Elastin-Like Polypeptide (ELP) Tags for Self-Assembly and High-Throughput Processing of Functional Protein Materials

by

Carolyn Elaine Mills

B. S., University of California, Santa Barbara, 2013
M. S. Chemical Engineering Practice, Massachusetts Institute of Technology, 2016

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

Doctor of Philosophy

at the

Massachusetts Institute of Technology

June 2019

Signature of Author: Signature redacted
Department of Chemical Engineering
May 28, 2019

Certified by: Signature redacted
Bradley D. Olsen
Associate Professor of Chemical Engineering
Thesis Supervisor

Accepted by: Signature redacted
Patrick S. Doyle
Robert T. Haslam (1911) Professor
Chairman, Committee for Graduate Students

Elastin-Like Polypeptide (ELP) Tags for Self-Assembly and High-Throughput Processing of Functional Protein Materials

Carolyn Elaine Mills

Submitted to the Department of Chemical Engineering on May 28, 2019 in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Chemical Engineering

Abstract

The diverse recognition and catalytic capabilities of globular proteins makes these biomolecules promising candidates in a broad array of applications, including industrial production of commodity chemicals, point-of-care diagnostics, and therapeutics. Self-assembly of globular protein-polymer bioconjugates into nanostructured materials is an attractive protein immobilization strategy that allows for both high protein packing density and control over protein orientation in the material. However, challenges associated with protein-polymer bioconjugate preparation limit the use of these materials in high-throughput processes. This thesis focuses on overcoming this challenge by replacing the polymer block of these bioconjugate materials with a genetically fused elastin-like polypeptide (ELP) tag.

The first part of this thesis explores the effects of ELP charge and hydrophobicity on the self-assembly of ELP-mCherry fusion proteins in concentrated solution. Concentrated solution characterization of fusion protein self-assembly showed that the addition of charge to the ELP block decreases the propensity for fusion protein self-assembly. Subsequent dilute solution measurements on the ELPs and ELP/mCherry blends revealed that these ELPs behave similarly to analogous charged polymers, but do not complex with mCherry in dilute solution. This combination of results leads to two main conclusions concerning the self-assembly of ELP-mCherry fusion proteins. First, the addition of charge to the ELP block decreases the preopensity for concentrated solution self-assembly because it reduces the effective repulsion between ELP and mCherry blocks by reducing charge cohesion asymmetry between the two protein blocks. Second, for fusions containing a negatively charged ELP block, repulsion between negatively charged ELPs further weakens self-assembly.

The second part of this thesis focused on development of a platform for high-throughput preparation of ELP-fusion materials. One key pitfall of existing ELP-based purification strategies (which can be implemented in high-throughput formats) is that they do not permit control over final protein solution salinity. To overcome this challenge, the existence of ELP cononsolvency in water/alcohol solutions was investigated. This resulted in the discovery and first reports of ELP cononsolvency, as well as the report of upper-critical solution temperature (UCST) transitions of ELPs under certain solvent conditions. Application of ethanol-induced ELP precipitation to desalting of protein materials was then investigated using a model fusion protein of ELP and superfolder green fluorescent protein (ELP-sfGFP). A combination of sodium chloride- and ethanol-induced precipitations was used to reliably purify and desalt ELP-sfGFP in a well-plate format. Finally, this purification procedure was applied to an ELP fusion construct capable of incorporating a library of Sso7d binding proteins with variate streptavidin binding affinity. Two Sso7d variants incorporated into this construct were found to have measurably different binding affinities in both dilute solution and in self-assembled films, demonstrating the capacity of this system to screen self-assembled materials for desired functional protein properties.

Thesis Supervisor: Bradley D. Olsen, Associate Professor of Chemical Engineering

Acknowledgements

There is an old proverb that says it takes a village to raise a child. With that same sentiment, I will say that it took an incredible community of people to accomplish everything I was able to do in this thesis. As such, I have many people to thank.

First, I would like to thank my advisor, Professor Bradley Olsen, for his support and guidance over these last six years. Brad has provided feedback and helped me grow as a scientist, a leader, and a human being. I could write a laundry list of the different things I have learned from Brad over the years, but I think what many of these things have in common is that he continually provided me not only with challenges, but also the faith that I would be able to handle them. While I didn't always believe him at first, seeing these challenges through has made me confident that I will be able to weather any future storm I come across. I would also like to thank Brad for the respectful culture he has built in his lab, and for creating a collaborative space for everyone to work in. As a woman in science and engineering, I can say that I have felt the effects of sexism in some spaces over the years. I am grateful that the space I came to work in everyday was not one of those. Finally, I want to thank Brad for his openness to change and new ideas. I respect so much that there is never an attitude of "this is just how it's always been done, so that's it" from him.

I would also like to thank my thesis committee for their support over the years. Professor Robert Cohen sparked my interest in polymer physical chemistry, and laid the groundwork for the knowledge I have built up in that area over the last four and a half years. Professor Hadley Sikes has been an amazing resource for all my protein engineering questions, and has been an enthusiastic supporter of the collaboration that made Chapter 6 of this thesis possible. Professor Alfredo Alexander-Katz has been an excellent resource for my polymer/protein physics questions. I need to thank every member of my committee for their research questions; the insight they have provided has always improved the quality of my work. Finally, I would like to thank my committee for their unwavering support over the years, both for the work I do as well as for my long-term career goals.

There are so many thanks to be extended to the past and present members of the Olsen lab. Mitchell, Shengchang, Dongsook, Chris, and Charlotte, I thank all of you for welcoming me into the lab when I started. I must extend a special thanks to Xuehui, Minkyu, Matt, and Guokui, who were also in the lab when I started, and who generously took the time to train me and answer my research questions. I would also like to thank Allie, who started in the lab shortly after me, for both mentoring me while she was in the Olsen lab, but also for continuing to offer advice and support after leaving. For those that sat near me and next to me in the offices over the years, I thank you for tolerating (or even appreciating!) all the bad jokes and terrible puns—Michelle (S. and C.), Helen, Aaron, Reggie, Danielle and Ameya. I extend special thanks to Michelle S. for her emotional support over the years, and for always making the time to listen when I needed it. Aaron, for sharing pink bench with me for many years, being an endless source of bad puns, and offering technical assistance whenever I needed to jerry-rig things in the lab. Reggie for advice and thoughts on everything from classes, to cloning problems, to statistics questions. Helen for all the random help in the lab, as well as always lending an ear, whether it be to talk out a scientific idea or venting about broken lab equipment. Danielle for her mentorship and advice. I would be remiss if I did not thank Justin for his help on the work done in Chapter 6, as well as the amazing job he has done as a member of the safety team over the years. To everyone else—Hursh, Irina, Daphne, Brian, Ameya, Vivi, Tzyy-Shyang, Celestine, Andreia, Fernando, Rogerio, Diego, Haley, Sieun, Nari, Angie, Yun-Jung, Wontae, Takuya, and Sarah (and anyone else whose name escapes me right

now)—every person who has been in the Olsen lab with me over the past six years has been a constant source of scientific discussion, support and laughter. I am grateful for how brilliant, generous, and fun everyone in the lab has been, and I know I have learned just as much (if not more) from my labmates as I did from my classes in graduate school. Thank you for making the lab a lovely place to be.

I am also incredibly grateful for the two undergraduates I had the opportunity to work with during my time at MIT—Zachary Michaud and Erika Ding. Both have put in many hours to make the work presented in this thesis possible. Working with both of them has made me a better scientist, and I have learned so much from the mentorship experience. I thank them both for their patience with the scientific process (and me), and for the opportunity to be a part of their learning experience at MIT.

The research presented in this thesis was supported by both the National Science Foundation's Graduate Research Fellowship Program and a grant from the National Science Foundation's Department of Materials Research (DMR-1253306). The small-angle x-ray scattering experiments presented in this work were performed at the Advanced Light Source at Lawrence Berkeley National Lab, and at the National Synchrotron Light Source II at Brookhaven National Lab, both of which are supported by the Department of Energy. The small-angle neutron scattering experiments presented in this work were performed at the Spallation Neutron Source at Oak Ridge National Laboratory, which is supported by the Department of Defense. Access to this beamtime, as well as the assistance from the beamline scientists at each beamline we used was essential to the success of this thesis. These scientists include Dr. William Heller, Dr. Wei-Ren Chen, Dr. Carrie Gao, and Dr. Changwoo Do at Oak Ridge, Dr. Lin Yang, Dr. Masa Fukuto, and Dr. Ruipeng Li at Brookhaven, and Dr. Eric Schaible and Dr. Chenhui Zhu at LBNL.

I have also had the opportunity to work with and spend time with many wonderful people outside the lab over the course of this PhD. In general, I would like to thank all my colleagues in the department who have provided their expertise and thoughts whenever I had questions. I must extend a huge thanks to Eric Miller, who was instrumental in getting the work presented in Chapter 6 off the ground. Both his experimental help and scientific input have been invaluable. I would also like to thank Kameron Conforti, both for his amazing friendship, jokes and playlists over the years, and his input on the best ways to evaluate solution salinity, which was essential for the work presented in Chapter 5. Outside of research-related things, I thank all the people I met at MIT and through people at MIT who have made Cambridge feel like home—Aaron, Kenny, Aurora, Kameron, Brittany, Naveed, Sam, Ben, Leslie, Orpheus, and Matt, among others.

Zach Sherman has been such an important part of my time at MIT, being my best friend here for the past six years, and my roommate for the past four. I could easily write an entire thesis chapter on all the things I appreciate about our friendship and Zach as a person. I am so grateful for everything I have learned from him in every facet of life, from scientific concepts to new music to balancing hard work with actual relaxing downtime. I thank Zach for filling my time at MIT with laughter, love, and unwavering support. He has played such a critical role in shaping the positive experience I have had throughout my PhD.

I would also like to thank my boyfriend, Andrew, for all his support throughout my PhD. In addition to being an incredible, fun, loving person, he has always been there to lift me up in whatever way I needed.

Last, but certainly not least, I would like to thank my family, and, specifically, my parents. They have been a constant source of love and support, not only during my PhD, but throughout each and every day of my life. They have always been proud of me not for my achievements alone, but for who I am when I achieve. They inspire me to work hard and be kind, and I hope to make them proud.

Table of Contents

List of Figures

List of Tables

Chapter 1 : Introduction

1.1 Globular Proteins

Globular proteins constitute a class of biomolecules whose 3-D folded structure enables highly specific catalytic or binding functionalities. Enzymes, or biocatalytic globular proteins, are capable of performing reactions with high turnover frequencies (as characterized by k_{cat}) while also retaining high substrate specificity (characterized by k_{cat}/K_m, where K_m is the Michaelis constant).[1]. Antibodies, a popular example of globular binding proteins, are capable of recognizing a diversity of antigens with high specificity (as characterized by the binding constant, K_D).[2]

The broad array of chemistries accessible by enzymatic catalysis makes these globular proteins relevant in many diverse applications. Industrially, enzymes are already used in the production of a variety of chemicals. In the food industry, for example, enzymes like glucose isomerase and glucoamylase are used in the production of sugars. [3-4] In commodity chemical production, nitrile hydratase is used in the production of acrylamide. [5-6] Enzymes have also garnered great interest as catalysts in the conversion of lignocellulosic waste streams to fuels and chemicals. [7-8] In the defense sector, enzymes like diisopropyl-fluorophosphatase (DFPase), mammalian serum paraoxonase 1 (PON1), and organophosphorus hydrolase (OPH) have been investigated as tools for nerve agent decontamination and prophylactics.[9-12] Enzymatic activity is also utilized as a metric in sensing applications. Perhaps the most widely known instance of this is measurement of glucose levels *via* glucose oxidase activity,[13] a common technology in home glucose testing strips.

Globular proteins with the ability to recognize and bind specifically to peptide motifs, other globular proteins, or other non-protein moieties (such as DNA) are also of high value in biotherapeutic and biosensing technologies. For example, immunoglobulins, or antibodies, are already used commercially in therapeutics and diagnostics and are the subject of ongoing research for these same applications.[14-19] In an effort to overcome some of the cost and stability issues associated with antibodies, researchers have also identified and engineered other functional binding proteins, including single-domain antibodies, single-chain variable fragments, and DNA-binding scaffolds from hypertherophilic microbes.[20-22] Other types of globular binding proteins have found uses as immobilization tags or as cross-linking groups in protein hydrogels. Examples of this include the cellulose-binding domain (CBD), which can be used to immobilize a fused protein partner to a cellulose substrate, and the SpyCatcher domain, which can be used to form

covalent linkages in protein hydrogels (or other environments) upon recognition of the SpyTag domain. [23-26]

Many of these globular proteins have been well-characterized with respect to their solution behavior; however, the possibility of immobilizing these functional proteins has garnered interest as a promising strategy for stabilizing globular proteins as well as potentially improving their function. Similar to the advantages of using a heterogeneous catalyst, enzyme immobilization in particular presents the opportunity for continuous production modes, improved enzyme stability, and the elimination of a purification step (removal of the enzyme from the product stream). [27] Enzyme immobilization is already utilized in certain industrial scale processes catalyzed by enzymes like glucose isomerase and glucoamylase. [28-29] Immobilization of binding proteins has also found widespread use in applications like enzyme-linked immunosorbent assay (ELISA).[30] These examples highlight the potential of immobilized globular proteins.

1.2 Immobilization of Globular Proteins

Protein immobilization can be achieved *via* a variety of strategies, and there are many factors to consider when selecting an immobilization strategy. Options include routes that generate a monolayer of protein as well as those that permit immobilization of proteins in a three dimensional space. When selecting between these strategies or considering development of a new strategy, several engineering features should be considered. First and foremost, the immobilization must not damage the fold or function of the globular protein. Control over protein orientation in the immobilized state is also desirable, as studies have shown that the orientation of a protein in the immobilized state can impact its functionality. [31] The spatial organization of the protein in the immobilized state may also confer advantages in some end-use applications. Finally, because globular proteins are quite large compared to small-molecule catalysts and binding moieties, high protein density in the immobilized material is an important requirement

1.2.1 Monolayer Immobilization of Globular Proteins

Construction of a protein monolayer on a surface or support particle/matrix is one mode of achieving functional protein immobilization, and, in some cases, spatial and orientation control. This can be achieved by leveraging physical interactions between the protein and surface of interest, or by forming a covalent linkage between the two. Many types of protein physisorption exist. If the protein carries sufficient charge, it can be adsorbed onto a (typically positively)

charged surface, such as a gold nanoparticle, and held there by electrostatics. [28, 32] Similar adsorption can be achieved by leveraging hydrophobic interactions, hydrogen bonding, and even van der Waals interactions. [33] In all of these cases, the interactions governing the immobilization of the protein are relatively nonspecific. As a result, these sorts of approaches typically do not permit control over the orientation of the immobilized protein. However, they are simple to execute in that they do not require any modification to the protein of interest. Binding of proteins to the surface can also be mediated by fusion of a protein binding domain (such as CBD or streptavidin/biotin) or conjugation of a DNA sequence to the target protein. [23, 34] While these approaches do require genetic or covalent modification of the target protein, they afford much more control over protein orientation in the monolayer formed. Pretreatment of the substrate by lithographic techniques or block copolymer self-assembly can also enable spatial control over proteins immobilized in this way. [35]

Proteins can also be linked to surfaces *via* covalent bonds. Non-specific covalent linkage of proteins to surfaces most often utilizes surface groups that are able to react with lysine residues (amines) on the surface of the protein. [31, 36-38] Because there are typically multiple lysine residues on the surface of a given protein, this approach does not yield a substantial degree of control over protein orientation. However, it is advantageous in that it works with almost all proteins with no need for protein engineering or modification. To better control orientation, genetic engineering of a unique surface reactive group into the protein can be employed. Several reactive groups have been used for this purpose, including cysteines and non-canonical amino acids. [31, 36-37] Other approaches include posttranslational modification of the protein to append a clickable group to the protein as well as a recognition tag enabling Sortase A mediated ligation of the protein to a surface. [39-40] In these cases, control over protein spatial organization can be achieved by applying techniques like photolithography to template the groups that the protein adsorbs to.

Monolayer immobilization of proteins onto synthetic supports is a powerful tool in protein immobilization. Control over protein spatial organization and orientation can be achieved by selecting the proper technique; however, because protein can only be packed in two dimensional space, the overall packing density of proteins immobilized in this fashion is inherently limited. This limitation can be overcome with techniques that immobilize proteins in three dimensional space.

1.2.2 Three Dimensional Immobilization of Globular Proteins

Globular protein immobilization can also be achieved in three dimensional formats, where the globular protein is incorporated into multiple layers of a given material. Because the protein incorporates into multiple layers of the material, much higher packing densities of proteins are possible. As with monolayers, the interactions utilized to incorporate proteins into three dimensional materials include both physical and covalent. In layer-by-layer (LBL) assembly, for example, the electrostatic interaction (or other physical interaction, such as hydrogen bonding) between the protein of interest and a polymer or other macromolecule is used to build up alternating layers of globular protein and the other interacting molecule.[41-42] LBL permits high packing density of the protein and spatial organization into protein-rich domains; however, there is little control of the protein orientation within these domains. Proteins can be covalently incorporated into polyurethane foams (PUFs), as urea-type linkages can be formed between the lysine residues on the protein surface and the isocyanate groups in the polyurethane prepolymer; this approach has been demonstrated to be effective for a wide variety of enzymes, including organophosphate hydrolase. [43-44] While varying levels of protein incorporation are possible using this approach, protein orientation within the polyurethane matrix is not well-controlled.

Protein incorporation into a three dimensional material can also be achieved using self-assembly strategies. One way to do this is using a co-assembly approach, in which self-assembly of a non-protein material is used to provide spatial organization that the protein can then integrate into. [45] For example, self-assembled block copolymers have been co-assembled with electrochemically active proteins to produce biocatalytic films that both preserved protein function and permitted electron transfer within the material. [46] Co-assembly of proteins with lipid/polymer systems has also been used to immobilize and stabilize proteins, including reconstituted membrane proteins, which do not fold in the absence of the co-assembled materials.[47-48] Co-assembly approaches are attractive in that they can enable stabilization of the protein of interest (as in the membrane protein example), and that they typically do not require engineering or modification of the protein. However, the success of co-assembly techniques in providing control over the protein orientation is highly dependent on the specific protein used, and the ability of the co-assembling material to guide protein orientation. Self-assembled protein materials can also be formed by bioconjugating a polymer, such as poly(*N*-isopropyl acrylamide) (PNIPAM) to a specific site on the protein of interest. These protein-polymer bioconjugates can then be cast into gel-like structures or thin films

and permanently immobilized via glutaraldehyde crosslinking. In this approach, self-assembly of the bioconjugates into protein-rich and polymer-rich domains provides control over the spatial arrangement and orientation of the protein in the material. Because the self-assembly occurs in three dimensions, high protein packing densities are possible.

1.3 Self-Assembly of Globular Protein-Polymer Block Copolymers

As previously mentioned, self-assembly of globular protein-polymer bioconjugates represents an attractive strategy for processing globular proteins into a 3-D material, offering control of both protein orientation and packing in 3-D space.[49] Because the self-assembly of these bioconjugates dictates the spatial arrangement and orientation of the globular proteins involved, understanding the driving forces for this self-assembly is crucial. The fundamental driving forces involved in bioconjugate self-assembly extend from principles previously established in polymer literature regarding self-assembly of block copolymers. Block copolymers consist of two or more homopolymer units covalently linked together. If the constituent homopolymers are immiscible (as many are), microphase separation of the polymers occurs in which domains rich in each of the different homopolymers form.[50] These domains can organize into periodic structures such as lamellae or hexagonally packed cylinders. depending on factors like the respective lengths of the different polymers involved and the strength of the repulsion between them.[51-52] The phase behavior of different types of diblock copolymers has been explored and characterized extensively. This includes systems where both blocks are uncharged and adopt random coil conformations, copolymers with a rod-like block, as well as copolymers containing a charged polymer block.[53-54] In simple AB diblock copolymer melt systems, mean field theories and self-consistent field theory have shown that self-assembly of these block copolymers can be predicted based on the Flory-Huggins parameter (which characterizes the effective repulsion between polymer blocks), the degree of polymerization, and the relative volume fraction of each of the constituent blocks. [52] As further complexity is added to the block copolymer system, additional factors must be considered. For example, in block copolymer systems containing a rod-like block, the effects of chain stretching also influence the phases formed.[55] Thus, in diblock copolymers where one block is a globular protein, more complex interactions must be considered.

Early efforts to incorporate proteins into block copolymer systems typically used a random coil polymer block attached to poly(amino acid) blocks composed of single amino acid (most commonly poly(γ-benzyl L-glutamate)) that folded into α-helices. For example, work by Perly *et.*

al. showed that polybutadiene-*b*-poly(γ-benzyl L-glutamate) block copolymers ordered into lamellae. [56] Interestingly, the α-helices in the polypeptide domain of this block copolymer are reported to pack hexagonally. Further work on similar systems reported that different self-assembled morphologies could be observed by using triblock instead of diblock systems as well as by changing the total length of the copolymer. [57-58] More recent work on these systems has explored their use in membrane filtration devices, where the hexagonally-packed α-helical domains are used as filtration channels. [59] Because the α-helical block in these copolymers is rod-like, these block copolymers are considered to be rod-coil block copolymers. [53] As such, the energetics governing their assembly is assumed to be the same; however, it is important to note that this treatment does not extend to all globular proteins.

Polymer-conjugated globular proteins have been shown to self-assemble in dilute and concentrated solution, as well as in the solid-state. In dilute solution, protein-polymer conjugates can form a variety of structures, from micelles to giant amphiphiles. [60-61] For example, a lipase enzyme conjugated to polystyrene was shown to form micellar rods in water that retained enzymatic functionality. [62] In concentrated solution (≥ 20 wt% bioconjugate in solution) and the solid state, model red fluorescent protein mCherry has been shown to assemble into many of the phases observed in traditional diblock copolymers, such as lamellae and hexagonally packed cylinders, when conjugated to PNIPAM. As with traditional block copolymers, parameters like coil-fraction and temperature play a strong role in the phase behavior of these systems. [63-64] Work on similar systems showed that polymer chemistry could also be used to tune the concentrated solution assembly of these bioconjugates, but that details of protein surface charge did not play a strong role unless there were strong anisotropies in charge distribution. [65-68] Finally, more recent work has shown that overall protein size can play a strong role in self-assembly of these systems, with the best self-assembly typically occurring when the globular protein of interest has a molecular weight of approximately 30 kDa.[69] It has also been shown that when the polymer block in these protein-polymer block copolymer is replaced with an elastin-like polypeptide (ELP), assembly into mesophases in concentrated solution and solid materials is observed, where the ELP behaves similarly to the polymer block.[70-71] Specific features of ELPs (described below) paired with the fact that these fusion protein systems can be made completely biosynthetically, make this ELP-fusion system a promising platform for study of protein-polymer block copolymers.

1.4 Elastin-Like Polypeptides

Elastin-like polypeptides (ELPs) are thermoresponsive polypeptides with a consensus repeat amino acid sequence of XPGZG, where X is either valine (V) or isoleucine (I), and Z is any amino acid (including non-canonical amino acids) except proline.[72] The ELP sequence is one of the consensus repeat sequences derived from tropoelastin.[73] ELPs undergo a lower critical solution temperature (LCST)-like transition in aqueous solution in which they are soluble at lower temperatures and become insoluble upon heating above a certain temperature, called the transition temperature (T_t). This transition to an insoluble state is accompanied by a secondary structure change in the ELP from a random coil conformation when the ELP is soluble to a β-turn conformation when the ELP is insoluble.[74-75] The ELP transition temperature is dependent on several factors. The first of these is the overall hydrophobicity of the ELP sequence, which is tuned by the amino acid composition in the X and Z positions.[72] Urry and coworkers showed that the transition temperature of an ELP could be roughly predicted by averaging the transition temperature contributions (on a molar basis) from different fourth position amino acids.[72] Later work by McDaniel *et al.* formulated a predictive model that also incorporated the effects of ELP chain length and ELP concentration on ELP transition temperature in solution.[76] The other key factors that influence the ELP transition temperature are the identities and concentrations of salts in solution. Cho *et al.* studied that effect of Hofmeister anions on ELP transition temperature.[77] This study showed that chaotropic anions exhibited competing salting-in and salting-out effects on ELP solubility, whereas kosmotropic anions only exhibited salting-out effects. Furthermore, they showed that the magnitude of these salting-in and salting-out effects was dependent on a given ion's position in the Hofmeister series. The predictability of ELP solubility behavior paired with the highly flexible sequence has made ELPs an attractive protein material for a variety of applications, including protein-based hydrogels, fibers, and underwater adhesives.[78-80] Using genetic engineering, ELP blocks can be fused to other protein genes of interest to generate fusion protein materials similar to the block copolymer or protein-polymer conjugates described above. In these systems, the thermoresponsiveness of the ELP blocks can be used to drive self-assembly of the fusion protein.

1.4.1 Self-Assembly of ELP Fusion Proteins

Incorporation of ELP domains into fusion proteins can enable self-assembly similar to that observed in block copolymers and bioconjugates; however, whether or not self-assembly occurs

is dependent on both the design of the fusion protein and the identity of the other fusion partners. In dilute solution, ELP fusions often take on solution properties similar to the ELP alone, in which the entire fusion precipitates above a certain temperature. This behavior has been used in both protein purification processes and to fabricate depots of ELP-fused therapeutic protein that release over time.[81-83] Alternatively, if two ELPs with differing transition temperatures are fused into diblocks, it has been shown that micelles can form at temperatures in between the two transition temperatures.[84-85] Assuming the physics governing these systems is similar to those of diblock copolymers solutions, the formation of micelles instead of insoluble precipitates is dependent on the interfacial tension between the fused proteins, the ability of the ELP chain(s) to stretch, and the difference in solvent quality for the fused proteins. [86] ELPs incorporated into a fusion protein as end blocks in a triblock fusion (*i.e.* ELP-(polymer-like protein)-ELP) can cause the fusion protein can self-assemble into a micellar gel above the transition temperature of the ELP. [84, 87-88] This behavior is analogous to that observed for triblock ABA copolymers in concentrated solution.[89] When photo-crosslinkable chemical groups are incorporated into the ELP sequence, it is possible to trigger self-assembly into the morphologies mentioned above, and then permanently set the fusion protein material into a given morphology.[90-91] As described above, ELPs fused to globular protein mCherry have been shown to self-assemble similarly to mCherry-PNIPAM bioconjugates in concentrated solution.[70-71] These initial studies indicate that fusion protein self-assembly is dependent on the length and hydrophobicity of the ELP. These ELP features are analogous to polymer volume fraction and polymer chemistry, both of which have been shown to play an important role in mCherry bioconjugate self-assembly.[63, 65]

1.4.2 ELPs as Protein Purification Tags

In addition to driving self-assembly in fusion protein constructs, ELPs can also be used as purification tags. As mentioned in the above section, when globular proteins are genetically fused to ELPs, the fusions typically retain the thermoresponsive behavior of the ELP. That is, the fusions become insoluble on heating above some transition temperature, T_t, though it is well-established that ELP fusion partners (even short peptide sequences) shift the ELP T_t. [92] Initial work by McPherson *et. al.* showed that the thermoresponsive behavior of ELPs could be leveraged to purify ELPs grown in *E. coli*. [93] This purification scheme selectively precipitates ELPs out of solution upon heating above the ELP T_t (where the ELP is insoluble) and collecting precipitates by centrifugation. The precipitate is then resolublized in aqueous solution at a temperature below the

ELP T_t (where the ELP is soluble), and centrifugation is performed again to remove any insoluble non-ELP aggregates.

Meyer and Chilkoti later showed that this same procedure could be used to purify globular proteins fused to an ELP tag.[81] Trabbic-Carlson *et. al.* established that ELP-based purification methods were comparable to His-tag methods in terms of purification yield and purity. [82] Further work out of the Chilkoti lab established that shorter-length ELPs and fusion of the ELP to the C-terminus of the globular protein both typically led to improved expression of ELP fusion proteins in *E. coli*.[94-95] In all these purification schemes, molar levels of sodium chloride (NaCl) or some other kosmotropic salt are commonly added as precipitants, as it is known that kosmotropic salts decrease the solubility of ELPs in aqueous solution.[77] This salt addition is often necessary as the fusion of the globular protein is expected to shift the ELP transition temperature, potentially to temperatures high enough to denature the globular protein of interest. [92] In cases where it is desirable to remove the ELP fusion tag, cleavable sites can be included in between the globular protein of interest and the ELP tag, allowing removal. Examples of this include tags that are cleaved by proteases like thrombin or tobacco etch virus (TEV), tags that allow transpeptidase rearrangements of the ELP tag onto a different protein using Sortase A, as well as self-cleaving intein tags. [96-99] ELP-based methods have also been used for monoclonal antibody purification, where the epitope for a given antibody is fused to an ELP, and binding to this epitope allows precipitation of the antibody of interest.[100] This ELP-based purification method is highly scalable and low cost compared to chromatography-based methods. For this reason, ELP-based precipitations represent an attractive, inexpensive strategy for performing protein purification in high-throughput formats.

1.5 Protein Engineering

The diverse functionality of globular proteins presents many opportunities in catalysis and biosensing; however, these globular proteins are often not optimized to operate under the conditions that they are used (i.e. at elevated temperatures or in the presence of solvents). [101-102] In particular, proteins are extremely rarely optimized to function in the aforementioned materials formats required for biocatalysis, bioelectrocatalysis, or biosensing. The field of protein engineering has emerged as a tool to alleviate these challenges by making changes to protein sequence to improve function. There are two main approaches to this challenge—rational design and directed evolution. [103] Rational design applies existing knowledge about a given protein's

function, such as the amino acid residues involved in its active site, or the protein crystal structure, to make advantageous changes to the protein's sequence in an effort to improve performance. [104-105] In contrast, directed evolution makes random changes to a protein's sequence paired with some selection for protein performance to improve protein qualities of interest. [106] Directed evolution approaches have proven effective on a diverse set of targets, and have been used to develop more stable enzymes for industrial use, novel enzyme chemistries, improved fluorescent proteins, and new paratopes for biosensing.[22, 107-111] One key consideration in this directed evolution approach is the way in which the screening for the property of interest is performed. That is, selecting for improved enzymatic activity in a crude cell lysate will ensure improved activity in a lysate; it is highly likely that this will translate into improved activity in solution, but it is not necessarily guaranteed. This consideration, paired with the knowledge that local protein environment can impact function leads to the conclusion that directed evolution screens should be performed in an environment as close to the end-use environment as possible. For self-assembled protein materials, this would be entail screening the purified protein in its self-assembled state.

Screening self-assembled globular protein materials in their self-assembled state presents several challenges. They must be pure, and they must be processed into their self-assembled state. Several approaches to improving protein purification throughput have been used. For example, affinity-based chromatography methods, including those that bind to 6xHistidine and glutathione S-transferase (GST) tags have been implemented on well-plate scale for protein purification. [112-113] Because these methods require the use of resin for every protein purified, however, the cost of implementing these types of purifications in high throughput can be prohibitively high. ELP-based purification methods can be used to overcome this challenge, as they do not require the use of specific resins, and precipitation-based methods can easily be implemented in a high-throughput format. [81] Another route to increasing purity for a given protein is to use protein excretion systems, which only require separation of the protein of interest from growth media instead of from cell lysate. This approach has been implemented in the production of protein-based materials, such as spider silk and ELPs. [114-115]

1.6 Thesis Overview

The goal of this thesis is to explore the ways in which ELP tags can be used to drive the self-assembly of globular proteins into block copolymer like structures and how these same tags can be used to enable high-throughput processing, including functional screening, of self-assembled

fusion proteins. The remaining chapters of this thesis are organized as follows. Chapter 2 provides in-depth, detailed descriptions of the methods and materials used in the completion of this thesis. Chapter 3 utilizes the model ELP-mCherry fusion protein system to explore the effects of ELP charge and hydrophobicity on fusion protein self-assembly. Specifically, four novel ELP sequences were designed for this study and the self-assembly of the corresponding ELP-mCherry fusion proteins was evaluated and compared to that of the previously studied ELP-mCherry. Dilute solution characterizations of the different ELPs studied are also provided. Chapter 4 reports the existence of cononsolvency behavior in ELP systems, and explores the effect of alcohol hydrophobicity on the cononsolvency behavior of an ELP. In chapter 5, this cononsolvency is applied to ELP-based protein purification systems, where a combination of NaCl- and ethanol-induced precipitations are used to purify and desalt ELP-sfGFP proteins expressed both at the fermenter and well-plate scale. Chapter 5 also reports the application of this procedure to an ELP-based protein gel. Finally, in chapter 6, this high-throughput protein purification and desalting method is applied to a library of engineered streptavidin binding proteins to evaluate the difference between binding protein performance in self-assembled and solution formats.

1.7 References

1. Michaelis, L.; Menten, M. L., The Kinetics of Invertase Action. *Biochem. Z.* **1913,** *49*, 333-369.

2. Davies, D. R.; Padlan, E. A.; Sheriff, S., Antibody-Antigen Complexes. *Annu. Rev. Biochem.* **1990,** *59*, 439-473.

3. Bound enzyme cuts starch conversion cost. *Chem. Eng. News* **1970,** *48* (41), 54.

4. Bhosale, S. H.; Rao, M. B.; Deshpande, V. V., Molecular and industrial aspects of glucose isomerase. *Microbiol. Rev.* **1996,** *60* (2), 280-300.

5. Nagasawa, T.; Yamada, H., Microbial Transformations of Nitriles. *Trends Biotechnol.* **1989,** *7* (6), 153-158.

6. Yamada, H.; Kobayashi, M., Nitrile hydratase and its application to industrial production of acrylamide. *Biosci., Biotechnol., Biochem.* **1996,** *60* (9), 1391-1400.

7. Merino, S. T.; Cherry, J., Progress and challenges in enzyme development for Biomass utilization. *Adv Biochem Eng Biot* **2007,** *108*, 95-120.

8. Solomon, K. V.; Haitjema, C. H.; Henske, J. K.; Gilmore, S. P.; Borges-Rivera, D.; Lipzen, A.; Brewer, H. M.; Purvine, S. O.; Wright, A. T.; Theodorou, M. K.; Grigoriev, I. V.; Regev, A.; Thompson, D. A.; O'Malley, M. A., Early-branching gut fungi possess a large, comprehensive array of biomass-degrading enzymes. *Science* **2016,** *351* (6278), 1192-1195.

9. Efremenko, E. N.; Sergeeva, V. S., Organophosphate hydrolase — an enzyme catalyzing degradation of phosphorus-containing toxins and pesticides. *Russ. Chem. Bull.* **2001,** *50* (10), 1826-1832.

10. Melzer, M.; Chen, J. C. H.; Heidenreich, A.; Gab, J.; Koller, M.; Kehe, K.; Blum, M. M., Reversed Enantioselectivity of Diisopropyl Fluorophosphatase against Organophosphorus Nerve Agents by Rational Design. *J. Am. Chem. Soc.* **2009,** *131* (47), 17226-17232.

11. Goldsmith, M.; Ashani, Y.; Simo, Y.; Ben-David, M.; Leader, H.; Silman, I.; Sussman, J. L.; Tawfik, D. S., Evolved Stereoselective Hydrolases for Broad-Spectrum G-Type Nerve Agent Detoxification. *Chem. Biol.* **2012,** *19* (4), 456-466.

12. Latifi, A. M.; Karami, A.; Khodi, S., Efficient Surface Display of Diisopropylfluorophosphatase (DFPase) in E. coli for Biodegradation of Toxic Organophosphorus Compounds (DFP and Cp). *Appl. Biochem. Biotechnol.* **2015,** *177* (3), 624-36.

13. Keilin, D.; Hartree, E. F., The Use of Glucose Oxidase (Notatin) for the Determination of Glucose in Biological Material and for the Study of Glucose-Producing Systems by Manometric Methods. *Biochem. J.* **1948,** *42* (2), 230-238.

14. Warnke, R. A.; Gatter, K. C.; Falini, B.; Hildreth, P.; Woolston, R. E.; Pulford, K.; Cordell, J. L.; Cohen, B.; Dewolfpeeters, C.; Mason, D. Y., Diagnosis of Human Lymphoma with Monoclonal Anti-Leukocyte Antibodies. *New Engl. J. Med.* **1983,** *309* (21), 1275-1281.

15. Waldmann, T. A., Monoclonal-Antibodies in Diagnosis and Therapy. *Science* **1991,** *252* (5013), 1657-1662.

16. Carter, P. J., Potent antibody therapeutics by design. *Nat. Rev. Immunol.* **2006,** *6* (5), 343-357.

17. Nelson, A. L.; Dhimolea, E.; Reichert, J. M., Development trends for human monoclonal antibody therapeutics. *Nat. Rev. Drug Discovery* **2010,** *9* (10), 767-774.

18. Chin, C. D.; Laksanasopin, T.; Cheung, Y. K.; Steinmiller, D.; Linder, V.; Parsa, H.; Wang, J.; Moore, H.; Rouse, R.; Umviligihozo, G.; Karita, E.; Mwambarangwe, L.; Braunstein, S. L.; van de Wijgert, J.; Sahabo, R.; Justman, J. E.; El-Sadr, W.; Sia, S. K., Microfluidics-based diagnostics of infectious diseases in the developing world. *Nat. Med.* **2011,** *17* (8), 1015-U138.

19. Sliwkowski, M. X.; Mellman, I., Antibody Therapeutics in Cancer. *Science* **2013,** *341* (6151), 1192-1198.

20. Goldman, E. R.; Anderson, G. P.; Conway, J.; Sherwood, L. J.; Fech, M.; Vo, B.; Liu, J. L.; Hayhurst, A., Thermostable Llama Single Domain Antibodies for Detection of Botulinum A Neurotoxin Complex. *Anal. Chem.* **2008,** *80* (22), 8583-8591.

21. Miller, B. R.; Demarest, S. J.; Lugovskoy, A.; Huang, F.; Wu, X. F.; Snyder, W. B.; Croner, L. J.; Wang, N.; Amatucci, A.; Michaelson, J. S.; Glaser, S. M., Stability engineering of scFvs for the development of bispecific and multivalent antibodies. *Protein Eng. Des. Sel.* **2010,** *23* (7), 549-557.

22. Miller, E. A.; Traxlmayr, M. W.; Shen, J.; Sikes, H. D., Activity-based assessment of an engineered hyperthermophilic protein as a capture agent in paper-based diagnostic tests. *Mol. Syst. Des. Eng.* **2016,** *1* (4), 377-381.

23. Levy, I.; Shoseyov, O., Cellulose-binding domains biotechnological applications. *Biotechnol. Adv.* **2002,** *20* (3-4), 191-213.

24. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (12), E690-E697.

25. Sun, F.; Zhang, W. B.; Mahdavi, A.; Arnold, F. H.; Tirrell, D. A., Synthesis of bioactive protein hydrogels by genetically encoded SpyTag-SpyCatcher chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111* (31), 11269-11274.

26. Miller, E. A.; Baniya, S.; Osorio, D.; Al Maalouf, Y. J.; Sikes, H. D., Paper-based diagnostics in the antigen-depletion regime: High-density immobilization of rcSso7d-cellulose-binding domain fusion proteins for efficient target capture. *Biosens. Bioelectron.* **2018,** *102*, 456-463.

27. DiCosimo, R.; McAuliffe, J.; Poulose, A. J.; Bohlmann, G., Industrial use of immobilized enzymes. *Chem. Soc. Rev.* **2013,** *42* (15), 6437-6474.

28. Bachler, M. J.; Strandberg, G. W.; Smiley, K. L., Starch Conversion by Immobilized Glucoamylase. *Biotechnol. Bioeng.* **1970,** *12* (1), 85-+.

29. Poulsen, P. B., European and American Trends in the Industrial Application of Immobilized Biocatalysts. *Enzyme Microb. Technol.* **1981,** *3* (3), 271-273.

30. Engvall, E., Enzyme immunoassay ELISA and EMIT. In *Method Enzymol*, Academic Press: 1980; Vol. 70, pp 419-439.

31. Wu, J. C. Y.; Hutchings, C. H.; Lindsay, M. J.; Werner, C. J.; Bundy, B. C., Enhanced Enzyme Stability Through Site-Directed Covalent Immobilization. *J. Biotechnol.* **2015,** *193*, 83-90.

32. Ghosh, P.; Han, G.; De, M.; Kim, C. K.; Rotello, V. M., Gold nanoparticles in delivery applications. *Adv. Drug Delivery Rev.* **2008,** *60* (11), 1307-1315.

33. Ekblad, T.; Liedberg, B., Protein adsorption and surface patterning. *Curr. Opin. Colloid Interface Sci.* **2010,** *15* (6), 499-509.

34. Boozer, C.; Ladd, J.; Chen, S. F.; Yu, Q.; Homola, J.; Jiang, S. Y., DNA directed protein immobilization on mixed ssDNA/oligo(ethylene glycol) self-assembled monolayers for sensitive biosensors. *Anal. Chem.* **2004,** *76* (23), 6967-6972.

35. Kumar, N.; Hahm, J. I., Nanoscale protein patterning using self-assembled diblock copolymers. *Langmuir* **2005,** *21* (15), 6652-6655.

36. Schoffelen, S.; Lambermon, M. H. L.; van Eldijk, M. B.; van Hest, J. C. M., Site-specific modification of Candida antarctica lipase B via residue-specific incorporation of a non-canonical amino acid. *Bioconjugate Chem.* **2008,** *19* (6), 1127-1131.

37. Raliski, B. K.; Howard, C. A.; Young, D. D., Site-Specific Protein Immobilization Using Unnatural Amino Acids. *Bioconjugate Chem.* **2014,** *25* (11), 1916-1920.

38. Kim, D.; Herr, A. E., Protein immobilization techniques for microfluidic assays. *Biomicrofluidics* **2013,** *7* (4).

39. Duckworth, B. P.; Xu, J. H.; Taton, T. A.; Guo, A.; Distefano, M. D., Site-specific, covalent attachment of proteins to a solid surface. *Bioconjugate Chem.* **2006,** *17* (4), 967-974.

40. Chan, L. Y.; Cross, H. F.; She, J. K.; Cavalli, G.; Martins, H. F. P.; Neylon, C., Covalent Attachment of Proteins to Solid Supports and Surfaces via Sortase-Mediated Ligation. *Plos One* **2007,** *2* (11).

41. Balabushevitch, N. G.; Sukhorukov, G. B.; Moroz, N. A.; Volodkin, D. V.; Larionova, N. I.; Donath, E.; Mohwald, H., Encapsulation of proteins by layer-by-layer adsorption of polyelectrolytes onto protein aggregates: Factors regulating the protein release. *Biotechnol. Bioeng.* **2001,** *76* (3), 207-213.

42. Ai, H.; Jones, S. A.; Lvov, Y. M., Biomedical applications of electrostatic layer-by-layer nano-assembly of polymers, enzymes, and nanoparticles. *Cell Biochem. Biophys.* **2003,** *39* (1), 23-43.

43. Hartdegen, F. J.; Swann, W. E. Immobilization of proteins with polyurethane polymers. 1976.

44. LeJeune, K. E.; Russell, A. J., Covalent binding of a nerve agent hydrolyzing enzyme within polyurethane foams. *Biotechnol. Bioeng.* **1996,** *51* (4), 450-457.

45. Presley, A. D.; Chang, J. J.; Xu, T., Directed co-assembly of heme proteins with amphiphilic block copolymers toward functional biomolecular materials. *Soft Matter* **2011,** *7* (1), 172-179.

46. Jia, S. S.; Fei, J. J.; Deng, J. J.; Cai, Y. L.; Li, J. A., Direct electrochemistry and electrocatalysis of hemoglobin immobilized in an amphiphilic diblock copolymer film. *Sensor Actuat B-Chem* **2009,** *138* (1), 244-250.

47. Laible, P. D.; Kelley, R. F.; Wasielewski, M. R.; Firestone, M. A., Electron-transfer dynamics of photosynthetic reaction centers in thermoresponsive soft materials. *J. Phys. Chem. B* **2005,** *109* (49), 23679-23686.

48. Sharma, K. P.; Collins, A. M.; Perriman, A. W.; Mann, S., Enzymatically Active Self-Standing Protein-Polymer Surfactant Films Prepared by Hierarchical Self-Assembly. *Adv. Mater.* **2013,** *25* (14), 2005-2010.

49. Huang, A.; Olsen, B. D., Self-Assembly of Differently Shaped Protein–Polymer Conjugates through Modification of the Bioconjugation Site. *Macromol. Rapid Commun.* **2016,** *37* (15), 1268-1274.

50. Hamley, I. W., *The physics of block copolymers*. Vol. 19.

51. Bates, F. S.; Fredrickson, G. H., Block Copolymer Thermodynamics - Theory and Experiment. *Annu. Rev. Phys. Chem.* **1990,** *41*, 525-557.

52. Matsen, M. W.; Bates, F. S., Unifying weak- and strong-segregation block copolymer theories. *Macromolecules* **1996,** *29* (4), 1091-1098.

53. Olsen, B. D.; Segalman, R. A., Self-assembly of rod–coil block copolymers. *Materials Science and Engineering: R: Reports* **2008,** *62* (2), 37-66.

54. Sing, C. E.; Zwanikken, J. W.; Olvera de la Cruz, M., Electrostatic control of block copolymer morphology. *Nat. Mater.* **2014,** *13* (7), 694-8.

55. Williams, D. R. M.; Fredrickson, G. H., Cylindrical Micelles in Rigid-Flexible Diblock Copolymers. *Macromolecules* **1992,** *25* (13), 3561-3568.

56. Perly, B.; Douy, A.; Gallot, B., Block Copolymers Polybutadiene-Poly(Benzyl-L-Glutamate) and Polybutadiene-Poly(N5-Hydroxypropylglutamine) Preparation and Structural Study by X-Ray and Electron-Microscopy. *Makromol. Chem.* **1976,** *177* (9), 2569-2589.

57. Nakajima, A.; Hayashi, T.; Kugo, K.; Shinoda, K., Synthesis and Structural Study of the a-B-a-Type Block Copolymer Consisting of Poly(Gamma-Benzyl L-Glutamate) as the a-Component and Polybutadiene as the B-Component. *Macromolecules* **1979,** *12* (5), 840-843.

58. Klok, H. A.; Langenwalter, J. F.; Lecommandoux, S., Self-assembly of peptide-based diblock oligomers. *Macromolecules* **2000,** *33* (21), 7819-7826.

59. Sutisna, B.; Bilalis, P.; Musteata, V.; Smilgies, D. M.; Peinemann, K. V.; Hadjichristidis, N.; Nunes, S. P., Self-Assembled Membranes with Featherlike and Lamellar Morphologies Containing alpha-Helical Polypeptides. *Macromolecules* **2018,** *51* (20), 8174-8187.

60. Dirks, A. J.; Nolte, R. J. M.; Cornelissen, J. J. L. M., Protein-Polymer Hybrid Amphiphiles. *Adv. Mater.* **2008,** *20* (20), 3953-3957.

61. Boyer, C.; Huang, X.; Whittaker, M. R.; Bulmus, V.; Davis, T. P., An overview of protein-polymer particles. *Soft Matter* **2011,** *7* (5), 1599-1614.

62. Velonia, K.; Rowan, A. E.; Nolte, R. J. M., Lipase polystyrene giant amphiphiles. *J. Am. Chem. Soc.* **2002,** *124* (16), 4224-4225.

63. Lam, C. N.; Olsen, B. D., Phase transitions in concentrated solution self-assembly of globular protein-polymer block copolymers. *Soft Matter* **2013,** *9* (8), 2393-2402.

64. Thomas, C. S.; Olsen, B. D., Coil fraction-dependent phase behaviour of a model globular protein-polymer diblock copolymer. *Soft Matter* **2014,** *10* (17), 3093-3102.

65. Chang, D.; Lam, C. N.; Tang, S.; Olsen, B. D., Effect of polymer chemistry on globular protein-polymer block copolymer self-assembly. *Polym. Chem.* **2014,** *5* (17), 4884-4895.

66. Lam, C. N.; Kim, M.; Thomas, C. S.; Chang, D.; Sanoja, G. E.; Okwara, C. U.; Olsen, B. D., The Nature of Protein Interactions Governing Globular Protein–Polymer Block Copolymer Self-Assembly. *Biomacromolecules* **2014,** *15* (4), 1248-1258.

67. Chang, D.; Olsen, B. D., Self-assembly of protein-zwitterionic polymer bioconjugates into nanostructured materials. *Polym. Chem.* **2016,** *7* (13), 2410-2418.

68. Lam, C. N.; Yao, H.; Olsen, B. D., The Effect of Protein Electrostatic Interactions on Globular Protein–Polymer Block Copolymer Self-Assembly. *Biomacromolecules* **2016,** *17* (9), 2820-2829.

69. Paloni, J. M.; Miller, E. A.; Sikes, H. D.; Olsen, B. D., Improved Ordering in Low Molecular Weight Protein–Polymer Conjugates Through Oligomerization of the Protein Block. *Biomacromolecules* **2018,** *19* (9), 3814-3824.

70. Qin, G.; Glassman, M. J.; Lam, C. N.; Chang, D.; Schaible, E.; Hexemer, A.; Olsen, B. D., Topological Effects on Globular Protein-ELP Fusion Block Copolymer Self-Assembly. *Adv. Funct. Mater.* **2015,** *25* (5), 729-738.

71. Qin, G.; Perez, P. M.; Mills, C. E.; Olsen, B. D., Effect of ELP Sequence and Fusion Protein Design on Concentrated Solution Self-Assembly. *Biomacromolecules* **2016,** *17* (3), 928-934.

72. Urry, D. W.; Gowda, D. C.; Parker, T. M.; Luan, C.-H.; Reid, M. C.; Harris, C. M.; Pattanaik, A.; Harris, R. D., Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers* **1992,** *32* (9), 1243-1250.

73. Foster, J. A.; Bruenger, E.; Gray, W. R.; Sandberg, L. B., Isolation and Amino-Acid Sequences of Tropoelastin Peptides. *J. Biol. Chem.* **1973,** *248* (8), 2876-2879.

74. Urry, D. W.; Long, M. M.; Cox, B. A.; Ohnishi, T.; Mitchell, L. W.; Jacobs, M., The synthetic polypentapeptide of elastin coacervates and forms filamentous aggregates. *Biochim. Biophys. Acta, Protein Struct.* **1974,** *371* (2), 597-602.

75. Urry, D. W.; Long, M. M.; Ohnishi, T.; Jacobs, M., Circular dichroism and absorption of the polytetrapeptide of elastin: A polymer model for the β-turn. *Biochem. Biophys. Res. Commun.* **1974,** *61* (4), 1427-1433.

76. McDaniel, J. R.; Radford, D. C.; Chilkoti, A., A Unified Model for De Novo Design of Elastin-like Polypeptides with Tunable Inverse Transition Temperatures. *Biomacromolecules* **2013,** *14* (8), 2866-2872.

77. Cho, Y.; Zhang, Y.; Christensen, T.; Sagle, L. B.; Chilkoti, A.; Cremer, P. S., Effects of Hofmeister anions on the phase transition temperature of elastin-like polypeptides. *J. Phys. Chem. B* **2008,** *112* (44), 13765-71.

78. Huang, L.; McMillan, R. A.; Apkarian, R. P.; Pourdeyhimi, B.; Conticello, V. P.; Chaikof, E. L., Generation of synthetic elastin-mimetic small diameter fibers and fiber networks. *Macromolecules* **2000,** *33* (8), 2989-2997.

79. Dooling, L. J.; Buck, M. E.; Zhang, W. B.; Tirrell, D. A., Programming Molecular Association and Viscoelastic Behavior in Protein Networks. *Adv. Mater.* **2016,** *28* (23), 4651-4657.

80. Brennan, M. J.; Kilbride, B. F.; Wilker, J. J.; Liu, J. C., A bioinspired elastin-based protein for a cytocompatible underwater adhesive. *Biomaterials* **2017,** *124*, 116-125.

81. Meyer, D. E.; Chilkoti, A., Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat. Biotechnol.* **1999,** *17* (11), 1112-1115.

82. Trabbic-Carlson, K.; Liu, L.; Kim, B.; Chilkoti, A., Expression and purification of recombinant proteins from Escherichia coli: Comparison of an elastin-like polypeptide fusion with an oligohistidine fusion. *Protein Sci.* **2004,** *13* (12), 3274-3284.

83. Gilroy, C. A.; Roberts, S.; Chilkoti, A., Fusion of fibroblast growth factor 21 to a thermally responsive biopolymer forms an injectable depot with sustained anti-diabetic action. *J. Controlled Release* **2018,** *277*, 154-164.

84. Wright, E. R.; Conticello, V. P., Self-assembly of block copolymers derived from elastin-mimetic polypeptide sequences. *Adv. Drug Delivery Rev.* **2002,** *54* (8), 1057-1073.

85. Hassouneh, W.; Fischer, K.; MacEwan, S. R.; Branscheid, R.; Fu, C. L.; Liu, R.; Schmidt, M.; Chilkoti, A., Unexpected Multivalent Display of Proteins by Temperature Triggered Self-

Assembly of Elastin-like Polypeptide Block Copolymers. *Biomacromolecules* **2012,** *13* (5), 1598-1605.

86. Tuzar, Z.; Kratochvil, P., Block and Graft Copolymer Micelles in Solution. *Adv. Colloid Interface Sci.* **1976,** *6* (3), 201-232.

87. Glassman, M. J.; Olsen, B. D., End Block Design Modulates the Assembly and Mechanics of Thermoresponsive, Dual-Associative Protein Hydrogels. *Macromolecules* **2015,** *48* (6), 1832-1842.

88. Sing, M. K.; Glassman, M. J.; Vronay-Ruggles, X. T.; Burghardt, W. R.; Olsen, B. D., Structure and rheology of dual-associative protein hydrogels under nonlinear shear flow. *Soft Matter* **2017,** *13* (45), 8511-8524.

89. Zhou, Z. K.; Yang, Y. W.; Booth, C.; Chu, B., Association of a triblock ethylene oxide (E) and butylene oxide (B) copolymer (B(12)E(260)B(12)) in aqueous solution. *Macromolecules* **1996,** *29* (26), 8357-8361.

90. Raphel, J.; Parisi-Amon, A.; Heilshorn, S., Photoreactive elastin-like proteins for use as versatile bioactive materials and surface coatings. *J. Mater. Chem.* **2012,** *22* (37), 19429-19437.

91. Costa, S. A.; Simon, J. R.; Amiram, M.; Tang, L.; Zauscher, S.; Brustad, E. M.; Isaacs, F. J.; Chilkoti, A., Photo-Crosslinkable Unnatural Amino Acids Enable Facile Synthesis of Thermoresponsive Nano- to Microgels of Intrinsically Disordered Polypeptides. *Adv. Mater.* **2018,** *30* (5).

92. Trabbic-Carlson, K.; Meyer, D. E.; Liu, L.; Piervincenzi, R.; Nath, N.; LaBean, T.; Chilkoti, A., Effect of protein fusion on the transition temperature of an environmentally responsive elastin-like polypeptide: a role for surface hydrophobicity? *Protein Eng. Des. Sel.* **2004,** *17* (1), 57-66.

93. McPherson, D. T.; Xu, J.; Urry, D. W., Product Purification by Reversible Phase Transition Following Escherichia coli Expression of Genes Encoding up to 251 Repeats of the Elastomeric Pentapeptide GVGVP. *Protein Expr. Purif.* **1996,** *7* (1), 51-57.

94. Meyer, D. E.; Trabbic-Carlson, K.; Chilkoti, A., Protein purification by fusion with an environmentally responsive elastin-like polypeptide: Effect of polypeptide length on the purification of thioredoxin. *Biotechnol. Progr.* **2001,** *17* (4), 720-728.

95. Christensen, T.; Amiram, M.; Dagher, S.; Trabbic-Carlson, K.; Shamji, M. F.; Setton, L. A.; Chilkoti, A., Fusion order controls expression level and activity of elastin-like polypeptide fusion proteins. *Protein Sci.* **2009,** *18* (7), 1377-1387.

96. Banki, M. R.; Feng, L.; Wood, D. W., Simple bioseparations using self-cleaving elastin-like polypeptide tags. *Nat. Methods* **2005,** *2*, 659.

97. Banki, M. R.; Wood, D. W., Inteins and affinity resin substitutes for protein purification and scale up. *Microb. Cell Fact.* **2005,** *4*.

98. Hassouneh, W.; Christensen, T.; Chilkoti, A., Elastin-like polypeptides as a purification tag for recombinant proteins. *Current protocols in protein science* **2010,** *Chapter 6*, Unit-6.11.

99. Bellucci, J. J.; Amiram, M.; Bhattacharyya, J.; McCafferty, D.; Chilkoti, A., Three-in-One Chromatography-Free Purification, Tag Removal, and Site-Specific Modification of Recombinant

Fusion Proteins Using Sortase A and Elastin-like Polypeptides. *Angew. Chem. Int. Ed.* **2013,** *52* (13), 3703-3708.

100. Kim, J.-Y.; Mulchandani, A.; Chen, W., Temperature-triggered purification of antibodies. *Biotechnol. Bioeng.* **2005,** *90* (3), 373-379.

101. Arnold, F. H.; Volkov, A. A., Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol.* **1999,** *3* (1), 54-59.

102. Cherry, J. R.; Fidantsef, A. L., Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **2003,** *14* (4), 438-443.

103. Kazlauskas, R. J.; Bornscheuer, U. T., Finding better protein engineering strategies. *Nat Chem. Biol.* **2009,** *5* (8), 526-529.

104. Marshall, S. A.; Lazar, G. A.; Chirino, A. J.; Desjarlais, J. R., Rational design and engineering of therapeutic proteins. *Drug Discovery Today* **2003,** *8* (5), 212-221.

105. Lutz, S., Beyond directed evolution-semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **2010,** *21* (6), 734-743.

106. Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **2015,** *16* (7), 379-394.

107. Jespers, L. S.; Roberts, A.; Mahler, S. M.; Winter, G.; Hoogenboom, H. R., Guiding the Selection of Human-Antibodies from Phage Display Repertoires to a Single Epitope of an Antigen. *Bio-Technol* **1994,** *12* (9), 899-903.

108. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N. G.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp red fluorescent protein. *Nat. Biotechnol.* **2004,** *22* (12), 1567-1572.

109. Eijsink, V. G. H.; Gaseidnes, S.; Borchert, T. V.; van den Burg, B., Directed evolution of enzyme stability. *Biomol. Eng.* **2005,** *22* (1-3), 21-30.

110. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H., Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013,** *339* (6117), 307-310.

111. McIsaac, R. S.; Engqvist, M. K. M.; Wannier, T.; Rosenthal, A. Z.; Herwig, L.; Flytzanis, N. C.; Imasheva, E. S.; Lanyi, J. K.; Balashov, S. P.; Gradinaru, V.; Arnold, F. H., Directed evolution of a far-red fluorescent rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111* (36), 13034-13039.

112. Scheich, C.; Sievert, V.; Bussow, K., An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography. *BMC Biotechnol.* **2003,** *3*, 12.

113. Hammarström, M.; Woestenenk, E. A.; Hellgren, N.; Härd, T.; Berglund, H., Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *J. Struct. Funct. Genomics* **2006,** *7* (1), 1-14.

114. Widmaier, D. M.; Tullman-Ercek, D.; Mirsky, E. A.; Hill, R.; Govindarajan, S.; Minshull, J.; Voigt, C. A., Engineering the Salmonella type III secretion system to export spider silk monomers. *Mol. Syst. Biol.* **2009,** *5*.

115. Azam, A.; Li, C.; Metcalf, K. J.; Tullman-Ercek, D., Type III secretion as a generalizable strategy for the production of full-length biopolymer-forming proteins. *Biotechnol. Bioeng.* **2016,** *113* (11), 2313-20.

Chapter 2 : Materials and Methods

2.1 Biosynthesis

2.1.1 Genetic engineering

2.1.1.1 General cloning procedures

Digests and dephosphorylations

Restriction enzymes used in DNA digests were purchased from New England Biolabs (NEB) and used, as specified, with the provided reaction buffers. Dephosphorylations were performed with Antarctic Phosphatase, also purchased from NEB. In both cases, enzyme or phosphatase was added to DNA at a concentration $\geq$ 1 unit enzyme/μg DNA. At least 1 μL of enzyme or phosphatase was added to all reactions. All DNA digests and dephosphorylations were carried out by placing Eppendorf tubes in an incubator without shaking at 37 °C, typically for 1-2 hours, but up to 12 hours. Digests carried out for longer than 2 hours were parafilmed to prevent evaporative losses.

DNA Purification

Plasmid DNA was purified out of cultures of NEB5α cells (a derivative of DH5α cells) grown in lysogeny broth (LB) supplemented with the relevant antibiotic for the plasmid of interest (50 μg/mL kanamycin or 100 μg/mL ampicillin) at 37 °C for 12-16 hours in an orbital incubator-shaker. Purification was performed either via miniprep (Omega BioTek D6942) of 5 mL of culture or midiprep (Sigma PLD35) of 30-50 mL of culture, following the kit instructions. Cultures for miniprep (5 mL) were grown in sterilized glass culture tubes and cultures for midiprep (30-50 mL) were grown in a sterilized 250 mL baffled Erlenmeyer flask. Typical minipreps and midipreps yielded 30-50 μL and 500 μL, respectively, of 60-100 ng/μL DNA.

DNA generated via polymerase chain reaction (PCR) was purified via spin column (Omega BioTek D6492), following the kit instructions. Digested plasmid DNA was purified by running agarose gel electrophoresis (1% in Tris-Acetate-EDTA (TAE) buffer: 40 mM tris, 20 mM acetic acid, 1 mM ethylene diaminetetraacetic acid, pH 8.0, stained with SYBR™ Safe DNA Gel Stain, ThermoFisher #S33102), followed by gel extraction on the DNA band of interest (Qiagen #28704). SYBR™ Safe Gel Stain was incorporated into the gel after dissolving agarose into TAE (via microwave heating) by diluting it from 10,000X (as ordered) to 1X. Gels were then cast by pouring the agarose/TAE/SYBR™ Safe solutions into molds and cooling them at room temperature for at least 45 min (or until the gel set).

Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was carried out using Q5® High-Fidelity 2X Master Mix (NEB #M0492). Primers for DNA prepared by polymerase chain reaction (PCR) were designed using the following criteria. First, the overhang sequence of 12-24 nucleotides was selected to ensure the melting temperature of the overhang for both primers was within 5 °C of each other (calculated using NEB Tm calculator, https://tmcalculator.neb.com/). The entire primer was also designed to avoid the formation of primer dimers. Likelihood of dimer formation was calculated using ThermoFisher Multiple Primer Analyzer, https://www.thermofisher.com/us/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html. If the calculator predicted dimer formation, primer overhangs were adjusted until no primer formation was predicted. In the case of primers that were used to append new restriction sites to a given gene, five base pairs were added upstream of the appended restriction site to ensure proper function of the restriction enzyme. These base pairs were selected randomly such that the same number of A/T and C/G pairs were present; however, if the primer analyzer indicated that these base pairs led to dimer formation, they were changed to avoid this. Annealing temperatures for PCR were selected by performing gradient PCR (35 cycles) 5 °C above and below the calculated melting temperature for the primers. Annealing temperature was selected to maximize the yield of desired PCR product while minimizing undesired PCR products. This procedure (including sample gels) is detailed in **Appendix 2A.1**. When preparing DNA for use in a ligation, 25 cycles of PCR were used to minimize the generation of off-target PCR products. The sequence of all PCR products was confirmed via Sanger sequencing. DpnI (NEB #R0176) was used in digests of PCR'd DNA that was not subject to purification via agarose gel electrophoresis to eliminate methylated template DNA.

Ligations

All ligations were carried out using T4 DNA Ligase at 2,000,000 units/mL (NEB #M202T/M). Use of this higher concentration ligase was particularly important in ligations involving ELP concatamerization. Ligase buffer was thawed once after receipt and separated into single use aliquots prior to refreezing to minimize degradation of buffer components due to multiple freeze/thaw cycles. Unless otherwise specified, total volume of ligation reactions was 10 μL with

a total DNA content of 60-120 ng. Unless otherwise indicated, all ligation reactions were performed such that there were three moles of insert for every one mole of vector (3:1 insert:vector ratio). All ligations were run with a control containing the vector only to determine if the vector was self-ligating without incorporation of the insert.

All ligations were carried out at 16 °C in a thermal cycler for 12-20 hours. Ligations were subsequently transformed into chemically competent NEB5α cells, prepared using the *Mix & Go! E. coli* kit from Zymo Research (#T3002) following the kit instructions. Transformations were carried out by adding 2 µL of the ligation mixture to 100 µL of competent cells (in 1.75 mL Eppendorf tubes) that were thawed from -80 °C on ice, where the ligation mixture was added as soon as the competent cells thawed. The competent cell/DNA mixture was left on ice for 30 minutes to allow transformation of the DNA. An outgrowth step was then performed by adding 400 µL of super optimal broth (SOB) to the cells and mixing them at 37 °C in an orbital shaker for 45-60 minutes. After outgrowth, the cells were gently pelleted by centrifugation, and 400 µL of supernatant was aspirated off. The cells were then resuspended in the remaining 100 µL of liquid and plated using glass beads onto a pre-warmed LB-agar plate supplemented with antibiotic to select for the vector of interest.

Screening of clones

Success of all ligations performed was screened in a two-step process. First, 4-10 colonies were selected from the LB-agar ligation transformation plate, streaked onto a fresh LB-agar plate, and grown up in 5 mL of LB supplemented with appropriate antibiotic for 12-20 hours. Plasmid DNA was isolated from the 5 mL LB cultures via miniprep (see **DNA Purification**). Analytical digestion was then performed on all the different clones as follows. Analytical digest "master mix" was prepared by first preparing 8 x (# of clones screened + 1) µL of 10x reaction buffer (typically NEB CutSmart Buffer #B7204) and water mixed at a 1:7 volumetric ratio.1 µL of each restriction enzyme of interest was then added into this reaction buffer/water mixture, mixed thoroughly, and centrifuged in a microcentrifuge. Analytical digests were then carried out by adding 8 µL of this analytical digest "master mix" to 2 µL of miniprepped DNA from each clone in a PCR tube. Digests were then incubated at 37 °C in a thermal cycler for 1 hour, and subsequently screened via agarose gel electrophoresis (1% in TAE stained with SYBR™ Safe, see **DNA Purification**). Detailed protocol for analytical digest can be found in **Appendix 2A.2**. One to two clones

exhibiting the expected bands in the analytical digest were subsequently verified *via* Sanger sequencing in the forward and reverse direction, typically using the T7 and T7-Term universal primers from Genewiz.

Transformations

Transformations of plasmid DNA (either final cloning products or as received from Genewiz) into expression and cloning *E. coli* cell lines was carried out as described in the protocol in **Appendix A.3**.

2.1.1.2 Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins

ELP Concatamerization

ELP genes encoding five repeats of each 25 amino acid long repeat sequence, comprised of 450 bp, were ordered from Genscript, and were subcloned into the pET28b vector between the 5' NcoI and 3' XhoI restriction sites. Here, the ATG sequence in the NcoI restriction site was used as the start codon for the amino acid sequence. A 6xHis tag was included at the N-terminus of the ELP sequence to enable purification. Upstream of the 5' end of the ELP sequence (but after the sequence encoding for the 6xHis tag) is a BamHI restriction site followed by a BsaI restriction site in the forward orientation (cut site is downstream). At the 3' end of each ELP gene is a BsaI restriction site in the reverse orientation (cut site is upstream), followed by an NheI restriction site and then a HindIII restriction site.

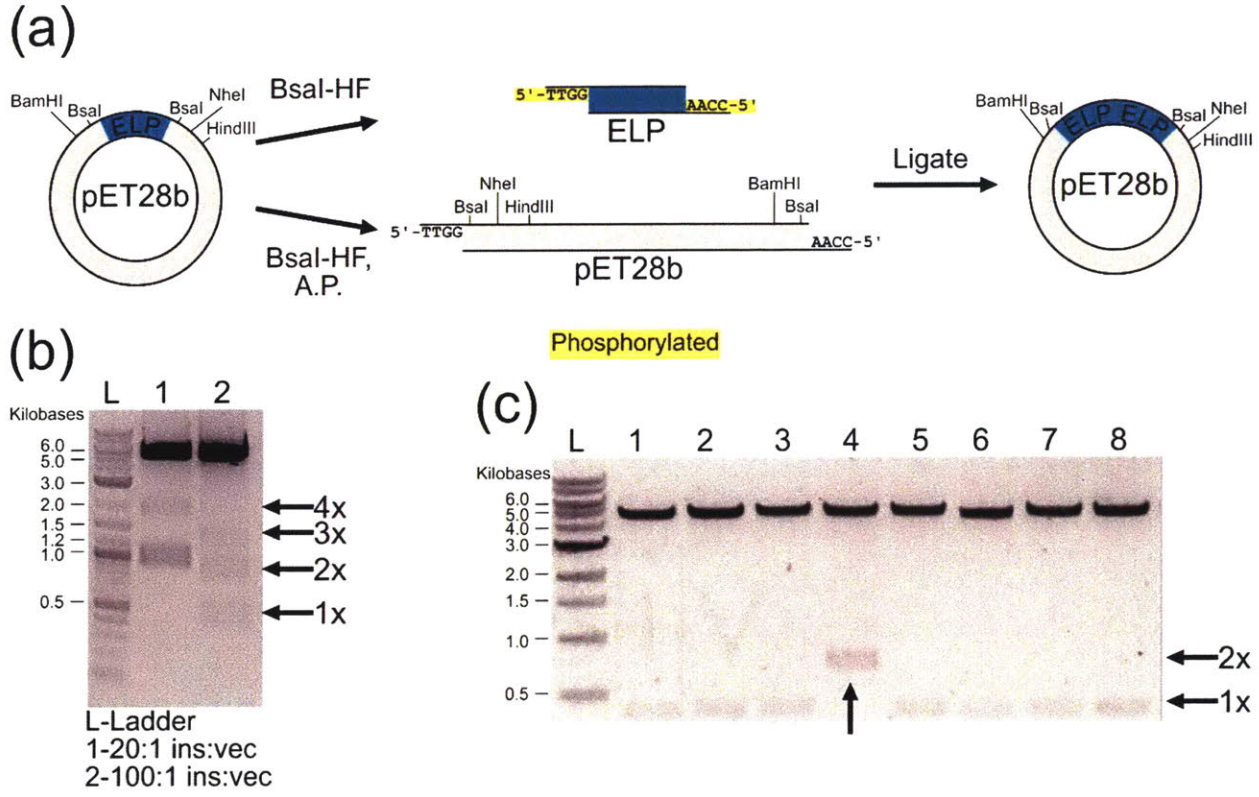


Figure 2-1. (a) Schematic of cloning scheme using to concatamerize ELP genes used in this study. (b) Analytical digest with NcoI-HF/HindIII-HF on ligation transformant of 375 bp bELP at two different insert:vector molar ratios. (c) Analytical digest with NcoI-HF/HindIII-HF on 8 different nELP clones (where each lane 1-8 represents a different selected clone) produced by a single round of concatamerization on a 375 bp nELP.

To achieve the desired gene length of 1554 bp, ELP genes were concatamerized using a cloning scheme similar to Golden Gate cloning, and previously used for ELP elongation. (**Figure 2-1**a). [1-2] To prepare the ELP insert for concatamerization, 500 µL of plasmid containing ELP genes was prepared by midiprepping out of 50 mL of culture. The ELP gene was digested out of midiprepped DNA with BsaI-HF (NEB #R3535) in CutSmart Buffer (NEB #B7204), leaving a TTGG at the 5' end of the gene and an AACC overhang at the 3' end of the gene (**Figure 2-1**a). After digestion, DNA was purified and concentrated down to 35 µL by spin column purification (Omega BioTek #D6492), using MilliQ® water as the eluent. Final concentrations after this step were typically 40-80 ng/µL after this step. The gene fragment of interest (either ~400 bp or ~800 bp) was isolated using agarose gel electrophoresis followed by spin column gel extraction. Acceptor plasmid for the ELP gene was prepared by first miniprepping 30 µL of plasmid out of 5 mL of culture. The

plasmid was digested with BsaI-HF (NEB #R3535) in CutSmart Buffer (NEB #B7204) (**Figure 2-1**a), and subsequently dephosphorylated with Antarctic Phosphatase (NEB #M0289) in Antarctic Phosphatase buffer (NEB #B0289). Digested and dephosphorylated vector (~5500 bp) was isolated and purified using agarose gel electrophoresis followed by gel extraction. Purified inserts and vectors were ligated with a high molar ratio of insert to vector (insert: vector ratios from 10:1 to 100:1). Initially, 100:1 insert to vector ratios were used, based on protocols published in the thesis of M. J. Glassman; however, in some cases, final concentration of the purified inserts was not sufficient to reach this insert to vector ratio.[3] In collaboration with R. K. Avery, several experiments were performed to evaluate doubling efficiency as a function of insert:vector ratio. The results of these experiments suggested that ratios of 10:1 and 20:1 were just as, if not more, efficient at producing a doubled product than the 100:1 ratio. One example of this is shown in **Figure 2-1**b. For new constructs being cloned using this same method, comparison of different insert:vector ratios is recommended. This can be achieved without individual colony selection by performing an analytical digest on the total ligation transformant as described below. Ligations were transformed into chemically competent NEBα cells as described above.

Prior to screening individual clones from the transformed concatamerization, the success of a given concatamerization was evaluated as follows. Transformation of the ligation into NEB5α chemically competent cells was performed as described up to the pelleting/resuspension step after outgrowth. Instead of plating the entire 100 μL of resuspended cells, a small amount of these cells (1-2 μL) was used to inoculate a 5 mL LB culture supplemented with 50 μg/mL kanamycin. The remaining resuspended cells were plated using glass beads as normal. The culture inoculated with the ligation transformant was grown up at 37 °C for 12-20 hours and subsequently miniprepped. The entirety of the miniprepped DNA (30 μL) was subsequently digested with NcoI-HF (NEB #R3193) and HindIII-HF (NEB #R3104) in CutSmart Buffer, and analyzed by agarose gel electrophoresis. The goal of this experiment was to determine the distribution of different concatamers in the final ligation transformant; if this analysis revealed the presence of 4x and 2x concatamers in the ligation transformant, for example (as in lane 1 of **Figure 2-1b**), colonies could be selected and screened until concatamers of both lengths were discovered. This analysis was also used as a tool to probe which ligation conditions produced concatamers of the desired length, as shown in **Figure 2-1b**, where two different insert:vector molar ratios were compared.

After screening the ligation transformant, individual clones were selected from the ligation transformation plate and screened by analytical digest with NcoI-HF/HindIII-HF in CutSmart Buffer as described above (see **Screening of clones** above). An example of such an analytical digest on eight clones from a given ligation is shown in **Figure 2-1c**, where lane 4 represents a successful ELP concatamer, doubled in length from 375 bp to 750 bp (ELP bp only). Final clones were confirmed by Sanger sequencing with universal T7 and T7-Term primers. It should be noted that because highly repetitive sequences typically cannot be fully sequenced with Sanger sequencing, the entirety of the sequence sometimes could not be confirmed; however, the combination of analytical digest and correct sequences at the 5' and 3' end of the ELP gene were considered sufficient to confirm the desired concatamer. Amino acid and DNA sequences for all ELP genes ordered for or generated by this cloning scheme are included in **Appendix 2B.1**, with their molecular weights included. All genes generated are listed below, and a plasmid map of pET28b indicating the cloning site used for all constructs in this study is shown in **Figure 2-2**. Some of these sequences were not used in the final study, but were cloned, and their sequences are included in this thesis for completeness.

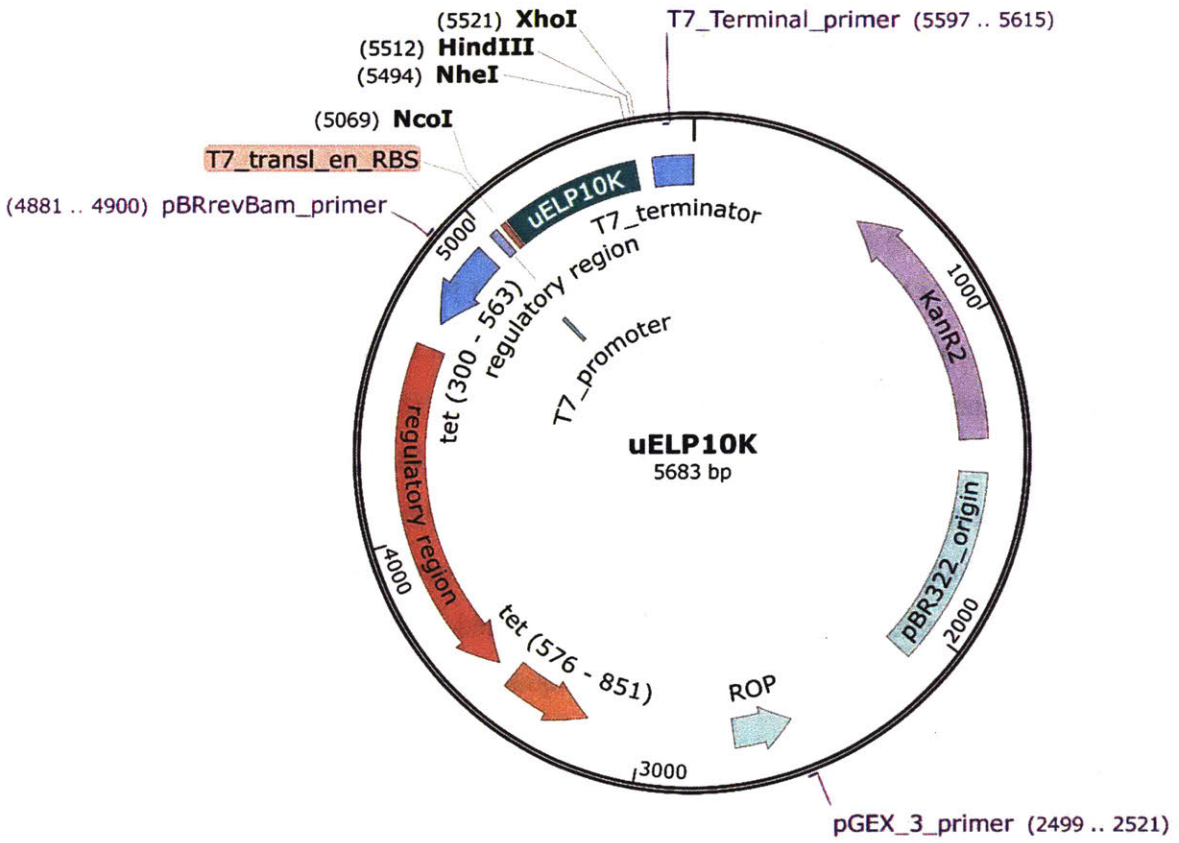


Figure 2-2. Sample plasmid map showing the cloning site (NcoI/XhoI) used in the pET28b constructs presented in this study (uELP10K)

Proteins used in study:

- Uncharged ELP, ordered gene (uELP10K)
- Uncharged ELP, doubled (uELP20K)
- Uncharged ELP, quadrupled (uELP40K)
- Charge balanced ELP, ordered gene (bELP10K)
- Charge balanced ELP, doubled (bELP20K)
- Charge balanced ELP, quadrupled (bELP40K)
- Charge balanced, increased hydrophobicity ELP, ordered gene (bhELP10K)
- Charge balanced, increased hydrophobicity ELP, doubled (bhELP20K)
- Charge balanced, increased hydrophobicity ELP, quadrupled (bhELP40K)
- Negatively charged ELP, doubled (nELP20K)
- Negatively charged ELP, quadrupled (nELP40K)

- Negatively charged, increased hydrophobicity ELP, ordered gene (nhELP10K)
- Negatively charged, increased hydrophobicity ELP, doubled (nhELP20K)
- Negatively charged, increased hydrophobicity ELP, quadrupled (nhELP40K)

Proteins not used in study, but cloned:

- Uncharged, most hydrophobic ELP, ordered gene (uhhELP10K)
- Uncharged, most hydrophobic ELP, doubled (uhhELP20K)
- Uncharged, most hydrophobic ELP, quadrupled (uhhELP40K)
- Charge balanced, most hydrophobic ELP, ordered gene (bhhELP10K)
- Charge balanced, most hydrophobic ELP, doubled (bhhELP20K)
- Charge balanced, most hydrophobic ELP, quadrupled (bhhELP40K)
- Negatively charged, most hydrophobic ELP, ordered gene (nhhELP10K)
- Negatively charged, most hydrophobic ELP, doubled (nhhELP20K)
- Negatively charged, most hydrophobic ELP, quadrupled (bhhELP40K)

Cloning of ELP-mCherry Fusions

Fusion of the mCherry gene to different ELP genes was performed using PCR cloning. The ELP genes were designed to have an NheI/HindIII cloning site after the full ELP gene but before the stop codon to allow for fusion of the PCR'd mCherry gene into the sequence. PCR was used to add NheI and HindIII restriction sites to the 5' and 3' ends of the mCherry sequence, respectively. Primers were designed according to principles listed above (see **Polymerase Chain Reaction (PCR)**). Primers used in this PCR can be found in **Appendix 2C**. mCherry DNA was prepared for ligation by performing two 50 µL PCR reactions and subsequently purifying them into one spin column purification to produce 30 µL of purified PCR'd DNA. The pET28b vector containing the full length (1500 bp) ELP was prepared via miniprep of 5 mL of culture. Both the purified PCR'd DNA as well as the vector containing the ELP were double digested with NheI-HF (NEB #R3131) and HindIII-HF (NEB #R3104) in CutSmart Buffer (NEB #B7204). The pET28b vector containing the ELP gene upstream of the cut site was then dephosphorylated with Antarctic Phosphatase (NEB #M0289). Both genes were then purified using agarose gel electrophoresis followed by gel extraction. The purified DNA was ligated at a 3:1 insert to vector molar ratio, where the insert in this case is the mCherry gene, and the vector is the pET28b vector containing the ELP gene.

Ligated plasmids were transformed into chemically competent NEB 5α cells under selection for kanamycin resistance (50 μg/mL). Clones were screened by BamHI/HindIII analytical digestion and confirmed by Sanger sequencing using universal T7 and T7 terminal primers.

2.1.1.3 High-Throughput Screening of Streptavidin Binder Library in Self-Assembled Thin Films

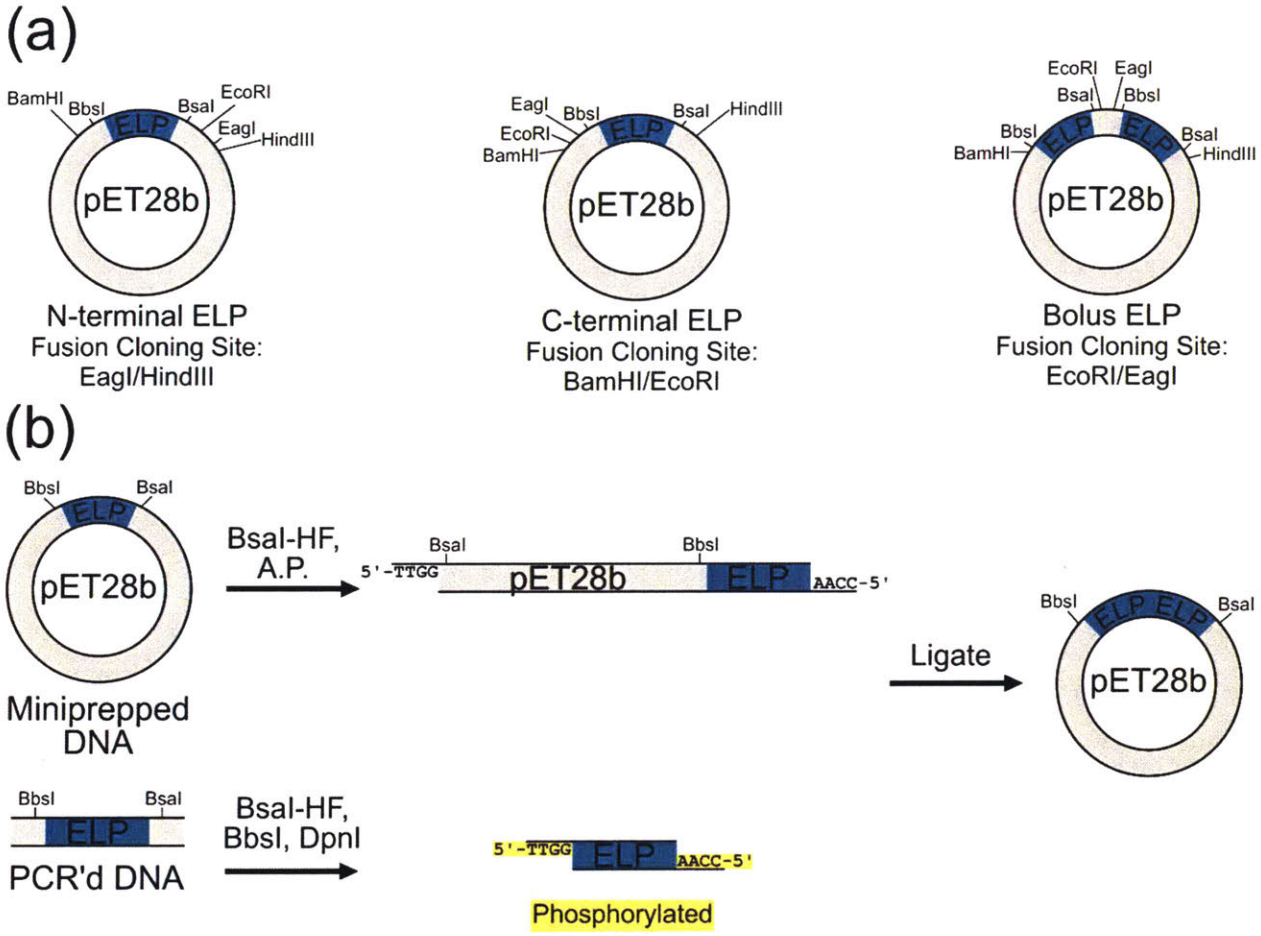


Figure 2-3. (a) Different ELP designs used in this study. (b) Cloning scheme for concatamerizing ELP in this study.

ELP Gene Design

The basis repeat unit of the ELP used in this study was the five pentapeptide repeat sequence previously found to yield the best self-assembly of ELP-mCherry fusions—(VPGVG VPGVG IPGVG VPGVG VPGVG)$_5$. [4-5] To enable efficient cloning of fusion proteins, the 375 bp ELP repeat unit was reordered from GenScript in pET28b between the NcoI and XhoI restriction sites with several considerations in mind (**Figure 2-3a**, N-terminal ELP). Upstream of the 5' end of the

ELP gene are BamHI and BbsI sites. Downstream of the 3' end of the ELP gene are BsaI, EcoRI, EagI, and HindIII sites. The BbsI and BsaI sites are both Type IIS restriction sites that were incorporated to enable concatamerization of the ELP gene, as described below. The other restriction sites (BamHI, EcoRI, EagI, and HindIII) were included to enable fusion of the ELP to the N-terminus, C-terminus or both of the termini of a target protein (**Figure 2-3a**). This cloning scheme required N-terminal and C-terminal ELP sequences to be concatamerized separately. As a result, assembly of the bolus ELP gene (as described below) would have to be performed after separate concatamerization of the desired N-terminal and C-terminal ELP sequences to the desired length, as described below.

Construction of C-terminal and Bolus ELP Genes

Construction of the C-terminal ELP pET28b plasmid (**Figure 2-3a**) was performed in a two-step process. First, an acceptor pET28b plasmid for the C-terminal ELP sequence was constructed by removing the ELP sequence between the BamHI and EcoI and replacing it with a 391 bp sequence (described below). The 375 bp ELP sequence was then reintroduced into the acceptor pET28b plasmid between the EagI and HindIII restriction sites. The details of this process were performed as follows.

The 391 bp sequence used to construct the acceptor plasmid was prepared for ligation *via* PCR of a random fragment of the mCherry gene where the primers in this PCR were designed to add a BamHI restriction site to the 5' end of the PCR product, and an EcoRI restriction site to the 3' end of the PCR product. Primers used in this PCR can be found in **A3. Table of Primers Used for PCR in this Thesis**. PCR product was purified by spin-column purification (Omega BioTek #D6492) and subsequently digested with BamHI-HF (NEB #R3136), EcoRI-HF (NEB #R3101), and DpnI (NEB #R0176) in CutSmart Buffer (NEB #B7204). The product of this digestion was purified by spin-column purification. Purification by agarose gel electrophoresis was not necessary here because the undesirable fragments cut out in the digest were sufficiently small to be eliminated via spin-column purification. The N-terminal ELP pET28b gene was prepared for ligation by double digest with BamHI-HF and EcoRI-HF in CutSmart Buffer, and subsequently dephosphorylated with Antarctic Phosphatase (NEB #M0289) in Antarctic Phosphatase Buffer (NEB #B0289). The desired 5272 bp band (including the pET28b vector) was isolated from the 396 bp ELP insert by agarose gel electrophoresis and subsequent gel extraction. The purified DNA

was ligated at a 3:1 insert to vector molar ratio, where the insert in this case is the random PCR'd DNA and the vector is the pET28b vector. Ligated plasmids were transformed into chemically competent NEB 5α cells under selection for kanamycin resistance (50 μg/mL). Clones were screened by BamHI/HindIII analytical digestion and confirmed by Sanger sequencing using universal T7 and T7 terminal primers. The sequence for this acceptor pET28b plasmid is shown in **A2.4 High-Throughput Screening of a Streptavidin Binder Library in Self-Assembled Solid Films** (Ct-vec).

The ELP gene to be reintroduced into the acceptor pET28b plasmid was prepared for ligation *via* PCR. The DNA for the N-terminal ELP (Nt-deELP10K, see below) was used as template DNA. Primers were designed to append an EagI restriction site upstream of the BbsI site at the 5' end of the ELP gene, and a HindIII restriction site downstream of the BsaI site at the 3' end of the ELP gene. The primers for this PCR can be found in **A3. Table of Primers Used for PCR in this Thesis**. PCR product was purified by spin-column purification and subsequently digested with EagI-HF (NEB #R3505), HindIII-HF (NEB #R3104), and DpnI in CutSmart Buffer. The acceptor pET28b plasmid described in the previous paragraph (Ct-vec) was prepared by miniprepping 5 mL of culture followed by double digestion with EagI-HF and HindIII-HF in CutSmart Buffer. The products of both digestions were purified by spin-column purification. Purification by agarose gel electrophoresis was not necessary here because the undesirable fragments cut out in the digests were sufficiently small to be eliminated via spin-column purification. The purified DNA was ligated at a 3:1 insert to vector molar ratio, where the insert in this case is the random PCR'd ELP gene and the vector is the acceptor pET28b vector. Ligated plasmids were transformed into chemically competent NEB 5α cells under selection for kanamycin resistance (50 μg/mL). Clones were screened by EcoRI/HindIII analytical digestion and confirmed by Sanger sequencing using universal T7 and T7 terminal primers. The sequence for this C-terminal ELP is shown below (Ct-deELP10K).

Construction of the Bolus ELP gene did not end up being necessary for completion of this thesis; however, N-terminal ELPs of ~10 kDa and ~20 kDa ELP molecular weight and C-terminal ELP sequences with ELPs of ~10 kDa, ~20 kDa, and ~30 kDa were generated (see below). Construction of a Bolus ELP gene can easily be achieved by starting with N-terminal ELP and C-terminal genes with an ELP of the desired length. To construct the Bolus ELP gene, one would digest the N-terminal ELP gene with EagI/HindIII and dephosphorylate the DNA. Simultaneously, the C-

terminal ELP would then be digested with EagI/HindIII, and the ELP band would be gel extracted. The C-terminal ELP would then be ligated in (as an insert) to the N-terminal ELP gene at a 3:1 insert:vector molar ratio.

