsive interactions of the negatively supercharged GFP resulted in the most poorly ordered nanostructures. In contrast, the Janus-like electrostatic surface potential of the charge-neutral GFP resulted in the direction and magnitude of its electric dipole favoring protein-protein interactions that enhanced nanostructure formation. Results demonstrated that the electrostatic environment of proteins may be designed to tune the morphologies of protein-polymer block copolymers, both enhancing self-assembly through
complementary packing geometries and orientations and suppressing ordered nanostructures due to repulsive interactions.

The final chapter of this thesis the complementary techniques of computer modeling and simulations and small-angle scattering to quantify the relationship between protein-polymer interactions and molecular shape. While the previous thesis objectives focused on understanding the role of protein interactions, four bioconjugates consisting of mCherry and four different polymers—poly(N-isopropylacrylamide), poly(hydroxypropyl acrylate), poly(oligoethylene glycol acrylate), and poly(ethylene glycol)—formed the set of bioconjugates that were modeled using coarse-gained molecular dynamics. The phase behaviors of mCherry-polymer conjugates to all of these polymers except poly(ethylene glycol) showed significant shifts in phase transition lines and different morphologies, motivating the quantification and comparison of their form factors in dilute solution. Model fitting using both a theoretical scattering function and coarse-grained MD simulations in which purely repulsive Weeks-Chandler-Andersen potentials were used to model protein-polymer interactions demonstrated great agreement with experimental results. The coarse-grained model also provides a simple way of modeling the conformation of protein-polymer conjugates with strong interactions that will be useful for systems with hydrophobic and electrostatic interactions between the two components.

7.2 Outlook

Enzymes are very diverse in their functional capabilities, and there is no lack of technological application potential for enzyme-containing materials. One day scientists will be able to synthesize enzyme mimics and de novo enzymes to meet specific needs and challenges that will further expand the technological relevance of these materials. But what's critical to
realize the technological application potential of enzymes are the underlying fundamental principles, physics, and interactions governing the functionality, stability, assembly, and dynamics of enzymes. The promising results of the subgroup focusing on globular protein-polymer block copolymers in the Olsen Lab have provided a foundation for understanding the thermodynamics of protein-polymer block copolymers, motivating countless new fundamental questions that remain to be explored to understand fully this fascinating class of molecules.

One key question that remains to be solved concerns the generality of globular protein-polymer block copolymer self-assembly: Are there universal principles that govern the phase behavior of these molecules? Selecting a protein at random from the vast number that exist, will it be able to self-assemble upon conjugation of a polymer coil, or could it be engineered to do so while retaining its functionality? Certainly, different proteins will have different surface potentials, but being able to decouple interaction potential and protein shape to understand the contribution of packing entropy in the thermodynamics of self-assembly. As proteins are also diverse in their shape and size, it would be helpful to have a method of quantifying the shape of individual proteins or classes of proteins.

Rich phase behavior is observed in concentrated solutions, motivating additional experiments and theoretical and computational work to complement and provide insight. A re-entrant order-disorder transition at very high concentrations has been observed to be a general phenomenon for mCherry-polymer block copolymers. While a re-entrant melting of crystalline colloidal phases has been observed, it is not clear what are the relevant interactions leading to the re-entrant ODT in protein-polymer block copolymers. While protein-polymer interactions appear to be important in self-assembly, it may be of interest to understand better the evolution of protein-protein interactions at very high concentration. As many proteins carry a certain
amount of charge and counterions, effects of counterion condensation and self-buffering may significantly affect protein electrostatic interactions and, consequently, self-assembly.

This thesis has also focused on model globular proteins that are generally quite stable in low ionic strength buffers and even in water and that do not exhibit strong interactions. Proteins with strong directional interactions such as S-layer proteins that form two-dimensional lattices can potentially lead to new morphologies and phase behaviors that are significantly different than those used in this work.

A hybrid particle-field theory that describes the fluid structure of the polymer in a field-theoretic context but models the particle such that the fluid components are excluded has been developed to simulate polymer nanocomposites that could also provide an elegant way of simulating colloid-polymer conjugates. Beginning with a sphere-coil particle, adding complexity to the colloid shape is possible for simulating different colloid shapes. Furthermore, SCFT can readily handle charge and solvent, which is both relevant and useful for modeling protein-polymer conjugates, as protein hydration could be an important consideration.

Fusion proteins of mCherry and elastin-like polypeptide have also demonstrated rich phase behavior in concentrated solution despite their chemical similarity. It is of fundamental interest to understand the governing self-assembly principles of these protein fusion systems, but of equal, if not greater, importance are the advantages of fusion systems for applications. Fusion proteins are fully biosynthetic analogues to protein-polymer conjugates but are entirely genetically encoded and can be synthesized as perfectly defined and monodispersed molecules. Scale-up of biosynthesis and purification is also more facile and economical than achieving site-specific, high-yield bioconjugation techniques. Fusion proteins and protein-polymer conjugates hold much potential for enzyme assembly and novel functional biomaterials, and coupled with
our current fundamental understanding of these systems and the questions that remain to be explored, it will be exciting to watch this area of research grow and evolve over the next couple of decades.
### Appendix A. List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFC</td>
<td>biofuel cell</td>
</tr>
<tr>
<td>RC</td>
<td>reaction center</td>
</tr>
<tr>
<td>EBL</td>
<td>electron beam lithography</td>
</tr>
<tr>
<td>DPN</td>
<td>dip-pen nanolithography</td>
</tr>
<tr>
<td>nCP</td>
<td>nanocontact printing</td>
</tr>
<tr>
<td>SPM</td>
<td>scanning probe microscopy</td>
</tr>
<tr>
<td>AFM</td>
<td>atomic force microscopy</td>
</tr>
<tr>
<td>PPL</td>
<td>polymer pen lithography</td>
</tr>
<tr>
<td>SAM</td>
<td>self-assembled monolayer</td>
</tr>
<tr>
<td>μCP</td>
<td>microcontact printing</td>
</tr>
<tr>
<td>LbL</td>
<td>layer-by-layer</td>
</tr>
<tr>
<td>ATRP</td>
<td>atom transfer radical polymerization</td>
</tr>
<tr>
<td>RAFT</td>
<td>reversible addition-fragmentation transfer</td>
</tr>
<tr>
<td>NMRP</td>
<td>nitroxide-mediated radical polymerization</td>
</tr>
<tr>
<td>ELP</td>
<td>elastin-like polypeptide</td>
</tr>
<tr>
<td>OOT</td>
<td>order-order transition</td>
</tr>
<tr>
<td>ODT</td>
<td>order-disorder transition</td>
</tr>
<tr>
<td>PPV</td>
<td>poly(alkoxyphenylenevinylene)</td>
</tr>
<tr>
<td>PI</td>
<td>polyisoprene</td>
</tr>
<tr>
<td>LCST</td>
<td>lower critical solution temperature</td>
</tr>
<tr>
<td>IQ</td>
<td>isoperimetric quotient</td>
</tr>
<tr>
<td>PRISM</td>
<td>polymeric reference interaction site model</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>PEO</td>
<td>poly(ethylene oxide)</td>
</tr>
<tr>
<td>SANS</td>
<td>small-angle neutron scattering</td>
</tr>
<tr>
<td>PNIPAM</td>
<td>poly(N-isopropylacrylamide)</td>
</tr>
<tr>
<td>EGFP</td>
<td>enhanced green fluorescent protein</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>MD</td>
<td>molecular dynamics</td>
</tr>
<tr>
<td>EMP</td>
<td>2-ethylsulfanylthiocarbonyl sulfanyl-2-methylpropionic acid</td>
</tr>
<tr>
<td>DCC</td>
<td>N,N’-dicyclohexylcarbodiimide</td>
</tr>
<tr>
<td>NIPAM</td>
<td>N-isopropylacrylamide</td>
</tr>
<tr>
<td>MEHQ</td>
<td>hydroquinone monomethyl ether</td>
</tr>
<tr>
<td>HPA</td>
<td>2-hydroxypropyl acrylate</td>
</tr>
<tr>
<td>MEA</td>
<td>2-methoxyethyl acrylate</td>
</tr>
<tr>
<td>MEEA</td>
<td>2-(2-methoxyethoxy)ethyl acrylate</td>
</tr>
<tr>
<td>CTA</td>
<td>chain transfer agent</td>
</tr>
<tr>
<td>AIBN</td>
<td>azobisisobutyronitrile</td>
</tr>
<tr>
<td>PHPA</td>
<td>poly(2-hydroxypropyl acrylate)</td>
</tr>
<tr>
<td>POEGA</td>
<td>poly(oligoethylene glycol acrylate)</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>PEG</td>
<td>poly(ethylene glycol)</td>
</tr>
<tr>
<td>sfGFP</td>
<td>superfolder GFP</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny broth</td>
</tr>
<tr>
<td>TB</td>
<td>Terrific broth</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>nickel-nitritriacetic acid</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>BME</td>
<td>β-mercaptoethanol</td>
</tr>
<tr>
<td>FPLC</td>
<td>fast protein liquid chromatography</td>
</tr>
<tr>
<td>SLS</td>
<td>static light scattering</td>
</tr>
<tr>
<td>CV</td>
<td>column volume</td>
</tr>
<tr>
<td>TCEP-HCl</td>
<td>tris(2-carboxyethyl)phosphine hydrochloride</td>
</tr>
<tr>
<td>SAXS</td>
<td>small-angle X-ray scattering</td>
</tr>
<tr>
<td>NSLS</td>
<td>National Synchrotron Light Source</td>
</tr>
<tr>
<td>APS</td>
<td>Advanced Photon Source</td>
</tr>
<tr>
<td>ALS</td>
<td>Advanced Light Source</td>
</tr>
<tr>
<td>GP-SANS</td>
<td>General Purpose Small-Angle Neutron Scattering Diffractometer</td>
</tr>
<tr>
<td>DPLS</td>
<td>depolarized light scattering</td>
</tr>
<tr>
<td>ND</td>
<td>neutral density</td>
</tr>
<tr>
<td>DLS</td>
<td>dynamic light scattering</td>
</tr>
<tr>
<td>DSC</td>
<td>differential scanning calorimetry</td>
</tr>
<tr>
<td>ISN</td>
<td>Institute for Solider Nanotechnologies</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier transform infrared spectroscopy</td>
</tr>
<tr>
<td>FSD</td>
<td>Fourier self-deconvolution</td>
</tr>
<tr>
<td>PBE</td>
<td>Poisson-Boltzmann equation</td>
</tr>
<tr>
<td>APBS</td>
<td>Adaptive Poisson-Boltzmann Solver</td>
</tr>
<tr>
<td>POV-Ray</td>
<td>Persistence of Vision Raytracer</td>
</tr>
<tr>
<td>ASP</td>
<td>atomic solvation parameter</td>
</tr>
<tr>
<td>WCA</td>
<td>Weeks-Chandler-Andersen</td>
</tr>
<tr>
<td>LJ</td>
<td>Lennard-Jones</td>
</tr>
<tr>
<td>MSD</td>
<td>mean squared displacement</td>
</tr>
<tr>
<td>VMD</td>
<td>Visual Molecular Dynamics</td>
</tr>
<tr>
<td>ODTC</td>
<td>order-disorder transition concentration</td>
</tr>
<tr>
<td>EGMEA</td>
<td>ethylene glycol methyl ether acrylate</td>
</tr>
<tr>
<td>SCFT</td>
<td>self-consistent field theory</td>
</tr>
<tr>
<td>PS</td>
<td>polystyrene</td>
</tr>
<tr>
<td>DOP</td>
<td>dioctyl phthalate</td>
</tr>
<tr>
<td>DBP</td>
<td>dibutyl phthalate</td>
</tr>
<tr>
<td>DEP</td>
<td>diethyl phthalate</td>
</tr>
<tr>
<td>DMP</td>
<td>dimethyl phthalate</td>
</tr>
<tr>
<td>TEM</td>
<td>transmission electron microscopy</td>
</tr>
<tr>
<td>DLVO</td>
<td>Derjaguin-Landau-Verwey-Overbeek</td>
</tr>
<tr>
<td>FWHM</td>
<td>full width at half maximum</td>
</tr>
<tr>
<td>HWHM</td>
<td>half-width at half maximum</td>
</tr>
<tr>
<td>WLC</td>
<td>worm-like chain</td>
</tr>
</tbody>
</table>
Appendix B. Supporting Information for Chapter 3