ELP Concatamerization

N-terminal and C-terminal ELP genes were concatamerized using the same cloning scheme (**Figure 2-3b**). The ELP insert used for concatamerization was produced by PCR of the ELP gene using the N-terminal ELP in pET28b as the template DNA. Primers for PCR were well outside the highly repetitive sequence of the ELP gene (342 and 38 bp away from the repetitive sequence for the forward and reverse primers, respectively), and thus did not produce multiple PCR products that might be expected when PCR'ing highly repetitive sequences. The PCR product was also confirmed by forward and reverse Sanger sequencing. It should be noted that using PCR to extract the ELP gene of interest is limited to repetitive ELP sequences that are 400 bp or less in length, as problems with PCR'ing highly repetitive genes starts to present an issue at larger ELP lengths. The primers used for this PCR can be found in **A3. Table of Primers Used for PCR in this Thesis**. The PCR product was digested with BsaI-HF (NEB #R3535), BbsI (NEB #R0539) and DpnI in NEBuffer 2.1 (NEB #B7202). While a high fidelity version of the BbsI enzyme exists from NEB, it was found that it did not consistently cut the PCR product as expected, likely due to sensitivities to buffer conditions; thus, it was critical in this process to use the non-HF BbsI enzyme in NEBuffer 2.1. The pET28b vector containing the ELP gene to be elongated was prepared by miniprepping out of 5 mL of culture. This gene was digested with BsaI-HF in CutSmart buffer and subsequently dephosphorylated with Antarctic Phosphatase. The products of both digestions were purified by spin-column purification. Purification by agarose gel electrophoresis was not necessary here because the undesirable fragments cut out in the digests were sufficiently small to be eliminated via spin-column purification. The purified DNA was ligated at a 10:1 insert to vector molar ratio, where the insert in this case is the PCR'd ELP DNA and the vector is the N-terminal or C-terminal pET28b vector. Ligated plasmids were transformed into chemically competent NEB 5α cells under selection for kanamycin resistance (50 μg/mL). Ligation transformant was subject to analytical digest with BamHI/HindIII as described in **Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins, ELP Concatamerization** to identify different concatamer lengths produced. In this case of this cloning scheme, the majority of the ELP clones produced were elongated by 375 bp (a doubling of the original ELP gene) with

a small fraction of them being elongated by 750 bp (a tripling of the original ELP gene). This is in contrast to the approach presented in the previous section, which typically resulted in a majority of genes being a single ELP repeat unit with uncommon instances of ELP doubling and rare instances of ELP tripling. Isolated clones were screened by BamHI/HindIII analytical digestion and confirmed by Sanger sequencing using universal T7 and T7 terminal primers. The sequences for the different ELPs produced by this method can be found in **A2.4 High-Throughput Screening of a Streptavidin Binder Library in Self-Assembled Solid Films**. The plasmid map for the sequences generated by this method is the same as that shown in **Figure 2-2**, with the only differences being in the inserted sequences.

Genes generated by this cloning scheme:

- N-terminal ELP gene, ordered (Nt-deELP10K)
- N-terminal ELP gene, doubled (Nt-deELP20K)
- C-terminal ELP vector (Ct-vec)
- C-terminal single ELP (Ct-deELP10K)
- C-terminal ELP, doubled (Ct-deELP20K)
- C-terminal LEP, tripled (Ct-deELP30K)

ELP-Sso7d-ZE Gene Design and Construction

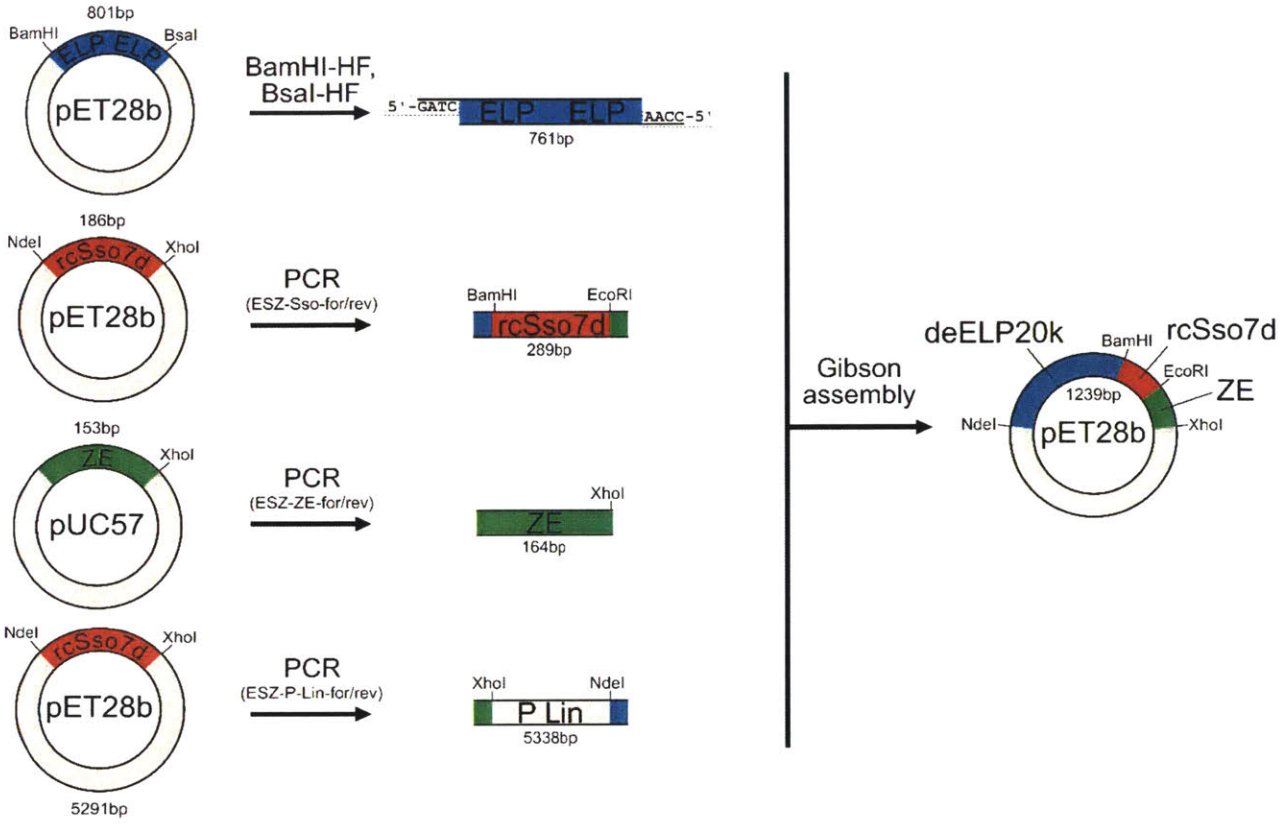


Figure 2-4. Schematic of the adapted Gibson assembly scheme used to produce the initial deELP20k-rcSso7d-ZE (ESZ) plasmid construct.

We sought to investigate the functional additivity of a triblock protein species incorporating the elastin-like protein, a streptavidin-binding rcSso7d variant, and an ordered-assembly-inducing ZE peptide. The construction of the ELP-Sso7d-ZE pET28b plasmid (**Figure 2-4**) was performed in a two-step Gibson assembly process, utilizing constructs isolated from three separate plasmids: pET28b-Nt-deELP20K, pET28b-rcSso7d.SA.1[7], and pUC57-ZE. Individual fragments were prepared via restriction digest (in the case of the Nt-deELP20K gene) or PCR, and these fragments were then combined in a Gibson assembly reaction in order to yield an in-frame ELP-Sso7d-ZE construct.

Due to the highly repetitive nature of the Nt-deELP20K construct, this gene could not be amplified via PCR. Instead, the deELP20k gene fragment was prepared by subjecting the pET28b-Nt-deELP20K plasmid to a restriction digest reaction, via a *BamHI*/*BsaI* double digest. This digested fragment would be included in the Gibson assembly reaction, and the 5' overhangs resulting from these restriction digestion reactions would be removed by the exonuclease function. These internal

restriction sites can then be destroyed (without changing the amino acid sequence of the encoded protein) by ensuring that the sequence overlaps of adjacent fragments substitute degenerate codons in place of the original DNA sequence. Both of the selected restriction sites were destroyed in order to ensure that those enzymes would uniquely digest the resulting construct at the desired locations, and the *BsaI* site was specifically selected as the 3' restriction site in order to ensure that the stop codon would also be removed from the digested gene fragment.

Amplicons of the linearized pET28b plasmid backbone, rcSso7d.SA.1 gene, and ZE gene were prepared via standard PCR, using the primer pairs denoted in **Table 2-1**. Primers were designed to append the appropriate sequence overlaps to the 5' and 3' termini of each PCR product, while maintaining a GC ratio within (or just above) the range of 40-60%. Specifically, the linearized plasmid backbone amplicon was modified to overlap with the 5' end of the ELP product (introducing a degenerate AGC codon (S) in the place of the TCC codon originally present in the *BamHI* site) and the 3' end of the ZE product. The rcSso7d.SA.1 amplicon was modified to overlap with the 3' end of the ELP product (introducing a degenerate GAA codon (E) in the place of the GAG codon originally present in the *BsaI* site and appending a flexible $(GGGGS)_2$ linker between the ELP and rcSso7d protein blocks) and the 5' end of the ZE product. The ZE amplicon was modified to overlap the 3' end of the rcSso7d product and the 5' end of the linearized plasmid backbone product. A stop codon was removed from the rcSso7d product, and introduced at the 3' end of the ZE amplicon. Unique restriction enzyme sites were introduced between each gene fragment, yielding the final construct: *NdeI*-ELP-*BamHI*-rcSso7d-*EcoRI*-ZE-*XhoI*. This construct will be referred to as ELP-rcSso7d-ZE, or may be shortened to ESZ.

Table 2-1. Oligonucleotide sequences of primers used in Gibson assembly of the ESZ construct

#	Oligo Name	DNA Sequence (5'-3') (*NdeI*, *XhoI*, *BamHI*, *and* *EcoRI* sites)	Annealing Temp. (°C)
1	ESZ-Sso-for	GGGTGTAGGCGTTCCAGGCGTTGGTGAAACCGAAT TTGTAGAACGGCCGGGTGGTGGTGGTAGCGGTGGT GGC**GGATCC**ATGGCAACCGTGAAATTC	74.5

2	ESZ-Sso-rev	GCCACCACCGCTACCACC**GAATTC**TTGCTTTTCCAG CATCTG	69.6
3	ESZ-ZE-for	TCGGTGGTAGCGGTGGTGGCGGTTCACTGGAGAT	71.2
4	ESZ-ZE-rev	**CTCGAG**TTACAGCGGACCATAACGGGTACG	64.3
5	ESZ-P-lin-for	ATGGTCCGCTGTAA**CTCGAG**CACCACCACCACCAC CACTGAGAT	56.8
6	ESZ-P-lin-rev	GTACACCAACAGTCTTCCCGCTTCCAAC**CATATG**GC TGCCGCGCGGCACCAGGCCGCTGC	65.0
7	rcSso7d-for	AGGCAGTCT**CATATG**GCAACCGTGAAAT	63.3

Reagents and materials

GenCatch Gel Extraction kits and GenCatch Plasmid Miniprep kits were purchased from Epoch Life Sciences (Missouri City, TX, USA). All cloning enzymes (BsaI-HF, BamHI-HF, Phusion High-Fidelity T7 Polymerase, and the Gibson Assembly Master Mix) were purchased from NEB. Unless otherwise noted, primers were synthesized by IDT.

Gibson assembly process

Plasmid DNA for each of the three constituent constructs was prepared by miniprepping DH5α cells grown to saturation in a 5 mL overnight culture. Purified plasmid DNA was quantified using a Tecan Plate Reader and three separate PCR mixtures were prepared, using 100 ng each of the pET28b-rcSso7d.SA.1 plasmid (including the ESZ-Sso-for/rev primer pair for amplification of the rcSso7d gene), the pUC57-ZE plasmid (including the ESZ-ZE-for/rev primer pair for amplification of the ZE gene), and the pET28b-rcSso7d.SA.1 plasmid (including the ESZ-P-lin-for/rev primer pair for linearization and amplification of the plasmid backbone). In addition to 100 ng of the appropriate miniprepped plasmid DNA, each PCR reaction contained 10 µL of 5x Phusion HF polymerase buffer, 1 µL of dNTP mix (10 mM of each base), 1 µL each of the forward

and reverse primers at a starting concentration of 10 μM, 1 μL of Phusion HF polymerase, and PCR-grade water, filled to a final volume of 50 μL.

Normally, each PCR process would be conducted separately, in order to incorporate an elongation time tailored to the length of the desired amplicon, and an annealing temperature which is 5°C below the lower of the two melting temperatures for the PCR primers. However, given significant time constraints imposed upon this process, acceptable shortcuts in the cloning process were taken. While this is not considered standard practice, intervening production steps (gel electrophoresis and DNA sequencing) were used to verify the size and identity of the various DNA fragments.

In light of these time constraints, two of the three PCR processes (the rcSso7d and ZE amplification reactions) were run together, with an elongation time of 30 seconds, and a melting temperature of 59.3°C (given the melting temperature of the ESZ-ZE-rev primer). The linearization of the plasmid backbone was conducted in a separate thermocycler, using an elongation time of 5.5 minutes (to accommodate the expected 5.5 kb product, assuming an elongation rate of 1 kb/minute), and a melting temperature of 51.8°C (given the low melting temperature of the ESZ-P-lin-for primer)

PCR was otherwise conducted via standard protocols, with thirty cycles of (i) denaturation at 95°C for 30 seconds, ii) primer annealing at the appropriate temperature for 30 seconds, and (iii) extension at 72°C for the pre-determined duration. Each PCR mixture was then loaded into a 1% agarose gel, and separated via gel electrophoresis, run at 100V for 45 minutes. Though unexpected product bands were observed for each reaction, strong bands were also observed at the expected product sizes (rcSso7d: 289 bp; P Lin: 5,338 bp; ZE: 164 bp) (**Figure 2-5**a). These product bands were excised using a clean razor blade, extracted from the agarose gel using a GenCatch Gel Extraction kit, and quantified using a Tecan plate reader.

In a separate process, the pET28b-deELP20k plasmid was digested in a BsaI-HF/BamHI-HF double digest reaction. Approximately 1,750 ng of plasmid was mixed with 5 μL of 10x Cutsmart Buffer, 2 μL of BsaI-HF, 2 μL of BamHI-HF, and PCR-grade water, filled to a final volume of 50 μL. This mixture was incubated at 37°C for one hour, after which the enzymes were inactivated via a ten-minute hold at 65°C. This mixture was also separate via a 1% agarose gel, and a faint

excision band was observed at the expected product size (ELP RD: 761 bp) (**Figure 2-5**b). This product band was also excised, gel extracted, and measured using a Tecan plate reader.

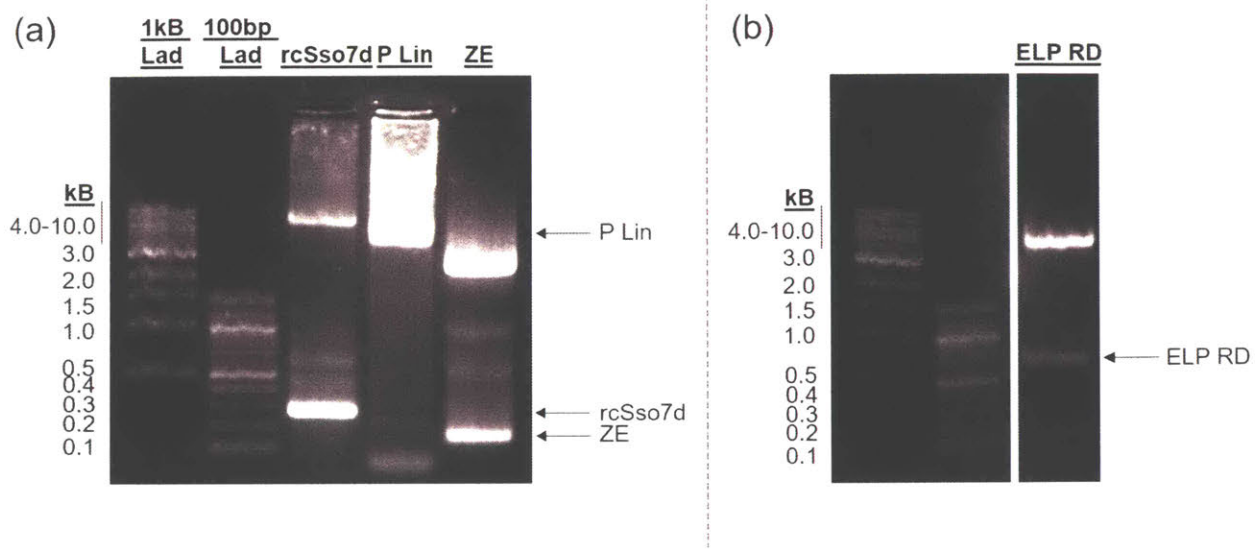


Figure 2-5. 1% agarose gels of (a) PCR amplicons of rcSso7d, the linearized pET28b backbone, and ZE, denoted by arrows pointing to strong bands in each product's respective lane, and (b) the digested ELP product, similarly denoted by an arrow indicating the desired product. Note that (b) was edited to remove unrelated products within internal lanes.

Gibson assembly

All gel extracted products were quantified, and a Gibson Assembly reaction with these four fragments was set-up, following manufacturer guidance where possible. In particular, NEB recommends that for the assembly of 4-6 DNA fragments, 0.2-1.0 pmol of total DNA be used in the reaction. Additionally, it is recommended that 50-100 ng of the vector be used, with a 2-3 fold excess of inserts (or 5-fold excess of inserts where the insert size is less than 200 bp). NEB also notes that the total volume of DNA should ideally not exceed 20% of the reaction volume. However, DNA quantification showed that the ELP restriction digest product was insufficiently concentrated to meet all of these criteria. Given project time constraints, the reaction was still assembled, using the parameters noted below:

Sample	Conc (ng/uL)	MW (ng/pmol)	Molar Conc (pmol/uL)	Moles Reqd (pmol)	Mass Reqd (ng)	Volume (uL)
ESZ P Full PCR	161.5666667	3300	4.90E-02	0.017	55.300	0.342
ELP RD	0.433333333	499.2	8.68E-04	0.008	4.19	9.680
ESZ Sso PCR	39.6	180.2	2.20E-01	0.084	15.10	0.381
ESZ ZE PCR	12	105.953	1.13E-01	0.084	8.88	0.740
			Total moles DNA	0.193	Total volume DNA	11.143

Sufficient linearized plasmid was used to have greater than 50 ng of plasmid DNA in the reaction, and a 5x molar excess of the rcSso7d and ZE PCR products was used. However, all of the available ELP RD product was used, given its low concentration, and still it was only present in a ~0.5-fold molar ratio, relative to the linearized plasmid. In total, 0.193 moles of DNA fragments were added to the reaction mixture. In order to adjust for the higher total volume of PCR-grade water/DNA fragments, 11.14 µL of Gibson Assembly Master Mix was added (rather than the standard 10 µL). This reaction was incubated at 50°C for one hour, and then placed on ice.

Three electrocompetent BL21(DE3) *E. coli* cells (40 µL) were thawed on ice, and 1 µL of the Gibson assembly reaction was diluted into 2 µL of PCR-grade water, in keeping with NEB protocols. Each bacterial aliquot was mixed with 1 µL of the diluted Gibson assembly product, and these mixtures were transferred to pre-chilled electroporation cuvettes. Cells were electroporated, and 1 mL of room-temperature SOC medium was added to the electroporation cuvette immediately after transformation. Transformation cultures were incubated at 250 rpm and 37°C for one hour, after which the entirety of each transformation reaction was plated onto a pre-warmed LB-kan agar plates.

Colony PCR

Following one day of outgrowth in a stationary incubator at 37°C, twenty *E. coli* colonies were observed across the three plates. All twenty bacterial colonies were picked with a pipette tip, and each colony was re-suspended in 50 µL of PCR-grade water, in order to prepare for colony PCR. A colony PCR master mix was prepared, with sufficient volume for twenty 10 µL reactions, each containing 0.25 µL of dNTPs (10 mM each), 2 µL of 5x Phusion Polymerase buffer, 0.6 µL of the rcSso7d-for primer (10 µM; see **Table 2-1**), 0.6 µL of the ESZ-ZE-rev primer (10 µM), 0.2 µL of Phusion polymerase, and 5.35 µL of PCR-grade water. This master mix was split into twenty 9-µL aliquots, and 1 µL of the bacterial colony resuspension was added to each aliquot. The use of the ESZ-Sso-for and ESZ-ZE-rev primers ensured that singular PCR products would only be generated from constructs in which the rcSso7d gene and the ZE gene had successfully been fused.

Colony PCR reactions were subjected to lysis at 95°C, followed by 35 amplification cycles incorporating (i) denaturation at 95°C for 30 seconds, (ii) primer annealing at 59.3°C for 30 seconds, and (iii) extension at 72°C for 30 seconds (for an expected product length of 366 bp). Of the twenty bacterial colonies, only one yielded a product band at the expected location; all others yielded a variety of product bands (**Figure 2-6**). The low efficiency of the Gibson assembly reaction is likely due to the very low abundance of the ELP restriction digest product. The colony which yielded this PCR product was grown overnight in a 5 mL LB-kan culture tube and miniprepped, and the sequence of the complete ESZ construct was confirmed via 5' and 3' sequence confirmation.

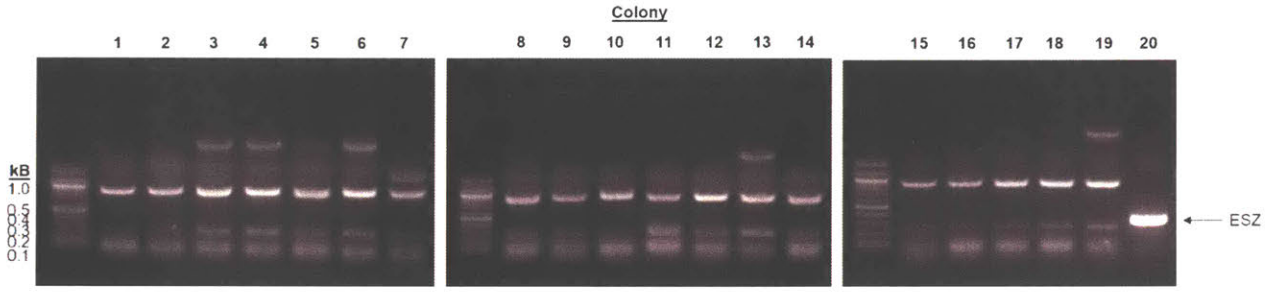


Figure 2-6. 1% agarose gels of colony PCR products from ESZ Gibson assembly transformants. Colonies 1-19 all show multiple nonspecific product bands, and only Colony 20 shows a single bright product band at the expected size (366 bp).

2.1.1.4 Cloning for Other Projects

Tyrosine-Containing ELPs

A second ELP was considered for use in the **High-Throughput Screening of Streptavadin Binder Library in Self-Assembled Thin Films** project that contained tyrosine residues in the ELP repeat sequence. This sequence was designed so that the tyrosine residues in the ELP and on the globular protein could be crosslinked *via* photoactivated RuBiPy crosslinking, and this crosslinking would render the self-assembled film permanently immobilized.[7] While this crosslinking strategy was eventually abandoned in favor of a glutaraldehyde strategy, the cloning for this tyrosine-containing ELPs (here called ELPY) is described here. A description of the work done on crosslinking of tyrosine-containing thin films can be found in **A4. Ru(bpy) Catalyzed Crosslinking of Tyrosine-Containing ELP Films**. The repeat unit for this ELP was (VPGVG VPGGG IPGYG VPGGG VPGVG)$_5$. The design and cloning scheme for this gene was identical

to that described in the previous section (see **High-Throughput Screening of Streptavadin Binder Library in Self-Assembled Thin Films**: ELP Gene Design, Construction of C-terminal and bolus ELP Genes, and ELP Concatamerization). The different ELPY genes generated for this project are listed below. Full DNA and amino acid sequences for these ELPs can be found in **A2.5 Tyrosine-Containing ELPs**.

In addition to the ELPs themselves, several N-terminal ELPY-mCherry fusion proteins were cloned. The cloning scheme used here was similar to that reported in **Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins**, Cloning of ELP-mCherry Fusions. First, PCR was used to append an EcoRI restriction site upstream of the 5' end of the mCherry gene and a HindIII restriction site downstream of the 3' end of the mCherry gene. Primers for this PCR can be found in **A3. Table of Primers Used for PCR in this Thesis**. This PCR product was subsequently purified via spin-column purification, and double digested with EcoRI-HF, HindIII-HF, and DpnI in CutSmart buffer. N-terminal ELPY genes with ELPs 750 bp, 1125 bp, and 1500 bp in length were prepared by miniprepping of separate 5 mL cultures. These different plasmids containing ELPY genes were double digested with EcoRI-HF and HindIII-HF in CutSmart buffer, and subsequently dephosphorylated with Antarctic phosphatase. Both the digested PCR'd mCherry DNA and the different N-terminal ELPY vectors were purified via spin-column purification. Purification by agarose gel electrophoresis was not necessary here because the undesirable fragments cut out in the digests were sufficiently small (< 100 bp) to be eliminated via spin-column purification. The purified DNA was ligated at a 3:1 insert to vector molar ratio, where the insert in this case is the PCR'd mCherry DNA and the vectors are the N-terminal ELPY plasmids. Ligated plasmids were transformed into chemically competent NEB 5α cells under selection for kanamycin resistance (50 μg/mL). Clones were screened by BamHI/HindIII analytical digestion and confirmed by Sanger sequencing using universal T7 and T7 terminal primers. The sequences for the different ELPY-mCherry fusions produced by this method are listed below. Full amino acid sequences can be found in **A2.5 Tyrosine-Containing ELPs**.

Genes generated for this section:

- N-terminal tyrosine-containing ELP gene, ordered (Nt-ELPY10K)
- N-terminal tyrosine-containing ELP gene, doubled (Nt-ELPY20K)
- N-terminal tyrosine-containing ELP gene, tripled (Nt-ELPY30K)

- N-terminal tyrosine-containing ELP gene, quadrupled (Nt-ELPY40K)
- C-terminal ELPY vector (Ct-ELPY vec)
- C-terminal single tyrosine-containing ELP (Ct-ELPY10K)
- Nt-ELPY10K-mCherry
- Nt-ELPY20K-mCherry
- Nt-ELPY30K-mCherry
- Nt-ELPY40K-mCherry

2.1.2 Protein Expression & Purification

2.1.2.1 General Techniques

All expressions were carried out in either Lysogeny Broth (LB) or Terrific Broth (TB). All media were prepared with MilliQ purified water and sterilized by autoclaving. All expressions were performed under antibiotic selection using 50 μg/mL kanamycin. Antibiotic stocks were prepared at 1000X concentrations (50 mg/mL kanamycin) in MilliQ water, sterile filtered using a syringe, aliquoted into sterile tubes in 1 or 5 mL aliquots, and stored at -20 °C. To prepare antibiotic-supplemented media, concentrated antibiotics were added to autoclaved media after cooling below 60 °C at a dilution of 1000:1. Isopropyl β-D-1-thiogalactopyranoside (IPTG) used to induce non-well plate expressions was prepared at 1 M in MilliQ water, sterile filtered, aliquoted, and stored at -20 °C prior to use.

2.1.2.2 Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins

ELP Expression and Purification

All plasmids containing the 1500 bp ELP genes of interest were transformed into *Escherichia coli* (*E. coli*) strain Tuner(DE3). Charged ELPs (bELP, bhELP, nELP, nhELP) were expressed in TB (1 L) supplemented with 50 μg/mL kanamycin in baffled shake flasks (2.5 L). Expressions were inoculated with 10 mL starter cultures grown in LB supplemented with 50 μg/mL kanamycin at 37 °C overnight. Expressions were grown to OD_{600} = 1.3-1.9 at 30 °C before induction with 0.5 mM IPTG. Immediately after induction, the temperature of incubator was dropped to 20 °C. Expressions were grown for 17-18 h after induction. Expression conditions for charged ELPs were selected to match the expression conditions that were most successful for corresponding ELP-mCherry fusions (which was evaluated based on the development of the characteristic pink

mCherry color after expression). Expression conditions were not optimized beyond this. The uncharged ELP (uELP) was expressed in 5 L TB in a fermenter with a 7 L working volume. Expression of uELP was inoculated with a 50 mL starter culture grown in LB supplemented with 50 μg/mL kanamycin overnight at 37 °C in an orbital shaker in a 250 mL baffled Erlenmeyer flask. After inoculation, uELP expression was carried out at 30 °C for 18 h with no induction. Expression conditions for uELP were based on procedures published by M. J. Glassman as well as reports out of the Chilkoti lab that ELP expression yields are often improved by not inducing and instead by allowing leaky expression from the T7 promoter to accumulate over time. [2, 8] Expression conditions were not optimized beyond this original procedure. Cells from all expressions were harvested by centrifugation (4,816xg for 15 min at 4 °C). uELP cells were resuspended in MENT buffer (3 mM $MgCl_2$, 1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 10 mM trizma, pH 7.5). bELP, bhELP, nELP, and nhELP cells were resuspended in 100 mL of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole) for every 30 g of wet cell mass. Lysis buffer used for bELP and bhELP resuspensions was adjusted to pH 8. Lysis buffer used for nELP and nhELP resuspensions was adjusted to pH 10 to decrease positive charge on lysine-rich proteins in the lysate, and thus decrease complexation between the ELP and lysine-rich proteins in the lysate. Resuspended cells were frozen at -80 °C overnight or longer. A table of expression conditions and yields is provided in **Table 2-2**. Analysis tracking protein expression over time by SDS-PAGE was not performed for any of the proteins produced in this work.

Table 2-2. Summary of expression and purification conditions, and yields for all proteins used in this study.

Protein	Media Used	Expression Format	Lysozyme Added?	Freeze Thaw Cycles	Final Yield
uELP	TB	Fermenter	Yes	1	76 mg/L
nELP	TB	Shake flask	No	3	99 mg/L
nhELP	TB	Shake flask	No	3	83 mg/L
bELP	TB	Shake flask	Yes	1	130 mg/L
bhELP	TB	Shake flask	Yes	1	96 mg/L

uELP-mCherry	LB	Shake flask	Yes	1	98 mg/L
nELP-mCherry	LB	Shake flask	Yes	1	38 mg/L
nhELP-mCherry	LB	Shake flask	Yes	1	50 mg/L
bELP-mCherry	LB	Shake flask	Yes	1	43 mg/L
bhELP-mCherry	LB	Shake flask	Yes	1	33 mg/L

Resuspended cells for uELP, bELP, and bhELP were thawed and incubated with 1 mg/mL lysozyme at 4 °C for 1-2 h. Resuspended cells for nELP and nhELP were thawed and refrozen at -80 °C three times to aid in cell lysis. Lysozyme was not used in lysis of cells containing negatively charged ELPs because the high amount of positive charge on lysozyme leads to aggregation with the negatively charged ELP, making it difficult to purify. After freeze/thaw cycles or incubation with lysozyme, resuspended cells were sonicated three times, 10 min each at power level 5, 50% duty cycle, with cooling to $\leq$ 15 °C between sonications. Lysates were clarified by high-speed centrifugation (26,700xg for 1 h at 4 °C).

uELP was purified first by two rounds of inverse transition cycling (ITC). [9-10] In a single ITC cycle, the uELP was first selectively precipitated by adjusting the NaCl concentration of the solution (in the first cycle, this would be the clarified lysate) to 1.5 M by adding 5 M NaCl in 20 mM pH 8 tris buffer. In the first cycle, DNase I and RNase A were also added to the clarified lysate at a concentration of ~0.1 mg/mL each to degrade any residual DNA and RNA in the lysate. Samples were incubated at 37 °C overnight to ensure complete protein precipitation. Precipitant was separated by centrifugation at 37 °C (1 h, 26,700xg), and supernatant was discarded. Pellets from this precipitation were resuspended in MENT buffer (3 mM $MgCl_2$, 1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 10 mM trizma, pH 7.5, typically added to be one half-two thirds the original clarified lysate volume) at 4 °C on a rocker for at least 3 h. Insoluble aggregates were separated from the solubilized uELP by centrifugation at 4 °C (1 h,

26,700xg). A sample SDS-PAGE gel showing the undesirable fractions produced in this process is shown in **Figure 2-7a**. The final uELP fraction produced after two ITC cycles was then dialyzed against MilliQ water (6,000-8,000 MWCO, 5-7 changes of MilliQ water separated by at least 3 h). Dialysis was set up such that ≤ 300 mL of protein solution was being dialyzed against 4 L of water for each change. Over the course of this first dialysis, some insoluble aggregate that did not contain the desired uELP product typically formed in the dialysis bag. Thus, after removal from dialysis, the product was subject to centrifugation at 4 °C (1 h, 26,700xg) to remove insoluble aggregates prior to any further purification. Solid urea was then dissolved in the dialyzed uELP solution to adjust the urea concentration to 6 M. 2 M pH 8 tris was also added to adjust the tris concentration of the solution to 20 mM, pH 8. uELP was then purified using fast protein liquid chromatography (FPLC) under denaturing conditions (using 20 mM tris, 6 M urea, pH 8 as the running buffer) with an anion exchange column (HiTrap Q HP anion exchange chromatography column, GE Life Sciences). After sample application to the column, it was washed with 40 mL of 20 mM tris, 6 M urea, pH 8 buffer prior to elution over 30 column volumes using a linear gradient that increased salt concentration from 0 to 200 mM NaCl, again, in 20 mM tris, 6 M urea, pH 8. Content and purity of uELP in different fractions were evaluated using SDS-PAGE that sampled every 3 wells over the entire gradient. It was found that uELP did not bind to the column and instead eluted in the flow through as the protein solution was applied to the column (**Figure 2-7**c). UV traces for the FPLC can be found in **Figure 2-7b**. Purified uELP was collected and dialyzed again into Milli-Q purified water (6,000-8,000 MWCO, 10 changes of MilliQ water separated by at least 3 h). Dialyzed uELP product was then lyophilized, and stored at -20 °C.

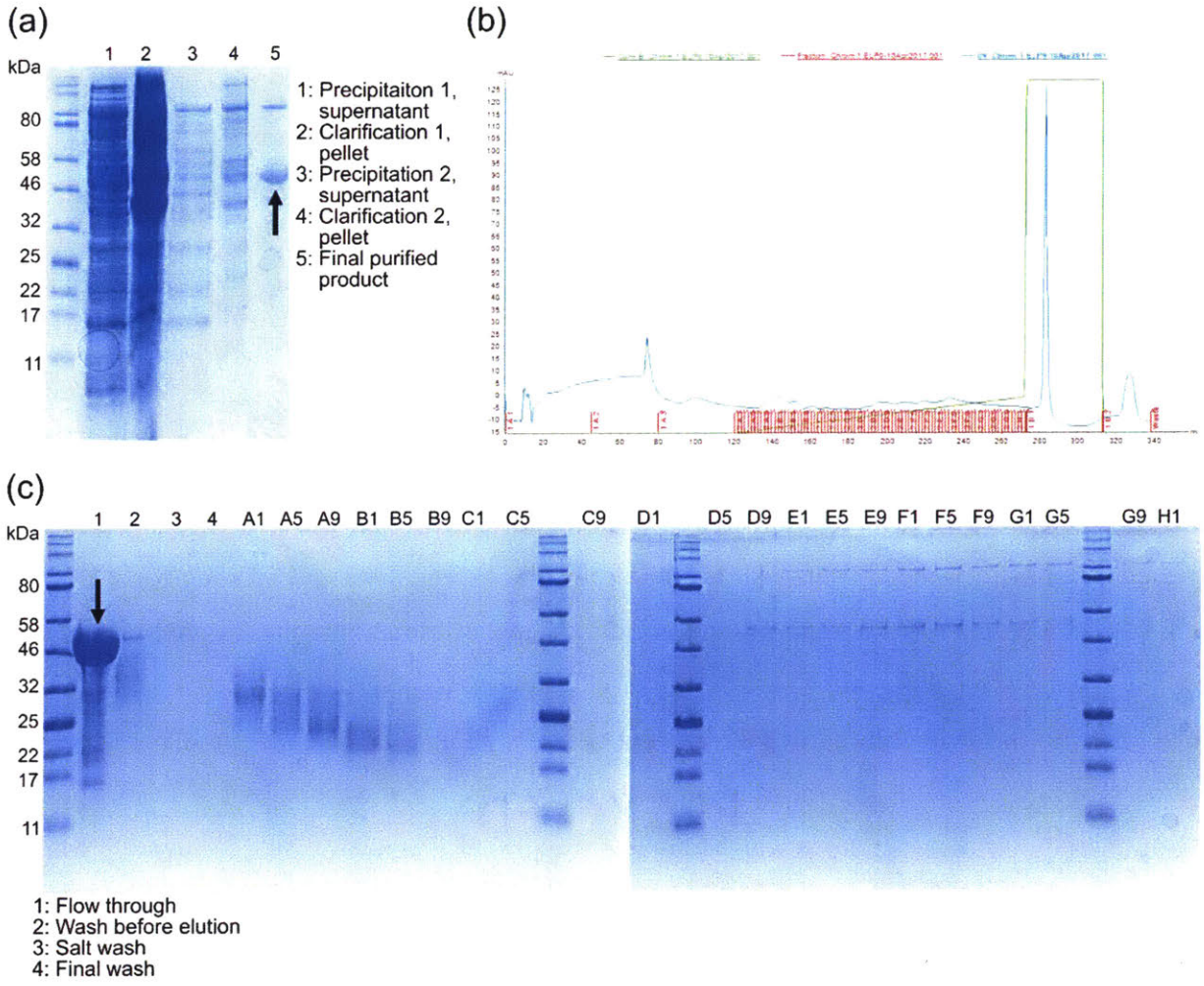


Figure 2-7. (a) Sample SDS-PAGE gel showing discarded fractions (lanes 1-4) and purified uELP (lane 5) after 2 rounds of NaCl-induced precipitation. (b) A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 added at each fraction in FPLC (fractions are labeled in red along x-axis). A zoomed trace where the elution fractions (120-250 mL) are visible is provided in **Figure A-8**). (c) SDS-PAGE gel showing different fractions produced by FPLC run in (b). The collected fraction here was the flow through (lane 1), as none of the uELP bound to the anion exchange column. Black arrows in PAGE gels indicate expected band for uELP.

bELP, bhELP, nELP, and nhELP were all purified first by Ni-NTA affinity chromatography (all buffers at pH 8 for bELP/bhELP, and all buffers at pH 10 for nELP/nhELP). Clarified lysates of all ELPs were bound to Ni-NTA agarose resin by mixing overnight at 4 °C on a rocker. After collecting flow through, resin was washed three times with 5 column volumes (25 mL bed volume

of Ni-NTA in a column of diameter 5.5 cm) of wash buffer for each wash (50 mM sodium phosphate, 300 mM sodium chloride, 20 mM imidazole, pH 8 or 10 for bELP/bhELP or nELP/nhELP, respectively). ELPs were then eluted in 5 fractions, 2 column volumes per fraction, with elution buffer (50 mM sodium phosphate, 300 mM sodium chloride, 250 mM imidazole, pH 8 or 10 for bELP/bhELP or nELP/nhELP, respectively). Fractions containing the ELP of interest were identified by SDS-PAGE (sample gels for a typical Ni-NTA purification for all ELPs discussed here are shown in **Figure 2-8**), combined, and dialyzed against MilliQ water (6,000-8,000 MWCO, 5-7 changes of MilliQ water separated by at least 3 hrs). Dialysis was set up such that ≤ 300 mL of protein solution was being dialyzed against 4 L of water for each change. Over the course of this first dialysis, some insoluble aggregate that did not contain the desired ELP product typically formed in the dialysis bag. Thus, after removal from dialysis, the product was subject to centrifugation at 4 °C (1 h, 26,700xg) to remove insoluble aggregates prior to any further purification.

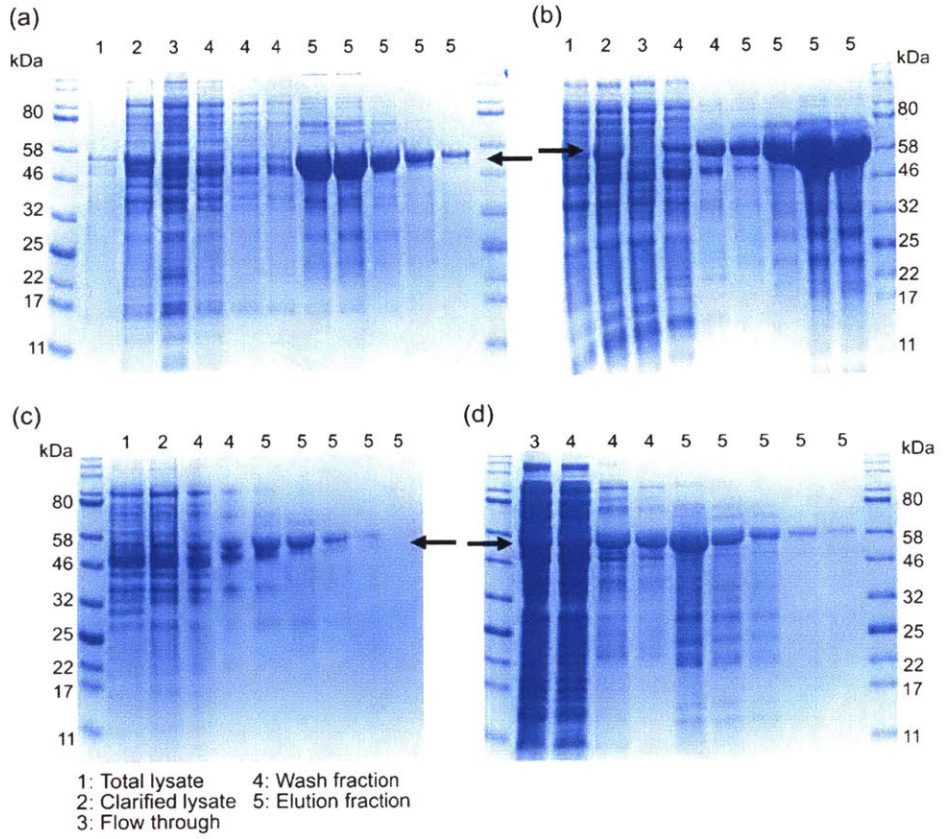


Figure 2-8. Coomassie blue stained SDS-PAGE gels on different fractions produced in Ni-NTA column purification of (a) bELP, (b) bhELP, (c) nELP, and (d) nhELP. Black arrows next

to/overlaid with each gel indicate the location in the gel where the desired product is expected to run.

To prepare bELP, bhELP, nELP, and nhELP solutions for FPLC, 2 M tris was added (pH 8 for bELP/bhELP, pH 10 for nELP/nhELP) to adjust final tris concentration of the solutions to 20 mM. FPLC was performed for these ELPs under native conditions (20 mM tris running buffer at pH 8 or pH 10) with an anion exchange column (HiTrap Q HP anion exchange chromatography column, GE Life Sciences). bELP/bhELP were eluted over 150 mL using a linear gradient that increased salt concentration from 0 to 200 mM NaCl in 20 mM tris (pH 8 for bELP/bhELP, pH 10 for nELP/nhELP). nELP/nhELP were eluted over 150 mL using two linear gradients. The first gradient increased salt concentration from 0 to 300 mM NaCl over 75 mL, and the second gradient increased salt concentration from 300 mM to 500 mM NaCl over 75 mL. Content and purity of ELP in different fractions were evaluated using SDS-PAGE that sampled every 3 wells over the entire gradient. Sample FPLC traces for all ELPs mentioned here as well as corresponding SDS-PAGE gels are shown in **Figure 2-9, Figure 2-10, Figure 2-11,** and **Figure 2-12**. Fractions containing pure ELP were collected, combined and dialyzed again into Milli-Q purified water (6,000-8,000 MWCO, 10 changes of MilliQ water separated by at least 3 h). Dialyzed ELPs were then lyophilized, and stored at -20 °C.

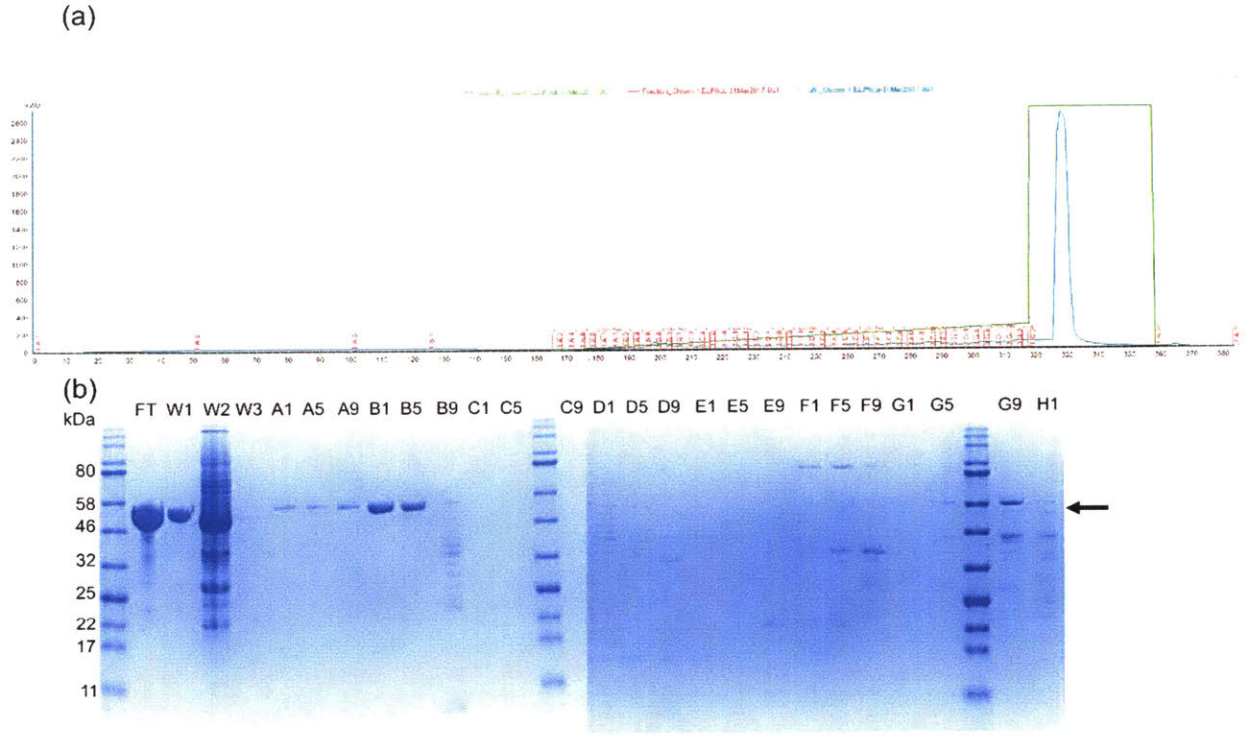


Figure 2-9. (a)A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 for bELP. Fractions corresponding to the gel below are indicated by labelled red lines along the x-axis on this plot. A zoomed trace where the elution fractions (160-320 mL) are visible is provided in **Figure A-9**). (b) Corresponding SDS-PAGE gels showing fractions produced in FPLC fractionation for bELP. Black arrow indicates expected band for this protein.

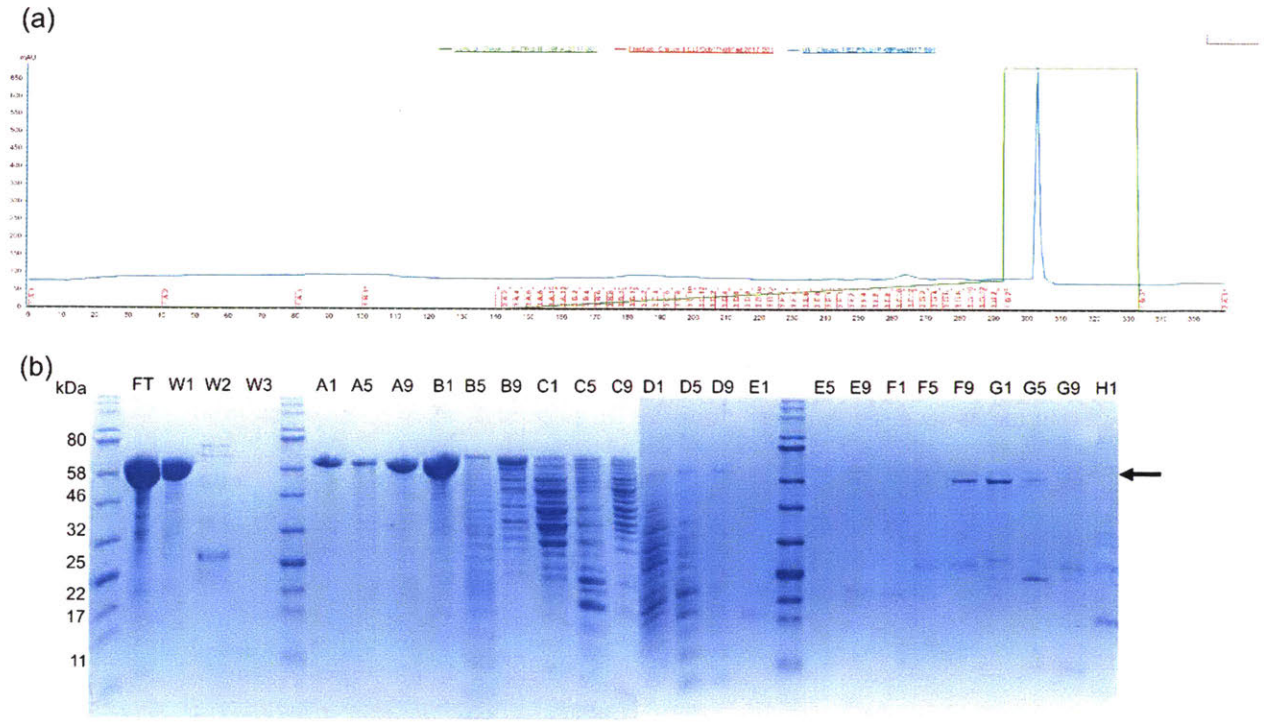


Figure 2-10. (a)A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 for bhELP. Fractions corresponding to the gel below are indicated by labelled red lines along the x-axis on this plot. A zoomed trace where the elution fractions (140-300 mL) are visible is provided in **Figure A-10**). (b) Corresponding SDS-PAGE gels showing fractions produced in FPLC fractionation for bhELP. Black arrow indicates expected band for this protein.

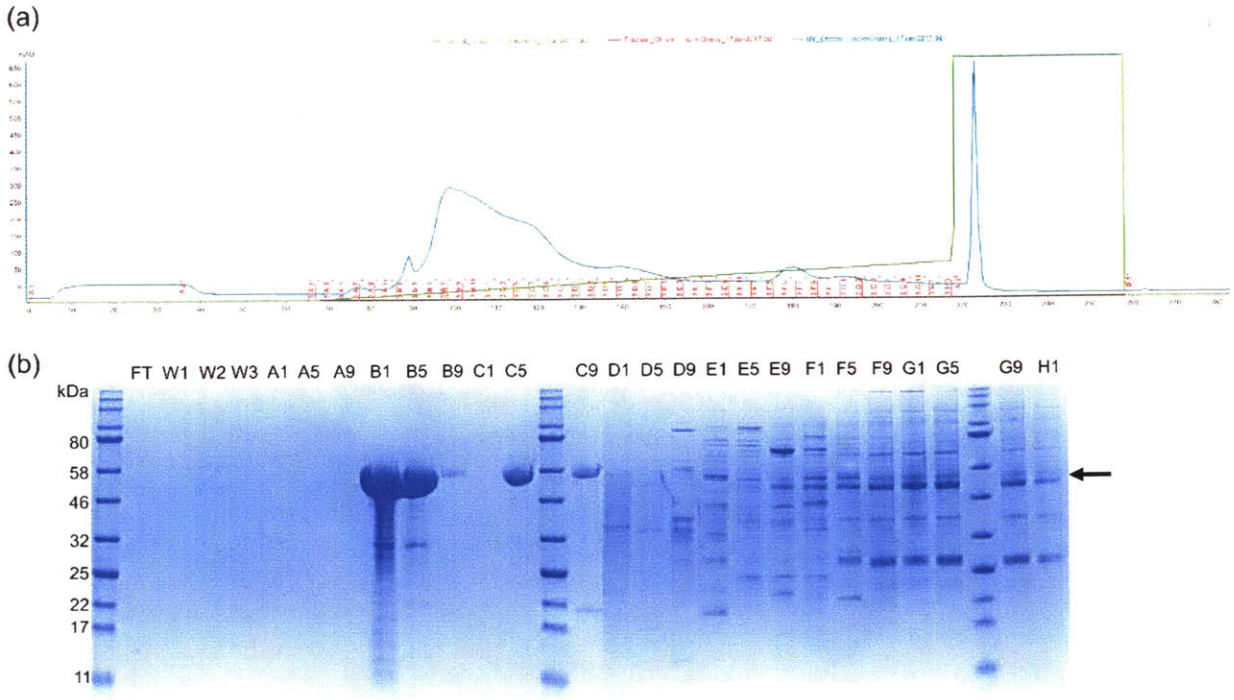


Figure 2-11. (a)A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 10 for nELP. Fractions corresponding to the gel below are indicated by labelled red lines along the x-axis on this plot. A zoomed trace where the elution fractions (190-350 mL) are visible is provided in **Figure A-11**. (b) Corresponding SDS-PAGE gels showing fractions produced in FPLC fractionation for nELP. Black arrow indicates expected band for this protein.

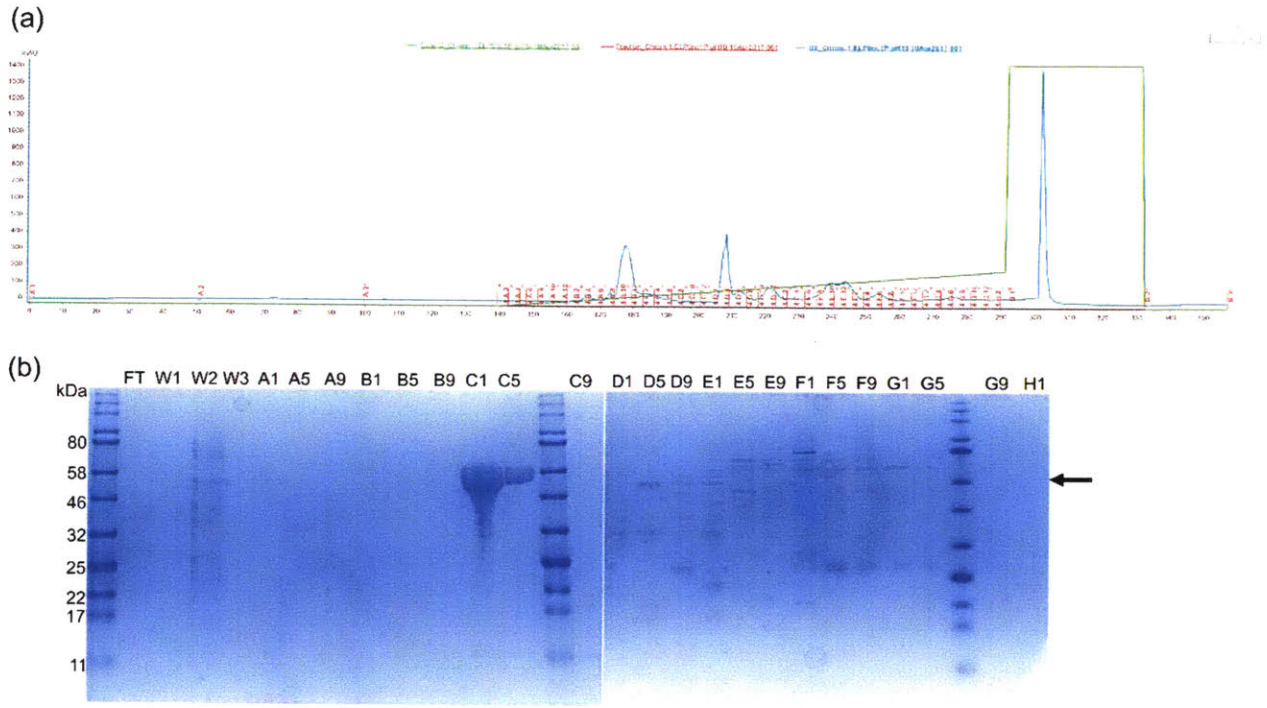


Figure 2-12. (a)A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 10 for nhELP. Fractions corresponding to the gel below are indicated by labelled red lines along the x-axis on this plot. A zoomed trace where the elution fractions (140-300 mL) are visible is provided in **Figure A-12**. (b) Corresponding SDS-PAGE gels showing fractions produced in FPLC fractionation for nhELP. Black arrow indicates expected band for this protein.

SDS-PAGE showing final purity of ELPs used in this study is shown in **Figure 2-13**.

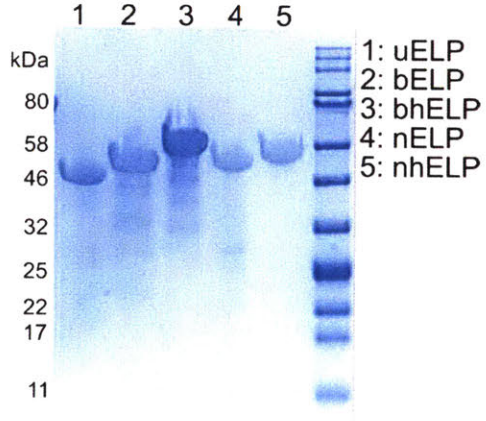


Figure 2-13. Coomassie blue stained SDS-PAGE gel on different ELPs produced and used in this study.

ELP-mCherry Expression and Purification

All plasmids containing the different ELP-mCherry genes of interest were transformed into *E. coli* strain Tuner(DE3). All mCherry-ELP fusion proteins were expressed in LB (1 L) supplemented with 50 μg/mL kanamycin in baffled shake flasks (2.5 L). Expressions were inoculated with 10 mL starter cultures grown in LB supplemented with 50 μg/mL kanamycin at 37 °C overnight. Expressions were grown to OD_{600} = 0.7-1.0 at 30 °C and then induced with 0.5 mM IPTG followed by a temperature drop to 20 °C. Expressions were grown until the pink color characteristic of mCherry was observed (17-24 h). Expression at 30 °C with no induction was also attempted for all fusion proteins, as these conditions were reported to produce ELP-mCherry fusions previously.[4-5] However, when attempting to replicate these expressions, even with previously expressed fusions, the pink color characteristic of mCherry was not consistently observed across all expression flasks. The procedure using induction with a temperature drop, in contrast, produced the characteristic pink color every time. Cells from all expressions were harvested by centrifugation (4,816xg for 15 min at 4 °C). Cells were resuspended in 100 mL of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 8) for every 30 g of wet cell mass. Resuspended cells were then frozen and stored at -80 °C. A table of expression conditions and yields is provided in **Table 2-2**.

Resuspended cells for all fusion proteins were thawed and incubated with 1 mg/mL lysozyme at 4 °C for 1-2 h. After incubation with lysozyme, resuspended cells were sonicated three times, 10 min each at power level 5, 50% duty cycle, with cooling to ≤ 15 °C between sonications. Between sonications, samples were cooled on ice. Cooling to ≤ 15 °C after a sonication was monitored using an infrared thermometer, and typically took approximately 10 minutes. Lysates were clarified by high-speed centrifugation (26,700xg for 1 h at 4 °C). All lysates were purified with two rounds of ammonium sulfate induced precipitation. Precipitations were performed by adjusting the ammonium sulfate concentration in the solution to 1.5 M by addition of 3 M ammonium sulfate in pH 8 sodium phosphate buffer. Precipitant was collected by high-speed centrifugation (26,700xg for 1 h at 4 °C). The supernatant from this spin was discarded and the pellet was resuspended in 50 mM pH 8 sodium phosphate buffer overnight at 4 °C on a rocker. The resuspended solution was then clarified by a second high-speed centrifugation (26,700xg for 1 h at 4 °C). In this case,

the supernatant (containing the product) was retained and the pellet discarded. SDS-PAGE gels showing the various fractions produced in this process for uELP-mCherry is shown in **Figure 2-14**. **Figure 2-15** shows fractions produced in purification of the other ELP-mCherry fusions; note that not all fractions are present for each purification step in these gels; however, they are included for completeness.

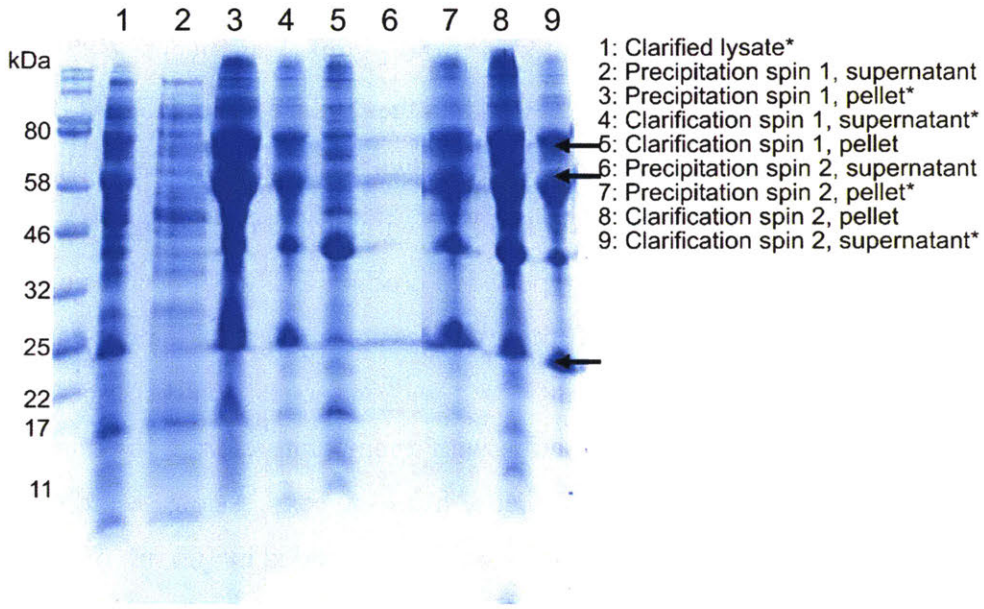


Figure 2-14. Coomassie blue stained SDS-PAGE gel showing the different fractions produced when purifying uELP-mCherry clarified lysate with two rounds of ammonium sulfate precipitation. Fractions that are expected to contain product are marked with an asterisk. Bands corresponding to the desired product are indicated by a black arrow. Note that pure proteins containing mCherry are expected to produce three distinct bands on a gel—one for the full length product and one for the two fragments produced by hydrolytic cleavage of the acylimine bond in the mCherry chromaphore. [11-13]

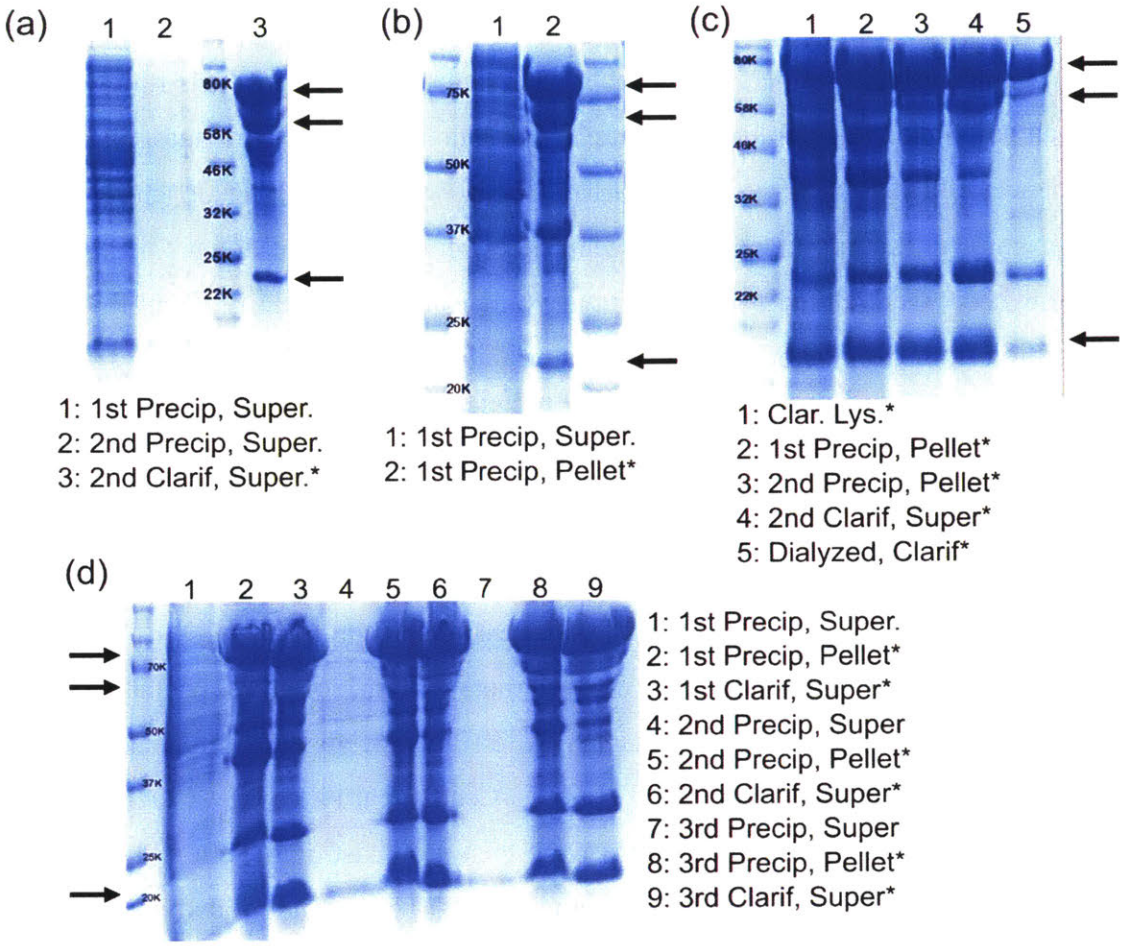


Figure 2-15. Coomassie blue stained SDS-PAGE gel showing fractions produced when purifying (a) bELP-mCherry (b) bhELP-mCherry (c) nhELP-mCherry and (d) nELP-mCherry with ammonium sulfate precipitation. Fractions that are expected to contain the desired product are marked with an asterisk. Bands corresponding to the desired product are indicated with black arrows (3 per mCherry-containing fusion, as described in caption for **Figure 2-14**).

After two rounds of ammonium sulfate induced precipitation, the final protein product was dialyzed into Milli-Q purified water (6,000-8,000 MWCO, 5-7 changes of MilliQ water separated by at least 3 h). Over the course of this first dialysis, some insoluble aggregate that did not contain the desired ELP-mCherry product typically formed in the dialysis bag. Dialysis was set up such that $\leq$ 300 mL of protein solution was being dialyzed against 4 L of water for each change. Thus, after removal from dialysis, the product was subject to centrifugation at 4 °C (1 h, 26,700xg) to remove insoluble aggregates prior to any further purification.

After dialysis and centrifugation, 2 M pH 8 tris was added to adjust final tris concentration of the solutions to 20 mM. FPLC was performed under native conditions (20 mM tris running buffer at pH 8) with an anion exchange column (HiTrap Q HP anion exchange chromatography column, GE Life Sciences). bELP-mCherry and bhELP-mCherry were eluted over 30 column volumes using a linear gradient that increased salt concentration from 0 to 200 mM NaCl. nELP-mCherry and nhELP-mCherry were eluted over 30 column volumes using a linear gradient that increased salt concentration from 0 to 300 mM NaCl. Fusion protein purity of ELP in different fractions was evaluated using SDS-PAGE that sampled every 2 wells in regions of the eluent fractions exhibiting the pink color characteristic of mCherry. Sample FPLC traces and corresponding SDS-PAGE gels for all fusion proteins expressed are shown in **Figure 2-16, Figure 2-17, Figure 2-18, Figure 2-19,** and **Figure 2-20.** Fractions containing pure fusion protein were collected, combined and dialyzed again into Milli-Q purified water (6,000-8,000 MWCO, 10 changes of MilliQ water separated by at least 3 hrs). Dialysis was set up such that $\leq$ 300 mL of protein solution was being dialyzed against 4 L of water for each change. After dialysis, fusion protein solutions were concentrated using ultrafiltration (Amicon ® Ultra-15 Centrifugal Filter, MWCO 30 kDa) to 10-20 wt%. Concentrated fusion protein solutions were then drop cast onto Teflon sheets. Samples were then dried at room temperature under vacuum ramp (50 Torr per hour down to 5 Torr) for 12 h. These conditions were selected based on methods published previously for producing solid bioconjugate samples. [14] SDS-PAGE and native PAGE on all proteins used are shown in **Figure 2-21**. After dehydration, storage and subsequent rehydration, the ratio of the molar concentration predicted by absorption at 587 nm (ε = 72,000 $M^{-1}cm^{-1}$) to the molar concentration predicted by absorption at 280 nm (ε = 32,430 M^{-1} cm^{-1}) was calculated to get an idea of the fraction of the protein that was functional. After several months storage, these ratios were found to be 0.14, 0.34, 0.25, 0.15, and 0.45 for uELP-mCherry, bELP-mCherry, bhELP-mCherry, nELP-mCherry and nhELP-mCherry, respectively.

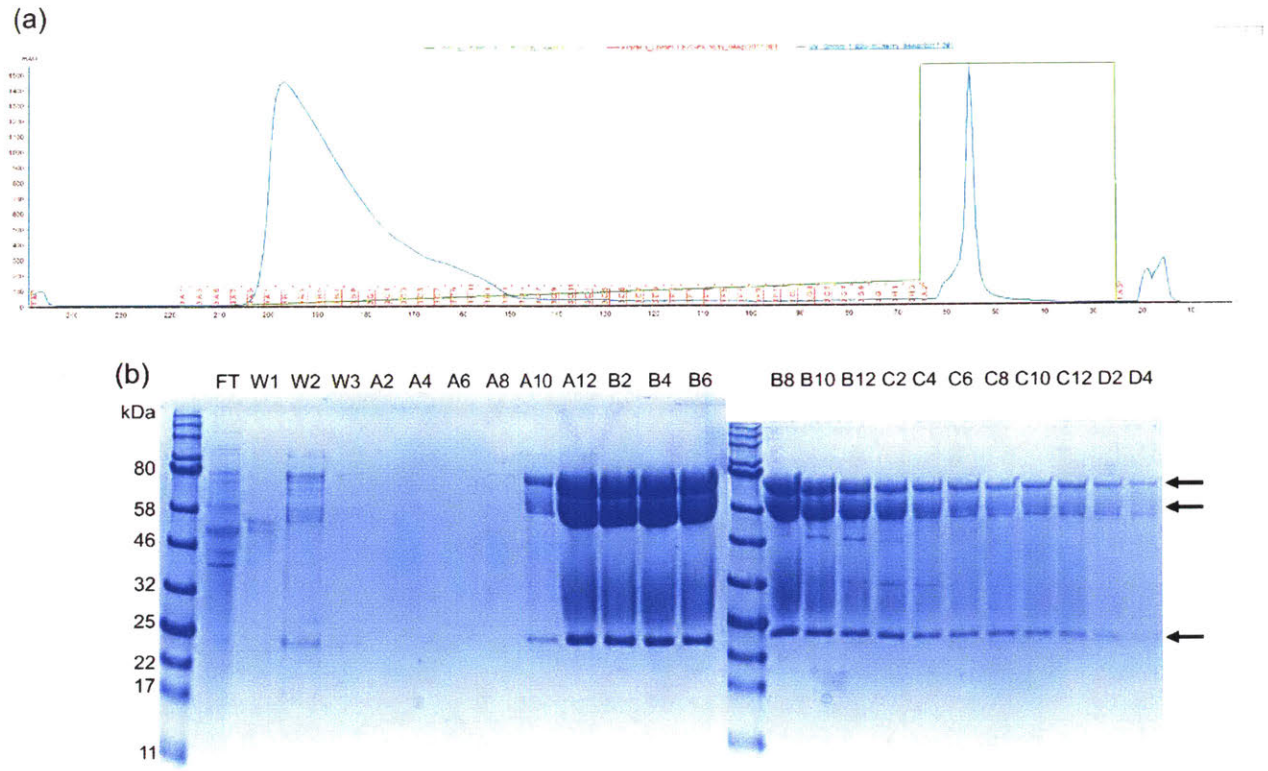


Figure 2-16. (a)A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 for uELP-mCherry. Fractions corresponding to the gel below are indicated by labelled red lines along the x-axis on this plot. A zoomed trace where the elution fractions (-220- -70 mL) are visible is provided in **Figure A-13.** (b) Corresponding SDS-PAGE gels showing fractions produced in FPLC fractionation for uELP-mCherry. Black arrow indicates expected band for this protein.

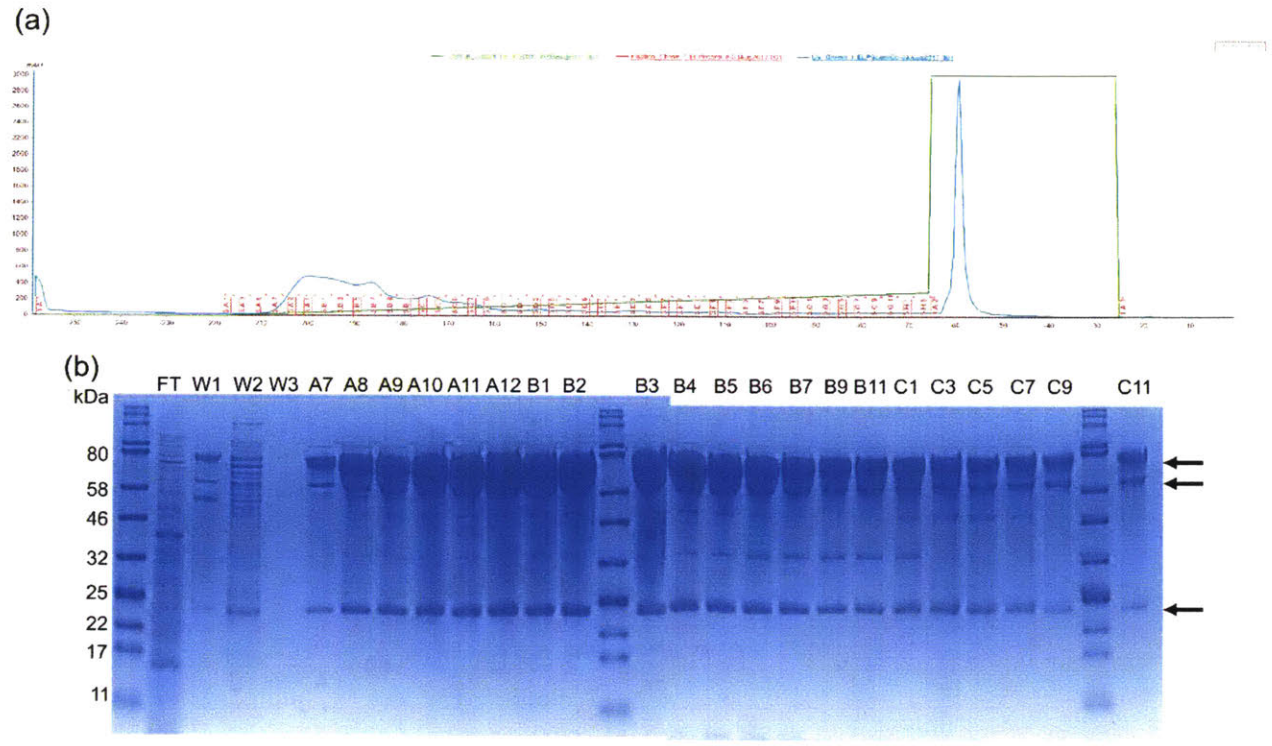


Figure 2-17. (a)A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 for bELP-mCherry. Fractions corresponding to the gel below are indicated by labelled red lines along the x-axis on this plot. A zoomed trace where the elution fractions (-220- -70 mL) are visible is provided in **Figure A-14**. (b) Corresponding SDS-PAGE gels showing fractions produced in FPLC fractionation for bELP-mCherry. Black arrow indicates expected band for this protein.

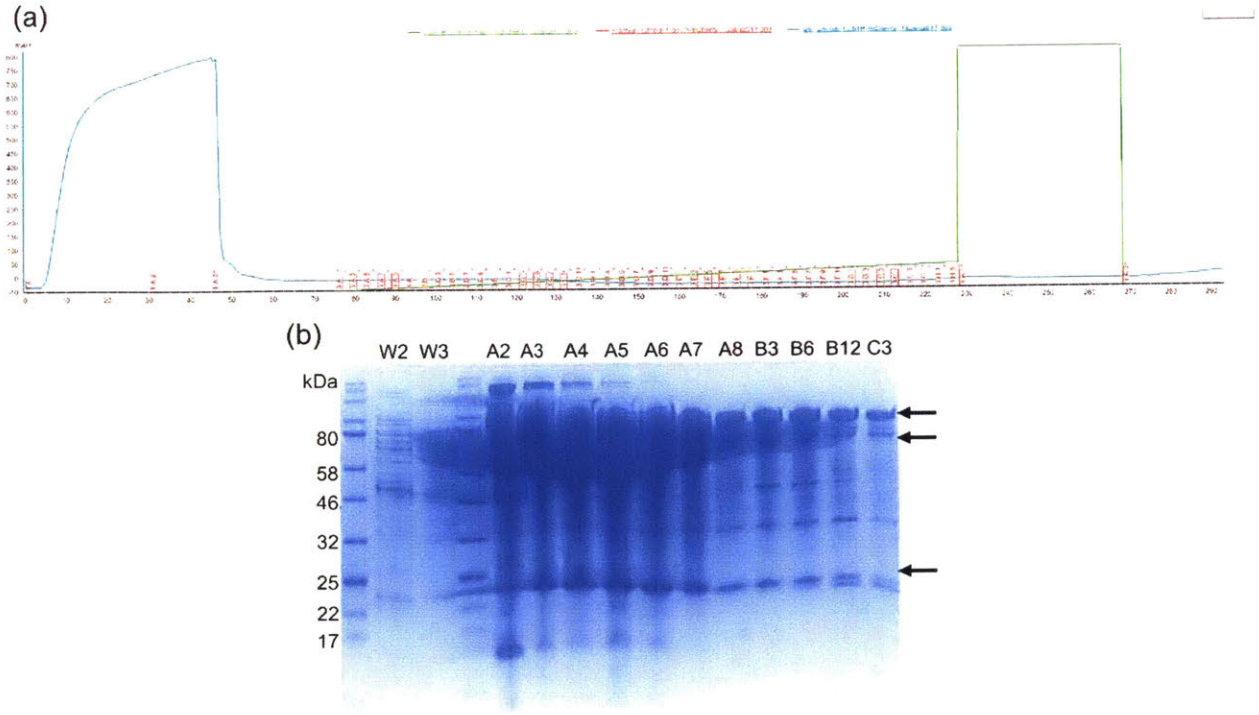


Figure 2-18. (a)A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 for bhELP-mCherry. Fractions corresponding to the gel below are indicated by labelled red lines along the x-axis on this plot. A zoomed trace where the elution fractions (70-230 mL) are visible is provided in **Figure A-15**. (b) Corresponding SDS-PAGE gels showing fractions produced in FPLC fractionation for bhELP-mCherry. Black arrow indicates expected band for this protein.

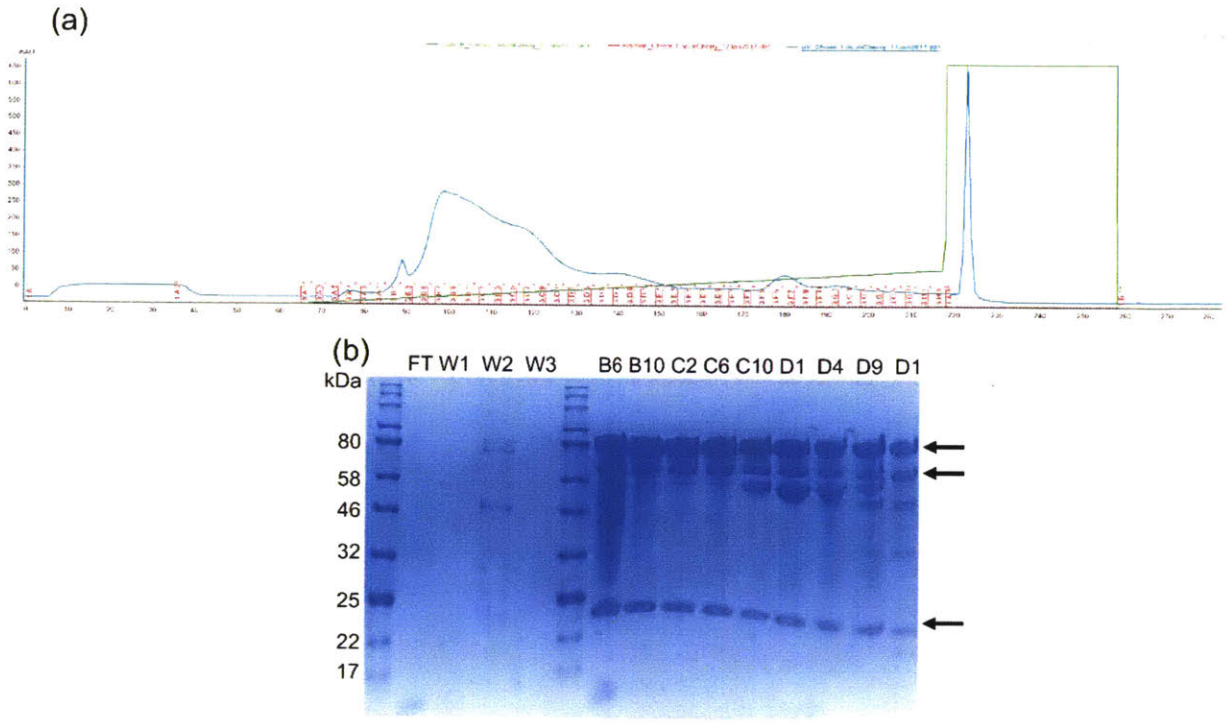


Figure 2-19. (a)A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 for nELP-mCherry. Fractions corresponding to the gel below are indicated by labelled red lines along the x-axis on this plot. A zoomed trace where the elution fractions (70-220 mL) are visible is provided in **Figure A-16**. (b) Corresponding SDS-PAGE gels showing fractions produced in FPLC fractionation for nELP-mCherry. Black arrow indicates expected band for this protein.

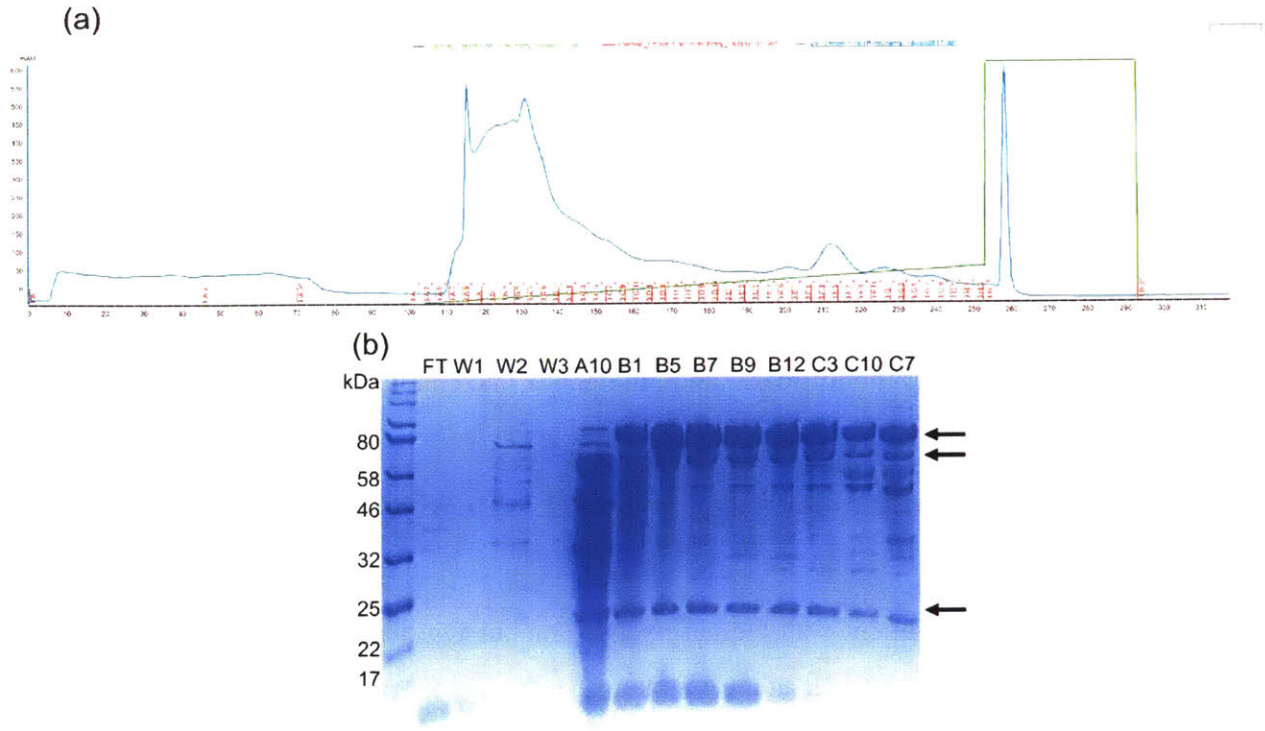


Figure 2-20. (a)A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 for nhELP-mCherry. Fractions corresponding to the gel below are indicated by labelled red lines along the x-axis on this plot. A zoomed trace where the elution fractions (100-260 mL) are visible is provided in **Figure A-17**. (b) Corresponding SDS-PAGE gels showing fractions produced in FPLC fractionation for nhELP-mCherry. Black arrow indicates expected band for this protein.

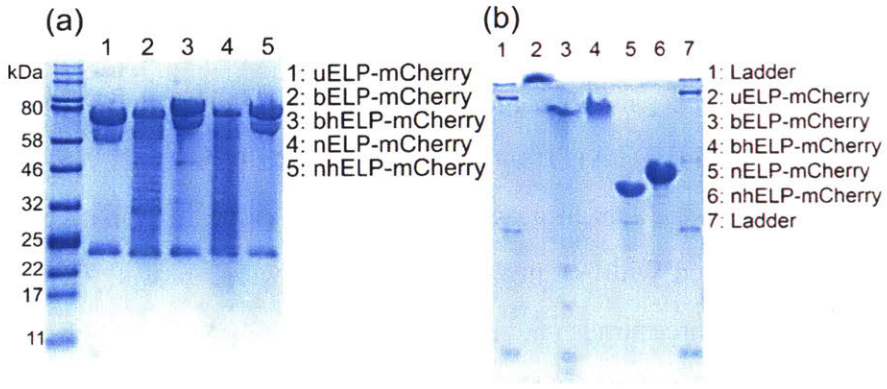


Figure 2-21. Coomassie blue stained (a) SDS-PAGE and (b) Native PAGE on final ELP-mCherry fusion proteins used in this study. In SDS-PAGE gels, three bands are observed for a

pure final product due to partial hydrolysis of the acylimine bond in the mCherry chromaphore.[11-13]

Of the other fusion proteins cloned, uhELP-mCherry, bhhELP-mCherry, and nhhELP-mCherry all appeared to express (as evidenced by the pink color characteristic of mCherry observed in the cell pellets). After lysis of the cells, however, the pink product localized to the lysate pellet upon clarification and could not be solubilized without the use of urea. Several solubilization agents were tried, including various amounts of ethanol, methanol, and chaotropic salts (potassium iodide, sodium iodide).

2.1.2.3 Cononsolvency of elastin-like polypeptides (ELPs) in water/alcohol solutions

ELP Expression and Purification

The ELP used in this experiment was cloned by Dr. Guokui Qin, and was used as received. A pET28b plasmid containing the ELP gene of interest (plasmid map, sequence, and amino acid sequence in **A2.2 Cononsolvency of Elastin-Like Polypeptides (ELPs) in Water/Alcohol Solutions**) was transformed into chemically competent Tuner(DE3) *E. coli* strain. The ELP was expressed in 5 L of TB supplemented with 50 μg/mL kanamycin in a fermenter with a 7 L working volume. The expression was inoculated with 50 mL of starter culture grown in LB supplemented with 50 μg/mL kanamycin, and was subsequently run at 30 °C for 18 h with no induction. Expression conditions were selected as described in the previous section (see **Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins, ELP Expression and Purification**), and no optimizations were made to the expression. After fermentation was complete, cells were harvested by centrifugation (4,816xg for 15 min at 4 °C). Cell pellets were resuspended in 100 mL of MENT buffer (3 mM $MgCl_2$, 1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 10 mM trizma, pH 7.5) per 30 g of wet cell mass. Resuspended cells were frozen at -80 °C overnight to aid in lysis. Resuspended cells were then thawed and incubated with ~1 mg/mL lysozyme for 1-2 h at 4 °C. Cells were then sonicated with a tip sonicator three times (10 min of sonication at power level 5, 50% duty cycle with cooling to $\leq$ 15 °C between sonications). Lysate was then clarified by high-speed centrifugation (26,700xg for 1.5 h at 4 °C). Subsequent purification steps were identical to those used for uELP in the previous section—two rounds of ITC followed by FPLC (see **Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins, ELP**

Expression and Purification). SDS-PAGE gels on the different ITC fractions, FPLC traces, and gels on FPLC product are nearly identical to those for uELP used in the previous study (see **Figure 2-7**). The final yield of this ELP was 230 mg protein/L expression.

2.1.2.4 Protein Purification by Ethanol-Induced Phase Transitions of the Elastin-Like Polypeptide (ELP)

Fermenter-Scale Expression of ELP-sfGFP

The ELP-sfGFP in pET28a gene used in this work was cloned by Dr. Guokui Qin. The DNA sequence, amino acid sequence, and plasmid map of this protein are shown in **A2.3 Protein Purification by Ethanol-Induced Phase Transitions of the Elastin-Like Polypeptide (ELP)**. ELP-sfGFP was expressed in the BL21(DE3) strain of *E. coli*. Expressions were carried out in 5 L of LB supplemented with 50 μg/mL kanamycin in a fermenter with a 7 L working volume. Cells were grown at 37 °C to OD_{600} = 0.7-0.9 and were then induced with 0.5 mM IPTG. After induction, temperature of the fermentation was dropped to 20 °C, and the expression was carried out for 16-20 h after induction. This expression conditions were chosen on the basis of the results of the previous section, noting that a temperature drop after induction appeared to yield the most consistent protein expression for ELP-mCherry fusions. In this case, it was assumed that growth at 37 °C instead of 30 °C would not impact the final expression yield and would allow the cells to reach their induction OD_{600} more quickly. No further optimizations to protein expression were performed beyond this. Cells were harvested by centrifugation (4,816xg for 15 min at 4 °C) and then resuspended in 100 mL of MENT buffer (3 mM $MgCl_2$, 1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 10 mM trizma, pH 7.5) for every 30 g of wet cell mass. Resuspended cells were then frozen at -80 °C overnight. Cells were then thawed and incubated with 1 mg/mL lysozyme at 4 °C for 1-2 h. Resuspended cells were then sonicated with a tip sonicator three times (5 min of sonication at power level 5, 50% duty cycle with cooling to ≤ 15 °C between sonications). After sonication, lysates were typically mixed at 4 °C for 30-60 min to ensure complete solubilization of the ELP-sfGFP after heating caused by sonication. Lysates were clarified by centrifugation at 26,700xg for 1 h at 4 °C.

Well-Plate Expression of ELP-sfGFP

ELP-sfGFP expressed in a well-plate format was also expressed in the BL21(DE3) strain of *E. coli*. Starter cultures (10 mL LB supplemented with 50 μg/mL kanamycin) were inoculated with

an isolated colony selected from an LB-agar plate supplemented with 50 μg/mL kanamycin produced from a transformation of ELP-sfGFP DNA into BL21(DE3) chemically competent cells. Starter culture was grown at 37 °C overnight in an orbital shaker. Expressions were carried out in cone-bottomed deep 96-well plates that were prepared with a single 3-5 mm diameter glass bead in each well, covered with foil, and subsequently autoclaved for sterilization (see **Figure 2-22** for technique used to add single glass bead to each well). Beads were used to improve the mixing in wells. This choice was made on the observation that adding beads to wells increased the average ELP-sfGFP yield in well-plate expressions (**Figure 2-23**, bar 1 versus bar 2). Each well was prepared with 1480 μL of LB supplemented with 50 μg/mL kanamycin. Each well was inoculated with 20 μL of the starter culture. Plates were then incubated at 37 °C with orbital shaking at 300 rpm. It should be noted that appreciable ELP-sfGFP were not observed in well-plate expressions until shaking speeds were increased from 250 rpm to 300 rpm; thus, it is the recommendation of the author that shaking speeds of 300 rpm are always used for well-plate expressions. Deep well plates were contained in square, 5.7 cup Tupperware containers (Sterilite 03324706 Ultra-Seal™ Containers, https://www.amazon.com/gp/product/B01B2409ZS/ref=ppx_yo_dt_b_asin_title_o03_s00?ie=UTF8&psc=1) during orbital shaking. The purpose of this containment was two-fold. First, these containers allowed well plates to fit into incubator slots designed for 2.5 L fernbach flasks. Second, sponges saturated with water could be placed in the containers with the well plate to pad the top and bottom of the well plate on the 8-well axis of the plate to hold the plate in place, as well as to provide a humid local environment for the deep well-plate to prevent volume losses from evaporation over the course of the expression. An image of this setup is shown in **Figure 2-22**. The blue tab on top of the Tupperware was left open during expression to permit oxygen exchange with the incubator/outside environment.

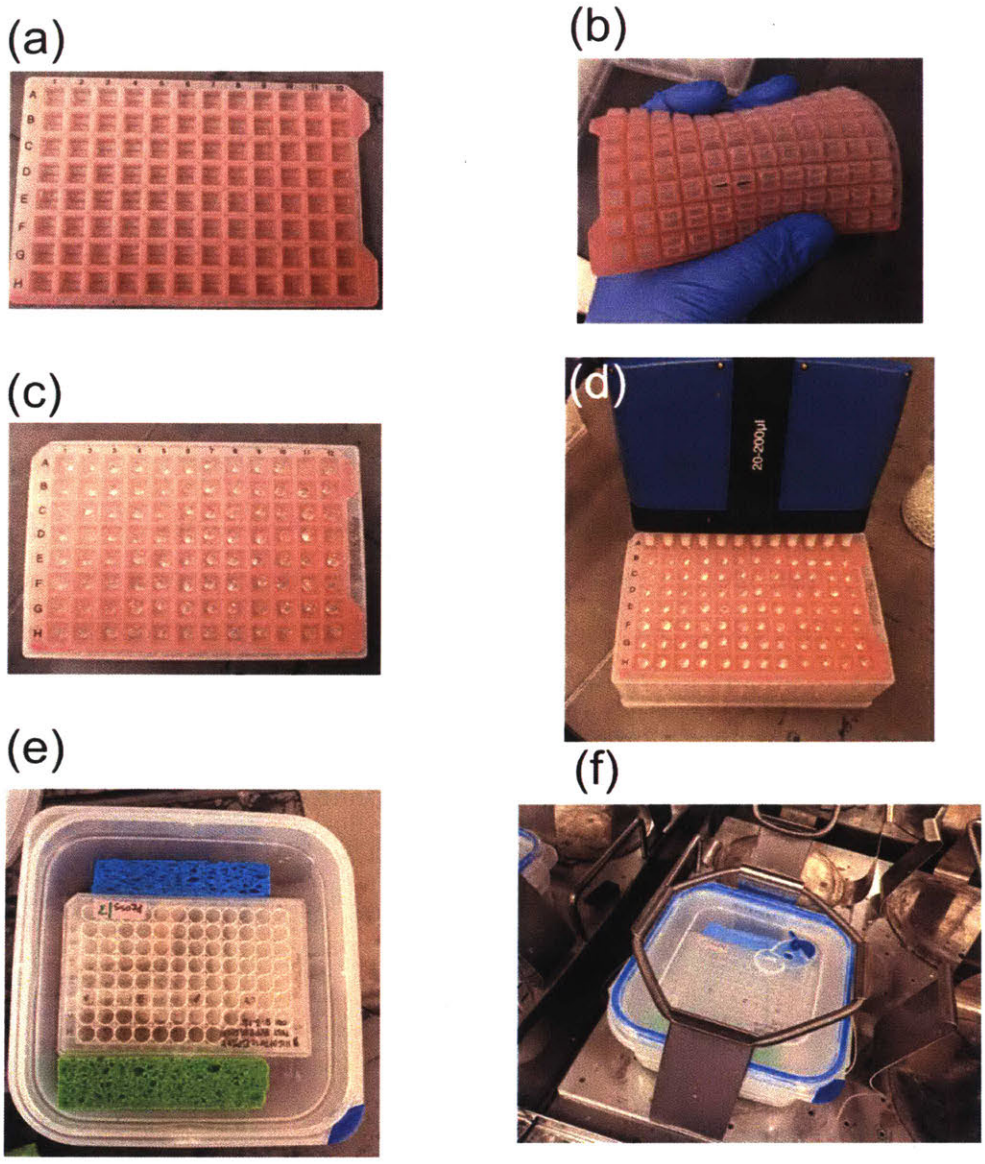


Figure 2-22. (a) Silicone mat used to put a single bead into each well of the plate (https://www.i-labpro.com/96-Square-Plug-Silicone-Sealing-Mat.html). At the bottom of the divet of each well in the silicone mat, a box cutter was used to cut into the mat to allow glass beads to pass through. This can be seen upon flexing of the mat as shown in (b). To load a plate with glass beads before autoclaving, the mat with cuts was loaded onto the plate and firmly pushed into place. Unsterilized (but clean) beads could then be poured on top of the mat and easily rolled around on top of the plate until there was one bead in each divet. This was found to be significantly less effort than dropping a single bead into each well of the plate by hand, as the beads were much easier to maneuver in this format. A mat with beads in every well is shown in (c). Finally, to get beads into a well, a P200 multichannel pipette was used to gently press down on the top of the

beads until they passed through the cuts in the membrane, taking care to not let the white cylinders of the multichannel pipette catch on the mat (and unseat the entire setup). An example of this is shown in (d). The entire process of loading beads into a single well plate with this procedure took about 5-7 minutes per plate. The entire plate could then be covered in foil and autoclaved. After autoclaving, a sterile lid, or a lid that had be resterilized with 70% ethanol was put on top of the plate to replace the foil. After loading the plate with media and inoculating the expression, the plate was placed in a square Tupperware and flanked with sponges as shown in (e). Sponges were hydrated with DI water until they were saturated, and small amount of water was added to the bottom of the Tupperware (sufficient to make small pools in the bottom of the Tupperware, but not enough to create an entire water layer at the bottom of the Tupperware). This water was added so that sponges could wick up moisture in the event that they started to dry over the course of the expression. These sponges humidified the local environment of the well-plate, minimizing evaporative losses in the expressions, and held the well-plate firmly in place. Use of containers of this shape also allowed the well plates to be used with incubator inserts for 2.5 L fernback flasks, as shown in (f).

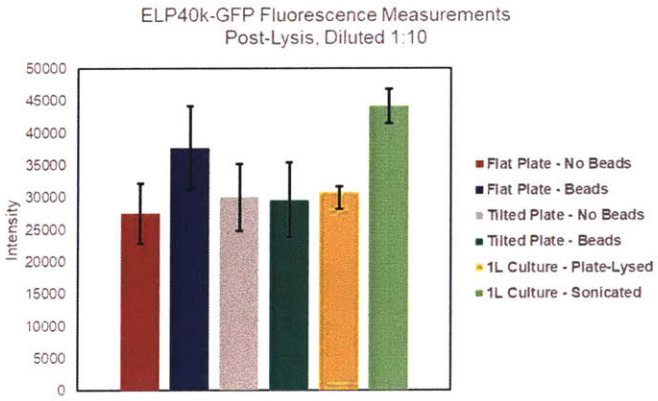


Figure 2-23. ELP-sfGFP fluorescence of clarified lysates, measured after trying different growth conditions, and performing one freeze-thaw cycle to lyse cells growth both in a well plate (first 4 bars) and in a 1 L flask (last 2 bars). The effectiveness of lysing in a well-plate and using a tip sonicator is also directly compared (last two bars).

After inoculation, OD_{600} was monitored by removing 150 µL of the expression from 8 wells of the expression plate into a sterile 96-well microtiter plate, measuring absorbance at 600 nm in a Tecan

Infinite M200 Pro plate reader, and then returning the 150 µL to the deep-well plate. Once the average OD_{600} of the sampled wells reached ~0.8, plates were put on ice or into the 4 °C fridge for 20-30 m. The decision to chill plates prior to induction was based on literature reports from the Arnold lab. In their protocols, when an expression is grown at 37 °C and subsequently dropped to 20 °C after induction, a chilling step is included prior to induction. [15] Expressions were then induced by adding 20 µL of sterile filtered 38 mM IPTG to each well to bring the final IPTG concentration to 0.5 mM. The concentration of the IPTG stock solution here was lower than that in the fermentation procedures to enable accurate pipetting (20 µL vs 0.15 µL). After induction, plates were incubated at 24 °C for 18-23 h with orbital shaking at 300 rpm. Cells were collected by centrifugation of the deep well plates (4,000xg, 15 m). Media was removed from cell pellets via aspiration with a multichannel pipettor. Cell pellets were then frozen at -20 °C overnight, and then thawed. Cells were resuspended in 140 µL of MENT buffer per well, and were lysed by incubation with 1 mg/mL lysozyme, ~0.1 mg/mL DNase I, and ~0.1 mg/mL RNase A at 37 °C for 3 h (with orbital shaking at 300 rpm), followed by a second freeze-thaw cycle at -20 °C (where freezing was performed overnight). This second freeze-thaw cycle was found to increase cell lysis slightly as compared to performing one freeze-thaw cycle (**Figure 2-24**). Cell lysates were then thawed and transferred to 96-well microtiter plates. Lysates were clarified by centrifugation at 6,000xg for 1 h at 4 °C.

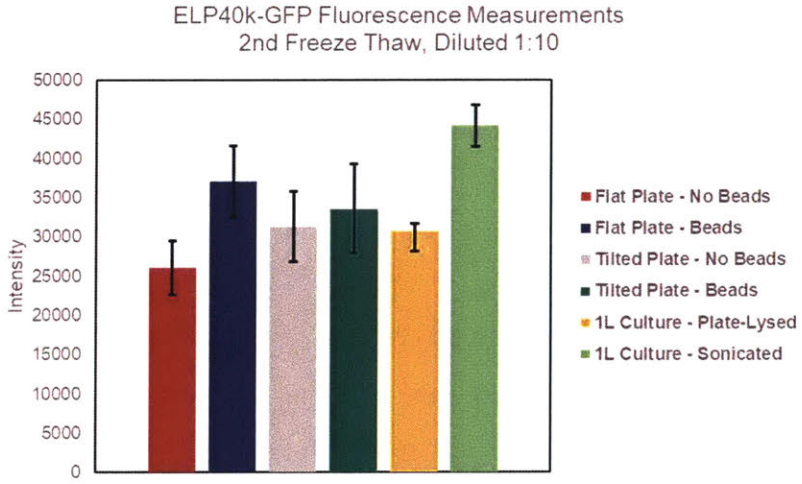


Figure 2-24. ELP-sfGFP fluorescence measured on clarified lysates produced after two freeze-thaw cycles on well-plate and liter-scale expressed ELP-sfGFP. Lysis efficiency is compared to that of a tip-sonicated liter-scale expression (last bar).

Purification of ELP-sfGFP

Clarified ELP-sfGFP lysates were purified by either two consecutive NaCl-induced precipitation cycles or one NaCl- then one ethanol-induced precipitation cycle. NaCl precipitations were carried out by adding 5 M NaCl in water to the protein solution of interest to bring the final NaCl concentration to 1.5 M. Samples were then incubated at 37 °C overnight to ensure complete protein precipitation. For fermenter-scale purification schemes, ~0.1 mg/mL DNase I and RNase A were added to the clarified lysate in the first NaCl-induced precipitation prior to incubation at 37 °C to encourage degradation of DNA and RNA in the lysate. After the overnight incubation, precipitant was collected by centrifugation at 37 °C (1 h, 26,700xg for fermenter-scale expressions, 6,000 xg for well-plate expressions), and supernatant was discarded. Pellets containing the ELP-sfGFP of interest were resuspended in MilliQ water at 4 °C for at least 3 h to ensure complete redissolution of the desired protein product. The resuspended protein was then clarified by centrifugation at 4 °C (1 h, 26,700xg for fermenter-scale expressions, 6,000 xg for well-plate expressions). In this case, the supernatant (containing the protein product of interest) was collected. This supernatant was the final product of a single NaCl-induced purification cycle.

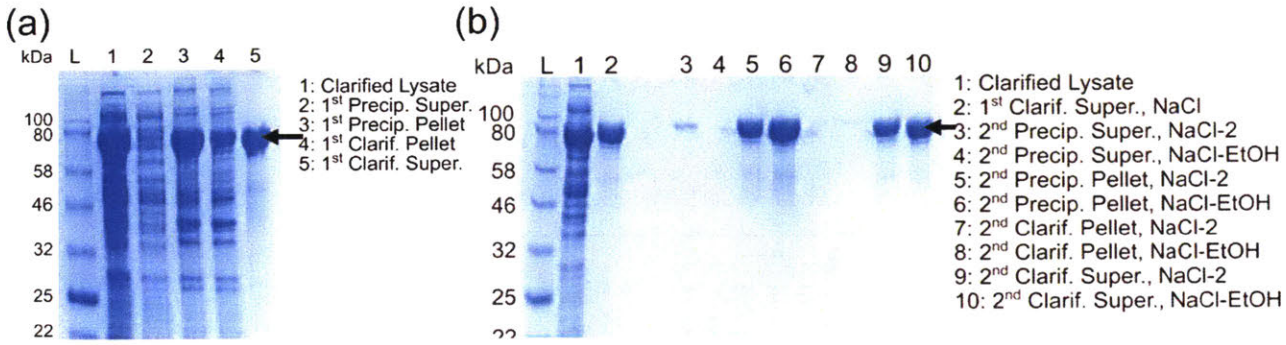


Figure 2-25. Coomassie blue stained SDS-PAGE gels on ELP-sfGFP (expected band at 73.3 kDa) produced in a 5 L fermentation for (a) the fractions produced after one cycle of NaCl-induced precipitation on an ELP-sfGFP clarified lysate and (b) the fractions produced after a second cycle of either NaCl- or EtOH-induced precipitation on the final product of the first NaCl-induced precipitation. Black arrows indicate the expected band in the gel.

Precipitations with ethanol (EtOH) were carried out similarly. Pure EtOH was added to the protein solution of interest to a final EtOH concentration of 30 vol% EtOH. Samples were then incubated at 4 °C for 16 h to allow complete precipitation. Precipitate was collected by centrifugation at 4 °C (1 h, 26,000xg for fermenter-scale expressions, 6,000xg for well-plate expressions). Supernatant produced in this centrifugation was discarded, and the pellet was resuspended in MilliQ water at 4 °C for at least 3 h to ensure complete redissolution of the ELP-sfGFP. The suspension was clarified by centrifugation at 4 °C (1 h, 26,700xg for fermenter-scale expressions, 6,000xg for well-plate expressions). The supernatant from this spin was collected as the final product of a single EtOH-induced precipitation cycle. SDS-PAGE gels of the different fractions produced by performing both these purification schemes on fermenter-scale expressions of ELP-sfGFP are shown in **Figure 2-25**. SDS-PAGE gels of the different fractions produced by performing these two purification schemes on well-plate scale expressions are shown in **Figure 2-26**.

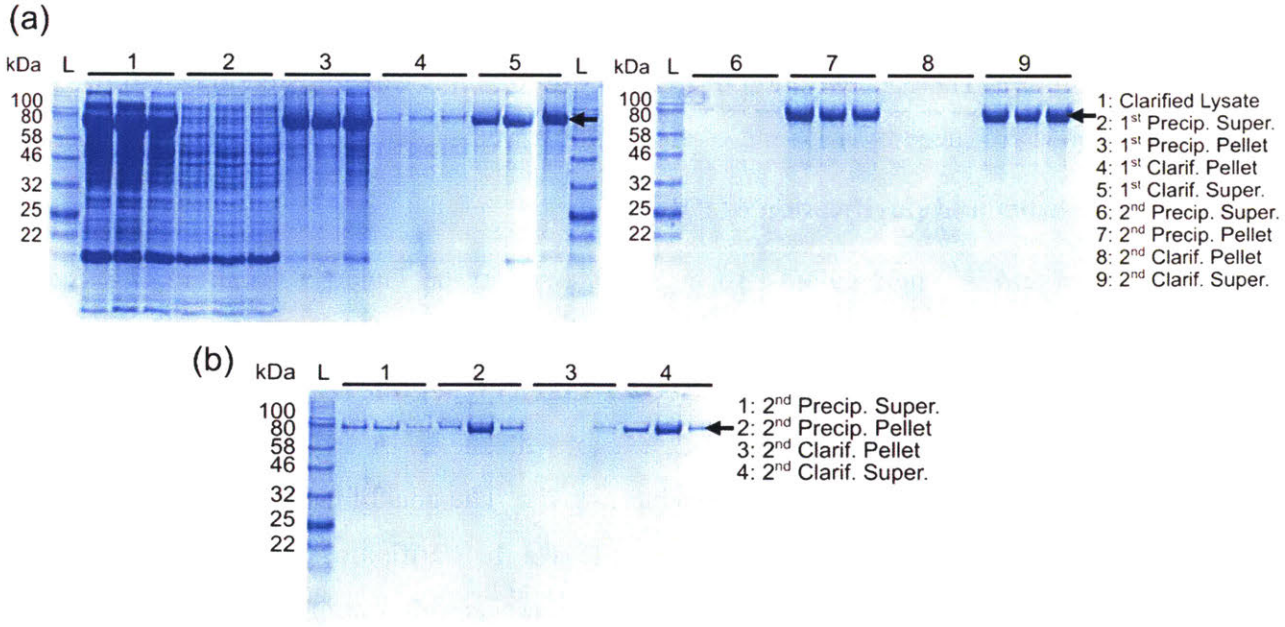


Figure 2-26. Coomassie blue stained SDS-PAGE gels on different fractions (in triplicate) produced by purification of clarified lysate from ELP-sfGFP expressed in a well-plate format for one cycle of NaCl-induced precipitation (a), lanes labeled 1-5. The final product from this first NaCl-induced precipitation (lanes labeled 5) was then subject to a second round of precipitation induced either with EtOH, shown in (a), lanes labeled 6-9 or with NaCl again, shown in (b). In the first NaCl precipitation cycle, and in the EtOH precipitation cycle, little well-to-well variability is observed. In contrast, high well-to-well variability is observed in the second NaCl

precipitation cycle, as seen in (b). In this second NaCl-induced precipitation cycle, the precipitated pellet delaminated from the bottom of the well when aspirating off the supernatant, resulting in high product loss and well-to-well variability. Black arrows indicate expected band for the desired product.

ELP-sfGFP used to construct the phase diagram quantifying protein partitioning into soluble and insoluble phases was purified using two NaCl-induced precipitations as described above. The protein was subsequently dialyzed against MilliQ water (6,000-8,000 MWCO, 7-10 changes of MilliQ water separated by at least 3 h). Over the course of this first dialysis, some insoluble aggregate that did not contain the desired ELP-sfGFP product typically formed in the dialysis bag. Dialysis was set up such that ≤ 300 mL of protein solution was being dialyzed against 4 L of water for each change. Thus, after removal from dialysis, the product was subject to centrifugation at 4 °C (1 h, 26,700xg) to remove insoluble aggregates prior to any further purification. sfGFP fluorescence (excitation 395 nm, emission 510 nm) was used to measure relative concentration of ELP-sfGFP in solution. The concentration of ELP-sfGFP solutions was adjusted to be 5 times the typical fluorescence produced by a 1.5 mL expression in a well-plate format.

Well-plate expression and purification of P10

An ELP-based protein gel, here called P10, was expressed in the Tuner(DE3) strain of *E. coli*. DNA and amino acid sequences for this protein are shown in **A2.3 Protein Purification by Ethanol-Induced Phase Transitions of the Elastin-Like Polypeptide (ELP).** The plasmid map for this gene is identical to that in **Figure 2-2**, except with the DNA sequence provided in the appendix substituted in between the NcoI/XhoI cloning site. The goal of using this protein was to determine whether ELP-based protein materials could also be purified using the cononsolvency approach. To make a starter culture for the P10 expression, a colony from an agar plate was used to inoculate a 10 mL starter culture of LB supplemented with 50 g/mL kanamycin. The starter culture was grown overnight at 37 °C in an orbital shaker. Expressions were carried out in a sterile deep-well plate as described in the section titled **Well-Plate Expression of ELP-sfGFP**. The only difference here is that TB was used as the expression media instead of LB. This expression protocol was selected based on discussions with Dr. Keun Ah Ryu that these conditions worked well for this particular protein. No expression optimization was performed. Cells were collected and lysed as described in the previous section (**Well-Plate Expression of ELP-sfGFP**). After incubation at

37 °C with lysozyme, DNase I, and RNase A, lysates were allowed to cool at 4 °C overnight (with intermittent mixing) to ensure complete resolubilization of the P10 protein prior to clarification. This step was implemented based on previous protocols for P10 expressed on the fermenter scale. In these experiments, the total lysate is cooled on the rocker at 4 °C overnight after tip sonication. This is done to ensure complete resolubilization of any P10 that may have aggregated in hot spots during sonication.

P10 was purified using either one cycle of NaCl- followed by one cycle of EtOH-induced purification (NaCl-EtOH), or two cycles of NaCl- followed by one cycle of EtOH-induced purification (NaCl-NaCl-EtOH). To ensure a high enough concentration to form gels, groups of three wells of clarified lysate were combined, and 5 M NaCl was added to a final concentration of 1.5 M (final volume 600 µL). Plates were incubated at 37 °C overnight. Pellets were separated by centrifugation (6,000 xg, 1 h, 20 °C) and resuspended in 105 µL at 4 °C for 3 h. Resuspensions were clarified by centrifugation (6,000 xg, 1 h, 4 °C). Groups of 6 wells were combined after this clarification step into a single well in a deep-well plate. For the NaCl-NaCl-EtOH samples, another NaCl-induced purification cycle was performed at the same conditions as above. For the final purification cycle by EtOH-induced phase transition, ethanol was added to 30 vol% (final volume of 900 µL for NaCl-EtOH samples, or 150 µL for NaCl-NaCl-EtOH samples). Plates were incubated for 16 h at 4 °C. Pellets were separated by centrifugation (6,000 xg, 1 h, 4 °C) and resuspended in 105 µL Milli-Q water for at least 3 h at 4 °C. Supernatants were separated by centrifugation (6,000 xg, 1 h, 4 °C) and were transferred to weighed Eppendorf tubes, combining three wells per tube. Tubes were frozen and lyophilized. Lyophilized protein was resuspended in 9 µL water to form solutions capable of reversible temperature-dependent gelation. Gels showing the different fractions produced in this process are shown in **Figure 2-27** and **Figure 2-28a**. Protein product produced by this procedure is shown in **Figure 2-28**b and c.

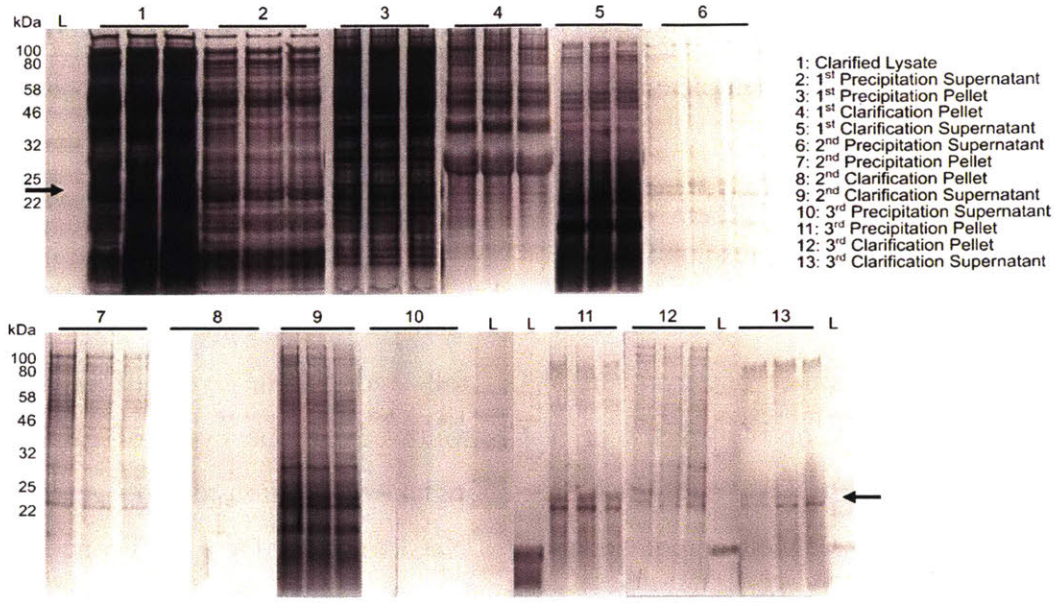


Figure 2-27. Silver-stained SDS-PAGE gels showing purification fractions of P10 purification at the well-plate scale. P10 runs at 25kDa under these conditions. Lane group 5 corresponds to final product after one cycle of purification (NaCl); group 9 corresponds to final product after two cycles (NaCl-NaCl); and group 13 to final product after three cycles (NaCl-NaCl-EtOH). Black arrow indicates the expected band from the desired product.

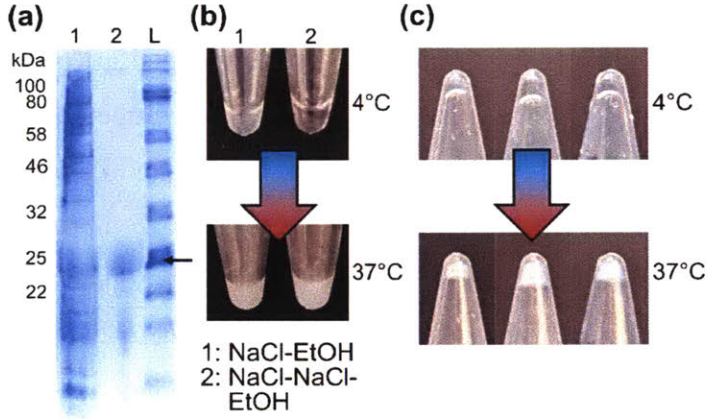


Figure 2-28. (a) Coomassie-stained SDS-PAGE gel showing P10 from a well-plate expression that has been purified by one cycle of NaCl precipitation and one cycle of EtOH precipitation (lane 1) and P10 from a well-plate expression that has been purified by two cycles of NaCl precipitation and one cycle of EtOH precipitation (lane 2) . Black arrow indicates the expected band from the desired product. (b) Images of P10 samples purified by the same methods described in (a) at 4 °C and 37 °C. (c) P10 samples purified from well-plate expressions by the

NaCl-NaCl-EtOH method that have been prepared at 10 wt% in water at 4 °C and 37 °C, demonstrating the ability of these purified proteins to form gels upon heating.

Well-plate expression and purification of ELP-OPH-ELP

Protein was expressed in Tuner(DE3) *E. coli* cells using the pET28b vector. DNA and amino acid sequences for ELP-OPH-ELP are provided in **A2.3 Protein Purification by Ethanol-Induced Phase Transitions of the Elastin-Like Polypeptide (ELP).** Starter cultures for protein production were inoculated from a single colony on an agar plate. A 10 mL starter culture was prepared in lysogeny broth (LB) supplemented with 50 μg/mL kanamycin and grown overnight at 37 ˚C in an orbital shaker. Expressions were conducted in a 96-well plates as described in the section above (**Well-Plate Expression of ELP-sfGFP**) with the following modifications. TB was used as the expression media, and expressions were induced at an average OD_{600} of ~0.5. Cells were collected, lysed and lysate clarified as described in **Well-Plate Expression of ELP-sfGFP**. Protein purification was performed using one NaCl-induced precipitation followed by one EtOH-induced precipitation. NaCl-induced precipitation was performed using 2.5 M NaCl at 20 ˚C for 16 h. Higher salt content and lower temperatures were used in this precipitation to avoid thermal denaturation of the OPH. This condition was based on previous literature on purification of ELP-tagged OPH. [16] EtOH-induced precipitations were performed using 30 vol% EtOH at 4 ˚C for 1 h. Gels on fractions produced by this purification procedure are shown in **Figure 2-29.**

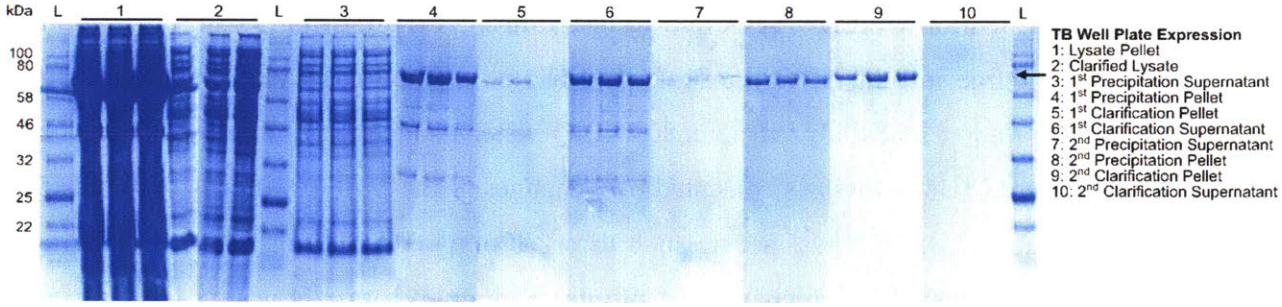


Figure 2-29. Coomassie-stained SDS-PAGE gels showing various fractions from two consecutive cycles of NaCl- then EtOH-induced precipitation (NaCl-EtOH method) using organophosphorus hydrolase (OPH) as the ELP-tagged protein of interest. Most product loss was observed in the lysate pellet and 2nd clarification pellet. The final product fraction (fraction 10) contains only a faint band at ~70 kDa, indicating that the protein was denatured or otherwise unable to be resuspended after the precipitation with ethanol. Product loss was confirmed by

BCA assay on the final product reporting negligible amounts of protein present (data not shown). The black arrow indicates the expected band from the desired product.

2.1.2.5 High-Throughput Screening of Streptavadin Binder Library in Self-Assembled Thin Films

Expression and purification of ESZ for preliminary SAXS experiments

ELP-Sso7d-ZE (ESZ) containing the engineered Sso7d variant used previously in our lab was expressed in BL21(DE3) *E. coli.* [16-17] ESZ was produced for small-angle x-ray scattering (SAXS) experiments to test the self-assembly of this construct as described subsequently. 50 mL of starter culture (inoculated from a single colony picked from an agar plate) in LB supplemented with 50 μg/mL kanamycin was grown in an Erlenmeyer flask overnight in an orbital shaker at 37 °C. Expression was carried out in a 7 L fermenter with a 5 L working volume in TB supplemented with 50 μg/mL kanamycin for 24 hours at 30 °C without induction. Cells were collected by centrifugation (4,000xg, 15 min at 4 °C), and resuspended in MENT buffer (100 mL/30 g wet cell mass). Resuspended cells were frozen at -80 °C and then thawed. Thawed cells were incubated for 2 h with 1 mg/mL lysozyme at 4 °C prior to tip sonication (3x5 min sonications at power level 5, 50% duty cycle with cooling to 15 °C between sonications). Lysates were clarified by high speed centrifugation (26,700xg for 1 h at 4 °C).

Clarified lysate was purified using the precipitation procedures described in **Purification of ELP-sfGFP** (including incubation with DNase A and RNase I during the first round of precipitation) where three different final products were produced. The first was purified with two rounds of NaCl-induced precipitation and dialyzed against MilliQ water using membranes with 6-8 kDa molecular weight cutoff, where dialysis consisted of 7 changes of the MilliQ water with at least 3 hrs between each change. Dialysis was set up such that $\leq$ 300 mL of protein solution was being dialyzed against 4 L of water for each change. Insoluble precipitates formed during dialysis were removed by high speed centrifugation (26,700xg for 1 h at 4 °C), and the supernatant was used as the final purified product. The second and third final products were purified with one round of NaCl-induced precipitation followed by one round of EtOH-induced precipitation. The second final product was then dialyzed and centrifuged to remove insoluble (as explained above for the first final product), and the third final product was not dialyzed. These three final products were then concentrated using centrifugal ultrafiltration (Amicon ® Ultra-15 Centrifugal Filter, MWCO

30 kDa) to 10-20 wt%. Concentrated fusion protein solutions were then drop cast onto Teflon sheets to prepare solid ESZ samples. Samples were then dried at room temperature under vacuum ramp (50 Torr per hour down to 5 Torr) for 12 h. SDS-PAGE on the different fractions produced by the purification processes described here are shown in **Figure 2-30**.

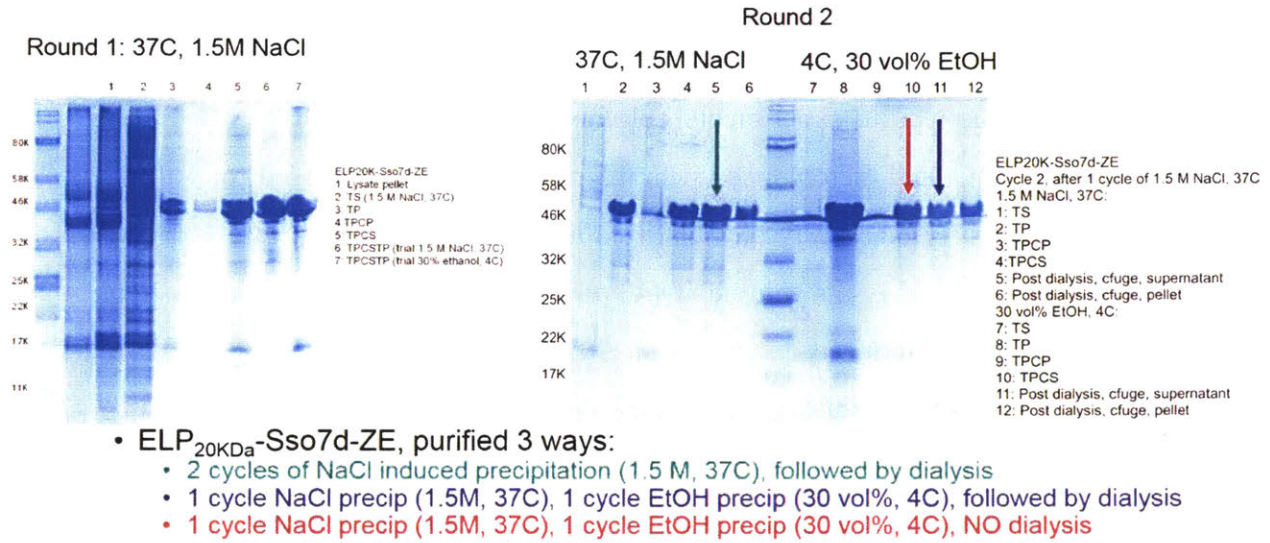


Figure 2-30. Coomassie blue stained SDS-PAGE gels showing the different fractions produced when purifying fermenter expressed ESZ using NaCl-induced and EtOH-induced precipitations. Two letters are associated with the fractions produced by each step, XY, where X is either T or C referring to a turbid/precipitation spin or clarification spin respectively. Y is either P or S, referring to the pellet or supernatant (respectively) produced by a given spin. For example, TP is the pellet produced by the precipitation/turbid spin and TPCS is the supernatant produced after clarifying the resuspended pellet from the turbid spin. Arrows on the gel indicate final products used in the SAXS study.

Well-plate expression and purification of ESZ

ELP-Sso7d-ZE (ESZ) constructs were expressed in BL21(DE3) *E. coli*. DNA and amino acid sequences for the engineered (expected to bind strongly to streptavidin) and naïve (expected not to bind to streptavidin) ESZ constructs used in preliminary experiments can be found in **A2.4 High-Throughput Screening of a Streptavidin Binder Library in Self-Assembled Solid Films**. Starter cultures for preliminary experiments expressing the same construct across the entire plate were prepared by inoculating 10 mL of LB supplemented with 50 μg/mL kanamycin with a small

amount of a previous starter culture produced in the Sikes lab by Eric Miller. Starter cultures were grown overnight in an orbital shaker at 37 °C. Expressions were carried out in a well-plate format using a procedure similar to that described in **Well-Plate Expression of ELP-sfGFP**. In this case, TB supplemented with 50 μg/mL kanamycin was used as the expression media, and 1000 μL of expression volume was used to produce the final product to minimize the possibility of spillover. Due to the lower expression volume, expressions were inoculated with 10 μL of starter culture in these experiments. As in previous experiments, cultures were grown to an OD_{600} of ~0.8 at 37 °C, chilled for 20 min at 4 °C, and then induced with 0.5 mM IPTG and expressed for 16-24 h at 24 °C. Cells were collected, lysed, and lysates clarified as described in **Well-Plate Expression of ELP-sfGFP** to produce a final lysate volume of 140 μL. Well plate expressions were then purified using one cycle of NaCl-induced precipitation followed by one cycle of EtOH-induced precipitation, where the final pellet here was resuspended in 125 μL of MilliQ water, to yield a total volume of 125 μL purified protein product per 1 mL protein expression. SDS-PAGE on the different fractions produced in this purification process (3 biological replicates per fraction) is shown in **Figure 2-31**. Protein yields, as measured by bicinchonic acid assay (BCA) are reported in **Figure 2-32**.

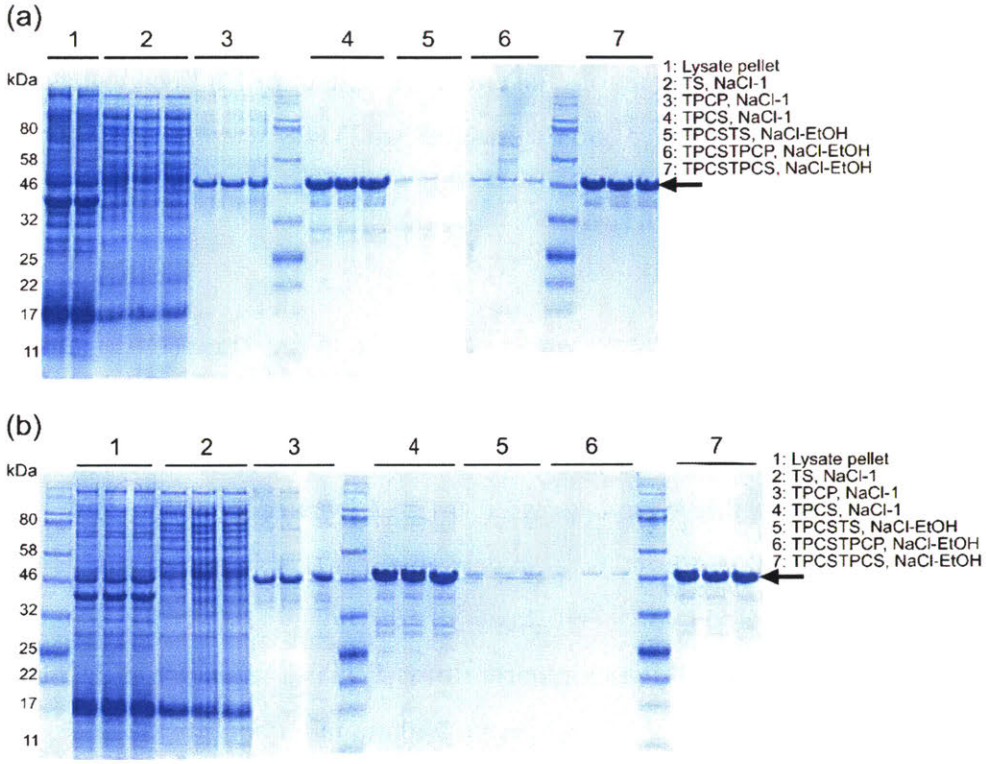


Figure 2-31. Coomassie blue stained SDS-PAGE gels on fractions produced during purification of ELP-Sso7d-ZEN fusion proteins containing a (a) naïve Sso7d sequence not expected to bind streptavidin and (b) an engineered Sso7d sequence that has been shown to strongly bind to streptavidin. Naming scheme for fractions is described in the caption of **Figure 2-30**. Black arrows indicate bands associated with the desired product.

(a)

Naive	1	2	3	4	5	6	7	8	9	10	11	12
A	0.523079	0.513611	0.442612	0.440833		0.521455	0.475717	0.508027		0.48333	0.438254	0.540524
B	0.551489	0.496159	0.42316	0.469277	0.549494	0.460345	0.528406	0.546684	0.530665	0.474644	0.51316	0.542516
C	0.54152	0.510999			0.534281	0.505058	0.495081	0.501011	0.46749	0.461774	0.514782	0.585244
D	0.487812	0.476433	0.439765	0.42795	0.444392	0.494632	0.483778	0.473838	0.499573	0.482344	0.541973	0.565112
E		0.496878			0.513611			0.482613			0.577128	
F	0.444214	0.506048	0.418285	0.444036		0.436298	0.396003	0.420855	0.440299	0.411202	0.413592	0.550854
G	0.536452	0.49158	0.452407	0.420944	0.442078	0.395385	0.411379	0.481717		0.43461	0.443502	0.521545
H	0.51307	0.506318	0.488709	0.476344	0.488888	0.409875	0.440299	0.429725	0.406603	0.444659	0.421121	0.500022

Engin.	1	2	3	4	5	6	7	8	9	10	11	12
A	0.639719	0.610856	0.463827	0.471602		0.438165	0.453299	0.440388		0.385598	0.419171	0.432212
B	0.78609	0.633548	0.406514	0.397062	0.435676	0.375741	0.378115	0.421919	0.355637	0.373806	0.44635	0.464631
C	0.550491	0.553666			0.467937	0.455884	0.42866	0.418817	0.430436	0.428128	0.498315	0.439143
D	0.642024	0.575123	0.50173	0.492837	0.546684	0.464363	0.422983	0.52669	0.496518	0.448576	0.461238	0.506858
E	0.496159	0.485839	0.396091	0.433455		0.442968	0.36081	0.400594	0.390005	0.405895	0.416691	0.470797
F	0.50209			0.448398			0.417665			0.50218		
G	0.59328	0.569114	0.501461	0.524433	0.472944	0.495979	0.494094	0.463024		0.499034	0.500742	0.654393
H	0.567113	0.488709	0.532473	0.478403	0.474017	0.451694	0.480552	0.526329	0.548043	0.516495	0.542154	0.663087

(mg/mL)

(b)

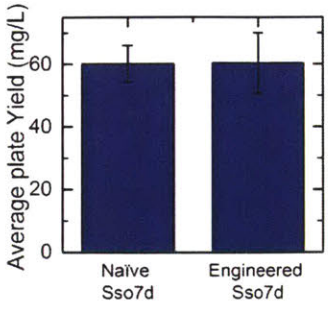


Figure 2-32. (a) Final concentration of two different ESZ variants (naïve and engineered) across the well plate in 125 L. The technical error on the concentrations reported here was measured to

be 0.016 ± .007 mg/mL, based on technical replicates ran on samples not included in this figure. (b) Average yield in mg/L expression across the entire well plate for ESZ constructs containing a naïve and engineered Sso7d.

2.1.2.6 Protein Expression for Other Projects

ELPY Expressions

Tyrosine containing ELPs of various lengths were expressed in the Tuner(DE3) strain of *E. coli*. Of the amino acid sequences listed above, four were expressed—Nt-ELPY10K, Nt-ELPY20K, Nt-ELPY30K, and Nt-ELPY40K. Nt-ELPY10K was expressed in a 7 L fermenter with a 5 L working volume, and the other three ELPs were expressed in 2.5 L baffled fernbach flasks with 1 L of expression media in them. All expressions were carried out in TB supplemented with 50 μg/mL kanamycin. Expressions were run for 17-20 h at 30 °C with no induction. These expression conditions were selected based on previous reports from M. J. Glassman and the Chilkoti lab that ELP expression yield can be improved by expressing without induction. Instead, leaky expression of the T7 promoter leads to accumulation of the ELP over time. [2, 7] No optimizations were performed on this expression. Cells were collected by centrifugation (4,000xg for 15 min at 4 °C), resuspended in MENT buffer (100 mL/30 g wet cell mass), and frozen at -80 °C. Resuspended cells were thawed, incubated with 1 mg/mL lysozyme at 4 C for 1-3 h, and then sonicated with a tip sonicator (2 times, 10 min each at power level 5, 50% duty cycle, with cooling to 15 °C in between sonications). Lysates were then clarified by centrifugation at 26,700xg for 1 h at 4 °C. Clarified lysates were subject to two rounds of NaCl-induced precipitation (see procedure in **Purification of ELP-sfGFP**) with 2.5 M NaCl at 37 °C used as the precipitant in this case. Note that RNase A and DNase I were added to the clarified lysate (as described in **Purification of ELP-sfGFP**) during the first incubation/precipitation at 37 °C to degrade any DNA/RNA left in the solution. Removal of residual DNA/RNA in this fashion is a crucial step in these precipitation-based purifications. After two rounds of NaCl-induced precipitation, solutions containing the ELPs of interest were dialyzed against MilliQ purified water (3-4 kDa molecular weight cutoff dialysis membranes, 7 changes of MilliQ water separated by at least 3 h each). Dialysis was set up such that ≤ 300 mL of protein solution was being dialyzed against 4 L of water for each change. Dialyzed products were then lyophilized, and solid protein samples were collected and weighed to determine final yield. SDS-PAGE gels showing the fractions collected during this process for one

cycle of the precipitations are shown in **Figure 2-33** for all four ELPs described here. Transition temperature measurements taken on Nt-ELPY10K in water and phosphate buffer, as well as a graph showing the change in protein yield with ELP length are shown in **Figure 2-34**.

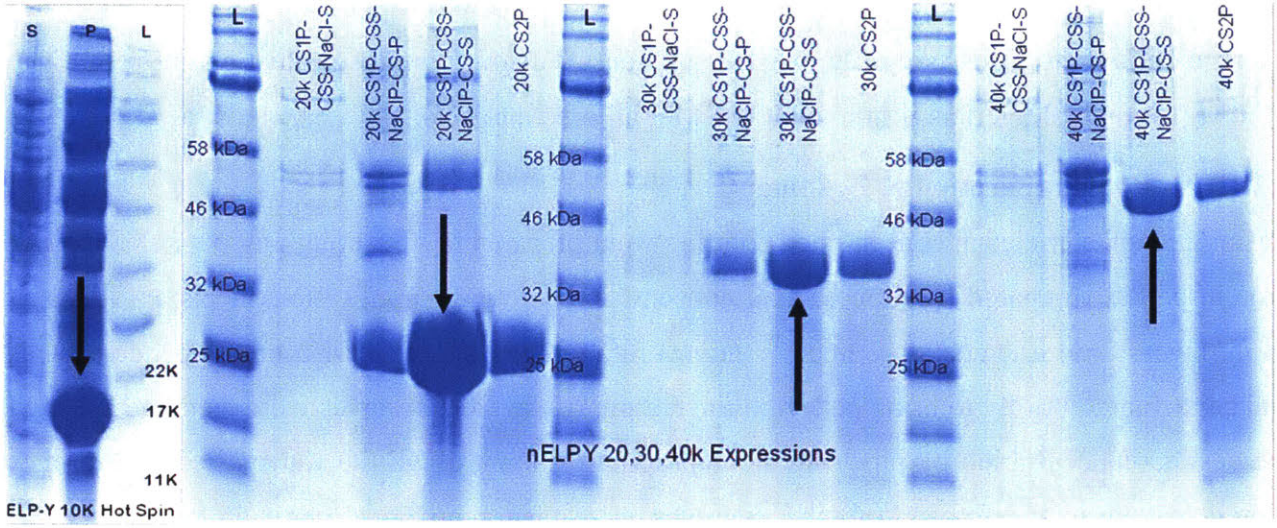


Figure 2-33. Coomassie blue stained SDS-PAGE gels showing different fractions produced when performing NaCl-induced precipitations on different tyrosine containing ELPs, where the fractions indicated with arrows are the ones collected with the desired products in them. From left to right, these arrows point to bands corresponding to Nt-ELPY10K, Nt-ELPY20K, Nt-ELPY30K, and Nt-ELPY40K.

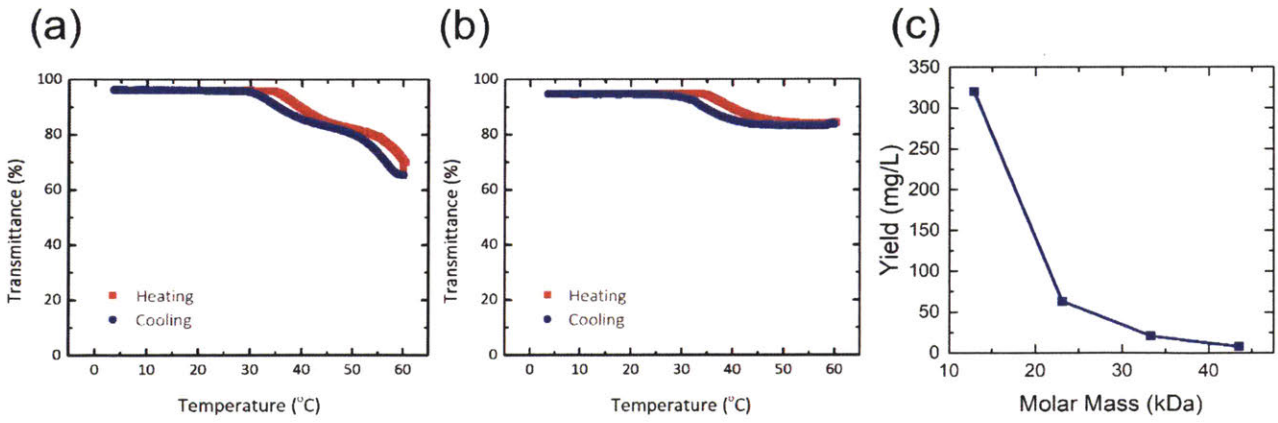


Figure 2-34. Transmittance versus temperature curves for Nt-ELPY10K in (a) water and (b) 20 mM pH 8 sodium phosphate buffer. (c) Final expression yield versus molar mass of tyrosine-containing ELPs of different lengths after 2 rounds of NaCl-induced precipitation.

2.2 Small-Angle Scattering

2.2.1 Small-Angle X-Ray Scattering (SAXS)

SAXS measurements were performed at the National Synchrotron Light Source II (NSLS-II) on beamline 16-ID (Life Science X-ray Scattering, LiX) for the **Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins** project and at beamline 11-BM (Complex Materials Scattering, CMS) for all other projects.

2.2.1.1 Sample Preparation & Loading

Fusion protein samples in concentrated solution were prepared by first weighing out solid fusion protein pellets (prepared by drop casting concentrated solutions onto Teflon sheets and drying them down at room temperature under vacuum) into either a 0.5 mL or 1.5 mL Eppendorf tube and centrifuging briefly to bring pellets to the bottom of the tube. Samples were then hydrated to the desired concentration by pipetting the appropriate amount of MilliQ water directly onto the protein pellets in the Eppendorf tube. Samples were then centrifuged again in a high-speed, refrigerated centrifuge ($\geq$ 6,000xg, 5-10 min at 4 °C) to ensure contact of the pellet and water. Tubes were then parafilmed, and samples were equilibrated at 4 °C for at least 30 min, and then centrifuged again prior to equilibration at 4 °C for at least 4 h (at least 12 h for samples $\geq$ 60 wt%). Samples were loaded into 1 mm thick anodized aluminum washers with a 2-3 mm bore and sealed between two pieces of Kapton® tape. To measure solid fusion protein samples, solid pellets were loaded into the bore of the same aluminum washers and sealed between two pieces of Kapton® tape. For these solid samples, a full, single pellet without bubbles was used (see **Figure 2-35a**)

Samples were then loaded into a brass sample stage that used a circulating chiller to control temperature (typically between 10 and 40 °C for samples measured in this work). An example of this process for uELP-mCherry is shown in **Figure 2-35b**.

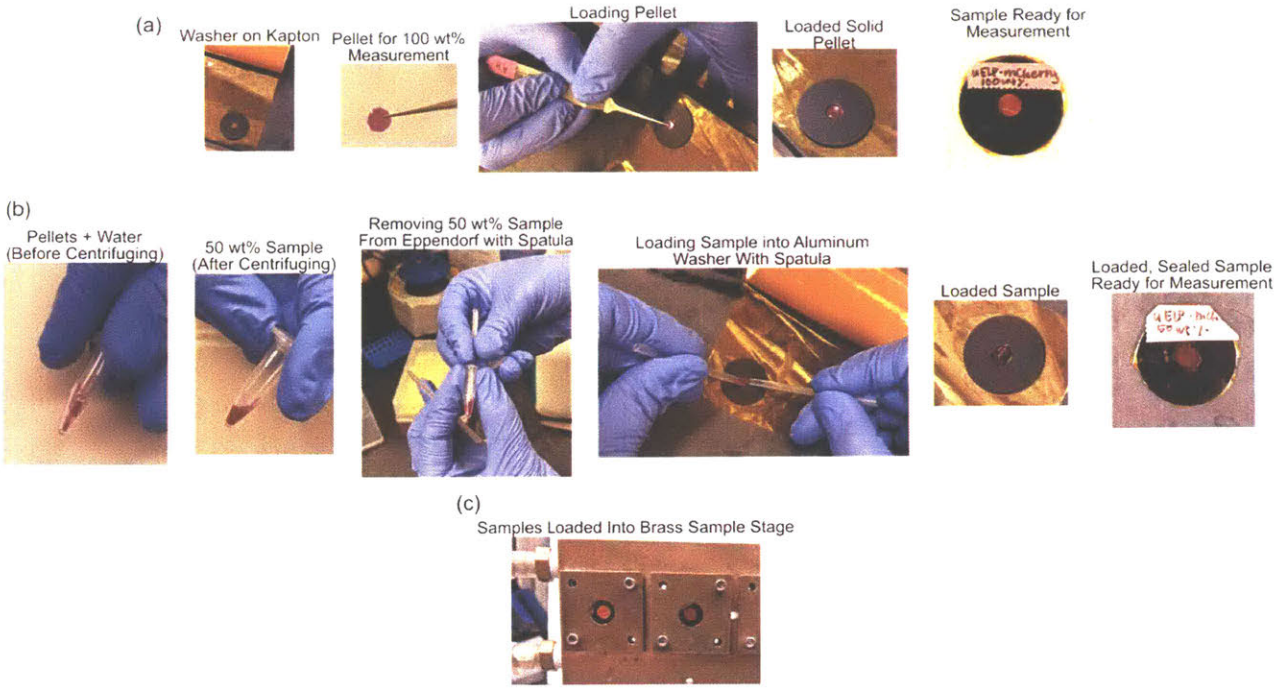


Figure 2-35. (a) Photos of the procedure used to load solid fusion protein samples into aluminum washers for SAXS measurements. Here, an aluminum washer with a 5 mm diameter bore is shown for ease of visualizing the process. (b) Photos of the procedure used to prepare and load uELP-mCherry samples into aluminum washer for SAXS measurements. Here, an aluminum washer with a 5 mm diameter bore is used to enable easy visualization of the process. (c) Example of the samples prepared in (a) and (b) loaded into the brass sample stage used to run SAXS measurements.

Fusion protein samples measured in a well-plate format were dried down from 75-100 μL of protein solution in a 96-well plate (and crosslinked, if applicable) as described below (see **2.10 Casting & Crosslinking of Fusion Protein Films**). Well plates were mounted into the sample environment by placing a plate lid on them that had the center cut out, parafilming around the edge (see **Figure 2-36**) and standing the plate up on its side. The plate was then held in place on a rubber mat (to prevent slipping) using a clamp to put pressure on the upper edge of the plate. Wells in the plate could then be navigated using normal sample stage commands by knowing the distance between wells in the plate. Measurements performed on ELP-sfGFP films prepared in this fashion did not produce sufficient scattering to detect any structure, but with thicker samples, it should be possible to make measurements on fusion protein films using this setup.

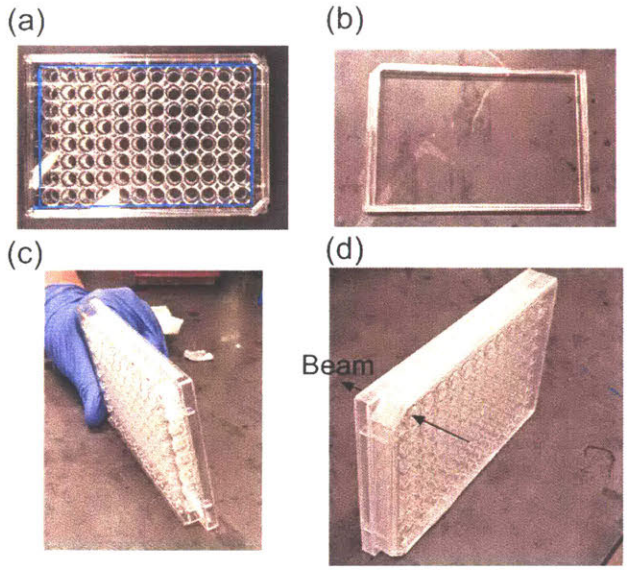


Figure 2-36. (a) Photo indicating the area to be cut out of a plate lid for well-plate SAXS measurements. (b) Photo of well-plate lid with center cut out for well-plate measurements. (c) Photo of plate trying to stand up with no lid on, demonstrating the difficulty of running measurements without a lid on the plate. (d) Well-plate with lid secured to the plate using parafilm. Plate is shown standing up, with beam direction indicated by arrows.

2.2.1.2 Data Collection & Processing

For concentrated solution samples where temperature was controlled, samples were equilibrated in the sample stage at 10 °C for 20 min and 10 min at all other temperatures measured. SAXS data was collected with a 10 sec acquisition time and corrected for empty cell scattering using the equation described below. This acquisition time was found to give good scattering signal without causing beam damage to the samples. Background correction was applied to 2D scattering patterns using the following equation

$$I(q) = \frac{I_{\text{sample}}}{T_{\text{sample}}} - \frac{t_{\text{sample}}}{t_{\text{empty}}} \frac{I_{\text{empty}}}{T_{\text{empty}}} \tag{2-1}$$

where I, t, and T are intensity, acquisition time, and transmission, respectively. The subscript sample refers to sample scattering intensity, acquisition time, or transmission, and the subscript empty refers to the empty cell scattering intensity, acquisition time, or transmission. Data was converted from 2D to 1D after background correction by taking an azimuthal average of the 2D data. Empty cells here were produced by leaving the center of the 1 mm thick aluminum washer

empty and encasing both sides with Kapton® tape. Background corrections and averages on data acquired at beamline 16-ID (**Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins**) were performed in MATLAB using code developed by M. K. Sing, published in her thesis.[18] Corrections to all other data were performed on-site using software provided by beamline scientists.

2.2.2 Small-Angle Neutron Scattering (SANS)

Measurements were performed at the Extended Q-Range Small-Angle Neutron Scattering Diffractometer (BL-6) at the Spallation Neutron Source (SNS) at Oak Ridge National Laboratory. Three detector configurations were used: two configurations at a 4 m detector distance for collecting 1.2 and .25 nm wavelength neutrons, and one configuration at 1.3 m detector distance, also for .25 nm wavelength neutrons. This set of configurations allowed data collection over a q-range of 0.0042-0.76 nm^{-1}.

Solid protein samples (either drop-cast fusion protein samples or lyophilized ELPs) were weighed out into 1.5 mL Eppendorf tubes and hydrated with deuterium oxide to a final concentration of 25 mg/mL. Tubes were sealed with parafilm and allowed to equilibrate for at least 12 hours at 4 °C prior to measurement. After complete dissolution, samples were passed through a 0.2 μm syringe filter to remove any dust. Samples were then loaded into glass banjo cells with a 1 mm pathlength and sealed with a stopper prior to loading into sample cells. Measurements were performed on samples equilibrated at 10 °C. Intensity data were corrected to absolute intensity units using a porous silica standard. 2D scattering data were azimuthally averaged to obtain 1D curves. Background data was taken on deuterium oxide alone in a 1 mm pathlength banjo cell of the same make. Background subtraction (using $I_{final} = I_{sample} - I_{background}$) was performed by subtracting 1D curves from each other.

2.3 Depolarized Light Scattering (DPLS)

Birefringence of fusion protein samples in concentrated solutions was determined by performing measurements of turbidimetry and depolarized light scattering. Samples were loaded into 1 mm thick, 3 mm diameter bore Teflon mold and sealed between two quartz disks in a brass sample cell. Measurements were performed using a coherent OBIS LX660 laser at wavelength 662 nm (beyond the absorption cutoff of mCherry) with continuous wave power output of 10 mW. All samples were equilibrated at 10 °C for 10 min and then heated at 1 °C min^{-1} to 40 °C, allowed to

equilibrate for 10 min at 40 °C, and then cooled back to 10 °C at 1 °C min^{-1}. Two of these temperature ramps were run in order to collect transmission and light polarization. Transmission was measured by passing the laser through a polarizer and then through the sample, and finally into the detector during the first temperature ramp. This measurement was performed first to anneal out any sample alignment that may have occurred due to shear during sample loading. Static birefringence was measured by the same setup as transmission, but with a second polarizer (perpendicular to the first polarizer) between the sample and the detector. Measured light polarization was corrected for transmission and dark field background using the equation below. All transitions were reproducible on repeated cycling.

$$\text{Power Fraction} = \frac{I_{\text{open}}}{I_{\text{transmitted}}}\left(I_{\text{polarized}} - I_{\text{cross}}\right) - \left(I_{\text{cross}} - I_{\text{dark}}\right) \qquad (2\text{-}2)$$

Where I_{open} is the intensity of the open beam, $I_{\text{transmitted}}$ is the intensity transmitted through the sample with no second polarizer, $I_{\text{polarized}}$ is the intensity transmitted through the sample and the second polarizer, I_{cross} is the intensity transmitted through the first and second polarizer with no sample, and I_{dark} is the measured intensity when no laser is on.

2.4 Dynamic Light Scattering (DLS)

DLS measurements on samples for the **Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins** projects were taken on a Wyatt DynaPro Plate Reader DLS instrument with an 850 nm laser on 50 μL of sample in a 384 well, glass-bottom plate. ELP samples were prepared by weighing lyophilized ELP samples into 15 mL conical tubes and hydrating them with MilliQ water to a concentration of 5 mg/mL. mCherry samples were prepared at 5 mg/mL by diluting down more concentrated samples whose concentrations were determined by measuring absorbance at 280 nm and using the molar extinction coefficient and molecular weight of mCherry (ε_{A280}=32,430 M^{-1} cm^{-1}, molecular weight = 26.722 kDa). All single-component solutions were filtered through a 0.45 μm syringe filter after adjustment of concentration. Blended solutions for this study were prepared by mixing filtered solutions together at a weight ratio of 63:37 ELP:mCherry (representative of the approximate fraction of ELP and mCherry in the fusion proteins). Each measurement was averaged over ten 5 sec acquisitions.

DLS measurements for the project **Protein Purification by Ethanol-Induced Phase Transitions of the Elastin-Like Polypeptide (ELP)** were performed on a Wyatt Möbiuζ DLS/zeta potential instrument with a 532 nm laser on samples in disposable polystyrene cuvettes with a 1 cm path length. In this case, measurements were performed directly on protein purification fractions of interest with no filtration or concentration adjustment. Size distributions reported in this study were produced by performing fits to autocorrelation curves in the Wyatt Dynamics software using their proprietary regularization algorithm.

2.5 Circular Dichroism (CD)

All samples measured by CD were dialyzed against MilliQ water prior to measurement (due to the strong effect buffer conditions can have on the noise observed in this measurement) and were prepared at 0.5 mg/mL. For ELPs, solutions were prepared first by weighing lyophilized protein into a 15 mL conical tube and hydrating the samples to 5 mg/mL in MilliQ water. Additional water and ethanol was then added to the sample to adjust the final protein concentration to 0.5 mg/mL and adjust the solvent composition to the ethanol/water ratio desired for measurement. ELP-sfGFP samples were prepared by first obtaining the ELP-sfGFP concentration via absorbance measurement at 280 nm with a cuvette held at 4 °C (to prevent any absorption due to ELP aggregation). Molar absorption and molecular weight of ELP-sfGFP were assumed to be ε_{A280} = 17,210 M^{-1} cm^{-1} and MW = 73.307 kDa for this measurement. ELP-sfGFP concentration was then adjusted to 0.5 mg/mL with MilliQ water prior to measurement.

Samples were loaded into a quartz cuvette with a 1 mm path length and sealed. They were scanned in a JASCO Model J-1500 circular dichroism spectrometer equipped with a Peltier temperature controller. Spectra were collected with a 0.5 nm resolution, with a bandwidth of 1 nm, and a scan speed of 50 nm per minute. All samples were background corrected by subtracting off the signal produced by the solvent alone in the same cuvette. Data was converted to mean residue molar ellipticity based on the prepared concentration (0.5 mg/mL) and the amino acid sequence of each protein.

2.6 Turbidimetry

All turbidimetry data taken in this thesis was performed on samples with a final protein concentration of 1 mg/mL. For the project **Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins**, ELP samples were prepared at 1 mg/mL by

weighing sample into a 15 mL conical tube and hydrating with either MilliQ or buffer to a final concentration of 1 mg/mL. Samples for the project **Cononsolvency of elastin-like polypeptides (ELPs) in water/alcohol solutions** were prepared by weighing lyophilized ELP protein into a 15 mL conical tube, hydrating it with MilliQ water to 5 mg/mL, and fully dissolving the ELP at 4 °C. 1 mL samples containing the final desired alcohol/salt contents were then prepared by starting with 200 μL of 5 mg/mL ELP solution in a 1.5 mL Eppendorf tube. MilliQ water, 500 mM NaCl in water, and alcohol were then added (in that order) to achieve a protein concentration of 1mg/mL and the final alcohol/salt content of interest.

After preparation, samples were loaded into a quartz cuvette and sealed with a cap and parafilm. All measurements were started at temperatures where the ELP solutions were soluble—either at the lowest temperature of the temperature ramp for solutions that became turbid upon heating, or at the highest temperature of the temperature ramp for solutions that became turbid upon cooling. Cuvettes were ramped at a rate of 1 °C/min using a Peltier temperature stage. Solution transmittance at 600 nm was monitored over the duration of the ramp. Transition temperatures were taken as the point in the turbidimetry curve where the transmittance reached 50% of the minimum transmittance.

2.7 Yield Measurements

For ELP-sfGFP, yield was measured by both ELP-sfGFP fluorescence and bicinchonic acid assay (BCA). All other protein yields reported here were determined by BCA. ELP-sfGFP fluorescence was measured on a Tecan Infinite M200 Pro plate reader with an excitation wavelength of 395 nm and an emission wavelength of 510 nm. Fluorescence parameters for this measurement were optimized based on the fluorescence of ELP-sfGFP at 1.5 mg/mL in solution.

When used to measure expression yields, BCA assays were performed following the manufacturer's instructions using provided bovine serum albumin (BSA) standards (Thermo Scientific #23227). Absorbance at 562 nm was measured using a Tecan Infinite M200 Pro plate reader. As suggested by the manufacturer, calibration curves were fit to a quadratic curve (where the absorbance at 562 nm produced by a given protein concentration was assumed to be a quadratic function of the protein concentration). An example calibration curve is shown in **Figure 2-37**. Calibration curves were generated every time the assay was run alongside samples to eliminate any variation in incubation time and/or temperature. BCA assays used for construction of the ELP-

sfGFP phase diagram used different calibration curves, as described in the section below (**2.8 Protein Partitioning into Soluble/Insoluble Phases as a Function of Salt/Ethanol Concentrations**).

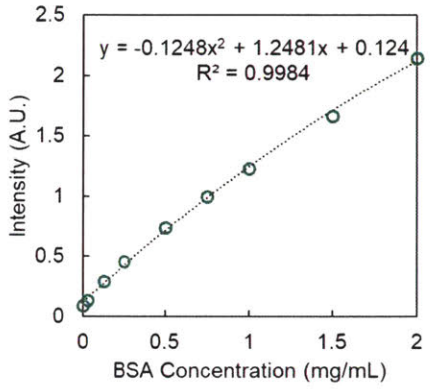


Figure 2-37. Sample BCA calibration curve generated with BSA standards provided by the manufacturer of the BCA assay kit. Data are fit to a quadratic curve (as suggested by the manufacturer). Fit coefficients for this curve are shown on the plot. A similar curve taken on a different day produced a fit of $y = -0.1497x^2 + 1.034x + 0.1134$, highlighting that, while there are not dramatic differences in these coefficients depending on the day the assay is run, there is enough variability that the curves should be run anew for each BCA.

2.8 Protein Partitioning into Soluble/Insoluble Phases as a Function of Salt/Ethanol Concentrations

ELP-sfGFP partitioning into soluble and insoluble phases was measured as a function of sodium chloride and ethanol concentration (see **Chapter 5**). Protein samples were prepared at the specified ethanol and sodium chloride concentrations using stock solutions of pure ethanol, 5 M NaCl in water, 0.5 M NaCl in water and purified, dialyzed protein. Reagents were added to wells in a 96-well microtiter plate to 200 μL in the following order: water, protein stock, NaCl, and then EtOH. Three replicate plates were prepared for each temperature condition. The same volume of protein (40 μL) was added to each well, and the volumes of other reagents were adjusted to produce the sodium chloride and ethanol concentrations of interest. Plates were sealed with Parafilm and enclosed in an airtight box to mitigate solvent evaporation, then incubated for 16 h at either 4 °C or 20 °C. Plates were centrifuged at 6,000 xg at the relevant temperature for 1 h. Supernatants were

removed to a new plate, and protein concentration was determined immediately (to mitigate solvent evaporative effects) by BCA assay (Pierce) following the manufacturer's instructions using an α-amylase standard curve (see below). The insoluble fractions were resuspended in 200 μL Milli-Q water by mixing with a multichannel pipette and rocking for 16 h at 4 °C. Protein concentration was again determined by BCA assay as for the supernatants.

BCA was used to determine protein concentration based on preliminary experiments constructing similar phase diagrams. In these experiments, three methods for measuring protein concentration were compared—BCA, Bradford, and sfGFP fluorescence. Methods were evaluated based on their ability to reproducibly close the mass balance; that is, whether or not the concentrations measured in the soluble and insoluble fractions summed to the same value across the different ethanol and sodium chloride concentrations used. **Figure 2-38** shows a summary of these preliminary experiments. In **Figure 2-38**a, a summary of each technique used is provided. The remaining panels of the figure summarize the results of a study on mass balance closure across the different conditions studied. For each technique, fluorescence or protein concentration in resuspended pellets and supernatants was measured across three replicate plates. Measured fluorescence/concentrations for pellets and supernatants were then summed for each plate. The average and standard deviations of these summed values over the three plates were then taken. Finally, the averages and standard deviations of the summed concentration/fluorescence in each plate were divided by the average fluorescence or protein concentration at 0 mM NaCl and 0% EtOH. These are the "avg" and "stddev" reported in each panel of **Figure 2-38**b-d.

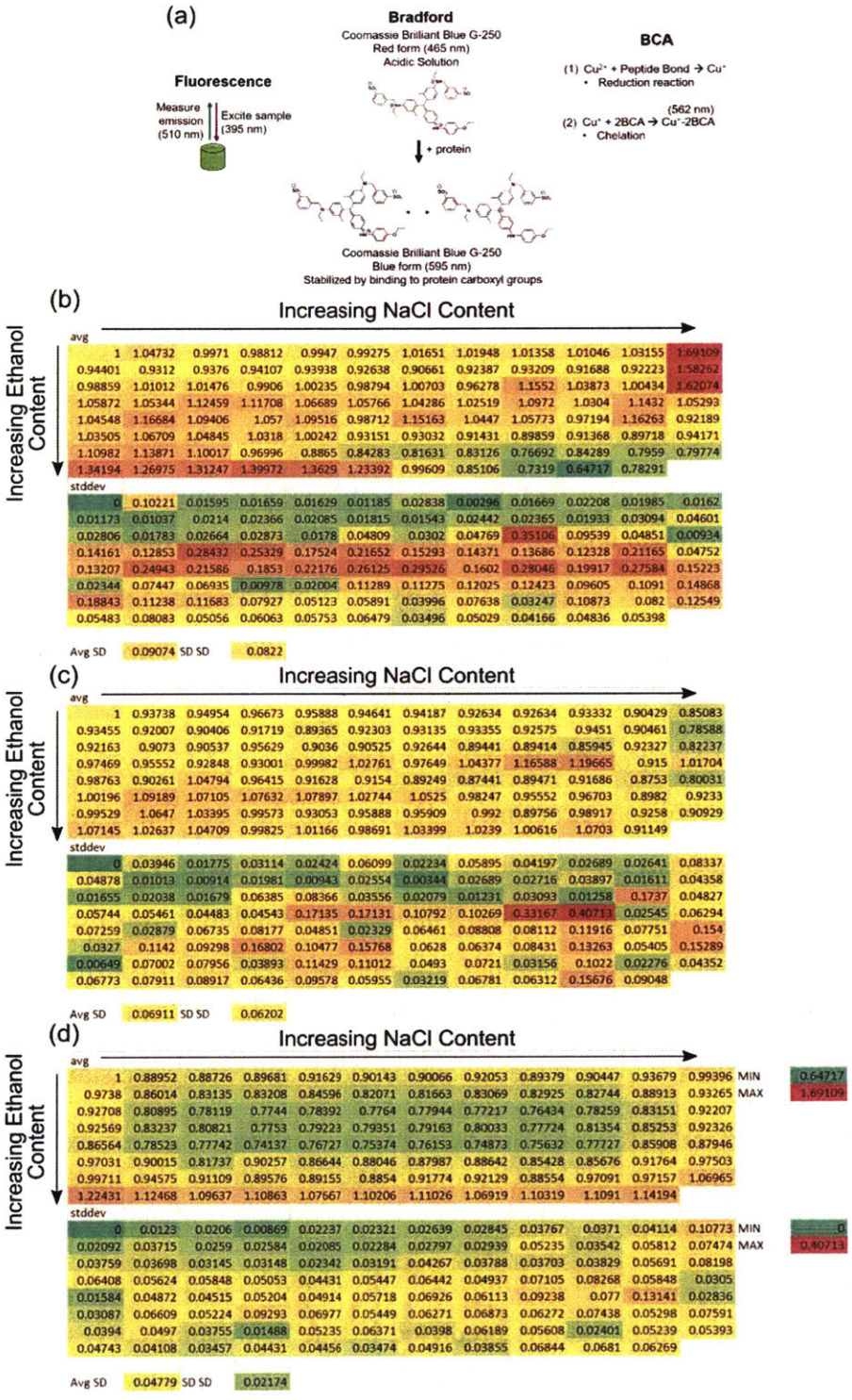


Figure 2-38. (a) Summary of different techniques used to measure ELP-sfGFP concentration in partitioning experiment. Mass balance experiment results (procedure described in main text) for ELP-sfGFP concentrations measured by (b) fluorescence, (c) Bradford assay, and (d) bicinchoninic acid (BCA) assay.

Figure 2-38b shows these results for protein concentration measured by sfGFP fluorescence. In this case, the mass balance deviates strongly from 1 at high NaCl concentrations, evidenced by the large red patch in the averaged mass balance in the upper right corner. At high ethanol contents (70 vol% ethanol), the fluorescence is also strongly affected by the addition of NaCl, as evidenced by the shift from orange to green in the bottom row of the averaged mass balance panel. When using fluorescence to calculate ELP-sfGFP fluorescence, across three well-plates, the average standard deviation in the mass balance was 0.09 ± 0.08. This combination of results indicated that, while this fluorescence technique gave reproducible results, addition of salt and ethanol had an impact on the measured fluorescence, leading to deviations from 1 in the mass balance at certain points in phase space. When comparing Bradford and BCA assays, the effects of ethanol and NaCl on mass balance closure are much less apparent, as evidenced by a lack of strong systematic trends in the averaged mass balance data; however, BCA led to a lower average standard deviation across replicate plates (0.05 ± 0.02 for BCA versus 0.07 ± 0.06 for Bradford). Thus, BCA was selected as the method of choice for the final partitioning study.

The effects of ethanol and NaCl on the BCA assay were accounted for by generating 9 standard curves at the time of each replicate experiment. Here, α-amylase standards were prepared in lab by weighing purchased α-amylase (Sigma-Aldrich #10065) into a 15 mL conical tube and hydrating it to ~10 mg/mL with MilliQ water. The α-amylase solution was then dialyzed against MilliQ water (3-4 kDa molecular weight cutoff membrane, 7 changes of MilliQ water separated by at least 3 hours) to remove any residual salt. The final concentration of the α-amylase was then measured by BCA using BSA standards (see above section), and α-amylase standard curves in the various solvents of interest were generated. Each curve was generated with 9 concentrations of α-amylase prepared in 0, 30, or 60 vol% EtOH and 0, 100, or 500mM NaCl. Each standard curve was generated in triplicate and fit to a quadratic function. The protein concentration in solvents that did not have a standard curve (for example, 20 vol% EtOH, 50 mM NaCl) was calculated by bilinearly interpolating the concentrations predicted by the calibration curves at the conditions bracketing the solvent condition of interest. For example, the concentration of protein at 20 vol% EtOH, 50 mM NaCl would be calculated by interpolating between the concentrations predicted at 0 vol% EtOH/0 mM NaCl, 0 vol% EtOH/100 mM NaCl, 30 vol% EtOH/0 mM NaCl, and 30 vol%/100 mM NaCl.

Examples of these different curves are shown in **Figure 2-39**. Curves of the form $y = ax^2 + bx + c$ were used to fit all curves. R^2 values for all curves presented here show that the curves fit well to this quadratic model, as expected. For the standards presented in this figure, the b and c parameters are relatively well conserved, with average values of 1.4 ± 0.1 and 0.013 ± 0.005 ($\pm$ indicates standard deviation). The a value, however, fluctuates substantially, as it takes on both negative, positive, as well as near-zero values. Fluctuations in a do not seem to trend with increasing NaCl or EtOH concentrations, however.

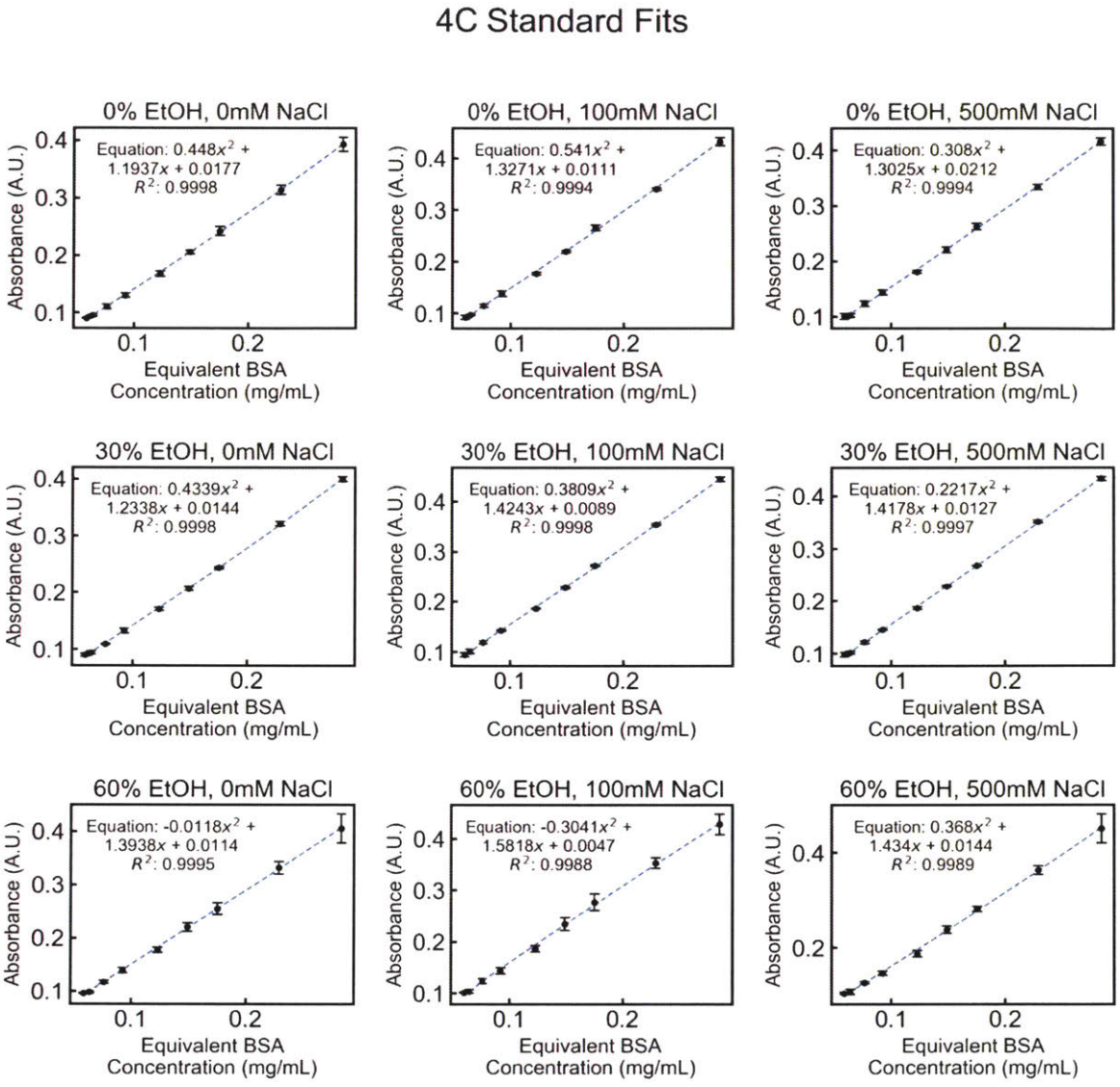


Figure 2-39. Standard curves used to calculate (and then interpolate for) ELP-sfGFP concentration, accounting for differences in solvent ethanol content and NaCl content.

2.9 Conductivity

Conductivity measurements were performed to determine residual salt levels in purified protein solutions after different purification schemes. Proteins were removed from conductivity samples using centrifugal ultrafiltration using filters with a 10 kDa molecular weight cutoff. Filtrates of

this centrifugal ultrafiltration were measured using a Mettler Toledo SevenCompact™ Duo S213 conductivity probe. This probe was used in the lab of Professor Martin Bazant with assistance from Dr. Kameron Conforti. The probe was calibrated using a NIST traceable 1413 μS/cm standard (692 ppm as NaCl). Measured conductivities were converted into molar equivalents of NaCl using a conversion curve generated in the following manner. Samples containing 0 mM, 10 mM, 50 mM, and 100 mM sodium chloride in MilliQ water were prepared in triplicate in 15 mL conical tubes. Their conductivities were measured to construct a conversion curve, which showed a linear relation between NaCl concentration and conductivity. This calibration curve is shown in **Figure 2-40**.

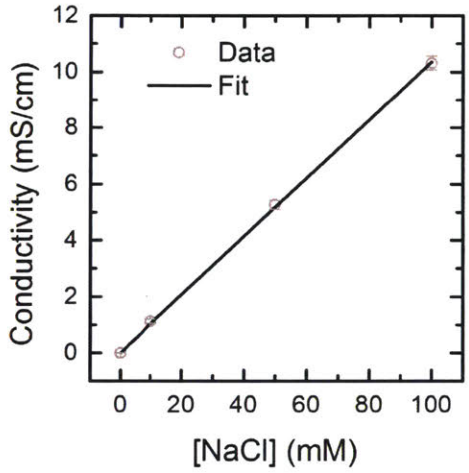


Figure 2-40. Conductivity versus concentration of sodium chloride data points (and fit to data points) used to convert conductivities of solutions to molar equivalents of NaCl. Error bars on data points indicate standard deviation of conductivity over three replicate samples.

2.10 Casting & Crosslinking of Fusion Protein Films

ELP fusion protein films were cast in well plates by the following procedure. After expression in a 96-well plate (as described above) and purification with one NaCl-induced and one EtOH-induced precipitation, a final purified, desalting protein solution with volume 75-100 μL remained in each well. These ELP-fused globular protein solutions were cast into films by drying down well plates at room temperature under a controlled vacuum ramp (50 Torr per hour down to 5 Torr) for 12-16 h, or until all liquid had been removed from the plate. Plates were dried down with lids on to prevent contamination of wells with dust.

After casting, films were crosslinked using glutaraldehyde chemistry to lightly crosslink lysine residues in the globular protein portion of the film (**Figure 2-41a**). Plates with cast films were placed in a stirred bath on a hot plate with the temperature held at 40 °C. The purpose of this bath was twofold. First, it accelerated the glutaraldehyde crosslinking reaction. Second, it prevented delamination of the films prior to crosslinking, as the ELP block in the fusion protein is insoluble in the crosslinking solution at 40 °C. A glass pyrex dish was used to accommodate two plates at a time, and aluminum bars were clamped in place to prevent plates from sinking into the water bath (**Figure 2-42**). To achieve a temperature of 40 °C in the well plates, a set point temperature of 43 °C was used for the water bath. Two plates were set up in the bath—one with the films to be crosslinked, and one with 200 μL of 1.4 wt% glutaraldehyde solution in each well. Plates were placed in the bath and allowed to come to thermal equilibrium (which typically took 10-15 min). After plates were equilibrated, the crosslinking reaction was initiated by using a multichannel pipette to move the glutaraldehyde solutions into the plate with the ELP-fusion protein films. Films were incubated with this glutaraldehyde solution for 1-2 min. The glutaraldehyde solutions were then removed with a multichannel pipette. Crosslinked films were removed from the heating bath after removal of glutaraldehyde. Films were then washed immediately three times with 200 μL cold MilliQ water used in each wash. A final overnight wash of films was then performed where 200 μL MilliQ water was added to the films and left overnight. Images of ELP-sfGFP films cast and crosslinked in this way are shown in **Figure 2-41b**. Film dissolution was also monitored in crosslinked and uncrosslinked films by tracking ELP-sfGFP fluorescence in the wash solution. **Figure 2-41c** shows the results of this experiment, demonstrating that minimal ELP-sfGFP dissolves into the wash buffer (even after overnight incubation at 4 °C), whereas uncrosslinked films completely dissolve after an overnight wash. It should be noted that no treatment of the plate was necessary to achieve adherent films.

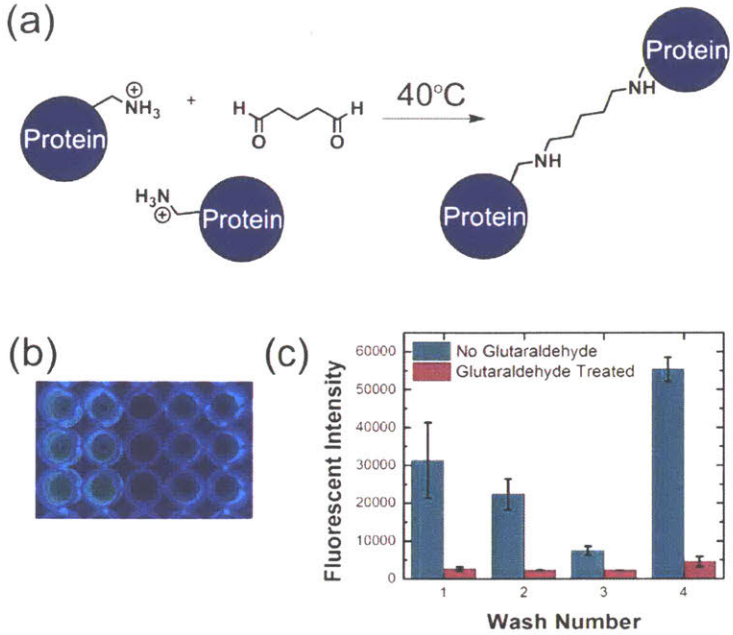


Figure 2-41. (a) Glutaraldehyde chemistry used to crosslink surface lysine residues in the ELP-globular protein films. (b) Image of ELP-sfGFP films after casting and crosslinking, where the two sets of six wells correspond to two different initial ELP-sfGFP concentrations in the casting solution. (c) Fluorescence of ELP-sfGFP in the wash solutions for both glutaraldehyde treated and non-glutaraldehyde treated films demonstrating that treatment with glutaraldehyde successfully immobilizes the films at the bottom of the well plate.

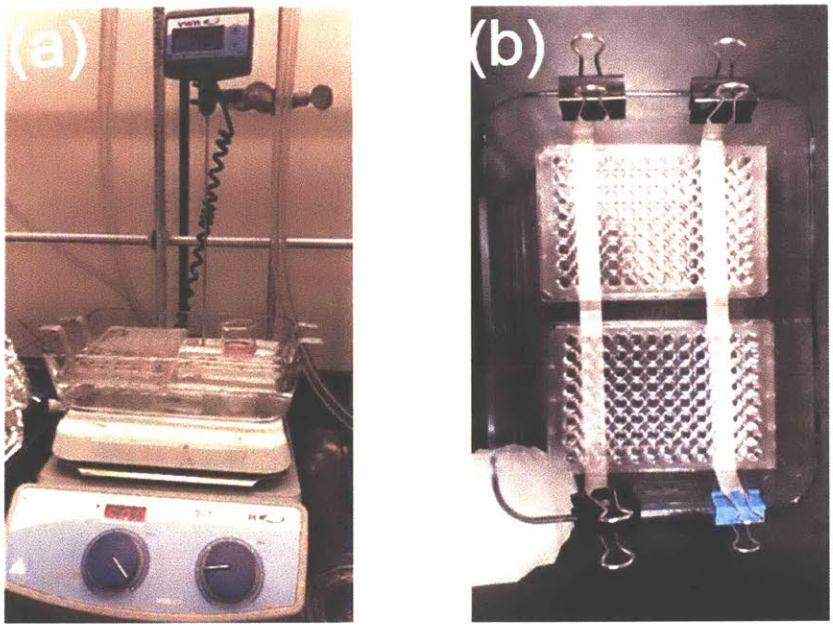


Figure 2-42. Photos of the setup used to control the temperature of films cast in 96-well plates. (a) Setup of the rectangular Pyrex dish on a hot plate with stirring and temperature control. (b) Top view of the Pyrex dish with two 96-well plates loaded in.

2.11 Enzyme Activity Assays

Colormetric assays for enzymatic activity were all carried out in clear, 96-well polystyrene microtiter plates using a Tecan Infinite M200 Pro plate reader.

2.11.1 Paraoxon-Based Assay

Activity of ELP-OPH-ELP as well as off-target activity of diisopropyl phosphatase (DFPase) was measured based on methyl paraoxon hydrolysis. Hydrolysis of methyl paraoxon can be tracked by monitoring the formation of the optically active hydrolysis product, *p*-nitrophenol. In these assays, *p*-nitrophenol concentration was tracked by measuring absorbance at 405 nm. Because of the high activity of ELP-OPH-ELP, serial dilutions of ELP-OPH-ELP (1:10, 1:50, 1:250, 1:1250, 1:6250) were assayed in triplicate over 5 minutes to ensure that a linear activity curve could be accessed. The highest ELP-OPH-ELP concentration that resulted in a linear absorbance versus time curve was used to calculate specific ELP-OPH-ELP activity. The slope of this line was determined using linear regression, and the specific activity was calculated using the following equation

$$\text{specific activity}\left(\frac{\mu\text{mol}}{\text{min mg}}\right) = \frac{\text{slope} \times 10^6 \times \text{dilution factor}}{1000 \times 17{,}100 \times P \times C} \quad (2\text{-}3)$$

where dilution factor refers to how much the OPH solution had to be diluted to achieve a linear slope (*i.e.*, if the 1:10 dilution produced a linear slope, the dilution factor would be 10), slope is the slope of the absorbance versus time curve (acquired by linear regression, as mentioned above), C is the concentration of ELP-OPH-ELP in the original solution, P is the path length of the light in cm, and 17,100 is the molar extinction coefficient of *p*-nitrophenol at 405 nm in M^{-1} cm^{-1}. All OPH-based assays were performed in buffer containing 50 mM (4-(2-hydroxyethyl)-1-piperzineethanesulfonic acid) (HEPES) and 100 μM cobalt (II) chloride at pH 8, where HEPES buffer was used to minimize background hydrolysis of paraoxon, and cobalt (II) chloride was included because it is a cofactor in the OPH active site and, as such, is required for OPH activity.

Assays for DFPase off-target activity were never converted to specific activities; however, a similar calculation (without any dilution factor) could be performed by monitoring methyl paraoxon degradation with DFPase. DFPase activity was measured in buffer containing 20 mM tris and 100 μM calcium chloride (because Ca^{2+} is a cofactor in the DFPase active site). Note that path length of a given volume of liquid in a microplate can be calculated by measuring the absorbance of the same liquid in both a cuvette and in a well plate at 975 and 900 nm and using the following equation

$$\text{Pathlength}^{\text{well plate}} = \frac{A_{975\text{ nm}}^{\text{well plate}} - A_{900\text{ nm}}^{\text{well plate}}}{A_{975\text{ nm}}^{\text{cuvette}} - A_{900\text{ nm}}^{\text{cuvette}}} \times \text{Pathlength}^{\text{cuvette}} \quad (2\text{-}4)$$

where A is measured absorbance at the wavelength indicated in the subscript and in the format indicated in the superscript.

2.12 References

1. Wright, E. R.; Conticello, V. P., Self-assembly of block copolymers derived from elastin-mimetic polypeptide sequences. *Adv. Drug Delivery Rev.* **2002,** *54* (8), 1057-1073.

2. Glassman, M. J.; Olsen, B. D., End Block Design Modulates the Assembly and Mechanics of Thermoresponsive, Dual-Associative Protein Hydrogels. *Macromolecules* **2015,** *48* (6), 1832-1842.

3. Glassman, M. J. Synthesis, nanostructure, and mechanics of thermoresponsively tough biomaterials from artificial polypeptides. Massachusetts Institute of Technology, Cambridge, Massachusetts, 2015.

4. Qin, G.; Glassman, M. J.; Lam, C. N.; Chang, D.; Schaible, E.; Hexemer, A.; Olsen, B. D., Topological Effects on Globular Protein-ELP Fusion Block Copolymer Self-Assembly. *Adv. Funct. Mater.* **2015,** *25* (5), 729-738.

5. Qin, G.; Perez, P. M.; Mills, C. E.; Olsen, B. D., Effect of ELP Sequence and Fusion Protein Design on Concentrated Solution Self-Assembly. *Biomacromolecules* **2016,** *17* (3), 928-934.