Figure B-1. Gel permeation chromatography of (a) PNIPAM8k: $M_n = 8,190$ g/mol; $M_w = 9,640$ g/mol; PDI = 1.18 (b) PNIPAM17k: $M_n = 17,200$ g/mol; $M_w = 19,300$ g/mol; PDI = 1.12 (c) PNIPAM27k: $M_n = 27,000$; $M_w = 29,400$; PDI = 1.09. (d) PNIPAM30k: $M_n = 29,700$; $M_w = 37,100$; PDI = 1.25 (e) PNIPAM57k: $M_n = 57,100$ g/mol; $M_w = 60,700$ g/mol; PDI = 1.06
Figure B-2. Denaturing protein gels of mCherryS131C-b-PNIPAM bioconjugates are shown in (a) – (c); a native protein gel of the bioconjugates is shown in (d). Lanes 1-4 in (a) represent ladder, mChP17, mChP30, mChP57, respectively. Lanes 1-2 in (b) represent ladder and mChP8, respectively. Lanes 1-2 in (c) represent ladder and mChP27, respectively. Lane 1 in (d) represents the native gel ladder, and lanes 2 – 6 correspond to mChP8, mChP17, mChP30, mChP57, and mChP27. Band at 30k is unconjugated mCherry impurity, and smaller molar mass bands correspond to mCherry fragments resulting from breaking of the chromophore acylimine bond during gel sample preparation. Purity estimates by mass percent for the bioconjugates are as follows: mChP8k: 85%; mChP17k: 95%; mChP27k: 94%; mChP30k: 97%; mChP57k: 97%. Explain why native gel molecular weights are artificially high.
Figure B-3. Representative depolarized light scattering (DPLS) measurements for (a) mChP8, (b) mChP17, (c) mChP30, and (d) mChP57. DPLS was measured from 10 to 40 °C and corrected for transmissive and reflective losses. Macrophase separation and a subsequent loss in transmission at high temperatures results in an artificially high power fraction after correction for transmissive losses; therefore, data has been omitted in mChP30 and mChP57 in the 20 wt.% curves beyond their macrophase separation transitions.
Figure B-4. Small-angle neutron scattering of mCherryS131C-b-PNIPAM27k in different H₂O/D₂O blend compositions at (a) T = 10 °C and (b) T = 40 °C. At 40 °C, higher scattering intensity with increasing D₂O content is observed as water favorably partitions into the protein block as the polymer domain collapses above the thermal transition temperature.

Figure B-5. Zimm plot of mCherry. From extrapolation to zero concentration, the weight-average molar mass is determined to be 3.8e4 g/mol, somewhat larger than the expected value of 2.8e4 g/mol. mCherry is observed to exert weak self-repulsion in solution. The second virial coefficient was determined to be A₂ = 1.1e-4 mol mL g⁻².
Table B-1. The concentration and temperature conditions corresponding to each representative SAXS pattern in Figure 3-3 are listed below.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Phase</th>
<th>Conc. (wt.%)</th>
<th>T (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mChP8</td>
<td>Hex</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>NB Lam</td>
<td>40</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>DM</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Dis</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>mChP17</td>
<td>HPC</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Hex</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>NB Lam</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>DM</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Dis</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>mChP30</td>
<td>PL</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Lam</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>NB Lam</td>
<td>33</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>DM</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Dis</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>mChP57</td>
<td>Lam</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>NB Lam</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>DM</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Dis</td>
<td>25</td>
<td>10</td>
</tr>
</tbody>
</table>
Table B-2. Scattering length densities (SLDs) of molecules. SLDs were computed using the Scattering Length Density Calculator from the NIST Center for Neutron Research:

http://www.ncnr.nist.gov/resources/sldcalc.html

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Density (g/cm³)</th>
<th>Molecular formula</th>
<th>SLD (Å⁻²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mCherryS131C</td>
<td>1.19</td>
<td>C₁₂₅₀H₁₉₁₅N₃₄₁O₃₇₇S₁₂</td>
<td>1.67e⁻⁶</td>
</tr>
<tr>
<td>PNIPAM</td>
<td>1.05</td>
<td>C₆H₁₁NO</td>
<td>7.8e⁻⁷</td>
</tr>
<tr>
<td>H₂O</td>
<td>1.00</td>
<td>H₂O</td>
<td>-5.6e⁻⁷</td>
</tr>
<tr>
<td>D₂O</td>
<td>1.107</td>
<td>D₂O</td>
<td>6.37e⁻⁶</td>
</tr>
</tbody>
</table>
Appendix C. Supporting Information for Chapter 4

C.1 Determining PNIPAM Coil Fraction

From published crystallographic data for EGFP (PDB 2Y0G) and mCherry (PDB 2H5Q), the crystallized monomeric molecular weights and crystal unit cell parameters and solvent fraction are tabulated as follows:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>EGFP</th>
<th>mCherry</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW (kg/mol)</td>
<td>28.5694</td>
<td>26.7404</td>
</tr>
<tr>
<td>Formula units z</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Unit cell length a (Å)</td>
<td>51.05</td>
<td>48.76</td>
</tr>
<tr>
<td>Unit cell length b (Å)</td>
<td>62.18</td>
<td>42.85</td>
</tr>
<tr>
<td>Unit cell length c (Å)</td>
<td>70.02</td>
<td>61.06</td>
</tr>
<tr>
<td>Unit cell angle α (°)</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>Unit cell angle β (°)</td>
<td>90.00</td>
<td>112.31</td>
</tr>
<tr>
<td>Unit cell angle γ (°)</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>Crystal solvent content $\phi_s$ (%)</td>
<td>40</td>
<td>44.28</td>
</tr>
</tbody>
</table>

The protein density is then calculated as follows:

$$\rho_{protein} = \left( \frac{1}{1 - \phi_s} \right) \left( \frac{zMW_{protein}}{N_A V_{unit cell}} \right)$$

where $N_A$ is Avogadro’s number. The densities are calculated to be

$$\rho_{EGFP} = 1.42 \, \text{g cm}^{-3}$$

$$\rho_{mCherry} = 1.35 \, \text{g cm}^{-3}$$

Using a density for PNIPAM\(^1\) of 1.05 g cm\(^{-3}\), the PNIPAM coil fraction can then calculated as below:

$$\phi_{PNIPAM} = \frac{MW_{PNIPAM}}{\rho_{PNIPAM}} - \frac{MW_{protein}}{\rho_{protein}}$$
C.2 Domain Spacing Scaling Analysis

Solid State

For a polymer containing a rigid block, the domain spacing $d$ can be calculated as $d = Na^2$ where $N$ is the number of volumetric repeat units and $a$ is the statistical segment length of the rigid block. The total number of volumetric repeat units $N$ comprises both the protein and polymer blocks, hence $N = N_{\text{protein}} + N_{\text{PNIPAM}}$. Therefore,

$$d = N_{\text{protein}} a \left( 1 + \frac{N_{\text{PNIPAM}}}{N_{\text{protein}}} \right) = N_{\text{protein}} a \left( \frac{N_{\text{protein}} + N_{\text{PNIPAM}}}{N_{\text{protein}}} \right) = N_{\text{protein}} a \left( \frac{1}{1 - \phi_{\text{PNIPAM}}} \right)$$

For the symmetric conjugates, $N_{\text{mCherry}} a = 4.2$ nm and $\phi_{\text{PNIPAM}} = 0.55$ for mCherry-$b$-PNIPAM26k, and $N_{\text{EGFP}} a = 4.3$ nm and $\phi_{\text{PNIPAM}} = 0.55$ for EGFP-$b$-PNIPAM26k. Substituting these values,

$$d_{\text{mChP26}} = 9.33 \text{ nm}$$
$$d_{\text{EGFP26}} = 9.56 \text{ nm}$$

for monolayer packing within lamellar nanodomains. These calculated domain spacings are slightly smaller than half of the domain spacings observed by SAXS: 23.2 and 21.3 nm for mCherry-$b$-PNIPAM26k and EGFP-$b$-PNIPAM26k, respectively, suggesting formation of bilayer lamellar morphologies.

Concentrated Solution

In the concentrated solution regime, the scaling law for a coil-coil block copolymer is applied as a zeroth-order approximation for the globular protein-polymer block copolymer. In a concentrated solution of $z$ wt. %, the increase in domain spacing due to swelling of the block copolymer domains scales as
\[ \Delta d = \left( \frac{1}{z / 100} \right)^{1/3} \]

Therefore, the domain spacing in concentrated solution, \( d' \), swells to

\[ d' = d \left( \frac{1}{z / 100} \right)^{1/3} \]

For example, for 50 wt.% samples for EGFP-b-PNIPAM26k and mCherry-b-PNIPAM26k,

\[ d'_{\text{EGFP26}} = 26.8 \text{ nm} \]
\[ d'_{\text{mCherry26}} = 29.2 \text{ nm} \]

These values are very similar to the domain spacings calculated from the primary peak observed by SAXS, suggesting the formation of bilayer lamellar morphologies in concentrated solution.

**C.3 Calculation of Solvation Energy \( \Delta G \).**

Following the method of Eisenberg and McLachlan,\(^3\) the solvation free energies for the globular proteins mCherry (PDB 2H5Q) and EGFP (PDB 2Y0G) are calculated using the software PyMOL.\(^4\) Modeller\(^4\) was used to add the missing residues and atoms to the PDB files for mCherry and EGFP, and PROPKA\(^6\) was used to assign charge states at pH = 7.0. The residues glutamic acid and aspartic acid with pKa values less than 5.5 were assumed to be in the conjugate base O\(^-\) state, and the residues lysine, arginine, and histidine with pKa values greater than 8.5 were assumed to be prononated as N\(^+\). Using PyMOL, the charge was assumed to be concentrated on the most solvent exposed O or N atom. In PyMOL, hydrogen atoms were added to the molecule, a solvent radius of 1.4 Å was specified to determine the solvent-accessible surface, and the dot density to be the maximum value of 4 for the highest achievable surface area calculations. The procedure consists of calculating the solvent-accessible atomic areas of the
proteins, weighting each atomic area by its atomic solvation parameter (ASP), followed by summation.