6. Miller, E. A.; Traxlmayr, M. W.; Shen, J.; Sikes, H. D., Activity-based assessment of an engineered hyperthermophilic protein as a capture agent in paper-based diagnostic tests. *Mol. Syst. Des. Eng.* **2016,** *1* (4), 377-381.

7. Fancy, D. A.; Kodadek, T., Chemistry for the analysis of protein-protein interactions: Rapid and efficient cross-linking triggered by long wavelength light. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96* (11), 6020-6024.

8. Trabbic-Carlson, K.; Liu, L.; Kim, B.; Chilkoti, A., Expression and purification of recombinant proteins from Escherichia coli: Comparison of an elastin-like polypeptide fusion with an oligohistidine fusion. *Protein Sci.* **2004,** *13* (12), 3274-3284.

9. McPherson, D. T.; Xu, J.; Urry, D. W., Product Purification by Reversible Phase Transition Following Escherichia coli Expression of Genes Encoding up to 251 Repeats of the Elastomeric Pentapeptide GVGVP. *Protein Expression Purif.* **1996,** *7* (1), 51-57.

10. Meyer, D. E.; Chilkoti, A., Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat. Biotechnol.* **1999,** *17* (11), 1112-1115.

11. Shu, X. K.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J., Novel chromophores and buried charges control color in mFruits. *Biochemistry* **2006,** *45* (32), 9639-9647.

12. Chung, J. A.; Wollack, J. W.; Hovlid, M. L.; Okesli, A.; Chen, Y.; Mueller, J. D.; Distefano, M. D.; Taton, T. A., Purification of prenylated proteins by affinity chromatography on cyclodextrin-modified agarose. *Anal. Biochem.* **2009,** *386* (1), 1-8.

13. Lee, B. C.; Zuckermann, R. N., Templated display of biomolecules and inorganic nanoparticles by metal ion-induced peptide nanofibers. *Chem. Commun.* **2010,** *46* (10), 1634-1636.

14. Lam, C. N. Interactions governing the self-assembly of globular protein-polymer block copolymers. Massachusetts Institute of Technology, Cambridge, Massachusetts, 2016.

15. Boville, C. E.; Romney, D. K.; Almhjell, P. J.; Sieben, M.; Arnold, F. H., Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from Thermotoga maritima. *J. Org. Chem.* **2018,** *83* (14), 7447-7452.

16. Shimazu, M.; Mulchandani, A.; Chen, W., Thermally triggered purification and immobilization of elastin–OPH fusions. *Biotechnol. Bioeng.* **2003,** *81* (1), 74-79.

17. Paloni, J. M.; Miller, E. A.; Sikes, H. D.; Olsen, B. D., Improved Ordering in Low Molecular Weight Protein–Polymer Conjugates Through Oligomerization of the Protein Block. *Biomacromolecules* **2018,** *19* (9), 3814-3824.

18. Sing, M. K. Structure-property relationships in physical polymer gels. Massachusetts Institute of Technology, Cambridge, Massachusetts, 2017.

Chapter 3 : Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins

3.1 Abstract

Self-assembly of protein-polymer bioconjugates presents an elegant strategy for controlling nanostructure and orientation of globular proteins in functional materials. Recent work has shown that genetic fusion of globular protein mCherry to an elastin-like polypeptide (ELP) yields similar self-assembly behavior to these protein-polymer bioconjugates. In the context of studying protein-polymer bioconjugate self-assembly, the flexibility of the ELP sequence allows several different properties of the ELP block to be tuned orthogonally while maintaining consistent polypeptide backbone chemistry. This work uses this ELP sequence flexbility in combination with the precise control offered by genetic engineering of an amino acid sequence to generate a library of four novel ELP sequences that are used to study the combined effect of charge and hydrophobicity on ELP-mCherry fusion protein self-assembly. Concentrated solution self-assembly is studied by small-angle X-ray scattering (SAXS) and depolarized light scattering (DPLS). These experiments show that fusions containing a negatively charged ELP block do not assemble at all, and fusions with a charge balanced ELP block exhibit a weak propensity for assembly. By comparison, the fusion containing an uncharged ELP block starts to order at 40 wt% in solution, and at all concentrations measured has sharper, more intense SAXS peaks than other fusion proteins. These experiments show that charge character of the ELP block is a stronger predictor of self-assembly behavior than the hydrophobicity of the ELP block. Dilute solution small-angle neutron scattering (SANS) on the ELPs alone suggests that all ELPs used in this study (including the uncharged ELP) adopt dilute solution conformations similar to those of traditional polymers, including polyampholytes and polyelectrolytes. Finally, dynamic light scattering studies on ELP/mCherry blends shows that there is no significant complexation between the charged ELPs and mCherry. Therefore, it is proposed that the superior self-assembly of fusion proteins containing uncharged ELP block is due to effective repulsions between charged and uncharged blocks due to local charge correlation effects and, in the case of anionic ELPs, repulsion between like charges within the ELP block.

3.2 Introduction

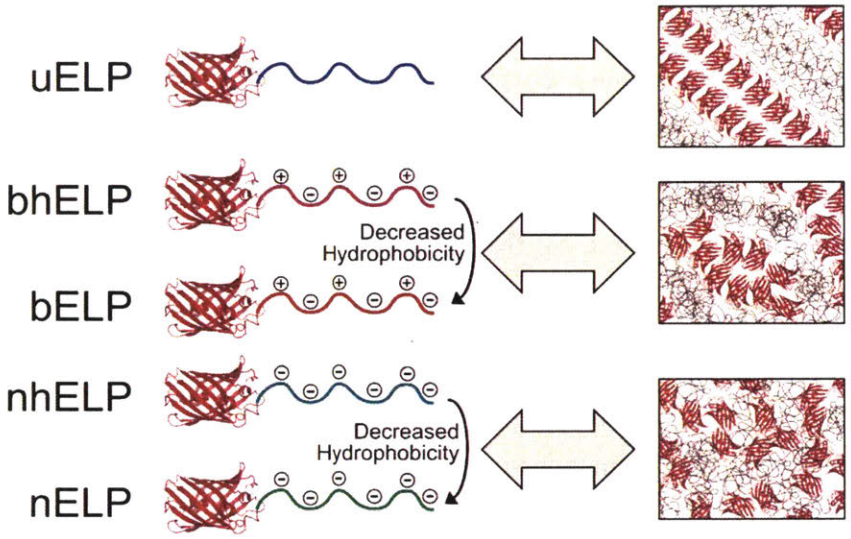
Globular proteins are capable of performing a wide variety of useful functions, including catalysis[1-4], sensing[5-6], and drug delivery[7-8]. The applications of globular proteins span many fields including transesterification processes for biodiesel production[1], monitoring of blood glucose levels[6], sensing and decontaminating organophosphates[9], and targeted delivery of cancer-treating drugs[7-8]. The ability of proteins to perform these functions under mild reaction conditions and aqueous environments makes them attractive candidates for the development of functional materials; however, there are several criteria that must be considered in the design of such materials. These include the preservation of the protein structure and function, the orientation of the protein in the material to allow access to the active site, and the ability of the material itself to transport substrates necessary for protein function[10-11].

The framework for nanostructure formation established in coil-coil block copolymer self-assembly provides an elegant strategy for addressing these material design criteria. It has been shown that protein-polymer conjugates are capable of assembling into many of the same phases formed by traditional coil-coil block copolymers[12-15]. These self-assembled bioconjugates have the potential to address the design concerns above, including substrate transport, control over protein orientation[16], and stabilization of protein structure and function[17-18]. In fact, recent work has shown that these bioconjugates can be used with enzymes to make both highly-active catalytic films[19] as well as with antibodies to make biosensors[20]. Previous work aimed at better understanding the phase behavior of these concentrated bioconjugate materials indicates that the most important factors governing their self-assembly are polymer chemistry[12, 21] and polymer volume fraction[14], whereas specifics of protein surface charge[22-23] and conjugation site[16] have relatively little impact. In addition, salt has only been found to have a significant impact on ordering when the conjugated polymer is zwitterionic, and in this case, addition of salt typically decreased ordering[21, 23]. However, the impact of different types of charge on the polymer block remains largely unexplored.

It has also been shown that nanoscale self-assembly of globular proteins can be achieved via genetic fusion of an elastin-like polypeptide (ELP) to red fluorescent protein mCherry[24-25]. Concentrated solution phase behavior of these ELP-mCherry fusions shares many similarities with the phase behavior of mCherry bioconjugates, and the topology of these fusion proteins had a large impact on the observed self-assembly[13, 24]. In addition to these concentrated solution self-assembly studies, there has been a large amount of work studying the dilute solution self-assembly of various

different proteins fused to ELPs, which have resulted in the observation of structures such as spherical and cylindrical micelles[26-27]. ELPs are polypeptides consisting of a pentapeptide repeat unit Xaa-Pro-Gly-Zaa-Gly (XPGZG), where Xaa can be either valine (V) or isoleucine (I), and Zaa can be any amino acid except proline[28]. The flexibility of the ELP sequence, in both the first (Xaa) and fourth (Zaa) positions allows for simultaneous adjustment of ELP hydrophobicity, charge, and functionality. This property of ELPs has been used in the literature to develop ELPs with a small fraction of crosslinkable lysine and cysteine residues[29-30], and to generate blocks of ELPs with different transition temperatures capable of undergoing multiple temperature induced transitions[26].

Here, we take advantage of the highly-flexible ELP sequence to systematically explore the impact of charge and hydrophobicity on the assembly of ELP-mCherry fusion proteins in concentrated solution. mCherry has been shown to have strong attractive interactions with polycations [31-32]. Due to the fact that attractive interactions have been shown to lead to poor self-assembly in block copolymers[33], this study focuses on ELP-mCherry fusions containing negatively charged and charge balanced ELPs. Using genetic engineering, we precisely control both the degree of charge and its spacing in the ELP block. The concentrated solution assembly as well as the dilute solution behavior of fusions with negatively charged and charge balanced ELP blocks with varying hydrophobicity are compared to determine the impact of both charge and hydrophobicity on ELP-mCherry fusion protein self-assembly.



uELP:[VPGVG IPGVG VPGVG IPGVG VPGVG]$_{20}$
bhELP:[IPGVG IPGEG IPGVG IPGKG IPGVG]$_{20}$
bELP:[VPGVG IPGEG VPGVG IPGKG VPGVG]$_{20}$
nhELP:[IPGVG IPGEG IPGVG IPGEG IPGVG]$_{20}$
nELP:[VPGVG IPGEG VPGVG IPGEG VPGVG]$_{20}$

Figure 3-1. Schematic illustration of different ELP-mCherry fusions including ELP amino acid sequences. Sequence design includes an uncharged ELP, two charge balanced ELPs, and two negatively charged ELPs. The charge balanced ELPs and the negatively charged ELPs all have the same net charge density. Within each set of charge type, there is a more hydrophobic and a less hydrophobic variant.

3.3 Materials & Methods

3.3.1 Biosynthesis of mCherry-ELP fusion proteins

Concatamerization of elastin-like polypeptide genes. Concatamerication of elastin-like polypeptide (ELP) genes was performed using previously described procedures[34-35]. Briefly, 458 bp ELP sequences (encoding for five repeats of the 25 amino acid long sequence shown in **Figure 3-1**) in were cloned into the pET28b plasmid between the 5' NcoI site and the 3' XhoI site. A 6xHis tag was included at the N-terminus of the ELP sequence for ease of purification. At the 5' end of each ELP gene are a BamHI restriction site and a BsaI restriction site in the forward orientation (cut site is downstream). At the 3' end of each ELP gene are a HindIII restriction site and a BsaI restriction site in the reverse orientation (cut site is upstream). To concatamerize the ELP genes, the vectors containing the genes were digested with BsaI, leaving TTGG overhangs

on both ends of the gene. In a separate reaction, the vectors were also digested with BsaI and then dephosphorylated with Antarctic phosphatase. Purified inserts and vectors were ligated with a high molar ratio of insert to vector (from 10:1 to 100:1 insert:vector). The ligations were transformed into chemically-competent NEB5α cells and clones were screened via analytical digestion with NcoI/HindIII. Successful clones were confirmed by sequencing. Because the BsaI cut sites exist outside of the ligation site, multiple rounds of concatamerization were possible. One round of concatamerization typically resulted in a doubling of the original ELP gene, so two rounds of concatamerization were used to elongate the ELP gene to the desired 1575 bp gene length. Verified DNA sequences and amino acid sequences for the full length ELPs can be found in **A2.1 Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins.**

Cloning of ELP-mCherry fusions. Fusion of the gene encoding mCherry to various ELP genes was also performed according to published protocols[24-25]. Briefly, an mCherry gene containing a NheI restriction site at the 5' end of the sequence and a HindIII restriction site at the 3' of the sequence was generated using polymerase chain reaction (PCR). ELP genes were designed to have an NheI/HindIII cloning site after the full ELP gene but before the stop codon to allow for addition of the fusion of the mCherry gene into the sequence. Both the PCR'd mCherry gene and the pET28b vector containing the ELP gene were digested with NheI/HindIII. The pET28b vector (again, containing the ELP gene upstream of the cut site) was then dephosphorylated with Antarctic phosphatase. Both genes were then purified using spin-column purification. Gel extraction was not necessary here because the undesirable fragments cut out in the digest were small enough to be eliminated via spin-column purification. Finally, the purified DNA was ligated at a 3:1 insert to vector molar ratio, where the insert here was the mCherry gene and the vector was the pET28b vector containing the ELP gene. Ligated plasmids were transformed into chemically-competent NEB5α cells under selection by kanamycin resistance (50 μg/mL). Clones were screened using BamHI/HindIII analytical digestion and confirmed by sequencing. Verified DNA sequences and amino acid sequences for the fusion proteins can be found in **A2.1 Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins.**

Protein expression. All expression vectors were transformed into *Escherichia coli* strain Tuner (DE3). All charged ELPs were expressed in Terrific Broth (TB, 1 L) in baffled shake flasks (2.5 L), and were grown to OD_{600} = 1.3-1.9 at 30 °C before induction. Expressions were then induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG), and the temperature was dropped to

20 °C. The uncharged ELP was expressed in 5 L TB in a fermenter with 7 L working volume at 30 °C for 18 h with no induction. All mCherry-ELP fusion proteins were expressed in Luria Broth (LB, 1 L) in baffled shake flasks (2.5 L), and were also grown to OD_{600} = 0.7-1.0 at 30 °C, followed by induction with 0.5 mM IPTG and a temperature drop to 20°C. All induced expressions were grown for 17-24 hours after induction (17-18 hours for charged ELPs, and until pink color characteristic of mCherry was observed for all ELP-mCherry fusion proteins), and cells from all expressions were harvested by centrifugation. Cells were resuspended in 100 mL lysis buffer for every 30 g of wet cell mass. Lysis buffer contained 50 mM sodium phosphate, 300 mM sodium chloride, and 10 mM imidazole. The pH of the lysis buffer was adjusted to pH 8 for all expressions except negatively charged ELPs. The pH of the lysis buffer used for negatively charged ELPs was adjusted to 10 to decrease complexation between the ELP and other positively charged proteins in the lysate. Resuspended cells were stored at -80 °C. A table of expression conditions and yields is provided in the Supplementary Information (**Table B-1**). mCherry alone was expressed and purified according to previously published protocols[31].

Protein purification. Resuspended cells for negatively charged ELPs were thawed and refrozen at -80 °C three times to aid in cell lysis. All other resuspended cells were thawed and incubated with 1 mg/mL lysozyme at 4 °C for 1-2 hours. Lysozyme was not added to negatively charged ELPs because the high amount of positive charge on lysozyme lead to aggregation with the negatively charged ELP, making it difficult to purify. All resuspended cells were then sonicated using a tip sonicator. The lysate was clarified by high-speed centrifugation at 4 °C. The uncharged ELP was purified by thermal cycling between 37 °C and 4 °C two times. All charged ELPs were purified by Ni-NTA affinity chromatography (pH 8 for charge balanced ELPs, pH 10 for negatively charged ELPs). All fusion proteins were purified using two cycles of ammonium sulfate precipitation, in which 1.5 M ammonium sulfate was used to precipitate out the fusion, and the resultant pellets (collected by centrifugation) were resuspended in lysis buffer and centrifuged again to remove insoluble contaminants.

All ELPs and fusions were subsequently purified using fast protein liquid chromatography (FPLC) under native conditions with an anion exchange column (HiTrap Q HP anion exchange chromatography column, GE Life Sciences). All proteins except the negatively charged ELPs and negatively charged ELP fusions were purified at pH 8. Negatively charged ELPs and negatively charged ELP fusions were purified at pH 10. All proteins except the negatively charged ELPs were

eluted over 30 column volumes using a linear gradient that increased salt concentration from 0 to 200 mM NaCl. Negatively charged ELPs were eluted over 30 column volumes using a linear gradient that increased salt concentration from 0 to 300 mM NaCl. Content and purity of ELP in different fractions was evaluated using SDS-PAGE that sampled every 3 wells over the entire gradient. For ELP-mCherry fusion proteins, only fractions that exhibited pink color characteristic of mCherry were run on SDS-PAGE. Elution fractions containing pure proteins of interest were collected and dialyzed into MilliQ purified water. Protein purity was confirmed by denaturing gel electrophoresis (SDS-PAGE). ELP samples were freeze dried after dialysis. Fusion protein samples were concentrated using ultrafiltration to 10-20 wt% solutions and then drop cast onto Teflon sheets. Samples were subsequently dried at room temperature under vacuum ramp (50 torr per hour down to 5 torr) for 12 hours. Summary of yields and purification conditions used for different proteins are listed in **Table B-1**, and SDS-PAGE of all different ELPs and fusion proteins are shown in **Figure B-1**.

3.3.2 Characterization

Small-angle X-ray scattering. Prepared solutions were used to fill 1 mm thick washers and sealed with Kapton tape. The fusion proteins were measured at Brookhaven National Lab at the National Synchrotron Light Source II (NSLS-II) on the 16-ID Life Science X-Ray Scattering (LiX) beamline. Samples were equilibrated at 10 °C for 20 min and 10 min at all other temperatures prior to data collection. SAXS data were collected and corrected for empty cell scattering. Acquisition times were selected such that damage to the samples' nanostructure by the beam was not detectable. Background correction was applied on 2D scattering patterns using the following equation

$$I(q) = \frac{I_{\text{sample}}}{T_{\text{sample}}} - \frac{t_{\text{sample}}}{t_{\text{empty}}}\frac{I_{\text{empty}}}{T_{\text{empty}}} \quad (1)$$

Where I, t, and T are intensity, acquisition time, and transmission, respectively. Subscripts sample and empty refer to either the sample or the empty cell scattering or transmission. After background correction, data was converted from 2D to 1D by taking an azimuthal average of 2D data.

Small-angle neutron scattering. All samples were prepared from solid samples at 25 mg/mL in deuterated water and dissolved overnight at 4 °C. After complete dissolution, all samples were passed through a 0.2 μm syringe filter before loading into 1 mm banjo cells. Measurements were performed at the Extended Q-Range Small-Angle Neutron Scattering Diffractometer (BL-6) at the Spallation Neutron Source (SNS) at Oak Ridge National Laboratories. Three detector

configurations were used—two configurations at a 4 m detector distance for collecting 12 and 2.5 Å wavelength neutrons, and one configuration at 1.3 m detector distance, also for 2.5 Å wavelength neutrons. This set of configurations allowed a q-range of 0.0042-0.76 nm^{-1}. All measurements were taken on samples held at 10 °C. 2D data was corrected for both empty cell and dark field scattering and azimuthally averaged to produce 1D curves.

Turbidimetry for transition temperature determination. The transition temperatures of the ELPs were determined as previously described in literature[28]. Briefly, lyophilized powders were dissolved at 1 mg/mL in water or 20 mM pH 8 tris overnight at 4 °C. Prior to measurement, samples were filtered through a 0.2 μm syringe filter. Samples were loaded into a quartz cuvette and sealed. The temperature of the cuvette was ramped up at a rate of 1 °C/min, and absorbance at 600 nm was monitored. Every sample went through two heating/cooling cycles, and the data from the second cycle was taken to eliminate contributions from irreversible aggregates to the turbidity.

Birefringence measurements. Turbidimetry and depolarized light scattering[36-37] were performed on concentrated solution samples to measure sample birefringence. Samples were loaded into 1 mm thick Teflon molds and sealed between two quartz disks. Measurements were performed with a coherent OBIS LX660 laser (λ = 662 nm, beyond the absorption cutoff of mCherry) with a continuous wave power output of 10 mW. All samples were equilibrated at 10 °C for 10 minutes and then heated at 1 °C min^{-1} to 40 °C, allowed to equilibrate for 10 minutes at 40 °C, and then cooled at 1 °C min^{-1} back down to 10 °C. Static birefringence was measured by passing the laser through a polarizer before the sample, and then a second polarizer, perpendicular in phase to the first polarizer, after the sample. Measured light polarization was corrected for transmission and dark field background. Transmission was measured in an identical setup to birefringence, but without a second polarizer after the sample. Transitions were reproducible on repeated cycling.

Turbidimetry to measure complex coacervation. 2 mg/mL solutions of mCherry (diluted from solution) and ELP (prepared by dissolving solid lyophilized powder) were mixed at ratios spanning from 100% ELP to 100% mCherry in 20 mM pH 8 tris buffer. Transmittance of the blends was measured in a 96-well plate at 700 nm. Error bars on these measurements are standard deviation from triplicate measurements performed at each weight ratio.

Dynamic light scattering. All solutions measured by dynamic light scattering (DLS) were prepared at 5 mg/mL in water. All single-component solutions were filtered through a 0.45 μm filter after preparation. Blended solutions were prepared by mixing filtered solutions together at a

ratio of 63:37 ELP:mCherry (representative of the approximate fraction of ELP and mCherry in the fusion proteins). Each measurement was averaged over ten 5 second acquisitions.

3.4 Results & Discussion

In this work, the effect of ELP charge and hydrophobicity on concentrated solution ELP-mCherry fusion self-assembly is explored using five different ELPs. The sequences for these ELPs and the abbreviations used to describe them in this paper are shown in **Figure 3-1**. The uELP sequence is the same uncharged ELP sequence that was found to lead to self-assembly of these fusion proteins previously.[25] nELP and nhELP are both negatively charged ELPs, and bELP and bhELP have an equivalent number of positive and negative charges in the repeat sequence, and are thus titled "charge balanced." nhELP and bhELP have increased hydrophobicity over their equivalently charged counterparts via changing of the first position residue in all pentapeptide repeat units to isoleucine (I). Self-assembly of mCherry fused to these different ELPs was studied using a combination of small-angle x-ray scattering (SAXS) and depolarized light scattering (DPLS). The better understand the interactions underpinning the concentrated solution self-assembly of the fusion proteins, dilute solution characterizations, including dynamic light scattering (DLS) and small-angle neutron scattering (SANS), were also performed on the ELPs alone as well as blends of ELP and mCherry.

3.4.1 Phase Behavior in Concentrated Solution

Neither fusion protein containing a negatively charged ELP block fused to mCherry shows the ability to form ordered nanostructures. **Figure 3-2** shows representative SAXS curves of nELP-mCherry and nhELP-mCherry at 10 °C at all concentrations measured for each sample. As shown by this representative data, a weak primary scattering peak begins to appear in the nhELP-mCherry data at 55 wt% and by 60 wt%, both nELP-mCherry and nhELP-mCherry have a broad primary peak; however, in both cases, the observed peak is broad and low intensity. Birefringence measurements reveal little to no birefringence in either of these fusion proteins at any of the concentrations measured (**Figure B-16, Figure B-17**).

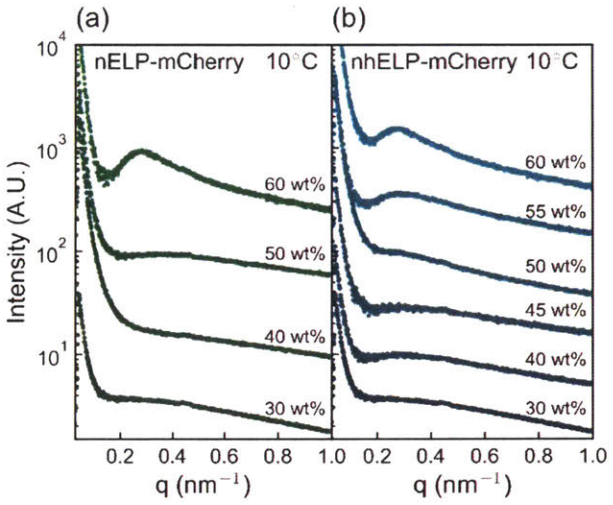


Figure 3-2. Representative SAXS curves of (a) nELP-mCherry and (b) nhELP-mCherry at 10 °C as a function of concentration in water. Curves are offset for clarity. SAXS data including all concentrations measured is available in **Figure B-6** and **Figure B-7**.

The two fusion proteins containing a charge-balanced ELP block fused to mCherry exhibit weak microphase separation in their concentrated solution phase behavior. **Figure 3-3** shows representative SAXS data of bELP-mCherry and bhELP-mCherry at 10 °C at various concentrations in water. In both variants, a primary scattering peak begins to appear around 45 wt%. In bhELP-mCherry, the intensity of this peak increases with increasing concentration, whereas the intensity of this peak remains about the same between 45 and 60 wt% for bELP-mCherry. This indicates that, at higher concentrations, some degree of microphase separation is occurring in both materials. The more hydrophobic variant (bhELP-mCherry) appears to order slightly better than the less hydrophobic variant at higher temperatures (**Figure B-4, Figure B-5**), evidenced by the sharper primary peaks in scattering at 50 wt% and above at these temperatures. Furthermore, at 60 wt%, bhELP-mCherry exhibits a secondary peak that indexes to $2q^*$, suggesting that this material is forming a weakly lamellar phase.

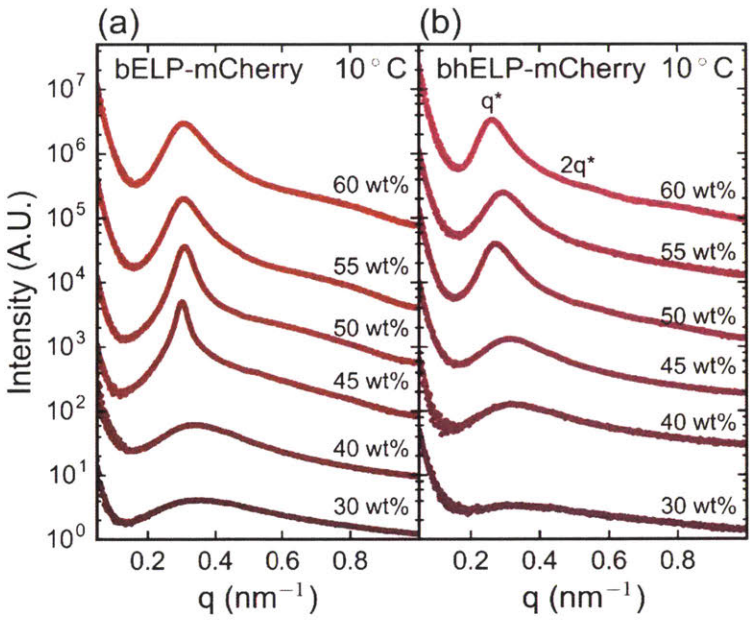


Figure 3-3. SAXS curves of (a) bELP-mCherry and (b) bhELP-mCherry at 10 °C various concentrations in water. Curves are offset for clarity. SAXS data including all concentrations measured is available in **Figure B-4** and **Figure B-5**.

Comparison of SAXS data for concentrated solutions of the different fusion proteins reveals that fusions with an uncharged ELP block assemble into the best ordered structures, followed by charge balanced fusions, followed by negatively charged fusions, and that the hydrophobicity of the fusion is a secondary effect relative to the charge state. **Figure 3-4a** shows representative SAXS curves of the different fusion proteins at 60 wt% in water and 10 °C. uELP-mCherry has the sharpest peaks of all the fusions studied, with peaks that index to q^* and $2q^*$, which is indicative of a lamellar phase. The two fusions with charge-balanced blocks have a broader, less intense primary peak, and bhELP-mCherry shows a weak secondary peak indexing to $2q^*$, again indicating a lamellar phase. The primary peak locations in the uELP-mCherry and bhELP-mCherry patterns correspond to lamellar domain spacings of 26.9 nm and 24.2 nm respectively. Finally, the fusions containing negatively-charged ELP blocks have much weaker, broader peaks than the other fusions, indicating that there is little to no microphase separation between the ELP and mCherry blocks in these variants. Data shown at 60 wt% mirrors trends observed at other concentrations by SAXS—uELP-mCherry exhibits secondary peaks in SAXS data starting at 40 wt%. Below 60

wt%, neither bhELP-mCherry nor bELP-mCherry have any secondary peaks by SAXS, but both have a primary peak starting at 45 wt%. And neither fusion containing an anionic ELP block exhibits any indication of microphase separation at the concentrations explored here (30-60 wt%). SAXS data taken on fusions between 10 °C and 40 °C did not show a strong dependence on temperature for any of the fusions except uELP-mCherry and bhELP-mCherry at 30 wt% (**Figure B-3**, **Figure B-4**, **Figure B-5**, **Figure B-6**, and **Figure B-7**). This temperature dependence is likely due to the thermoresponsive behavior of the respective ELPs (**Table 3-1**). In the case of bhELP-mCherry, we hypothesize that the temperature response is eliminated at higher concentrations due to an increase in the local charge concentration around the ELP chains at higher concentrations. This is consistent with results presented below showing that small increases in the ionic strength of the solution dramatically increase the transition temperature of bhELP, but not uELP (**Table 3-1**).

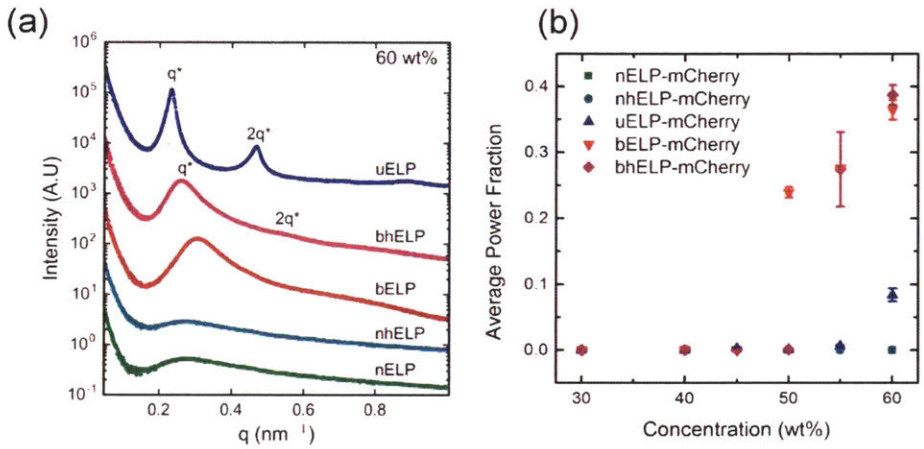


Figure 3-4. Representative (a) SAXS and (b) birefringence data on ELP-mCherry fusions. (a) Representative SAXS data of different ELP-mCherry fusion proteins at 10 °C and 60 wt% in water. Curves are offset for clarity. A full SAXS dataset for uELP-mCherry can be found in **Figure B-3**. (b) Average birefringence of different ELP-mCherry fusion proteins at different concentrations in water. Note that data points for nELP-mCherry are not visible due to overlap with values for nhELP-mCherry at all concentrations measured.

While the SAXS data indicates uELP-mCherry assembles into the most well-defined periodic structures of all the fusions studied (consistent with previous SAXS data taken on this same fusion protein[25]), fusions containing a charge-balanced ELP block exhibit much stronger birefringence. uELP-mCherry birefringence ranges from power fractions of ~0.001 to ~0.1 above background in

samples that also show ordering by SAXS (**Figure 3-4b**). By comparison, power fractions between 0.2 and 0.4 were observed in birefringence measurements on bhELP-mCherry and bELP-mCherry in samples that showed no periodic ordering by SAXS. It is possible that the fusions containing charge-balanced ELPs form a liquid crystalline phase that is highly birefringent, but does not have a strong propensity to form periodic structures that would produce secondary scattering peaks in SAXS. Polarized optical microscopy images (not included) on bELP-mCherry and bhELP-mCherry samples exhibiting high birefringence did not show clearly identifiable liquid crystalline textures, likely due to small grain sizes. Thus, we hypothesize that the charge-balanced blocks introduce some degree of molecular alignment that strongly polarizes light but does not result in strong periodic ordering.

3.4.2 Interactions Between the ELP and mCherry Blocks

The concentrated solution self-assembly results presented above suggest that there is a large difference in the interactions between mCherry and the different ELPs. The broad range over which self-assembly is observed in uELP-mCherry suggests that the effective ELP/mCherry repulsion is strongest for uELP. The absence of self-assembly (within the experimental range presented above) in fusions containing a negatively-charged ELP block suggests that the ELP/mCherry repulsion is weakest in negatively charged ELPs. To understand how sequence translates into a difference in protein-polymer interactions, electrostatic attractions, sequence hydrophobicity, and chain conformation in solution are systematically investigated.

Interactions of proteins with charged polymers is a well-established phenomenon in the literature [38], often manifest in macrophase separation of protein/polymer aggregates termed complex coacervation. It was recently established that the propensity of a globular protein to form a complex coacervate with a polycation could be predicted by the ratio of negatively charged to positively charged amino acids in the globular protein.[32] Since mCherry is negatively charged, it is shown to coacervate primarily with cationic polymers.[31] For this reason, anionic and charge balanced ELPs were specifically selected in this study. Turbidimetry experiments on ELP/mCherry blends below the ELP thermal transition temperature (**Figure B-19**) confirm that mCherry does not coacervate with any of the ELPs in this study. However, this does not exclude the possibility of soluble complex formation between mCherry and the charged ELPs. [38]

Dynamic light scattering (DLS) measurements on mCherry/ELP blends indicate that soluble complexes do not form between mCherry and any of the ELPs (**Figure 3-5**). CONTIN fits (as

performed by Wyatt's Dynamics® software) of DLS measurements on uELP, bELP, bhELP, and mCherry confirm the existence of a single species in solution (**Figure B-20**). For the ELPs, the predicted hydrodynamic radius (R_H), approximated by applying the Stokes-Einstein relation to diffusivity measured by DLS, is 5-6 nm, consistent with the radius of gyration measured by SANS. Measurements on mCherry yield an R_H of 2.56 ± 0.05 nm, which is consistent with literature reported values.[39] Negatively charged ELPs exhibit behavior consistent with the literature reports on semidilute polyelectrolyte solutions, where a single polyelectrolyte produces both a fast and a slow diffusive mode in DLS. [40-41] The two diffusion coefficients observed in these experiments--D_{fast}~10^{-6} cm^2 s^{-1} and D_{slow}~10^{-8} cm^2 s^{-1}—are consistent with literature values on other semidilute polyelectrolyte solutions.[42] Blends containing uncharged or charge balanced ELPs produce autocorrelation curves (**Figure 3-5**) that fit well to a weighted average of the individual components. The fraction attributed to each component agrees well with the actual fraction in solution, indicating that the ELP and mCherry are diffusing without interaction in these blends. While autocorrelation curves for blends containing negatively charged ELPs do not fit well to a weighted average of individual components, indicating that the electrostatic effects on ELP diffusivity are modified by the presence of the charged protein, the timescale on which the autocorrelation decays occur in the blend are between those of the individual ELPs and mCherry. This also suggests that mCherry/ELP complexes are not present. Therefore, it is unlikely that charge-based association between mCherry and ELPs plays a significant role in the observed self-assembly behavior.

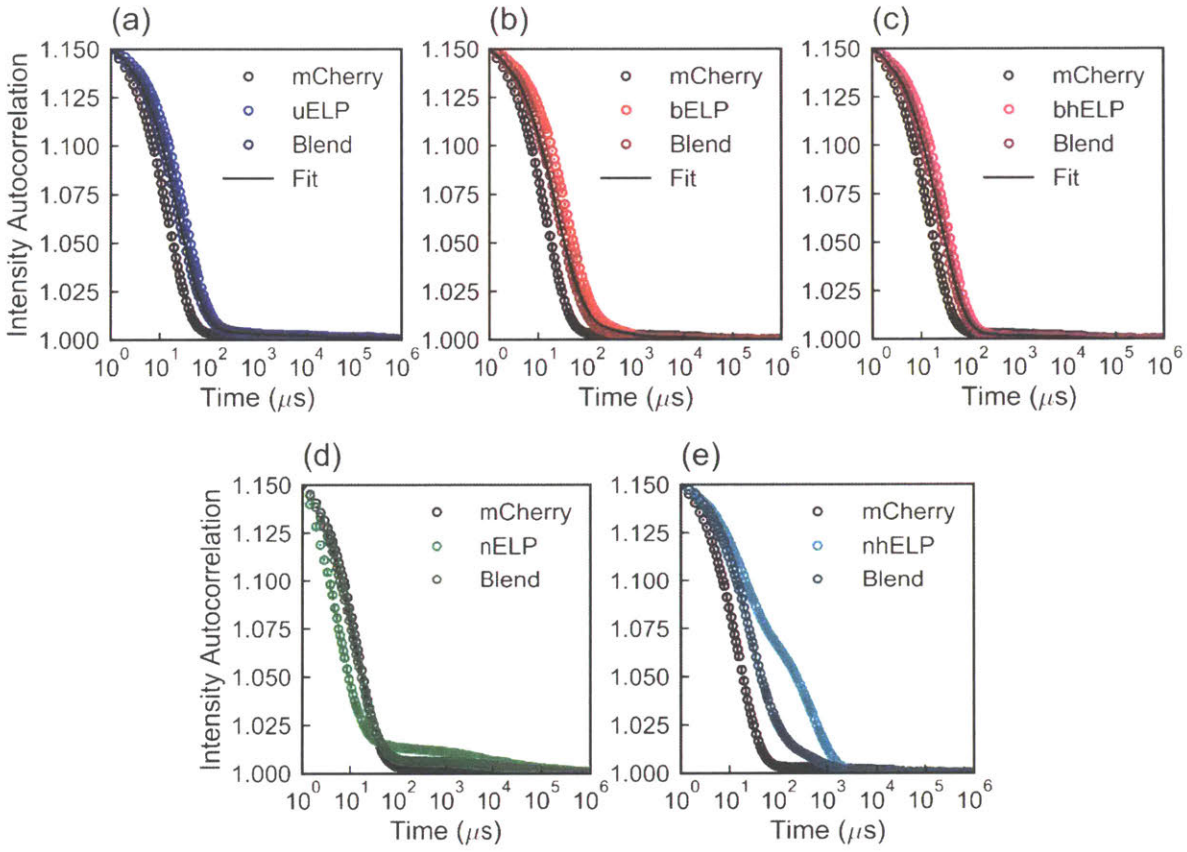


Figure 3-5. Dynamic light scattering (DLS) autocorrelation curves for (a) uELP, (b) bELP, (c) bhELP, (d) nELP, (e) nhELP, mCherry, and respective ELP/mCherry blends. Error bars represent standard deviation over measurements on 3 different samples. Fit curves represent fits performed on blend data that was fit to a weighted average of the individual component autocorrelation curves.

Table 3-1. Measured ELP transition temperatures

	Water		20 mM pH 8 Tris	
	Heating	Cooling	Heating	Cooling
uELP	30.2 °C	24.9 °C	30.9 °C	26.4 °C
bELP	75.5 °C	71.6 °C	> 90 °C	
bhELP	41.5 °C	36.3 °C	70-90 °C*	
nELP	> 90 °C		> 90 °C	
nhELP	50-90 °C*		> 90 °C	

*Indicates that ~10-40% drop in transmittance observed over this temperature range

The addition of charge to ELP sequences also has a large impact on hydrophobicity, which may play a role in determining the propensity for self-assembly. Dilute solution thermal transition temperatures of the ELPs used in this study provide a means for assessing the hydrophobicity of the different ELPs. In pure water, the negatively charged ELP shows no transition within the experimental range, while nhELP shows a broad transition between 50 and 90 °C that does not drop below 50% transmittance (**Figure B-18**). However, in 20 mM buffer, only the uncharged ELP shows a thermal transition, indicating that all of the charged ELPs have strong salting-in behavior, consistent with the salting-in behavior of proteins.[43-44] While the transition temperatures of the ELPs in pure water suggest that the hydrophobic versions of both charge balanced and anionic ELPs are substantially more hydrophobic than their similarly charged partners without hydrophobic amino acid substitutions, the self-assembly behavior shows only small differences. This is consistent with the salting in behavior observed, as the ionic strength in the highly concentrated self-assembled structures should be much higher than in dilute aqueous solution due to the high concentration of proteins and counterions. Therefore, modifications in chain hydrophobicity also cannot fully explain the self-assembly behavior.

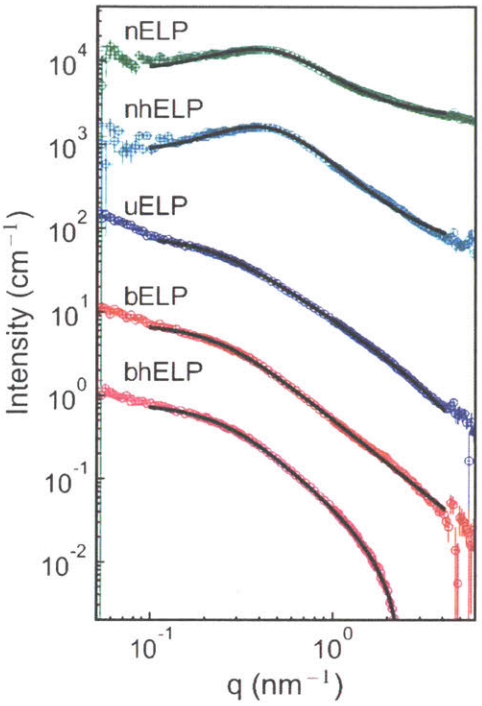


Figure 3-6. SANS data taken on bhELP, bELP, uELP, nhELP, and nELP at 25 mg/mL in deuterium oxide at 10 °C. For uELP, bELP, and bhELP, black lines represent best fits of data to a model for swollen/collapsed Gaussian polymers[45]. For nELP and nhELP, black lines represent best fits of data to a model for weak polyelectrolytes in dilute solution[46].

Finally, SANS data on the different ELPs in dilute solution shows that they adopt conformations similar to those of analogous synthetic polymers. Fits of the excluded volume polymer model to SANS data on uELP, bELP, and bhELP show that all three ELPs adopt a swollen or ideal Gaussian conformation in solution. Data for these ELPs were fit to a model for swollen/collapsed Gaussian polymers, as described by Hammouda[45] (see **B1.1 Models for Small-Angle Neutron Scattering** for details). The data and fits are shown in **Figure 3-6** and the final fit parameters are given in **Table 3-2**. These fits give a similar radius of gyration for all three ELPs—5.9 ± 0.1 nm for uELP, 6.0 ± 0.2 nm for bELP, and 5.8 ± 0.1 nm for bhELP. In contrast, the Porod exponents, *m*, are markedly different between the charge balanced ELPs and the uncharged ELP. The Porod exponent of uELP is 1.49 ± 0.01, corresponding to a fully swollen chain, and the Porod exponents of bELP and bhELP are 1.78 ± 0.02 and 1.87 ± 0.02, which both lie between the Porod exponents for a fully swollen chain and an ideal Gaussian coil. This result suggests that the uncharged ELP

is better solvated by water than the charge balanced ELPs, due to some combination of self-repulsion within the uELP chain and favorable interactions within the charge balanced ELP chains. The latter hypothesis is supported by the observation of smaller domain spacing in the lamellar phase formed by bhELP-mCherry at 60 wt%, as compared to the domain spacings in uELP-mCherry at the same concentration (24.2 nm vs 26.9 nm, respectively). Interestingly, the degree of ELP chain solvation suggested by this fit (uELP > bELP > bhELP) is markedly different from the hydrophobicity predicted by the transition temperature measurements of these same ELPs in dilute solution (uELP > bhELP > bELP). It should be noted that the fact that SANS data on the charge balanced ELPs fits well to a swollen/collapsed Gaussian coil agrees with previous theoretical predictions of polyampholyte behavior in solution, where, at low charge density and balanced positive and negative charges, the polyampholyte is expected to adopt a Gaussian coil formation[47].

Table 3-2. Results of SANS fit to different ELPs in dilute solution

ELP	Contrast Factor C_1, cm^{-1}	Radius of Gyration R_g, nm	Porod Exponent m	Background C_2, cm^{-1}		Reduced χ^2 of Fit
uELP	0.82 ± 0.02	5.9 ± 0.1	1.49 ± 0.01	-0.0041 ± 0.0003		1.3
bELP	0.73 ± 0.02	6.0 ± 0.2	1.78 ± 0.02	0 ± 0.02		3.7
bhELP	0.81 ± 0.02	5.8 ± 0.1	1.87 ± 0.02	-0.0108 ± 0.0002		1.7
	Contrast Factor C_1, cm^{-1}	**Segment Length a, nm**	**Linear Charge Density Parameter m**	**Flory Parameter χ**	**Background C_2, cm^{-1}**	**Reduced χ^2 of Fit**
nELP	0.32 ± 0.06	3.2 ± 0.7	0.5 ± 0.2	0.45 ± 0.04	0.0235 ± 0.0005	9.7
nhELP	0.35 ± 0.05	3.0 ± 0.4	0.5 ± 0.1	0.48 ± 0.02	0.0059 ± 0.0004	6.4

*Reported errors are 95% confidence intervals on fitted parameters

SANS data on negatively charged ELPs was fit to the scattering model proposed by Borue and Erukhimovich for weakly charged polyelectrolytes in the semidilute regime[46]. The fits of SANS data taken on negatively charged ELPs fits well to the polyelectrolyte model posed (see **B1.1 Models for Small-Angle Neutron Scattering** for details), indicating that these ELPs adopt a

conformation similar to that of traditional weakly charged polyelectrolytes in solution. **Figure 3-6** shows the fits, and **Table 3-2** shows the corresponding fit parameters. The Flory parameter, χ, for both fits is close to 0.5, indicating that these ELPs adopt a slightly swollen conformation in solution[48]. Of greater interest is the linear charge density parameter, m, and segment length a. In a traditional polymer system, the segment length, a, would typically be set to the root mean-square distance between monomers, and (m-1) would represent the number of uncharged monomers between each charged monomer; however, because of the nature of the ELP system, a single monomer cannot be well-defined *a priori*. Instead, we can use the results of this fit to draw some conclusions about the nature of this system in relation to typical polymers. Assuming both ELPs have a charge of -40 at pH 7, and that the amino acids making up a single "monomer" unit in the charged ELP chain adopt a Gaussian conformation, the fit values for m can be used to predict a segment length in the ELPs (see **B1.2 Estimation of monomer-monomer spacing from linear charge density parameter** for details). This estimation based on m yields a segment length of 1.6 nm for both anionic ELPs, which is smaller than those from the fit (3.2 nm and 3.0 nm for nELP and nhELP, respectively). This difference in size can be explained by two main factors. Firstly, charging a polymer backbone is well-known to lead to extended chain conformations, especially in low salt conditions[49-50]; because there are 2-3 charged residues per "monomer," the larger observed segment length is not surprising. Secondly, based on the SANS fit of uELP, it is possible that the amino acids within each "monomer" unit are adopting a swollen conformation, which would also increase the average size of each unit. Thus, we conclude that these negatively charged ELPs adopt conformations in solution that are analogous to those of traditional weakly charged polyelectrolytes, but with monomer units that are composed of ~25 amino acids.

Ultimately, the combination of different characterizations presented above reveal some key differences between the charged ELPs introduced in this study and the uncharged ELP studied previously. Firstly, while charge balanced ELPs adopt a conformation similar to uELP in dilute solutions as measured by SANS, the combination of SANS and transition temperature measurements do not indicate a hydrophobicity/hydration trend that explains the concentrated solution phase behavior of the corresponding fusion proteins. Furthermore, dilute solution measurements on negatively charged ELPs indicate that these proteins adopt conformations similar to those of weakly charged polyelectrolytes in solution. This is further evidence indicating that while these ELPs have low charge density, they still show behavior consistent with charged

polymers in dilute solution. Finally, dilute solution turbidimetry and DLS on blends of the different ELPs and mCherry show that there are no strong attractions between these two molecules in dilute solution, suggesting that the poor ordering observed in fusions containing a charged ELP block is not due to attractive interactions between the charged ELPs and mCherry. This combination of results suggests that the concentrated solution phase behavior trends can be explained by a combination of two factors. The first of these arises from the discovery that, in block copolymers with one charged block and one uncharged block, repulsion between blocks is enhanced due to a combination of counterion solubility and electrostatic cohesion[51]. Thus, because the globular protein, mCherry, is charged, once charge is added to the ELP block, both blocks of the fusion protein become charged, decreasing the effective repulsion between blocks. In the case of the fusions containing negatively charged ELPs, there is the added effect of repulsions between the like-charged anionic ELP chains. This combination of considerations would explain why charge balanced ELPs exhibit poorer assembly than the uncharged ELP and why negatively charged ELPs do not assemble at all.

3.5 Conclusions

The concentrated solution phase behavior of five different ELP-mCherry fusions was studied to elucidate the effects of ELP charge and hydrophobicity on fusion protein self-assembly. The fusion with an uncharged ELP block gave the best periodic nanoscale ordering, and fusions with a charge balanced ELP block, while having little to no apparent long-range order, were highly birefringent at higher concentrations. Fusions containing a negatively charged ELP block showed little to no microphase separation and no birefringence. For each level of charge, sequences with and without hydrophobic amino acid substitutions showed little difference in self-assembly behavior. Dilute solution SANS of the individual ELPs suggests that the charge balanced ELPs used in this study behave similarly to traditional polyampholytes, and that the negatively charged ELPs behave similarly to traditional weakly charged polyelectrolytes. Dilute solution characterization of ELP/mCherry blends shows that there is no significant attraction between the ELPs and mCherry that would lead to coacervation. Thus, we conclude that the poor assembly of ELP-mCherry fusion proteins containing a charged ELP block is a consequence of two main factors. First, that effective repulsion that exists between charged and uncharged polymer blocks is eliminated once both blocks are charged. Second, that anionic-anionic repulsions between ELP chains promotes mixing between the ELP and mCherry blocks (for the negatively charged ELPs). This combination of

conclusions suggests that electrostatic cohesion effects play a key role in the self-assembly of protein-polymer block copolymer materials.

3.6 References

1. Bisen, P. S.; Sanodiya, B. S.; Thakur, G. S.; Baghel, R. K.; Prasad, G. B. K. S., Biodiesel production with special emphasis on lipase-catalyzed transesterification. *Biotechnol. Lett.* **2010,** *32* (8), 1019-1030. DOI: 10.1007/s10529-010-0275-z

2. Hari Krishna, S., Developments and trends in enzyme catalysis in nonconventional media. *Biotechnol. Adv.* **2002,** *20* (3), 239-267. DOI: 10.1016/S0734-9750(02)00019-8

3. Kirk, O.; Borchert, T. V.; Fuglsang, C. C., Industrial enzyme applications. *Curr. Opin. Biotechnol.* **2002,** *13* (4), 345-351. DOI: 10.1016/S0958-1669(02)00328-2

4. Schoemaker, H. E.; Mink, D.; Wubbolts, M. G., Dispelling the Myths--Biocatalysis in Industrial Synthesis. *Science* **2003,** *299* (5613), 1694-1697. DOI: 10.1126/science.1079237

5. Byrne, B.; Stack, E.; Gilmartin, N.; O'Kennedy, R., Antibody-based sensors: principles, problems and potential for detection of pathogens and associated toxins. *Sensors* **2009,** *9* (6), 4407-4445. DOI: 10.3390/s90604407

6. Rapp, B. E.; Gruhl, F. J.; Länge, K., Biosensors with label-free detection designed for diagnostic applications. *Anal. Bioanal. Chem.* **2010,** *398* (6), 2403-2412. DOI: 10.1007/s00216-010-3906-2

7. Chari, R. V. J., Targeted Cancer Therapy: Conferring Specificity to Cytotoxic Drugs. *Acc. Chem. Res.* **2008,** *41* (1), 98-107. DOI: 10.1021/ar700108g

8. Sliwkowski, M. X.; Mellman, I., Antibody Therapeutics in Cancer. *Science* **2013,** *341* (6151), 1192-1198. DOI: 10.1126/science.1241145

9. Efremenko, E. N.; Sergeeva, V. S., Organophosphate hydrolase — an enzyme catalyzing degradation of phosphorus-containing toxins and pesticides. *Russ. Chem. Bull.* **2001,** *50* (10), 1826-1832. DOI: 10.1023/a:1014377912147

10. Rotticci, D.; Norin, T.; Hult, K., Mass Transport Limitations Reduce the Effective Stereospecificity in Enzyme-Catalyzed Kinetic Resolution. *Org. Lett.* **2000,** *2* (10), 1373-1376. DOI: 10.1021/ol005639m

11. Lee, S.; Ringstrand, B. S.; Stone, D. A.; Firestone, M. A., Electrochemical activity of glucose oxidase on a poly(ionic liquid)-Au nanoparticle composite. *ACS Appl. Mater. Interfaces* **2012,** *4* (5), 2311-7. DOI: 10.1021/am300629n

12. Chang, D.; Lam, C. N.; Tang, S.; Olsen, B. D., Effect of polymer chemistry on globular protein-polymer block copolymer self-assembly. *Polym. Chem.* **2014,** *5* (17), 4884-4895. DOI: 10.1039/C4PY00448E

13. Lam, C. N.; Olsen, B. D., Phase transitions in concentrated solution self-assembly of globular protein-polymer block copolymers. *Soft Matter* **2013,** *9* (8), 2393-2402. DOI: 10.1039/C2SM27459K

14. Thomas, C. S.; Olsen, B. D., Coil fraction-dependent phase behaviour of a model globular protein-polymer diblock copolymer. *Soft Matter* **2014,** *10* (17), 3093-3102. DOI: 10.1039/C3SM52531G

15. Thomas, C. S.; Xu, L.; Olsen, B. D., Kinetically Controlled Nanostructure Formation in Self-Assembled Globular Protein–Polymer Diblock Copolymers. *Biomacromolecules* **2012,** *13* (9), 2781-2792. DOI: 10.1021/bm300763x

16. Huang, A.; Olsen, B. D., Self-Assembly of Differently Shaped Protein–Polymer Conjugates through Modification of the Bioconjugation Site. *Macromol. Rapid Commun.* **2016,** *37* (15), 1268-1274. DOI: 10.1002/marc.201500744

17. Holm, L. S.; McUmber, A.; Rasmussen, J. E.; Obiols-Rabasa, M.; Thulstrup, P. W.; Kasimova, M. R.; Randolph, T. W.; van de Weert, M., The Effect of Protein PEGylation on Physical Stability in Liquid Formulation. *J. Pharm. Sci.* **2014,** *103* (10), 3043-3054. DOI: 10.1002/jps.24094

18. Shakya, A. K.; Sami, H.; Srivastava, A.; Kumar, A., Stability of responsive polymer–protein bioconjugates. *Prog. Polym. Sci.* **2010,** *35* (4), 459-486. DOI: 10.1016/j.progpolymsci.2010.01.003

19. Huang, A.; Qin, G.; Olsen, B. D., Highly Active Biocatalytic Coatings from Protein–Polymer Diblock Copolymers. *ACS Appl. Mater. Interfaces* **2015,** *7* (27), 14660-14669. DOI: 10.1021/acsami.5b01884

20. Dong, X.-H.; Obermeyer, A. C.; Olsen, B. D., Three-Dimensional Ordered Antibody Arrays Through Self-Assembly of Antibody–Polymer Conjugates. *Angew. Chem. Int. Ed.* **2017,** *56* (5), 1273-1277. DOI: 10.1002/anie.201607085

21. Chang, D.; Olsen, B. D., Self-assembly of protein-zwitterionic polymer bioconjugates into nanostructured materials. *Polym. Chem.* **2016,** *7* (13), 2410-2418. DOI: 10.1039/C5PY01894C

22. Lam, C. N.; Kim, M.; Thomas, C. S.; Chang, D.; Sanoja, G. E.; Okwara, C. U.; Olsen, B. D., The Nature of Protein Interactions Governing Globular Protein–Polymer Block Copolymer Self-Assembly. *Biomacromolecules* **2014,** *15* (4), 1248-1258. DOI: 10.1021/bm401817p

23. Lam, C. N.; Yao, H.; Olsen, B. D., The Effect of Protein Electrostatic Interactions on Globular Protein–Polymer Block Copolymer Self-Assembly. *Biomacromolecules* **2016,** *17* (9), 2820-2829. DOI: 10.1021/acs.biomac.6b00522

24. Qin, G.; Glassman, M. J.; Lam, C. N.; Chang, D.; Schaible, E.; Hexemer, A.; Olsen, B. D., Topological Effects on Globular Protein-ELP Fusion Block Copolymer Self-Assembly. *Adv. Funct. Mater.* **2015,** *25* (5), 729-738. DOI: 10.1002/adfm.201403453

25. Qin, G.; Perez, P. M.; Mills, C. E.; Olsen, B. D., Effect of ELP Sequence and Fusion Protein Design on Concentrated Solution Self-Assembly. *Biomacromolecules* **2016,** *17* (3), 928-934. DOI: 10.1021/acs.biomac.5b01604

26. Hassouneh, W.; Fischer, K.; MacEwan, S. R.; Branscheid, R.; Fu, C. L.; Liu, R.; Schmidt, M.; Chilkoti, A., Unexpected Multivalent Display of Proteins by Temperature Triggered Self-Assembly of Elastin-like Polypeptide Block Copolymers. *Biomacromolecules* **2012,** *13* (5), 1598-1605. DOI: 10.1021/bm300321n

27. Weitzhandler, I.; Dzuricky, M.; Hoffmann, I.; Garcia Quiroz, F.; Gradzielski, M.; Chilkoti, A., Micellar Self-Assembly of Recombinant Resilin-/Elastin-Like Block Copolypeptides. *Biomacromolecules* **2017,** *18* (8), 2419-2426. DOI: 10.1021/acs.biomac.7b00589

28. Urry, D. W.; Gowda, D. C.; Parker, T. M.; Luan, C.-H.; Reid, M. C.; Harris, C. M.; Pattanaik, A.; Harris, R. D., Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers* **1992,** *32* (9), 1243-1250. DOI: 10.1002/bip.360320913

29. Trabbic-Carlson, K.; Setton, L. A.; Chilkoti, A., Swelling and Mechanical Behaviors of Chemically Cross-Linked Hydrogels of Elastin-like Polypeptides. *Biomacromolecules* **2003,** *4* (3), 572-580. DOI: 10.1021/bm025671z

30. Xu, D.; Asai, D.; Chilkoti, A.; Craig, S. L., Rheological Properties of Cysteine-Containing Elastin-Like Polypeptide Solutions and Hydrogels. *Biomacromolecules* **2012,** *13* (8), 2315-2321. DOI: 10.1021/bm300760s

31. Kim, B.; Lam, C. N.; Olsen, B. D., Nanopatterned Protein Films Directed by Ionic Complexation with Water-Soluble Diblock Copolymers. *Macromolecules* **2012,** *45* (11), 4572-4580. DOI:

32. Obermeyer, A. C.; Mills, C. E.; Dong, X.-H.; Flores, R. J.; Olsen, B. D., Complex coacervation of supercharged proteins with polyelectrolytes. *Soft Matter* **2016,** *12* (15), 3570-3581. DOI: 10.1039/C6SM00002A

33. Hamley, I. W., *The Physics of Block Copolymers*. Oxford University Press: 1998.

34. Glassman, M. J.; Olsen, B. D., End Block Design Modulates the Assembly and Mechanics of Thermoresponsive, Dual-Associative Protein Hydrogels. *Macromolecules* **2015,** *48* (6), 1832-1842. DOI: 10.1021/ma502494s

35. Wright, E. R.; Conticello, V. P., Self-assembly of block copolymers derived from elastin-mimetic polypeptide sequences. *Adv. Drug Delivery Rev.* **2002,** *54* (8), 1057-1073. DOI: 10.1016/S0169-409X(02)00059-5

36. Balsara, N. P.; Perahia, D.; Safinya, C. R.; Tirrell, M.; Lodge, T. P., Birefringence detection of the order-to-disorder transition in block copolymer liquids. *Macromolecules* **1992,** *25* (15), 3896-3901. DOI: 10.1021/ma00041a011

37. Olsen, B. D.; Segalman, R. A., Phase Transitions in Asymmetric Rod–Coil Block Copolymers. *Macromolecules* **2006,** *39* (20), 7078-7083. DOI: 10.1021/ma060994z

38. Cooper, C. L.; Dubin, P. L.; Kayitmazer, A. B.; Turksen, S., Polyelectrolyte-protein complexes. *Curr. Opin. Colloid Interface Sci.* **2005,** *10* (1-2), 52-78. DOI: 10.1016/j.cocis.2005.05.007

39. Lam, C. N.; Chang, D.; Wang, M.; Chen, W.-R.; Olsen, B. D., The shape of protein–polymer conjugates in dilute solution. *J. Polym. Sci., Part A: Polym. Chem.* **2016,** *54* (2), 292-302. DOI: 10.1002/pola.27975

40. Tanahatoe, J. J.; Kuil, M. E., Light scattering on semidilute polyelectrolyte solutions: Ionic strength and polyelectrolyte concentration dependence. *J Phys. Chem. B* **1997,** *101* (50), 10839-10844. DOI: 10.1021/jp9723776

41. Tanahatoe, J. J.; Kuil, M. E., Light scattering on semidilute polyelectrolyte solutions: Molar mass and polyelectrolyte concentration dependence. *J Phys. Chem. B* **1997,** *101* (45), 9233-9239. DOI: 10.1021/jp972227v

42. Muthukumar, M., Ordinary-extraordinary transition in dynamics of solutions of charged macromolecules. *Proc. Natl. Acad. Sci. U. S. A.* **2016**. DOI: 10.1073/pnas.1612249113

43. Arakawa, T.; Timasheff, S. N., Mechanism of Protein Salting in and Salting out by Divalent-Cation Salts - Balance between Hydration and Salt Binding. *Biochemistry* **1984,** *23* (25), 5912-5923. DOI: 10.1021/bi00320a004

44. Melander, W.; Horvath, C., Salt Effects on Hydrophobic Interactions in Precipitation and Chromatography of Proteins - Interpretation of Lyotropic Series. *Arch. Biochem. Biophys.* **1977,** *183* (1), 200-215. DOI: 10.1016/0003-9861(77)90434-9

45. Hammouda, B., SANS from homogeneous polymer mixtures: A unified overview. In *Polymer Characteristics*, Springer Berlin Heidelberg: Berlin, Heidelberg, 1993; pp 87-133.

46. Borue, V. Y.; Erukhimovich, I. Y., A statistical theory of weakly charged polyelectrolytes: fluctuations, equation of state and microphase separation. *Macromolecules* **1988,** *21* (11), 3240-3249. DOI: 10.1021/ma00189a019

47. Dobrynin, A. V.; Rubinstein, M.; Joanny, J.-F., Adsorption of a Polyampholyte Chain on a Charged Surface. *Macromolecules* **1997,** *30* (15), 4332-4341. DOI: 10.1021/ma9703057

48. Flory, P. J., *Principles of polymer chemistry*. Ithaca : Cornell University Press, 1953.: 1953.

49. Khokhlov, A. R., On the collapse of weakly charged polyelectrolytes. *J. Phys. A: Math. Gen.* **1980,** *13* (3), 979. DOI: 10.1088/0305-4470/13/3/030

50. De Gennes, P. G.; Pincus, P.; Velasco, R. M.; Brochard, F., Remarks on polyelectrolyte conformation. *J. Phys. France* **1976,** *37* (12), 1461-1473. DOI: 10.1051/jphys:0197600370120146100

51. Sing, C. E.; Zwanikken, J. W.; Olvera de la Cruz, M., Electrostatic control of block copolymer morphology. *Nat. Mater.* **2014,** *13* (7), 694-8. DOI: 10.1038/nmat4001

Chapter 4 : Cononsolvency of Elastin-Like Polypeptides (ELPs) in Water/Alcohol Solutions

4.1 Abstract

Elastin-like polypeptides (ELPs) are one of the most widely-studied classes of protein material due to their lower critical solution temperature (LCST)-like thermoresponsive behavior in aqueous solutions. Here, it is shown that ELPs also exhibit cononsolvency effects, similar to many other water-soluble polymers. The effect of solvent composition on the dilute solution phase behavior of an elastin-like polypeptide (ELP) is studied here in water/alcohol blends that contain 0-40 vol% methanol, ethanol, isopropanol, or 1-propanol. In all systems studied, the ELP exhibits cononsolvency behavior at low alcohol content, as indicated by a decrease in the transition temperature of the ELP. When the alcohol added is ethanol, isopropanol, or 1-propanol, the decrease in transition temperature is followed by a region of complete ELP insolubility, and finally, the emergence of upper-critical solution transition (UCST)-like behavior. The ELP is completely soluble at all temperatures measured at alcohol contents above 40 vol%. The effect of sodium chloride on this ELP cononsolvency in water/ethanol blends was also studied. Unlike the previously studied polymer poly(*N*-isoropylacrylamide) (PNIPAM), ELP exhibits nonmonotonic changes in transition temperature with the addition of sodium chloride at ethanol contents that produce UCST-like transitions of the ELP. This discovery of ELP cononsolvency in water/alcohol systems introduces a new handle with which the solubility of ELPs can be tuned.

4.2 Introduction

Elastin-like polypeptides (ELPs) are a family of biopolymers derived from tropoelastin with the repeat unit motif Val-Pro-Gly-Xaa-Gly (VPGXG), where Xaa is known as the "guest residue" and can be any amino acid except proline.[1] One key feature of these polypeptides is their lower critical solution temperature (LCST) behavior in aqueous solution: they are soluble at low temperatures and become insoluble upon heating.[2] The thermoresponsive nature of ELPs has inspired their use in applications in a broad spectrum of fields. For example, ELPs have found use in tissue engineering and drug delivery, where they remain soluble at low temperatures outside the body, and gel into a matrix at body temperature.[3-7] This temperature-induced transition has also been used to selectively precipitate and solubilize ELP-tagged proteins to develop a chromatography-free route to globular protein purification. [8-9]

Several features of ELP dilute and semidilute phase behavior in aqueous solutions have been investigated extensively, including ELP sequence, solution pH, and ionic strength. Here, dilute refers to solutions in which the ELP is below its overlap concentration, and semidilute indicates the solution is above the overlap concentration, but the ELP concentration in solution is still low. [10] Work by Urry established that the transition temperature of an ELP could be tuned by changing the hydrophobicity of the fourth position guest residue in the repeat sequence.[2] In addition to sequence, solution conditions also have a strong effect on ELP phase behavior in solution. In particular, it has been shown that the addition of different salts can strongly modulate the transition temperature of ELPs.[11] The impact of a given anion on ELP transition temperature can be predicted by the kosmotropic or chaotropic nature of the anion used.[12] Much of the work done on ELPs in dilute solution, including studies of ELP conformation, suggests that they share many similarities with synthetic polymers, including poly(2-oxazaline)s and polypeptoids.[13-17] However, the polypeptide nature of ELPs makes certain features of ELPs inherently different from synthetic polymers. For example, it is well-established that the phase transition of ELPs upon heating results from the formation of secondary structure in the ELP. [18] Furthermore, changes to the specific amino acid sequence of the ELP can also dramatically change ELP solution properties, including the introduction of strong hysteresis in the ELP phase transition. [19-21] However, little is known about the phase behavior of ELPs in nonaqueous solutions or in blends of water and other solvents. Behavior of ELPs in blends of water and nonaqueous solvents has implications in the use of ELP materials for delivery of drugs requiring cosolvents for dissolution, as well as in purification of ELP-based materials.

Blending of solvents for polymer dissolution can enhance or decrease the solubility of a given polymer. For example, poly(methyl methacrylate) (PMMA) experiences an increase in solubility in certain water/ethanol blends (as compared to water or ethanol alone), as evidenced by the decrease in the transition temperature of PMMA, which exhibits upper critical solution temperature (UCST) behavior.[22] This enhanced solubility in blended solvents has been termed cosolvency.[23-24] In contrast, poly(*N*-isoropylacrylamide) (PNIPAM) exhibits a decrease in solubility in certain water/methanol blends, relative to its solubility in either of the individual solvents, as indicated by a decrease in the LCST transition temperature of PNIPAM in these solutions. [25] This behavior has accordingly been titled cononsolvency, and has since been observed for PNIPAM in a broad array of water/solvent blends.[26]

PNIPAM cononsolvency in water/alcohol blends has been the subject of several experimental and theoretical studies. Initially, PNIPAM cononsolvency was reported in methanol/water systems, where the PNIPAM transition temperature was found to be reduced in certain methanol/water blends in dilute solution.[25, 27] Later studies by Costa and Freitas showed that PNIPAM cononsolvency could be observed in a broad range of water-miscible solvents, including ethanol, isopropanol, and 1-propanol. Interestingly, in these more hydrophobic alcohols, a transition from LCST-like behavior to UCST-like behavior occurred with increasing alcohol content.[26] Initial work attempting to describe this behavior with a Flory-Huggins formalism observed that the phase behavior of PNIPAM in methanol/water mixtures could be predicted by allowing the solvent-solvent interaction parameter, χ_{12}, vary with solvent fraction.[25, 28-30] Later work by Tanaka and coworkers has proposed that competition between PNIPAM-water and PNIPAM-alcohol hydrogen bonding and the cooperativity of these hydrogen bonds along the PNIPAM chain drive cononsolvency.[31-33] Recently, Fukai *et. al.* showed that this model could also predict the appearance of UCST behavior of PNIPAM in water-ethanol and water-isopropanol blends. [34]

Here, we report, for the first time, cononsolvency behavior of an ELP in four different water-alcohol systems (**Figure 4-1**). The resulting phase behavior is compared to that of PNIPAM to understand the similarities and differences in the thermodynamic driving forces of ELP and PNIPAM cononsolvency. The effect of sodium chloride on ELP cononsolvency in ethanol is also characterized. This work introduces a new handle for inducing phase transitions in ELP systems, with potential applications in protein purification as well as drug delivery depots for alcohol-soluble drugs.

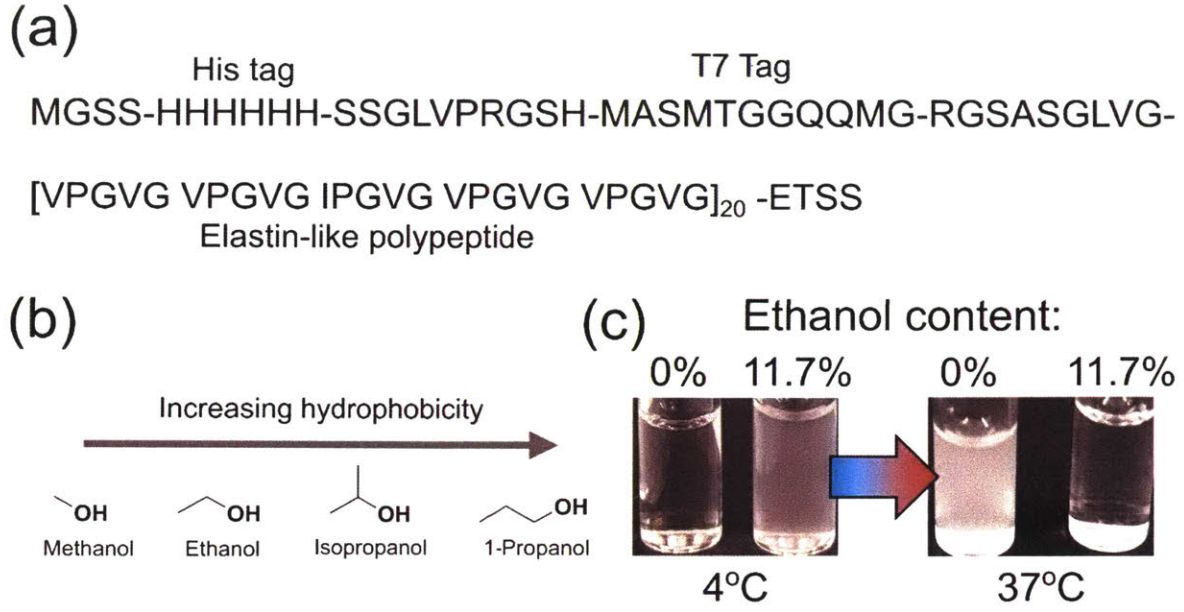


Figure 4-1. (a) Amino acid sequence of ELP used in this study. (b) Chemical structures of alcohols investigated in this study. (c) Photographs of 1 mg/mL ELP in ethanol/water mixtures containing 0 mol% and 11.7 mol% ethanol at 4 °C and 37 °C.

4.3 Materials & Methods

Materials. All solvents and salts used were reagent grade. Water used in all experiments was purified using a Millipore Milli-Q® system and passed through a 0.22 μm filter.

Protein expression and purification. The ELP sequence used in this study can be found in **Figure 4-1** and **Figure C-1**. The T7 tag preceding the ELP repeat sequence was included as it has been reported to improve expression yield of ELP-based fusion proteins.[35] The 6xHis tag was included as a part of the pET28b expression vector, and is also expected to improve soluble yield of the ELP. Protein was expressed in *Escherichia coli* strain BL21(DE3). Expression was performed using 5 L of Terrific Broth (TB) in a fermenter with 7 L working volume at 30 °C for 18 h with no induction. Cells were harvested by centrifugation. Cells were resuspended in 100 mL lysis buffer (10 mM tris, 1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 5 mM $MgCl_2$, pH 7.5) for every 30 g of wet cell mass and subsequently incubated with 1 mg/mL lysozyme at 4 °C for 2 hours. The resuspended cells were sonicated using a tip sonicator. This lysate was clarified by high-speed centrifugation at 4 °C.

The clarified lysate was subjected to two rounds of inverse transition cycling (ITC),[8-9] where precipitation was induced by the haddition of 1.5 M sodium chloride and incubation at 37 °C, and

resolubilization was achieved by resuspension in lysis buffer at 4 °C. Samples were then dialyzed into 20 mM pH 8 tris buffer. After dialysis, urea was added to the protein solution to a final concentration of 6 M to prevent protein aggregation. The protein solution was then purified using fast protein liquid chromatography (FPLC) with an anion exchange column (HiTrap Q HP anion exchange chromatography column, GE Life Sciences). Protein was eluted over 30 column volumes using a linear gradient that increased salt concentration from 0 to 200 mM NaCl in 20 mM pH 8 tris with 6 M urea. Content and purity of the ELP in different fractions were evaluated using SDS-PAGE that sampled every 3 wells over the entire gradient. Elution fractions containing the protein of interest were collected and dialyzed into MilliQ purified water. After dialysis, the ELP solution was freeze dried. Final protein purity was confirmed by SDS-PAGE (**Figure C-2**).

Sample preparation. Lyophilized proteins were dissolved at 5 mg/mL in water overnight at 4 °C. Samples were then diluted with a combination of water, alcohol, and 0.5 M sodium chloride or 5 M sodium chloride to bring the solution to the correct solvent content and salt concentration, as well as to bring the total protein concentration in solution to 1 mg/mL (for turbidimetry) or 0.5 mg/mL (for circular dichroism).

Turbidimetry. Samples (1 mg/mL) were loaded into a quartz cuvette and sealed. The cuvette temperature was ramped at a rate of 1 °C/min, and transmittance at 600 nm was monitored. Transition temperature was taken as the point where a 50% change in transmittance was reached. Error bars on transition temperatures correspond to standard deviations over 3 measurements, and error bars on mol% alcohol and mM NaCl were calculated by implementing error propagation on pipetting error (see **C.1 Propagation of Pipetting Error**).

Circular dichroism. Samples (0.5 mg/mL) were loaded into a quartz cuvette with a 1 mm path length and sealed. Samples were scanned in a JASCO Model J-1500 circular dichroism spectrometer equipped with a Peltier temperature controller. CD spectra were collected with a 0.5 nm resolution, with a bandwidth of 1 nm, and a scan speed of 50 nm per minute. Data was converted into mean residue molar ellipticity based on the prepared concentration (0.5 mg/mL).

4.4 Results & Discussion

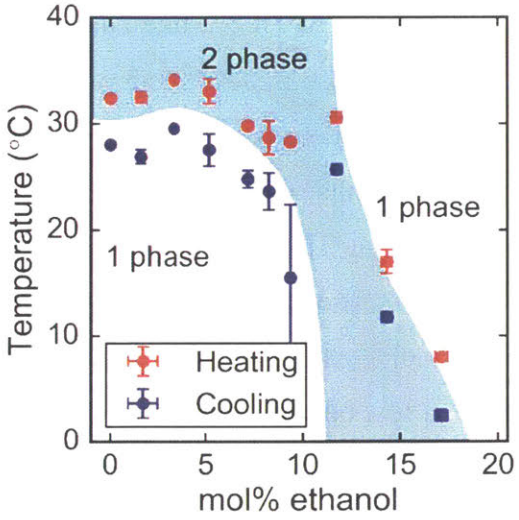


Figure 4-2. Transition temperatures measured by turbidimetry on heating and cooling of 1 mg/mL ELP in a range of water/ethanol solvents.

The elastin-like polypeptide (ELP) studied here exhibits cononsolvency behavior in ethanol/water blends (**Figure 4-2**). In pure water, the transition temperature of the ELP upon heating is 32.4 ± 0.2 °C. As the ethanol content increases to 3.3 mol%, the solubility of the ELP increases slightly, as indicated by the increase in the transition temperature to 34.1 ± 0.2 °C. As the ethanol content increases further, the solubility of the ELP in the binary solution decreases, with the transition temperatures dropping to 29.8 ± 0.4 °C and 28.3 ± 0.1 °C at 7.2 and 9.4 mol% ethanol, respectively. At 11.7 mol% ethanol, the phase separation behavior of the ELP changes from LCST-like, where the ELP is soluble at low temperatures and insoluble at high temperatures, to UCST-like, where the ELP solubilizes upon heating. At 11.7 mol% ethanol, the transition temperature upon cooling is 25.7 ± 0.4 °C. This transition temperature decreases sharply as ethanol content increases to 14.3 and 17.1 mol%. Above 17.1 mol% ethanol, the ELP is completely soluble at temperatures above 0 °C.

ELP phase behavior in binary water/ethanol solutions exhibits qualitative similarities, but quantitative differences from that observed in PNIPAM. Cononsolvency behavior of PNIPAM in several different binary aqueous solutions was previously characterized by Costa and Freitas.[26] Their work shows that in ethanol/water solutions, PNIPAM solubility decreases with increasing

ethanol content, as indicated by a decrease in the LCST cloud point. PNIPAM subsequently enters a region of complete insolubility between ethanol contents of 15 and 28 mol%. As ethanol content increases beyond 28 mol%, the emergence of UCST-like behavior is observed, similar to that reported here for the ELP system. The precise mole percent of ethanol at which LCST behavior vanishes is different for the two materials—between 9.4 and 11.7 mol% for the ELP, compared to 15 mol% for PNIPAM. The ethanol contents over which the ELP and PNIPAM are completely insoluble are also quantitatively different. The ELP is insoluble over a very narrow range of ethanol/water contents (9.4-11.7 mol% ethanol), whereas PNIPAM is insoluble over a broad range of ethanol/water contents (15-28 mol% ethanol). PNIPAM in ethanol/water also does not experience an increase in solubility at low ethanol contents as the ELP does at 3.3 mol% ethanol. Interestingly, PNIPAM does exhibit this same behavior (slight increase in solubility at low nonaqueous solvent content) in DMF/water blends. PNIPAM in DMF/water blends also has a narrower window of complete PNIPAM insolubility, much like that observed for the ELP in ethanol/water blends.

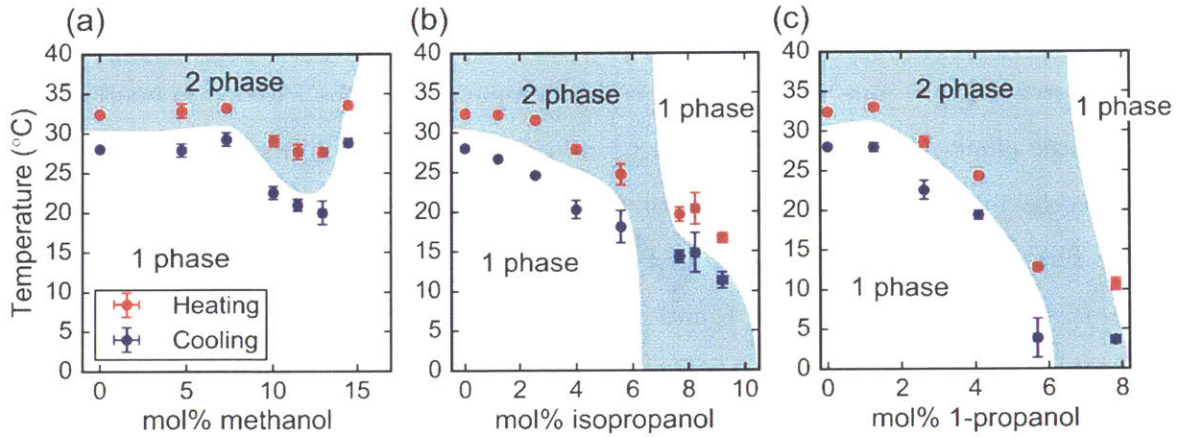


Figure 4-3. Transition temperatures measured by turbidimetry on heating and cooling of 1 mg/mL ELP binary aqueous solutions at various (a) methanol, (b) isopropanol, and (c) 1-propanol contents.

ELP cononsolvency is also observed in blends of water with other alcohols, as shown in **Figure 4-3**. For blends of methanol/water, as the methanol content increases from 0 to 7.3 mol%, there is no detectable change in the ELP transition temperature (within error of measurement), suggesting

that solubility remains the same in this solvent region. Between 7.3 and 10.0 mol% methanol, the ELP transition temperature begins to decrease with increasing solvent content, indicating the onset of cononsolvency behavior of the ELP in methanol/water blends. The transition temperature continues to decrease until it reaches a minimum of 27.7 ± 0.5 °C (on heating) at 13.0 mol% methanol, which appears to be the point of maximal cononsolvency. After this point, the transition temperature (and thus ELP solubility) increases rapidly. At methanol contents greater than 14.5 mol%, the ELP is miscible across the entire experimentally accessible temperature range.

ELP phase behavior in binary aqueous solutions containing isopropanol and 1-propanol shares many qualitative similarities to that of the ELP in ethanol/water solutions (**Figure 4-3b,c**). First, at lower alcohol contents, the LCST-like transition temperature of the ELP decreases with increasing alcohol content. After a certain threshold alcohol content is reached (9.4 mol%, 5.6 mol%, and 5.7 mol% for ethanol, isopropanol, and 1-propanol, respectively), the LCST-like behavior vanishes, and a region of complete ELP insolubility appears. Finally, upon further increasing alcohol content, UCST-like phase transitions begin to appear, and eventually the ELP is completely soluble at all temperatures measured. One interesting difference between ethanol and the other three alcohols studied here is that ethanol is the only alcohol that appears to cause an observable increase in ELP solubility at low solvent content (3.3 mol% ethanol).

While the phase behavior of the ELP in different alcohol/water solutions shares some qualitative similarities, the differences in quantitative phase behavior suggest trends in phase behavior with alcohol hydrophobicity. The alcohols used in this study, in order of increasing hydrophobicity are methanol, ethanol, isopropanol, and 1-propanol. It has been shown that while 1-propanol has the same carbon number as isopropanol, the linear arrangement of its carbon chain makes it more difficult for water to structure around it, thus making it more hydrophobic.[36] The least hydrophobic alcohol, methanol, does not strongly perturb ELP solubility, as the maximum change in transition temperature observed is 4.7 °C. Methanol also does not produce UCST-like transitions in the ELP at any concentrations. While ethanol, isopropanol, and 1-propanol all exhibit a transition from LCST-like behavior to UCST-like behavior with increasing solvent content, the onset of this transition as well as the region of complete insolubility between the LCST-like and UCST-like behavior changes with alcohol hydrophobicity. In ethanol, isopropanol, and 1-propanol, respectively, the LCST-like behavior vanishes after 9.4 mol%, 5.6 mol%, and 5.7 mol%, respectively, indicating that ELP solubility is suppressed more rapidly with alcohol content as the

hydrophobicity of the alcohol increases. Finally, as alcohol hydrophobicity increases, the region over which UCST-like behavior is observed before the onset of complete solubility decreases. In ethanol, UCST-like behavior can be observed in the range of 11.7 to 17.1 mol% alcohol. In isopropanol, UCST-like behavior can only be observed between 7.7 and 9.2 mol% alcohol. In the most hydrophobic alcohol, 1-propanol, the UCST-like behavior is only observed at 7.8 mol% alcohol at 3.7 ± 0.5 °C upon cooling.

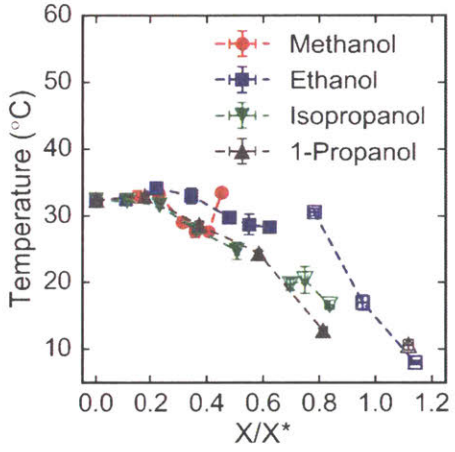


Figure 4-4. ELP transition temperatures upon heating in different solvents plotted against mole fraction solvent divided by mole fraction alcohol at which water-alcohol mixing enthalpy is minimized, for methanol, ethanol, isopropanol, and 1-propanol. Filled symbols correspond to LCST-like transitions and open symbols correspond to UCST-like transitions.

Comparison of ELP cononsolvency data to published PNIPAM cononsolvency data in the low alcohol regime suggests that the principle driving force for cononsolvency in these systems is similar. Previously published cloud point experiments on PNIPAM in the same aqueous/alcohol blends studied here (methanol, ethanol, isopropanol, 1-propanol) show that PNIPAM follows similar trends in phase behavior with respect to alcohol hydrophobicity.[26, 37] In both ELP and PNIPAM systems, the change in transition temperature per mole alcohol added increases with increasing alcohol hydrophobicity. This leads to an earlier onset of complete ELP and PNIPAM insolubility as alcohol hydrophobicity increases.[26, 37] Two groups have attempted to explain these trends in PNIPAM cononsolvency behavior with respect to alcohol hydrophobicity in the context of thermodynamic driving forces. Work by Fukai *et. al.* has shown that this shift in PNIPAM phase

behavior with alcohol hydrophobicity can be predicted by a thermodynamic model proposing that PNIPAM cononsolvency in alcohol/water mixtures arises as a result of cooperative hydrogen bonding along the polymer chain.[33-34, 38] The fact that similar trends are observed for the ELP suggests that this mechanism may also apply to this system; however, a lack of requisite model parameters does not permit direct comparison of this model to our ELP data. Bischofberger *et. al.* hypothesized that the phase behavior at low alcohol content in different alcohols could be explained by the ability of different alcohols to stabilize bulk water, as quantified by their excess enthalpies of mixing with water. They support this claim by rescaling alcohol mole fraction by the mole fraction at which excess mixing enthalpy of a given water/alcohol mixture is minimized, and then plotting the transition temperatures in different solvents with respect to this rescaled mole fraction. In the case of PNIPAM, this rescaling leads to a collapse of the low-alcohol content data onto a single curve for all four solvents studied.[37] If this same rescaling is applied to data taken on the ELP, similar collapse is observed (**Figure 4-4**). It should be noted that this suggested mechanism of cononsolvency is consistent with literature reports that the LCST-like transition of ELPs is a result of disruption of the hydrogen bonding network surrounding the polypeptide, as a decrease in the enthalpy of bulk water would be expected to decrease the net enthalpic gain associated with the structuring of water around the ELP (resulting in an earlier onset of entropically driven collapse). [39] Interestingly, the data for ethanol does not collapse as well as the other three alcohols. The source of this discrepancy can likely be attributed to whatever specific interaction is responsible for the increase in ELP solubility at 3.3 mol% ethanol, which, as noted previously, is not observed for other alcohols.

Ultimately, these comparisons in the LCST-like region of phase space indicate that the cononsolvency behavior of ELPs and PNIPAM in the LCST regime are likely governed by similar underlying physics. As such, it is likely that the observed cononsolvency transitions are not a strong function of molecular weight of the ELP in this regime, as previous studies on PNIPAM have shown that the transition temperature is not a function of PNIPAM molecular weight. [37] It is also worth noting that some degree of this same cononsolvency is observed in poly(2-propyl-2-oxazaline) (P(PropOx)), as its LCST-like transition temperature decreases with increasing ethanol content in ethanol/water mixtures. In contrast to the ELPs studied here, however, P(PropOx) does not pass through a region of complete insolubility as ethanol content increases further; instead, it becomes completely soluble at ethanol contents above 20 mol%.[40]

Trends in ELP phase behavior at high alcohol content show qualitative and quantitative differences from those observed in PNIPAM, indicating that the thermodynamics that govern the phase behavior of the ELP in this region differ from PNIPAM. Literature reports on the thermodynamics of PNIPAM cononsolvency in the high alcohol content regime propose that phase behavior is dominated by mixing entropy.[34, 37] Bischofberger *et al.* observe that if the geometry and size of the alcohols do not significantly change the Flory-Huggins parameter governing PNIPAM-alcohol interactions, the high-alcohol content UCST phase boundary of PNIPAM should collapse with respect to alcohol volume fraction. [37] These authors observe evidence of such collapse in their PNIPAM system in the UCST region for ethanol, isopropanol, and 1-propanol. Such a collapse does not occur in the ELP system, where the slopes of the UCST-like phase boundary vary strongly regardless of abscissa choice (mole percent, X/X*, or volume percent, **Figure 4-4, Figure C-3**). Assuming that the driving force for phase separation in this region is still mixing entropy (which is in agreement with other thermodynamic models for emergent UCST cononsolvency[34]), this would suggest that the ELP-alcohol Flory-Huggins parameters vary substantially with alcohol hydrophobicity.

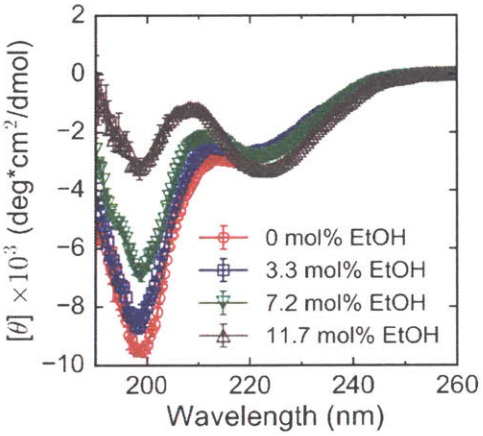


Figure 4-5. Circular dichroism data (in mean residue molar ellipticity) of the ELP (at 0.5 mg/mL) in the 1-phase region at different ethanol contents. 0-7.2 mol% ethanol measurements were taken at 10 °C and 11.7 mol% ethanol measurement was taken at 37 °C to ensure ELP solubility.

Circular dichroism (CD) spectra taken on the ELP in various water/ethanol solutions suggest that the differences between ELP and PNIPAM phase behavior at high alcohol contents are a result of differences in the way the two macromolecules form intramolecular hydrogen bonds. Previous literature has established that ELPs, unlike PNIPAM, form a beta-turn structure in solution when they coacervate upon heating. [18, 41] Here, CD spectra were taken on ELPs in several water/ethanol solvents (**Figure 4-5**) under conditions which the ELP was completely soluble (10 °C for 0-7.2 mol% ethanol and 37 °C for 11.7 mol% ethanol). Similar to previously published results on α-elastin and shorter ELPs, the ELP secondary structure changes as the ethanol content of the solvent increases. [18, 42-43] Several features of the CD data indicate this shift. As ethanol content increases, the intensity of the negative peak at 196 nm decreases, suggesting that the random coil content in solution is decreasing. [44] Concomitantly, the increase in ethanol content in the solvent leads to an increase in intensity of a negative peak at 218 nm and what appears to be a positive peak at 205 nm, both of which are associated with the formation of a beta turn. [43-44] Thus, it is concluded that the fraction of ELPs involved in the formation of a beta turn structure increases with increasing ethanol content, indicating an increase in the number of ELP intramolecular hydrogen bonds. Similarly, Fourier-transform infrared (FTIR) spectroscopy studies on PNIPAM in various methanol-water blends have shown that the hydrogen bonding in the high-alcohol regime is predominantly intramolecular between PNIPAM chains, and that in this regime, the chain is soluble due to solvation by the alcohol instead of by hydrophobic hydration.[45] Thus, while both the ELP and PNIPAM form intramolecular hydrogen bonds at high alcohol contents, the structure of these hydrogen bonding networks is different, potentially causing the differences in the observed phase behavior. The ELP forms hydrogen bonds along its backbone (resulting in the observed secondary structure change), and the alcohol solvates the hydrophobic side chains. This is different from PNIPAM, which forms hydrogen bonds along its side chains and the alcohol solvates the backbone.