C.4 Experimental Methods

Polymer Synthesis

Poly(N-isopropylacrylamide) (PNIPAM) was synthesized by reversible addition-fragmentation chain transfer (RAFT) polymerization as previously described. The monomer to chain transfer agent (CTA) ratios were 600:1 and 300:1, and reactions were performed at 55 °C for 110 and 85 min. to synthesize PNIPAM with molar masses of 26 kDa and 13 kDa, respectively. The absolute molar mass and polydispersity were characterized by gel permeation chromatography (Figure C-2) using an Agilent Technologies 1260 Infinity system, a Wyatt miniDAWN TREOS MALS light scattering detector, and a Wyatt Optilab T-rEX refractometer. N,N-dimethylformamide with 0.02 M LiBr was used as the mobile phase.

DNA cloning of mCherry variants

mCherry variants were constructed from a gene sequence of mCherry containing a single cysteine in pQE9 plasmid through a single-site mutation method. The mCherry gene was amplified by polymerase chain reaction (PCR) with primers containing mutation sequences (Table S1). After DpnI enzyme and T4 Polynucleotide Kinase treatments (product #: R0176S, M0201S, NEB, Ipswich, MA, USA) to remove template plasmids and add phosphates in PCR products, gel purified (QIAquick Gel Extraction kit, QIAGEN, Valencia, CA, USA) PCR products were ligased by T4 DNA ligase (product #: M0202S, NEB, Ipswich, MA, USA). The ligated plasmids were transformed into NEB turbo competent Escherichia coli (E. coli) (product#: C2984I, NEB, Ipswich, MA, USA) and grown on LB agar plates with ampicillin and
kanamycin. Plasmids from grown colonies were extracted and the PCR fidelity of the mutated plasmids was confirmed by gene sequencing (Genewiz, USA).

**Protein expression**

The proteins mCherry, all mCherry variants and enhanced green fluorescent protein (EGFP), each containing an N-terminal 6xHis tag, were expressed in the *E. coli* strain SG13009 containing the pREP4 repressor plasmid. A freshly grown bacterial colony was inoculated in 5ml Lysogeny broth (LB) medium supplemented with 0.2 mg/mL ampicillin (amp) and 0.05 mg/mL kanamycin (kan) at 37°C overnight. For pre-culture, 1 L LB medium (amp+kan) and 10 ml of the overnight culture were incubated until an OD600 of ~1.0 at 37°C for mCherry and the variants, and 29°C for EGFP. For mCherry variants, 2xYT medium (amp+kan) was used instead of LB medium. Protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and cells were harvested by centrifugation after 6 hours at 37°C (mCherry) and 12 hours at 29°C (mCherry variants and EGFP) and then precipitated cells were stored at -80°C. The frozen cells were resuspended in lysis buffer (50 mM NaH2PO4, 300 mM sodium chloride, 10 mM imidazole, 10 mM β-mercaptoethanol (BME), pH 8.0), incubated with 1 mg/mL lysozyme at 4 °C for 30 min, and sonicated. The lysate was clarified, and the protein was purified using Ni–NTA metal affinity chromatography. Throughout the purification, 10 mM BME was used in all buffers. Elution fractions containing purified protein were dialyzed into 20 mM Tris-Cl buffer, pH = 8.0. The yield in the elution fractions was determined spectrophotometrically using the absorbance peak at 586 nm (extinction coefficient of 72,000 M⁻¹ cm⁻¹) for mCherry and variants and 488 nm (extinction coefficient of 56,000 M⁻¹ cm⁻¹) for EGFP. Typical yields were 100 mg/L of cell culture for mCherry and variants (70 mg/L for the
least expressible mCherry variant K123E) and 50 mg/L for EGFP. The purity was confirmed by denaturing gel electrophoresis (SDS-PAGE). The variants EGFPC14S (N-terminal reactive cysteine mutated to a serine to remove the single reactive thiol) and mCherry (without any cysteine residues) were expressed similarly to mCherry and EGFP and used for macrophase separation studies. DNA and amino acid sequences for EGFP and the variant EGFPC14S are included in Figure C-1.

The coupling reactions between EGFP or mCherry and maleimide end-functionalized PNIPAM were performed in 20 mM Tris buffer, pH 8.0, as described previously. Although the amino acid sequence for EGFP contains three cysteine residues, only one of them was found to be solvent-accessible and reactive (Figure C-3). Following the thiol-maleimide coupling reaction, at least two cycles of precipitation in 1 M (NH₄)₂SO₄ were performed, yielding a clear supernatant in the final precipitation cycle to ensure complete removal of unconjugated protein. Excess polymer was then removed by Ni-NTA metal affinity chromatography; wash fractions were monitored by checking for precipitation above 32°C to ensure complete removal of excess PNIPAM. Purity was analyzed using native and SDS-PAGE (Figure C-4).

Sample Preparation and Characterization

Conjugate solutions were concentrated to approximately 100 mg/mL using Millipore Ultra-15 centrifugal filters with a molecular weight cutoff of 10 kDa. Bulk solid samples were then prepared by drop-casting aliquots of the concentrated bioconjugate solution onto Teflon sheets and drying under vacuum overnight at room temperature. For solution studies, the solid material was rehydrated to the desired concentration. Solid-state samples were cast from concentrated aqueous solutions at room temperature using a vacuum controller to control the solvent evaporation rate with a ramp rate of 300 Torr/hr to a final set point of 50 Torr, followed
by allowing the samples to dry further under full vacuum overnight. Samples were then annealed for 8 hrs in water at room temperature.

Small-angle X-ray scattering (SAXS) samples were prepared in 1 mm thick washers and sealed with Kapton tape. The bioconjugates EGFPP26.3 and mChP26.3 were measured at the Argonne National Laboratory Advanced Photon Source at Beamline 12-ID-B; bioconjugates EGFPP12.8, mChP12.8 and the conjugates of the mCherry mutants—mChK123EP26.0, mChK123SP26.0, mChE26SP26.0, mChE26KP26.0—were measured at Brookhaven NSLS X27C. Samples were equilibrated for 20 minutes at 10 °C and for 10 minutes at all other temperatures prior to data collection. Data of mChP26.0 (variants) were only collected at 15 °C, a low temperature where well-ordered structures had previously been observed in mCherry-b-PNIPAM and EGFP-b-PNIPAM. SAXS data were corrected for empty cell and dark field scattering, and acquisition times were minimized to prevent beam damage. All observed transitions were reversible with temperature.

Turbidimetry and depolarized light scattering (DPLS)\(^\text{11}\) were performed on samples loaded into a 1 mm thick Teflon mold and sealed between two quartz disks. A Coherent OBIS LX660 laser was used with wavelength \(\lambda = 662\) nm (beyond the absorption cutoff of mCherry and EGFP) and continuous wave output power 20 mW. Samples were equilibrated at 10 °C for 20 minutes and then heated at 0.5 °C/min. to 40 °C, allowed to equilibrate for 10 minutes, and then cooled at 0.5 °C/min. to 10 °C. The static depolarized light scattering signal was corrected for transmission and dark field background. For turbidimetry measurements, the same apparatus was used without the rear polarizer to enable measurement of sample transmission. Macrophase separation transitions were defined as the temperatures \(T_1\) corresponding to a 10 % reduction in
the initial sample transmittance, according to literature methods. Transitions were reproducible upon repeated cycling, and data analysis was performed on the first heating cycle.

For differential scanning calorimetry (DSC) measurements, bioconjugate solution was loaded into a hermetically sealed aluminum pan. Data was acquired using a Texas Instruments Q-10 differential scanning calorimeter. The sample was equilibrated at 5 °C for 10 minutes, followed by two cycles of ramping to 45 °C at 5 °C/min., holding isothermally for 2 minutes, cooling at 5 °C/min. to 5 °C, and holding isothermally for 2 minutes. Temperature transition values were extracted from the peak point in the initial heating cycle.

Static light scattering was performed on EGFPC14S, and the second virial coefficient was compared to that of mCherry as reported previously. Measurements were performed using a Wyatt Technology DAWN 8 equipped with a 120 mW solid-state GaAs laser at 658 nm configured in offline-batch mode. Four different concentrations at 3.97, 4.85, 8.04, and 10.22 mg/mL were measured in 10 mM Tris Cl pH = 7.0. Measurements were made at 25 °C, and $dn/dc = 0.185 \text{ mL/g}$ for proteins was used for data analysis.

Zeta potential measurements were made of EGFPC14S, mCherry, mCherryE26S, mCherryE26K, mCherryK123S and mCherryK123E using a Zetasizer Nano ZS (Malvern Instruments Ltd). Protein samples (0.25 mg/mL) were measured in 20 mM Tris-Cl (pH 8) with 5 mM Tcep (product #77720, Thermo Scientific), and the Smoluchovski model for samples in aqueous solutions is used to analyze results.

The macrophase separation behavior of mCherry/PNIPAM and EGFPC14S/PNIPAM blends was studied by UV-Vis turbidimetry. Protein solutions in 10 mM Tris Cl pH = 7.0 were concentrated using Millipore Ultra-15 centrifugal filters with a molecular weight cutoff of 10 kDa and then diluted to 250 mg/mL. A 250 mg/mL solution of PNIPAM of molar mass 27 kDa
(Figure C-2) in 10 mM Tris Cl pH = 7.0 was also prepared. Protein and polymer solutions were mixed in a 1:1 ratio volumetrically and allowed to equilibrate at 4 °C. The cloud point was then measured by UV-Vis using a temperature ramp of 0.1 °C/min. from 4 to 35 °C. Blend compositions from mixing protein and polymer solutions each at 200, 150, 100, and 50 mg/mL were prepared and measured. The cloud point was determined as the temperature at which the sample was 50 % transmissive.

Bulk samples for transmission electron microscopy were cryo-microtomed using a Leica EM UC6 at -110 °C to a thickness of 60 nm and stained with ruthenium tetroxide vapors from a 0.5% aqueous solution for 20 min. This treatment preferentially stains the protein domains, making them appear darker in images. A JEOL 2011 TEM was used to obtain bright field images using an accelerating voltage of 120 kV and a LaB6 filament. Images were captured using a camera in a fixed bottom mount configuration.
**EGFP DNA sequence**
ATGAGAGGATCGCAGCAGAGAGACGCTCATGCGAGCTGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAACTGGACGGCGACGTAAACGGCCACAAGTTCAGC

**EGFPC14S DNA sequence**
ATGAGAGGATCGCAGCAGAGAGACGCTCATGCGAGCTGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAACTGGACGGCGACGTAAACGGCCACAAGTTCAGC

**Figure C-1.** DNA and amino acid sequences of EGFP and mutant EGFPC14S in this study.
Figure C-2. Gel permeation chromatography of (a) PNIPAM13k: $M_n = 12,830$ g/mol; $M_w = 14,230$ g/mol; PDI = 1.11 (b) PNIPAM26k: $M_n = 26,290$ g/mol; $M_w = 28,870$ g/mol; PDI = 1.10 (c) PNIPAM27k for protein/polymer macrophase blend studies: $M_n = 27,010$ g/mol; $M_w = 29,930$ g/mol; PDI = 1.11 (d) PNIPAM26k for mCherry mutant conjugation: $M_n = 26,040$ g/mol; $M_w = 27,230$ g/mol; PDI = 1.05.
Table C-1. mCherry protein sequence and primers for preparation of mCherry variants in this study. Mutation locations are highlighted in blue at the mCherry protein sequence.

<table>
<thead>
<tr>
<th>mCherry variants</th>
<th>Primer sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>mCherryE26S</td>
<td>5'-GCGTTAATGGCCACAGCTTTGAATTTGAAG-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-TACCTTCCATATGCACTTTGAACGC-3' (reverse)</td>
</tr>
<tr>
<td>mCherryE26K</td>
<td>5'-GGAGTTGGAATTAGCCGCGGAAGTGAAAGTC-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-TACCTTCCATATGCACCTTTGAAACGC-3' (reverse)</td>
</tr>
<tr>
<td>mCherryE30S</td>
<td>5'-GAGTTTGAATTAGCCGCGGAAGTGAAAGTC-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-TACCTTCCATATGCACCTTTGAAACGC-3' (reverse)</td>
</tr>
<tr>
<td>mCherryE30K</td>
<td>5'-GAAGCCGGAAGTGAAAGTGAAGTGTTAC-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-TACCTTCCATATGCACCTTTGAAACGC-3' (reverse)</td>
</tr>
<tr>
<td>mCherryE34S</td>
<td>5'-AATTTCAATGCGGCCATTAACGCTACC-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-GAAGGCGAAGTGAAAGTGAAGTGTTAC-3' (reverse)</td>
</tr>
<tr>
<td>mCherryE34K</td>
<td>5'-AATTTCAATGCGGCCATTAACGCTACC-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-GAAGGCGAAGTGAAAGTGAAGTGTTAC-3' (reverse)</td>
</tr>
<tr>
<td>mCherryE100S</td>
<td>5'-GTGATGAACTCTAACGCGATAGCGGCTGTTAC-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-ATCTCACAAAGTGACCGGCGGTACCAAT-3' (reverse)</td>
</tr>
<tr>
<td>mCherryE100K</td>
<td>5'-GTGATGAACTCTAACGCGATAGCGGCTGTTAC-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-ATCTCACAAAGTGACCGGCGGTACCAAT-3' (reverse)</td>
</tr>
<tr>
<td>mCherryK123S</td>
<td>5'-ATCTCACAAAGTGACCGGCGGTACCAAT-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-ATCTCACAAAGTGACCGGCGGTACCAAT-3' (reverse)</td>
</tr>
<tr>
<td>mCherryK123E</td>
<td>5'-ATCTCACAAAGTGACCGGCGGTACCAAT-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-ATCTCACAAAGTGACCGGCGGTACCAAT-3' (reverse)</td>
</tr>
<tr>
<td>mCherryE160S</td>
<td>5'-GCACGTGAAAAGCCACAGCATCAAACACGCGGCTG-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-ACCATCCTCCGCGGTACCATACGTTGCTAC-3' (reverse)</td>
</tr>
<tr>
<td>mCherryE160K</td>
<td>5'-GCACGTGAAAAGCCACAGCATCAAACACGCGGCTG-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-ACCATCCTCCGCGGTACCATACGTTGCTAC-3' (reverse)</td>
</tr>
<tr>
<td>mCherryD200S</td>
<td>5'-GTTCACATCAAAACTGACCCACCACCCATAAC-3' (forward)</td>
</tr>
<tr>
<td>mCherryD200K</td>
<td>5'-GTTGTATGCACCCGCGCAGCTGCACCGG-3' (reverse)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>5'-GTAAACATCAAACCTGAAAATCACCAGCCAATAAC-3' (forward)</td>
</tr>
<tr>
<td></td>
<td>5'-GTTGTATGCACCCGCGCAGCTGCACCGG-3' (reverse)</td>
</tr>
</tbody>
</table>
Figure C-3. Denaturing gel of test bioconjugation of wild-type EGFP with maleimide-PEG 5k. Lanes 1-3 represent ladder, crude bioconjugation reaction with EGFP, and crude bioconjugation reaction with EGFP14S, respectively. Results indicate that although EGFP contains 3 cysteine residues, only 1 is solvent-accessible for thiol-maleimide bioconjugation.

Figure C-4. (a) Native protein gel of the bioconjugates. Lanes 1-5 represent ladder, EGFP12.8, mChP12.8, EGFP26.3, mChP26.3, respectively. Native PAGE shows undetectable levels of impurity, suggesting that the bioconjugates are > 98 % pure. (b) SDS-PAGE of the bioconjugates. Lanes 1-4 represent EGFP12.8, EGFP26.3, mChP12.8, and mChP26.3, respectively, juxtaposed with protein ladders. Bands visible at 30k in the denaturing gels represent unconjugated mCherry or EGFP, resulting from cleavage of protein and polymer during gel boiling due to hydrolysis of the ester bond linking the two blocks. For the mCherry-PNIPAM samples, denaturing gels often show two smaller molar mass bands corresponding to mCherry fragments resulting from cleavage of the chromophore acylimine bond during gel sample preparation.²
Figure C-5. Depolarized light scattering (left) and turbidimetry (right) results for concentrated solutions of (a) mChP12.8 (b) mChP26.3 (c) EGFP12.8 and (d) EGFP26.3 show thermal reversibility. DPLS was measured from 10 to 40 °C and corrected for transmissive and reflective losses. Macrophase separation and a subsequent loss in transmission at high temperatures results in an artificially high power fraction after correction for transmissive losses; therefore, data has been omitted in the DPLS results for mChP12.8 and mChP26.3 at temperatures beyond their macrophase separation transitions.
Figure C-6. Zimm plot of EGFPC14S. From extrapolation to zero concentration, the weight-average molar mass is determined to be $39.1 \pm 0.1$ kDa, slightly greater than the expected value of 29.5 kDa. The second virial coefficient was determined to be $A_2 = 1.0 \pm 0.2$ mol mL g$^{-2}$.

Figure C-7. Cloud point of EGFPC14S/PNIPAM and mCherry/PNIPAM blends as a function of blend concentration.
**Figure C-8.** UV-Vis of mCherry and its variants showing minimal change in the ratio between the maximum absorption peak of mCherry (586 nm) relative to the A280 absorption peak.

**Figure C-9.** Concentrated solution SAXS of mCherry and mutants prepared with 200 mM NaCl, showing that the order-disorder transition concentration ($C_{ODT}$) has increased by approximately 5 wt.\% relative to self-assembly in H$_2$O.
After dialysis
Room temperature water anneal 8 hr
Room temperature water cast
Rehydrated

(a)

(b)

Figure C-10. UV-Vis absorption spectra for a) EGFP-b-PNIPAM13k and b) EGFP-b-PNIPAM26k after dialysis, after solid-state casting from room temperature water, after annealing in room temperature water for 8 hours, and after rehydration, showing that solid-state self-assembly and processing conditions have minimal effect on the protein optical activity. Upon conjugation of EGFP to PNIPAM, a secondary absorbance peak is observed near 395 nm, which is the primary peak observed in wild-type GFP, possibly resulting from slight perturbation of the chromophore.15

Figure C-11. Representative solid-state FTIR spectra showing the amide I region for the EGFP-PNIPAM conjugates as cast and after annealing in room temperature water for 8 hours. Fourier self-deconvolution (FSD) analysis indicates a high fraction of β-sheet structure for both PNIPAM coil fractions with the larger coil fraction showing a lower fraction of β-sheet structure due to the larger fraction of random coil polymer present.
Table C-2. Fourier self-deconvolution analysis of FTIR spectra.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Average fraction helix</th>
<th>Average fraction sheet</th>
<th>Average fraction turn</th>
<th>Average fraction unordered</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFP-PNIPAM13 As cast</td>
<td>0.213</td>
<td>0.494</td>
<td>0.088</td>
<td>0.205</td>
</tr>
<tr>
<td>eGFP-PNIPAM13 Room temp water annealed</td>
<td>0.242</td>
<td>0.424</td>
<td>0.116</td>
<td>0.218</td>
</tr>
<tr>
<td>eGFP-PNIPAM26 As cast</td>
<td>0.253</td>
<td>0.414</td>
<td>0.086</td>
<td>0.246</td>
</tr>
<tr>
<td>eGFP-PNIPAM26 Room temp water annealed</td>
<td>0.266</td>
<td>0.411</td>
<td>0.105</td>
<td>0.219</td>
</tr>
<tr>
<td>Average</td>
<td>0.243</td>
<td>0.436</td>
<td>0.099</td>
<td>0.222</td>
</tr>
</tbody>
</table>

Figure C-12. Circular dichroism spectroscopy of EGFP (in 20 mM Tris Cl pH = 8.0) and EGFP conjugates with a) \( \phi_{PNIPAM} = 0.37 \) and b) \( \phi_{PNIPAM} = 0.55 \) after purification and after rehydration from solid-state samples cast and annealed in room temperature water. The secondary structure of EGFP is preserved during the conjugation and self-assembly process. The concentration of EGFP is 0.2 mg/mL and for the EGFP-PNIPAM conjugates, solution concentrations were prepared at 0.25 mg/mL.
C.5 References

4. *The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.*
5. Sali, A.; Blundell, T., Comparative protein modelling by satisfaction of spatial restraints. *Protein structure by distance analysis* 1994, 64, C86.
Appendix D. Supporting Information for Chapter 5

Sample Preparation and Characterization

Conjugate solutions were concentrated to approximately 100 mg/mL using Millipore Ultra-15 centrifugal filters with a molecular weight cutoff of 10 kDa. Bulk solid samples were then prepared by drop-casting aliquots of the concentrated bioconjugate solution onto Teflon sheets and drying under vacuum overnight at room temperature. For solution studies, the solid material was rehydrated to the desired concentration.

Small-angle X-ray scattering (SAXS) samples were prepared in 1 mm thick washers and sealed with Kapton tape. The bioconjugates with $\phi_{\text{PNIPAM}} = 0.50$ were measured at the Argonne National Laboratory Advanced Photon Source at Beamline 12-ID-C,D; bioconjugates with $\phi_{\text{PNIPAM}} = 0.59$ were measured at the Stanford Synchrotron Radiation Lightsource (SLAC), and divalent cation salt studies were performed at Beamline 7.3.3 at the Advanced Light Source. Samples were equilibrated for 10 minutes prior to data collection. SAXS data were corrected for empty cell and dark field scattering, and acquisition times were minimized to prevent beam damage. All observed transitions were reversible with temperature.