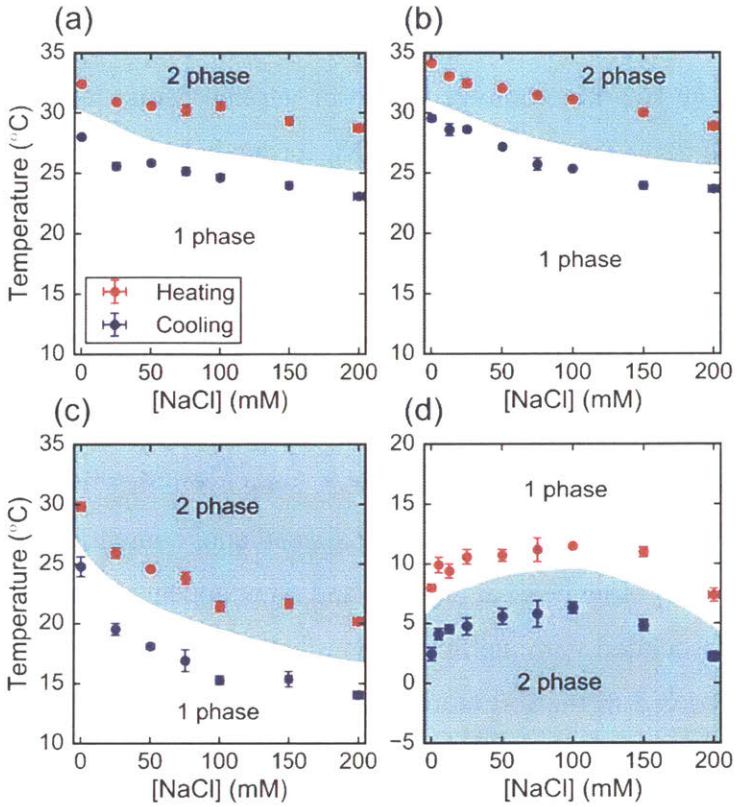


Figure 4-6. Transition temperatures measured by turbidimetry on heating and cooling of 1 mg/mL ELP solutions at various sodium chloride concentrations in (a) 0 mol% ethanol, (b) 3.3 mol% ethanol, (c) 7.2 mol% ethanol, and (d) 17.1 mol% ethanol

In the region of ethanol/water phase space where the ELP exhibits LCST-like behavior, the addition of sodium chloride (NaCl) to the solution produces a monotonic decrease in ELP solubility where the magnitude of the solubility change depends on the ethanol content in the solvent. As expected from other reports in the literature, the addition of NaCl to the ELP solution when there is no ethanol present results in a monotonic decrease in the transition temperature of the ELP (**Figure 4-6**a). [14] Similarly, the addition of NaCl to ELP in 3.3 mol% and 7.2 mol% EtOH also results in a monotonic decrease of ELP transition temperature. Interestingly, in the region of phase space where the ELP exhibits strong cononsolvency behavior (7.2 mol% EtOH), the addition of NaCl causes a much larger drop in the ELP transition temperature compared to regions where ELP cononsolvency is not observed (0 and 3.3 mol% EtOH). For example, the transition temperature of the ELP in 3.3 mol% EtOH is depressed by 3.6 ± 0.2 °C with the addition of 200

mM NaCl to the solvent. In contrast, the ELP transition temperature in 7.2 mol% EtOH, 200 mM NaCl is decreased by 9.6 ± 0.4 °C.

When NaCl is added to ELP solutions in ethanol/water regions where UCST-like behavior is observed, NaCl induces nonmonotonic changes in the ELP transition temperature not observed in the LCST-like region of the ELP/ethanol/water phase space. When NaCl is added to solutions containing 17.1 mol% ethanol (**Figure 4-6d**), the solubility of the ELP decreases at low NaCl concentrations, evidenced by a ~4 °C increase in the ELP transition temperature between 5 and 150 mM NaCl. The ELP does not recover its 0 mM NaCl transition temperature until 200 mM NaCl is added.

The addition of NaCl to the ELP/ethanol/water system causes nonmonotonic changes in ELP phase behavior in the UCST-like region of phase space that are not observed when the same amount of NaCl is added to PNIPAM/ethanol/water systems. Previous work investigating the effect of different Hofmeister anions to PNIPAM in ethanol/water and DMSO/water blends showed that, while the different Hofmeister anions could increase or decrease the PNIPAM transition temperature in these ethanol/water and DMSO/water blends (depending on the salt used), all salts investigated caused linear, monotonic changes in the PNIPAM transition temperature with increasing salt concentration.[46] In contrast, addition of NaCl to ELP in water/ethanol blends causes nonmonotonic behavior at ethanol contents where UCST-like behavior is observed (**Figure 4-6**). Previous studies on the effects of the Hofmeister anion series on PNIPAM and ELP phase behavior have noted that both PNIPAM and ELPs exhibit similar trends in transition temperature behavior with respect to the chaotropic and kosmotropic nature of anions.[13, 47] However, chaotropic anions cause much stronger salting in of ELPs than of PNIPAM, as indicated by the increase in ELP/PNIPAM transition temperature at intermediate anion concentrations.[13, 47] It is possible that whatever mechanism causes stronger salting in behavior for chaotropic anions in ELPs is also responsible for the nonmonotonicities observed in the data presented here. It should be noted that the addition of NaCl to ELP/water/ethanol systems in the LCST-like region of phase space produces similar trends with added NaCl to those observed in PNIPAM. Thus, it is also possible that the deviations in phase behavior with respect to salt addition between the ELP and PNIPAM in the UCST-like region are similar in origin to the differences noted earlier (such as hydrogen bonding network around the ELP/PNIPAM chains).

Finally, it is important to contextualize these results by noting that this study was performed using an ELP that has 40 and 4 amino acids flanking the repetitive sequence at the N- and C-termini of the polypeptide, respectively. Previous work by Trabbic-Carlson *et. al.* has demonstrated that flanking peptide sequences can have a significant impact on ELP solubility.[48] Based on the work by Trabbic-Carlson *et. al.*, the amino acid sequence flanking the ELP used here is expected to make the ELP more hydrophilic. Thus, in future studies looking to develop understanding of ELP cononsolvency further, the impact of flanking peptide sequence will be an important parameter to consider.

4.5 Conclusions

Cononsolvency of an ELP in various water/alcohol blends is reported, representing a new handle for the tuning of ELP phase behavior. The phase behavior of the ELP is found to be qualitatively similar to that of PNIPAM in methanol, ethanol, isopropanol, and 1-propanol. At low alcohol contents, changing alcohol hydrophobicity has similar effects on both ELP and PNIPAM phase behavior. In the regime where cononsolvency causes UCST-like behavior of the ELP and PNIPAM; however, alcohol hydrophobicity changes the phase boundaries of the ELP and PNIPAM differently. This indicates that differences in the ELP and PNIPAM, such as the ability of the polymer backbone to hydrogen bond, play a more important role in the thermodynamics governing cononsolvency in the higher alcohol regime. The effect of sodium chloride on ELP cononsolvency in water/ethanol blends was also investigated. The ELP transition temperature scales monotonically with NaCl concentration in the LCST-like region of phase space. In contrast, the ELP transition temperature changed non-monotonically with NaCl concentration in the UCST-like region of phase space, in contrast to monotonic dependencies on NaCl concentration observed in PNIPAM/water/ethanol systems in this same region.

4.6 References

1. Foster, J. A.; Bruenger, E.; Gray, W. R.; Sandberg, L. B., Isolation and Amino-Acid Sequences of Tropoelastin Peptides. *J. Biol. Chem.* **1973,** *248* (8), 2876-2879 DOI:

2. Urry, D. W.; Gowda, D. C.; Parker, T. M.; Luan, C.-H.; Reid, M. C.; Harris, C. M.; Pattanaik, A.; Harris, R. D., Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers* **1992,** *32* (9), 1243-1250 DOI: 10.1002/bip.360320913.

3. Glassman, M. J.; Avery, R. K.; Khademhosseini, A.; Olsen, B. D., Toughening of Thermoresponsive Arrested Networks of Elastin-Like Polypeptides To Engineer Cytocompatible Tissue Scaffolds. *Biomacromolecules* **2016,** *17* (2), 415-426 DOI: 10.1021/acs.biomac.5b01210.

4. Kyle, S.; Aggeli, A.; Ingham, E.; McPherson, M. J., Production of self-assembling biomaterials for tissue engineering. *Trends Biotechnol.* **2009,** *27* (7), 423-433 DOI: 10.1016/j.tibtech.2009.04.002.

5. Nettles, D. L.; Chilkoti, A.; Setton, L. A., Applications of elastin-like polypeptides in tissue engineering. *Adv. Drug Delivery Rev.* **2010,** *62* (15), 1479-1485 DOI: 10.1016/j.addr.2010.04.002.

6. Nicol, A.; Gowda, C.; Urry, D. W., Elastic Protein-Based Polymers as Cell Attachment Matrices. *J Vasc Surg* **1991,** *13* (5), 746-748 DOI: 10.1016/0741-5214(91)90373-3.

7. Rodríguez-Cabello, J. C.; Arias, F. J.; Rodrigo, M. A.; Girotti, A., Elastin-like polypeptides in drug delivery. *Adv. Drug Delivery Rev.* **2016,** *97*, 85-100 DOI: 10.1016/j.addr.2015.12.007.

8. Meyer, D. E.; Chilkoti, A., Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat. Biotechnol.* **1999,** *17* (11), 1112-1115 DOI: 10.1038/15100.

9. Trabbic-Carlson, K.; Liu, L.; Kim, B.; Chilkoti, A., Expression and purification of recombinant proteins from Escherichia coli: Comparison of an elastin-like polypeptide fusion with an oligohistidine fusion. *Protein Science* **2004,** *13* (12), 3274-3284 DOI: 10.1110/ps.04931604.

10. Gennes, P. G. d., *Scaling concepts in polymer physics / Pierre-Gilles de Gennes*. Cornell University Press: Ithaca, N.Y, 1979; DOI:

11. Urry, D. W., Molecular Machines - How Motion and Other Functions of Living Organisms Can Result from Reversible Chemical Changes. *Angew. Chem. Int. Edit.* **1993,** *32* (6), 819-841 DOI: 10.1002/anie.199308191.

12. Cho, Y. H.; Zhang, Y. J.; Christensen, T.; Sagle, L. B.; Chilkoti, A.; Cremer, P. S., Effects of Hofmeister Anions on the Phase Transition Temperature of Elastin-like Polypeptides. *J. Phys. Chem. B* **2008,** *112* (44), 13765-13771 DOI: 10.1021/jp8062977.

13. Zhang, Y.; Furyk, S.; Bergbreiter, D. E.; Cremer, P. S., Specific ion effects on the water solubility of macromolecules: PNIPAM and the Hofmeister series. *J. Am. Chem. Soc.* **2005,** *127* (41), 14505-10 DOI: 10.1021/ja0546424.

14. Cho, Y.; Zhang, Y.; Christensen, T.; Sagle, L. B.; Chilkoti, A.; Cremer, P. S., Effects of Hofmeister anions on the phase transition temperature of elastin-like polypeptides. *J. Phys. Chem. B* **2008,** *112* (44), 13765-71 DOI: 10.1021/jp8062977.

15. Tao, X. F.; Du, J. W.; Wang, Y. X.; Ling, J., Polypeptoids with tunable cloud point temperatures synthesized from N-substituted glycine N-thiocarboxyanhydrides. *Polym. Chem.* **2015,** *6* (16), 3164-3174 DOI: 10.1039/c5py00191a.

16. Hoogenboom, R.; Schlaad, H., Thermoresponsive poly(2-oxazoline)s, polypeptoids, and polypeptides. *Polym. Chem.* **2017,** *8* (1), 24-40 DOI: 10.1039/c6py01320a.

17. Mills, C. E.; Michaud, Z.; Olsen, B. D., Elastin-like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins. *Biomacromolecules* **2018,** *19* (7), 2517-2525 DOI: 10.1021/acs.biomac.8b00147.

18. Urry, D. W.; Long, M. M.; Cox, B. A.; Ohnishi, T.; Mitchell, L. W.; Jacobs, M., The synthetic polypentapeptide of elastin coacervates and forms filamentous aggregates. *Biochim. Biophys. Acta, Protein Struct.* **1974,** *371* (2), 597-602 DOI: 10.1016/0005-2795(74)90057-9.

19. Urry, D. W.; Luan, C.-H.; Harris, C. M.; Parker, T. M., Protein-Based Materials with a Profound Range of Properties and Applications: The Elastin ΔTt Hydrophobic Paradigm. In *Protein-Based Materials*, McGrath, K.; Kaplan, D., Eds. Birkhäuser Boston: Boston, MA, 1997; pp 133-177.

20. Glassman, M. J.; Olsen, B. D., Arrested Phase Separation of Elastin-like Polypeptide Solutions Yields Stiff, Thermoresponsive Gels. *Biomacromolecules* **2015,** *16* (12), 3762-73 DOI: 10.1021/acs.biomac.5b01026.

21. Li, N. K.; Roberts, S.; Quiroz, F. G.; Chilkoti, A.; Yingling, Y. G., Sequence Directionality Dramatically Affects LCST Behavior of Elastin-Like Polypeptides. *Biomacromolecules* **2018,** *19* (7), 2496-2505 DOI: 10.1021/acs.biomac.8b00099.

22. Hoogenboom, R.; Becer, C. R.; Guerrero-Sanchez, C.; Hoeppener, S.; Schubert, U. S., Solubility and Thermoresponsiveness of PMMA in Alcohol-Water Solvent Mixtures. *Aust. J. Chem.* **2010,** *63* (8), 1173-1178 DOI: 10.1071/Ch10083.

23. Shultz, A. R.; Flory, P. J., Phase Equilibria in Polymer—Solvent Systems. III. Three-component Systems1. *J. Am. Chem. Soc.* **1953,** *75* (22), 5681-5685 DOI: 10.1021/ja01118a061.

24. Wolf, B. A.; Blaum, G., Measured and calculated solubility of polymers in mixed solvents: Monotony and cosolvency. *J. Polym. Sci., Part B: Polym. Phys.* **1975,** *13* (6), 1115-1132 DOI: doi:10.1002/pol.1975.180130605.

25. Schild, H. G.; Muthukumar, M.; Tirrell, D. A., Cononsolvency in mixed aqueous solutions of poly(N-isopropylacrylamide). *Macromolecules* **1991,** *24* (4), 948-952 DOI: 10.1021/ma00004a022.

26. Costa, R. O. R.; Freitas, R. F. S., Phase behavior of poly(N-isopropylacrylamide) in binary aqueous solutions. *Polymer* **2002,** *43* (22), 5879-5885 DOI: 10.1016/S0032-3861(02)00507-4.

27. Winnik, F. M.; Ringsdorf, H.; Venzmer, J., Methanol Water as a Co-Nonsolvent System for Poly(N-Isopropylacrylamide). *Macromolecules* **1990,** *23* (8), 2415-2416 DOI: 10.1021/ma00210a048.

28. Amiya, T.; Hirokawa, Y.; Hirose, Y.; Li, Y.; Tanaka, T., Reentrant Phase-Transition of N-Isopropylacrylamide Gels in Mixed-Solvents. *J. Chem. Phys.* **1987,** *86* (4), 2375-2379 DOI: 10.1063/1.452740.

29. Hirotsu, S., Critical-Points of the Volume Phase-Transition in N-Isopropylacrylamide Gels. *J. Chem. Phys.* **1988,** *88* (1), 427-431 DOI: 10.1063/1.454619.

30. Katayama, S.; Hirokawa, Y.; Tanaka, T., Reentrant Phase-Transition in Acrylamide-Derivative Copolymer Gels. *Macromolecules* **1984,** *17* (12), 2641-2643 DOI: 10.1021/ma00142a032.

31. Kojima, H.; Tanaka, F.; Scherzinger, C.; Richtering, W., Temperature dependent phase behavior of PNIPAM microgels in mixed water/methanol solvents. *J. Polym. Sci., Part B: Polym. Phys.* **2013,** *51* (14), 1100-1111 DOI: 10.1002/polb.23194.

32. Tanaka, F.; Koga, T.; Kojima, H.; Xue, N.; Winnik, F. M., Preferential Adsorption and Co-nonsolvency of Thermoresponsive Polymers in Mixed Solvents of Water/Methanol. *Macromolecules* **2011,** *44* (8), 2978-2989 DOI: 10.1021/ma102695n.

33. Tanaka, F.; Koga, T.; Winnik, F. M., Temperature-Responsive Polymers in Mixed Solvents: Competitive Hydrogen Bonds Cause Cononsolvency. *Phys. Rev. Lett.* **2008,** *101* (2), 028302 DOI: 10.1103/PhysRevLett.101.028302.

34. Fukai, T.; Shinyashiki, N.; Yagihara, S.; Kita, R.; Tanaka, F., Phase Behavior of Co-Nonsolvent Systems: Poly(N-isopropylacrylamide) in Mixed Solvents of Water and Methanol. *Langmuir* **2018,** *34* (9), 3003-3009 DOI: 10.1021/acs.langmuir.7b03815.

35. Heilshorn, S. C.; DiZio, K. A.; Welsh, E. R.; Tirrell, D. A., Endothelial cell adhesion to the fibronectin CS5 domain in artificial extracellular matrix proteins. *Biomaterials* **2003,** *24* (23), 4245-4252 DOI: 10.1016/S0142-9612(03)00294-1.

36. Hayashi, H.; Nishikawa, K.; Iijima, T., Small-angle x-ray scattering study of fluctuations in 1-propanol-water and 2-propanol-water systems. *The Journal of Physical Chemistry* **1990,** *94* (21), 8334-8338 DOI: 10.1021/j100384a062.

37. Bischofberger, I.; Calzolari, D. C. E.; Trappe, V., Co-nonsolvency of PNiPAM at the transition between solvation mechanisms. *Soft Matter* **2014,** *10* (41), 8288-8295 DOI: 10.1039/c4sm01345j.

38. Tanaka, F.; Koga, T.; Winnik, F. M., Competitive Hydrogen Bonds and Cononsolvency of Poly(N-isopropylacrylamide)s in Mixed Solvents of Water/Methanol. *Prog Coll Pol Sci S* **2009,** *136*, 1-7 DOI: 10.1007/2882_2009_1.

39. Li, N. K.; Quiroz, F. G.; Hall, C. K.; Chilkoti, A.; Yingling, Y. G., Molecular Description of the LCST Behavior of an Elastin-Like Polypeptide. *Biomacromolecules* **2014,** *15* (10), 3522-3530 DOI: 10.1021/bm500658w.

40. Lambermont-Thijs, H. M. L.; van Kuringen, H. P. C.; van der Put, J. P. W.; Schubert, U. S.; Hoogenboom, R., Temperature Induced Solubility Transitions of Various Poly(2-oxazoline)s in Ethanol-Water Solvent Mixtures. *Polymers-Basel* **2010,** *2* (3), 188-199 DOI: 10.3390/polym2030188.

41. Muiznieks, L. D.; Weiss, A. S.; Keeley, F. W., Structural disorder and dynamics of elastin. *Biochemistry and Cell Biology* **2010,** *88* (2), 239-250 DOI: 10.1139/O09-161.

42. Mammi, M.; Gotte, L.; Pezzin, G., Evidence for Order in the Structure of α-Elastin. *Nature* **1968,** *220*, 371 DOI: 10.1038/220371b0.

43. Urry, D. W.; Long, M. M.; Ohnishi, T.; Jacobs, M., Circular dichroism and absorption of the polytetrapeptide of elastin: A polymer model for the β-turn. *Biochem. Biophys. Res. Commun.* **1974,** *61* (4), 1427-1433 DOI: 10.1016/S0006-291X(74)80442-0.

44. Nuhn, H.; Klok, H.-A., Secondary Structure Formation and LCST Behavior of Short Elastin-Like Peptides. *Biomacromolecules* **2008,** *9* (10), 2755-2763 DOI: 10.1021/bm800784y.

45. Liu, M.; Bian, F.; Sheng, F., FTIR study on molecular structure of poly(N-isopropylacrylamide) in mixed solvent of methanol and water. *Eur. Polym. J.* **2005,** *41* (2), 283-291 DOI: 10.1016/j.eurpolymj.2004.09.008.

46. Liu, L. D.; Wang, T.; Liu, C.; Lin, K.; Liu, G. M.; Zhang, G. Z., Specific Anion Effect in Water-Nonaqueous Solvent Mixtures: Interplay of the Interactions between Anion, Solvent, and Polymer. *J. Phys. Chem. B* **2013,** *117* (37), 10936-10943 DOI: 10.1021/jp406215c.

47. Cho, Y.; Zhang, Y.; Christensen, T.; Sagle, L. B.; Chilkoti, A.; Cremer, P. S., Effects of Hofmeister Anions on the Phase Transition Temperature of Elastin-like Polypeptides. *J. Phys. Chem. B* **2008,** *112* (44), 13765-13771 DOI: 10.1021/jp8062977.

48. Trabbic-Carlson, K.; Meyer, D. E.; Liu, L.; Piervincenzi, R.; Nath, N.; LaBean, T.; Chilkoti, A., Effect of protein fusion on the transition temperature of an environmentally responsive elastin-like polypeptide: a role for surface hydrophobicity? *Protein Eng. Des. Sel.* **2004,** *17* (1), 57-66 DOI: 10.1093/protein/gzh006.

Chapter 5 : Protein Purification by Ethanol-Induced Phase Transitions of the Elastin-Like Polypeptide (ELP)

5.1 Abstract

Thermoresponsive elastin-like polypeptide (ELP) tags provide an efficient and highly scalable route to purifying recombinantly expressed proteins without the need for chromatography. By changing temperature and/or salt content, ELPs can be selectively precipitated and resolubilized, which enables the purification of the fused target protein. Despite its advantages, the method often requires addition of molar level salts for precipitation, making it difficult to control the final product salinity. Control over product salinity is especially important when fabricating protein-based materials, such as hydrogels, which often contain protein concentrations $\geq$ 10 wt% in their final state. One way to circumvent the issue of high-salt protein solutions is to take advantage of ELP cononsolvency in water/ethanol solvents. This work investigates ethanol-induced ELP precipitations as a route to purify and desalt ELP-tagged proteins. Superfolder green fluorescent protein (sfGFP) tagged with an ELP (ELP-sfGFP) was used as a model protein to develop this ethanol-based purification process. A final protein product with high purity and low salinity was obtained after one cycle of sodium chloride (NaCl)-induced precipitation followed by one cycle of ethanol-induced precipitation. ELP-sfGFP purified using this method was found to have 0.9 $\pm$ 0.4 mM and 1.48 $\pm$ 0.05 mM equivalents of NaCl when purified from liter-scale and well-plate scale expressions, respectively. Achieving low salinities from well-plate expression is particularly useful as this format does not permit use of other desalting techniques such as dialysis. However, the procedure is only viable for proteins that are robust to incubation in moderate concentrations of ethanol. The method's applicability in generating protein materials was demonstrated by the well-plate scale purification and desalting of an ELP-based protein hydrogel. This new process for selectively precipitating and desalting ELP-tagged proteins will enable high-throughput screening of protein materials that require tight control of buffer conditions.

5.2 Introduction

Purification of recombinantly-expressed proteins is critical in many applications of interest at the industrial and laboratory scale. Affinity-based chromatography is a common strategy for purifying proteins of interest.[1-2] In these purification strategies, a protein or peptide tag is genetically fused to the protein of interest, and the affinity between the tag and its binding partner is used to selectively immobilize, and subsequently elute, the protein of interest on a column containing a

resin functionalized with an affinity tag.[3] These methods are capable of producing high-purity protein products with excellent yield.[4-6] However, affinity chromatography-based methods can prove challenging with respect to scalability. In small-volume, high-throughput formats, when many different proteins are purified separately, the cost per purification can be high, as an individual column/resin setup is required for each protein of interest.[7-9] At the industrial scale, the cost of large volumes of resin can become prohibitively expensive, especially for lower profit margin protein products.[10-11]

Meyer and Chilkoti showed that an elastin-like polypeptide (ELP) tag can be used to enable scalable protein purification in a process that is both low-cost and high-throughput.[12-13] ELPs are a class of highly-repetitive polypeptides that undergo a reversible, temperature-dependent inverse phase transition in aqueous solution.[14] That is, ELPs are insoluble at high temperatures and soluble at low temperatures. The temperature at which a given ELP becomes insoluble corresponds to its transition temperature (T_t), and is determined by the specific ELP sequence and resulting hydrophobicity. Meyer and Chilkoti showed that the temperature-dependent solubility of these ELPs could be used to purify an ELP-tagged protein of interest by a method called inverse transition cycling (ITC). In ITC, complex mixtures of proteins can be heated and cooled above and below the T_t of the ELP to selectively precipitate and then re-solubilize the ELP-tagged protein. A schematic of this process, which uses heat as a precipitant, is shown in **Figure 5-1**a. Because separation of the target protein from contaminants only requires heating, cooling, and centrifugation, ITC is simpler and more economical to implement on a high-throughput scale than other common protein purification methods. ITC has been used to purify a variety of ELP-tagged proteins, including thioredoxin, mCherry, organophosphorus hydrolase (OPH), and even antibodies.[15-20]

Precipitation of ELP-tagged proteins in ITC is typically not performed using heat as the sole precipitant. In most ITC protocols, the T_t of ELP-tagged proteins is adjusted by adding a salt, such as sodium chloride, to the protein solution. The addition of salt depresses the T_t of the ELP to a range that keeps the target protein folded and is also within range of lab-scale process equipment.[21] Molar-level salt concentrations are typically employed to improve the yield of recovered ELP-tagged protein. However, the addition of salt results in uncontrolled levels of residual salt ions in the final protein solution. Residual salt is undesirable in cases where the tight control of buffer conditions is necessary, such as in the preparation of protein materials or other highly concentrated

systems. In these cases, salinity can be lowered in several ways. It is possible to use more kosmotropic salts that have a stronger effect on ELP solubility on a per mole basis; however, stronger kosmotropes are also more prone to precipitate non-target proteins, leading to lower final product purity.[22-23] Salinity can also be adjusted by dialysis, but dialysis can prove unwieldy in high-volume or high-throughput systems. Recent work by VerHeul *et al.* has shown that liquid-liquid extraction using organic solvents can be used to selectively extract and purify ELPs and ELP fusions from crude cell lysates.[24] This strategy represents an attractive alternative for producing pure, desalted ELPs in high-volume systems; however, liquid-liquid extractions can also be challenging in high-throughput formats. Thus, it is of interest to develop a purification procedure that uses precipitations, like ITC, but also permits tight control over final buffer conditions.

To this end, a different feature of ELP solution phase behavior can be exploited— cononsolvency. Cononsolvency is a phenomenon observed in some polymer solutions, in which the polymer is less soluble in a blend of two solvents than it is in either of the individual solvents.[25-26] In the case of ELPs, this cononsolvency behavior has been demonstrated in binary mixtures of water with various alcohols (methanol, ethanol, isopropanol, 1-propanol).[27] In ethanol/water mixtures, for example, it was shown that adding ethanol decreases the T_t of the ELP (constituting a decrease in ELP solubility), but only up to a certain ethanol content. Above this ethanol content, the nature of the ELP phase transition inverts—the ELP becomes insoluble at low temperatures and soluble at high temperatures, indicating the onset of upper-critical solution transition (UCST)-like behavior. These data suggest that it should be possible to purify ELP-tagged proteins with ethanol-induced phase transitions at low temperatures in a manner analogous to ITC.

Here, the use of ethanol-induced ELP precipitations for protein purification and desalting is investigated using an ELP-superfolder green fluorescent protein fusion (ELP-sfGFP, **Figure 5-1**b) as a model protein. First, the degree of ELP-sfGFP partitioning into soluble and insoluble phases is measured at 4 °C and 20 °C across a range of ethanol (EtOH) and sodium chloride (NaCl) concentrations to identify a suitable temperature/ethanol condition for ELP-sfGFP purification. After identifying a precipitation condition (4 °C, 30 vol% EtOH), EtOH-induced precipitation is evaluated on its ability to purify and desalt ELP-sfGFP lysates, as compared to conventional NaCl-induced precipitation. This comparison shows that NaCl-induced precipitation yields higher purity protein products with uncontrolled salinity, while EtOH-induced precipitation produces less-pure

ELP-sfGFP with low, well-controlled salinity. Based on these observations, a second precipitation scheme in investigated, where one cycle of NaCl-induced precipitation is followed by one cycle of EtOH-induced precipitation. The performance of this second purification method is characterized for both liter-scale and well-plate scale protein expressions. The method is then applied to an ELP-OPH fusion protein, which is found to denature during the EtOH precipitation, highlighting that this approach is limited to EtOH-stable fusions. However, the method is found to be effective for high-throughput purification and desalting of ELP-based protein gels, showing the method's promise in high-throughput processing of protein-based materials.

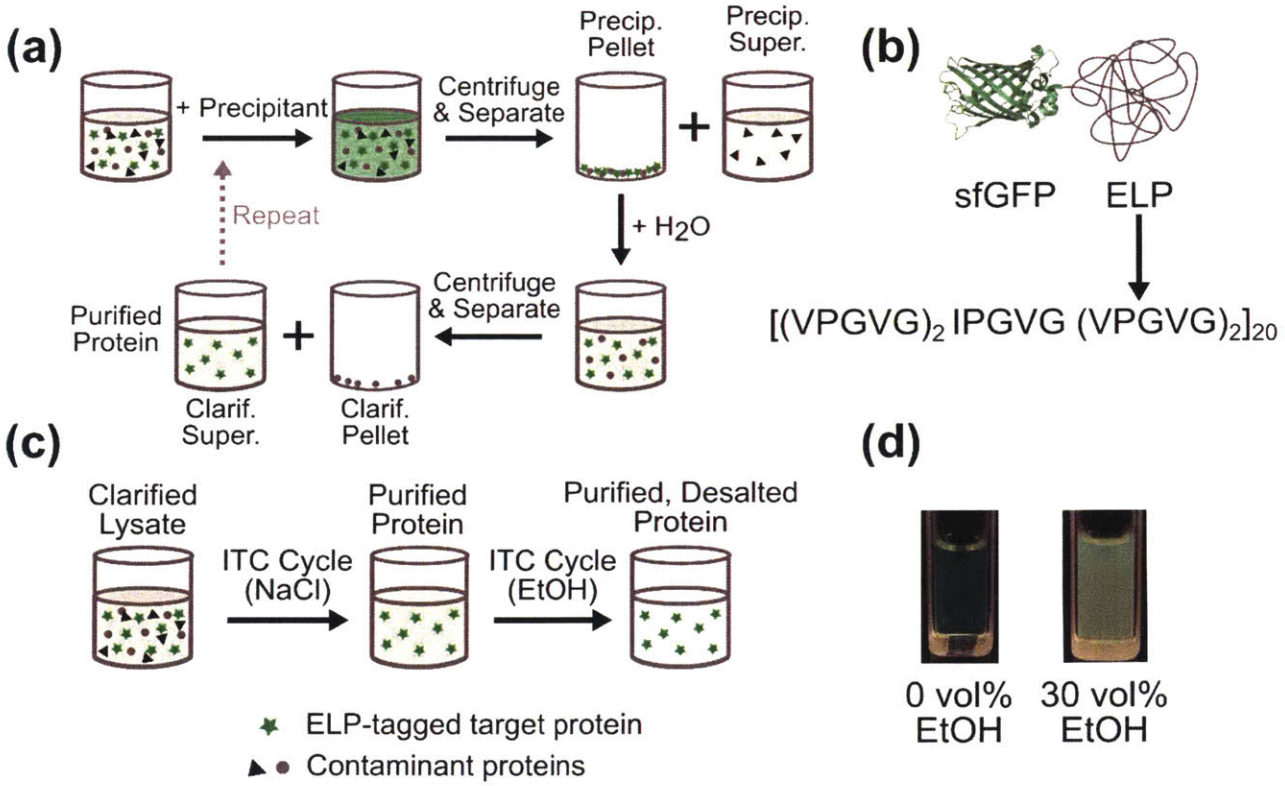


Figure 5-1. (a) Schematic representation of how a single precipitation cycle is used to purify an ELP-tagged protein of interest. (b) Schematic of ELP-sfGFP used as the model protein in this work, including the ELP repeated sequence. (c) Overview of the final process for ELP-tagged protein purification and desalting proposed in this work. (d) Images of ELP-sfGFP in solution at 4 °C with 0 and 30 vol% ethanol, demonstrating the ethanol-induced precipitation of this protein.

5.3 Materials & Methods

Liter-scale expression of ELP-sfGFP. Complete amino acid and DNA sequences for ELP-sfGFP used in this work can be found in **Figure D-1** and **Figure D-2**. For phase diagram and liter-scale

experiments, ELP-sfGFP was expressed in the BL21(DE3) strain of *Escherichia coli* (*E. coli*). Expressions were carried out in 5 L of Lysogeny Broth (LB), supplemented with 50 μg/mL kanamycin in a fermenter with a 7 L working volume. Cells were grown at 37 °C to OD_{600} = 0.7-0.9 and were then induced with 0.5 mM β-D-1-thiogalactopyranoside (IPTG). After induction, the temperature of the fermentation was lowered to 20 °C. Expressions were grown for 16 hours after induction. Cells were harvested by centrifugation and then resuspended in buffer containing 3 mM $MgCl_2$, 1 mM EDTA, 100 mM NaCl, and 10 mM trizma at pH 7.5 (MENT buffer). Resuspended cells were frozen at -80 °C, then thawed and incubated with 1 mg/mL lysozyme at 4 °C for 1-2 h. Cells were sonicated with a tip sonicator, and lysates were clarified by centrifugation at 26,700 xg for 1 h at 4 °C.

Well-plate expression of ELP-sfGFP. For well-plate experiments, ELP-sfGFP was expressed in the BL21(DE3) strain of *E. coli*. A colony from an agar plate was used to inoculate a 10 mL starter culture containing LB supplemented with 50 μg/mL kanamycin. The starter culture was grown to stationary growth phase at 37 °C overnight. Expressions were carried out in a sterile deep-well 96-well plate, where each well was prepared with 1480 μL LB supplemented with 50 μg/mL kanamycin and one 3-5 mm diameter glass bead to improve mixing. Wells were inoculated with 20 μL of the starter culture and plates were incubated at 37 °C with orbital shaking at 300 rpm. OD_{600} of well-plate cultures was monitored by removing 150 μL from 8 wells in the expression plate into a sterile 96-well microtiter plate, measuring absorbance, and returning the 150 μL to the deep-well expression plate. Once the average OD_{600} of the measured wells reached ~0.8, plates were put on ice for 20 min before induction by addition of IPTG to 0.5 mM. After induction, plates were incubated at 20 °C for 18-23 h. Cells were harvested by centrifugation (4,000 xg, 15 min), and media was removed via aspiration with a multichannel pipettor. Cell pellets were frozen at -20 °C for at least 12 h, then thawed. Cells were resuspended in 140 μL of MENT buffer (3 mM $MgCl_2$, 1 mM EDTA, 100 mM NaCl, 10 mM trizma; pH 7.5) per well and were lysed by incubation with 1 mg/mL lysozyme, ~0.1 mg/mL DNase, and ~0.1 mg/mL RNase at 37 °C for 3 h, followed by a second freeze-thaw cycle at -20 °C. Cell lysates were thawed and transferred to 96-well microtiter plates. Lysates were clarified by centrifugation at 6,000 xg for 1 h at 4 °C.

Protein purification. To compare purification methods, clarified lysates containing target proteins were purified by either two consecutive NaCl-induced precipitation cycles, two consecutive

ethanol-induced precipitation cycles, or one NaCl-induced cycle followed by one ethanol-induced precipitation cycle.

Precipitations with NaCl were carried out as follows. 5 M NaCl in water was added to the protein solution to bring the total NaCl concentration to 1.5 M. The samples were then incubated at 37 °C overnight to ensure complete precipitation of the protein. Precipitant was collected by centrifugation at 37 °C (1 h, 26,700 xg for liter-scale expressions or 6,000 xg for well-plate expressions), and supernatant was removed. Pellets were resuspended in Milli-Q water at 4 °C for at least 3 h to ensure complete redissolution of the desired protein product. The resuspension was then clarified by centrifugation at 4 °C (1 h, 26,700 xg for liter-scale expressions or 6,000 xg for well-plate expressions), and the supernatant was collected. This supernatant constituted the final product of a single NaCl-induced purification cycle.

For precipitations with ethanol (EtOH), pure EtOH was added to the protein solution to a final concentration of 30 vol% EtOH, and samples were incubated at 4 °C for 16 h to allow complete precipitation of the protein. Precipitate was separated by centrifugation at 4 °C (1 h, 26,700 xg for liter-scale expressions or 6,000 xg for well-plate expressions), and supernatant was discarded. The pellets were resuspended in Milli-Q water at 4 °C for at least 3 h, after which the suspension was clarified by centrifugation (1 h, 26,700 xg for liter-scale expressions or 6,000 xg for well-plate expressions). The purified supernatant was transferred to a new tube or well plate, constituting the final product of a single EtOH-induced purification cycle.

ELP-sfGFP purification for phase diagram construction. Clarified cell lysate containing ELP-sfGFP fusion protein was purified using two NaCl-induced precipitations as described above. The protein was dialyzed extensively against water using a membrane with a 6-8 kDa molecular weight cutoff. sfGFP fluorescence (excitation 395 nm, emission 510 nm) was used to measure relative concentration. The concentration of ELP-sfGFP solutions was adjusted to reflect 5 times the typical fluorescence produced by a 1.5 mL expression in a well-plate format.

Phase diagram construction. Protein samples were prepared at the specified ethanol and sodium chloride concentrations using stock solutions of pure ethanol, 5 M NaCl in water, 0.5 M NaCl in water, and purified, dialyzed ELP-sfGFP. Reagents were added to 200 μL per well in a 96-well microtiter plate in the following order: water, ELP-sfGFP protein stock, NaCl, and then EtOH. Three replicate plates were prepared for each temperature condition. The same volume of ELP-

sfGFP (40 μL) was added to each well, and the volumes of other reagents were adjusted to achieve the desired sodium chloride and ethanol concentrations. Plates were sealed with Parafilm and enclosed in an airtight box to mitigate solvent evaporation. Plates were incubated for 12 h at either 4 °C or 20 °C to ensure complete precipitation of the ELP-sfGFP. Soluble and insoluble fractions were separated by centrifugation of the plates at 6,000 xg at the relevant temperature (4 °C or 20 °C) for 1 h. Supernatants were removed to a new plate, and the protein concentration of this soluble fraction was determined immediately (to mitigate solvent evaporative effects) by bicinchoninic acid (BCA) assay (Pierce) following the manufacturer's instructions. The pelleted insoluble fractions were resuspended in 200 μL Milli-Q water by mixing with a multichannel pipette followed by 12 h incubation at 4 °C. Protein concentration of these resolubilized pellets was then determined by BCA assay.

The effects of ethanol and NaCl on the BCA assay chemistry were accounted for by generating 9 standard curves at the time of each experiment. Each curve was generated using 9 concentrations of α-amylase prepared in 0, 30, or 60 vol% EtOH and 0, 100, or 500 mM NaCl. Standard curves for this experiment were generated with α-amylase because the dilutions required to adjust solvent conditions for different standard curves required preparation of protein solutions that were above the BSA standard concentration provided by the manufacturer. The α-amylase standards were prepared by weighing purchased α-amylase (Sigma-Aldrich #10065) into a 15 mL conical tube and hydrating to ~10 mg/mL with MilliQ water. The α-amylase solution was then dialyzed against MilliQ water (3-4 kDa molecular weight cutoff membrane, 7 changes of MilliQ water separated by at least 3 hours) to remove any residual salt. The concentration of the α-amylase after dialysis was measured by BCA using the BSA standard provided by the manufacturer. α-Amylase solutions were then diluted with MilliQ water, 5 M NaCl, and EtOH to produce solutions of 0.28 mg/mL α-amylase in the nine different solvent conditions of interest (all permutations of 0, 30, 60 vol% EtOH and 0, 100, 500 mM NaCl). Standard curves in the various solvents of interest were generated by performing serial dilutions on these 0.28 mg/mL α-amylase solutions. Each standard curve was generated in triplicate and fit to a quadratic function as suggested in the manufacturer's manual. All calibration curves and fit parameters can be found in Figures S6-S8. The protein concentration in solvents that did not have a standard curve (for example, 20 vol% EtOH, 50 mM NaCl) was calculated by bilinearly interpolating the concentrations predicted by the calibration curves at the conditions bracketing the solvent condition of interest. For example, the concentration

of protein at 20 vol% EtOH, 50 mM NaCl would be calculated by interpolating between the concentrations predicted at 0 vol% EtOH/0 mM NaCl, 0 vol% EtOH/100 mM NaCl, 30 vol% EtOH/0 mM NaCl, and 30 vol%/100 mM NaCl. The unique composition of ELPs can cause anomalous behavior in dye-binding assays such as BCA. However, while the absolute concentration predicted by BCA may be inaccurate due to the ELP tag, the relative response of the ELP-sfGFP in this assay should still be proportional to its concentration. For this reason, ELP-sfGFP concentrations calculated from these calibration curves (Table S1) are reported as normalized values here. Concentrations were normalized to the sum of the calculated ELP-sfGFP concentration in soluble and insoluble fractions in water at the relevant temperature (4 °C or 20 °C).

Yield measurements. Expression yield was measured for all proteins by BCA assay following the manufacturer's instructions using provided bovine serum albumin (BSA) standards. ELP-sfGFP fluorescence was measured using a Tecan Infinite M200 Pro plate reader with an excitation wavelength of 395 nm and an emission wavelength of 510 nm.

Circular dichroism. All samples measured by circular dichroism (CD) were dialyzed against water prior to measurement and centrifuged (26,700 xg, 15 minutes, 4 °C) to remove insoluble aggregates that formed during dialysis. The concentration of the final product was then adjusted to 0.5 mg/mL with MilliQ water. Protein samples (0.5 mg/mL) were loaded into a quartz cuvette with a 1 mm path length and sealed. Scans were performed in a JASCO Model J-1500 circular dichroism spectrometer equipped with a Peltier temperature controller. Spectra were collected at 10 °C with 0.5 nm resolution, bandwidth of 1 nm, and scan speed of 50 nm per minute. Data were converted to mean residue molar ellipticity based on the sample concentration (0.5 mg/mL) and protein molecular weight.

Conductivity. Protein was removed from conductivity samples using centrifugal ultrafiltration (10 kDa MWCO). Filtrates were then measured with a Mettler Toledo SevenCompact^{TM} Duo S213 conductivity probe. The probe was calibrated using a NIST traceable 1413 μS/cm standard (692 ppm as NaCl). To convert data into molar equivalents of NaCl, solutions containing 0, 10 mM, 50 mM, and 100 mM sodium chloride were prepared in triplicate, and their conductivities were measured to construct a conversion curve, which showed a linear relation between NaCl concentration and conductivity (Figure S9).

Dynamic light scattering. Dynamic light scattering (DLS) measurements were performed directly on protein purification fractions of interest with no filtration or concentration adjustment. Measurements were performed on a Wyatt Möbiuζ DLS and Zeta Potential instrument with a 532 nm laser. Each measurement was averaged over ten 5 s acquisitions. Size distributions reported were produced by performing fits to autocorrelation curves using the Regularization algorithm in the Wyatt Dynamics software.

Well-plate expression and purification of ELP-based protein gel. The ELP-based protein gel, here called P10, was expressed in the Tuner (DE3) strain of *E. coli*. Sequence information for this protein can be found in previously published work, where it is called $([I^{0.6}V^{0.4}]PAVG)_{50}$.[28] Cells were grown and lysed as described above for the well-plate expression of ELP-sfGFP, with the modification that lysates were allowed to cool at 4 °C for 16 h before the second freeze-thaw cycle, to prevent P10 from aggregating.

P10 was purified using either one cycle of NaCl- followed by one cycle of EtOH-induced purification (NaCl-EtOH method), or two cycles of NaCl- followed by one cycle of EtOH-induced purification (NaCl-NaCl-EtOH method). To achieve a high enough protein concentration to form gels, groups of three wells of clarified lysate were combined, and 5 M NaCl was added to a final concentration of 1.5 M (final volume 600 μL). Plates were then incubated at 37 °C for 12 h. Pellets were separated by centrifugation (6,000 xg, 1 h, 37 °C) and resuspended in 105 μL of Milli-Q water at 4 °C for 3 h. Soluble protein was separated by centrifugation (6,000 xg, 1 h, 4 °C) followed by aspiration with a multichannel pipettor. Six product wells from this first NaCl-induced precipitation cycle were then combined into a single well to yield a volume of 630 μL. For the NaCl-NaCl-EtOH samples, a second NaCl-induced purification cycle was performed on this 630 μL volume, again using 105 μL of MilliQ water for pellet resuspension. For the final purification cycle by EtOH-induced precipitation, ethanol was added to 30 vol% (final volume of 900 μL for NaCl-EtOH samples, or 150 μL for NaCl-NaCl-EtOH samples). Plates were incubated for 12 h at 4 °C. Precipitated proteins were separated by centrifugation (6,000 xg, 1 h, 4 °C) and subsequently resuspended in 105 μL Milli-Q water for at least 3 h at 4 °C. Insoluble aggregates were removed by centrifugation (6,000 xg, 1 h, 4 °C). Purified protein products were transferred to weighed Eppendorf tubes, combining three wells per tube. Samples were then frozen and lyophilized. Lyophilized protein was resuspended in 9 μL water to form solutions capable of reversible temperature-dependent gelation.

5.4 Results & Discussion

Ethanol-induced phase transitions of ELPs provide a viable thermodynamic basis for protein purification, both for ELPs and for globular proteins with fused ELP tags. A protein composed of an ELP fused to superfolder green fluorescent protein (ELP-sfGFP, **Figure 5-1**b) is used as a model system to assess the feasibility of a purification process that utilizes ethanol (EtOH)-induced phase transitions. Previous work by Trabbic-Carlson *et al.* demonstrated that the transition temperature, and thus the phase behavior, of a given ELP can be altered when fused to a globular protein.[17] To understand how globular protein fusion changes the cononsolvency behavior of the ELP, the phase behavior of ELP-sfGFP is compared to previously published data on the ELP alone.[27] sfGFP has been shown to be both soluble and stable at all the ethanol contents at which ELP cononsolvency occurs.[29] Thus, differences in ELP and ELP-sfGFP cononsolvency can be attributed to the modulation of ELP phase behavior by the sfGFP fusion partner. ELP-sfGFP partitioning into soluble and insoluble phases is measured as a function of the concentrations of ethanol (EtOH) and sodium chloride (NaCl) in the solvent at two temperatures, 4 °C and 20 °C. This solubility information is used to identify a suitable condition for ethanol-induced ELP-sfGFP precipitation. An ideal EtOH/temperature condition is one that precipitates the ELP-sfGFP over a wide range of salt concentrations, as it is not always possible to tightly control the salt content during the protein purification process.

The phase behavior of ELP-sfGFP as a function of ethanol concentration is similar to that of the ELP alone, but with a larger apparent insoluble region. At 4 °C, the ELP alone is soluble from 0-22.5 vol% (0-8.2 mol%) ethanol. At 4 °C, the ELP-sfGFP is predominantly soluble (**Figure 5-2**a) in the absence of NaCl and at 0 and 10 vol% (0 and 3.3 mol%) ethanol, but at 20 vol% (7.2 mol%) ethanol, half of the ELP-sfGFP partitions into the insoluble phase. Similarly, at 20 °C, the ELP alone is completely soluble at 20 vol% (7.2 mol%) ethanol, but at these conditions ELP-sfGFP partitions predominantly into the insoluble phase. Also at 20 °C, the ELP alone is completely soluble at 35 and 40 vol% (14.3 and 17.1 mol%, respectively) ethanol. In contrast, the ELP-sfGFP is predominantly segregated into the insoluble phase at 20 °C for both of these ethanol concentrations (35 and 40 vol% ethanol). These observations indicate that the fusion of the sfGFP to the ELP shifts the phase boundaries of the ELP. However, the trends in phase behavior are still similar—the ELP-sfGFP becomes less soluble at intermediate ethanol contents and then completely resolubilizes at 60 vol% ethanol at both 4 °C and 20 °C.

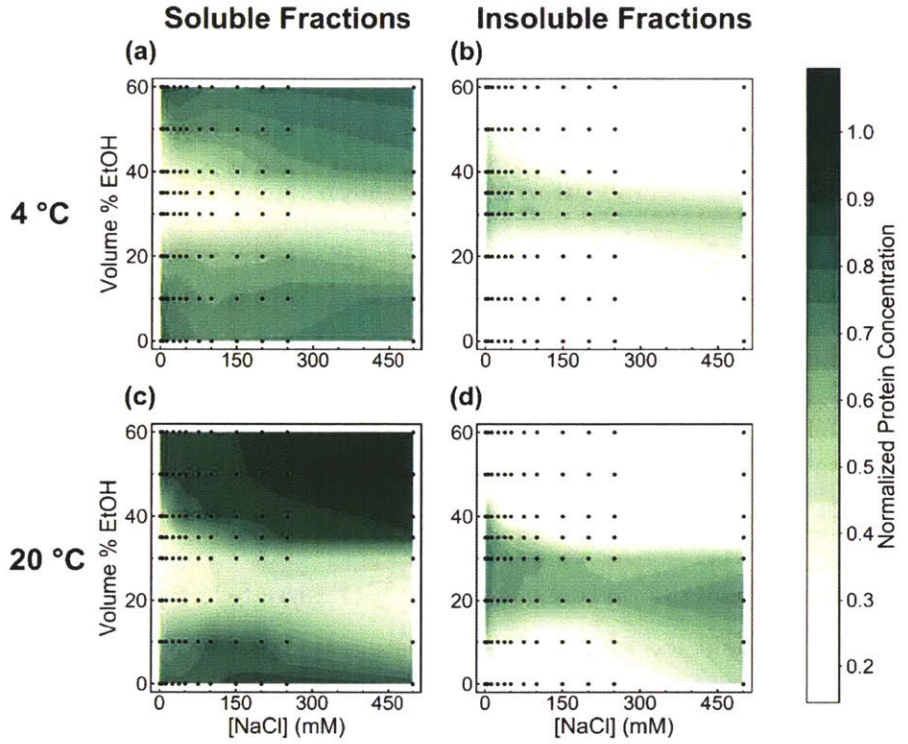


Figure 5-2. ELP-sfGFP partitions into soluble and insoluble fractions after precipitation and centrifugation at various ethanol and sodium chloride concentrations. Panels (a) and (b) correspond to precipitation and centrifugation at 4 °C, and panels (c) and (d) correspond to precipitation and centrifugation at 20 °C. Protein concentration data were taken for both the soluble (a, c) and insoluble (b, d) fractions. The black points indicate solution conditions at which data were collected. Between these points, the contours correspond to linearly interpolated estimates of protein concentration in the indicated phase. Raw protein concentration data used to produce these contour plots are reported in **Table D-1**, and the standard deviations across replicate measurements for each data point are plotted in **Figure D-5** and values are reported in **Table D-2**.

At and below NaCl concentrations of 200 mM, the effects of NaCl on ELP-sfGFP solubility are different from those observed for the ELP alone. At NaCl concentrations of 200 mM and lower, the addition of NaCl to the ELP-sfGFP solution increases the solubility of the ELP-sfGFP, regardless of EtOH concentration, at both 4 ˚C and 20 ˚C. In ELP-only solutions, an opposite trend is observed. For 10 and 20 vol% EtOH solutions (3.3 and 7.2 mol%, respectively), the addition of NaCl leads to a monotonic decrease in ELP solubility. In 40 vol% EtOH solutions (17.1 mol%), the solubility of the ELP alone decreases with increasing NaCl content up to a NaCl concentration

of 100 mM; at 150 and 200 mM NaCl, the solubility of the ELP increases again, albeit very slightly, with the addition of NaCl. With the exception of this 40 vol% EtOH/150-200 mM NaCl solvent condition, all of the reported trends in ELP solubility with respect to NaCl concentration are opposite to those observed for ELP-sfGFP. This observation suggests that the sfGFP dominates the system response when the solvent contains 5-200 mM NaCl. This behavior is likely because the addition of NaCl helps to solubilize sfGFP (which has a charged surface), rendering the entire fusion protein more soluble with the addition of salt in this regime.[30] Such an increase in sfGFP solubility could solubilize the entire protein or cause the ELP-sfGFP to form soluble micelles, such as those reported to form in certain ELP-thioredoxin, ELP-mCherry, and ELP-fibronectin type III domain fusions. [20, 31-32]

Based on the results of the ELP-sfGFP phase behavior study (**Figure 5-2**), a precipitation condition of 30 vol% EtOH at 4 °C was selected for further purification studies. This precipitation condition was chosen principally for its insensitivity to the salt content in solution. The results of the ELP-sfGFP phase behavior study indicate that at 20 °C, there is no single ethanol concentration at which the ELP-sfGFP solubility is completely unchanged by the addition of ≤ 500 mM NaCl. In contrast, at 30 vol% EtOH and 4 °C, the ELP-sfGFP is predominantly in the insoluble phase across all NaCl concentrations tested. These data suggest that, at this condition, the precipitation of ELP-sfGFP will not be sensitive to the salt content in solution when performing selective precipitation of the protein. Additionally, the ethanol used for precipitation is less prone to evaporation at 4 °C, which is a pertinent concern when this process is performed in well-plate formats where evaporation is more difficult to control.

It is expected that a precipitation condition of 30 vol% EtOH at 4 °C would work for a variety of fusion proteins, based on the fact that both the ELP alone and ELP-sfGFP are completely insoluble at this condition. However, the fusion of different peptide tags and globular proteins to ELPs is known to shift ELP phase behavior in aqueous solution differently.[33] Thus, this condition may not hold for all fusion partners of interest. For new ELP-protein fusions, optimal purification conditions can be selected in the following manner. After lysate clarification, or after one cycle of NaCl-induced precipitation, small samples of protein solution should be incubated with a variety of ethanol concentrations (0-40 vol%) at 4 °C. Then, precipitated and soluble proteins should be separated by centrifugation and analyzed by SDS-PAGE to determine which ethanol concentration produced the highest yield of desired protein product in the insoluble pellet fraction.

The feasibility of ethanol-induced precipitations in ELP-sfGFP purification was assessed by comparing the performance of the new precipitant, ethanol, to that of NaCl, which is commonly used in ELP-based purification schemes.[12-13] The process, as shown in Figure 1a, was performed as follows. Either 1.5 M NaCl at 37 ˚C or 30 vol% EtOH at 4 ˚C was added to a clarified lysate containing ELP-sfGFP, causing the ELP-sfGFP to selectively precipitate out of solution while leaving other contaminant proteins soluble. The precipitate was separated from the soluble fraction by centrifugation. The precipitate was resuspended in Milli-Q water at 4 ˚C to selectively re-solubilize the ELP-sfGFP. The sample was centrifuged again to separate the ELP-sfGFP, now soluble, from any insoluble protein contaminants. This process constituted one cycle of either NaCl- or EtOH-induced purification. To further purify the protein of interest, the process was repeated starting with the partially purified product instead of clarified lysate, resulting in a total of two purification cycles. Processes with EtOH as the precipitant, with NaCl as the precipitant, and with NaCl followed by EtOH as precipitant were examined. To determine differences between the NaCl- and EtOH-induced precipitations, several performance criteria were evaluated, including protein yield (by sfGFP fluorescence), product loss (by SDS-PAGE), and desalting (by solution conductivity). These performance criteria were measured using at least three replicate experiments, each from a separate protein expression. Experiments were performed at both the liter-scale and well-plate scale. To compare yield across replicates, sfGFP fluorescence for each replicate experiment was normalized to the fluorescence of the clarified lysate from that replicate for liter-scale expressions) or the average fluorescence of the clarified lysate over 12 replicate wells (for well-plate expressions).

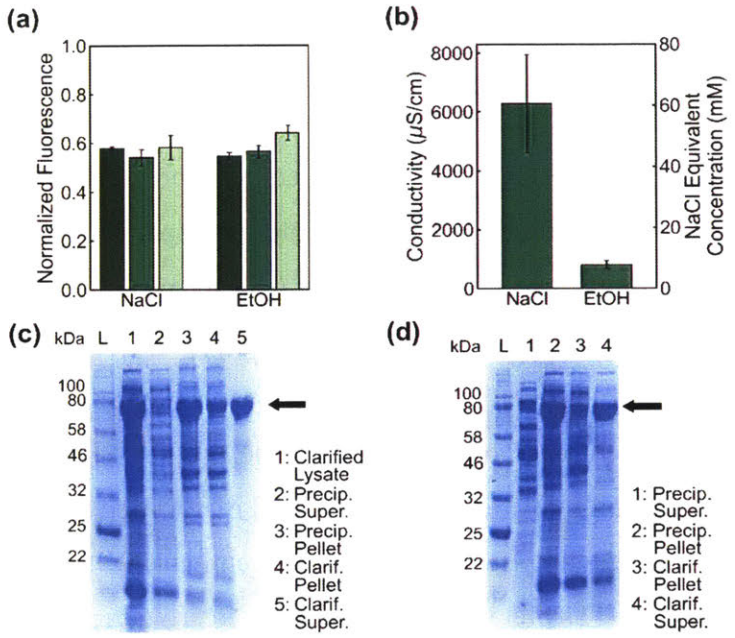


Figure 5-3. (a) Normalized fluorescence of samples after one cycle of NaCl- or EtOH-induced precipitation. Fluorescence intensity is normalized to clarified lysate fluorescence. Each bar represents the average of three fluorescence measurements on the same purified protein product, and each different bar color represents protein purified from a separate protein expression experiment. Error bars correspond to one standard deviation from the mean, calculated from the three replicate fluorescence measurements taken on a given sample. (b) Conductivity of samples after NaCl- or EtOH-induced precipitation, averaged over three replicate protein expression experiments; error bars represent one standard deviation from the mean, calculated over three replicate protein expression experiments. (c, d) Coomassie-stained SDS-PAGE gel showing fractions from either (c) one NaCl-induced precipitation cycle or (d) one EtOH-induced precipitation cycle. The arrow indicates the apparent molecular weight of ELP-sfGFP. Fractions are labeled as in Figure 1a: "Precip." refers to samples after incubation with precipitant (30 vol% EtOH or 1.5 M NaCl) and subsequent centrifugation to separate soluble supernatant ("Super.") and insoluble ("Pellet") fractions. "Clarif." refers to samples after resuspension in water and clarification by centrifugation to separate the soluble supernatant ("Super.") and insoluble ("Pellet") fractions.

Samples purified by one cycle of either EtOH- or NaCl-induced precipitation produced similar yields and product losses, but different purities and significantly different salinities. Purification yields after one cycle, as measured by sfGFP fluorescence, were 57 ± 2% and 62 ± 5% for NaCl- and EtOH-induced precipitations, respectively (**Figure 5-3**a). Similar product losses were also observed for both precipitation methods, as can be seen in lanes 2 and 4 in **Figure 5-3**c for the NaCl-induced precipitation and lanes 1 and 3 in Figure 3d for the EtOH-induced precipitation, where the band corresponding to ELP-sfGFP is indicated by an arrow. While the product losses in these two precipitation methods were similar, the NaCl-induced precipitation yielded a higher purity protein product, as evidenced by the comparison of clarification supernatant fractions for the two precipitants (**Figure 5-3**c lane 5 and **Figure 5-3**d lane 4). Another difference between the two methods was observed in conductivity measurements; one cycle of EtOH-based purification resulted in significantly lower salinity in the purified product than one cycle of NaCl-based purification (**Figure 5-3**b). The solvent conductivities of the NaCl- and EtOH-purified products correspond to 61 ± 16 mM and 8 ± 1 mM equivalents of NaCl, respectively, constituting a >7-fold difference in solvent salinity.

This combination of results indicates that the purification and desalting capabilities of EtOH-induced precipitations and NaCl-induced precipitations differ. NaCl-induced precipitations yield higher purity protein products, while EtOH-induced precipitations yield protein products with well-controlled salinity (**Figure 5-3**). Thus, while an EtOH-induced precipitation cycle does improve the purity of the final product, it is most effective at desalting.

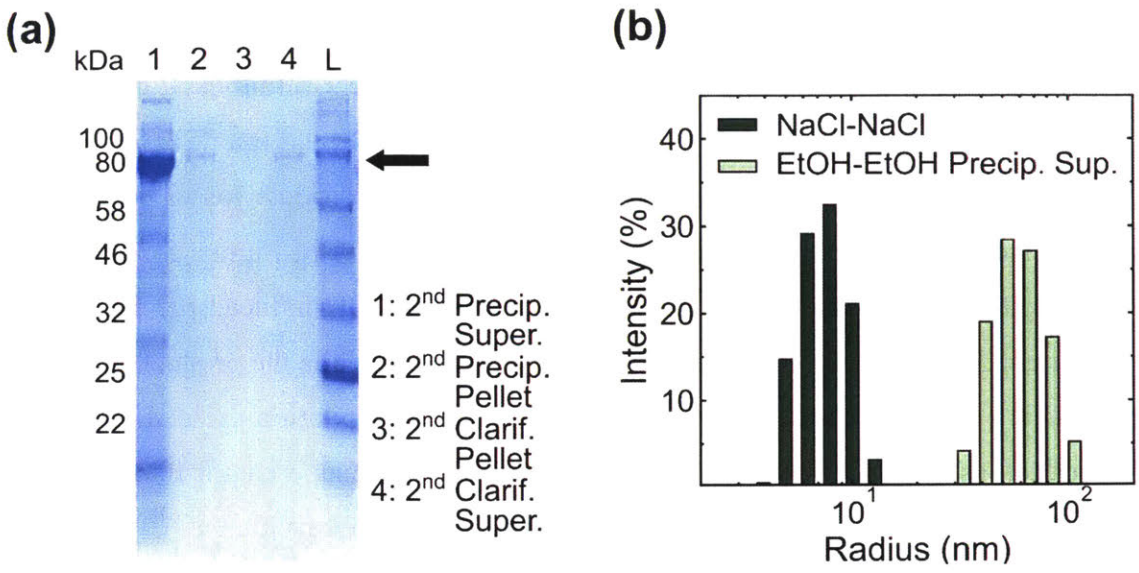


Figure 5-4. Results of attempted ELP-sfGFP purification using two consecutive EtOH-induced precipitations. (a) Coomassie-stained SDS-PAGE gel showing protein content in different fractions produced by a second EtOH-induced precipitation cycle on ELP-sfGFP, labeled as in Figure 1a: "Precip." refers to portion of the cycle where the precipitant (30 vol% EtOH) is added. "Super." and "Pellet" refer to supernatant and pellet, respectively, collected after centrifugation of this precipitant/protein solution. (b) Dynamic Light Scattering (DLS) measurements comparing ELP-sfGFP purified by a conventional two-cycle method (NaCl-NaCl) with unprecipitated ELP-sfGFP resulting from an attempted second EtOH-induced precipitation (EtOH-EtOH Precip. Sup.). The EtOH-EtOH Precip. Sup. sample was analyzed by SDS-PAGE in lane 1 of panel (a).

Previous ELP-based purification methods have shown that multiple cycles of heat- or NaCl-induced precipitation can be used to improve product purity when one cycle is insufficient. Thus, a second consecutive EtOH-induced precipitation was attempted after one EtOH purification cycle had been performed, to explore whether a low-salinity, high purity product could be obtained solely with EtOH-induced precipitations. However, the majority of the ELP-sfGFP remained in the soluble supernatant fraction after the precipitant (30 vol% EtOH, 4 °C) was added (**Figure 5-4**a, lanes 1 and 2), causing a low final purification yield across all replicate experiments (1.4 ± 0.4% by ELP-sfGFP fluorescence). This observation does not agree with the results of the ELP-sfGFP phase diagram. According to conductivity measurements (**Figure 5-3**b), the salinity of the sample after one EtOH cycle and precipitant addition should be equivalent to <10 mM NaCl. The

phase diagram predicts that the ELP-sfGFP should be almost completely insoluble at <10 mM NaCl, 30 vol% EtOH and 4 °C (**Figure 5-2**b), but the ELP-sfGFP partitioned almost entirely into the soluble phase in the purification experiment (**Figure 5-4**a). This discrepancy may arise from differences in the samples used in the purification and phase diagram experiments. First, it is possible that different types of salt (other than NaCl) and small molecule contaminants, which are present in partially purified samples, impact the phase behavior of the ELP-sfGFP and increase its solubility at this condition. A more likely hypothesis is that the discrepancy arises from a difference in protein concentration between the two experiments. While the phase diagram samples contained pure ELP-sfGFP at 0.38 ± 0.04 mg/mL, the purification samples contained 1.8 ± 0.4 mg/mL total protein on average, including ELP-sfGFP and other contaminant proteins.

The second consecutive EtOH-induced precipitation was unsuccessful because ELP-sfGFP formed soluble aggregates instead of precipitating. To determine the aggregation state of the ELP-sfGFP that failed to precipitate, the supernatant of the second ethanol precipitation (where most of the ELP-sfGFP localized, **Figure 5-4**a, lane 1) was measured by dynamic light scattering (DLS) (Figure 4b). The size distribution observed in this sample was compared to that of ELP-sfGFP purified using conventional methods, which is expected to contain free ELP-sfGFP in aqueous solution. The free ELP-sfGFP produced a size distribution with a peak radius of ~8 nm. In contrast, the unprecipitated ELP-sfGFP (in 30 vol% EtOH at 4 °C) formed species at a much higher radius of ~60 nm. The observation of a higher-radius species in the soluble fraction indicates that the ELP-sfGFP formed soluble aggregates instead of insoluble precipitates when a second consecutive EtOH-induced precipitation was attempted. Because the ELP-sfGFP was in a soluble aggregate conformation after the addition of EtOH, it was not possible to collect purified ELP-sfGFP in a precipitated form.

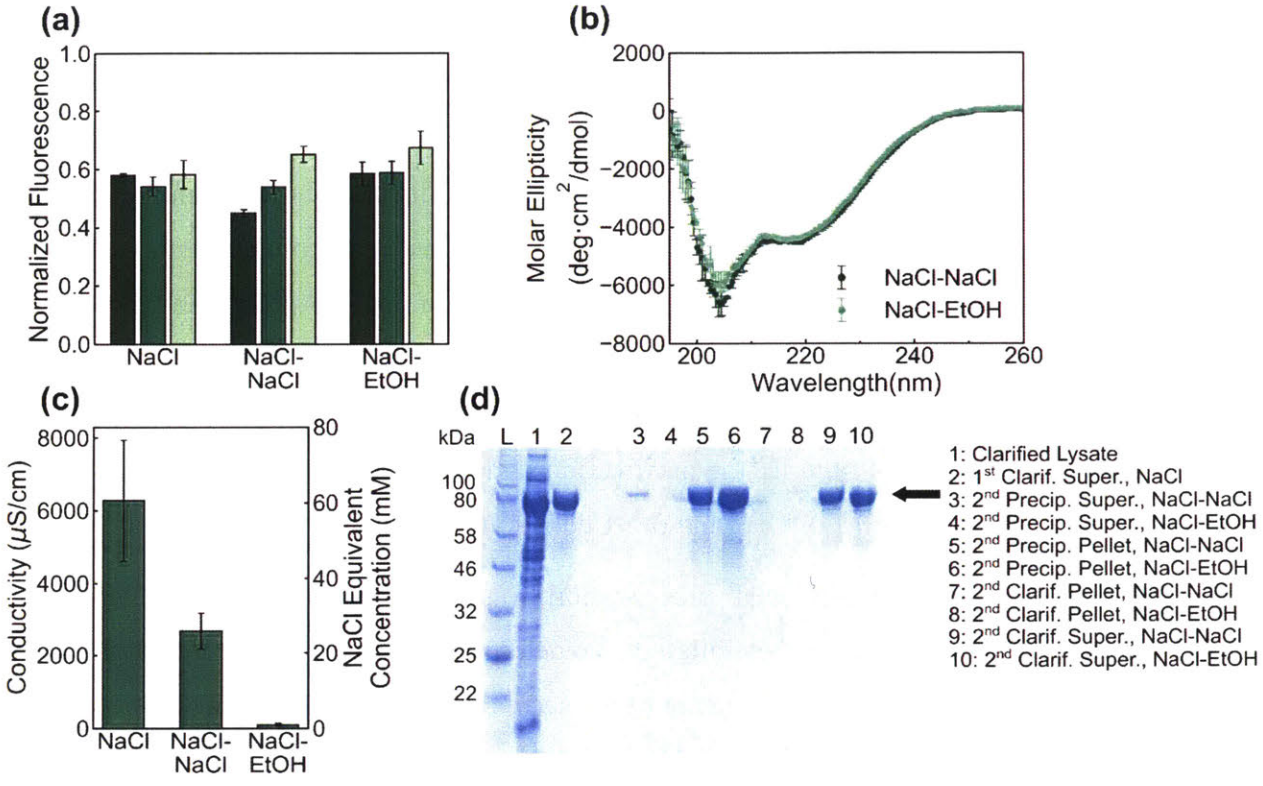


Figure 5-5. (a) Normalized fluorescence of samples after one cycle of NaCl-induced precipitation (NaCl), two cycles of NaCl-induced precipitation (NaCl-NaCl), or one cycle of NaCl- and one of EtOH-induced precipitation (NaCl-EtOH). Fluorescence intensity was normalized to clarified lysate fluorescence. As in **Figure 5-3**a, each bar color corresponds to a separate replicate purification experiment starting from a different fermentation, and the height of each bar corresponds to the average value over three replicate fluorescence measurements. Error bars represent one standard deviation above and below the average, based on the three replicate fluorescence measurements. (b) Circular dichroism measurements of samples purified by either the NaCl-NaCl method or the NaCl-EtOH method. (c) Conductivity of samples after various purification methods, bar heights representing average values over three replicate protein expression experiments. Error bars correspond to one standard deviation over three replicate expression experiments. (d) Coomassie-stained SDS-PAGE gel comparing product losses between the NaCl-NaCl and NaCl-EtOH methods. In both cases, the NaCl-based first cycle had already been performed. The arrow indicates the apparent molecular weight of ELP-sfGFP. In the lane labels, 1st and 2nd refer to the first or second purification cycle performed. Fractions are labeled as in **Figure 5-1**a—"Precip." refers to samples after incubation with precipitant (1.5 M

NaCl or 30 vol% EtOH) and subsequent centrifugation to separate soluble supernatant ("Super.") and insoluble ("Pellet") fractions. "Clarif." refers to samples after resuspension in water and clarification by centrifugation to separate the soluble supernatant ("Super.") and insoluble ("Pellet") fractions.

To achieve a strong desalting effect but still maintain an acceptable product yield, a hybrid method was explored in which one cycle of NaCl-induced precipitation was followed by one cycle of EtOH-induced precipitation (NaCl-EtOH method). The results of the NaCl-EtOH purification scheme were compared to a conventional ELP-based purification scheme in which two cycles of NaCl-induced precipitation were performed (NaCl-NaCl method). The NaCl-EtOH scheme was confirmed to produce yield and purity comparable to the conventional NaCl-NaCl ITC purification. The purification schemes were compared by performing the first NaCl cycle on a large sample, then dividing the semi-purified product, using part of it to perform a second NaCl-based cycle (NaCl-NaCl), and the other part to perform an EtOH-based cycle (NaCl-EtOH). Purification yields using the NaCl-EtOH method were very similar to that of the NaCl-NaCl method. About 60% of clarified lysate fluorescence intensity was retained after the first NaCl-induced precipitation. In the second precipitation cycle, the retained fluorescence remained ~60% regardless of the precipitant used (**Figure 5-5**a). This finding is corroborated by SDS-PAGE, which shows that both methods result in minimal product losses in the second cycle (**Figure 5-5**d, lanes 3, 4, 7, and 8). Circular dichroism (CD) measurements on ELP-sfGFP purified by each method confirm that the use of ethanol as a precipitant does not unfold the sfGFP. As shown in **Figure 5-5**b, the molar ellipticity traces of ELP-sfGFP purified by the two different purification methods are nearly identical. Further, DLS data taken on protein purified by each method indicates that the EtOH-induced precipitation does not affect the final aggregation state of the ELP-sfGFP (**Figure D-14**).

The key difference between the NaCl-NaCl and the NaCl-EtOH purification schemes is that the use of EtOH as a precipitant in the second cycle results in a mostly desalted final protein product. The desalting effect allows for fine control over the salt content of the final solution. Conductivity measurements averaged over three replicate experiments indicate that the NaCl-NaCl and NaCl-EtOH methods result in solutions with 26. ± 5. mM and 0.9 ± 0.4 mM equivalents of NaCl, respectively (**Figure 5-5**c). Importantly, the conductivity of NaCl-EtOH purified solutions was

consistent across all replicate experiments, indicating reproducible control over the final salt content. In contrast, the NaCl-NaCl method produced solutions with a comparatively wider spread of conductivities in the final product. Ultimately, these data demonstrate that the NaCl-EtOH method is a promising approach in applications where fine control over salt content is desired, especially in situations where dialysis or buffer exchange columns are not feasible.

To demonstrate the promise of this method in high-throughput protein processing, the NaCl-EtOH method was used to purify and desalt ELP-sfGFP expressed in a well-plate format. ELP-sfGFP was used as a model protein because it is produced in high yield and because purification and desalting of this protein by the NaCl-EtOH purification method was demonstrated on liter-scale expressions (**Figure 5-5**). Thus, any complications in processing can be attributed to the scale-down of the process rather than changes to the protein system. To compare the NaCl-EtOH method with the NaCl-NaCl method, two independent experiments were performed for each method. Each experiment consisted of with 84 replicate wells out of a 96-well plate expression. The metrics considered were the same as those in liter-scale experiments: yield/function retention (measured by fluorescence and BCA assay), final sample purity and well-to-well variability in purification performance (measured by SDS-PAGE), and average purified product salinity (measured by conductivity).

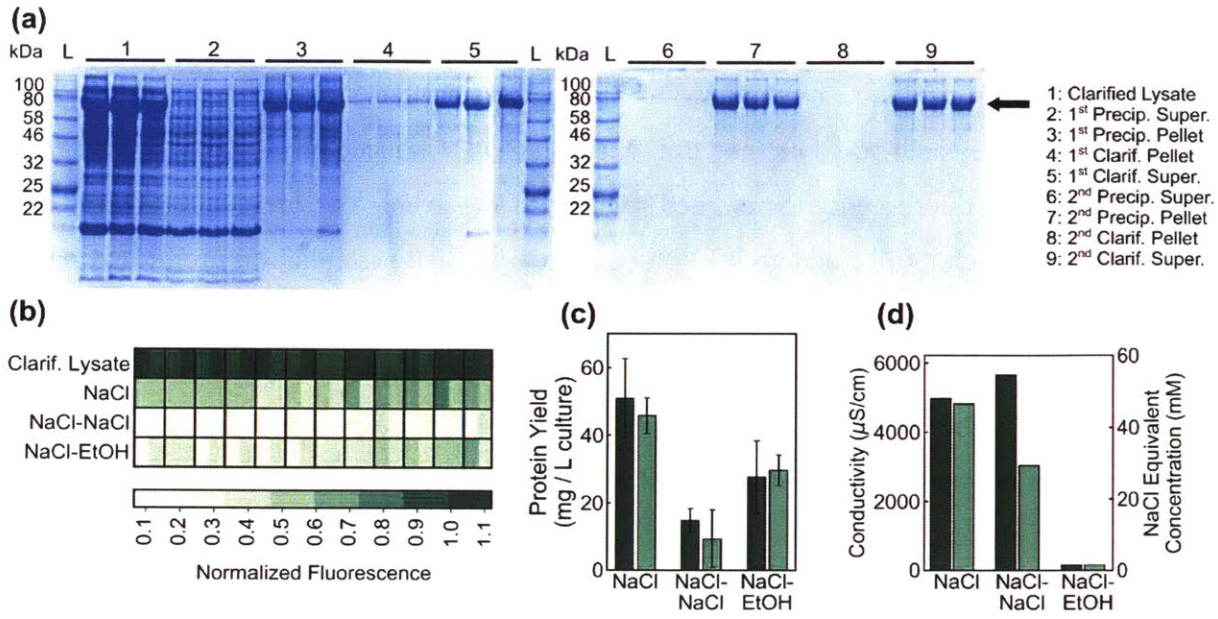


Figure 5-6. (a) Coomassie-stained SDS-PAGE gels showing well-to-well variability in ELP-sfGFP well plate expressions purified by the NaCl-EtOH method. The arrow indicates the apparent molecular weight of ELP-sfGFP. In the lane labels, 1st and 2nd refer to the first and second purification cycles. Labels of fractions are the same as those shown in **Figure 5-1**a: "Precip." refers to samples after incubation with precipitant (1.5 M NaCl or 30 vol% EtOH) and subsequent centrifugation to separate soluble supernatant ("Super.") and insoluble ("Pellet") fractions. "Clarif." refers to samples after resuspension in water and clarification by centrifugation to separate the soluble supernatant ("Super.") and insoluble ("Pellet") fractions. (b) Normalized sfGFP fluorescence after one cycle of NaCl-induced precipitation, the NaCl-NaCl method, or the NaCl-EtOH method. Each condition includes measurements of 12 replicate samples (squares) from each of the two separate experiments (half-squares). Fluorescence intensity was normalized to the average clarified lysate fluorescence of a given well-plate expression experiment. (c) Protein yield after one cycle of NaCl-induced precipitation, the NaCl-NaCl method, or the NaCl-EtOH method, as measured by BCA. Each bar color corresponds to a well-plate expression experiment. The height of each bar is the average yield over 12 replicate expression wells (*i.e.* 12 biological replicates). Error bars represent one standard deviation calculated from these 12 replicate expression wells. (d) Average solvent conductivity in purified

protein samples after one cycle of NaCl-induced precipitation, the NaCl-NaCl method, or the NaCl-EtOH method, each bar representing an average over 30 replicate wells.

The NaCl-EtOH precipitation method led to higher final protein yields of well-plate expressed ELP-sfGFP than the NaCl-NaCl method did. After one NaCl-induced precipitation cycle, 54 ± 9% of the ELP-sfGFP fluorescence is retained relative to the average fluorescence of the clarified lysate. A second cycle performed with EtOH- or NaCl-induced precipitation produces final ELP-sfGFP yields of 41 ± 9% and 22 ± 6%, respectively (individual replicates shown in **Figure 5-6**b). Importantly, much of the well-to-well variability did not arise from differences in protein expression level—the variation in the fluorescence of clarified lysates was, on average, 5%. Instead, most of the variability in yield was introduced in the first NaCl-induced precipitation while aspirating the supernatant fraction after separating the precipitated ELP-sfGFP. In these precipitations, even careful aspiration of the supernatant often resulted in some degree of delamination of the pellets from the bottom of the well, which was easily visualized with the green ELP-sfGFP. The pellets produced by the EtOH-induced precipitation did not exhibit this sensitivity to pipetting. Evidence of well-to-well variability in the NaCl-NaCl method was also observed by SDS-PAGE in **Figure D-10**, as a significant amount of the ELP-sfGFP product localized to the soluble fraction of the second NaCl-induced precipitation. In contrast, little to no product was observed in the same fraction for the EtOH-induced precipitation (lane group 6 in **Figure 5-6**a). Ultimately, the average final yields of the NaCl-EtOH and NaCl-NaCl methods were 29 ± 8 and 12 ± 7 mg protein/L culture, respectively (**Figure 5-6**c).

The NaCl-EtOH purification method also reproducibly yielded final protein products with low salinities. Because conductivity measurements required more sample volume than could be collected from a single well, the salinities reported represent the average salinities over 30 replicate wells from each experiment. These measurements show that, after one NaCl-induced precipitation cycle, the product has a salinity equivalent to 47 ± 1 mM NaCl on average (**Figure 5-6**d). In the second cycle, products of the two different precipitation methods have significantly different salinities. The product of the NaCl-NaCl method exhibited high variability in final salinity, with an average salinity over the two separate experiments of 40 ± 20 mM equivalents of NaCl. In contrast, the final NaCl-EtOH product had an average final salinity of 1.48 ± 0.05 mM equivalents NaCl, indicating salinity both low and consistent across experiments.

To demonstrate that a combination of NaCl and EtOH precipitations could be used to effectively purify a different ELP-based material, purification was performed on well-plate expressions of an ELP called P10, which is known to form gels upon heating. [34] After expressing this protein in a 96 well plate, P10 was purified using one cycle of NaCl-induced precipitation (37 °C, 1.5 M NaCl) and one cycle of EtOH-induced precipitation (4 °C, 30 vol% EtOH). However, after lyophilizing and resuspending the material produced by this purification scheme (NaCl-EtOH), the product remained turbid and did not fully resuspend at 4 °C as expected (**Figure 5-7**c, tube 1). Analysis of this product by SDS-PAGE showed that, while it contained the protein of interest (**Figure 5-7**b lane 1, indicated by a band at ~25 kDa) there were many contaminant proteins in the sample. To increase sample purity, a second NaCl-induced precipitation was performed prior to the EtOH-induced precipitation. This purification scheme is referred to as NaCl-NaCl-EtOH.

P10 purified by the NaCl-NaCl-EtOH purification scheme was found to be pure and capable of forming stiff gels upon heating. Lane 2 in Figure 7b shows the final purified sample, with a strong band at ~25 kDa and few other bands, indicating high protein purity. In contrast to the material purified by the NaCl-EtOH method (**Figure 5-7**c tube 1), the final protein product of the NaCl-NaCl-EtOH method produced a clear solution upon rehydration at 4 °C (**Figure 5-7**c tube 2). This purity was reproduced across replicate wells in the well plate (**Figure D-11**). After verifying the NaCl-NaCl-EtOH purification scheme, the gelation properties of the well-plate-expressed P10 were evaluated. Purified P10 samples were first lyophilized into a solid. Samples were then rehydrated to a concentration of 10 wt% in water. At 4 °C, these samples formed a viscous solution, but upon heating to 37 °C they formed stiff gels (**Figure 5-7**d). This finding is consistent with previous reports on P10[34] and demonstrates the capability of these ethanol-induced precipitations to purify concentrated ELP-based materials, in addition to ELP-fused proteins.

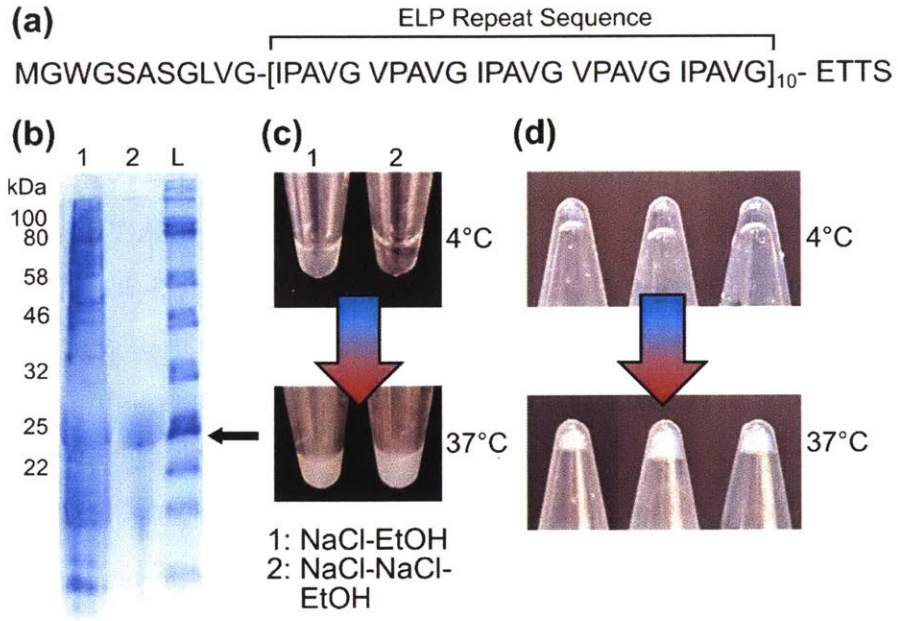


Figure 5-7. (a) Amino acid sequence of the P10 protein used in this study. (b) Coomassie-stained SDS-PAGE gel showing P10 from a well-plate expression that has been purified by the NaCl-EtOH method (lane 1) and P10 from a well-plate expression that has been purified by the NaCl-NaCl-EtOH method (lane 2). The arrow indicates the apparent molecular weight of P10. (c) Images of P10 samples purified by the same methods described in (b) at 4 °C and 37 °C. (d) P10 samples purified from well-plate expressions by the NaCl-NaCl-EtOH method that have been prepared at 10 wt% in water at 4 °C and 37 °C. The change in the meniscus shape upon heating in these images indicates a liquid-solid transition, demonstrating the ability of these purified proteins to form gels upon heating.

While the use of EtOH as a precipitation agent offers several advantages, this method will not generalize to all proteins. Some proteins may show sensitivity to the levels of ethanol required to induce the ELP thermoresponsive transition. For example, extending this procedure to an ELP-fused organophosphate hydrolase (OPH) enzyme was unsuccessful, likely because the ethanol-induced precipitation resulted in unfolding of the OPH. The protein used here was an OPH enzyme with ELP tags fused to both termini of the enzyme—ELP-OPH-ELP (**Figure D-12**). The protein was expressed in a 96-well plate format and purified by the NaCl-EtOH purification scheme. SDS-PAGE analysis showed that the addition of EtOH reproducibly precipitated the ELP-OPH-ELP. However, after the EtOH-induced precipitation, the ELP-OPH-ELP did not resolubilize in water

or buffer at 4 °C, as it did after the NaCl-induced precipitation. Instead, ELP-OPH-ELP remains insoluble, suggesting that the OPH has unfolded and rendered the fusion insoluble (**Figure D-13**). This result indicates an important limitation of this approach—that this purification procedure is unlikely to work if the fused protein is unstable in the concentration of ethanol used for precipitation, as is the case for OPH.[35]

5.5 Conclusions

Ethanol-induced ELP phase transitions were investigated as a new route to selective precipitation of ELP-tagged proteins. Using ELP-sfGFP as a model protein, it was shown that two cycles of EtOH-induced precipitation (EtOH-EtOH) resulted in low protein yields, as the ELP-sfGFP formed soluble aggregates instead of precipitating during the second addition of EtOH. However, one cycle of NaCl-induced precipitation followed by one cycle of EtOH-induced precipitation (NaCl-EtOH) yielded pure, desalted protein for both liter-scale and well-plate scale protein expressions. When an ELP-OPH fusion was precipitated using ethanol, the protein did not resolubilize due to denaturation, demonstrating that the use of ethanol as a precipitating agent will not be effective for all proteins. However, the NaCl-EtOH purification scheme was shown to be effective in simultaneous purification and desalting of well-plate expressions of an ELP-based protein gel. Ethanol-induced ELP precipitation constitutes an important tool enabling high-throughput protein purification and desalting, particularly for applications in protein-based materials.

5.6 References

(1) Terpe, K., Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **2003,** *60* (5), 523-533.

(2) Lichty, J. J.; Malecki, J. L.; Agnew, H. D.; Michelson-Horowitz, D. J.; Tan, S., Comparison of affinity tags for protein purification. *Protein Expr. Purif.* **2005,** *41* (1), 98-105.

(3) Smith, M. C.; Furman, T. C.; Ingolia, T. D.; Pidgeon, C., Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. *J. Biol. Chem.* **1988,** *263* (15), 7211-7215.

(4) Hammarström, M.; Woestenenk, E. A.; Hellgren, N.; Härd, T.; Berglund, H., Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *J. Struct. Funct. Genomics* **2006,** *7* (1), 1-14.

(5) Lin, C.-T.; Moore, P. A.; Auberry, D. L.; Landorf, E. V.; Peppler, T.; Victry, K. D.; Collart, F. R.; Kery, V., Automated purification of recombinant proteins: Combining high-throughput with high yield. *Protein Expr. Purif.* **2006,** *47* (1), 16-24.

(6) Scheich, C.; Sievert, V.; Bussow, K., An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography. *BMC Biotechnol.* **2003,** *3*, 12.

(7) Lesley, S. A., High-Throughput Proteomics: Protein Expression and Purification in the Postgenomic World. *Protein Expr. Purif.* **2001,** *22* (2), 159-164.

(8) Braun, P.; LaBaer, J., High throughput protein production for functional proteomics. *Trends Biotechnol.* **2003,** *21* (9), 383-388.

(9) Kim, Y.; Babnigg, G.; Jedrzejczak, R.; Eschenfeldt, W. H.; Li, H.; Maltseva, N.; Hatzos-Skintges, C.; Gu, M.; Makowska-Grzyska, M.; Wu, R.; An, H.; Chhor, G.; Joachimiak, A., High-throughput protein purification and quality assessment for crystallization. *Methods* **2011,** *55* (1), 12-28.

(10) Houde, A.; Kademi, A.; Leblanc, D., Lipases and their industrial applications. *Appl. Biochem. Biotechnol.* **2004,** *118* (1), 155-170.

(11) de Souza, P. M.; de Oliveira Magalhães, P., Application of microbial α-amylase in industry - A review. *Braz. J. Microbiol.* **2010,** *41* (4), 850-861.

(12) Meyer, D. E.; Chilkoti, A., Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat. Biotechnol.* **1999,** *17* (11), 1112-1115.

(13) Trabbic-Carlson, K.; Liu, L.; Kim, B.; Chilkoti, A., Expression and purification of recombinant proteins from Escherichia coli: Comparison of an elastin-like polypeptide fusion with an oligohistidine fusion. *Protein Sci.* **2004,** *13* (12), 3274-3284.

(14) Urry, D. W.; Gowda, D. C.; Parker, T. M.; Luan, C.-H.; Reid, M. C.; Harris, C. M.; Pattanaik, A.; Harris, R. D., Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers* **1992,** *32* (9), 1243-1250.

(15) McPherson, D. T.; Xu, J.; Urry, D. W., Product Purification by Reversible Phase Transition Following Escherichia coli Expression of Genes Encoding up to 251 Repeats of the Elastomeric Pentapeptide GVGVP. *Protein Expr. Purif.* **1996,** *7* (1), 51-57.

(16) Shimazu, M.; Mulchandani, A.; Chen, W., Thermally triggered purification and immobilization of elastin–OPH fusions. *Biotechnol. Bioeng.* **2003,** *81* (1), 74-79.

(17) Trabbic-Carlson, K.; Meyer, D. E.; Liu, L.; Piervincenzi, R.; Nath, N.; LaBean, T.; Chilkoti, A., Effect of protein fusion on the transition temperature of an environmentally responsive elastin-like polypeptide: a role for surface hydrophobicity? *Protein Eng. Des. Sel.* **2004,** *17* (1), 57-66.

(18) Banki, M. R.; Feng, L.; Wood, D. W., Simple bioseparations using self-cleaving elastin-like polypeptide tags. *Nat. Methods* **2005,** *2*, 659.

(19) Kim, J.-Y.; Mulchandani, A.; Chen, W., Temperature-triggered purification of antibodies. *Biotechnol. Bioeng.* **2005,** *90* (3), 373-379.

(20) Qin, G.; Glassman, M. J.; Lam, C. N.; Chang, D.; Schaible, E.; Hexemer, A.; Olsen, B. D., Topological Effects on Globular Protein-ELP Fusion Block Copolymer Self-Assembly. *Adv. Funct. Mater.* **2015,** *25* (5), 729-738.

(21) Cho, Y. H.; Zhang, Y. J.; Christensen, T.; Sagle, L. B.; Chilkoti, A.; Cremer, P. S., Effects of Hofmeister Anions on the Phase Transition Temperature of Elastin-like Polypeptides. *J. Phys. Chem. B* **2008,** *112* (44), 13765-13771.

(22) Cho, Y.; Zhang, Y.; Christensen, T.; Sagle, L. B.; Chilkoti, A.; Cremer, P. S., Effects of Hofmeister anions on the phase transition temperature of elastin-like polypeptides. *J. Phys. Chem. B* **2008,** *112* (44), 13765-71.

(23) Hassouneh, W.; Christensen, T.; Chilkoti, A., Elastin-like polypeptides as a purification tag for recombinant proteins. *Current protocols in protein science* **2010,** *Chapter 6*, Unit-6.11.

(24) VerHeul, R.; Sweet, C.; Thompson, D. H., Rapid and simple purification of elastin-like polypeptides directly from whole cells and cell lysates by organic solvent extraction. *Biomater. Sci.* **2018,** *6* (4), 863-876.

(25) Schild, H. G.; Muthukumar, M.; Tirrell, D. A., Cononsolvency in mixed aqueous solutions of poly(N-isopropylacrylamide). *Macromolecules* **1991,** *24* (4), 948-952.

(26) Costa, R. O. R.; Freitas, R. F. S., Phase behavior of poly(N-isopropylacrylamide) in binary aqueous solutions. *Polymer* **2002,** *43* (22), 5879-5885.

(27) Mills, C. E.; Ding, E.; Olsen, B. D., Cononsolvency of elastin-like polypeptides (ELPs) in water/alcohol solutions. *Biomacromolecules* **2019,** *In press*.