Turbidimetry and depolarized light scattering (DPLS) were performed on samples loaded into a 1 mm thick Teflon mold and sealed between two quartz disks. A Coherent OBIS LX660 laser was used with wavelength $\lambda = 662$ nm (beyond the absorption cutoff of the GFP variants) and continuous wave output power 20 mW. Samples were equilibrated at 10 °C for 20 minutes and then heated at 0.5 °C/min. to 40 °C, allowed to equilibrate for 10 minutes, and then cooled at 0.5 °C/min. to 10 °C. The static depolarized light scattering signal was corrected for transmission and dark field background. For turbidimetry measurements, the same apparatus was used without the rear polarizer to enable measurement of sample transmission. Macrophase separation transitions were defined as the temperatures $T_t$ corresponding to a 10 % reduction in the initial sample transmittance, according to literature methods. Transitions were reproducible upon repeated cycling, and data analysis was performed on the first heating cycle.

For differential scanning calorimetry (DSC) measurements, bioconjugate solution was loaded into a hermetically sealed aluminum pan. Data for five of the six conjugates were acquired using a TA Instruments Discovery differential scanning calorimeter at the Institute for Soldier Nanotechnologies (ISN). Data for the conjugate GFP(-21)-PNIPAM21k were collected using a Texas Instruments Q-10 differential scanning calorimeter in Professor Swager’s laboratory in the Department of Chemistry at MIT. The sample was equilibrated at 5 °C for 10 minutes, followed by two cycles of ramping to 45 °C at 5 °C/min., holding isothermally for 2 minutes, cooling at 5 °C/min. to 5 °C, and holding isothermally for 2 minutes. The two measurement cycles overlap, and temperature transition values were extracted from the onset point in the initial heating cycle.

Circular dichroism (CD) spectroscopy was performed using an Aviv model 202 CD spectrometer to measure far UV CD spectroscopy between 190 - 250 nm in a 0.1 cm path length quartz cuvette at a scan rate of 6 nm/min. UV-Vis spectrophotometry was performed using a Varian Cary 50 over the wavelength range 190 - 800 nm in a 1.0 cm path length quartz cuvette at a scan
rate of 600 nm/min. GFP(0) was measured in 20 mM Tris-Cl, 250 mM NaCl, pH = 8.0; GFP(-8) and GFP(-21) were measured in 20 mM Tris-Cl, pH = 8.0; and GFP-PNIPAM conjugates were measured in milliQ water. Measurements were made at T = 10 °C.

**Determining PNIPAM Coil Fraction**

From published crystallographic data for the superfolder GFP (PDB 2B3P), the crystallized monomeric molecular weights and crystal unit cell parameters and solvent fraction are tabulated as follows:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GFP(-8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW (kg/mol)</td>
<td>29.1444</td>
</tr>
<tr>
<td>Formula units z</td>
<td>6</td>
</tr>
<tr>
<td>Unit cell length a (Å)</td>
<td>88.46</td>
</tr>
<tr>
<td>Unit cell length b (Å)</td>
<td>88.46</td>
</tr>
<tr>
<td>Unit cell length c (Å)</td>
<td>69.35</td>
</tr>
<tr>
<td>Unit cell angle α (°)</td>
<td>90.00</td>
</tr>
<tr>
<td>Unit cell angle β (°)</td>
<td>90.00</td>
</tr>
<tr>
<td>Unit cell angle γ (°)</td>
<td>120.00</td>
</tr>
<tr>
<td>Crystal solvent content φs (%)</td>
<td>57.92</td>
</tr>
</tbody>
</table>

The protein density is then calculated as follows:

\[
\rho_{\text{protein}} = \left( \frac{1}{1 - \phi_s} \right) \left( \frac{z MW_{\text{protein}}}{NA V_{\text{unit cell}}} \right)
\]

where \( N_A \) is Avogadro's number. The density is calculated to be

\[ \rho_{\text{GFP(-8)}} = 1.47 \text{ g cm}^{-3} \]

and this density value is also used for the other two GFP variants, GFP(0) and GFP(-21). Using a density for PNIPAM\(^4\) of 1.05 g cm\(^{-3}\), the PNIPAM coil fraction can then calculated as below:

\[
\phi_{\text{PNIPAM}} = \frac{MW_{\text{PNIPAM}}}{MW_{\text{PNIPAM}} + MW_{\text{protein}}}
\]

\[
\rho_{\text{PNIPAM}} \quad \rho_{\text{protein}}
\]

\[
\rho_{\text{PNIPAM}} = \frac{MW_{\text{PNIPAM}}}{\rho_{\text{PNIPAM}}} + \frac{MW_{\text{protein}}}{\rho_{\text{protein}}}
\]

250
Figure D-1. Gel permeation chromatography of (a) PNIPAM30k: $M_n = 30,720$ g/mol; $M_w = 33,580$ g/mol; $D = 1.093$ and (b) PNIPAM21k: $M_n = 21,650$ g/mol; $M_w = 22,910$ g/mol; $D = 1.058$. 
GFP(0) DNA sequence
ATGGGACATCACATCACCACTACGATCGGATGCAATCATCTGACGCATGGCAGCTGCGTGGTCCGCTTGTGAACTGATGGTTTCCAAAGGTGAAGAACTGTTCACCGGTGTGGTGCCGATCCTGGTGGAGCTGGATGGCGATGTTAACGGTCACAAATTTTCTGTTCGTGGTGAAGGCGAAGGTGACGCTACTAACGGTAAACTGACTCTGAAATTCATCTGCACCACCGGCAAGCTGCCGGTCCCATGGCCGACCCTGGTAACCACCCTGACCTATGGTGTTCAGTGCTTTTCTCAGTTATCACCACCGCACCAGATGAAACAGCACGATTTCTTTAAGTCCGCGATGCCGGGAAGGCTACGTTCAGGAACGCACCATCTCTTTCAAAGACGACGGCACTTACAAAACTCGCGCCGAAGTGAAATTGAAGGCGACACCCTGGTGAACCGTATTGAACTGAAAGGTATCGACTTCAAAGAGGATGGCAACATCCTGGGTCATAAAGCTGGAGTACAACTTCAACTCTCACAACGTATATATCACCGCAGACAAACGTAAAAATGGCATCAAGGCTAACTTCAAAAATTCGCCACAACGTGAAGGACGGTTCTGTTCAACTGGCTGACCATTACCATCAGCAGAATACTCCTATCGGTGACG GCCCGGTCTCTGCTGCGCACCATTACCTGTCCACTCAGTCTGCCAAAGACCCAAAAGAAAAACGTGACCATATGGTGCTGCTGGAATTTGTGACCGCGGCGGGTATTACCCATGGCATGGATGAGCTGTATAAGTAA

GFP(0) amino acid sequence
MGHHHHHHGSACELMVSKGEELFTGVVPILVELDGDVGKFSVREGEDEGATNGKLFLKICTTGTGKLVPVPWPTLVTTLTYVGQCFSRYPDHMKQHDFFKSAMPEGYQERTISFKDDGTYKRTAEVKFEGETLTVNRIELKIDFKEGKNIPLGHELYSNFNSHVYITADKQKNGIKANFKIRNHVNDGVQPLADHYQQNTPIGDPVPLLDNYLSTQALSSDKPKEKRDMVLLFEVTAGITHGMDELYK

GFP(-8) DNA sequence
ATGGGACATCACATCACCACTACGATCGGATGCAATCATCTGACGCATGGCAGCTGCGTGGTCCGCTTGTGAACTGATGGTTTCCAAAGGTGAAGAACTGTTCACCGGTGTGGTGCCGATCCTGGTGGAGCTGGATGGCGATGTTAACGGTCACAAATTTTCTGTTCGTGGTGAAGGCGAAGGTGACGCTACTAACGGTAAACTGACTCTGAAATTCATCTGCACCACCGGCAAGCTGCCGGTCCCATGGCCGACCCTGGTAACCACCCTGACCTATGGTGTTCAGTGCTTTTCTCAGTTATCACCACCGCACCAGATGAAACAGCACGATTTCTTTAAGTCCGCGATGCCGGGAAGGCTACGTTCAGGAACGCACCATCTCTTTCAAAGACGACGGCACTTACAAAACTCGCGCCGAAGTGAAATTGAAGGCGACACCCTGGTGAACCGTATTGAACTGAAAGGTATCGACTTCAAAGAGGATGGCAACATCCTGGGTCATAAAGCTGGAGTACAACTTCAACTCTCACAACGTATATATCACCGCAGACAAACGTAAAAATGGCATCAAGGCTAACTTCAAAAATTCGCCACAACGTGAAGGACGGTTCTGTTCAACTGGCTGACCATTACCATCAGCAGAATACTCCTATCGGTGACG GCCCGGTCTCTGCTGCGCACCATTACCTGTCCACTCAGTCTGCCAAAGACCCAAAAGAAAAACGTGACCATATGGTGCTGCTGGAATTTGTGACCGCGGCGGGTATTACCCATGGCATGGATGAGCTGTATAAGTAA

GFP(-8) amino acid sequence
MGHHHHHHGSACELMVSKGEELFTGVVPILVELDGDVGKFSVREGEDEGATNGKLFLKICTTGTGKLVPVPWPTLVTTLTYVGQCFSRYPDHMKQHDFFKSAMPEGYQERTISFKDDGTYKRTAEVKFEGETLTVNRIELKIDFKEGKNIPLGHELYSNFNSHVYITADKQKNGIKANFKIRNHVNDGVQPLADHYQQNTPIGDPVPLLDNYLSTQALSSDKPKEKRDMVLLFEVTAGITHGMDELYK

GFP(-21) DNA sequence
ATGGGACATCACATCACCACTACGATCGGATGCAATCATCTGACGCATGGCAGCTGCGTGGTCCGCTTGTGAACTGATGGTTTCCAAAGGTGAAGAACTGTTCACCGGTGTGGTGCCGATCCTGGTGGAGCTGGATGGCGATGTTAACGGTCACAAATTTTCTGTTCGTGGTGAAGGCGAAGGTGACGCTACTAACGGTAAACTGACTCTGAAATTCATCTGCACCACCGGCAAGCTGCCGGTCCCATGGCCGACCCTGGTAACCACCCTGACCTATGGTGTTCAGTGCTTTTCTCAGTTATCACCACCGCACCAGATGAAACAGCACGATTTCTTTAAGTCCGCGATGCCGGGAAGGCTACGTTCAGGAACGCACCATCTCTTTCAAAGACGACGGCACTTACAAAACTCGCGCCGAAGTGAAATTGAAGGCGACACCCTGGTGAACCGTATTGAACTGAAAGGTATCGACTTCAAAGAGGATGGCAACATCCTGGGTCATAAAGCTGGAGTACAACTTCAACTCTCACAACGTATATATCACCGCAGACAAACGTAAAAATGGCATCAAGGCTAACTTCAAAAATTCGCCACAACGTGAAGGACGGTTCTGTTCAACTGGCTGACCATTACCATCAGCAGAATACTCCTATCGGTGACG GCCCGGTCTCTGCTGCGCACCATTACCTGTCCACTCAGTCTGCCAAAGACCCAAAAGAAAAACGTGACCATATGGTGCTGCTGGAATTTGTGACCGCGGCGGGTATTACCCATGGCATGGATGAGCTGTATAAGTAA