(28) Glassman, M. J.; Avery, R. K.; Khademhosseini, A.; Olsen, B. D., Toughening of Thermoresponsive Arrested Networks of Elastin-Like Polypeptides To Engineer Cytocompatible Tissue Scaffolds. *Biomacromolecules* **2016,** *17* (2), 415-426.

(29) Raghunathan, G.; Sokalingam, S.; Soundrarajan, N.; Munussami, G.; Madan, B.; Lee, S. G., A comparative study on the stability and structure of two different green fluorescent proteins in organic co-solvent systems. *Biotechnol. Bioprocess Eng.* **2013,** *18* (2), 342-349.

(30) Royant, A.; Noirclerc-Savoye, M., Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein. *J. Struct. Biol.* **2011,** *174* (2), 385-390.

(31) Meyer, D. E.; Trabbic-Carlson, K.; Chilkoti, A., Protein purification by fusion with an environmentally responsive elastin-like polypeptide: Effect of polypeptide length on the purification of thioredoxin. *Biotechnol. Progr.* **2001,** *17* (4), 720-728.

(32) Hassouneh, W.; Fischer, K.; MacEwan, S. R.; Branscheid, R.; Fu, C. L.; Liu, R.; Schmidt, M.; Chilkoti, A., Unexpected Multivalent Display of Proteins by Temperature Triggered Self-Assembly of Elastin-like Polypeptide Block Copolymers. *Biomacromolecules* **2012,** *13* (5), 1598-1605.

(33) Trabbic-Carlson, K.; Meyer, D. E.; Liu, L.; Piervincenzi, R.; Nath, N.; LaBean, T.; Chilkoti, A., Effect of protein fusion on the transition temperature of an environmentally responsive elastin-like polypeptide: a role for surface hydrophobicity? *Protein Eng. Des. Sel.* **2004,** *17* (1), 57-66.

(34) Glassman, M. J.; Olsen, B. D., Arrested Phase Separation of Elastin-like Polypeptide Solutions Yields Stiff, Thermoresponsive Gels. *Biomacromolecules* **2015,** *16* (12), 3762-73.

(35) Mills, C. E.; Obermeyer, A.; Dong, X. H.; Walker, J.; Olsen, B. D., Complex Coacervate Core Micelles for the Dispersion and Stabilization of Organophosphate Hydrolase in Organic Solvents. *Langmuir* **2016,** *32* (50), 13367-13376.

Chapter 6 : High-Throughput Screening of a Streptavidin Binder Library in Self-Assembled Solid Films

6.1 Abstract

Proteins capable of molecular recognition and binding, such as antibodies, are of great interest in a variety of biosensing technologies. [1-5] Immobilization of such binding proteins is an important consideration in fabricating sensing devices, as the immobilization strategy can influence the protein performance. Self-assembly of protein-polymer bioconjugates and elastin-like polypeptide (ELP) fusion proteins represents a promising strategy for achieving high protein packing densities while retaining control over protein orientation. While this self-assembled format represents a promising strategy for globular protein immobilization, current strategies for engineering functional proteins do not screen for the ways in which this self-assembled environment may alter protein function. The goal of this work is to establish a platform for comparing the binding affinity of a library of Sso7d proteins in dilute solution and self-assembled formats. A fusion protein containing an ELP block, an Sso7d block, and a coiled-coil order directing sequence (ZE) is constructed for this purpose, and is shown to self-assemble into lamellae. Constructs containing two Sso7d variants (one with high streptavidin affinity, and one with no streptavidin affinity) are expressed and purified in a well-plate format using ELP-based precipitation techniques. This process is shown to produce proteins with high yield and purity with low well-to-well variability. Finally, the binding affinities of the two Sso7d variants are measured in dilute solution using biolayer interferometry (BLI) and in self-assembled films using fluorescence microscopy. These measurements indicate that the affinities of the two Sso7d variants are differentiable in both dilute solution and self-assembled formats. By combining this ELP-Sso7d-ZE system with an existing library of Sso7d variants (with varying streptavidin affinities), it will be possible to answer the question of how self-assembly influences protein function.

6.2 Introduction

Proteins capable of binding to specific molecular motifs are an essential component in the development of biosensors. [6-8] Binding proteins have found applications in a variety of formats, including paper-based sensors, electrochemical devices, and microfluidic devices. [4-5, 9-13] The simplest strategy for covalent immobilization of proteins onto a surface is to react lysine residues (primary amine groups) on the protein with some reactive group on a substrate of interest. [14-15] However, because these amine groups are typically distributed randomly over the protein surface,

this technique does not permit control over protein orientation in the immobilized state. Alternative strategies for protein immobilization can overcome this challenge. For example, the protein can be site-specifically conjugated to a molecule, such as ssDNA, that then binds to the surface of interest. [16] Incorporation of non-canonical amino acids into the protein sequence at specific locations on the protein surface can also enable access to a broad array of chemistries for protein surface immobilization. [17-19] Alternatively, the protein of interest can be genetically fused to a peptide or protein tag that can specifically bind to a functionalized surface. [20] Candidates for this strategy include the cellulose binding domain (CBD), coiled coil sequences, and poly-histidine tags. [5, 13, 21-22] While all these strategies allow fabrication of immobilized protein monolayers and afford control over protein orientation, many of them rely on genetic and/or covalent modification of the protein of interest, as well as pretreatment of the substrate.

Block copolymer inspired self-assembly represents an alternative approach to immobilizing binding proteins while controlling their orientation in the immobilized state. It has been shown that when globular protein mCherry is conjugated to the polymer poly(*N*-isopropyl acrylamide) (PNIPAM), microphase separation of the protein and polymer occur, both in concentrated solution ($\geq$ 20 wt% bionconjugate in water) and the solid state.[23-24] Depending on the temperature, concentration in solution, and the length of the PNIPAM polymer, this microphase separation can result in the formation of periodic nanostructures, such as lamellae. [23-25] Control over protein orientation in these periodic structures can be achieved by changing the attachment location of the polymer to the protein.[26] Furthermore, it has been shown that glutaraldehyde crosslinking of these bioconjugates in the solid state can yield a permanently immobilized protein material. [27] This strategy has been successfully employed to fabricate immobilized biosensors, both with the antibody human immunoglobulin G (IgG), as well as an engineered streptavidin binding protein. [28-29] While this bioconjugate self-assembly is a promising platform for protein immobilization, bioconjugate preparation requires chemical attachment of a polymer to the protein of interest. Thus, synthesis of these materials can be complicated by challenges such as low coupling efficiency of larger polymers to proteins. [30] Genetic fusion of an elastin-like polypeptide (ELP) to a globular protein of interest circumvents some of the synthetic challenges of bioconjugate preparation while still retaining the key desirable features of protein-polymer bioconjugates. Work out of the Olsen lab has shown that fusion of an ELP to model protein mCherry results in self-

assembly in concentrated solution that is similar to that observed in PNIPAM-mCherry bioconjugates. [31-32]

One key feature of binding proteins that makes them attractive candidates for many applications is the diversity of accessible targets. This is limited not only to naturally occurring binding proteins. Protein engineering has enabled the rapid development of proteins capable of binding a broad array of non-native targets. [33-36] In directed evolution strategies, many variants of a given protein are screened using some type of display (such as yeast/bacterial surface display) combined with fluorescence-activated cell sorting (FACS) for improved binding affinity towards a given target. [37-39] One key assumption of this approach is that binding affinity of the protein in the screened format, such as on a cell surface, is the same (or, at the very least, a good approximation) for the binding affinity of the protein in the final immobilized format. However, studies have shown that the way in which proteins are immobilized can directly impact their performance.[19] In particular, recent studies in the Olsen lab have demonstrated that proteins immobilized in bioconjugate films perform differently than those immobilized in a monolayer format. [27-29] This combination of observations raises the question of whether the tools for screening protein performance during directed evolution are representative of the protein behavior in the target environment.

The goal of this work is to develop a method for screening of a library of binding proteins in a self-assembled film. This is done using rcSso7d as the binding protein fused to an ELP tag that enables high-throughput purification, and self-assembly of the Sso7d binding protein. This work assesses the self-assembly of this ELP-Sso7d construct and establishes that this construct can be expressed and purified in a high-throughput format. The viability of this technique is demonstrated on two Sso7d variants—one engineered to have high streptavidin affinity ($K_d = 5.5 \times 10^{-10}$ M) and one with no engineered affinity. [5] Finally, streptavidin binding affinity is quantified in both dilute solution and self-assembled formats using these two variants.

6.3 Methods

6.3.1 Biosynthesis

6.3.1.1 Molecular Cloning

Details of the concatamerization of the ELP gene used in this study can be found in **2.1.1.3 High-Throughput Screening of Streptavidin Binder Library in Self-Assembled Thin Films**. The

construction of the ELP-Sso7d-ZE pET28b plasmid was performed in a two-step Gibson assembly process, utilizing constructs isolated from three separate plasmids: pET28b-Nt-deELP20K, pET28b-rcSso7d.SA.1,[13] and pUC57-ZE. Individual fragments were prepared via restriction digest (in the case of the ELP gene) or PCR, and these fragments were then combined in a Gibson assembly reaction in order to yield an in-frame ELP-Sso7d-ZE construct.

Due to the highly repetitive nature of the ELP construct, this gene could not be amplified via PCR. Instead, the ELP gene fragment was prepared by subjecting the pET28b-ELP plasmid to a restriction digest reaction, via a *BamHI*/*BsaI* double digest. This digested fragment would be included in the Gibson assembly reaction, and the 5' overhangs resulting from these restriction digestion reactions would be removed by the exonuclease function. These internal restriction sites can then be destroyed (without changing the amino acid sequence of the encoded protein) by ensuring that the sequence overlaps of adjacent fragments substitute degenerate codons in place of the original DNA sequence. Both of the selected restriction sites were destroyed in order to ensure that those enzymes would uniquely digest the resulting construct at the desired locations, and the *BsaI* site was specifically selected as the 3' restriction site in order to ensure that the stop codon would also be removed from the digested gene fragment.

Amplicons of the linearized pET28b plasmid backbone, rcSso7d.SA.1 gene, and ZE gene were prepared via standard PCR, using the primer pairs tabulated in **Table 2-1**. Primers were designed to append the appropriate sequence overlaps to the 5' and 3' termini of each PCR product, while maintaining a GC ratio within (or just above) the range of 40-60%. Specifically, the linearized plasmid backbone amplicon was modified to overlap with the 5' end of the ELP product (introducing a degenerate AGC codon (S) in the place of the TCC codon originally present in the *BamHI* site) and the 3' end of the ZE product. The rcSso7d.SA.1 amplicon was modified to overlap with the 3' end of the ELP product (introducing a degenerate GAA codon (E) in the place of the GAG codon originally present in the *BsaI* site and appending a flexible $(GGGGS)_2$ linker between the ELP and rcSso7d protein blocks) and the 5' end of the ZE product. The ZE amplicon was modified to overlap the 3' end of the rcSso7d product and the 5' end of the linearized plasmid backbone product. A stop codon was removed from the rcSso7d product, and introduced at the 3' end of the ZE amplicon. Unique restriction enzyme sites were introduced between each gene fragment, yielding the final construct: *NdeI*-ELP-*BamHI*-rcSso7d-*EcoRI*-ZE-*XhoI*. This construct

will be referred to as ELP-Sso7d-ZE, or may be shortened to ESZ. Full sequences for both of the rcSso7d variants used in this chapter can be found in **E.1 DNA and Amino Acid Sequences**.

6.3.1.2 Protein Expression and Purification

Fermenter Scale

ELP-Sso7d-ZE expression for small-angle X-ray scattering experiments was carried out using Sso7d variant rcSso7d.SA, which has been previously reported in studies of PNIPAM-Sso7d bioconjugate self-assembly. [29] A starter culture consisting of 50 mL lysogeny broth (LB) supplemented with 50 μg/mL kanamycin was inoculated from a single colony selected off an LB-agar plate supplemented with 50 μg/mL kanamycin. The starter culture was grown in a baffled 250 mL conical flask overnight at 37 °C in an orbital shaker. The expression was carried out in a 7 L fermenter with a 5 L working volume in terrific broth (TB) supplemented with 50 μg/mL kanamycin. The expression was inoculated with the starter culture and grown without induction at 30 °C for 24 hrs. Cells were then collected by centrifugation (4,000xg at 4 °C for 15 min). Cells were resuspended in MENT buffer (3 mM $MgCl_2$, 1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 10 mM trizma, pH 7.5) at a ratio of 100 mL buffer per 30 g wet cell mass. Resuspended cells were frozen overnight at -80 °C. Resuspended cells were then thawed and incubated with lysozyme (1 mg/mL) at 4 °C for 2 hrs. To ensure complete cell lysis, cells were then disrupted by tip sonication and subsequently cooled for at least 1 hr. Lysates were then clarified by centrifugation (26,700 xg for 1.5 hrs at 4 °C). After clarification, lysates were incubated with ~0.1 mg/mL DNase I and RNase A at 37 °C for 2 hrs to ensure degradation of nucleic acids in the lysate.

Lysates were purified by two rounds of ELP-based precipitation using one NaCl-induced precipitation and one EtOH-induced precipitation as described in **Chapter 5**. Briefly, 5 M NaCl in water was added to the clarified lysate to adjust the final NaCl concentration to 1.5 M. This solution was then incubated at 37 °C overnight to ensure complete precipitation of the desired protein product. Precipitant was collected by centrifugation at 37 °C (26,700 xg for 1 hr) and supernatant was discarded. Precipitant was then resuspended in MilliQ® purified water (2/3 of the volume of the protein solution at the beginning of the cycle) at 4 °C for at least 3 hrs to ensure complete redissolution of the desired ELP-tagged product. This resuspended product was then clarified by centrifugation to remove any insoluble aggregates (26,700 xg for 1 hr at 4 °C). The

pellet produced by this centrifugation was discarded, and the supernatant was retained as the final product of the first NaCl-induced purification cycle. The product of this first purification cycle was then subjected to a second purification cycle using ethanol-induced precipitation. Precipitation was induced by adjusting the EtOH content in the solution to 30 vol% by addition of 100 vol% EtOH. Complete precipitation was ensured by incubation of the sample at 4 °C for at least 3 hrs. Precipitant was collected by centrifugation at 4 °C (26,700 xg for 1 hr). Resuspension of the pellet and subsequent removal of insoluble aggregates by centrifugation was carried out as in the first cycle (see above).

After purification, protein was concentrated to ~10-20 wt% in solution by ultrafiltration (molecular weight cutoff of 10 kDa). Concentrated protein solutions were then drop cast onto a Teflon sheet and dried at room temperature under a vacuum ramp (50 Torr per hour down to 5 Torr), for 12 hours. Solid samples were then collected and stored at -20 °C until needed for further experiments.

Well-Plate Scale

Well-plate scale expressions of ELP-Sso7d-ZE were carried out for two Sso7d variants—one engineered, well-characterized variant, and one naïve variant, expected to have no binding affinity for streptavidin. Starter cultures for well-plate expressions were prepared in 10 mL of LB supplemented with 50 μg/mL kanamycin. Cultures were inoculated using a previous overnight culture and grown overnight at 37 °C with orbital shaking. Well-plate expressions were carried out in 96 deep-well plates with cone-shaped bottoms. Each well contained 1 mL of TB supplemented with 50 μg/mL kanamycin, and one 3-5 mm diameter glass bead (to promote mixing). Each well was inoculated with 10 μL of starter culture at the start of expression. Plates were grown at 37 °C with shaking at 300 rpm. OD_{600} was monitored in 7 of the 96 wells of the plate by using a multi-channel pipette to move 150 μL of expression culture into a sterile microtiter plate. The absorbance at 600 nm was then read using a Tecan Infinite Pro M200 plate reader, and a path length correction was applied to calculate equivalent OD_{600} for a 1 cm path length. After average OD_{600} in the monitored wells reached 0.6-0.8, plates were removed from the incubator and cooled at 4 °C for 30 min. Expressions were then induced with 0.5 mM IPTG and moved back into the orbital shaker, where expression was allowed to proceed for 16 hrs at 24 °C with shaking at 300 rpm. After expression, cells were harvested by centrifugation (4,000xg, 15 m, 4 °C). Cell pellets were frozen overnight at -20 °C to aid in lysis. Cell pellets were then resuspended in 140 μL of MENT buffer

(3 mM $MgCl_2$, 1 mM EDTA, 100 mM NaCl, 10 mM trizma, pH 7.5) containing 1 mg/mL lysozyme and ~0.1 mg/mL DNase I and RNase A. Resuspended cell pellets were incubated at 37 °C with orbital shaking for 2 hrs and then moved to the refridgerator and allowed to cool at 4 °C for 2 hrs. Lysates were then frozen again at -20 °C overnight. Lysates were then thawed at 4 °C and clarified by centrifugation (6,000 xg, 4 °C, 1 hr).

Clarified lysates were purified by two rounds of ELP-based precipitation, similar to that described above. Briefly, after clarification, 60 μL of 5 M NaCl in water was added to each well of lysate to bring the total NaCl concentration to 1.5 M. Plates were then incubated overnight at 37 °C to ensure complete precipitation. Precipitation and clarification were carried out as described above, except with a centrifugation speed of 6,000 xg. In this case, precipitants were resuspended in 140 μL of MilliQ water per well. The second precipitation was carried out by adding 60 μL of EtOH to each well (bringing the final concentration to 30 vol% ethanol, as above). Again, precipitation and clarification were carried out using centrifugation at 6,000 xg. In this case, the precipitant pellet was resuspended in 125 μL of MilliQ water. The supernatant produced by the clarification step after this EtOH-induced precipitation was the final product used to cast films. Protein yield in each well was measured by sacrificing 25 μL from each well to perform a bicinchoninic acid (BCA) assay. Of the remaining 100 μL of purified protein in each well, 25 μL was used to measure the dilute solution binding of the protein by bio-layer interferometry (BLI). The final 75 μL was used to cast films at the bottom of 96-well clear polystyrene microtiter plates. To form these films, protein solutions were dried down into films at room temperature under a vacuum ramp (50 Torr per hour down to 5 Torr) for 12 hrs.

6.3.2 Film Crosslinking

After drying of films under vacuum, films were permanently immobilized *via* glutaraldehyde crosslinking. Plates containing protein films were equilibrated in a 42 °C water bath. This ensured that the ELP portion of the protein would remain insoluble for the duration of the crosslinking reaction, which kept the entire protein immobilized at the bottom of the well while the glutaraldehyde chemistry proceeded. A solution of 1.4 wt% glutaraldehyde in water was brought to equilibrium in the same 42 °C water bath for 10 minutes prior to crosslinking. Once both the plate containing the films and the glutaraldehyde solution were thermally equilibrated, 200 μL of 1.4 wt% glutaraldehyde solution was added to each well with a multichannel pipette and allowed to sit for 10-20 sec. Glutaraldehyde solution was then removed with a multichannel pipette and the

plate containing the protein films was removed from the warm bath. Protein films were then washed immediately two times with 200 µL of room temperature MilliQ water. After these initial two washes, 200 µL of MilliQ water was added to each well. Plates were then allowed to equilibrate overnight at 4 °C with the MilliQ water to ensure the removal of all uncrosslinked protein from the films. BCA performed on the washes from nine replicate films from each binding protein (rcSso7d.SA and the naïve variant) did not detect any protein in these washes, suggesting that the crosslinking was successful.

6.3.3 Small-Angle X-Ray Scattering

Concentrated solution samples were prepared for small-angle X-ray scattering (SAXS) by weighing out pellets into Eppendorf tubes and then hydrating with MilliQ water to the desired concentration (50 or 60 wt%). Solid samples were measured in pellet form produced by vacuum drying (see above). All samples were loaded into 1 mm thick anodized aluminum washers with a 3 mm bore and sealed with Kapton tape. SAXS measurements were performed at Brookhaven National Lab at the National Synchrotron Light Source II (NSLS-II) on the 11-BM Complex Materials Scattering (CMS) beamline. Samples were equilibrated at 10 °C for 20 min prior to measurement. Samples were equilibrated for 10 min for all other temperatures measured. Data was acquired for 10 seconds for all samples. No damage to nanostructure by the beam was detectable for this acquisition time. 2D data was corrected for empty cell scattering using the following equation

$$I(q) = \frac{I_{\text{sample}}}{T_{\text{sample}}} - \frac{t_{\text{sample}}}{t_{\text{empty}}}\frac{I_{\text{empty}}}{T_{\text{empty}}} \tag{6.1}$$

Where I, t, and T refer to scattering intensity, acquisition time, and transmission, respectively. Subscripts sample and empty refer to the values of these variables for the sample and the empty cell, respectively.

6.3.4 Bio-Layer Interferometry

The binding activity of concentration-quantified samples of this purified ESZ species was assessed via biolayer interferometry (BLI). Protein concentrations were assessed using a reducing agent compatible bicinchoninic acid (BCA) assay from Pierce In short, 200 µL volumes of the three ESZ species were prepared at soluble concentrations of 100 nM. Streptavidin-coated BLI sensors were hydrated in kinetics buffer (1x PBS, 0.1% w/v BSA, and 0.02% v/v Tween-20) for ten minutes at

room temperature. BLI tips were assayed for function by a ForteBio Octet Red reader, and then these tips were dipped into fresh kinetics buffer for 60 seconds in order to establish a signal baseline. Tips were then transferred into the 100 nM sample wells, and incubated with mixing at 1000 rpm for 5 minutes. Ligand-loaded tips were then transferred back into the equilibration buffer for a 10-minute dissociation step.

6.3.5 Film Assay

The streptavidin affinity of the two Sso7d variants in well-plate cast films was assayed using streptavidin Alexa Fluor 647 (ThermoFisher #S32357). To perform the assay, solutions of labeled streptavidin were prepared at 1 μM in phosphate buffered saline (PBS). 75 μL of this solution was loaded into each well with a film and allowed to equilibrate for 5 hours. Streptavidin solutions were then removed, and the films were washed three times with PBS by using a multichannel pipette to add and then 150 μL of PBS to each well. The PBS added to the wells for the third wash was allowed to equilibrate with the films overnight to allow non-specifically bound species to diffuse out of the films. Films were then rinsed with water and dried under filtered airflow.

Fluorescence microscopy images of films were acquired on an Olympus IX-81 inverted fluorescence microscope equipped with an AxioCam HRC CCD camera. Images were taken at 4x magnification using a Cy5 filter set and an exposure time of 5000 ms. Fluorescence intensity was calculated using ImageJ software by determining the area averaged intensity in a rectangular area of the film free of defects. [40]

6.4 Results & Discussion

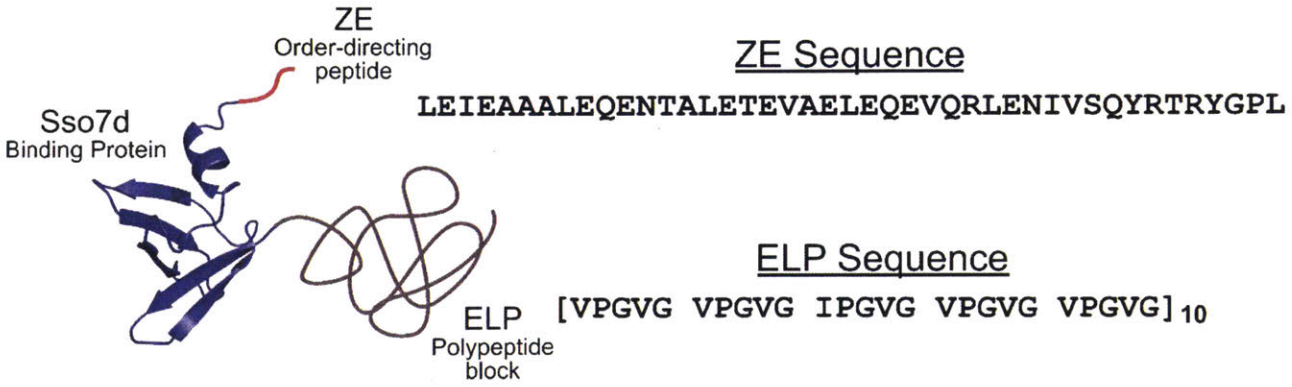


Figure 6-1. Schematic image of the protein design for this study, including the sequences of ZE peptide and ELP used.

The goal of this work was to define a system which could be used to address the question of how incorporation into a self-assembled material impacts the affinity of a library of different rcSso7d variants. The proposed construct for conducting this study is comprised of an rcSso7d binding protein fused to an ELP at its N-terminus and an order-directing coiled coil sequence (ZE) at its C-terminus (**Figure 6-1**). The ELP repeat sequence used in this study was [VPGVG VPGVG IPGVG VPGVG VPGVG]$_{10}$. This ELP repeat sequence was selected based on previous studies (published and unpublished) showing that this ELP sequence yields self-assembly in concentrated solution for mCherry and sfGFP fusions. [31-32, 41] One challenge in the design of this system is the fact that monomeric rcSso7d has been shown to yield poor self-assembly in PNIPAM conjugate systems. [29] Previous work has overcome this challenge by multimerization of the Sso7d domain; however, for ease of library incorporation into a genetic construct, it is nonideal to have to incorporate mutations in multiple copies of the same Sso7d gene. Thus, an alternative strategy was necessary to promote ordering in this system. Unpublished work by Justin Paloni has shown that fusion of a coiled coil sequence, here called "ZE" (LEIEAAALEQENTALETEVAELEQEVQRLENIVSQYRTRYGPL), to the C-terminus domain of dimeric Sso7d greatly improved the self-assembly of PNIPAM-2xSso7d bioconjugates. It was hypothesized that fusion of the same domain to monomeric Sso7d would yield similar improvements to self-assembly of ELP-Sso7d fusion proteins. Thus, the final design for the fusion protein was ELP-Sso7d-ZE, where the ELP acts as a polymer-like block, the Sso7d is the binding protein of interest, and the ZE peptide chain serves as an order-directing sequence. (**Figure 6-1**)

The potential of this construct to answer the scientific question of interest was evaluated based on the following three criteria. First, ELP-Sso7d-ZE constructs must self-assemble; otherwise, the question of how self-assembly impacts performance will not be answered. Second, ELP-Sso7d-ZE variants must reproducibly express well in a well-plate format and be stable enough to withstand NaCl-induced and EtOH-induced precipitations required to purify and desalt ELP-tagged constructs. Otherwise, the requisite library size ($\geq$ 80 variants) will not be accessible. Finally, the binding activity of these films must be reproducibly measurable by some technique that can be executed with reasonable throughput. Ideally, this would be achievable in a plate reader; however, microscopy-based techniques are also feasible. The remainder of this chapter provides evidence that these criteria are met for the ELP-Sso7d-ZE system using two rcSso7d

variants—one that has been engineered previously to have high streptavidin affinity, and one that has no specific streptavidin affinity. [13, 29]

6.4.1 Self-Assembly of ELP-Sso7d-ZE

The fusion protein ELP-Sso7d-ZE was designed based on previously published and unpublished results from the Olsen lab to maximize the likelihood that this construct would self-assemble in concentrated solution. In this construct, the ELP serves as a polymer-like block, and the coiled-coil ZE sequence acts as an order-directing moiety. Small-angle X-ray scattering (SAXS) data taken on the ELP-Sso7d-ZE construct in concentrated solution is shown in **Figure 6-2a**. At 10 °C at both concentrations, a primary and secondary scattering peak are observed. In both cases, these peaks index to q^* and $2q^*$, indicating that these materials form lamellae at 50 and 60 wt% in solution. The intensity of the scattering peaks decreases with increasing temperature at both concentrations. A similar trend has been observed for ELP-mCherry systems, and is thought to be a result of a change in the lamellar form factor. [31] This change in form factor is thought to be a result of the water selectively partitioning into the mCherry phase at temperatures above the ELP transition temperature; when the water partitions into the mCherry phase, it creates a minimum in the lamellar form factor that coincides with the primary scattering peak, resulting in the decreased intensity at higher temperatures. In the samples measured here, this is likely the case for the 50 wt% ELP-Sso7d-ZE sample, where the intensity of the secondary peak does not appear to decrease with decreasing primary peak intensity. **Figure 6-2**b shows SAXS data taken on the solid ELP-rcSso7d-ZE sample. In this sample, only one scattering peak is observed. This indicates that the sample is experiencing microphase separation, but it is unclear whether or not periodic structures exist in this solid material, as changes to form factor (similar to those described above) could result in loss of a secondary scattering peak. The location of the primary scattering peak in this data occurs at 0.35 nm^{-1}, which corresponds to an approximate domain spacing in this material of 18 nm. To assess whether this is reasonable, the dimensions of the individual fusion protein components can be estimated. PyMol calculations on the crystal structure of Sso7d indicate that it has a radius of gyration, R_g, of 1.1 nm.[42] Based on a SANS study of a larger ELP, [41] and assuming ELP R_g scales with chain length as a polymer chain in a good solvent, the R_g of the ELP used here is estimated to be ~4 nm. Not including the ZE peptide domain, two layers of the ELP-Sso7d fusion, based on these R_g approximations, would span 20.4 nm. Taking into account the fact that

ELP chains should interpenetrate in the ELP-rich domain of the lamellar structure, the 18 nm domain spacing reported here is quite reasonable.

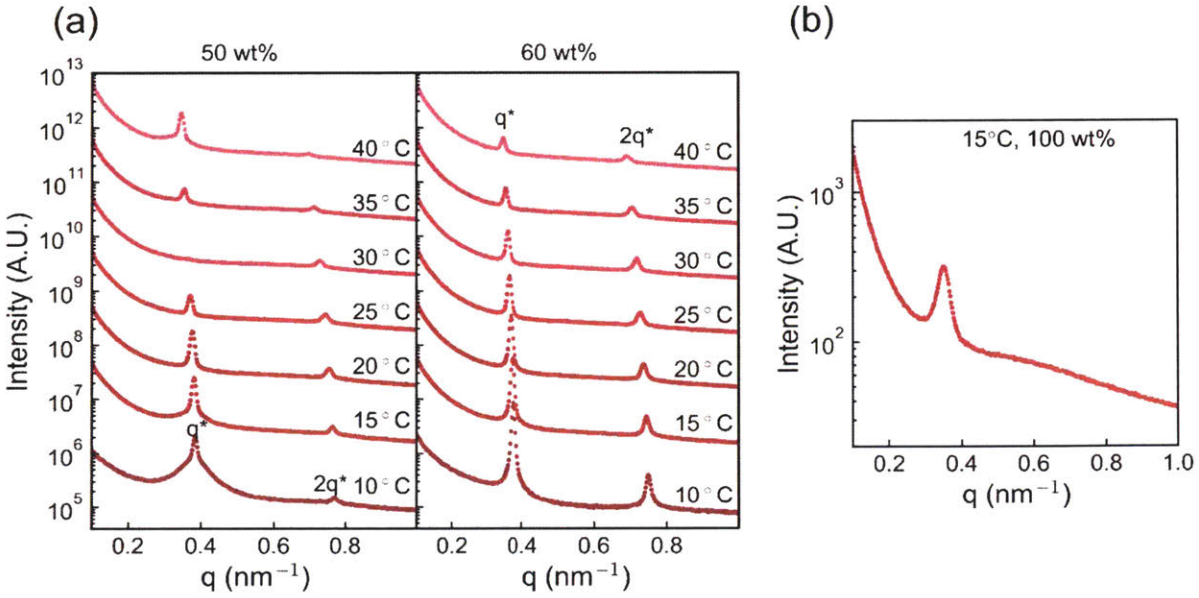


Figure 6-2. SAXS data taken on ELP-rcSso7d-ZE. (a) Data on ELP-rcSso7d-ZE in concentrated solution (50 and 60 wt% in water). Peak indexing (marked as q^* and $2q^*$ on these plots) indicates the formation of lamellar phases at both these concentrations. (b) Data taken on ELP-rcSso7d-ZE in the solid state. The primary scattering peak occurs at 0.35 nm^{-1}, corresponding to an approximate domain spacing of 18 nm.

This combination of results suggests that the ELP-Sso7d-ZE construct self-assembles sufficiently well for use in this study. The sharpness of the peaks observed by SAXS in concentrated solution are associated with a well-ordered material. While the lack of secondary peaks (corresponding to periodic ordering) in the solid state sample is initially concerning, it is possible that the secondary scattering peak is being masked by form factor scattering. Furthermore, the confinement of these materials to thin-film formats can induce periodic ordering in materials that are already microphase separated, though this would need to be verified experimentally for this system. It is important to note that these results are expected to hold for all Sso7d variants explored in this study as previous work in the Olsen lab has shown that minor changes to protein sequence do not strongly impact self-assembly behavior of bioconjugates. [43]

6.4.2 Well-Plate Expression, Purification and Processing of ELP-Sso7d-ZE

Reproducibility of well-plate scale expression and purification of ELP-Sso7d-ZE was assessed by performing a well-plate scale expression on constructs containing two Sso7d variants. One variant (rcSso7d.SA) is the Sso7d used in the self-assembly experiments, known to have a strong binding affinity ($K_d = 5.5 \times 10^{-10}$ M) for streptavidin, both in solution and in self-assembled formats. [5, 29] The second is a variant from the naïve Sso7d library constructed by Miller *et. al.*, which (as shown later in this work) does not have any binding affinity for streptavidin. [13] The goal of this portion of the study was to confirm that ELP-Sso7d-ZE expression, purification, and processing into crosslinked films could be implemented in a well plate format with sufficient yield and consistency to conduct the remainder of the study. This was achieved by running a full well plate's worth of replicates for each variant. Out of the 96 wells in each plate, 88 expression wells were run; 8 wells were run without inoculation to confirm that there was no cross contamination between wells during expression.

Analysis by SDS-PAGE indicates that there is little well-to-well variability in the purification of ELP-Sso7d-ZE for either Sso7d variant. Purification was carried out according to the procedure established in **Chapter 5** (recapitulated here in **Section 6.3.1.2**). Briefly, samples were subject to two rounds of purification by ELP-based precipitation. The first and second rounds used NaCl and EtOH as precipitants, respectively. For each fraction produced in the purification (except the precipitation pellet), three samples were selected from across the well plate for each variant to evaluate well-to-well variability in the purification procedure. This depleted an additional 6 samples from the total final number of product wells—Three for each clarification supernatant (of which there were two—one from the NaCl-induced precipitation and one from the EtOH-induced precipitation). Coomassie blue stained SDS-PAGE on the different fractions produced in this process is shown in **Figure 6-3**. Notably, these gels look nearly identical for both Sso7d variants used in this study, indicating that small changes to the Sso7d domain, as expected, do not impact the purification process. There is also little to no well-to-well variability in the samples measured. Both the impurity lanes and product lanes in each gel show bands in the same location of similar intensity. This combination of results indicates that this purification procedure should work reproducibly across an array of different Sso7d variants.

SDS-PAGE analysis indicates that the majority of the product localizes to purified fractions with only minor product losses. Lanes 1, 2, 3, 5, and 6 on each of the gels shown in **Figure 6-3** are

fractions that should not contain the product of interest. For both Sso7d variants, lanes 1 and 3 both have bands corresponding to the desired ELP-Sso7d-ZE product. The presence of desired product in the pellet produced during lysate clarification (Lane 1) is not surprising, and is likely due to two main factors. First, because lysis of cells in the well-plate format does not include a tip sonication step, it is quite possible that the cells were not completely lysed prior to clarification. If this were the case, the ELP-Sso7d-ZE contained in those unlysed cells would appear in the lysate pellet fraction. The second possibility is that a portion of the protein product expressed is unfolded and thus localizes to the pellet during clarification. Product loss is also observed in lane 3 of each gel (**Figure 6-3**). This fraction corresponds to the pellet (or insoluble fraction) produced when clarifying the resuspended precipitant produced by the NaCl-induced precipitation. As with the lysate pellet, product observed in this fraction could be unfolded protein that denatured during the precipitation. However, by nature, the Sso7d scaffold is thermally stable for extended periods of time at temperatures far above the 37 °C used in the precipitation here. Thus, an alternative explanation is more likely. The other possibility is that residual salt in the precipitated pellet raised the overall salinity of the suspension, and that this residual salinity caused a decrease in the solubility of the ELP tag. Regardless of the source of these product losses, the majority of the desired product localizes in the collected fractions (lanes 4 and 7). As described below, these final products have a sufficiently high yield to perform the desired experiments, and, as such, no further optimization was performed on the purification process.

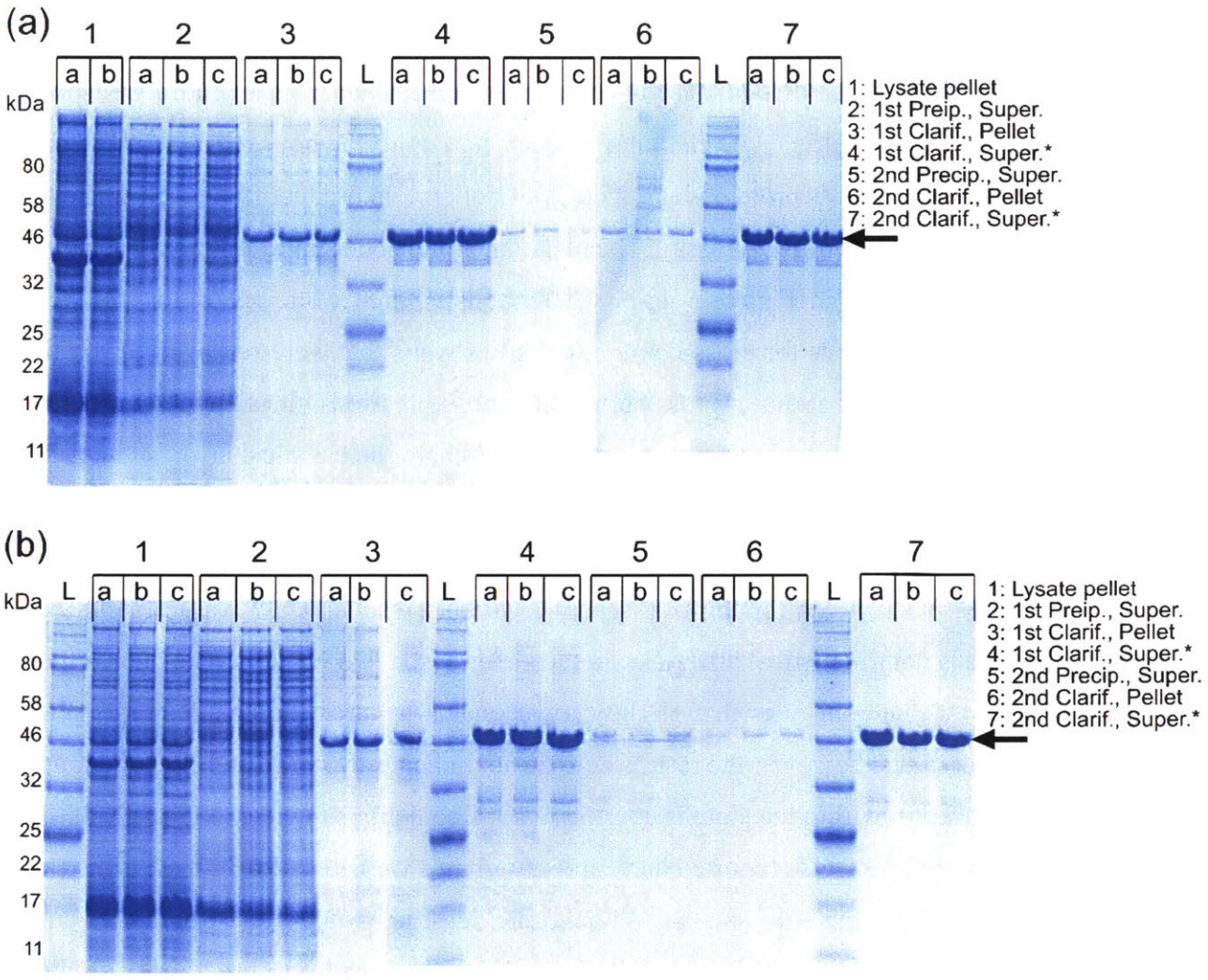


Figure 6-3. Coomassie blue stained SDS-PAGE gels on ELP-Sso7d-ZE proteins containing either an (a) engineered Sso7d (rcSso7d.SA) or a (b) non-binding rcSso7d variant. Lysate pellet refers to the solid pellet produced upon lysate clarification. In the lane labels, 1st and 2nd refer to the first and second purification cycles. Fractions are labeled as in Chapter 5, **Figure 5-6**a: "Precip." refers to samples after incubation with precipitant (1.5 M NaCl, 37 ˚C or 30 vol% EtOH, 4 ˚C) and subsequent centrifugation to separate soluble supernatant ("Super.") and insoluble ("Pellet") fractions. "Clarif." refers to samples after resuspension in water and clarification by centrifugation to separate the soluble supernatant ("Super.") and insoluble ("Pellet") fractions. The black arrow indicates the expected location of the product band. In each gel, fractions 4 and 7 are expected to contain the majority of the desired product (as indicated by the asterisk). For each fraction, labels a, b, and c indicate replicate samples from random locations in the plate.

Final protein yields for both ELP-Sso7d-ZE constructs were consistent across an entire well-plate expression. Yields were also sufficient to run binding characterizations in both self-assembled and dilute solution formats. To assess protein yield, 25 μL of final protein product from each well was used to run a bicinchnoninic acid (BCA) assay. The resulting protein concentrations measured over the well plate are represented in **Figure 6-4**a-b, where wells colored white are wells that were either not inoculated (as controls for cross-contamination) or wells that were removed to run the SDS-PAGE analysis described above. Across both well plate expressions, the average expression yield was calculated to be 60. ± 8 mg protein/ L expression (where the technical error in the BCA assay for this run, on average, was 1.9 ± 0.8 mg/L for samples of similar concentration). This corresponds to an average protein concentration of 0.48 ± 0.06 mg/mL in each well. An unpaired *t* test on the average plate yields for the two separate variants results in a *P* value of 0.8291, suggesting that there is not a statistical difference in the yields of the two variants. The distribution of yield over the well plate suggests that slightly lower yields towards the center of the plate compared to the outer wells; however, the results here do not indicate any trend significant enough to suggest that the location within the plate is causing major fluctuations in yield (**Figure 6-4a-b**). Based on the average protein yields reported here, and assuming a well area of 0.32 cm^2, films cast out of 75 μL of purified protein solutions should have site densities of 3.0 ± 0.4 μmol sites/cm^2. This site density is the same order of magnitude as that calculated for previously studied rcSso7d.SA bioconjugate films, and, as such, it is expected that the binding will be detectable by previously utilized microscopy methods. [29] Importantly, because the final volume of the protein solution produced in these experiments was 125 μL, there is sufficient solution to cast films (75 μL), run BCA (25 μL) to establish yield, and run dilute solution BLI measurements (≤ 25 μL) for each individual well.

(a)

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.64	0.61	0.46	0.47		0.44	0.45	0.44		0.39	0.42	0.43
B	0.79	0.63	0.41	0.40	0.44	0.38	0.38	0.42	0.36	0.37	0.45	0.46
C	0.55	0.55			0.47	0.46	0.43	0.42	0.43	0.43	0.50	0.44
D	0.64	0.58	0.50	0.49	0.55	0.46	0.42	0.53	0.50	0.45	0.46	0.51
E	0.50	0.49	0.40	0.43		0.44	0.36	0.40	0.39	0.41	0.42	0.47
F	0.50			0.45			0.42			0.50		
G	0.59	0.57	0.50	0.52	0.47	0.50	0.49	0.46		0.50	0.50	0.65
H	0.57	0.49	0.53	0.48	0.47	0.45	0.48	0.53	0.55	0.52	0.54	0.66

(b)

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.52	0.51	0.44	0.44		0.52	0.48	0.51		0.48	0.44	0.54
B	0.55	0.50	0.42	0.47	0.55	0.46	0.53	0.55	0.53	0.47	0.51	0.54
C	0.54	0.51			0.53	0.51	0.50	0.50	0.47	0.46	0.51	0.59
D	0.49	0.48	0.44	0.43	0.44	0.49	0.48	0.47	0.50	0.48	0.54	0.57
E		0.50			0.51			0.48			0.58	
F	0.44	0.51	0.42	0.44		0.44	0.40	0.42	0.44	0.41	0.41	0.55
G	0.54	0.49	0.45	0.42	0.44	0.40	0.41	0.48		0.43	0.44	0.52
H	0.51	0.51	0.49	0.48	0.49	0.41	0.44	0.43	0.41	0.44	0.42	0.50

(c)

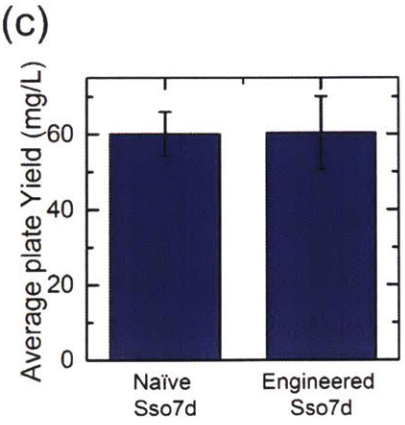


Figure 6-4. ELP-Sso7d-ZE yield in final protein product determined by BCA. White boxes in (a) and (b) correspond to wells containing no protein sample. Color scale in (a) and (b) ranges from green (highest protein concentration) to red (lowest protein concentration). Final protein yield across an entire well-plate expression for ELP-Sso7d-ZE with an (a) engineered Sso7d variant, rcSso7d.SA and a (b) naïve Sso7d variant with no specific streptavidin affinity. All values reported in these tables are in mg/mL. (b) Average yield over the entire plate for the two Sso7d variants. Yield is reported in mg protein produced per L culture.

6.4.3 Quantification of ELP-Sso7d-ZE Binding

The final viability assessment for the ELP-Sso7d-ZE system was to establish that streptavidin binding affinity of different Sso7d variants can be quantified and compared in both self-assembled and dilute solution formats. This was demonstrated for both formats using the same two Sso7d variants described above—an engineered variant, rcSso7d.SA and a naïve variant with no specific streptavidin affinity. [13, 29] The main goal of this portion of the study was not to characterize the binding affinities of these two variants in detail. Rather, it was to determine whether available techniques could distinguish between these two Sso7d variants in both dilute solution and self-assembled formats. Thus, after determining the yield of the purified ELP-Sso7d-ZE products, three biological replicates of each variant (naïve and engineered) were selected, and their binding affinities in dilute solution and self-assembled formats were characterized using bio-layer interferometry (BLI) and fluorescence microscopy, respectively.

Dilute solution binding affinity characterized by BLI showed a detectable difference in response between the two Sso7d variants. Purified protein samples for BLI were taken from the well-plate expressions characterized in the previous section (3 biological replicates per variant). Accurate binding characterization by BLI requires knowledge of the binding protein concentration; thus, it was quite important to have sufficient yield not only to run BLI measurements, but also to run a BCA assay to determine the protein concentration. After determining protein concentration in each well, the final concentration of each replicate was adjusted to 100 nM in 1x PBS, 0.1% w/v BSA, and 0.02% v/v Tween-20. Binding affinity of the ELP-Sso7d-ZE constructs was then evaluated by performing BLI with streptavidin (SA) biosensor tips. The relative response of the SA tips to each ELP-Sso7d-ZE was then calculated using curve fitting. The results of this fit are shown in **Figure 6-5**. As seen in this plot, the relative response of the constructs with the different Sso7d sequences is quite distinct. The response of the naïve Sso7d variant is, within error, zero. In contrast, the rc.Sso7d.SA variant shows strong binding affinity, with a normalized response of 0.37 ± 0.2. Notably, the error reported here is the standard deviation in the measured affinity over the three different biological replicates measured. This indicates that the purification strategy detailed above paired with concentration adjustment after BCA assay allows for accurate determination of relative dilute solution binding affinity of these ELP-Sso7d-ZE constructs. While this technique gives a readout of relative binding affinity, comparison to previously reported binding affinities would require performing this BLI measurement at various ELP-Sso7d-ZE concentrations.

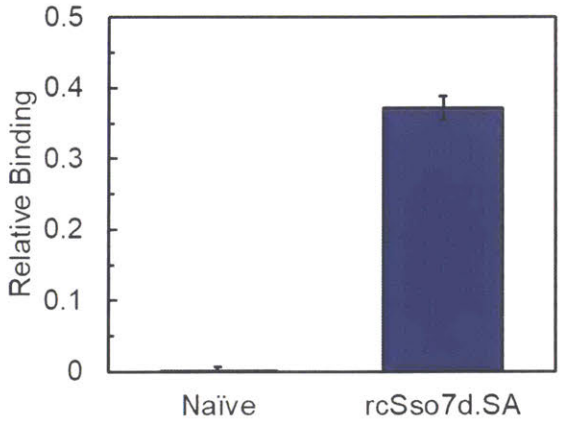


Figure 6-5. Relative response (normalized to ELP-Sso7d-ZE concentration) as calculated by curve fits to BLI data taken on ELP-Sso7d-ZE containing an engineered Sso7d sequence (rcSso7d.SA) previously shown to have strong streptavidin binding, and a naïve Sso7d sequence

expected not to bind streptavidin. Error bars represent standard deviation over three biological replicates.

Fluorescence microscopy on immobilized, self-assembled ELP-Sso7d-ZE films also produced a differential response for the two different Sso7d variants. Self-assembled films of ELP-Sso7d-ZE (again, 3 biological replicates per Sso7d variant) were generated by vacuum drying 75 μL of purified protein solution (see **Sections 6.4.2, 6.3.1**) in a 96-well microtiter plate. Films were then immobilized at the bottom of the wells using glutaraldehyde crosslinking. After crosslinking, streptavidin binding affinity of the films was evaluated by first incubating the films with fluorescently labeled streptavidin, and (after washing of unbound streptavidin) measuring the bound streptavidin by fluorescence microscopy. Fluorescence intensity in the films after binding for each of the Sso7d variants is shown in **Figure 6-6**. Here, mean fluorescence intensity is averaged over a section of the imaged film. As with the BLI, there is a distinct difference in the signal produced by the two Sso7d variants. However, this measurement produces a much higher noise level in the film expected not to bind streptavidin, as evidenced by the large error bar on the average fluorescent intensity produced by the naïve Sso7d variant. Based on this result, the procedure used to detect streptavidin binding in films should be subject to further optimization.

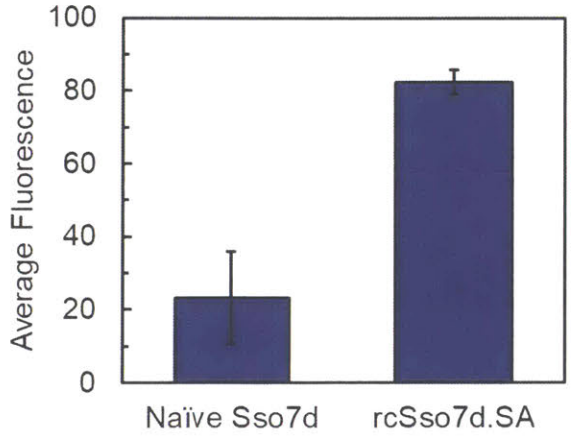


Figure 6-6. Mean fluorescence intensity (measured by fluorescence microscopy) of ELP-Sso7d-ZE films containing two Sso7d variants—a naïve Sso7d variant, expected not to bind streptavidin, and an engineered variant, rcSso7d.SA, which has been demonstrated to bind

streptavidin in thin film formats previously. Error bars correspond to standard deviation over three replicate films.

The characterizations described here enable the relative quantitation of rcSso7d streptavidin binding in both dilute solution and in self-assembled films. It is important to note that these relative values do not permit direct comparison of the K_d of a given rcSso7d in both formats. However, comparison of the relative binding affinities in these two formats should provide insight into whether the relative binding affinities in dilute solution and self-assembled formats scale similarly. From this study, a set of Sso7d variants with disparate performance in self-assembled films and dilute solution can be identified, and further characterized, thus enabling what protein features are key to performance in these different environments.

6.5 Conclusions

A system for comparing binding affinity of a library of streptavidin binding proteins in dilute solution and self-assembled formats was developed. This was done using a fusion protein construct of an Sso7d library (generated in a previous study) fused to an ELP and a coiled-coil sequence. In this construct, the ELP served as a polymer-like block to enable self-assembly into block copolymer like structures, and the coiled-coil sequence (ZE) acted as an order-directing peptide sequence. This work demonstrated that this construct self-assembles in concentrated solution and forms microphase separated domains in the solid state. Constructs containing two different Sso7d sequences—one engineered to have binding affinity (rcSso7d.SA) and one naïve variant—were expressed and purified in a well-plate format. Measurements on final protein products indicate low well-to-well variability in yield and purity, indicating that these constructs are amenable to this high-throughput processing. Finally, BLI and fluorescence microscopy were used to characterize the binding affinities of the two Sso7d variants in dilute solution and self-assembled films, respectively. Results of these characterizations suggest that both characterizations are capable of differentiating binding affinities of the two constructs. However, the procedure used for detection of film binding, including binding of the fluorescently labeled streptavidin as well as the microscopy process should be subject to further optimization prior to running a full-scale study.

6.6 References

1. Engvall, E., Enzyme immunoassay ELISA and EMIT. In *Method Enzymol*, Academic Press: 1980; Vol. 70, pp 419-439.

2. Waldmann, T. A., Monoclonal-Antibodies in Diagnosis and Therapy. *Science* **1991,** *252* (5013), 1657-1662.

3. Goldman, E. R.; Anderson, G. P.; Conway, J.; Sherwood, L. J.; Fech, M.; Vo, B.; Liu, J. L.; Hayhurst, A., Thermostable Llama Single Domain Antibodies for Detection of Botulinum A Neurotoxin Complex. *Anal. Chem.* **2008,** *80* (22), 8583-8591.

4. Chin, C. D.; Laksanasopin, T.; Cheung, Y. K.; Steinmiller, D.; Linder, V.; Parsa, H.; Wang, J.; Moore, H.; Rouse, R.; Umviligihozo, G.; Karita, E.; Mwambarangwe, L.; Braunstein, S. L.; van de Wijgert, J.; Sahabo, R.; Justman, J. E.; El-Sadr, W.; Sia, S. K., Microfluidics-based diagnostics of infectious diseases in the developing world. *Nat. Med.* **2011,** *17* (8), 1015-U138.

5. Miller, E. A.; Baniya, S.; Osorio, D.; Al Maalouf, Y. J.; Sikes, H. D., Paper-based diagnostics in the antigen-depletion regime: High-density immobilization of rcSso7d-cellulose-binding domain fusion proteins for efficient target capture. *Biosens. Bioelectron.* **2018,** *102*, 456-463.

6. Borrebaeck, C. A. K., Antibodies in diagnostics - from immunoassays to protein chips. *Immunol Today* **2000,** *21* (8), 379-382.

7. Hey, T.; Fiedler, E.; Rudolph, R.; Fiedler, M., Artificial, non-antibody binding proteins for pharmaceutical and industrial applications. *Trends Biotechnol.* **2005,** *23* (10), 514-522.

8. Pluckthun, A., Designed Ankyrin Repeat Proteins (DARPins): Binding Proteins for Research, Diagnostics, and Therapy. *Annu Rev Pharmacol* **2015,** *55*, 489-511.

9. Grant, S.; Davis, F.; Law, K. A.; Barton, A. C.; Collyer, S. D.; Higson, S. P. J.; Gibson, T. D., Label-free and reversible immunosensor based upon an ac impedance interrogation protocol. *Anal. Chim. Acta* **2005,** *537* (1-2), 163-168.

10. Zou, Z. W.; Kai, J. H.; Rust, M. J.; Han, J.; Ahn, C. H., Functionalized nano interdigitated electrodes arrays on polymer with integrated microfluidics for direct bio-affinity sensing using impedimetric measurement. *Sensor Actuat a-Phys* **2007,** *136* (2), 518-526.

11. Cheng, C. M.; Martinez, A. W.; Gong, J. L.; Mace, C. R.; Phillips, S. T.; Carrilho, E.; Mirica, K. A.; Whitesides, G. M., Paper-Based ELISA. *Angew Chem Int Edit* **2010,** *49* (28), 4771-4774.

12. Gu, Z.; Zhao, M. X.; Sheng, Y. W.; Bentolila, L. A.; Tang, Y., Detection of Mercury Ion by Infrared Fluorescent Protein and Its Hydrogel-Based Paper Assay. *Anal. Chem.* **2011,** *83* (6), 2324-2329.

13. Miller, E. A.; Traxlmayr, M. W.; Shen, J.; Sikes, H. D., Activity-based assessment of an engineered hyperthermophilic protein as a capture agent in paper-based diagnostic tests. *Mol. Syst. Des. Eng.* **2016,** *1* (4), 377-381.

14. Zhu, H.; Snyder, M., Protein arrays and microarrays. *Curr. Opin. Chem. Biol.* **2001,** *5* (1), 40-45.

15. Kim, D.; Herr, A. E., Protein immobilization techniques for microfluidic assays. *Biomicrofluidics* **2013,** *7* (4).

16. Boozer, C.; Ladd, J.; Chen, S. F.; Yu, Q.; Homola, J.; Jiang, S. Y., DNA directed protein immobilization on mixed ssDNA/oligo(ethylene glycol) self-assembled monolayers for sensitive biosensors. *Anal. Chem.* **2004,** *76* (23), 6967-6972.

17. Schoffelen, S.; Lambermon, M. H. L.; van Eldijk, M. B.; van Hest, J. C. M., Site-specific modification of Candida antarctica lipase B via residue-specific incorporation of a non-canonical amino acid. *Bioconjugate Chem.* **2008,** *19* (6), 1127-1131.

18. Raliski, B. K.; Howard, C. A.; Young, D. D., Site-Specific Protein Immobilization Using Unnatural Amino Acids. *Bioconjugate Chem.* **2014,** *25* (11), 1916-1920.

19. Wu, J. C. Y.; Hutchings, C. H.; Lindsay, M. J.; Werner, C. J.; Bundy, B. C., Enhanced Enzyme Stability Through Site-Directed Covalent Immobilization. *J. Biotechnol.* **2015,** *193*, 83-90.

20. Hutsell, S. Q.; Kimple, R. J.; Siderovski, D. P.; Willard, F. S.; Kimple, A. J., High-Affinity Immobilization of Proteins Using Biotin- and GST-Based Coupling Strategies. *Methods Mol Biol* **2010,** *627*, 75-90.

21. Mateo, C.; Fernandez-Lorente, G.; Cortes, E.; Garcia, J. L.; Fernandez-Lafuente, R.; Guisan, J. M., One-step purification, covalent immobilization, and additional stabilization of poly-His-tagged proteins using novel heterofunctional chelate-epoxy supports. *Biotechnol. Bioeng.* **2001,** *76* (3), 269-276.

22. Zhang, K. C.; Diehl, M. R.; Tirrell, D. A., Artificial polypeptide scaffold for protein immobilization. *J. Am. Chem. Soc.* **2005,** *127* (29), 10136-10137.

23. Lam, C. N.; Olsen, B. D., Phase transitions in concentrated solution self-assembly of globular protein-polymer block copolymers. *Soft Matter* **2013,** *9* (8), 2393-2402.

24. Thomas, C. S.; Olsen, B. D., Coil fraction-dependent phase behaviour of a model globular protein-polymer diblock copolymer. *Soft Matter* **2014,** *10* (17), 3093-3102.

25. Thomas, C. S.; Xu, L.; Olsen, B. D., Kinetically Controlled Nanostructure Formation in Self-Assembled Globular Protein–Polymer Diblock Copolymers. *Biomacromolecules* **2012,** *13* (9), 2781-2792.

26. Huang, A.; Olsen, B. D., Self-Assembly of Differently Shaped Protein–Polymer Conjugates through Modification of the Bioconjugation Site. *Macromol. Rapid Commun.* **2016,** *37* (15), 1268-1274.

27. Huang, A.; Qin, G.; Olsen, B. D., Highly Active Biocatalytic Coatings from Protein–Polymer Diblock Copolymers. *ACS Appl. Mater. Interfaces* **2015,** *7* (27), 14660-14669.

28. Dong, X.-H.; Obermeyer, A. C.; Olsen, B. D., Three-Dimensional Ordered Antibody Arrays Through Self-Assembly of Antibody–Polymer Conjugates. *Angew. Chem. Int. Ed.* **2017,** *56* (5), 1273-1277.

29. Paloni, J. M.; Miller, E. A.; Sikes, H. D.; Olsen, B. D., Improved Ordering in Low Molecular Weight Protein–Polymer Conjugates Through Oligomerization of the Protein Block. *Biomacromolecules* **2018,** *19* (9), 3814-3824.

30. Obermeyer, A. C.; Olsen, B. D., Synthesis and Application of Protein-Containing Block Copolymers. *Acs Macro Lett* **2015,** *4* (1), 101-110.

31. Qin, G.; Glassman, M. J.; Lam, C. N.; Chang, D.; Schaible, E.; Hexemer, A.; Olsen, B. D., Topological Effects on Globular Protein-ELP Fusion Block Copolymer Self-Assembly. *Adv. Funct. Mater.* **2015,** *25* (5), 729-738.

32. Qin, G.; Perez, P. M.; Mills, C. E.; Olsen, B. D., Effect of ELP Sequence and Fusion Protein Design on Concentrated Solution Self-Assembly. *Biomacromolecules* **2016,** *17* (3), 928-934.

33. Nord, K.; Gunneriusson, E.; Ringdahl, J.; Stahl, S.; Uhlen, M.; Nygren, P. A., Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat. Biotechnol.* **1997,** *15* (8), 772-777.

34. Schneider, S.; Buchert, M.; Georgiev, O.; Catimel, B.; Halford, M.; Stacker, S. A.; Baechi, T.; Moelling, K.; Hovens, C. M., Mutagenesis and selection of PDZ domains that bind new protein targets. *Nat. Biotechnol.* **1999,** *17* (2), 170-175.

35. Hecht, M. H.; Das, A.; Go, A.; Bradley, L. H.; Wei, Y., De novo proteins from designed combinatorial libraries. *Protein Sci.* **2004,** *13* (7), 1711-1723.

36. Binz, H. K.; Amstutz, P.; Pluckthun, A., Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **2005,** *23* (10), 1257-1268.

37. Boder, E. T.; Wittrup, K. D., Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **1997,** *15* (6), 553-557.

38. Wentzel, A.; Christmann, A.; Adams, T.; Kolmar, H., Display of passenger proteins on the surface of Escherichia coli K-12 by the enterohemorrhagic E-coli intimin EaeA. *J. Bacteriol.* **2001,** *183* (24), 7273-7284.

39. Bessette, P. H.; Rice, J. J.; Daugherty, P. S., Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Eng. Des. Sel.* **2004,** *17* (10), 731-739.

40. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012,** *9* (7), 671-675.

41. Mills, C. E.; Michaud, Z.; Olsen, B. D., Elastin-like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins. *Biomacromolecules* **2018,** *19* (7), 2517-2525.

42. Baumann, H.; Knapp, S.; Lundback, T.; Ladenstein, R.; Hard, T., Solution Structure and DNA-Binding Properties of a Thermostable Protein from the Archaeon Sulfolobus-Solfataricus. *Nature Structural Biology* **1994,** *1* (11), 808-819.

43. Lam, C. N.; Kim, M.; Thomas, C. S.; Chang, D.; Sanoja, G. E.; Okwara, C. U.; Olsen, B. D., The Nature of Protein Interactions Governing Globular Protein–Polymer Block Copolymer Self-Assembly. *Biomacromolecules* **2014,** *15* (4), 1248-1258.

Chapter 7 : Conclusions

7.1 Summary

The goal of this thesis was to expand the ways in which elastin-like polypeptides (ELPs) could be leveraged in both the self-assembly and high-throughput processing of functional protein materials. Previous work by Qin *et. al.* demonstrated that an ELP-mCherry model system self-assembled in a manner similar to that of poly(*N*-isopropyl acrylamide) (PNIPAM)-mCherry bioconjugates. [1-2] Unlike the bioconjugate system, however, these ELP-mCherry fusion proteins were completely biosynthetic, and, as such, could be manipulated precisely using biomolecular engineering. This thesis aimed to exploit these advantages for two different scientific goals. First, using the precise control afforded by genetic engineering techniques, the role of ELP charge and hydrophobicity on ELP-mCherry self-assembly was assessed. Second, a platform for high-throughput fabrication of self-assembled fusion protein materials was developed.

A panel of ELPs with varying charge and hydrophobicity was used to systematically explore the impact of ELP charge and hydrophobicity on fusion protein self-assembly. Four new ELP sequences were designed to this end—two negatively charged and two charge balanced, with each charge type having a more and less hydrophobic ELP variant. These ELPs were then fused to mCherry, and the self-assembly of these fusion proteins in concentrated solution was evaluated. The addition of charge to the ELP block, regardless of whether it was net negative or net neutral, decreased the propensity for fusion protein self-assembly. Dilute solution characterization indicated that there were no strong interactions between the ELPs and the mCherry that explained this decrease in self-assembly. Ultimately, this study suggested that addition of charge to the ELP block likely decreases the charge cohesion asymmetry in the fusion protein, which leads to a decrease in the effective repulsion between the ELP and mCherry domains. [3] This is in agreement with a previous study in which a bioconjugate composed of a zwitterionic polymer, poly-3-dimethyl(methacryloyloxyethyl)ammonium propane sulfonate (PDMAPS) conjugated to mCherry, also showed comparably lower propensity for self-assembly than bioconjugates containing an uncharged polymer block. [1, 4-5]

The second goal of this thesis was to develop a platform for high-throughput fabrication of these self-assembled fusion protein materials. Realization of this goal required high-throughput purification of fusion proteins. ELP-based precipitation methods have the capability of addressing the purification challenge; however, they typically result in final protein solutions with relatively

high salinity. This high salinity can impact the performance of the highly concentrated or solid materials of interest in this work. Thus, a strategy for desalting these ELP fusion proteins in high throughput was necessary. With this goal in mind, the possibility of cononsolvency behavior in ELP solutions was investigated. ELP cononsolvency was found to exist in blends of water with four different alcohols—methanol, ethanol, isopropanol, and 1-propanol. In certain water/alcohol blends, ELPs were found to exhibit upper critical solution temperature (UCST) transitions, in which the ELP was soluble at high temperatures and insoluble at low temperatures. Trends in ELP cononsolvency with alcohol hydrophobicity were compared to those observed in PNIPAM to probe whether the mechanism for cononsolvency is similar in these two systems. These comparisons led to the conclusion that the driving force for cononsolvency in the ELP and PNIPAM are similar in the region where lower critical solution temperature (LCST) transitions are observed. However, where UCST transitions occur, the driving force for this cononsolvency differs between the ELP and PNIPAM, likely due to the existence of secondary structure in ELP systems.

Next, this ELP cononsolvency was evaluated for use in high throughput protein purification and desalting. This was done using a model system of superfolder green fluorescent protein (sfGFP) fused to an ELP, with ethanol as the cononsolvent. These studies showed that ethanol-induced precipitations of ELP-sfGFP did not purify the protein as effectively as NaCl-induced precipitations. However, using NaCl-induced precipitation for purification followed by ethanol-induced precipitation for desalting results in a purified, desalted protein product. This was demonstrated to be effective in purification of ELP-sfGFP expressions produced both on the well-plate scale and on the fermenter scale. This procedure was also effective in purification and desalting of an ELP-based protein gel (with applications in tissue engineering). This purification/desalting approach did not work when the ELP was fused to the enzyme organophosphate hydrolase (OPH), which is known to be unstable in ethanol. This indicates a key limitation of the approach—that the various components in the fusion protein of interest must be stable to ethanol.

This procedure for fusion protein purification and desalting represents a powerful tool for fabrication ELP-based protein materials in a high-throughput format. The final portion of this thesis assessed whether this platform could be used to answer the question of how self-assembly impacts globular protein function. This was done using a fusion protein composed of an ELP fused

a small streptavidin binding protein, Sso7d, engineered in the Sikes lab to have affinity for streptavidin. [6-7] This construct was confirmed to assemble in concentrated solution. Two different Sso7d variants were then incorporated into this construct—one with strong affinity for streptavidin, and one with no affinity. These variants were expressed and purified in a well-plate format using the previously developed purification/desalting procedure, which reproducibly gave high yield/purity products for both Sso7d variants. The affinity of these two variants was found to be differentiable, both in dilute solution and in self-assembled films. These results indicate that this system will enable a binding affinity comparison of a library of different Sso7d variants (available from previous engineering efforts in the Sikes lab) in dilute solution and self-assembled formats.

Overall, this work demonstrates the ability of ELP tags to both drive self-assembly and enable high-throughput processing of functional protein materials.

7.2 Outlook

ELP-based protein materials have garnered increasing interest in recent years, and, as such, the engineering applications of these materials has grown. The work presented here focuses on self-assembled fusion proteins comprised of an ELP and a globular protein partner. However, similar principles to those used in the design of the charged ELP panel could be applied to ELPs used in other fusion constructs. Such comparative studies afford opportunities to learn important design principles for assembled protein materials, not only in the context shown here, but also in protein hydrogels, dilute solution protein assemblies, and in layer-by-layer assemblies. One drawback to these combinatorial studies, however, is their relatively low throughput. Given that the building blocks for these proteins are 20 different naturally-occurring amino acids, and that there are hundreds of amino acids per protein, there exists an immense number of potential candidates for comparison.

For globular proteins (both binding proteins and enzymes), this challenge has been addressed using directed evolution, which is performed in the following manner.[8-10] Starting with an existing protein sequence, random changes are made to building a library of protein sequences. These new sequences are then screened for some feature of interest (such as improved thermal stability, or activity towards a new substrate), and protein sequences that exhibit improvement in the feature of interest are selected for further iterations of this same process.[11] The success of this approach

relies on production of a large library of candidates and the ability to reliably screen these candidates for the property of interest. For protein materials, these criteria present several engineering challenges. The ability to perform high-throughput screening for the properties of interest, for example, is highly dependent on the material of interest. More universal, however, is the purification challenge. In the case of globular proteins, protein performance in crude lysates or on cell surfaces is considered adequately representative of the protein performance in the application of interest. For protein materials, however, the properties of interest are inherently related to the final purity and processing of the material. Thus, control over protein purity and final solution conditions is critical for the success of directed evolution studies on protein materials. In 2014, work out of the Kaplan group used ELP-based purification to enable high throughput production of ELP-silk fusion proteins.[12] The authors then screened the library for stimuli responsive properties. However, this study did not consider how final solution properties, such as salinity, might affect the observed responses.

The high throughput procedure for the purification and desalting of ELP-based materials developed in this thesis overcomes this challenge. As a result, this work presents numerous exciting opportunities for future research. For example, the effects of self-assembly on enzyme performance could also be evaluated, as the impacts of local environment likely have different impacts on enzymes than binding proteins. If these studies indicate that self-assembly impacts the final properties of the enzyme or binding protein, this high-throughput processing platform would enable directed evolution on the globular protein of interest. This platform could even be used to evolve globular proteins that are more stable to ethanol-based processing. Paired with efforts to enable high-throughput screening of self-assembled materials at synchrotrons, this platform may also find uses in producing libraries of materials that can subsequently be screened for certain self-assembly features. The scalability of this precipitation-based desalting method may also find uses in industrial scale purification and production of protein materials, especially in cases where profit margins are insufficient to justify chromatography-based methods. This goes beyond globular protein-ELP fusion proteins, and includes any protein material that incorporates an ELP block, including those considered in tissue engineering applications.[13]

More broadly, there are many opportunities in improving throughput for the combinatorial studies of protein materials, extending beyond those that can incorporate an ELP fusion. In these cases, it may be possible to consider the approach used here to develop a high throughput desalting method,

in which known polymer solution phenomena inspired investigations into polypeptide solution phase behavior. In this paradigm, more rapid understanding of novel protein material systems can be achieved by considering what is already known (but also what is distinctly different) about similar synthetic polymer systems.

On a fundamental level, further exploration of the ELP cononsolvency phenomenon reported here also opens the door to many interesting scientific questions. Using the precise control offered by genetic engineering, the effect of different ELP features, such as charge and hydrophobicity, on this cononsolvency behavior can be studied. Such studies paired with ongoing efforts to understand the driving force for PNIPAM cononsolvency could deepen fundamental understanding of the thermodynamic underpinnings of this phenomenon.

7.3 References

1. Lam, C. N.; Olsen, B. D., Phase transitions in concentrated solution self-assembly of globular protein-polymer block copolymers. *Soft Matter* **2013,** *9* (8), 2393-2402.

2. Qin, G.; Glassman, M. J.; Lam, C. N.; Chang, D.; Schaible, E.; Hexemer, A.; Olsen, B. D., Topological Effects on Globular Protein-ELP Fusion Block Copolymer Self-Assembly. *Adv. Funct. Mater.* **2015,** *25* (5), 729-738.

3. Sing, C. E.; Zwanikken, J. W.; Olvera de la Cruz, M., Electrostatic control of block copolymer morphology. *Nat. Mater.* **2014,** *13* (7), 694-8.

4. Chang, D.; Lam, C. N.; Tang, S.; Olsen, B. D., Effect of polymer chemistry on globular protein-polymer block copolymer self-assembly. *Polym. Chem.* **2014,** *5* (17), 4884-4895.

5. Chang, D.; Olsen, B. D., Self-assembly of protein-zwitterionic polymer bioconjugates into nanostructured materials. *Polym. Chem.* **2016,** *7* (13), 2410-2418.

6. Miller, E. A.; Traxlmayr, M. W.; Shen, J.; Sikes, H. D., Activity-based assessment of an engineered hyperthermophilic protein as a capture agent in paper-based diagnostic tests. *Mol. Syst. Des. Eng.* **2016,** *1* (4), 377-381.

7. Paloni, J. M.; Miller, E. A.; Sikes, H. D.; Olsen, B. D., Improved Ordering in Low Molecular Weight Protein–Polymer Conjugates Through Oligomerization of the Protein Block. *Biomacromolecules* **2018,** *19* (9), 3814-3824.

8. Arnold, F. H.; Volkov, A. A., Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol.* **1999,** *3* (1), 54-59.

9. Marshall, S. A.; Lazar, G. A.; Chirino, A. J.; Desjarlais, J. R., Rational design and engineering of therapeutic proteins. *Drug Discovery Today* **2003,** *8* (5), 212-221.

10. Lutz, S., Beyond directed evolution-semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **2010,** *21* (6), 734-743.

11. Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **2015,** *16* (7), 379-394.

12. Wang, Q.; Xia, X. X.; Huang, W. W.; Lin, Y. N.; Xu, Q. B.; Kaplan, D. L., High Throughput Screening of Dynamic Silk-Elastin-Like Protein Biomaterials. *Adv. Funct. Mater.* **2014,** *24* (27), 4303-4310.

13. Glassman, M. J.; Avery, R. K.; Khademhosseini, A.; Olsen, B. D., Toughening of Thermoresponsive Arrested Networks of Elastin-Like Polypeptides To Engineer Cytocompatible Tissue Scaffolds. *Biomacromolecules* **2016,** *17* (2), 415-426.

Appendix A. Supporting Information for Chapter 2

A1. Lab Protocols

A1.1: PCR with Q5 Polymerase Master Mix

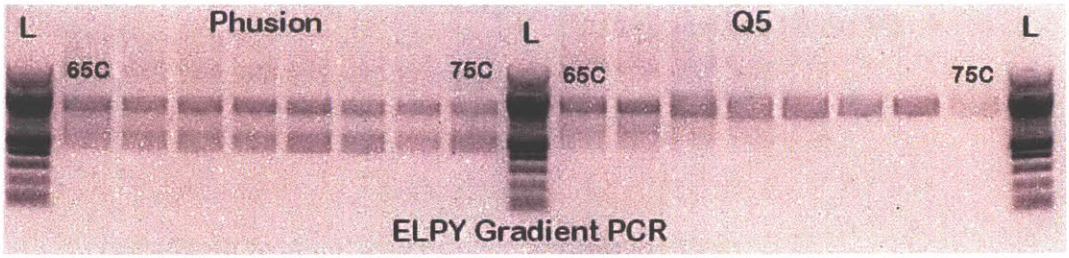
Goal

To perform PCR on a gene of interest

Materials Needed

- **Primers** (Forward and Reverse)
    - Stock solutions of the primers (as received from IDT) should be prepared at 100 μM. Solutions that you will use in PCR should be diluted to 10 μM for everyday use.
    - Ensure that the overhang sequence in your forward and reverse primers is specific to the part of the sequence you are trying to PCR (that is, it does not have high similarity to a different part of the same sequence
    - Ensure that the part of the primers that will be annealing to your template DNA have similar melting temperatures. This can be calculated using an online Tm calculator
        - https://tmcalculator.neb.com/#!/
    - https://www.addgene.org/protocols/pcr-cloning/
    - To get the best melting temperature, run a gradient PCR with your primers for 35 cycles (or the max cycles for your polymerase), from temperatures that are 5C above and 5C below the recommended annealing temperature from the calculator. You want to use a temperature that maximizes your yield without creating any side products. Lower temp → Better yield, but nonspecific products, higher temp → lower yield, but less/no nonspecific products. After the gradient PCR, run your products on an agarose gel to visualize the results.

      Example:



Here, the phusion polymerase always resulted in multiple PCR products. In the case of Q5, multiple products are observed at lower temperatures, but eliminated at higher temperatures. Thus, a temperature closer to 75 °C using the Q5 polymerase is probably best in this instance.

- **Template DNA**
  - This is the DNA that contains the fragment that you wish to amplify. For best PCR results, the amount of template DNA in your reaction mixture should be kept within the range recommended by the manufacturer of the polymerase (typically 1 pg – 1 ng for plasmid DNA). You may need to do serial dilutions of your DNA to achieve this. Generally, you can prep a stock of 0.1 ng/µL of your template DNA and use 10 uL of this stock in each of your PCR reactions
- **Q5 MasterMix**
  - This contains all the necessary enzyme, buffers, and nucleic acids for PCR. This mastermix should be kept in the -20 °C freezer when not in use and should be thawed at 4 °C prior to use for best results. This is easiest to do by placing the master mix vial on the rocker in the 4 °C fridge ~1 hr prior to running PCR. Never let the mastermix heat up to room temperature or above. When thawed, this mastermix sometimes has some solids in it. Gently flick the tube to redissolve them and then use the small centrifuge to collect any liquid that may have gotten into the lid

Procedure

1. Prepare a mastermix of your primers and template DNA:
   - **25 uL PCR Reactions**
     - 10 µL of 0.1 ng/µL template DNA x (number of reactions + 1)
     - 1.25 µL of 10 µM forward primer x (number of reactions + 1)
     - 1.25 µL of 10 µM reverse primer x (number of reactions + 1)

- **50 uL PCR Reactions**
    - 10 µL of 0.1 ng/uL template DNA x (number of reactions + 1)
    - 10 µL of MilliQ water x (number of reactions + 1)
    - 2.5 µL of 10 uM forward primer x (number of reactions + 1)
    - 2.5 µL of 10 uM reverse primer x (number of reactions + 1)

2. Thaw the Q5 mastermix on ice or in water that has been brought to 4 °C. If you see solid particulate formed upon thawing, mix gently by flicking the tube until they redissolve
3. Set up your PCR program on the thermal cycler using the following setup with the melting temperature determined from your gradient PCR (Tm is the starred temperature in the table below). For the best results, run 25 cycles.

STEP	**TEMP**	**TIME**
Initial Denaturation	98 °C	30 seconds
25–35 Cycles	98 °C	5–10 seconds
	*50–72 °C	10–30 seconds
	72 °C	20–30 seconds/kb
Final Extension	72 °C	2 minutes
Hold	4–10 °C	

4. Pipette Q5 PCR mastermix (12.5 for 25 µL reaction, 25 µL for 50 µL reaction) into PCR tubes
5. Add your primer/template DNA mastermix to the PCR tubes (12.5 µL for 25 µL reaction, 25 µL for 50 µL reaction)
6. Mix by flicking gently and spin down in small PCR tube centrifuge before putting into the thermal cycler. Make sure the heated lid function is on
7. Once PCR is done, store PCR reactions at 4C or use PCR cleanup kit (cycle pure kit from epoch) if you will be using your PCR product for further cloning applications

Hazards

- Working with DNA

  As with any DNA work, never throw DNA down the drain, and always handle with gloves
- Hot lid

  This is pretty low risk, but keep in mind that the thermal cycler does use a heated lid, so do not ever touch the lid of the thermal cycler when you open it.
- Damage to common enzymes

  Do not let the mastermix get above 4 °C, and do not thaw it by heating it, as it will become less effective over time if you do

A1.2: Analytical Digest for DNA Screening

Goal

To use a restriction digest with known enzymes to determine some feature of a DNA sequence, such as whether a ligation was successful, or how long a specific insert sequence is.

Materials Needed

- Restriction enzymes of interest (~1 µL of each)
  - Make sure they are not expired
- Proper buffer for restriction digest (see NEB website for compatible buffers with your enzyme of interest. Usually CutSmart buffer will work)
- DNA of interest (2 µL of plasmid DNA at ~80 ng/µL should be fine)
- Agarose
- 1xTAE buffer (non-recycled)
- SYBR safe dye
- PCR tubes in strip form

Procedure

Setting up the digest

1. Make a mastermix for the digest (if running more than 1 digest):
   - 7 x (# of digests + 1) µL of water

- (# of digests + 1) μL of CutSmart Buffer
- 1 μL of each restriction enzyme of interest

  Enzymes should NOT BE THAWED. They should be kept in cold blocks at ALL TIMES unless you are actively capping or uncapping them.

  Please ensure that the molecular biology freezer is completely closed when you are done using it. These reagents are very sensitive to heat/cool cycles and will prematurely stop working if you are not careful.

2. Mix the mastermix above by flicking the PCR tube gently and then centrifuge in a small benchtop centrifuge to collect all liquid at the bottom of the tube
3. Label as many PCR tubes as you have analytical digests to run accordingly. Add 8 μL of mastermix to each PCR tube
4. Add 2 μL of DNA of interest to each PCR tube
5. Mix by flicking the tubes gently and then centrifuge to collect all liquid at the bottom of the tube. Put in one of the thermal cyclers (with heated lid) for 1 hr at 37 °C to digest the DNA. Do not run for more than 2 hrs, or evaporative losses will affect your reaction.

Making an agarose gel for DNA analysis

1. Typically in our lab we will run a standard 1 wt% agarose gel. This is good for most DNA applications and can resolve bands at 500 bp up to ~3 kb. If you are trying to resolve things that are 5 kb vs 2 kb (for example), I would recommend using a 0.7 wt% gel. You can scale the recipe below accordingly
2. See the table below for the components needed to make 1 wt% of different sized gels

Gel Dimension (width cm x length cm)	Sample Capacity (μL)	Agarose (g)	1xTAE Buffer (mL)	SYBR Safe Dye (μL)
6 x 7	5-30	0.35	35	3.5
	30-55	0.5	50	5
6 x 10	5-30	0.5	50	5
	30-55	0.7	70	7
15 x 7	5-30	0.77	77	7.7
	30-55	1.2	120	12
15 x 10	5-30	1.1	110	11
	30-55	1.7	170	17

3. Ensure you have a clean 250 mL Erlenmeyer flask (should be on the shelf with agarose).
4. Weigh agarose into Erlenmeyer flask
5. Add appropriate amount of 1xTAE buffer (not recycled)
6. Microwave in 30 second increments to dissolve agarose. Mix by swirling between microwaving. Continue until there is not visible turbidity in the sample
7. While Erlenmeyer flask is microwaving, set up DNA gel caster and adjust feet to level the caster appropriately
8. Once agarose is completely dissolved (free of small bubbles and small solids!), remove from microwave. Let cool for at least 30 seconds.
9. Add SYBR safe dye to Erlenmeyer flask using pipette.
10. Add the appropriate comb to the gel caster.
11. Slowly pour the contents of the Erlenmeyer flask into the gel caster. If any bubbles linger in the gel after pouring, poke them out with a gel loading pipette tip
12. The gel will be cast once it cools back down to room temperature. This will take ~30 minutes

Running an agarose gel for DNA analysis

1. After your digest has run for an appropriate amount of time, remove it from the 37 °C environment
2. Add appropriate amount of 6x loading dye to each of your samples and mix them by gently flicking the tubes
    a. You can use the purple or orange loading dyes, but the orange loading dye will run at ~100 bp, which means it will not interfere with gel imaging for most samples. This is why I recommend this loading dye
3. Use a small centrifuge to collect your sample/dye mixture at the bottom of the tube
4. Remove the gel from the caster and put it into the appropriate agarose gel runner. Make SURE that the gel is in the correct direction. There is an arrow on the side of both of the gel runners that will tell you which direction your DNA will move once you start the gel.
5. If you are only doing an analytical gel, add recycled 1xTAE buffer just until your gel is covered and the buffer can travel freely between the wells in the runner
    a. If you will need to do a gel extraction, use non-recycled 1xTAE buffer to run your gel

6. Carefully load the appropriate ladder (I usually use 5 μL of a 2log ladder that covers a broad range of bps, but the 1kb and 100bp ladders from NEB work fine as well) using a 10 μL pipette tip. When loading these wells, ensure that you do not pierce the bottom of the well (you will poke a hole for your sample to leak out of), and, as with any gel, move the pipette tip out of the well slowly after loading so you do not drag your sample out of the well.
7. Load your samples. Make sure you take note of which lanes are which.
8. Once everything is loaded, put the lid on the runner, lining up the black and red parts on the lid with the black and red parts in the runner.
9. Plug in the lid to the power source. Make SURE you do this in the correct direction (red to red, black to black). If you do not, it won't throw an error, it will just move the things on your gel in the wrong direction (as in, off the gel and into your running buffer), which would result in sample loss.
10. Use the following settings to run your gel. If you have not been trained on gel electrophoresis, speak to the steward to make sure you are using the equipment correctly.
    100-110 V
    400 mA
    20-30 min
11. Once your run is done image your gel:
    a. Use the black/purple conversion screen on the gel imager
    b. In ImageLabs, select the Nucleic Acids → SYBR Safe option
    c. Optimize for faint or intense bands correctly. For analytic gels like this one, the DNA concentration is relatively low, so I would recommend optimizing for faint bands
    d. Position your gel and image it
    e. Label your ladder accordingly, and draw conclusions

<u>Hazards</u>

- Heat

  When casting the agarose gel, you will be dealing with a microwaved liquid. Always use a hot hand to handle the flask (not your glove alone!) so that you do not risk dropping it and shattering it. Also be aware that there is steam coming out of the flask.

Do your best not to over-microwave the flask. Once you see it at a rolling boil, stop the microwave.

- UV exposure

  The gel imager has the potential to expose you to UV light. Always cover your arms completely if you are using the transUV option on the gel imager, and keep your face at least 1.5 ft from the light source.

- DNA work

  DNA is a biological substance. Dispose of all gloves that you use during this procedure in biological waste.

A1.3 DNA Transformation into Lab-Prepped Competent Cells

Goal

To take DNA ordered from Genscript and transform it into cells that have been prepared in-lab using the Zymogen *Mix and Go*® kit.

Materials Needed

- **DNA of Interest**
  - Shipped from Genscript, in solid form
- **Competent Cells**
  - Stored in bottom shelf of -80 °C freezer
  - If using SG13009C, make sure it contains the pREP4 plasmid as well
    - If using SG13009, also transform into NEB5α for purposes of making more DNA
    - SG13009C contains inherent kanamycin resistance from pREP4, so you should use kan + vector antibiotic for this cell line
- **LB-Agar Plates with Appropriate Antibiotic**
- **SOB or SOC Media** (if antibiotic being used is not ampicillin)
  - Should be refrigerated when not in use
- **Ice**
- **Glass beads** (sterilized)

Procedure

1. Dissolve your DNA according to the Genscript protocol:
    a. Centrifuge at ≥ 6000 xg for 1 minute at 4 °C
    b. Open the vial and add 20 µL of MilliQ water to dissolve the DNA
    c. Close the lid and vortex the vial for 1 minute
        i. Vortex both the lid and the bottom of the vial to ensure that any DNA that may have gotten stuck to the lid is also dissolved
2. Thaw 1 eppendorf tube of the competent cells of interest on ice. Label.
3. Gently flick the tube with cells to ensure they are suspended
4. Set up a sterile environment
    a. Spray down with ethanol
    b. Turn on flame
    c. Use sterile pipette tips
5. Add 1 µL of dissolved DNA to cell vial, pipette up and down in cell slurry to ensure that DNA has been removed from pipette tip, close competent cell vial.
6. Gently flick cell tube to make sure that cells and DNA are adequately mixed and place back on ice.
7. Wait 30 mins.
8. If antibiotic is not JUST ampicillin:
    a. Add 500 µL of SOB/SOC to competent cell vial (again, in sterile environment)
    b. Shake the cell vial in the 37 °C incubator for 30 min-1 hr
    c. While cells are in incubator, you should also prewarm the agar plates you will be using by putting them on the top shelf of the nonreservable incubator.
        i. Make sure the water baths in the top shelf are full
        ii. You will need two plates for each transformation you are doing
        iii. Unless you are plating cells or selecting colonies, plates should always be stored with the agar surface facing down.
9. Plate your cells:
    a. Set up sterile environment
    b. Remove plates and cells from incubator
    c. Label plates appropriately—Include
        i. Gene that you are adding, including vector
        ii. Cell line
        iii. Your initials
        iv. Date
        v. Antibiotic
    d. Flip each plate over and gently shake 4-6 glass beads onto each plate.
        i. Unless you are actively adding beads/media to plate, keep the lid on it; pass the lid through a flame before recapping plate to resterilize
    e. Pipette cells onto plate using sterile pipette tips
        i. 30 µL onto first plate
        ii. 70 µL onto second plate
    f. Shake plates for 1 minute to spread cells
        i. Make sure that beads don't get caught around edge of plate when shaking

g. Remove beads from plate
    i. Tilt plate sideways over dirty bead bin
    ii. Crack plate lid (opening fully risks exposure of plate to contamination)
    iii. Tap gently to get beads to fall out into bin
h. Flip plate to agar side down

10. Parafilm ~half-2/3 of your plates to prevent from drying out
    a. Requires about half of one square of parafilm
11. Incubate plates at 37 °C overnight (12 h) or until visible colonies start to appear
    a. Incubating for too long, especially with ampicillin, will cause the growth of satellite colonies, so it is important to remove plates from the incubator within 16 hours
12. Once you have colonies, you can use these to make starter cultures for expression, but on your first expression, you should make 1 extra 10 mL culture and 1 extra 5 mL culture:
    a. Use 2 mL to make 2 glycerol stocks (mix 1 mL sterilized 80% glycerol in water with 1 mL overnight culture)
    b. Miniprep the 10 mL culture so that you have DNA stocks for future use
        i. Use these stocks for all future cloning work and transformations to preserve the purity of your stock gene.
    c. If you transformed into SG13009 and NEB5α, use the SG13009 for all expressions and glycerol stocks and the NEB5α cells to miniprep and make your DNA

Hazards

- Open flame

  Part of setting up a sterile environment requires working with an open flame. Be aware of any hair/loose clothing when working near the flame to prevent any accidents. Also ensure that any ethanol used to sterilize the work area is *completely* evaporated before starting the Bunsen burner

- Biologics

  This work is BL1. Dispose of any gloves used in this process in the biological waste stream. Bleach any media for 20 minutes prior to disposing down the drain.

A1.4 ELP-mCherry Purification by Ammonium Sulfate Precipitation

Goal

To purify ELP-mCherry constructs using ammonium sulfate precipitation

Required Materials

- Cells from protein expression

  These should be resuspended in ~100 mL MENT buffer (see below) per 30 g of wet cell mass. All solutions containing fluorescent proteins should be kept away from light during processing. This can be achieved by wrapping bottles in foil, as noted in the procedure.
- MENT buffer
  - 5 mM $MgCl_2$
  - 1 mM ethylenediaminetetraacetic acid (EDTA)
  - 100 mM NaCl
  - 20 mM trizma
- 3 M ammonium sulfate in 20 mM tris, pH 8
  - To make this concentration of ammonium sulfate, weigh out the ammonium sulfate first in a bottle marked to the desired volume (usually 1 L), and add 20 mM tris buffer after to the marked volume
  - Readjust the pH of the solution after dissolving ammonium sulfate
- Lysozyme, DNase I, RNase A
  - Lysozyme and RNase A can be found in the door of the -20 °C freezer.
  - DNase I is stored in the top shelf of the 4 °C delicase fridge with the dialysis stirplate in it
- Ice (day 1 only)

Required Equipment

- High-speed centrifuge
- Tip sonicator

Procedure

Day 1: Lyse Cells, Clarify Lysate, and Perform First Precipitation

1. Remove your frozen resuspended cells from the -80 °C freezer. Thaw them in a bucket of water that you refill periodically with new water, and remove any ice layers that form around the previously frozen conicals.
   - The water will help thaw the resuspended cells more quickly due to better heat transfer.

2. Once the cells are thawed, weigh out 50 mg of lysozyme per L of original culture into a 250 mL plastic beaker. Add a small amount of MENT and a stir bar to this beaker, and stir it on a stir plate either in or out of the fridge until the lysozyme is dissolved. Once it is dissolved, add the thawed resuspended cells in while the beaker is stirring.
    - I have noticed that sometimes the lysozyme clumps up and doesn't disperse if you add it directly to the thawed cells, which is why I prefer to dissolve it first.
    - If the protein you are trying to purify is highly negatively charged, lysozyme can complex with it and becomes very difficult to remove (because it is highly positively charged). In this instance, consider using several freeze-thaw cycles to help with cell lysis and omit the addition of lysozyme.
3. Wrap the beaker in foil and cover with foil (label!), and move to a stir spot in the 4 °C fridge. Let this stir for 1-2 hrs.
    - Depending on the stability of your protein, you can also stir at room temperature, or at slightly elevated temperature. This accelerates the enzyme kinetics
4. Remove the beaker from the fridge and remove the stir bar using another larger stir bar. Keep it for later by storing it in the foil you previously used to cover the beaker. Remove the foil from the outside of the beaker (but try to save it for later, minimize waste).
5. Sonicate cell lysate 3 times at 50% duty cycle, power level 5, for 7 minutes at a time. Keep the beaker submerged in ice water while sonicating and remove from sonicator and allow to cool down to at least 15 °C between sonications (measure with IR gun). Cover the beaker while it sits on the bench to prevent contamination. Occassional stirring with a glass rod can help to speed up this cooling. Clean the sonicator tip with 70% ethanol between uses.
    - Sonication helps to break up the remaining cell membranes and the genomic DNA of the cell, which constitutes a highly entangled polymer (which in turn increases the viscosity of the lysate considerably). Less sonication may be necessary depending on how new the sonicator tip is, and how much sample you have. Typically it is "done" when the solution becomes visibly less viscous.
    - It is important to add water to the ice bath you use to cool your beaker. Addition of water will improve the thermal contact between your lysate solution and the bath, and thus will better prevent your lysate from heating up too much. Keep in

mind that ice will melt as you sonicate, though, so you may have to replenish ice and/or pour out water as you go so your beaker does not float and tip over.

6. Once the sonications are done, re-foil the beaker and put the stir bar back in. Put it back on a stir location in the 4 °C fridge for 30 minutes to an hour
    - I've found that allowing the lysate to cool completely and stir helps minimize product loss in the clarification step.
7. Clarify the lysate by centrifuging (preferably in 50 mL conical tubes) at 12500 rpm for 90 minutes at 4 °C. If you end up using 500 mL centrifuge bottles, spin for 2 hours at 9500 rpm.
    - Make sure you follow the centrifuge SOP! You can balance your tubes using MENT buffer
    - In the case of mCherry, if you have lost a lot of product, or not completely lysed the cells, you will see a very dark pink layer in the pellet of the clarified lysate.
    - The centrifugation will be more effective if you keep the volume in each tube to 40 mL or less.
8. Decant the clarified lysate into a 500 mL centrifuge bottle (labeled!), taking care to minimize the amount of pellet dragged with the lysate (pour slowly).
9. Add ~ 1 mg per L of expression of DNase and RNase to the clarified lysate and wrap the bottle in foil. Put it in the 37 °C non-reservable incubator for 1 hour to allow the enzymes to break down the DNA and RNA in your lysate.
10. Add a stir bar to the centrifuge bottle and move it to a stir plate in the fridge.
11. Using the markings on the side of the centrifuge bottle, approximate the volume in the bottle. Add an equivalent volume of 3 M ammonium sulfate in 20 mM tris, pH 8 to bring the total ammonium sulfate concentration to ~1.5 M.
12. Allow the lysate-ammonium sulfate mixture to stir for 30 minutes - overnight.
13. **Remove the stir bar**, and balance the centrifuge bottle with one filled with water.
14. Centrifuge the bottle to collect the precipitate at 9500 rpm for 45mins-1 hr at 4 °C or 20 °C (you can allow the centrifuge to warm up after your clarification step, and run it at whatever temperature it reaches after that.)
15. Decant off the supernatant into a separate, labelled bottle.

- If this is your second+ time running this purification, and you know you don't have any product loss in this supernatant, you can discard of it down the drain. DO NOT MIX IT WITH BLEACH. (ammonium + bleach = XXXX)

16. Add ~70 mL of MENT and a stir bar to the centrifuge bottle with the pellet. Use a spatula to scrape the pellet from the walls of the bottle.
17. Let the bottle stir (covered in foil!) overnight at 4 °C to allow resuspension of the pellet.

Day 2: Cold Spin & Second Ammonium Sulfate Precipitation

1. Move your resuspended pellets into 50 mL conicals (balance with MENT), and centrifuge at 4 °C, 12500 rpm, 90 min.
2. Collect the supernatant of this spin in a new 500 mL centrifuge bottle.
3. Add a stir bar to the centrifuge bottle and again adjust the ammonium sulfate concentration to 1.5 M by adding 3 M ammonium sulfate.
4. Allow this to stir at 4 °C for 30 minutes (covered in foil)
5. REMOVE THE STIR BAR, balance with a bottle filled with water, and centrifuge at any temperature (≤ 37 °C) for 45-60 mins, 9500 rpm.
6. Repeat steps 15-17 from *Day 1*.

Day 3: Final Cold Spin & Dialysis

1. Repeat step 1 from Day 2
2. Collect the supernatant from this spin in any bottle.
3. Set up dialysis into water (10 x 3 h)

Hazards

- Centrifuge use

  Take care to secure all rotors used in the centrifuge; damage to the centrifuge and personal harm can occur if rotors are not properly seated and secured to spindle.
- Protein solutions with ammonium sulfate

  Do not bleach solutions with ammonium sulfate in them. This could potentially generate chloramine gas.

A2. DNA and Amino Acid Sequences for All Proteins Used in Thesis

A2.1 Elastin-Like Polypeptide (ELP) Charge Influences Self-Assembly of ELP-mCherry Fusion Proteins

Uncharged ELP, ordered gene (uELP10K)

Vector: pET28b; Cloning site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATT
CCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGGATCCC
GGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGGATTCCGG
GTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGTATCCCGGGT
GTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGT
GGGCGTTCCGGGTGTGGGTGTTCCGGGTGTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

Legend: NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (143 AA, 12257.311 Da, Estimated pI = 7.14)

MVHHHHHHGSGLVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPG
VGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGETAS*

Uncharged ELP, doubled (uELP20K)

Vector: pET28b; Cloning site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATT
CCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGGATCCC
GGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGGATTCCGG
GTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGTATCCCGGGT
GTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGT
GGGCGTTCCGGGTGTGGGTGTTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGTGG
GCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGGATCCCGGGTGTGGGC
GTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGGATTCCGGGTGTGGGCGT
TCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGTATCCCGGGTGTGGGCGTTC
CGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGTGGGCGTTCCG
GGTGTGGGTGTTCCGGGTGTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (268 AA, 22544.5742 Da, Estimated pI = 7.14)

MVHHHHHHGSGLVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPG
VGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVG

VPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVP
GVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVETAS*

Uncharged ELP, quadrupled (uELP40K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATT
CCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGGATCCC
GGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGGATTCCGG
GTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGTATCCCGGGT
GTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGT
GGGCGTTCCGGGTGTGGGTGTTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGTGG
GCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGGATCCCGGGTGTGGGC
GTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGGATTCCGGGTGTGGGCGT
TCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGTATCCCGGGTGTGGGCGTTC
CGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGTGGGCGTTCCG
GGTGTGGGTGTTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGTGGGCGTTCCGGG
TGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGGATCCCGGGTGTGGGCGTTCCGGGTG
TGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGGATTCCGGGTGTGGGCGTTCCGGGTGTG
GGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGTATCCCGGGTGTGGGCGTTCCGGGTGTGGG
CGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGTGGGCGTTCCGGGTGTGGGTG
TTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTT
CCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGGATCCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCC
GGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGGATTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGG
GTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGTATCCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGT
GTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTTGGTATTCCGGGTGTGGGCGTTCCGGGTGTGGGTGTTCCGGGTGT
TGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (518 AA, 43159.0625 Da, Estimated pI = 7.14)

MVHHHHHHGSGLVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPG
VGVPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIPGVG
VPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVP
GVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGV
GVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGV
PGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPG
VGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVETAS*

Charge balanced ELP, ordered gene (bELP10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATT
CCGGGTGTGGGCGTTCCAGGCAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGTGAAGGCATCCC
GGGTGTGGGCGTTCCCGGTAAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGG
GTGTGGGCGTTCCAGGCAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGGGAGGGTATCCCGGGT
GTGGGCGTTCCAGGTAAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGCATTCCGGGTGT
GGGCGTTCCAGGCAAGGGTGTTCCGGGTGTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (143 AA, 12532.4521 Da, Estimated pI = 7.14)

MVHHHHHHGSGLVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPG
VGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGETAS*

Charge balanced ELP, doubled (bELP20K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATT
CCGGGTGTGGGCGTTCCAGGCAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGTGAAGGCATCCC
GGGTGTGGGCGTTCCCGGTAAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGG
GTGTGGGCGTTCCAGGCAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGGGAGGGTATCCCGGGT
GTGGGCGTTCCAGGTAAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGCATTCCGGGTGT
GGGCGTTCCAGGCAAGGGTGTTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATTCCGGGTGTGG
GCGTTCCAGGCAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGTGAAGGCATCCCGGGTGTGGGC
GTTCCCGGTAAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGGGTGTGGGCGT
TCCAGGCAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGGGAGGGTATCCCGGGTGTGGGCGTTC
CAGGTAAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGCATTCCGGGTGTGGGCGTTCCA
GGCAAGGGTGTTCCGGGTGTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (268 AA, 23134.8184 Da, Estimated pI = 7.14)

MVHHHHHHGSGLVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPG
VGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVG
VPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVP
GKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGETAS*

Charge balanced ELP, quadrupled (bELP40K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATT
CCGGGTGTGGGCGTTCCAGGCAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGTGAAGGCATCCC
GGGTGTGGGCGTTCCCGGTAAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGG
GTGTGGGCGTTCCAGGCAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGGGAGGGTATCCCGGGT
GTGGGCGTTCCAGGTAAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGCATTCCGGGTGT
GGGCGTTCCAGGCAAGGGTGTTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATTCCGGGTGTGG
GCGTTCCAGGCAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGTGAAGGCATCCCGGGTGTGGGC
GTTCCCGGTAAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGGGTGTGGGCGT
TCCAGGCAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGGGAGGGTATCCCGGGTGTGGGCGTTC
CAGGTAAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGCATTCCGGGTGTGGGCGTTCCA
GGCAAGGGTGTTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATTCCGGGTGTGGGCGTTCCAGG
CAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGTGAAGGCATCCCGGGTGTGGGCGTTCCCGGTA
AAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGGGTGTGGGCGTTCCAGGCAAG
GGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGGGAGGGTATCCCGGGTGTGGGCGTTCCAGGTAAAGG
CGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGCATTCCGGGTGTGGGCGTTCCAGGCAAGGGTG
TTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATTCCGGGTGTGGGCGTTCCAGGCAAGGGCGTT
CCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGTGAAGGCATCCCGGGTGTGGGCGTTCCCGGTAAAGGCGTGCC
GGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGGGTGTGGGCGTTCCAGGCAAGGGCGTTCCGG
GTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGGGAGGGTATCCCGGGTGTGGGCGTTCCAGGTAAAGGCGTGCCGGGT
GTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGCATTCCGGGTGTGGGCGTTCCAGGCAAGGGTGTTCCGGGTGT
TGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (518 AA, 44339.5508 Da, Estimated pI = 7.14)

MVHHHHHHGSGLVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPG
VGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVG
VPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVP
GKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGK
GVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGV
PGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPG
VGVPGVGVPGEGIPGVGVPGKGVPGVGVPGVGVPGEGIPGVGVPGKGVPGVGETAS*

Charge balanced, increased hydrophobicity ELP, ordered gene (bhELP10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATC
CCGGGTGTTGGCATTCCGGGCAAGGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGTGAAGGCATTCC

GGGTGTGGGCATCCCGGGTAAAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGTATCCCGG
GTGTTGGAATCCCTGGCAAGGGCATCCCGGGTGTGGGTATTCCGGGTGTTGGGATTCCGGGCGAGGGTATCCCGGGT
GTGGGAATCCCTGGTAAAGGTATTCCGGGTGTGGGTATACCGGGTGTTGGTATTCCGGGCGAAGGTATACCGGGTGT
TGGCATACCTGGCAAGGGTATCCCGGGTGTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (143 AA, 12812.9883 Da, Estimated pI = 7.14)

MVHHHHHHGSGLVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPG
VGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGETAS*

Charge balanced, increased hydrophobicity ELP, doubled (bhELP20K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATC
CCGGGTGTTGGCATTCCGGGCAAGGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGTGAAGGCATTCC
GGGTGTGGGCATCCCGGGTAAAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGTATCCCGG
GTGTTGGAATCCCTGGCAAGGGCATCCCGGGTGTGGGTATTCCGGGTGTTGGGATTCCGGGCGAGGGTATCCCGGGT
GTGGGAATCCCTGGTAAAGGTATTCCGGGTGTGGGTATACCGGGTGTTGGTATTCCGGGCGAAGGTATACCGGGTGT
TGGCATACCTGGCAAGGGTATCCCGGGTGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATCCCGGGTGTTG
GCATTCCGGGCAAGGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGTGAAGGCATTCCGGGTGTGGGC
ATCCCGGGTAAAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGTATCCCGGGTGTTGGAAT
CCCTGGCAAGGGCATCCCGGGTGTGGGTATTCCGGGTGTTGGGATTCCGGGCGAGGGTATCCCGGGTGTGGGAATCC
CTGGTAAAGGTATTCCGGGTGTGGGTATACCGGGTGTTGGTATTCCGGGCGAAGGTATACCGGGTGTTGGCATACCT
GGCAAGGGTATCCCGGGTGTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (268 AA, 23695.8906 Da, Estimated pI = 7.14)

MVHHHHHHGSGLVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPG
VGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVG
IPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIP
GKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGETAS*

Charge balanced, increased hydrophobicity ELP, quadrupled (bhELP40K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATC
CCGGGTGTTGGCATTCCGGGCAAGGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGTGAAGGCATTCC
GGGTGTGGGCATCCCGGGTAAAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGTATCCCGG
GTGTTGGAATCCCTGGCAAGGGCATCCCGGGTGTGGGTATTCCGGGTGTTGGGATTCCGGGCGAGGGTATCCCGGGT

GTGGGAATCCCTGGTAAAGGTATTCCGGGTGTGGGTATACCGGGTGTTGGTATTCCGGGCGAAGGTATACCGGGTGT
TGGCATACCTGGCAAGGGTATCCCGGGTGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATCCCGGGTGTTG
GCATTCCGGGCAAGGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGTGAAGGCATTCCGGGTGTGGGC
ATCCCGGGTAAAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGTATCCCGGGTGTTGGAAT
CCCTGGCAAGGGCATCCCGGGTGTGGGTATTCCGGGTGTTGGGATTCCGGGCGAGGGTATCCCGGGTGTGGGAATCC
CTGGTAAAGGTATTCCGGGTGTGGGTATACCGGGTGTTGGTATTCCGGGCGAAGGTATACCGGGTGTTGGCATACCT
GGCAAGGGTATCCCGGGTGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATCCCGGGTGTTGGCATTCCGGG
CAAGGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGTGAAGGCATTCCGGGTGTGGGCATCCCGGGTA
AAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGTATCCCGGGTGTTGGAATCCCTGGCAAG
GGCATCCCGGGTGTGGGTATTCCGGGTGTTGGGATTCCGGGCGAGGGTATCCCGGGTGTGGGAATCCCTGGTAAAGG
TATTCCGGGTGTGGGTATACCGGGTGTTGGTATTCCGGGCGAAGGTATACCGGGTGTTGGCATACCTGGCAAGGGTA
TCCCGGGTGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATCCCGGGTGTTGGCATTCCGGGCAAGGGCATT
CCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGTGAAGGCATTCCGGGTGTGGGCATCCCGGGTAAAGGTATCCC
GGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGTATCCCGGGTGTTGGAATCCCTGGCAAGGGCATCCCGG
GTGTGGGTATTCCGGGTGTTGGGATTCCGGGCGAGGGTATCCCGGGTGTGGGAATCCCTGGTAAAGGTATTCCGGGT
GTGGGTATACCGGGTGTTGGTATTCCGGGCGAAGGTATACCGGGTGTTGGCATACCTGGCAAGGGTATCCCGGGTGT
TGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (518 AA, 45461.6914 Da, Estimated pI = 7.14)

MVHHHHHHGSGLVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPG
VGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVG
IPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIP
GKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGK
GIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGI
PGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPG
VGIPGVGIPGEGIPGVGIPGKGIPGVGIPGVGIPGEGIPGVGIPGKGIPGVGETAS*

Negatively charged ELP, ordered gene (nELP10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATT
CCGGGTGTGGGCGTTCCAGGTGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGCGAGGGCATCCC
GGGTGTGGGCGTTCCCGGTGAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGG
GTGTGGGCGTTCCAGGGGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGAATCCCGGGT
GTGGGCGTTCCAGGAGAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGGATACCGGGTGT
GGGCGTTCCAGGGGAAGGTGTTCCGGGTGTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (143 AA, 12537.1602 Da, Estimated pI = 4.48)

MVHHHHHHGSGLVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPG
VGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGETAS*

Negatively charged ELP, doubled (nELP20K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATT
CCGGGTGTGGGCGTTCCAGGTGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGCGAGGGCATCCC
GGGTGTGGGCGTTCCCGGTGAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGG
GTGTGGGCGTTCCAGGGGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGAATCCCGGGT
GTGGGCGTTCCAGGAGAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGGATACCGGGTGT
GGGCGTTCCAGGGGAAGGTGTTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATTCCGGGTGTGG
GCGTTCCAGGTGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGCGAGGGCATCCCGGGTGTGGGC
GTTCCCGGTGAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGGGTGTGGGCGT
TCCAGGGGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGAATCCCGGGTGTGGGCGTTC
CAGGAGAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGGATACCGGGTGTGGGCGTTCCA
GGGGAAGGTGTTCCGGGTGTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (268 AA, 23144.2344 Da, Estimated pI = 4.01)

MVHHHHHHGSGLVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPG
VGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVG
VPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVP
GEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGETAS*

Negatively charged ELP, quadrupled (nELP40K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATT
CCGGGTGTGGGCGTTCCAGGTGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGCGAGGGCATCCC
GGGTGTGGGCGTTCCCGGTGAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGG
GTGTGGGCGTTCCAGGGGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGAATCCCGGGT
GTGGGCGTTCCAGGAGAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGGATACCGGGTGT
GGGCGTTCCAGGGGAAGGTGTTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATTCCGGGTGTGG
GCGTTCCAGGTGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGCGAGGGCATCCCGGGTGTGGGC
GTTCCCGGTGAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGGGTGTGGGCGT
TCCAGGGGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGAATCCCGGGTGTGGGCGTTC

CAGGAGAAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGGATACCGGGTGTGGGCGTTCCA
GGGGAAGGTGTTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATTCCGGGTGTGGGCGTTCCAGG
TGAAGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGCGAGGGCATCCCGGGTGTGGGCGTTCCCGGTG
AAGGCGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGGGTGTGGGCGTTCCAGGGGAA
GGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGAATCCCGGGTGTGGGCGTTCCAGGAGAAGG
CGTGCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGGATACCGGGTGTGGGCGTTCCAGGGGAAGGTG
TTCCGGGTGTTGGTGTTCCGGGTGTGGGCGTTCCGGGCGAGGGTATTCCGGGTGTGGGCGTTCCAGGTGAAGGCGTT
CCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCTGGCGAGGGCATCCCGGGTGTGGGCGTTCCCGGTGAAGGCGTGCC
GGGTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAAGGCATTCCGGGTGTGGGCGTTCCAGGGGAAGGCGTTCCGG
GTGTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGAATCCCGGGTGTGGGCGTTCCAGGAGAAGGCGTGCCGGGT
GTGGGCGTTCCGGGTGTGGGCGTTCCAGGCGAGGGGATACCGGGTGTGGGCGTTCCAGGGGAAGGTGTTCCGGGTGT
TGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (518 AA, 44358.3828 Da, Estimated pI = 3.65)

MVHHHHHHGSGLVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPG
VGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVG
VPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVP
GEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGE
GVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGV
PGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPG
VGVPGVGVPGEGIPGVGVPGEGVPGVGVPGVGVPGEGIPGVGVPGEGVPGVGETAS*

Negatively charged, increased hydrophobicity ELP, ordered gene (nhELP10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATC
CCGGGTGTTGGCATTCCGGGTGAAGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGCGAGGGCATTCC
GGGTGTGGGCATCCCGGGCGAAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGCATTCCGG
GTGTTGGTATTCCGGGCGAAGGCATCCCGGGTGTGGGAATCCCGGGTGTTGGAATTCCGGGCGAGGGTATCCCGGGT
GTGGGAATACCGGGTGAAGGTATTCCGGGTGTGGGTATCCCGGGTGTTGGGATCCCTGGCGAAGGTATTCCGGGTGT
TGGTATCCCTGGTGAAGGTATCCCGGGTGTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (143 AA, 12817.6963 Da, Estimated pI = 4.48)

MVHHHHHHGSGLVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPG
VGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGETAS*

Negatively charged, increased hydrophobicity ELP, doubled (nhELP20K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATC
CCGGGTGTTGGCATTCCGGGTGAAGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGCGAGGGCATTCC
GGGTGTGGGCATCCCGGGCGAAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGCATTCCGG
GTGTTGGTATTCCGGGCGAAGGCATCCCGGGTGTGGGAATCCCGGGTGTTGGAATTCCGGGCGAGGGTATCCCGGGT
GTGGGAATACCGGGTGAAGGTATTCCGGGTGTGGGTATCCCGGGTGTTGGGATCCCTGGCGAAGGTATTCCGGGTGT
TGGTATCCCTGGTGAAGGTATCCCGGGTGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATCCCGGGTGTTG
GCATTCCGGGTGAAGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGCGAGGGCATTCCGGGTGTGGGC
ATCCCGGGCGAAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGCATTCCGGGTGTTGGTAT
TCCGGGCGAAGGCATCCCGGGTGTGGGAATCCCGGGTGTTGGAATTCCGGGCGAGGGTATCCCGGGTGTGGGAATAC
CGGGTGAAGGTATTCCGGGTGTGGGTATCCCGGGTGTTGGGATCCCTGGCGAAGGTATTCCGGGTGTTGGTATCCCT
GGTGAAGGTATCCCGGGTGTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (268 AA, 23705.3066 Da, Estimated pI = 4.01)

MVHHHHHHGSGLVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPG
VGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVG
IPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIP
GEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGETAS*

Negatively charged, increased hydrophobicity ELP, quadrupled (nhELP40K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATC
CCGGGTGTTGGCATTCCGGGTGAAGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGCGAGGGCATTCC
GGGTGTGGGCATCCCGGGCGAAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGCATTCCGG
GTGTTGGTATTCCGGGCGAAGGCATCCCGGGTGTGGGAATCCCGGGTGTTGGAATTCCGGGCGAGGGTATCCCGGGT
GTGGGAATACCGGGTGAAGGTATTCCGGGTGTGGGTATCCCGGGTGTTGGGATCCCTGGCGAAGGTATTCCGGGTGT
TGGTATCCCTGGTGAAGGTATCCCGGGTGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATCCCGGGTGTTG
GCATTCCGGGTGAAGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGCGAGGGCATTCCGGGTGTGGGC
ATCCCGGGCGAAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGCATTCCGGGTGTTGGTAT
TCCGGGCGAAGGCATCCCGGGTGTGGGAATCCCGGGTGTTGGAATTCCGGGCGAGGGTATCCCGGGTGTGGGAATAC
CGGGTGAAGGTATTCCGGGTGTGGGTATCCCGGGTGTTGGGATCCCTGGCGAAGGTATTCCGGGTGTTGGTATCCCT
GGTGAAGGTATCCCGGGTGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATCCCGGGTGTTGGCATTCCGGG
TGAAGGCATTCCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGCGAGGGCATTCCGGGTGTGGGCATCCCGGGCG
AAGGTATCCCGGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGCATTCCGGGTGTTGGTATTCCGGGCGAA

GGCATCCCGGGTGTGGGAATCCCGGGTGTTGGAATTCCGGGCGAGGGTATCCCGGGTGTGGGAATACCGGGTGAAGG
TATTCCGGGTGTGGGTATCCCGGGTGTTGGGATCCCTGGCGAAGGTATTCCGGGTGTTGGTATCCCTGGTGAAGGTA
TCCCGGGTGTTGGTATCCCGGGTGTGGGCATTCCGGGCGAGGGCATCCCGGGTGTTGGCATTCCGGGTGAAGGCATT
CCGGGTGTGGGTATTCCGGGTGTTGGCATCCCGGGCGAGGGCATTCCGGGTGTGGGCATCCCGGGCGAAGGTATCCC
GGGTGTGGGTATCCCGGGTGTTGGTATCCCGGGCGAAGGCATTCCGGGTGTTGGTATTCCGGGCGAAGGCATCCCGG
GTGTGGGAATCCCGGGTGTTGGAATTCCGGGCGAGGGTATCCCGGGTGTGGGAATACCGGGTGAAGGTATTCCGGGT
GTGGGTATCCCGGGTGTTGGGATCCCTGGCGAAGGTATTCCGGGTGTTGGTATCCCTGGTGAAGGTATCCCGGGTGT
TGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (518 AA, 45480.5234 Da, Estimated pI = 3.65)

MVHHHHHHGSGLVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPG
VGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVG
IPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIP
GEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGE
GIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGI
PGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPG
VGIPGVGIPGEGIPGVGIPGEGIPGVGIPGVGIPGEGIPGVGIPGEGIPGVGETAS*

Proteins not used in study, but cloned:

Uncharged, most hydrophobic ELP, ordered gene (uhhELP10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCTTTGGTATCCCGGGTTTCGGCATTCCGGGTGTGGGCATC
CCGGGTTTTGGCATTCCGGGTGTTGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGTGGGCATTCC
GGGTTTCGGCATCCCGGGTGTTGGTATCCCGGGCTTTGGTATTCCGGGCTTTGGTATCCCGGGTGTGGGTATTCCGG
GCTTTGGCATCCCGGGTGTTGGCATCCCGGGCTTTGGCATTCCTGGCTTCGGTATTCCGGGTGTGGGTATCCCGGGC
TTCGGTATCCCGGGTGTTGGTATTCCGGGTTTCGGTATTCCTGGCTTCGGCATCCCGGGTGTGGGTATACCTGGCTT
TGGTATTCCGGGTGTTGGCATACCTGGCTTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (143 AA, 13286.5703 Da, Estimated pI = 7.14)

MVHHHHHHGSGLFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPG
FGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGETAS*

Uncharged, most hydrophobic ELP, doubled (uhhELP20K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCTTTGGTATCCCGGGTTTCGGCATTCCGGGTGTGGGCATC

CCGGGTTTTGGCATTCCGGGTGTTGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGTGGGCATTCC
GGGTTTCGGCATCCCGGGTGTTGGTATCCCGGGCTTTGGTATTCCGGGCTTTGGTATCCCGGGTGTGGGTATTCCGG
GCTTTGGCATCCCGGGTGTTGGCATCCCGGGCTTTGGCATTCCTGGCTTCGGTATTCCGGGTGTGGGTATCCCGGGC
TTCGGTATCCCGGGTGTTGGTATTCCGGGTTTCGGTATTCCTGGCTTCGGCATCCCGGGTGTGGGTATACCTGGCTT
TGGTATTCCGGGTGTTGGCATACCTGGCTTTGGTATCCCGGGTTTCGGCATTCCGGGTGTGGGCATCCCGGGTTTTG
GCATTCCGGGTGTTGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGTGGGCATTCCGGGTTTCGGC
ATCCCGGGTGTTGGTATCCCGGGCTTTGGTATTCCGGGCTTTGGTATCCCGGGTGTGGGTATTCCGGGCTTTGGCAT
CCCGGGTGTTGGCATCCCGGGCTTTGGCATTCCTGGCTTCGGTATTCCGGGTGTGGGTATCCCGGGCTTCGGTATCC
CGGGTGTTGGTATTCCGGGTTTCGGTATTCCTGGCTTCGGCATCCCGGGTGTGGGTATACCTGGCTTTGGTATTCCG
GGTGTTGGCATACCTGGCTTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (268 AA, 24595.0117 Da, Estimated pI = 7.14)

MVHHHHHHGSGLFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPG
FGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFG
IPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIP
GVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGETAS*

Uncharged, most hydrophobic ELP, quadrupled (uhhELP40K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCTTTGGTATCCCGGGTTTCGGCATTCCGGGTGTGGGCATC
CCGGGTTTTGGCATTCCGGGTGTTGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGTGGGCATTCC
GGGTTTCGGCATCCCGGGTGTTGGTATCCCGGGCTTTGGTATTCCGGGCTTTGGTATCCCGGGTGTGGGTATTCCGG
GCTTTGGCATCCCGGGTGTTGGCATCCCGGGCTTTGGCATTCCTGGCTTCGGTATTCCGGGTGTGGGTATCCCGGGC
TTCGGTATCCCGGGTGTTGGTATTCCGGGTTTCGGTATTCCTGGCTTCGGCATCCCGGGTGTGGGTATACCTGGCTT
TGGTATTCCGGGTGTTGGCATACCTGGCTTTGGTATCCCGGGTTTCGGCATTCCGGGTGTGGGCATCCCGGGTTTTG
GCATTCCGGGTGTTGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGTGGGCATTCCGGGTTTCGGC
ATCCCGGGTGTTGGTATCCCGGGCTTTGGTATTCCGGGCTTTGGTATCCCGGGTGTGGGTATTCCGGGCTTTGGCAT
CCCGGGTGTTGGCATCCCGGGCTTTGGCATTCCTGGCTTCGGTATTCCGGGTGTGGGTATCCCGGGCTTCGGTATCC
CGGGTGTTGGTATTCCGGGTTTCGGTATTCCTGGCTTCGGCATCCCGGGTGTGGGTATACCTGGCTTTGGTATTCCG
GGTGTTGGCATACCTGGCTTTGGTATCCCGGGTTTCGGCATTCCGGGTGTGGGCATCCCGGGTTTTGGCATTCCGGG
TGTTGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGTGGGCATTCCGGGTTTCGGCATCCCGGGTG
TTGGTATCCCGGGCTTTGGTATTCCGGGCTTTGGTATCCCGGGTGTGGGTATTCCGGGCTTTGGCATCCCGGGTGTT
GGCATCCCGGGCTTTGGCATTCCTGGCTTCGGTATTCCGGGTGTGGGTATCCCGGGCTTCGGTATCCCGGGTGTTGG
TATTCCGGGTTTCGGTATTCCTGGCTTCGGCATCCCGGGTGTGGGTATACCTGGCTTTGGTATTCCGGGTGTTGGCA
TACCTGGCTTTGGTATCCCGGGTTTCGGCATTCCGGGTGTGGGCATCCCGGGTTTTGGCATTCCGGGTGTTGGCATT
CCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGTGGGCATTCCGGGTTTCGGCATCCCGGGTGTTGGTATCCC
GGGCTTTGGTATTCCGGGCTTTGGTATCCCGGGTGTGGGTATTCCGGGCTTTGGCATCCCGGGTGTTGGCATCCCGG

GCTTTGGCATTCCTGGCTTCGGTATTCCGGGTGTGGGTATCCCGGGCTTCGGTATCCCGGGTGTTGGTATTCCGGGT
TTCGGTATTCCTGGCTTCGGCATCCCGGGTGTGGGTATACCTGGCTTTGGTATTCCGGGTGTTGGCATACCTGGCTT
TGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (518 AA, 47211.8906 Da, Estimated pI = 7.14)

MVHHHHHHGSGLFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPG
FGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFG
IPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIP
GVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGV
GIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGI
PGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPG
FGIPGFGIPGVGIPGFGIPGVGIPGFGIPGFGIPGVGIPGFGIPGVGIPGFGETAS*

Charge balanced, most hydrophobic ELP, ordered gene (bhhELP10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATC
CCGGGTTTTGGCATTCCGGGCAAGGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGAAGGCATTCC
GGGTTTCGGCATCCCGGGTAAGGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGTATTCCGG
GCTTTGGCATCCCGGGCAAGGGCATCCCGGGCTTTGGTATACCTGGCTTTGGCATTCCGGGCGAAGGTATCCCGGGC
TTTGGCATTCCGGGTAAGGGTATTCCGGGTTTTGGTATTCCGGGTTTCGGTATCCCGGGCGAAGGCATTCCGGGCTT
TGGTATTCCGGGCAAGGGTATCCCGGGTTTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (143 AA, 13494.6143 Da, Estimated pI = 7.14)

MVHHHHHHGSGLFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPG
FGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGETAS*

Charge balanced, most hydrophobic ELP, doubled (bhhELP20K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATC
CCGGGTTTTGGCATTCCGGGCAAGGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGAAGGCATTCC
GGGTTTCGGCATCCCGGGTAAGGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGTATTCCGG
GCTTTGGCATCCCGGGCAAGGGCATCCCGGGCTTTGGTATACCTGGCTTTGGCATTCCGGGCGAAGGTATCCCGGGC
TTTGGCATTCCGGGTAAGGGTATTCCGGGTTTTGGTATTCCGGGTTTCGGTATCCCGGGCGAAGGCATTCCGGGCTT
TGGTATTCCGGGCAAGGGTATCCCGGGTTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATCCCGGGTTTTG
GCATTCCGGGCAAGGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGAAGGCATTCCGGGTTTCGGC

ATCCCGGGTAAGGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGTATTCCGGGCTTTGGCAT
CCCGGGCAAGGGCATCCCGGGCTTTGGTATACCTGGCTTTGGCATTCCGGGCGAAGGTATCCCGGGCTTTGGCATTC
CGGGTAAGGGTATTCCGGGTTTTGGTATTCCGGGTTTCGGTATCCCGGGCGAAGGCATTCCGGGCTTTGGTATTCCG
GGCAAGGGTATCCCGGGTTTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (268 AA, 25185.2559 Da, Estimated pI = 7.14)

MVHHHHHHGSGLFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPG
FGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFG
IPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIP
GKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGETAS*

Charge balanced, most hydrophobic ELP, quadrupled (bhhELP40K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATC
CCGGGTTTTGGCATTCCGGGCAAGGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGAAGGCATTCC
GGGTTTCGGCATCCCGGGTAAGGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGTATTCCGG
GCTTTGGCATCCCGGGCAAGGGCATCCCGGGCTTTGGTATACCTGGCTTTGGCATTCCGGGCGAAGGTATCCCGGGC
TTTGGCATTCCGGGTAAGGGTATTCCGGGTTTTGGTATTCCGGGTTTCGGTATCCCGGGCGAAGGCATTCCGGGCTT
TGGTATTCCGGGCAAGGGTATCCCGGGTTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATCCCGGGTTTTG
GCATTCCGGGCAAGGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGAAGGCATTCCGGGTTTCGGC
ATCCCGGGTAAGGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGTATTCCGGGCTTTGGCAT
CCCGGGCAAGGGCATCCCGGGCTTTGGTATACCTGGCTTTGGCATTCCGGGCGAAGGTATCCCGGGCTTTGGCATTC
CGGGTAAGGGTATTCCGGGTTTTGGTATTCCGGGTTTCGGTATCCCGGGCGAAGGCATTCCGGGCTTTGGTATTCCG
GGCAAGGGTATCCCGGGTTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATCCCGGGTTTTGGCATTCCGGG
CAAGGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGAAGGCATTCCGGGTTTCGGCATCCCGGGTA
AGGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGTATTCCGGGCTTTGGCATCCCGGGCAAG
GGCATCCCGGGCTTTGGTATACCTGGCTTTGGCATTCCGGGCGAAGGTATCCCGGGCTTTGGCATTCCGGGTAAGGG
TATTCCGGGTTTTGGTATTCCGGGTTTCGGTATCCCGGGCGAAGGCATTCCGGGCTTTGGTATTCCGGGCAAGGGTA
TCCCGGGTTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATCCCGGGTTTTGGCATTCCGGGCAAGGGCATT
CCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGTGAAGGCATTCCGGGTTTCGGCATCCCGGGTAAGGGTATCCC
GGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGTATTCCGGGCTTTGGCATCCCGGGCAAGGGCATCCCGG
GCTTTGGTATACCTGGCTTTGGCATTCCGGGCGAAGGTATCCCGGGCTTTGGCATTCCGGGTAAGGGTATTCCGGGT
TTTGGTATTCCGGGTTTCGGTATCCCGGGCGAAGGCATTCCGGGCTTTGGTATTCCGGGCAAGGGTATCCCGGGTTT
TGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (518 AA, 48392.3789 Da, Estimated pI = 7.14)

MVHHHHHHGSLFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPG
FGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFG
IPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIP
GKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGK
GIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGI
PGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPG
FGIPGFGIPGEGIPGFGIPGKGIPGFGIPGFGIPGEGIPGFGIPGKGIPGFGETAS*

Negatively charged, most hydrophobic ELP, ordered gene (nhhELP10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATC
CCGGGTTTTGGCATTCCGGGTGAAGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGCGAGGGCATTCC
GGGTTTCGGCATCCCGGGCGAAGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGCATTCCGG
GCTTTGGTATTCCGGGTGAAGGCATCCCGGGCTTTGGCATCCCGGGCTTCGGTATACCTGGCGAAGGCATCCCGGGC
TTCGGAATCCCTGGCGAAGGTATCCCGGGTTTTGGTATTCCGGGCTTTGGCATTCCGGGCGAAGGTATTCCGGGTTT
CGGTATCCCGGGTGAAGGTATCCCGGGTTTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (143 AA, 13586.4004 Da, Estimated pI = 4.48)

MVHHHHHHGSLFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPG
FGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGETAS*

Negatively charged, most hydrophobic ELP, doubled (nhhELP20K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATC
CCGGGTTTTGGCATTCCGGGTGAAGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGCGAGGGCATTCC
GGGTTTCGGCATCCCGGGCGAAGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGCATTCCGG
GCTTTGGTATTCCGGGTGAAGGCATCCCGGGCTTTGGCATCCCGGGCTTCGGTATACCTGGCGAAGGCATCCCGGGC
TTCGGAATCCCTGGCGAAGGTATCCCGGGTTTTGGTATTCCGGGCTTTGGCATTCCGGGCGAAGGTATTCCGGGTTT
CGGTATCCCGGGTGAAGGTATCCCGGGTTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATCCCGGGTTTTG
GCATTCCGGGTGAAGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGCGAGGGCATTCCGGGTTTCGGC
ATCCCGGGCGAAGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGCATTCCGGGCTTTGGTAT
TCCGGGTGAAGGCATCCCGGGCTTTGGCATCCCGGGCTTCGGTATACCTGGCGAAGGCATCCCGGGCTTCGGAATCC
CTGGCGAAGGTATCCCGGGTTTTGGTATTCCGGGCTTTGGCATTCCGGGCGAAGGTATTCCGGGTTTCGGTATCCCG
GGTGAAGGTATCCCGGGTTTTGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (268 AA, 25194.6719 Da, Estimated pI = 4.01)

MVHHHHHHGSGLFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPG
FGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFG
IPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIP
GEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGETAS*

Negatively charged, most hydrophobic ELP, quadrupled (bhhELP40K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGTCTCTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATC
CCGGGTTTTGGCATTCCGGGTGAAGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGCGAGGGCATTCC
GGGTTTCGGCATCCCGGGCGAAGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGCATTCCGG
GCTTTGGTATTCCGGGTGAAGGCATCCCGGGCTTTGGCATCCCGGGCTTCGGTATACCTGGCGAAGGCATCCCGGGC
TTCGGAATCCCTGGCGAAGGTATCCCGGGTTTTGGTATTCCGGGCTTTGGCATTCCGGGCGAAGGTATTCCGGGTTT
CGGTATCCCGGGTGAAGGTATCCCGGGTTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATCCCGGGTTTTG
GCATTCCGGGTGAAGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGCGAGGGCATTCCGGGTTTCGGC
ATCCCGGGCGAAGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGCATTCCGGGCTTTGGTAT
TCCGGGTGAAGGCATCCCGGGCTTTGGCATCCCGGGCTTCGGTATACCTGGCGAAGGCATCCCGGGCTTCGGAATCC
CTGGCGAAGGTATCCCGGGTTTTGGTATTCCGGGCTTTGGCATTCCGGGCGAAGGTATTCCGGGTTTCGGTATCCCG
GGTGAAGGTATCCCGGGTTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATCCCGGGTTTTGGCATTCCGGG
TGAAGGCATTCCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGCGAGGGCATTCCGGGTTTCGGCATCCCGGGCG
AAGGTATCCCGGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGCATTCCGGGCTTTGGTATTCCGGGTGAA
GGCATCCCGGGCTTTGGCATCCCGGGCTTCGGTATACCTGGCGAAGGCATCCCGGGCTTCGGAATCCCTGGCGAAGG
TATCCCGGGTTTTGGTATTCCGGGCTTTGGCATTCCGGGCGAAGGTATTCCGGGTTTCGGTATCCCGGGTGAAGGTA
TCCCGGGTTTTGGTATCCCGGGTTTCGGCATTCCGGGCGAGGGCATCCCGGGTTTTGGCATTCCGGGTGAAGGCATT
CCGGGCTTCGGCATTCCGGGTTTTGGCATCCCGGGCGAGGGCATTCCGGGTTTCGGCATCCCGGGCGAAGGTATCCC
GGGCTTCGGTATTCCGGGCTTTGGTATCCCGGGCGAAGGCATTCCGGGCTTTGGTATTCCGGGTGAAGGCATCCCGG
GCTTTGGCATCCCGGGCTTCGGTATACCTGGCGAAGGCATCCCGGGCTTCGGAATCCCTGGCGAAGGTATCCCGGGT
TTTGGTATTCCGGGCTTTGGCATTCCGGGCGAAGGTATTCCGGGTTTCGGTATCCCGGGTGAAGGTATCCCGGGTTT
TGGTGAGACCGCTAGCTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (518 AA, 48411.2109 Da, Estimated pI = 3.65)

MVHHHHHHGSGLFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPG
FGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFG
IPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIP
GEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGE
GIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGI

PGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPG
FGIPGFGIPGEGIPGFGIPGEGIPGFGIPGFGIPGEGIPGFGIPGEGIPGFGETAS*

A2.2 Cononsolvency of Elastin-Like Polypeptides (ELPs) in Water/Alcohol Solutions

E20

Vector: pET28b; Cloning Site: BamHI/XhoI

DNA Sequence:

GGATCCATGGTGTCTAGCGGTCTCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGT
ACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTAC
CGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCG
GGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGG
TGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTG
TAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTA
GGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGG
CGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCG
TTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTT
CCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCC
AGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAG
GCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGC
GTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGT
TGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTG
GTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGT
GTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGT
ACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTAC
CTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCT
GGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGG
TGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGAGACCACTA
GTTAAATGAATAAGCTTTAACTCGAGCTCGAG

BamHI/XhoI sites

Amino Acid Sequence: (544 AA, 45693.8320 Da, Estimated pI = 10.00)

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSMVSSGLVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVG
IPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIP
GVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIPGV
GVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIPGVGV

PGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPG
VGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVG
VPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVP
GVGETTSGSACELMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTT
LTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHK
LEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDH
MVLLEFVTAAGITHGMDELYK*

Plasmid Map:

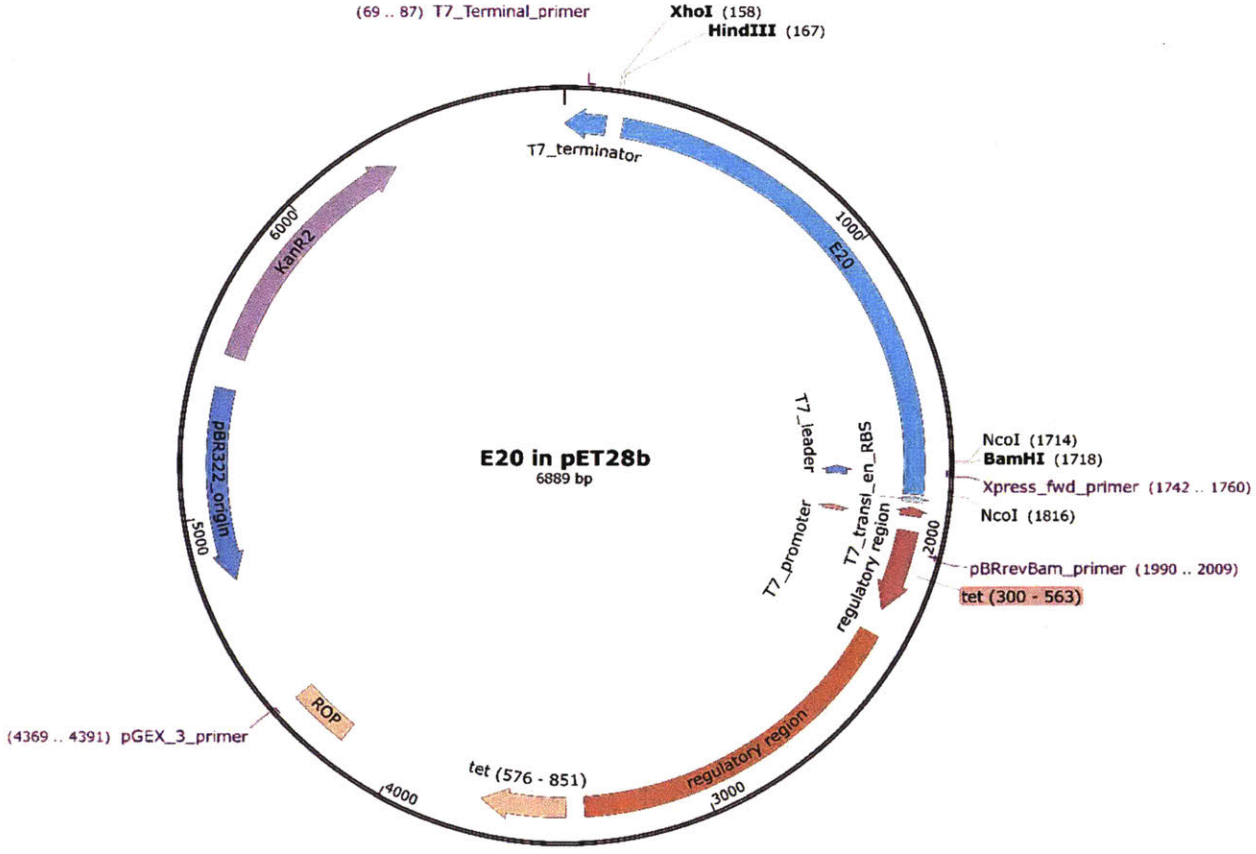


Figure A-1. Plasmid map for E20 in pET28b, cloned between BamHI/XhoI restriction sites.

A2.3 Protein Purification by Ethanol-Induced Phase Transitions of the Elastin-Like Polypeptide (ELP)

ELP-sfGFP

Vector: pET28a; Cloning Site: BamHI/HindIII

DNA Sequence:

GGATCCATGGTGTCTAGCGGTCTCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGT
ACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTAC
CGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCG
GGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGG
TGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTG
TAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTA
GGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGG
CGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCG
TTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTT
CCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCC
AGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAG
GCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGC
GTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGT
TGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTG
GTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGT
GTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGT
ACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTAC
CTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCT
GGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGG
TGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGAGACCACTA
GTGGATCCGCCTGCGAACTGATGGTGAGCAAAGGTGAAGAACTGTTTACTGGTGTAGTTCCGATCCTGGTAGAACTG
GACGGCGATGTTAACGGCCACAAGTTCAGCGTGCGTGGCGAAGGTGAAGGTGATGCAACCAACGGCAAACTGACCCT
GAAATTCATCTGTACTACCGGCAAACTGCCGGTGCCTTGGCCGACTCTGGTCACCACGCTGACTTACGGTGTTCAGT
GTTTCTCTCGTTACCCGGATCACATGAAACAGCACGATTTCTTCAAATCTGCTATGCCGGAAGGTTATGTGCAAGAG
CGTACCATCTCCTTCAAAGATGATGGCACCTACAAAACCCGTGCGGAAGTTAAATTCGAAGGTGACACGCTGGTCAA
CCGCATTGAACTGAAAGGCATCGATTTCAAAGAAGATGGCAACATCCTGGGCCACAAGCTGGAATACAACTTTAATT
CCCACAACGTATATATCACTGCGGATAAACAGAAAAACGGCATTAAAGCTAACTTCAAGATCCGCCACAACGTTGAA
GATGGTTCCGTTCAACTGGCGGACCACTACCAGCAGAACACTCCAATCGGTGACGGCCCAGTTCTGCTGCCAGATAA
CCACTACCTGTCTACCCAAAGCGCGCTGTCTAAAGACCCGAACGAAAAACGTGACCATATGGTTCTGCTGGAATTTG
TGACCGCGGCGGGCATCACGCATGGCATGGACGAGCTGTACAAGTAAAAGCTT

BamHI/HindIII sites

Amino Acid Sequence: (791 AA, 73306.9219 Da, Estimated pI = 6.66)

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSMVSSGLVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVG
IPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIP
GVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIPGV
GVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIPGVGV
PGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIPGVGVPG

VGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVG
VPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVP
GVGETTSGSACELMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTT
LTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHK
LEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDH
MVLLEFVTAAGITHGMDELYK*

Plasmid map:

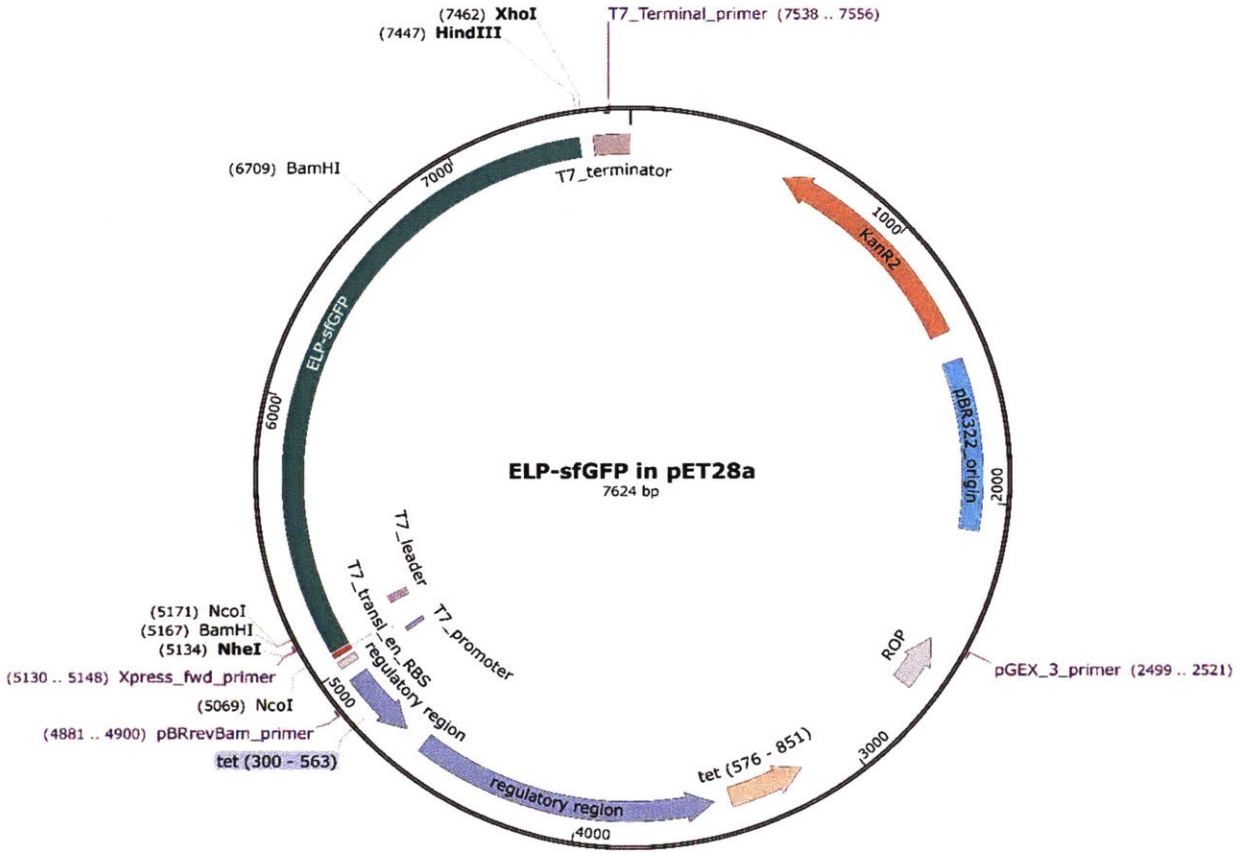


Figure A-2. Plasmid map for ELP-sfGFP in pET28a, cloned between BamHI/HindIII restriction sites.

P10

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CCATGGGCGGATCCGCTAGCGGTCTCGTTGGTATTCCTGCTGTTGGTGTGCCGGCTGTTGGTATCCCAGCTGTTGGC
GTTCCGGCTGTAGGTATTCCGGCTGTTGGTATTCCTGCTGTTGGTGTGCCGGCTGTTGGTATCCCAGCTGTTGGCGT
TCCGGCTGTAGGTATTCCGGCTGTTGGTATTCCTGCTGTTGGTGTGCCGGCTGTTGGTATCCCAGCTGTTGGCGTTC

CGGCTGTAGGTATTCCGGCTGTTGGTATTCCTGCTGTTGGTGTGCCGGCTGTTGGTATCCCAGCTGTTGGCGTTCCG
GCTGTAGGTATTCCGGCTGTTGGTATTCCTGCTGTTGGTGTGCCGGCTGTTGGTATCCCAGCTGTTGGCGTTCCGGC
TGTAGGTATTCCGGCTGTTGGTATTCCTGCTGTTGGTGTGCCGGCTGTTGGTATCCCAGCTGTTGGCGTTCCGGCTG
TAGGTATTCCGGCTGTTGGTATTCCTGCTGTTGGTGTGCCGGCTGTTGGTATCCCAGCTGTTGGCGTTCCGGCTGTA
GGTATTCCGGCTGTTGGTATTCCTGCTGTTGGTGTGCCGGCTGTTGGTATCCCAGCTGTTGGCGTTCCGGCTGTAGG
TATTCCGGCTGTTGGTATTCCTGCTGTTGGTGTGCCGGCTGTTGGTATCCCAGCTGTTGGCGTTCCGGCTGTAGGTA
TTCCGGCTGTTGGTATTCCTGCTGTTGGTGTGCCGGCTGTTGGTATCCCAGCTGTTGGCGTTCCGGCTGTAGGTATT
CCGGCTGTTGGTGAGACCACTAGTTAAATGAATAAGCTTTAACTCGAG

NcoI/XhoI Sites; **Start codon**

Amino Acid Sequence: (264 AA, 22849.7090 Da, Estimated pI = 3.81)

MGGSASGLVGIPAVGVPAVGIPAVGVPAVGIPAVGIPAVGVPAVGIPAVGVPAVGIPAVGIPAVGVPAVGIPAVGVP
AVGIPAVGIPAVGVPAVGIPAVGVPAVGIPAVGIPAVGVPAVGIPAVGVPAVGIPAVGIPAVGVPAVGIPAVGVPAV
GIPAVGIPAVGVPAVGIPAVGVPAVGIPAVGIPAVGVPAVGIPAVGVPAVGIPAVGIPAVGVPAVGIPAVGVPAVGI
PAVGIPAVGVPAVGIPAVGVPAVGIPAVGETTS*

ELP-OPH-ELP

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GGCTGGGGATCCGCTAGCGGTCTCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTT
GGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGG
TGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTG
TACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTA
CCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACC
GGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGG
GTGTAGGCGTTCCAGGCGTTGGCGAGACCACTAGCAGCGGCAGCATTGGCACCGGCGACCGTATCAACACCGTCCGT
GGCCCGATTACCATTAGCGAAGCAGGTTTTACCCTGACCCATGAACATATTTGCGGCAGCAGCGCGGGCTTTCTGCG
TGCTTGGCCGGAATTTTTCGGCAGCCGCAAAGCTCTGGCCGAAAAGGCGGTTCGCGGTCTGCGTCGCGCACGTGCGG
CCGGTGTTCGTACCATCGTGGATGTTTCTACGTTTGATATTGGCCGCGACGTCTCACTGCTGGCGGAAGTGTCGCGT
GCAGCTGATGTTCATATCGTTGCCGCAACCGGTCTGTGGTTTGACCCGCCGCTGAGTATGCGTCTGCGCTCCGTGGA
AGAACTGACCCAGTTTTTCCTGCGCGAAATTCAATATGGCATCGAAGATACGGGCATTCGTGCGGGTATTATCAAAG
TTGCGACCACGGGCAAGGCCACCCCGTTTCAGGAACTGGTTCTGCGTGCAGCTGCGCGTGCGAGCCTGGCAACCGGT
GTCCCGGTGACCACCCATACGGCCGCAAGTCAACGTGGCGGTGAACAGCAAGCTGCGATCTTCGAATCCGAAGGCCT
GTCACCGTCGCGCGTTTGTATTGGTCACAGCGATGACACCGATGACCTGTCTTATCTGACCGCCCTGGCGGCACGTG
GTTACCTGATTGGTCTGGATCGCATCCCGTTTTCAGCCATTGGCCTGGAAGACAACGCCAGCGCATCTGCTCTGCTG
GGTAACCGTTCGTGGCAGACCCGCGCACTGCTGATTAAAGCTCTGATCGATCAGGGCTATATGAAGCAAATCCTGGT
CAGCAATGACTGGCTGTTTGGTTTCAGTTCCTACGTGACGAACATTATGGATGTGATGGACAGCGTTAATCCGGATG

GTATGGCATTTATCCCGCTGCGTGTTATTCCGTTCCTGCGCGAAAAAGGCGTCCCGCAGGAAACCCTGGCCGGTATT
ACCGTGACGAACCCGGCTCGTTTCCTGTCCCCGACCCTGCGTGCGTCCCTGCAGGGCAGCACTAGCGGTCTCGTTGG
TGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTG
TACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTA
CCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACC
TGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTG
GTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGT
GTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGAGACCACTAG
TTAAATGAATAAGCTTTAACTCGAG

NcoI/XhoI Sites; **Start codon**

Amino Acid Sequence: (667 AA, 63843.7422 Da, Estimated pI = 6.72)

MGWGSASGLVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIPGVGV
PGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIPGVGVPG
VGVPGVGETTSSGSIGTGDRINTVRGPITISEAGFTLTHEHICGSSAGFLRAWPEFFGSRKALAEKAVRGLRRARAA
GVRTIVDVSTFDIGRDVSLLAEVSRAADVHIVAATGLWFDPPLSMRLRSVEELTQFFLREIQYGIEDTGIRAGIIKV
ATTGKATPFQELVLRAAARASLATGVPVTTHTAASQRGGEQQAAIFESEGLSPSRVCIGHSDDTDDLSYLTALAARG
YLIGLDRIPFSAIGLEDNASASALLGNRSWQTRALLIKALIDQGYMKQILVSNDWLFGFSSYVTNIMDVMDSVNPDG
MAFIPLRVIPFLREKGVPQETLAGITVTNPARFLSPTLRASLQGSTSGLVGVPGVGVPGVGIPGVGVPGVGVPGVGV
PGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPG
VGVPGVGIPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGETTS*

A2.4 High-Throughput Screening of a Streptavidin Binder Library in Self-Assembled Solid Films

A2.4.1 ELP and ELP-mCherry Genes

N-terminal ELP gene, ordered (Nt-deELP10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTGGATCCGGGAAGACTGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTA
CCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACC
GGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGG
GTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGT
GTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGT
AGGCGTTCCAGGCGTTGGTGAGACCGAATTCGTAGAACGGCCGTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (142 AA, 12130.2910 Da, Estimated pI = 4.70)

MVGSGKTVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPG
VGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGETEFVERP*

N-terminal ELP gene, doubled (Nt-deELP20K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTGGATCCGGGAAGACTGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTA
CCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACC
GGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGG
GTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGT
GTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGT
AGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAG
GCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGC
GTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGT
TCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTC
CAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCA
GGCGTTGGTGAGACCGAATTCGTAGAACGGCCGTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (267 AA, 22437.5352 Da, Estimated pI = 4.70)

MVGSGKTVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPG
VGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPG
VPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVP
GVGVPGVGVPGVGIPGVGVPGVGVPGVGETEFVERP*

C-terminal ELP vector (Ct-vec)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTGGATCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTC
GAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCC
CCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACA
TCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTG
GTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCC
CTCCGACGGCCCCGTAATGCGAATTCGTAGAACGGCCGGGGAAGACTGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGT
AGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAG
GTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGT
ATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTAT
CCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGAGACCAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (144 AA, 16194.4619 Da, Estimated pI = 6.81)

MVGSIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADI
PDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMRIRRTAVNIG*

C-terminal single ELP (Ct-deELP10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTGGATCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTC
GAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCC
CCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACA
TCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTG
GTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCC
CTCCGACGGCCCCGTAATGCGAATTCGTAGAACGGCCGTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (270 AA, 32968.0156 Da, Estimated pI = 10.56)

MVGSIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADI
PDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMRIRRTAGEDCWCTWCWRPGC
RYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRY
PRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRW*

C-terminal ELP, doubled (Ct-deELP20K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTGGATCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTC
GAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCC
CCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACA
TCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTG
GTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCC
CTCCGACGGCCCCGTAATGCGAATTCGTAGAACGGCCGGGGAAGACTGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGT
AGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAG
GTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGT
ATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTAT
CCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCC
CAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCA
GGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGG

CGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCG
TTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTT
GGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGAGACCAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (395 AA, 49534.4062 Da, Estimated pI = 11.32)

MVGSIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADI
PDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMRIRRTAGEDCWCTWCWRPGC
RYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRY
PRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWCTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPR
RWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPRRW
CTGCRRSRRW*

C-terminal ELP, tripled (Ct-deELP30K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTGGATCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTC
GAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCC
CCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACA
TCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTG
GTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCC
CTCCGACGGCCCCGTAATGCGAATTCGTAGAACGGCCGGGGAAGACTGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGT
AGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAG
GTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGT
ATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTAT
CCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCC
CAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCA
GGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGG
CGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCG
TTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTT
GGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGG
TGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTG
TACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTA
CCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACC
GGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGG
GTGTAGGCGTTCCAGGCGTTGGTGAGACCAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (520 AA, 66100.7969 Da, Estimated pI = 11.51)

MVGSIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADI
PDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMRIRRTAGEDCWCTWCWRPGC
RYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRY
PRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWCTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPR
RWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPRRW
CTGCRRSRRWCTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCT
GCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRWRTWCWRPGCRYPRRWCTGCRRSRRW*

A2.4.2 DNA Used for Gibson Assembly of ELP-rcSso7d-ZE Plasmid

pET28b Backbone GA Amplicon

Flanking restriction sites: XhoI/NdeI

Primers: ESZ-P-lin-for & ESZ-P-lin-rev

5' Overlap: ZE GA Amplicon 3' terminus (**bolded**)

3' Overlap: ELP GA Fragment 5' terminus (**bolded**)

Sequence (5'-3'):

ATGGTCCGCTGTAACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGA
AGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTG
AGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATT
AAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCT
TTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCC
CTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACG
TAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGA
CTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGC
CGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATT
AACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAA
ATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGC
AATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT
CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAAT
ACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAA
TCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGT
CATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCG
ATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCA

ACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGG
TGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAG
CCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAAC
TCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCC
ATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTG
AATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAAT
CCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGAT
CCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGC
CGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGT
CCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTG
CTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGAT
AGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAAC
GACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAG
GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACG
CCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCT
TTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAG
CTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCT
GATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAA
TCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCG
CCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGAC
AAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCA
GCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGC
TCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTT
CCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACG
AGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAA
CAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATA
CAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGG
CGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTC
GCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAA
GGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCAT
GCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGG
GCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGC
CGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGC
GGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGT

CGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGT
CGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGG
GCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTG
AGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAAC
GGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCA
GCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGG
AACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGT
TCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGA
CAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAG
TCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAAC
GCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCA
GCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTAC
CATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGC
GCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCA
CGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTG
GCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAAC
GTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGG
TTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGC
CCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAA
CAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGA
GCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGC
CGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGG
GAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
GGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGTTGGAAG**CG**
GGAAGACTGTTGGTGTAC

ELP GA Fragment (following GA excision)

Flanking restriction sites: N/A

Primers: N/A

5’ Overlap: pET28b Backbone GA Amplicon 3’ terminus (**bolded**)

3’ Overlap: rcSso7d.SA.1 GA Amplicon 5’ terminus (**bolded**)

Sequence (5’-3’):

CGGGAAGACTGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGT
GTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTAC
CGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGG
TGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGC
GTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCC
CAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGG
TATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCC
CGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGG
CGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGT
GTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACC**GGGTGTAGGCGTTCCAGGCG**

rcSso7d.SA.1 GA Amplicon

Flanking restriction sites: BamHI/EcoRI

Primers: ESZ-Sso-for & ESZ-Sso-rev

5' Overlap: ELP GA Fragment 3' terminus (**bolded**)

3' Overlap: ZE GA Amplicon 5' terminus (**bolded**)

Sequence (5'-3'):

GGTGTAGGCGTTCCAGGCGTTGGTGAAACCGAATTTGTAGAACGGCCGGGTGGTGGTGGTAGCGGTGGTG
GCGGATCCATGGCAACCGTGAAATTCACATACCAAGGCGAAGAAAAACAGGTGGATATTAGCAAAATCAA
GATCGTGGCTCGTGACGGCCAGTACATTGACTTTAAATATGATGAAGGTGGTGGTGCCTATGGTTATGGT
TGGGTGAGCGAAAAAGATGCACCGAAAGAACTGCTGCAGATGCTGGAAAAGCAAGAATTC**GGTGGTAGCG**
GTGGTGGC

ZE GA Amplicon

Flanking restriction sites: XhoI

Primers: ESZ-ZE-for & ESZ-ZE-rev

5' Overlap: rcSso7d.SA.1 GA Amplicon 3' terminus (**bolded**)

3' Overlap: pET28b Backbone GA Amplicon 5' terminus (**bolded**)

Sequence (5'-3'):

TCGGTGGTAGCGGTGGTGGCGGTTCACTGGAGATCGAAGCGGCGGCGCTGGAGCAGGAAAACACCGCGCT
GGAAACCGAGGTGGCGGAGCTGGAACAGGAAGTGCAACGTCTGGAAAACATTGTGAGCCAATACCGTACC
CGTT**ATGGTCCGCTGTAA**CTCGAG

Assembled deELP20k-rcSso7d.SA.1-ZE Construct

Sequence (5'-3'):

CATATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGTTG
GAAGCGGGAAGACTGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACC
GGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGT
GTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCG
TTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCC
AGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGT
ATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGTG
TAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGCGTCCC
GGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTGTTGGC
GTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCTGGTG
TTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACC
TGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGT
GAAACCGAATTTGTAGAACGGCCGGGTGGTGGTGGTAGCGGTGGTGGCGGATCCATGGCAACCGTGAAAT
TCACATACCAAGGCGAAGAAAAACAGGTGGATATTAGCAAAATCAAGATCGTGGCTCGTGACGGCCAGTA
CATTGACTTTAAATATGATGAAGGTGGTGGTGCCTATGGTTATGGTTGGGTGAGCGAAAAAGATGCACCG
AAAGAACTGCTGCAGATGCTGGAAAAGCAAGAATTCGGTGGTAGCGGTGGTGGCGGTTCACTGGAGATCG
AAGCGGCGGCGCTGGAGCAGGAAAACACCGCGCTGGAAACCGAGGTGGCGGAGCTGGAACAGGAAGTGCA
ACGTCTGGAAAACATTGTGAGCCAATACCGTACCCGTTATGGTCCGCTGTAA

Plasmid Map:

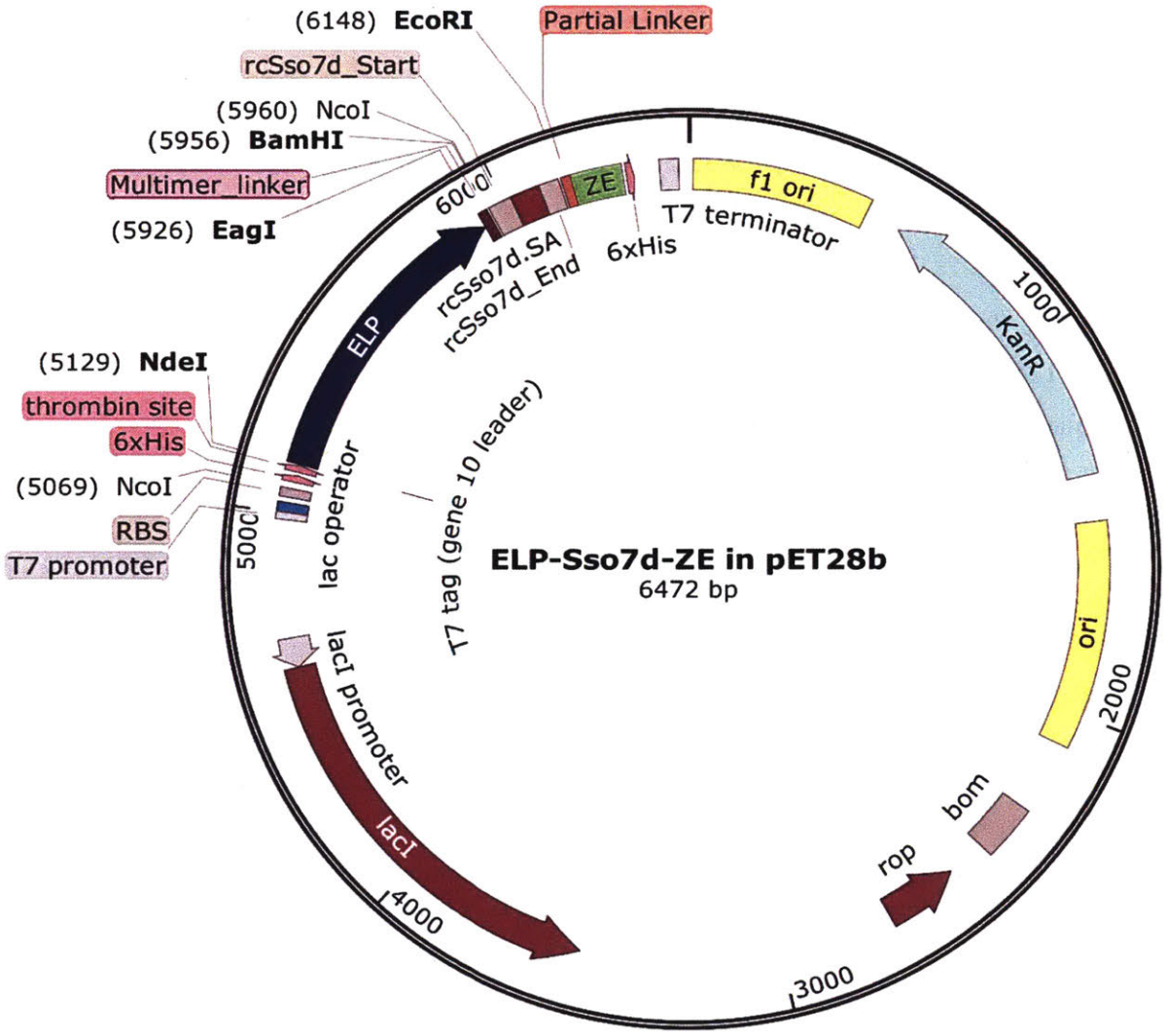


Figure A-3. Plasmid map for ELP-rcSso7d-ZE gene, cloned between NdeI/XhoI sites in pET28b plasmid.

A2.5 Tyrosine-Containing ELPs

N-terminal tyrosine-containing ELP gene, ordered (Nt-ELPY10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGGAAGACTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGT
ATTCCGGGCTACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTAT
CCCGGGTTATGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTC
CGGGGTACGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCG

GGGTACGGCGTCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGG
CTATGGCGTGCCGGGTGGTGGTGTGCCGGGTGTTGGTGAGACCGAATTCGTAGAACGGCCGTAAACATAGGATAAGC
TTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (148 AA, 12852.5488 Da, Estimated pI = 7.14)

MVHHHHHHGSGKTVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIP
GYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGETEFVERP*

N-terminal tyrosine-containing ELP gene, doubled (Nt-ELPY20K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGGAAGACTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGT
ATTCCGGGCTACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTAT
CCCGGGTTATGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTC
CGGGGTACGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCG
GGGTACGGCGTCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGG
CTATGGCGTGCCGGGTGGTGGTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCT
ACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTAT
GGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGG
TGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCG
TCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTG
CCGGGTGGTGGTGTGCCGGGTGTTGGTGAGACCGAATTCGTAGAACGGCCGTAAACATAGGATAAGCTTTAACTCGA
G

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (273 AA, 23059.2051 Da, Estimated pI = 7.14)

MVHHHHHHGSGKTVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIP
GYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGY
GVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGV
PGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGETEFVERP*

N-terminal tyrosine-containing ELP gene, tripled (Nt-ELPY30K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGGAAGACTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGT
ATTCCGGGCTACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTAT
CCCGGGTTATGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTC
CGGGGTACGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCG

GGGTACGGCGTCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGG
CTATGGCGTGCCGGGTGGTGGTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCT
ACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTAT
GGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGG
TGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCG
TCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTG
CCGGGTGGTGGTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCTACGGTGTTCC
GGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTATGGTGTGCCGG
GCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGGTGTGCCGGGC
GGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCGTCCCGGGCGG
TGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTGCCGGGTGGTG
GTGTGCCGGGTGTTGGTGAGACCGAATTCGTAGAACGGCCGTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (398 AA, 33265.8594 Da, Estimated pI = 7.14)

MVHHHHHHGSGKTVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIP
GYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGY
GVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGV
PGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPG
GGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGG
VPGVGETEFVERP*

N-terminal tyrosine-containing ELP gene, quadrupled (Nt-ELPY40K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGGAAGACTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGT
ATTCCGGGCTACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTAT
CCCGGGTTATGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTC
CGGGGTACGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCG
GGGTACGGCGTCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGG
CTATGGCGTGCCGGGTGGTGGTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCT
ACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTAT
GGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGG
TGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCG
TCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTG
CCGGGTGGTGGTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCTACGGTGTTCC
GGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTATGGTGTGCCGG
GCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGGTGTGCCGGGC
GGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCGTCCCGGGCGG

TGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTGCCGGGTGGTG
GTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCTACGGTGTTCCGGGCGGTGGC
GTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTATGGTGTGCCGGGCGGTGGCGT
TCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGGTGTGCCGGGCGGTGGCGTTC
CGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCGTCCCGGGCGGTGGCGTTCCG
GGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTGCCGGGTGGTGGTGTGCCGGG
TGTTGGTGAGACCGAATTCGTAGAACGGCCGTAAACATAGGATAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (523 AA, 43472.5156 Da, Estimated pI = 7.14)

MVHHHHHHGSGKTVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIP
GYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGY
GVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGV
PGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPG
GGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGG
VPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVP
GVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGETEFVERP*

C-terminal ELPY vector (Ct-ELPY vec)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCC
GTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAA
GGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACG
TGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAAC
TTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCT
GCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCGAATTCGTAGAACGGCCGTAAACATAGGATAAGCTTTAA
CTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (150 AA, 17017.3066 Da, Estimated pI = 7.33)

MVHHHHHHGSIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYV
KHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMRIRRTAVNIG*

C-terminal single tyrosine-containing ELP (Ct-ELPY10K)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCC
GTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAA

GGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACG
TGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAAC
TTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCT
GCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCGAATTCGTAGAACGGCCGGGAAGACTGTTGGTGTGCCGG
GTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCTACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGC
GTTGGTGTGCCGGGTGGCGGTATCCCGGGTTATGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGT
GGGCGTTCCGGGTGGCGGTATTCCGGGGTACGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGG
GCGTTCCGGGTGGCGGTATCCCGGGGTACGGCGTCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGC
GTGCCGGGTGGCGGTATTCCGGGCTATGGCGTGCCGGGTGGTGGTGTGCCGGGTGTTGGTGAGACCAAGCTTTAACT
CGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (280 AA, 27754.5859 Da, Estimated pI = 7.75)

MVHHHHHHGSIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYV
KHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMRIRRTAVNIG*

Nt-ELPY10K-mCherry

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGGAAGACTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGT
ATTCCGGGCTACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTAT
CCCGGGTTATGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTC
CGGGGTACGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCG
GGGTACGGCGTCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGG
CTATGGCGTGCCGGGTGGTGGTGTGCCGGGTGTTGGTGAGACCGAATTCATGGTGAGCAAGGGCGAGGAGGATAACA
TGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAG
GGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTT
CGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACT
ACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTG
ACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGG
CCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGG
GCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAG
AAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCAT
CGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAGAAGCTTTAAC
TCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (380 AA, 39075.0898 Da, Estimated pI = 6.40)

MVHHHHHHGSGKTVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIP

GYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGETEFMVSKGEEDNM
AIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDY
LKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKG
EIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK*

Nt-ELPY20K-mCherry

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGGAAGACTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGT
ATTCCGGGCTACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTAT
CCCGGGTTATGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTC
CGGGGTACGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCG
GGGTACGGCGTCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGG
CTATGGCGTGCCGGGTGGTGGTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCT
ACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTAT
GGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGG
TGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCG
TCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTG
CCGGGTGGTGGTGTGCCGGGTGTTGGTGAGACCGAATTCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCAT
CAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCG
AGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGAC
ATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCT
GTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACT
CCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATG
CAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAA
GCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGC
AGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAG
TACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAGAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (505 AA, 49281.7500 Da, Estimated pI = 6.40)

MVHHHHHHGSGKTVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIP
GYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGY
GVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGV
PGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGETEFMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGE
GRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDS
SLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQ
LPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK*

Nt-ELPY30K-mCherry

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGGAAGACTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGT
ATTCCGGGCTACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTAT
CCCGGGTTATGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTC
CGGGGTACGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCG
GGGTACGGCGTCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGG
CTATGGCGTGCCGGGTGGTGGTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCT
ACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTAT
GGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGG
TGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCG
TCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTG
CCGGGTGGTGGTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCTACGGTGTTCC
GGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTATGGTGTGCCGG
GCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGGTGTGCCGGGC
GGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCGTCCCGGGCGG
TGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTGCCGGGTGGTG
GTGTGCCGGGTGTTGGTGAGACCGAATTCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTC
ATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCC
CTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCC
CTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCC
GAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCA
GGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGA
CCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTG
AAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGG
CGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCG
CCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAGAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (630 AA, 59488.4062 Da, Estimated pI = 6.40)

MVHHHHHHGSGKTVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIP
GYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGY
GVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGV
PGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPG
GGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGG
VPGVGETEFMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSP
QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKT

MGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERA
EGRHSTGGMDELYK*

Nt-ELPY40K-mCherry

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GTTCATCATCATCATCATCACGGATCCGGGAAGACTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGT
ATTCCGGGCTACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTAT
CCCGGGTTATGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTC
CGGGGTACGGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCG
GGGTACGGCGTCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGG
CTATGGCGTGCCGGGTGGTGGTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCT
ACGGTGTTCCGGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTAT
GGTGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGG
TGTGCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCG
TCCCGGGCGGTGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTG
CCGGGTGGTGGTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCTACGGTGTTCC
GGGCGGTGGCGTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTATGGTGTGCCGG
GCGGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGGTGTGCCGGGC
GGTGGCGTTCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCGTCCCGGGCGG
TGGCGTTCCGGGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTGCCGGGTGGTG
GTGTGCCGGGTGTTGGTGTGCCGGGTGTGGGTGTGCCGGGCGGTGGTATTCCGGGCTACGGTGTTCCGGGCGGTGGC
GTTCCGGGTGTTGGCGTGCCGGGCGTTGGTGTGCCGGGTGGCGGTATCCCGGGTTATGGTGTGCCGGGCGGTGGCGT
TCCGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATTCCGGGGTACGGTGTGCCGGGCGGTGGCGTTC
CGGGTGTGGGCGTTCCGGGTGTGGGCGTTCCGGGTGGCGGTATCCCGGGGTACGGCGTCCCGGGCGGTGGCGTTCCG
GGTGTGGGCGTTCCAGGTGTTGGCGTGCCGGGTGGCGGTATTCCGGGCTATGGCGTGCCGGGTGGTGGTGTGCCGGG
TGTTGGTGAGACCGAATTCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCA
AGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC
ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCAT
GTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCA
AGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAG
TTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTG
GGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGG
ACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAC
GTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCG
CCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAGAAGCTTTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (755 AA, 69695.0703 Da, Estimated pI = 6.40)

MVHHHHHHGSGKTVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIP
GYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGY
GVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGV
PGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPG
GGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGG
VPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVP
GVGVPGVGVPGGGIPGYGVPGGGVPGVGVPGVGVPGGGIPGYGVPGGGVPGVGETEFMVSKGEEDNMAIIKEFMRFK
VHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFK
WERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKD
GGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK*

A3. Table of Primers Used for PCR in this Thesis

Chapter or Project	Name	Sequence	Overhang	Restriction Site(s)	Recommended Tm (degC)	Template DNA
3	mCherry-Rev-HindIII	CAAGCTCAGCTAATTAAGCTTCTACTTGTA	CAAGCTCAGCTAATTAAGCTTCTACTTGTA	HindIII	67	mCherry in pQE9
3	NheI-mCherry-Fwd	TAAGCAGCTAGCATGGTGAGCAAGGGCGAG	ATGGTGAGCAAGGGCGAG	NheI	67	mCherry in pQE9
6, Tyrosine-Containing ELPs	ELP-concat-fwd	tagtaggttgaggccgttgagc	tagtaggttgaggccgttgagc		71	Nt-ELPY10K, Nt-deELP10K
6, Tyrosine-Containing ELPs	ELP-concat-rev	gtggtggtgctcgagTTAAAGC	gtggtggtgctcgagTTAAAGC		71	Nt-ELPY10K, Nt-deELP10K
6, Tyrosine-Containing ELPs	BamHI-mChFrag-Fwd	GATCAggatccATCATCAAGGAGTTCATGCG	ATCATCAAGGAGTTCATGCG	BamHI	70	mCherry in pQE9
6, Tyrosine-Containing ELPs	EcoRI-mChFrag-Rev	TTACAgaattcGCATTACGGGGCCG	GCATTACGGGGCCG	EcoRI	70	mCherry in pQE9
6, Tyrosine-Containing ELPs	EagI-ELPY10K-Fwd	CATTAcggccgGGGAAGACTGTTGG	GGGAAGACTGTTGG	EagI	56	Nt-ELPY10K, Nt-deELP10K
6, Tyrosine-Containing ELPs	HindIII-ELPY10K-Rev	TTACCaagcttGGTCTCACCAACg	GGTCTCACCAACg	HindIII	56	Nt-ELPY10K, Nt-deELP10K
Tyrosine-Containing ELPs	EcoRI-mCherry-Fwd	TAGCAgaattcATGGTGAGCAAGG	ATGGTGAGCAAGG	EcoRI	54	mCherry in pQE9
Tyrosine-Containing ELPs	HindIII-mCherry-Rev	GAATTaagcttCTACTTGTACAG	CTACTTGTACAG	HindIII	54	mCherry in pQE9

A4. Ru(bpy) Catalyzed Crosslinking of Tyrosine-Containing ELP Films

Tyrosine-containing ELPs were originally designed with the end goal of enabling photo-crosslinking of ELP and globular protein domains in a thin-film format. Literature established that tyrosine residues in proteins could be crosslinked by combining ultraviolet light irradiation with the tris(bipyridine)ruthenium(II) ($Ru(bpy)_3^{2+}$) catalyst in the presence of a persulfate.[1] To assess the feasibility of crosslinking fusion protein films using this chemistry, thin films were prepared by flow coating films out of of 15 wt% protein solutions onto silicon wafers. Film thickness was controlled by coating speed. Based on the color of the film, the thickness can be estimated based on thin-film interference with the light. The color of film was used in these preliminary experiments to gauge the effectiveness of the crosslinking. **Figure A-4** shows the different film colors produced by flow coating of Nt-ELPY10K, and the approximate estimated film thicknesses based on film color.

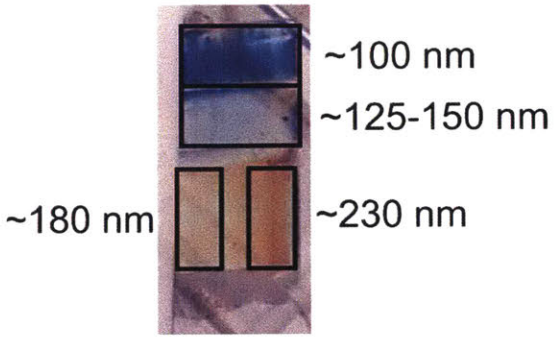


Figure A-4. Film coated with Nt-ELPY10K, showing the different colors and estimated thicknesses based on thin-film interference.

A4.1 Crosslinking of Nt-ELPY10K

Protein films were flow coated out of 15 wt% Nt-ELPY10K solution to produce films of varying thicknesses. After flow coating, films were submerged in a thin layer of aqueous solution containing 2 M NaCl (which was expected to keep the ELP immobilized), 0.1 or 0.5 mM $Ru(bpy)_3^{2+}$, and 100 mM sodium persulfate. Films were irradiated with a dental curing lamp (something like https://www.amazon.com/NSKI-Wireless-Cordless-1500mw-Silver/dp/B075D6PFTS/ref=asc_df_B075D6PFTS/?tag=hyprod-20&linkCode=df0&hvadid=221784482639&hvpos=1o1&hvnetw=g&hvrand=17314808092157

966634&hvpone=&hvptwo=&hvqmt=&hvdev=c&hvdvcmdl=&hvlocint=&hvlocphy=9002000&hvtargid=pla-379992223596&psc=1) for 90 seconds. Films were then washed with deionized water three times and tried under airflow.

Figure A-5 shows films of variate thickness before and after crosslinking using 0.1 and 0.5 mM $Ru(bpy)_3^{2+}$ solutions. In both cases, the thinnest part of the film (~100 nm, dark blue before crosslinking) appears to completely vanish after the crosslinking/washing procedure. The color change in the thicker films indicates loss of film thickness after this process, indicating that not all the film was immobilized after crosslinking. However, of the two conditions, the 0.5 mM $Ru(bpy)_3^{2+}$ solution appeared to produce sturdier films; thus, this solution was used for the remainder of the experiments. Film crosslinking was also attempted with untreated and poly(ethylene glycol) (PEG)-coated wafers. Wafers were coated with PEG as described in Dr. Stewart-Sloan's thesis.[2] **Figure A-6** shows films before and after crosslinking on wafters that were coated with PEG and uncoated. This data shows that there is little to no difference between the adhesion of the film to the wafer in the presence and absence of a PEG coating, and that, ultimately, only films that are ~175 nm in thickness or more adhere to the wafer after crosslinking. Even then, however, the films appear to be slightly flaky.

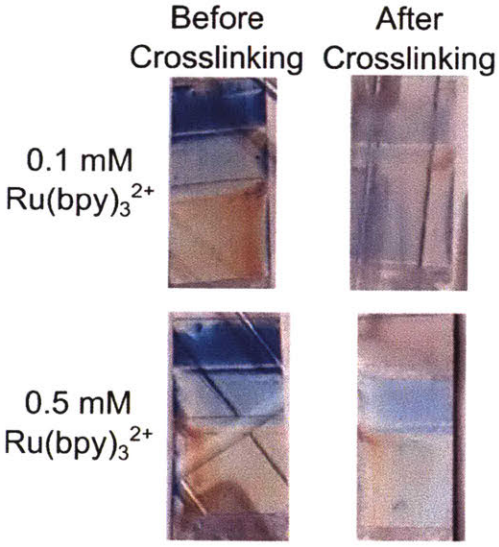


Figure A-5. Nt-ELPY10K thin films before and after $Ru(bpy)_3^{2+}$ catalyzed crosslinking.

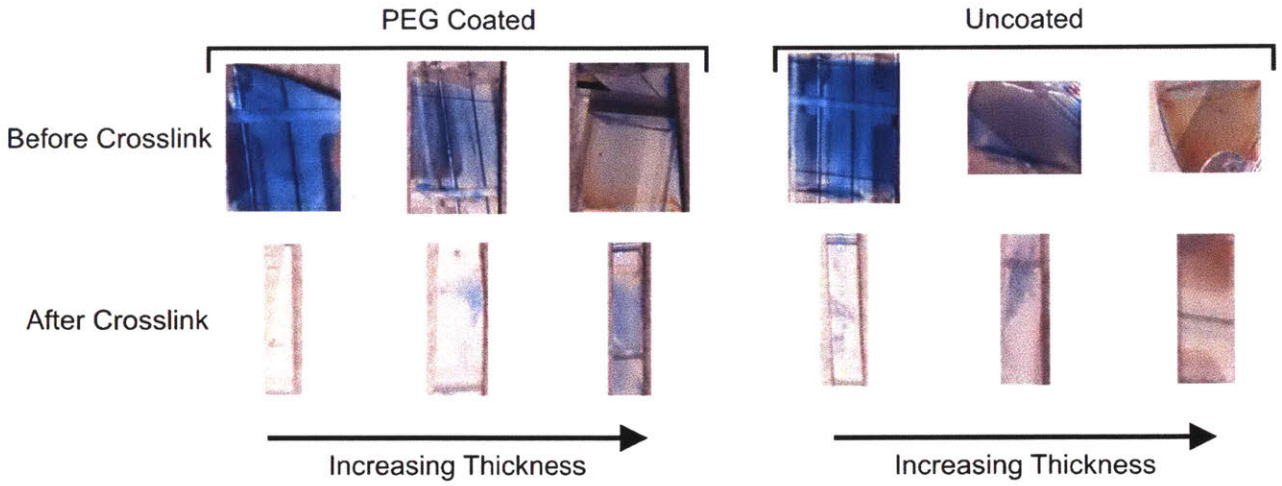


Figure A-6. Nt-ELPY10K thin films before and after $Ru(bpy)_3^{2+}$ catalyzed crosslinking on uncoated and PEG-coated wafers.

A4.2 Crosslinking of ELPY20K-mCherry and E20-mCherry Films

The previous section demonstrated that $Ru(bpy)_3^{2+}$ catalyzed crosslinking appeared to not be a robust method for achieving good immobilization of thin protein films for the ELPY construct alone. However, the possibility remained that crosslinking efficiency could be improved by crosslinking with tyrosine residues in the globular protein portion of the fusion protein. Thus, $Ru(bpy)_3^{2+}$ catalyzed crosslinking was also investigated for films coated from ELP-mCherry fusion proteins. Two different ELP sequences were used for this purpose—One containing an ELP with no tyrosine residues (E20, see **A2.2 Cononsolvency of Elastin-Like Polypeptides (ELPs) in Water/Alcohol Solutions** for sequence), and one containing tyrosine residues with the same frequency as Nt-ELPY10K. The sequence of this ELPY-mCherry fusion can be found in **A2.5 Tyrosine-Containing ELPs** under ELPY20K-mCherry. The length of the ELP block in these two fusion proteins differed; however, that was not expected to change the conclusions drawn from this experiment. In brief, the main goal of this experiment was to compare the crosslinking efficiency of $Ru(bpy)_3^{2+}$ catalyzed crosslinking to glutaraldehyde mediated crosslinking, which has been shown to be an effective route for immobilizing bioconjugate thin films.[3-5] **Figure A-7** shows the results of this experiment. While the photos might be slightly difficult to visualize, several key conclusions were reached. First, the $Ru(bpy)_3^{2+}$ catalyzed crosslinking, as expected, was more efficient at immobilizing ELPY20K-mCherry than E20-mCherry, likely due to the higher abundance of tyrosine residues in the ELPY20K-mCherry construct. However, even though

there was more film retention in this case, the film that was retained was splotchy and flaky in places. By comparison, in both cases, the glutaraldehyde crosslinking lead to smoother more reliably crosslinked films, regardless of the amino acid content of the ELP. This ultimately led to the decision to use glutaraldehyde mediated crosslinking of ELP-fusion protein films.