GFP(-21) amino acid sequence
MGHHHHHHGSACELMVSKGEELFTGVVPILVELDGDVGKFSVREGEDEGATNGKLFLKICTTGTGKLVPVPWPTLVTTLTYVGQCFSRYPDHMKQHDFFKSAMPEGYQERTISFKDDGTYKRTAEVKFEGETLTVNRIELKIDFKEGKNIPLGHELYSNFNSHVYITADKQKNGIKANFKIRNHVNDGVQPLADHYQQNTPIGDPVPLLDNYLSTQALSSDKPKEKRDMVLLFEVTAGITHGMDELYK

Figure D-2. DNA and amino acid sequences of GFP(0), GFP(-8), and GFP(-21).
Figure D-3. SDS-PAGE of GFP variants. Lanes 1-3 represent GFP(0), GFP(-8), and GFP(-21), respectively.
Figure D-4. Native PAGE of GFP conjugated to (a) PNIPAM21k and (b) PNIPAM30k. In (a), lanes 1-3 represent GFP(O)PNIPAM21k, GFP(-8)PNIPAM21k, and GFP(-21)PNIPAM21k, respectively. In (b), lanes 1-3 represent GFP(-21)PNIPAM30k, GFP(-8)PNIPAM30k, and GFP(0)PNIPAM30k, respectively. In native PAGE, conjugates of the neutral variant, GFP(0), do not migrate toward the positive electrode and are observed to remain in the stacking gel. Native PAGE shows undetectable levels of impurity, suggesting that the bioconjugates are >98% pure. (c) SDS-PAGE of bioconjugates. Lanes 1-6 represent GFP(O)PNIPAM21k, GFP(-8)PNIPAM21k, GFP(-21)PNIPAM21k, GFP(0)PNIPAM30k, GFP(-8)PNIPAM30k, and GFP(-21)PNIPAM30k, respectively.
Figure D-5. UV-Vis absorption spectra for GFP variants and conjugates to PNIPAM21k and PNIPAM30k for (a) GFP(0), (b) GFP(-8), and (c) GFP(-21). The three different GFP variants have different ratios of absorbance at 395 nm and 488 nm, which correspond to protonated and deprotonated tyrosyl hydroxyl groups, respectively. The different electrostatic properties of the GFP variants is observed to affect the relative ratio of absorbance at 395 and 488 nm. The hydroxyl group is part of a network of hydrogen bonds that is also observed to be perturbed upon conjugation. Similar UV-Vis spectra to those of GFP(-8)-PNIPAM has been observed in previously studied EGFP-PNIPAM conjugates.

Figure D-6. UV-Vis absorption spectra for GFP variants and conjugates to PNIPAM21k and PNIPAM30k after rehydration of solid conjugate pellets for (a) GFP(0), (b) GFP(-8), and (c) GFP(-21), showing minimal change in protein function after self-assembly and rehydration.
Figure D-7. Circular dichroism spectroscopy of GFP variants and conjugates to PNIPAM21k and PNIPAM30k for (a) GFP(0), (b) GFP(-8), and (c) GFP(-21) confirm that the secondary structure is preserved in the conjugates. GFP(0) was measured in 20 mM Tris-Cl, 250 mM NaCl, pH = 8.0, and GFP(-8) and GFP(-21) were measured in 20 mM Tris-Cl, pH = 8.0. Conjugates were measured in milliQ water. Measurements were performed at T = 10 °C.

Figure D-8. Circular dichroism spectroscopy of GFP variants and conjugates to PNIPAM21k and PNIPAM30k after rehydration of solid conjugate pellets for (a) GFP(0), (b) GFP(-8), and (c) GFP(-21) confirm minimal change in secondary structure after self-assembly and rehydration. GFP(0) was measured in 20 mM Tris-Cl, 250 mM NaCl, pH = 8.0, and GFP(-8) and GFP(-21) were measured in 20 mM Tris-Cl, pH = 8.0. Conjugates were measured in milliQ water. Measurements were performed at T = 10 °C.
Figure D-9. Radially averaged SAXS patterns for (top) GFP(0)-PNIPAM21k, (middle) GFP(-8)-PNIPAM21k, and (bottom) GFP(-21)-PNIPAM21k. The curves are offset vertically for clarity. The seven curves in each plot increase in temperature upward from 10 to 40 °C in 5 °C intervals.
Figure D-10. Radially averaged SAXS patterns for (top) GFP(0)-PNIPAM30k, (middle) GFP(-8)-PNIPAM30k, and (bottom) (GFP(-21))-PNIPAM30k. The curves are offset vertically for clarity. The seven curves in each plot increase in temperature upward from 10 to 40 °C in 5 °C intervals.
Figure D-11 a-c. DPLS and turbidimetry of GFP(0)-PNIPAM21k at (a) 25 wt.%, (b) 27 wt.%, and (c) 30 wt.%. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-11 d-f. DPLS and turbidimetry of GFP(0)-PNIPAM21k at (d) 33 wt.%, (e) 35 wt.%, and (f) 37 wt.%. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-11 g-i. DPLS and turbidimetry of GFP(0)-PNIPAM21k at (g) 40 wt.%, (h) 45 wt.%, and (i) 50 wt.%. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-12 a-c. DPLS and turbidimetry of GFP(-8)-PNIPAM21k at (a) 25 wt.%, (b) 27 wt.%, and (c) 30 wt.%. The increase in power fraction at concentrations of 25-40 wt.% results from the decrease in transmission. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-12 d-f. DPLS and turbidimetry of GFP(-8)-PNIPAM21k at (d) 33 wt.%, (e) 35 wt.%, and (f) 37 wt.%. The increase in power fraction at concentrations of 25-40 wt.% results from the decrease in transmission. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-12 g-i. DPLS and turbidimetry of GFP(-8)-PNIPAM21k at (g) 40 wt.%, (h) 45 wt.%, and (i) 50 wt.%. The increase in power fraction at concentrations of 25-40 wt.% results from the decrease in transmission. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-13 a-c. DPLS and turbidimetry of GFP(-21)-PNIPAM21k at (a) 25 wt.%, (b) 27 wt.%, and (c) 30 wt.%. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-13 d-f. DPLS and turbidimetry of GFP(-21)-PNIPAM21k at (d) 33 wt.%, (e) 35 wt.%, and (f) 37 wt.%. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-13 g-i. DPLS and turbidimetry of GFP(-21)-PNIPAM21k at (g) 40 wt.%, (h) 45 wt.%, and (i) 50 wt.%. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-14 a-c. DPLS and turbidimetry of GFP(0)-PNIPAM30k at (a) 20 wt.%, (b) 25 wt.%, and (c) 27 wt.%. The sharp drop in transmission results in the increase in power fraction at high temperature. The birefringence results can only be interpreted up to the thermal transition temperature. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-14 d-f. DPLS and turbidimetry of GFP(0)-PNIPAM30k at (d) 30 wt.%, (e) 33 wt.%, and (f) 37 wt.%. The sharp drop in transmission results in the increase in power fraction at high temperature. The birefringence results can only be interpreted up to the thermal transition temperature. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-14 g-h. DPLS and turbidimetry of GFP(0)-PNIPAM30k at (g) 40 wt.% and (h) 50 wt.% The sharp drop in transmission results in the increase in power fraction at high temperature. The birefringence results can only be interpreted up to the thermal transition temperature. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-15 a-c. DPLS and turbidimetry of GFP(-8)-PNIPAM30k at (a) 27 wt.%, (b) 30 wt.%, and (c) 33 wt.%. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-15 d-f. DPLS and turbidimetry of GFP(-8)-PNIPAM30k at (d) 37 wt.%, (e) 40 wt.%, and (f) 43 wt.%. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-15 g-h. DPLS and turbidimetry of GFP(-8)-PNIPAM30k at (g) 47 wt.% and (h) 50 wt.%. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-16 a-c. DPLS and turbidimetry of GFP(-21)-PNIPAM30k at (a) 25 wt.%, (b) 30 wt.%, and (c) 33 wt.%. The moderate increases in power fraction result from the moderate decrease in transmission. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-16 d-f. DPLS and turbidimetry of GFP(-21)-PNIPAM30k at (d) 40 wt.%, (e) 43 wt.%, and (f) 47 wt.%. The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-16 g-h. DPLS and turbidimetry of GFP(-21)-PNIPAM30k at (g) 50 wt.% and (h) 55 wt.% The red curve represents the heating cycle, and the blue curve represents the cooling cycle. The black curve represents the 10-minute equilibration at the end of the heating cycle before cooling.
Figure D-17. DSC of (a) GFP(0)-PNIPAM21k, (b) GFP(0)-PNIPAM30k, (c) GFP(-8)-PNIPAM21k, (d) GFP(-8)-PNIPAM30k, (e) GFP(-21)-PNIPAM21k, and (f) GFP(-21)-PNIPAM30k.
Figure D-18. Cartoon schematic representations of results from ClusPro 2.0\textsuperscript{7,8} of the ten largest clusters representing the most probable configurations for interactions between two GFP molecules as a function of total net charge. The orientation angle between proteins is displayed underneath each configuration and is calculated as the angle between the vectors in the directions of the principal axes aligned with the beta barrel of the two proteins, represented as the black arrows. The cluster size and corresponding energy for these clusters are displayed in Table S1.
Table D-1. Tabulated values for cluster scores from ClusPro 2.0 of the ten largest clusters representing the most probable configurations for interactions between two GFP molecules as a function of total net charge.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Members</th>
<th>Representative</th>
<th>Weighted Score</th>
<th>Members</th>
<th>Representative</th>
<th>Weighted Score</th>
<th>Members</th>
<th>Representative</th>
<th>Weighted Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>134</td>
<td>Center</td>
<td>-505.1</td>
<td>136</td>
<td>Center</td>
<td>-588.5</td>
<td>74</td>
<td>Center</td>
<td>-456.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest energy</td>
<td>-614.1</td>
<td></td>
<td>Lowest energy</td>
<td>-615.0</td>
<td></td>
<td>Lowest energy</td>
<td>-456.3</td>
</tr>
<tr>
<td>2</td>
<td>53</td>
<td>Center</td>
<td>-612.1</td>
<td>70</td>
<td>Center</td>
<td>-566.4</td>
<td>50</td>
<td>Center</td>
<td>-400.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest energy</td>
<td>-628.5</td>
<td></td>
<td>Lowest energy</td>
<td>-566.4</td>
<td></td>
<td>Lowest energy</td>
<td>-464.3</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>Center</td>
<td>-569.3</td>
<td>58</td>
<td>Center</td>
<td>-616.5</td>
<td>42</td>
<td>Center</td>
<td>-425.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest energy</td>
<td>-569.3</td>
<td></td>
<td>Lowest energy</td>
<td>-616.5</td>
<td></td>
<td>Lowest energy</td>
<td>-445.5</td>
</tr>
<tr>
<td>4</td>
<td>47</td>
<td>Center</td>
<td>-541.2</td>
<td>44</td>
<td>Center</td>
<td>-501.7</td>
<td>41</td>
<td>Center</td>
<td>-501.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest energy</td>
<td>-541.2</td>
<td></td>
<td>Lowest energy</td>
<td>-544.3</td>
<td></td>
<td>Lowest energy</td>
<td>-501.9</td>
</tr>
<tr>
<td>5</td>
<td>43</td>
<td>Center</td>
<td>-473.9</td>
<td>37</td>
<td>Center</td>
<td>-495.6</td>
<td>40</td>
<td>Center</td>
<td>-466.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest energy</td>
<td>-546.6</td>
<td></td>
<td>Lowest energy</td>
<td>-557.7</td>
<td></td>
<td>Lowest energy</td>
<td>-466.9</td>
</tr>
<tr>
<td>6</td>
<td>41</td>
<td>Center</td>
<td>-532.2</td>
<td>36</td>
<td>Center</td>
<td>-510.7</td>
<td>39</td>
<td>Center</td>
<td>-417.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest energy</td>
<td>-613.4</td>
<td></td>
<td>Lowest energy</td>
<td>-532.1</td>
<td></td>
<td>Lowest energy</td>
<td>-451.2</td>
</tr>
<tr>
<td>7</td>
<td>33</td>
<td>Center</td>
<td>-504.5</td>
<td>36</td>
<td>Center</td>
<td>-510.8</td>
<td>38</td>
<td>Center</td>
<td>-454.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest energy</td>
<td>-524.6</td>
<td></td>
<td>Lowest energy</td>
<td>-587.0</td>
<td></td>
<td>Lowest energy</td>
<td>-454.2</td>
</tr>
<tr>
<td>8</td>
<td>33</td>
<td>Center</td>
<td>-571.2</td>
<td>36</td>
<td>Center</td>
<td>-531.8</td>
<td>37</td>
<td>Center</td>
<td>-447.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest energy</td>
<td>-571.2</td>
<td></td>
<td>Lowest energy</td>
<td>-601.3</td>
<td></td>
<td>Lowest energy</td>
<td>-471.4</td>
</tr>
<tr>
<td>9</td>
<td>29</td>
<td>Center</td>
<td>-475.8</td>
<td>34</td>
<td>Center</td>
<td>-536.1</td>
<td>33</td>
<td>Center</td>
<td>-407.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest energy</td>
<td>-606.6</td>
<td></td>
<td>Lowest energy</td>
<td>-536.1</td>
<td></td>
<td>Lowest energy</td>
<td>-449.0</td>
</tr>
<tr>
<td>10</td>
<td>29</td>
<td>Center</td>
<td>-539.2</td>
<td>30</td>
<td>Center</td>
<td>-538.1</td>
<td>29</td>
<td>Center</td>
<td>-468.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest energy</td>
<td>-562.2</td>
<td></td>
<td>Lowest energy</td>
<td>-538.1</td>
<td></td>
<td>Lowest energy</td>
<td>-468.5</td>
</tr>
</tbody>
</table>
Figure D-19. SAXS patterns of GFP(-21), $\phi_{\text{PNIPAM}} = 0.59$, at 47 wt.% show lamellar morphologies in water and in the presence of NaCl at 20, 50, and 200 mM. FWHM values of the primary peak from fitting a Lorentzian show that NaCl is not observed to improve the ordering relative to the self-assembly in H$_2$O.
Estimation of counterion concentration for GFP(-21)-PNIPAM

The effect of divalent cations on the self-assembly of GFP(-21)-PNIPAM, $\phi_{\text{PNIPAM}} = 0.60$, is studied at the C$_{\text{ODT}}$ of 47 wt.%. Assuming a basis of 100 g, 47 g of GFP(-21)-PNIPAM, $\phi_{\text{PNIPAM}} = 0.60$ corresponds to $\left(\frac{28,430}{28,430 + 30,720}\right)(47) = 22.6$ g of protein, with GFP(-21) having a MW of 28,430 g/mol and PNIPAM30k having a MW of 30,720 g/mol. This corresponds to $7.94 \times 10^4$ mol of protein and a concentration of $\left(\frac{7.94 \times 10^4 \text{ mol}}{0.053 \text{ L}}\right) = 0.015$ M of GFP(-21).

GFP(-21) has 22 positive residues and 43 negative residues. Assuming that charge neutrality is maintained with 22 molecules of NaCl for 22 positive and negative residues and an additional 21 Na$^+$ ions for the remaining 21 negative residues, the total concentration of counterions is 0.975 M.

References

Appendix E. Supporting Information for Chapter 6

![Gel permeation chromatography graphs](image)

**Figure E-1.** Gel permeation chromatography of (a) PNIPAM26k: $M_n = 26,300 \text{ g/mol}; M_w = 28,400 \text{ g/mol}; D_M = 1.08$ (b) PHPA30k: $M_n = 29,900 \text{ g/mol}; M_w = 36,200 \text{ g/mol}; D_M = 1.21$ (c) POEGA26k: $M_n = 26,400 \text{ g/mol}; M_w = 29,800 \text{ g/mol}; D_M = 1.13$ (d) PEG28k: $M_n = 28,000 \text{ g/mol}; M_w = 29,100 \text{ g/mol}; D_M = 1.04$. PEG28k was purchased from Creative PEGWorks (product number PSB-230, mPEG-MAL, MW 30k).
Figure E-2. (a) SDS-PAGE and (b) native PAGE. Lanes 1-5 represent (1) mCherry, (2) mChPEG28, (3) mChPN26, (4) mChPOE26, and (5) mChPHA30.
Figure E-3. UV-Vis spectroscopy of mCherry and mCherry-polymer bioconjugates after purification and buffer exchange into 10 mM Tris-Cl, pH = 8.0 and D$_2$O, respectively, showing that the optical functionality of the protein is retained upon conjugation. The ratio of the absorbance at 586 nm to 280 nm ($A_{586}/A_{280}$) for mCherry and the bioconjugates (and the percentage of the ratio relative to mCherry) are as follows: mCherry: 1.82; mCherry-b-PEG: 1.71 (94%); mCherry-b-PHPA: 1.54 (85%); mCherry-b-POEGA: 1.49 (82%); mCherry-b-PNIPAM: 1.43 (79%). EMP has a broad absorption peak at approximately 310 nm that overlaps with the protein absorption at 280 nm. The molar absorptivity of EMP at 280 nm was measured to be \( \varepsilon_{\text{EMP,280 nm}} = 1,830 \pm 194 \text{ M}^{-1} \text{ cm}^{-1} \). With the molar absorptivity of mCherry at 280 nm being \( \varepsilon_{\text{mCh,280 nm}} = 34,380 \text{ M}^{-1} \text{ cm}^{-1} \), the absorbance at 280 nm due to the protein only is approximately 94.95% of the total absorbance of the bioconjugate. Therefore, a corrective multiplicative factor of \( 1/0.9495 = 1.053 \) is applied to the ratio of the absorbance at 586 nm to 280 nm ($A_{586}/A_{280}$) for the bioconjugates mCherry-b-PNIPAM, mCherry-b-PHPA, and mCherry-b-POEGA that contain the EMP chain-transfer agent; this results in the following corrected $A_{586}/A_{280}$ ratios relative to mCherry: mCherry-b-PHPA: 89%, mCherry-b-POEGA: 86%, mCherry-b-PNIPAM: 83%.
Guinier Analysis

Guinier analysis was performed by using weighted linear least-squares regression to fit the slope and intercept of the following equation to determine the particle radius of gyration, \( R_G \), and zero-angle scattering intensity, \( I(Q=0) \), respectively:

\[
\log I(Q) = \log I(0) - \frac{R_G^2}{3} Q^2, \quad \text{in the limit } Q R_G << 1
\]
Figure E-4. (Left column) shows the Guinier analysis fits and (right column) shows the bracketed data range used for Guinier analysis (dashed vertical lines) and the zero-angle scattering intensity $I(0)$ (horizontal line with dashed lines representing statistical error bounds) for (a) mCherry, (b) PN26, (c) PH30, (d) POE26, (e) PEG28, (f) mChPN26, (g) mChPH30, (h) mChPOE26, and (i) mChPEG28. The Q-range over which Guinier analysis was performed was chosen such that $Q_{\text{max}}R_G < 0.8$. Because the Guinier regime is not well established for mChPOE26 due to the upturn in intensity at low Q, Guinier analysis resulted in $Q_{\text{max}}R_G = 0.94$. The statistical error bounds correspond to 95% confidence intervals.
Figure E-5. Model fitting results from fitting the form factor of a homogeneous cylinder to mCherry SANS form factor data. The fitting results for the radius and length of the cylinder, 15.06 Å and 47.36 Å, respectively, agree well with values estimated from the crystal structure (PDB 2H5Q) of mCherry. In the model fitting analysis, a contrast factor (CT) parameter that accounts for the sample concentration and particle volume and a fitting parameter for the incoherent background (B) are included: \( I_{\text{cyl}}(Q) = CT \times P_{\text{cyl}}(Q, R, L) + B \).
Figure E-6. Model fitting results from fitting the form factor for a worm-like chain model to SANS data of (a) PNIPAM 26k, (b) PHPA 30k, (c) POEGA 26k, and (d) PEG 28k. In the model fitting analysis, a contrast factor (CT) parameter that accounts for the sample concentration and particle volume and a fitting parameter for the incoherent background (B) are included:

\[ I_{WLC}(Q) = CT \times P_{WLC}(Q, L, a_x, R_{CS}) + B. \]

The 95% confidence interval for the zero-angle scattering intensity is determined from Guinier analysis. Data points at low Q that trend above 10% of the upper bound are excluded from model fitting. Because the Guinier region for PEG is not established, Q values below which there appears to be a clear upward trend in scattering intensity are discarded in the analysis.
Calculation of Polymer Contour Length

The number average molecular weight of the polymer measured by GPC and the molecular weight of the repeat unit is used to determine the number of repeat units \( N_{\text{repeat}} \), rounded to the nearest integer. The number of repeat units calculated for the four polymers are the following: PNIPAM (232), PHPA (230), POEGA (175), and PEG (635). The repeat unit for PNIPAM, PHPA, and POEGA includes two sp³ hybridized carbon-carbon bonds. Using a value of 1.54 Å for the carbon-carbon bond length, the contour lengths of PNIPAM, PHPA, and POEGA are estimated by

\[
L = \sin \left( \frac{109.5^\circ}{180^\circ} \pi \right) \times 1.54 \ \text{Å} \times 2 \times N_{\text{repeat}}
\]

The repeat unit for PEG has two carbon-carbon single bonds and a carbon-oxygen single bond. The contour length for PEG is thus estimated similarly as

\[
L = \left[ \sin \left( \frac{109.5^\circ}{180^\circ} \pi \right) \times 1.54 \ \text{Å} + \sin \left( \frac{112^\circ}{180^\circ} \pi \right) \times 1.43 \ \text{Å} \times 2 \right] \times N_{\text{repeat}}
\]

Calculation of mCherry Scattering Length Density (SLD)

SANS contrast matching of mCherry was performed in D\(_2\)O/H\(_2\)O compositions of 100/0, 90/10, 80/20, 70/30, and 10/90. The values of \( l(0) \) were determined from Guinier analysis, and the contrast match point was determined from performing a linear fit to \( \sqrt{l(0)} \) as a function of %H\(_2\)O, finding that the match point occurs at a solvent composition of 43.2% D\(_2\)O and 56.8% H\(_2\)O (Figure E-7).
Figure E-7. Determination of the contrast point of mCherry by linear regression to $\sqrt{i(0)}$ as a function of $\%$H$_2$O.

The values of the scattering length densities for H$_2$O, D$_2$O, fully hydrogenated mCherry ($h$-mCherry), and fully deuterated mCherry ($d$-mCherry) were obtained from the online NIST scattering length density calculator and values are summarized in Table S1.

Table E-1. Summary of chemical composition and scattering length density values.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Chemical composition</th>
<th>MW (g/mol)</th>
<th>Density (g cm$^{-3}$)</th>
<th>SLD ($\text{Å}^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H$_2$O</td>
<td>H$_2$O</td>
<td>18.0146</td>
<td>1.00</td>
<td>-5.61e-7</td>
</tr>
<tr>
<td>D$_2$O</td>
<td>D$_2$O</td>
<td>20.0302</td>
<td>1.11</td>
<td>6.39e-6</td>
</tr>
<tr>
<td>$h$-mCherry</td>
<td>C$<em>{250}$H$</em>{913}$N$<em>{341}$O$</em>{378}$S$_{11}$</td>
<td>28120.411</td>
<td>1.35</td>
<td>1.90e-6</td>
</tr>
<tr>
<td>$d$-mCherry</td>
<td>C$<em>{250}$D$</em>{913}$N$<em>{341}$O$</em>{378}$S$_{11}$</td>
<td>30050.348</td>
<td>1.44</td>
<td>7.65e-6</td>
</tr>
</tbody>
</table>

The SLD of 43.2% D$_2$O / 56.8% H$_2$O is calculated to be 2.44e-6 Å$^2$. This SLD is matched by a mixture of $h$-mCherry and $d$-mCherry at a volume fraction ratio of $f_{h\text{-mCherry}} = 0.91$ and $f_{d\text{-mCherry}} = 0.09$. This is equivalent to approximately 9.4% of the mCherry hydrogen atoms exchanging with deuterium. Assuming that the energy of exchanging hydrogen and deuterium atoms is similar, then dividing this value by the volume fraction of D$_2$O at the match point, 0.432, gives a final
estimate of the percentage of exchangeable hydrogen atoms in mCherry to be 21.8%. Therefore, in 100% D₂O, the scattering length density of mCherry is estimated to be 3.16 x 10⁻⁶ Å².

Table E-2. Estimated values of scattering length densities of polymers from Guinier analysis.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>conc. (mg/mL)</th>
<th>ρpoly (x 10⁻⁶ Å²)</th>
<th>Vpolymer (x 10⁻¹ Å⁴)</th>
<th>MW (g/mol)</th>
<th>(N/V) x Vpolymer (cm⁻¹)</th>
<th>I(0) (cm⁻¹)</th>
<th>(ρpoly - ρD₂O)² (x 10⁻¹¹ Å⁻⁴)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNIPAM</td>
<td>8</td>
<td>0.55</td>
<td>4.1</td>
<td>26300</td>
<td>0.00744</td>
<td>0.50121</td>
<td>1.66</td>
</tr>
<tr>
<td>PHPA</td>
<td>8</td>
<td>0.55</td>
<td>4.2</td>
<td>29900</td>
<td>0.00672</td>
<td>0.84686</td>
<td>3.02</td>
</tr>
<tr>
<td>POEGA</td>
<td>8</td>
<td>0.55</td>
<td>4.1</td>
<td>26400</td>
<td>0.00742</td>
<td>0.87022</td>
<td>2.88</td>
</tr>
<tr>
<td>PEG</td>
<td>8</td>
<td>0.53</td>
<td>3.8</td>
<td>28000</td>
<td>0.00655</td>
<td>0.46323</td>
<td>1.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Molecule</th>
<th>ρpoly (x 10⁻⁶ Å²)</th>
<th>NIST calculator ρpoly (x 10⁻⁶ Å²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNIPAM</td>
<td>2.32</td>
<td>0.777</td>
</tr>
<tr>
<td>PHPA</td>
<td>0.897</td>
<td>1.07</td>
</tr>
<tr>
<td>POEGA</td>
<td>1.03</td>
<td>1.05</td>
</tr>
<tr>
<td>PEG</td>
<td>2.08</td>
<td>0.679</td>
</tr>
</tbody>
</table>

From fitting a homogeneous cylinder model to the scattering of mCherry (Figure E-5), the best fit radius and length values for mCherry are used to calculate the volume of mCherry, VₘCh = 33723.79 Å³. The polymer volume fractions of the mCherry-polymer conjugates are used to estimate the homopolymer volumes. The scattering length density of the polymer is calculated according to the following equation:

\[ I(Q = 0) = \left( \frac{N}{V} \right) \times V_{polymer}^2 \times (\rho_{polymer} - \rho_{D₂O})^2 \]

\( \frac{N}{V} \) represents the number concentration of the polymer solution.
Figure E-8. Log-log plots of nonlinear least-squares regression of cylinder-polymer scattering function (black lines) to experimental mCherry-polymer bioconjugate SANS data (open circles) for (a) mChPN26, (b) mChHPA30, (c) mChPOE26, and (d) mChPEG28. The dashed vertical lines denote data points at low Q that are omitted in the model fitting. Slight aggregation is observed in mChPOE26 and mChPEG28, represented by a slight upturn in intensity at low Q.
Figure E-9. Form factor time correlation functions as a function of Q for cylinder-coil bead model with (a) 28, (b) 45, (c) 46, and (d) 64 polymer beads attached to represent mChPOEGA, mChPNIPAM, mChPHPA, and mChPEG, respectively. Each curve corresponds to a different experimental Q value ranging between 0.03478 nm\(^{-1}\) and 5.18 nm\(^{-1}\). Coarse-grained simulation configurations are averaged every 1000\(\tau\), an interval greater than the correlation time of the coarse-grained conjugate models.
Figure E-10. Plots of the reduced $\chi^2$ values from fitting the coarse-grained form factor, $I_{\text{model}}(Q) = C T \times P_{\text{model}}(Q) + B$, to SANS form factor data for (a) PNIPAM26k, (b) PHPA30k, (c) POEGA26k, and (d) PEG28k to determine the number of coarse-grained beads $N_{\text{coil}}$ to represent each homopolymer for constructing the coarse-grained cylinder-coil bead model. The coarse-grained model scattering intensity is smeared by the NG-7 SANS instrument resolution function in the model fitting procedure.
Figure E-11. Plots of the SANS data and the corresponding best fits for $N_{\text{coil}}$ beads that minimizes the reduced $\chi^2$ values in Figure E-10. (a) PNIPAM26k, (b) PHPA30k, (c) POEGA26k, and (d) PEG28k. Similar to the model fitting described above for fitting to mCherry and polymer SANS data, a contrast factor ($CT$) and an incoherent background ($B$) are also incorporated into the fitting analysis. Data for PNIPAM, PHPA, and POEGA show a slight increase in intensity for a few data points at low $Q$ values due to potential small aggregates. The upturn in scattering intensity for PEG at low $Q$ clearly indicates aggregation, which has been reported previously in the literature for PEG chains with hydrophobic end groups. Model fitting is performed over the same range as in Figure E-6.
Zero-Angle Scattering Analysis

The zero-angle scattering intensity is given by

\[ I(Q = 0) = \frac{N}{V} \left( \rho_{part} - \rho_s \right)^2 V_{part}^2, \]

where \( N/V \) is the particle concentration in number density. The volume of the polymers are calculated from estimating the volume of the protein assuming the protein is a homogeneous cylinder of length \( L \) and radius \( R \) determined from model fitting (Figure E-5) and knowing the polymer volume fractions of the protein-polymer conjugates and the densities of mCherry and the polymers: \( \rho_{mCh} = 1.35 \text{ g cm}^{-3} \), \( \rho_{PNIPAM} = 1.05 \text{ g cm}^{-3} \), \( \rho_{PHPA} = 1.16 \text{ g cm}^{-3} \), \( \rho_{POEGA} = 1.05 \text{ g cm}^{-3} \), and \( \rho_{PEG} = 1.2 \text{ g cm}^{-3} \). The ratios of the scattering length contrast of polymer to mCherry, estimated from

\[ \frac{(\rho_{poly} - \rho_s)V_{poly}}{(\rho_{mCh} - \rho_s)V_{mCh}} = \frac{M_{poly}}{M_{mCh}}, \]

are calculated to be 1.52, 2.10, 2.00, and 1.50 for PNIPAM, PHPA, POEGA, and PEG, respectively. These ratios are larger than the ratios obtained from model fitting (Figure 6-4). Their use decreases the quality of fits to the data, shown in Figure E-12.
Figure E-12. Nonlinear least-squares regression of coarse-grained cylinder-coil form factors (black lines) to experimental mCherry-polymer bioconjugate SANS data (open circles) for (a) mChPN26, (b) mChPHPA30, (c) mChPOE26, and (d) mChPEG28. The dashed vertical lines denote data points at low Q that are omitted in the model fitting. Compared to Figure 6-4 in the main text, the fitting results here are worse, as the contrast factor (which accounts for the conjugate concentration) and the incoherent background are the only two fitting parameters used, whereas in the fitting results shown in Figure 6-4, the scattering length contrast of the coil beads is an additional fitting parameter. The coarse-grained model scattering intensity is smeared by the NG-7 SANS instrument resolution function in the model fitting procedure.
Figure E-13. (a) Truncated and shifted Lennard-Jones potential at different cutoff distances. The cutoff distances shown here correspond to the same cylinder-coil cutoff distances at which the simulations in Figure 6-4b of the manuscript are performed. (b) Second virial coefficient between a cylinder and a coil bead calculated as a function of cutoff distance in the truncated and shifted Lennard-Jones potential in terms of $\sigma_{\text{cyl-coil}}$. These values were calculated for a coarse-grained cylinder-coil model with 46 coil beads attached to the 28-bead cylinder model. The coil volume fraction is set to be 0.5, and the ratio of the coil/cylinder bead density is taken to be $1/1.35 = 0.741$. The cylinder and coil bead size and mass are then $\sigma_{\text{cyl}} = 1$, $m_{\text{cyl}} = 1$ and $\sigma_{\text{coil}} = 0.8475$, $m_{\text{coil}} = 0.4508$. 
Figure E-14. Average form factors of a coarse-grained cylinder-coil bead model with 46 coil beads at a coil volume fraction of 0.5, simulated using a truncated and shifted Lennard-Jones potential with cutoff radii of (a) $2^{1/6} \sigma_{\text{cyl-coil}}$, (b) $1.6041 \sigma_{\text{cyl-coil}}$, (c) $1.6840 \sigma_{\text{cyl-coil}}$, and (d) $2.5 \sigma_{\text{cyl-coil}}$. Increasing the interaction strength between cylinder and coil beads at cutoff distances greater than $2^{1/6} \sigma_{\text{cyl-coil}}$ causes the coil beads to wrap around the cylinder beads due to a transition from net repulsive to net attractive interactions. This is observed in the form factor shifting to resemble more closely to that of a homogeneous sphere.

References