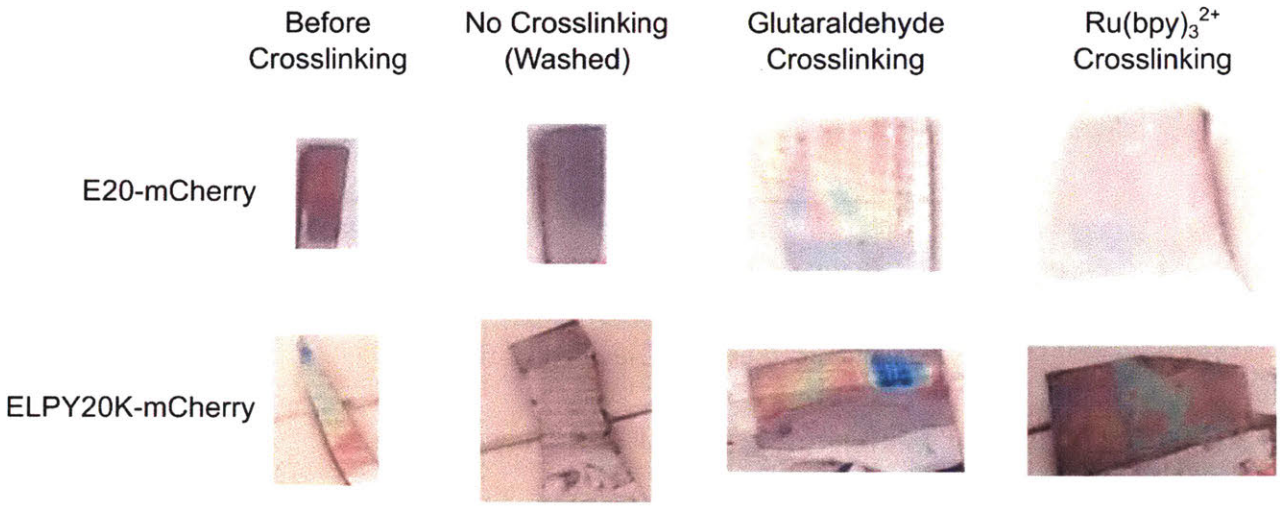


Figure A-7. Images of thin films before and after crosslinking and washing with MilliQ water. A negative control in which no crosslinking was performed is also included.

A4.3 References

1. Fancy, D. A.; Kodadek, T., Chemistry for the analysis of protein-protein interactions: Rapid and efficient cross-linking triggered by long wavelength light. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96* (11), 6020-6024.

2. Stewart-Sloan, C. Understanding the effect of protonation on the self-assembly of a model polyelectrolyte-neutral block copolymer. Massachusetts Institute of Technology, Cambridge, MA, 2016.

3. Huang, A.; Qin, G.; Olsen, B. D., Highly Active Biocatalytic Coatings from Protein–Polymer Diblock Copolymers. *ACS Appl. Mater. Interfaces* **2015,** *7* (27), 14660-14669.

4. Dong, X.-H.; Obermeyer, A. C.; Olsen, B. D., Three-Dimensional Ordered Antibody Arrays Through Self-Assembly of Antibody–Polymer Conjugates. *Angew. Chem. Int. Ed.* **2017,** *56* (5), 1273-1277.

5. Paloni, J. M.; Miller, E. A.; Sikes, H. D.; Olsen, B. D., Improved Ordering in Low Molecular Weight Protein–Polymer Conjugates Through Oligomerization of the Protein Block. *Biomacromolecules* **2018,** *19* (9), 3814-3824.

A5. Detailed FPLC Traces for Elution Fractions

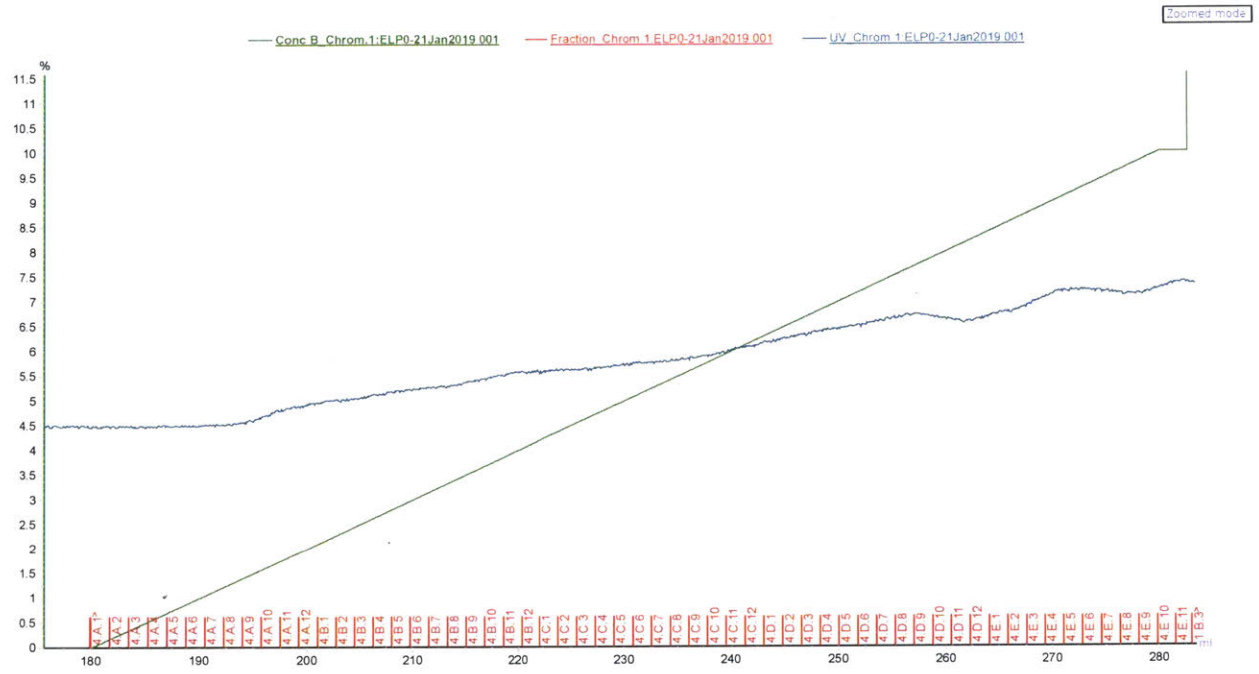


Figure A-8. A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, 6 M urea, pH 8 added at each elution fraction in FPLC (fractions are labeled in red along x-axis) for uELP.

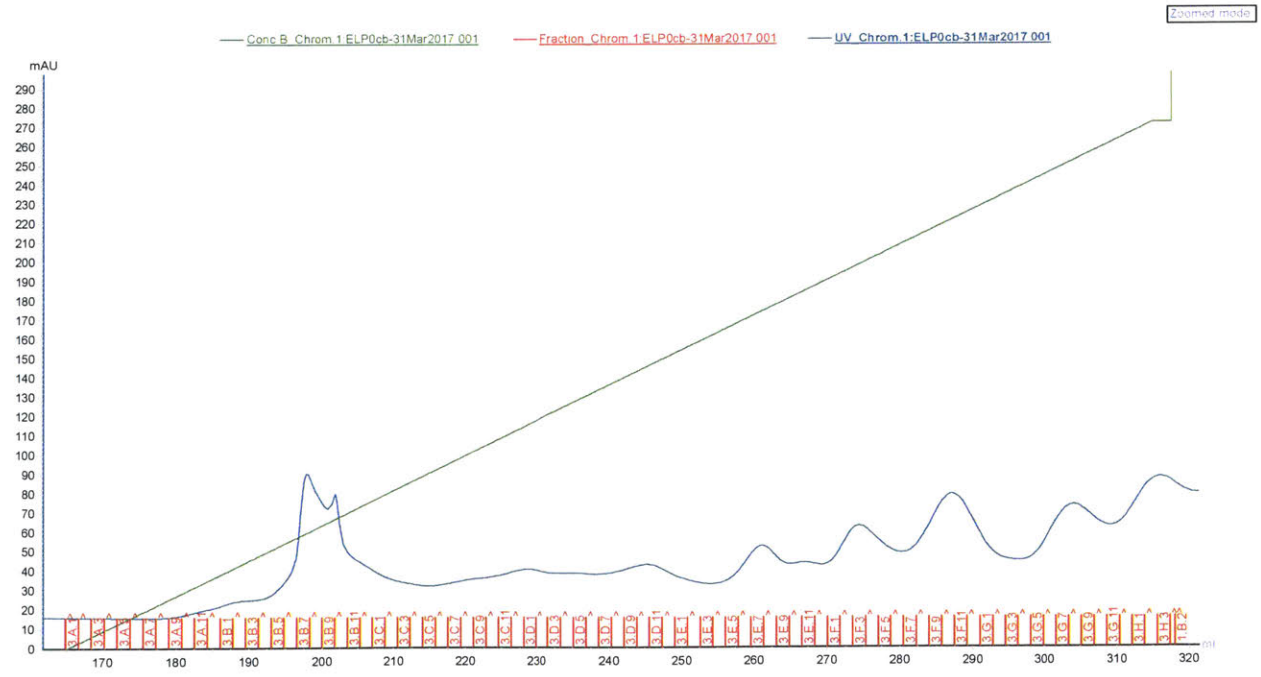


Figure A-9. A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 added at each elution fraction in FPLC (fractions are labeled in red along x-axis) for bELP.

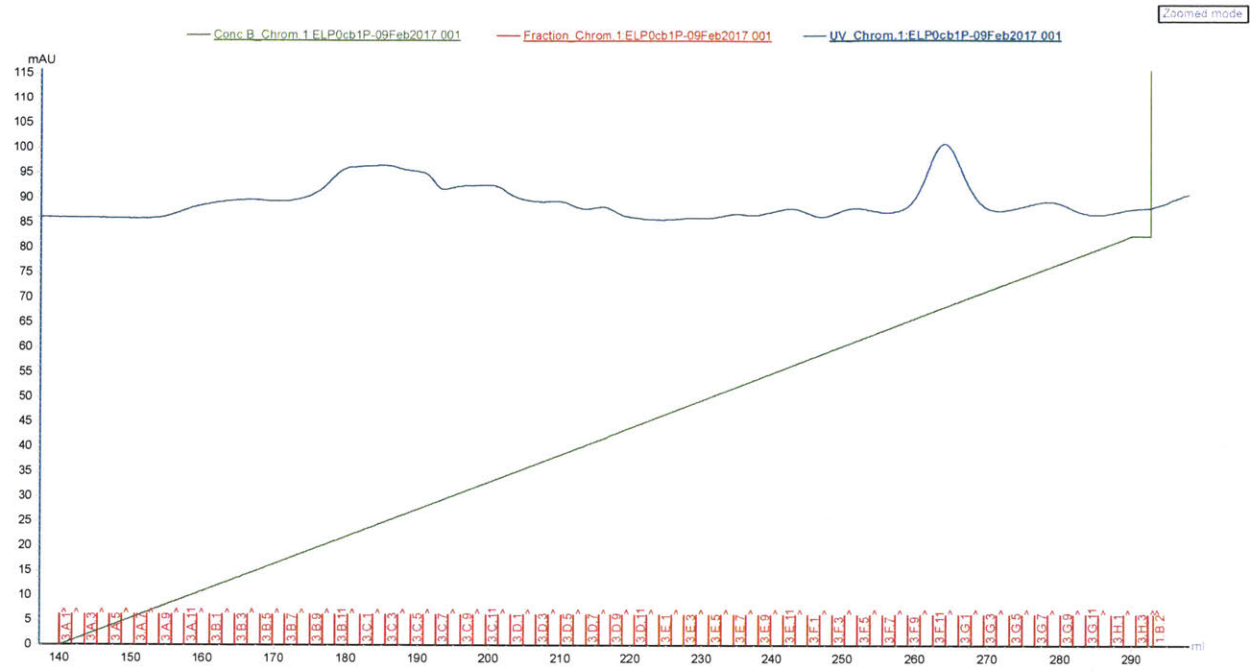


Figure A-10. A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 added at each elution fraction in FPLC (fractions are labeled in red along x-axis) for bhELP.

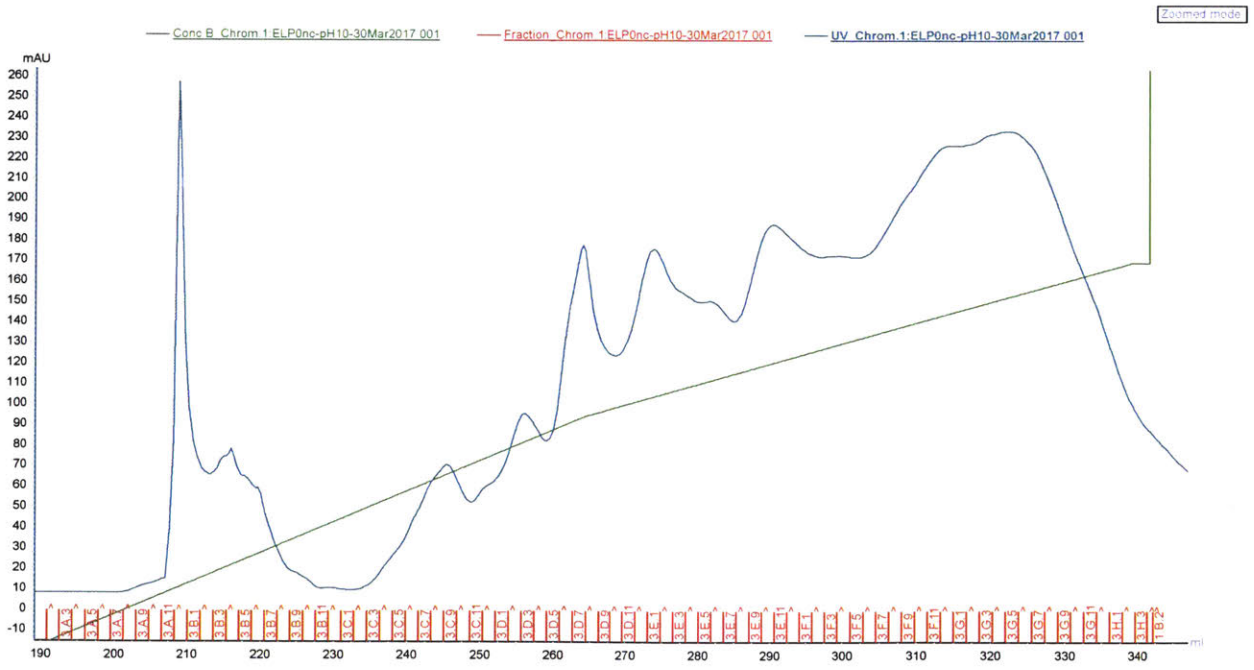


Figure A-11. A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 10 added at each elution fraction in FPLC (fractions are labeled in red along x-axis) for nELP.

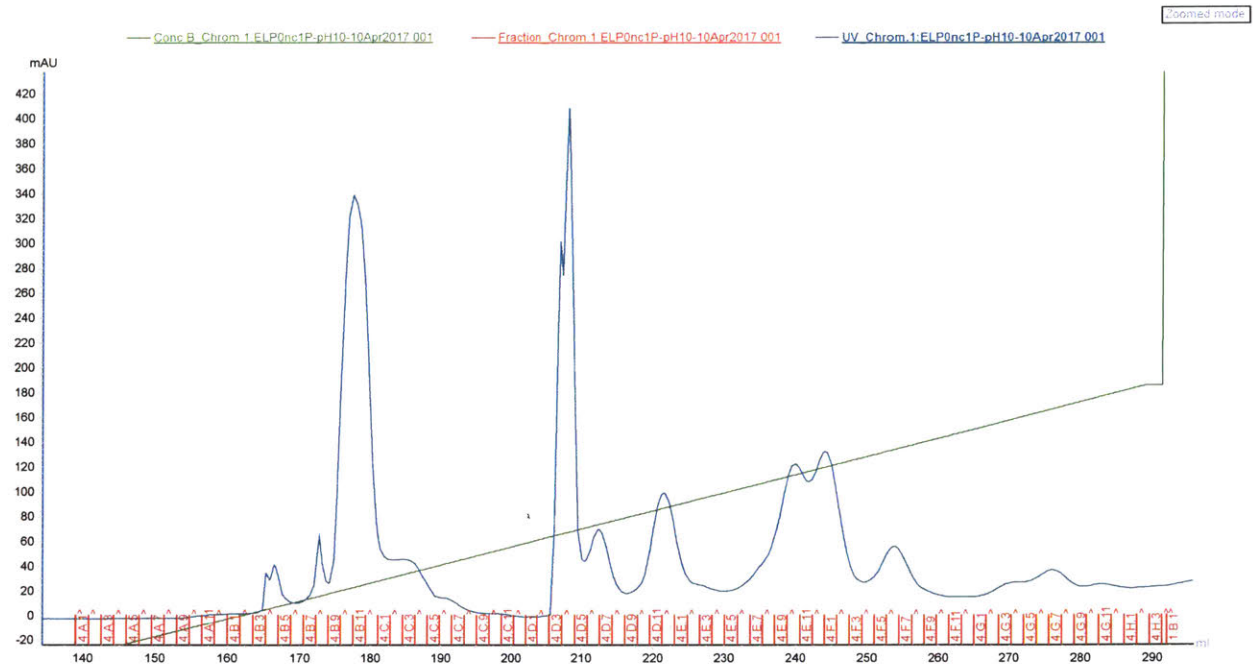


Figure A-12. A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 10 added at each elution fraction in FPLC (fractions are labeled in red along x-axis) for nhELP.

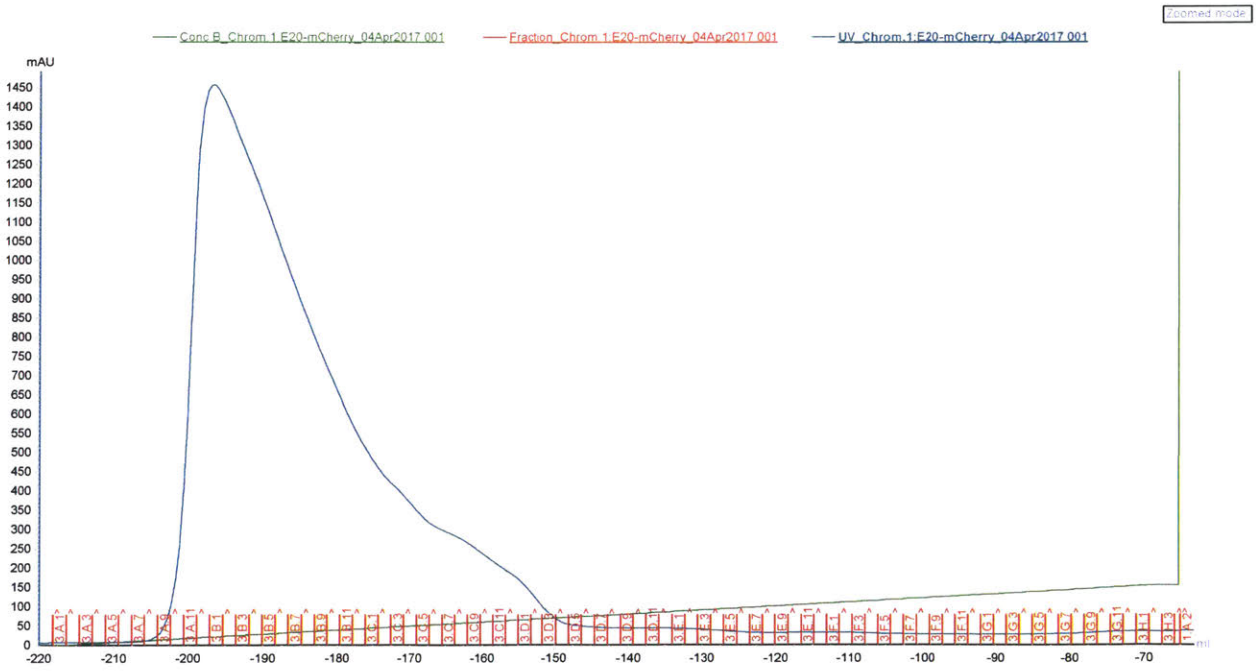


Figure A-13. A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 added at each elution fraction in FPLC (fractions are labeled in red along x-axis) for uELP-mCherry.

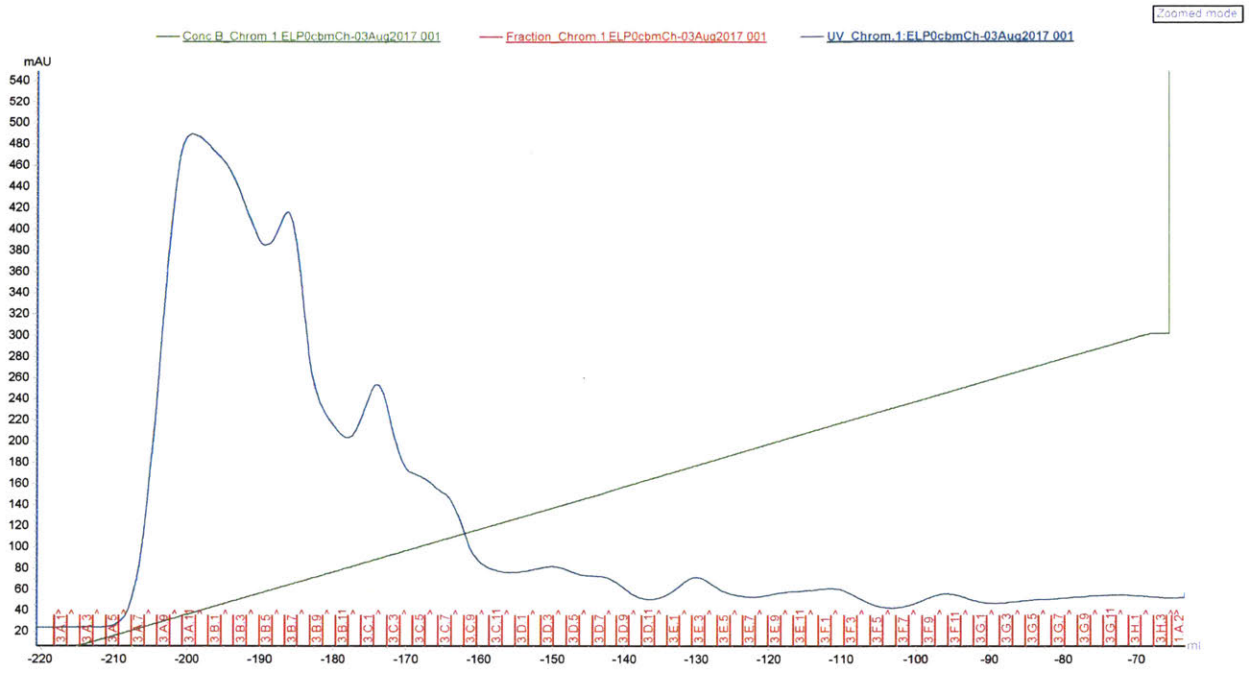


Figure A-14. A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 added at each elution fraction in FPLC (fractions are labeled in red along x-axis) for bELP-mCherry.

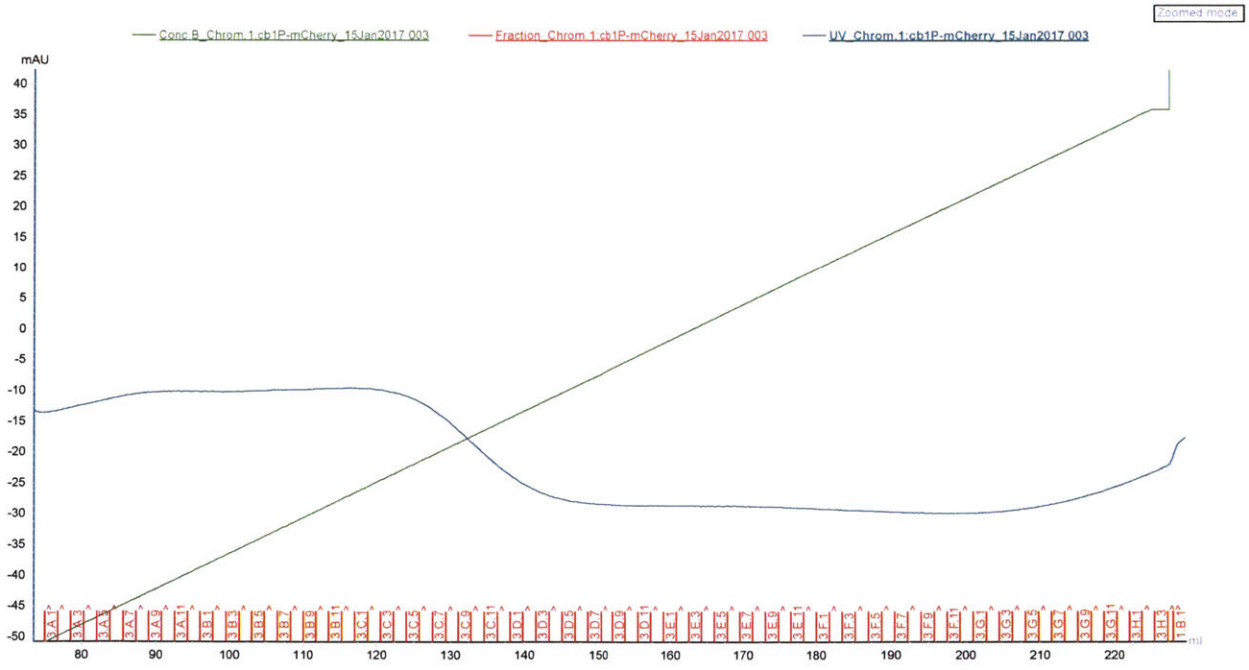


Figure A-15. A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 added at each elution fraction in FPLC (fractions are labeled in red along x-axis) for bhELP-mCherry.

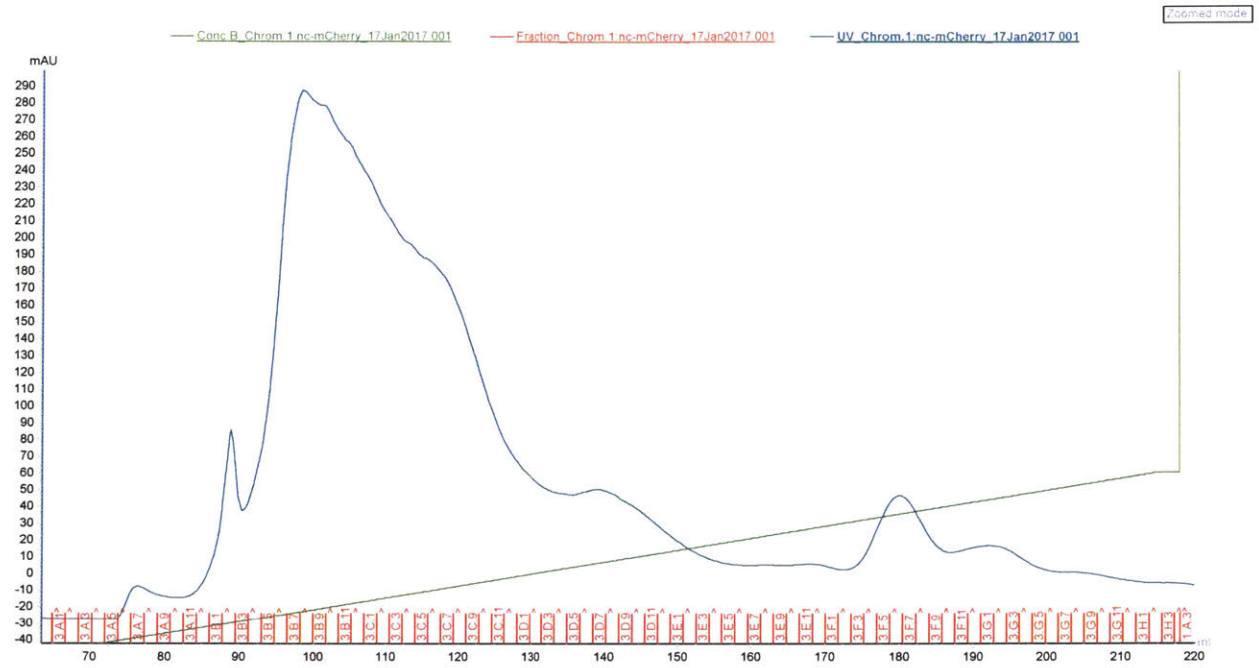


Figure A-16. A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 added at each elution fraction in FPLC (fractions are labeled in red along x-axis) for nELP-mCherry.

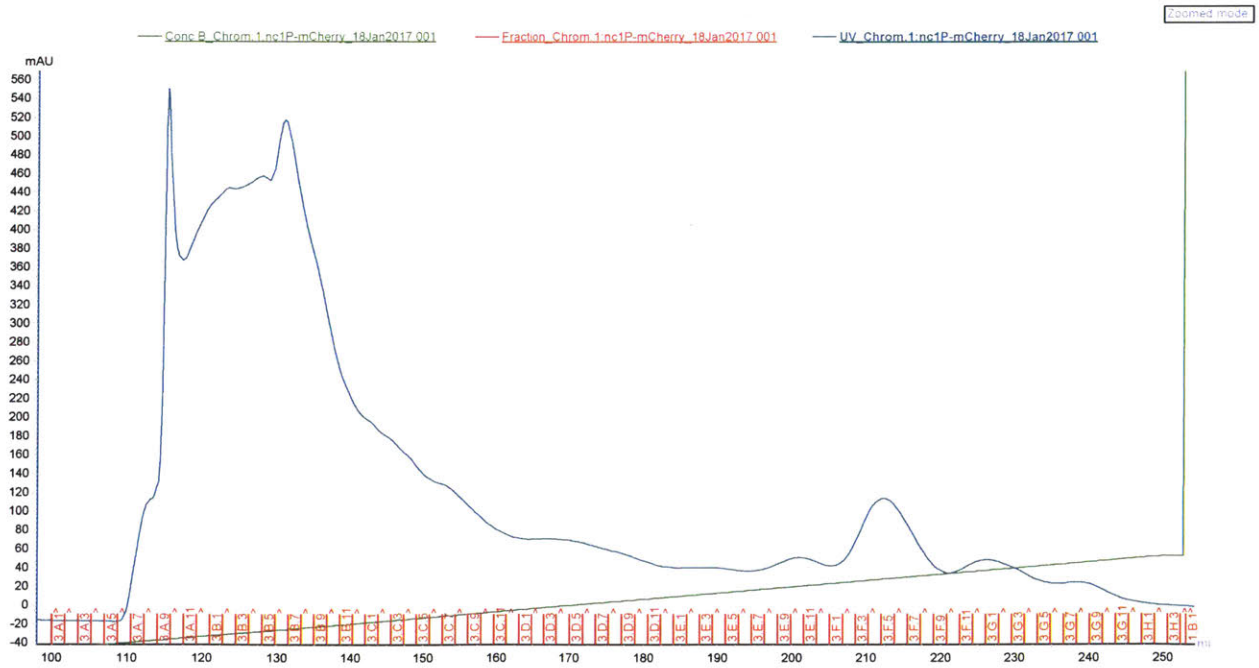


Figure A-17. A280 trace overlaid with percentage of 2 M NaCl 20 mM tris, pH 8 added at each elution fraction in FPLC (fractions are labeled in red along x-axis) for nhELP-mCherry.

Appendix B. Supporting Information for Chapter 3

B1. Supplementary Information

B1.1 Models for Small-Angle Neutron Scattering

B1.1.1 Swollen/collapsed Gaussian polymer

SANS data on uELP, bELP, and bhELP was fit to a model for a nonideal Gaussian chain that can be either swollen or collapsed. The data range for fitting was selected based on two criteria. One, that the fit excluded data points with large error bars. Two, to avoid fitting upturns in the low q region. Previous work on dilute polymer samples in our lab has found that upturn at low q is indicative of aggregate formation[1]. Because the model posed here only accounts for the polymer conformation, and not scattering contributions from aggregates, this upturn reigon was excluded. This model was implemented as described by Hammouda[2]. The form factor for this model is given by:

$$P_{\text{excluded volume}}(q) = 2\int_0^1 \mathrm{d}x(1-x)\exp\left[-\frac{1}{6}a^2N^{2\nu}\,q^2x^{2\nu}\right] \tag{B-1}$$

where a is the polymer chain statistical segment length, N is the degree of polymerization, q is the scattering vector, and ν is the excluded volume parameter, related to the Porod exponent m by the expression

$$\nu = \frac{1}{m} \tag{B-2}$$

The form of the model does not allow independent fitting of a and N on a single data set; thus, these parameters were lumped into a single term, the radius of gyration, R_g, defined in the following expression

$$P_{\text{excluded volume}}(q) = 2\int_0^1 \mathrm{d}x(1-x)\exp\left[-\frac{1}{6}\,(2\nu+1)(2\nu+2)R_g^2q^2x^{2\nu}\right] \tag{B-3}$$

where the integral is evaluated numerically using a global adaptive quadrature algorithm. The total scattering intensity is fit to the equation

$$I(q) = C_1 * P_{\text{excluded volume}}(q) + C_2 \quad \text{(B-4)}$$

where C_1 is the contrast factor and C_2 is the background. Thus, the fit parameters are R_g, m, C_1, and C_2.

B1.1.2 Borue-Erukhimovich model for weakly charged polyelectrolytes

SANS data on negatively charged ELPs was fit to the scattering model proposed by Borue and Erukhimovich for weakly charged polyelectrolytes in the semidilute regime[3]. The data range for fitting was selected as described above. In this case, low q upturn in SANS data taken on polyelectrolyte solutions has been reported in several instances in literature[4]. The structure factor for this model is given by

$$S(x) = C_1 \frac{x^2 + s}{(x^2 + t)(x^2 + s) + 1} + C_2 \quad \text{(B-5)}$$

where C_1 is the contrast factor which accounts for contrast and concentration, C_2 is background, and x is the scaled scattering vector, scaled by the characteristic length scale of screening in a saltless polyelectrolyte solution, r_0

$$r_0 = a\left(\frac{48\pi\ell_B}{am^2}\phi\right)^{-1/4} \quad \text{(B-6)}$$

where a is the statistical segment length in the polymer, l_B is the Bjerrum length, φ is the polymer volume fraction, and m is the linear charge density parameter, defined as

$$m = \frac{eN}{Q} \quad \text{(B-7)}$$

where e is electron charge, N is the degree of polymerization, and Q is the total macromolecule charge. Because a monomer unit is not well-defined in the ELPs (charges in the sequence are non-uniformly spaced along the backbone), N is poorly defined. The parameter s is the effective reduced charge concentration, given by

$$s = \kappa^2 r_0^2 \quad \text{(B-8)}$$

Where r_0 is given above, and κ^{-1} is the Debye length, defined by the equation below

$$\kappa^2 = 4\pi \frac{\ell_B}{a^3} \sum_i \left(z_i \phi_{s,i} + \frac{\phi}{m} \right) \quad \text{(B-9)}$$

where z_i is the valency of salt i, which has concentration $\varphi_{s,i}$. The parameter t represents the effective solvent quality, where a good solvent is indicated by $t > 0$ and a poor solvent is indicated by $t < 0$. The variable t is defined by the following equation

$$t = 12 \left(\frac{r_0}{a} \right)^2 \left[(1 - 2\chi) + \frac{3B_3}{a^3} \phi \right] \quad \text{(B-10)}$$

where χ is the Flory parameter, and B_3 is the third virial coefficient. To reduce the number of fit parameters in this model, several assumptions were made. First, a salt concentration of 0 ($\varphi_s = 0$) is assumed, as samples were dialyzed extensively against water prior to lyophilization, and were resuspended in pure deuterium oxide. A Bjerrum length of 0.704 nm was used for deuterium oxide [5]. The polymer volume fraction was estimated to be 0.0203, assuming an elastin density of 1.232 g/mL, which is based on literature reported values[6]. Finally, the third virial coefficient was assumed to be zero, as other interactions are expected to dominate the behavior in dilute solution. This assumption has been made in previous reports fitting to this model[4].

B1.2 Estimation of monomer-monomer spacing from linear charge density parameter

Recalling that $m = eN/Q$, and knowing that the total charge of both negatively charged ELPs is estimated at approximately -40 at pH 7, we can use the value of m in the fit to approximate the effective "degree of polymerization" in these ELPs. Using this route of estimation, we get degrees of polymerization of 20 for both negatively charged ELPs. The total number of amino acids per ELP is 508; thus, by this fit, there are 25.4 amino acids per "monomer" in both negatively charged ELPs. If the amino acids within each "monomer" unit assumed a Gaussian conformation ($R_g^2 = Na^2/6$), and assuming each amino acid is 0.40 nm in length[7], the radius of gyration of each "monomer" would be 0.82 nm. These R_g would yield monomer-monomer spacings 1.6 nm.

Table B-1. Summary of expression and purification conditions, and yields used for all proteins used in this paper.

Protein	Media Used	Lysozyme Added?	Freeze Thaw Cycles	Final Yield

uELP	TB	Yes	1	76 mg/L
nELP	TB	No	3	99 mg/L
nhELP	TB	No	3	83 mg/L
bELP	TB	Yes	1	130 mg/L
bhELP	TB	Yes	1	96 mg/L
uELP-mCherry	LB	Yes	1	98 mg/L
nELP-mCherry	LB	Yes	1	38 mg/L
nhELP-mCherry	LB	Yes	1	50 mg/L
bELP-mCherry	LB	Yes	1	43 mg/L
bhELP-mCherry	LB	Yes	1	33 mg/L

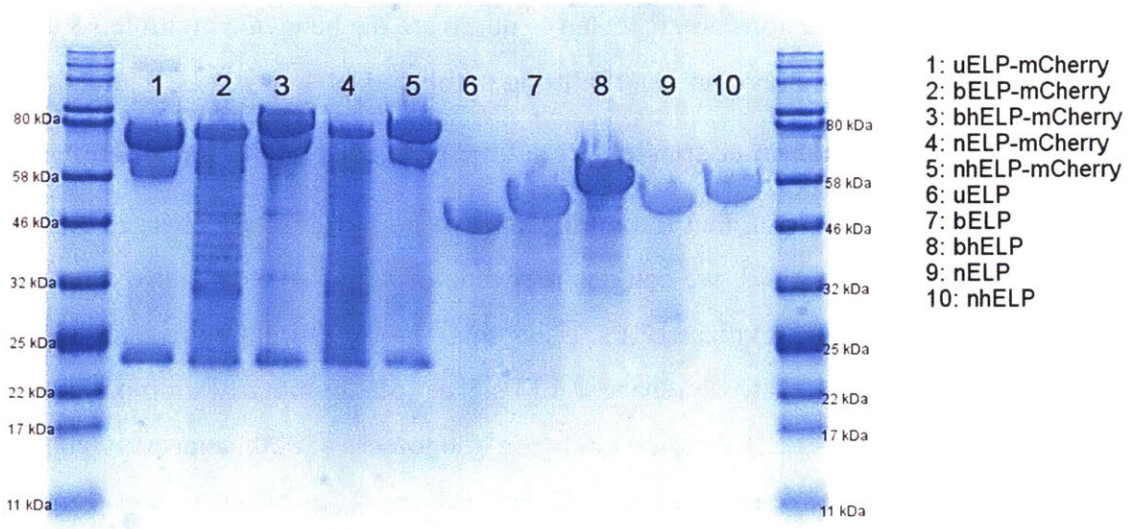


Figure B-1. SDS-PAGE of all proteins used in this study. Note that the appearance of three bands in lanes containing fusion proteins is due to partial hydrolysis of the acylimine bond in the mCherry chromophore[8-10].

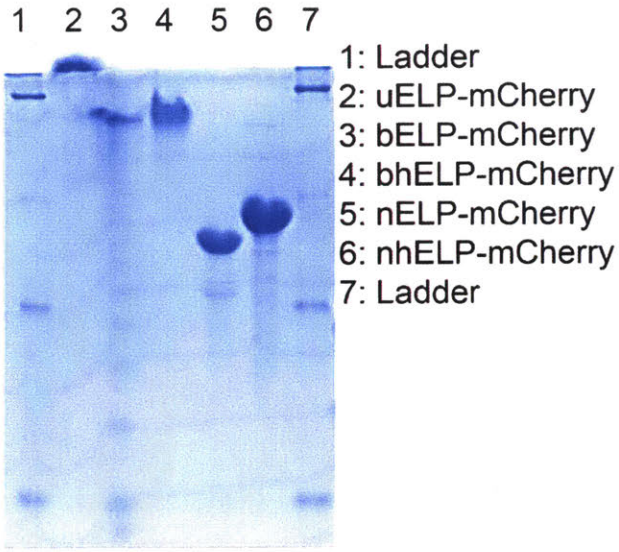


Figure B-2. Native PAGE of fusion proteins used in this study, evidencing the existence of a single species in these purified samples, and supporting the claim in **Figure B-1** that the observation of 3 bands by SDS-PAGE is due to partial cleavage of fusion proteins across the mCherry chromaphore.

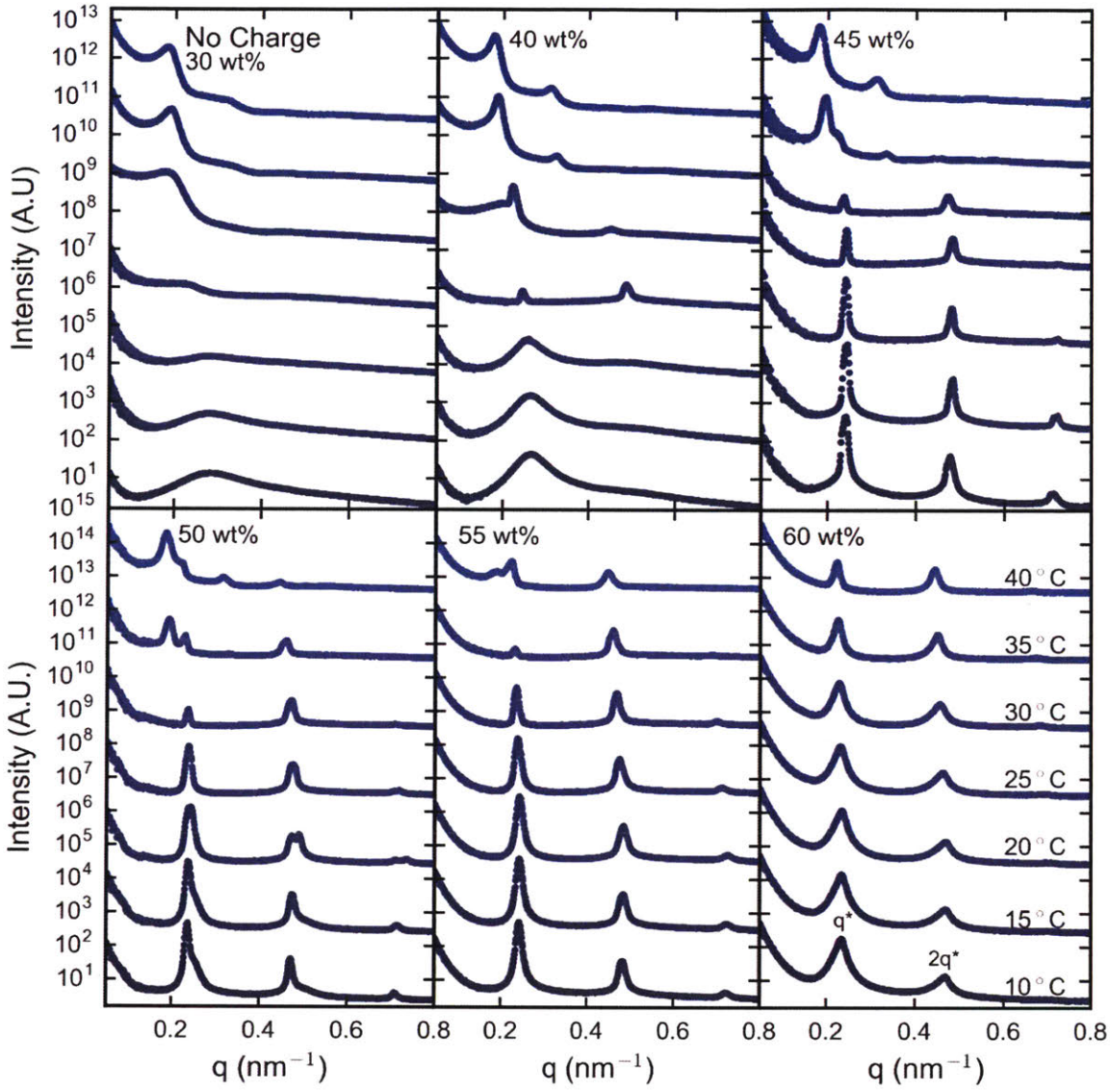


Figure B-3. Small-angle x-ray scattering as a function of temperature for uELP-mCherry from 30-60 wt% in water between 10 °C and 40 °C. Each panel is labeled with the relevant concentration and the curves are offset for clarity.

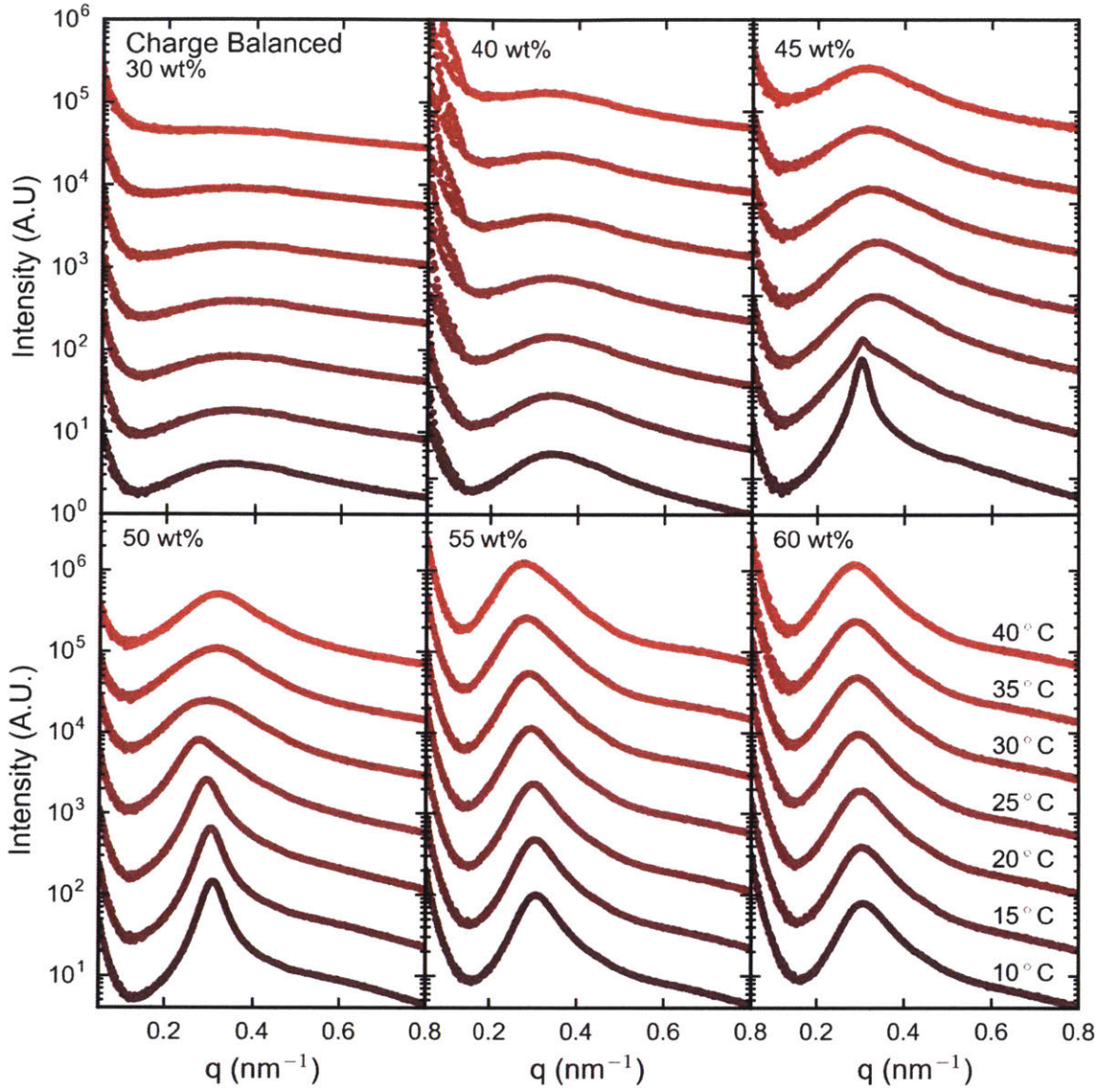


Figure B-4. Small-angle x-ray scattering as a function of temperature for bELP-mCherry from 30-60 wt% in water between 10 ˚C and 40 ˚C. Each panel is labeled with the relevant concentration and the curves are offset for clarity.

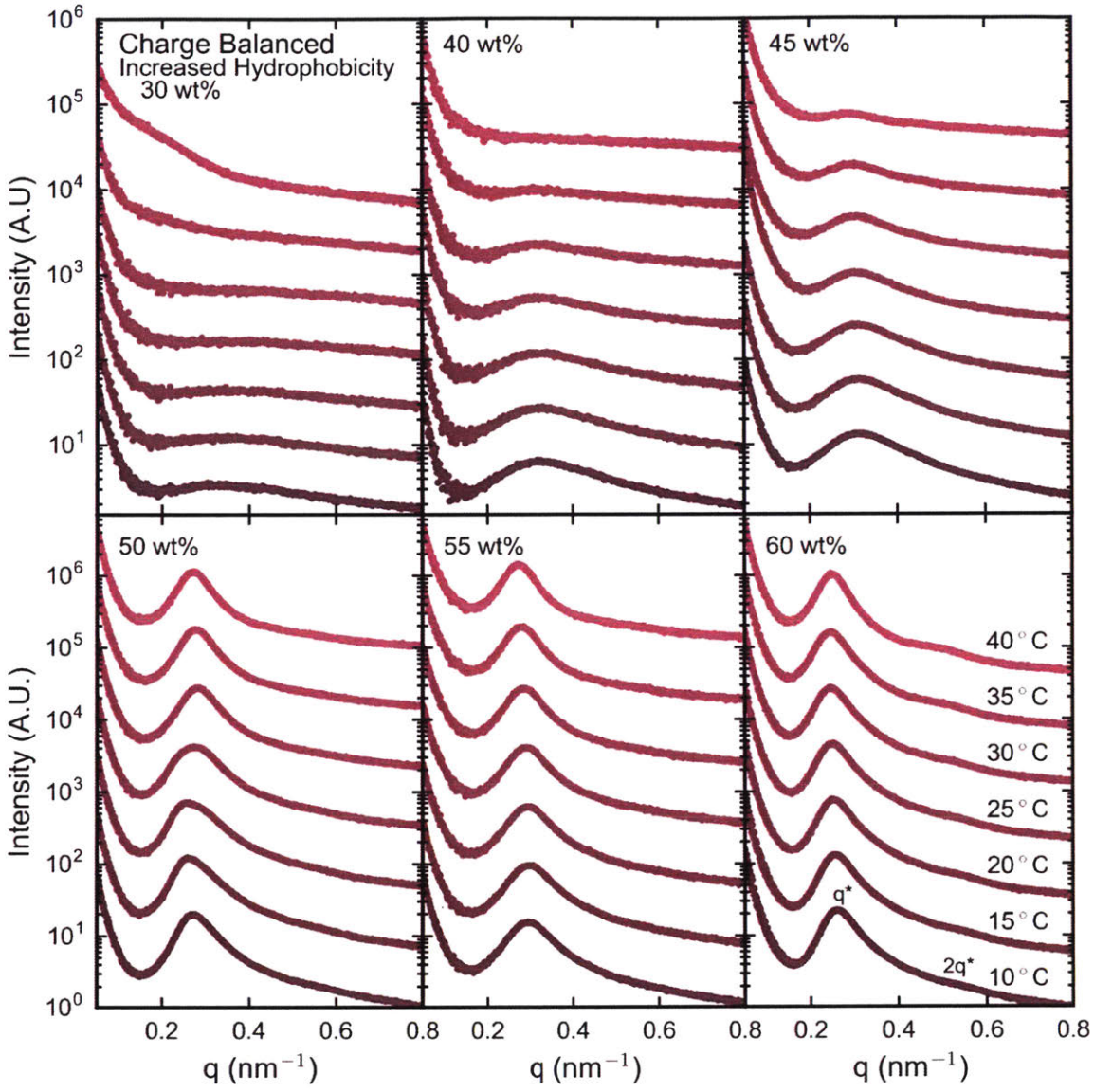


Figure B-5. Small-angle x-ray scattering as a function of temperature for bhELP-mCherry from 30-60 wt% in water between 10 °C and 40 °C. Each panel is labeled with the relevant concentration and the curves are offset for clarity.

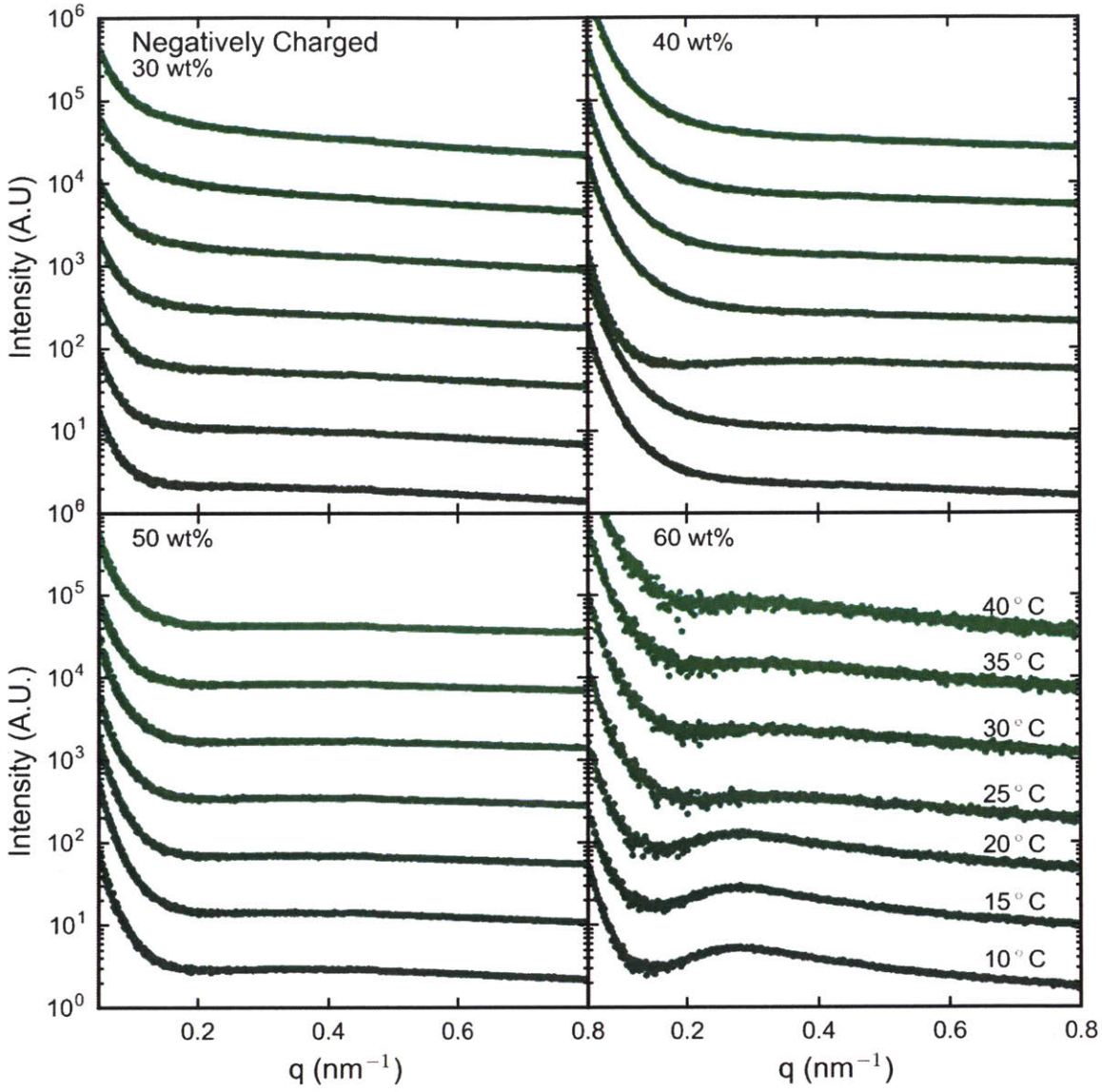


Figure B-6. Small-angle x-ray scattering as a function of temperature for nELP-mCherry from 30-60 wt% in water between 10 ˚C and 40 ˚C. Each panel is labeled with the relevant concentration and the curves are offset for clarity.

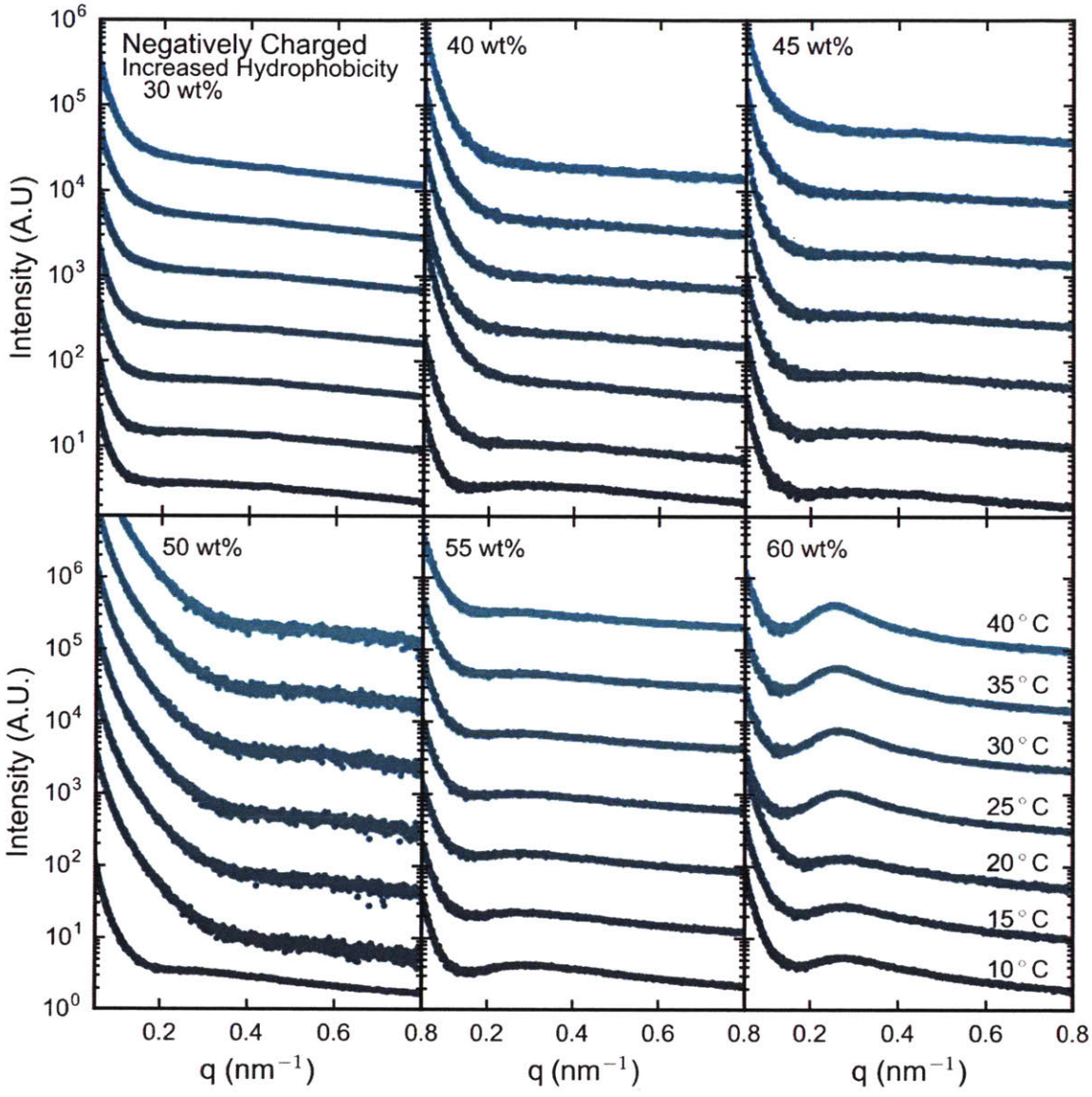


Figure B-7. Small-angle x-ray scattering as a function of temperature for nhELP-mCherry from 30-60 wt% in water between 10 °C and 40 °C . Each panel is labeled with the relevant concentration and the curves are offset for clarity.

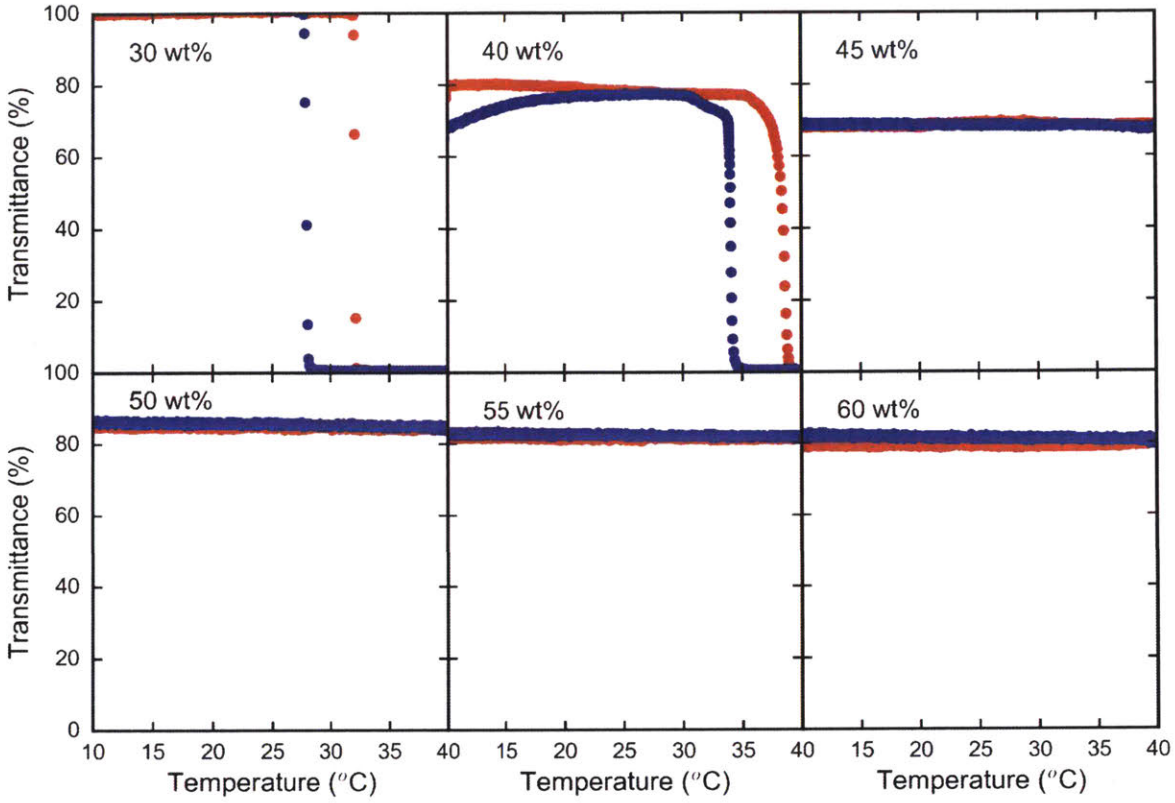


Figure B-8. Transmittance data of uELP-mCherry between 30-60 wt% in water from 10-40 ˚C. Each panel is labelled with the relevant concentration. Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

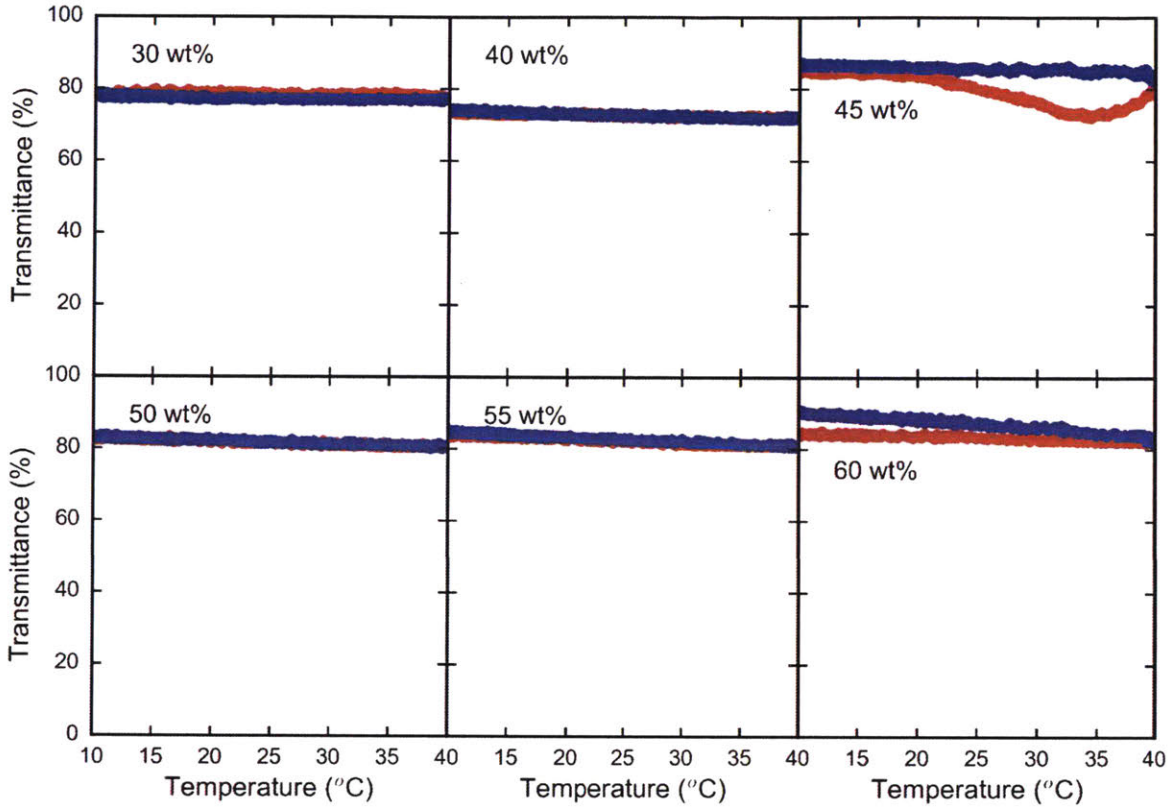


Figure B-9. Transmittance data of bELP-mCherry between 30-60 wt% in water from 10-40 °C. Each panel is labelled with the relevant concentration. Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

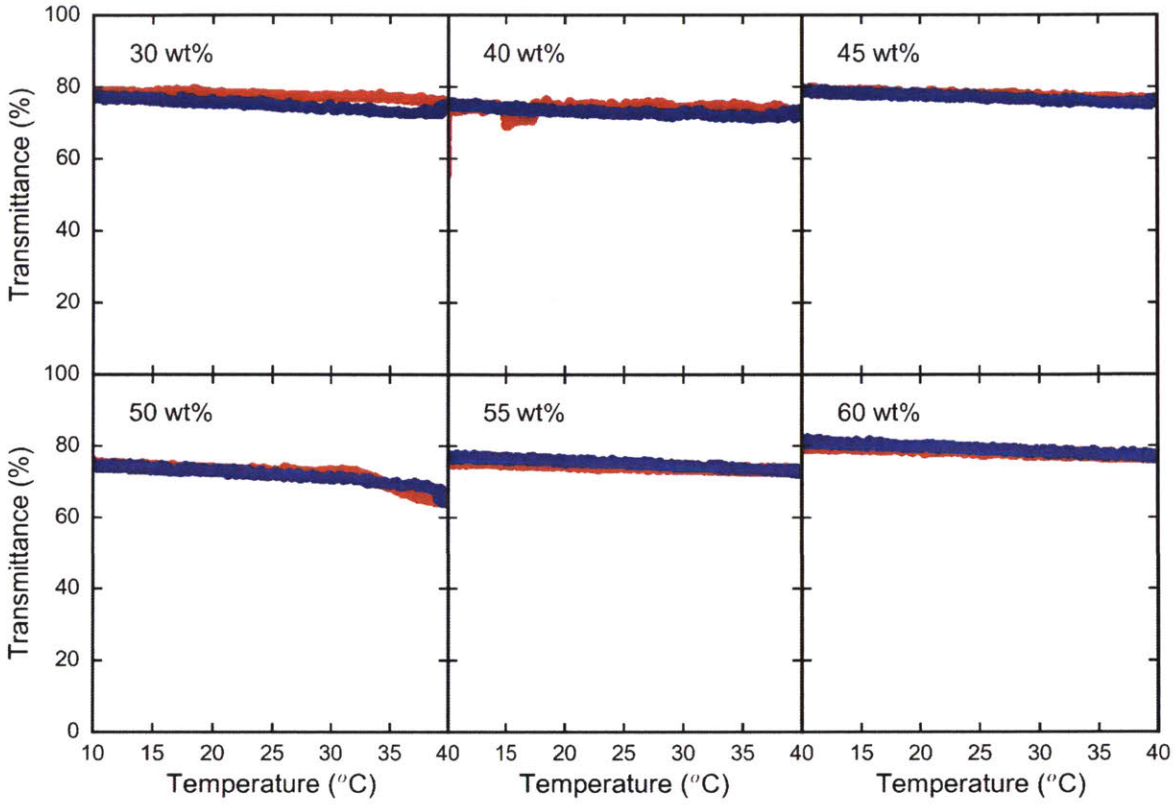


Figure B-10. Transmittance data of bhELP-mCherry between 30-60 wt% in water from 10-40 ˚C. Each panel is labelled with the relevant concentration. Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

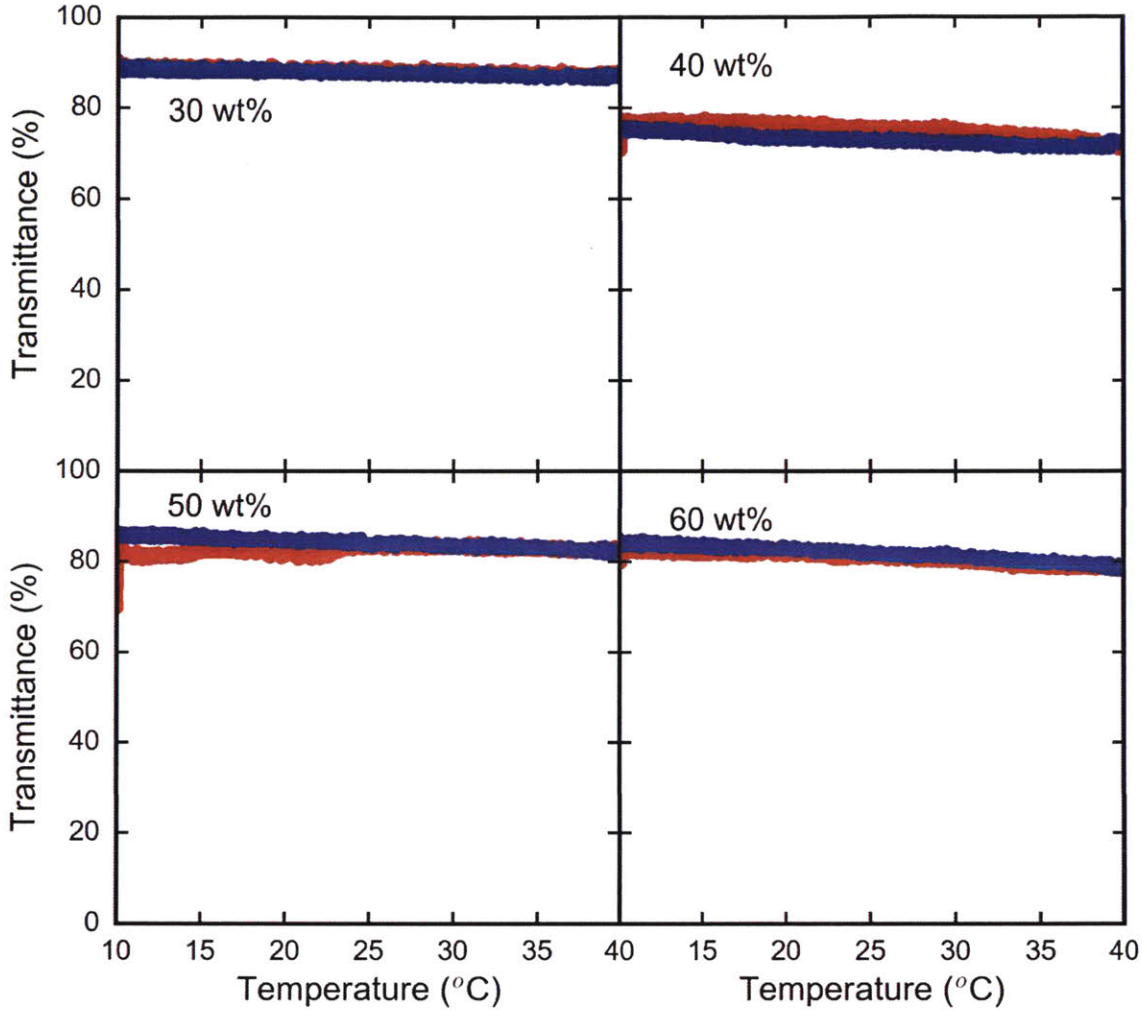


Figure B-11. Transmittance data of nELP-mCherry between 30-60 wt% in water from 10-40 ˚C. Each panel is labelled with the relevant concentration. Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

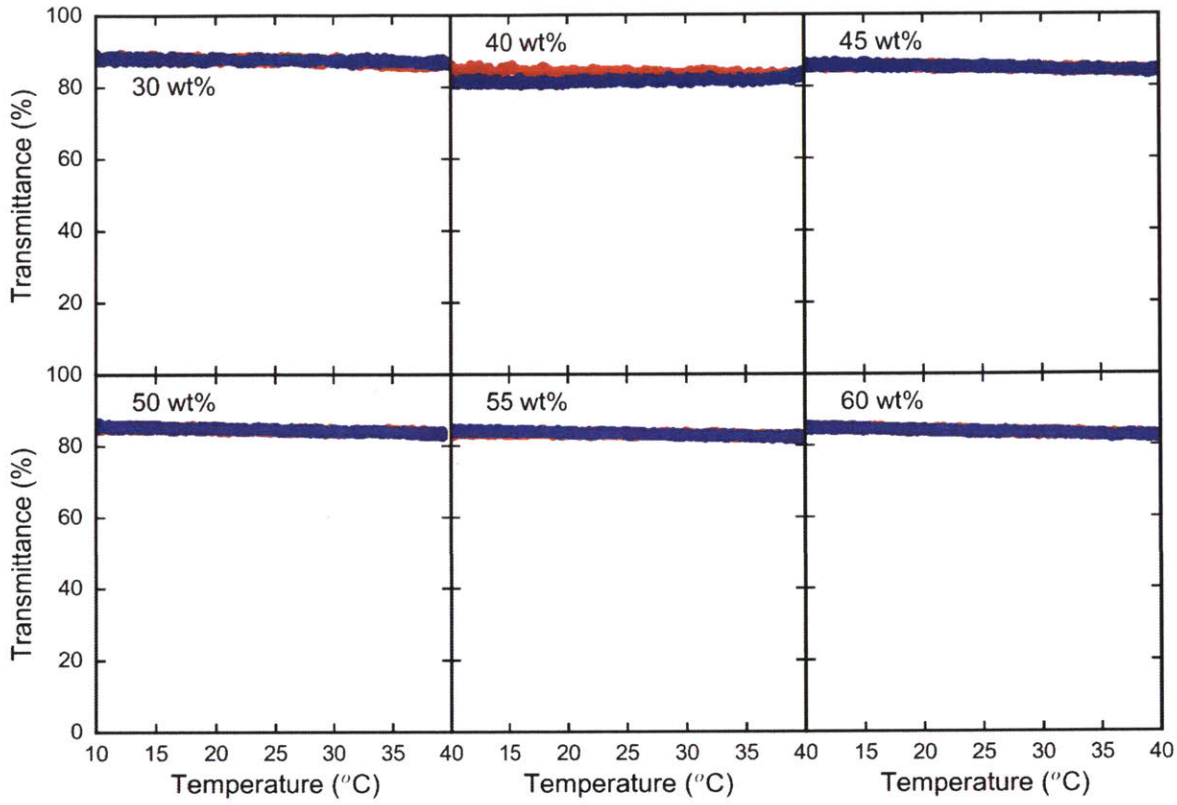


Figure B-12. Transmittance data of nhELP-mCherry between 30-60 wt% in water from 10-40 °C. Each panel is labelled with the relevant concentration. Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

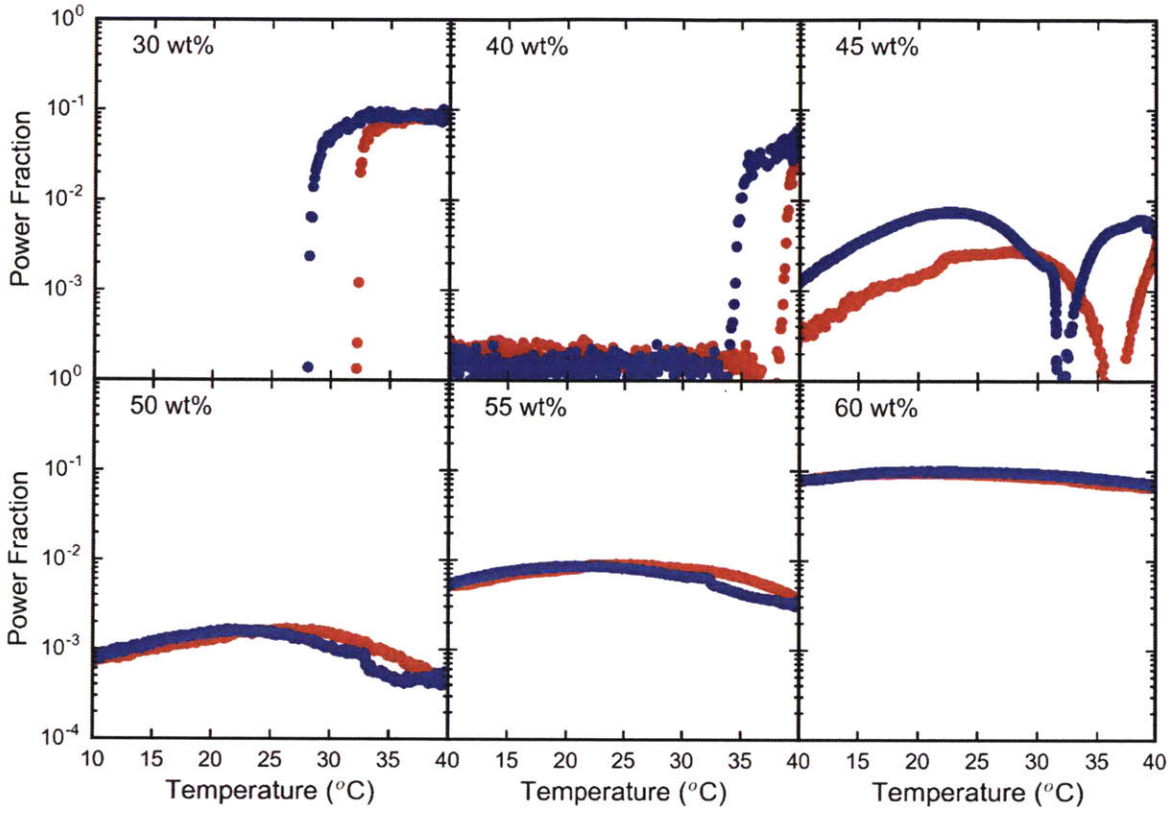


Figure B-13. Birenfringence of uELP-mCherry at concentrations between 30 and 60 wt% in water from 10 ˚C to 40 ˚C, demonstrating the low birefringence of this material at higher concentrations (45-60 wt%). Each panel is labelled with the relevant concentration. Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

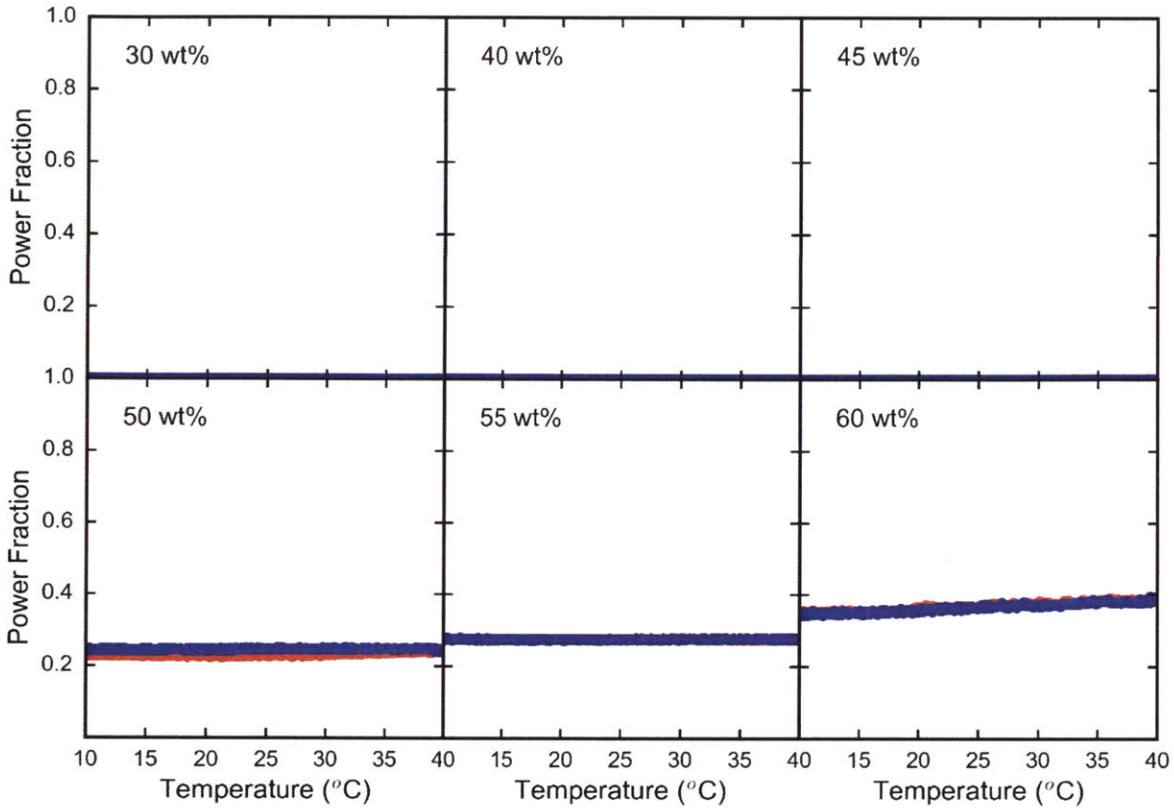


Figure B-14. Birenfringence of bELP-mCherry at concentrations between 30 and 60 wt% in water from 10 °C to 40 °C, demonstrating the high birefringence of this material at higher concentrations (50-60 wt%) in water. Each panel is labelled with the relevant concentration. Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

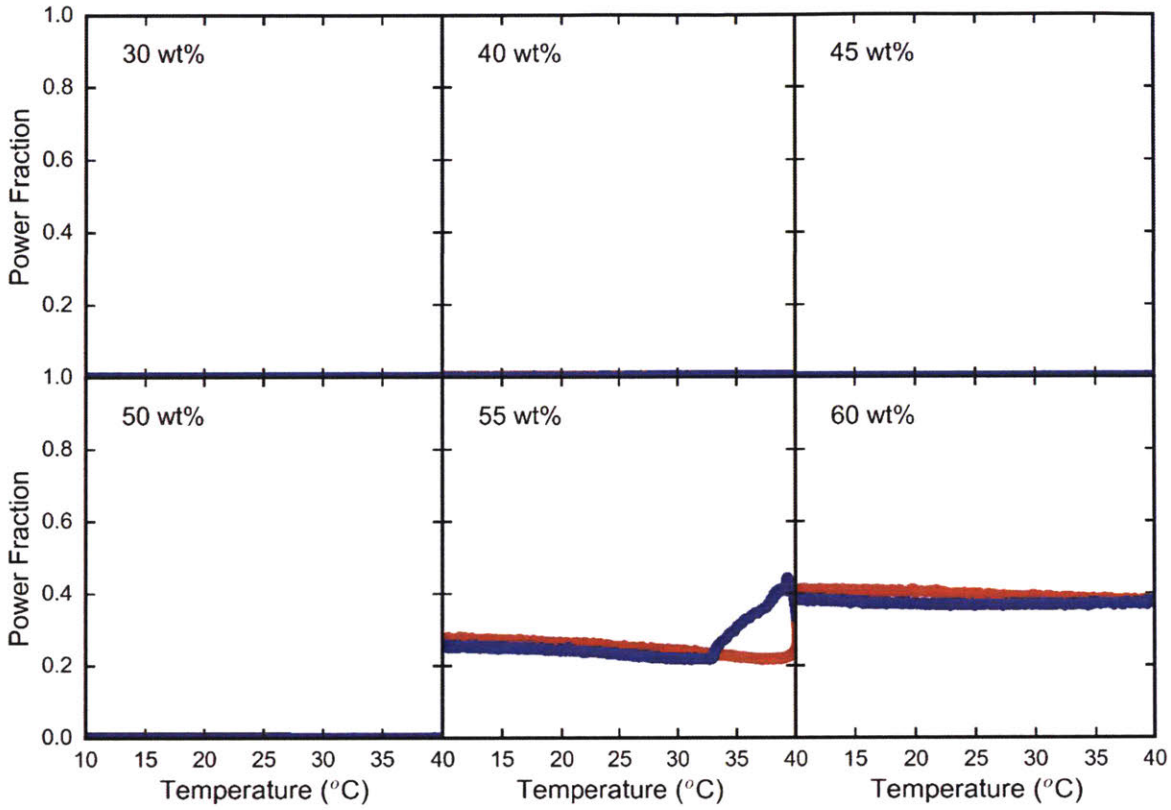


Figure B-15. Birenfringence of bhELP-mCherry at concentrations between 30 and 60 wt% in water from 10 ˚C to 40 ˚C, demonstrating the high birefringence of this material at higher concentrations (55-60 wt%) in water. Each panel is labelled with the relevant concentration. Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

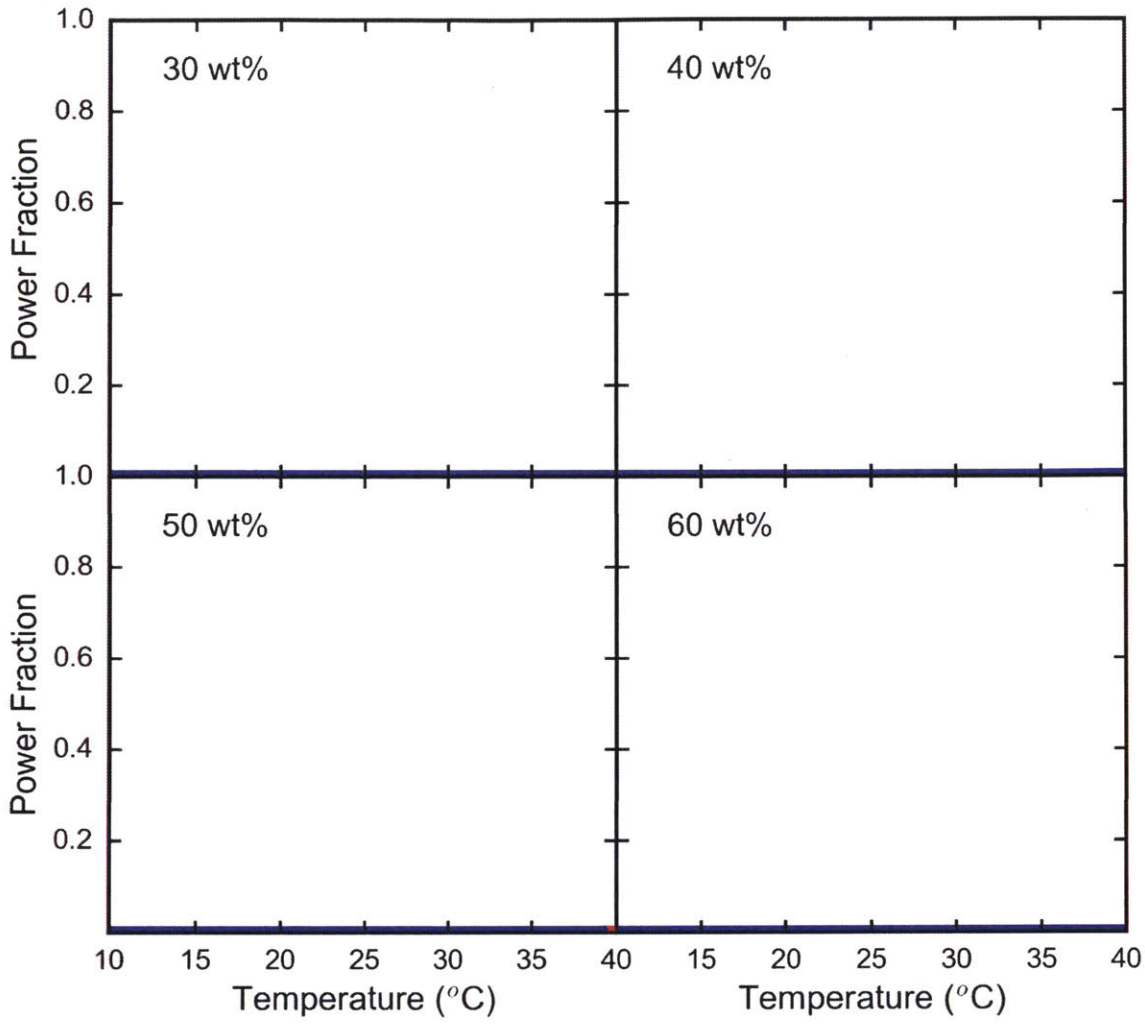


Figure B-16. Birenfringence of nELP-mCherry at concentrations between 30 and 60 wt% in water from 10 °C to 40 °C , demonstrating the absence of birefringence of this material at all concentrations. Each panel is labelled with the relevant concentration. Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

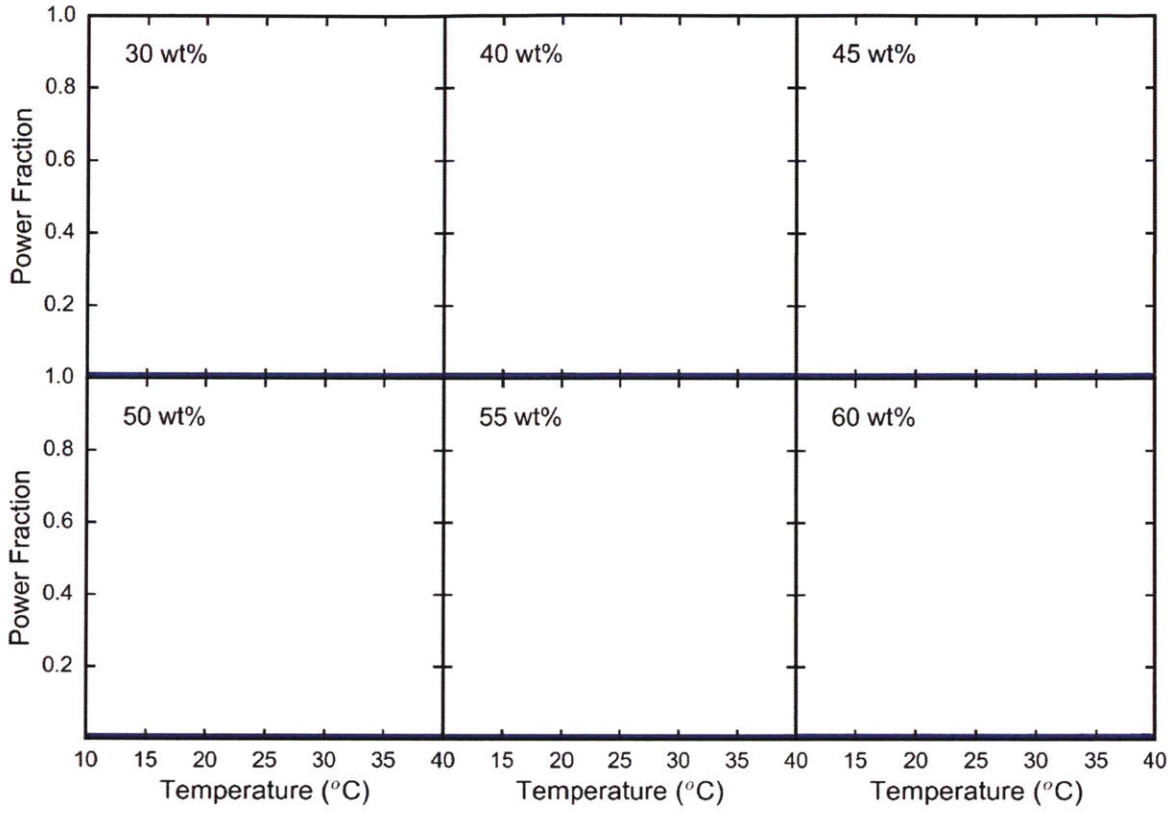


Figure B-17. Birenfringence of nhELP-mCherry at concentrations between 30 and 60 wt% in water from 10 ˚C to 40 ˚C, demonstrating the absence of birefringence of this material at all concentrations. Each panel is labelled with the relevant concentration. Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

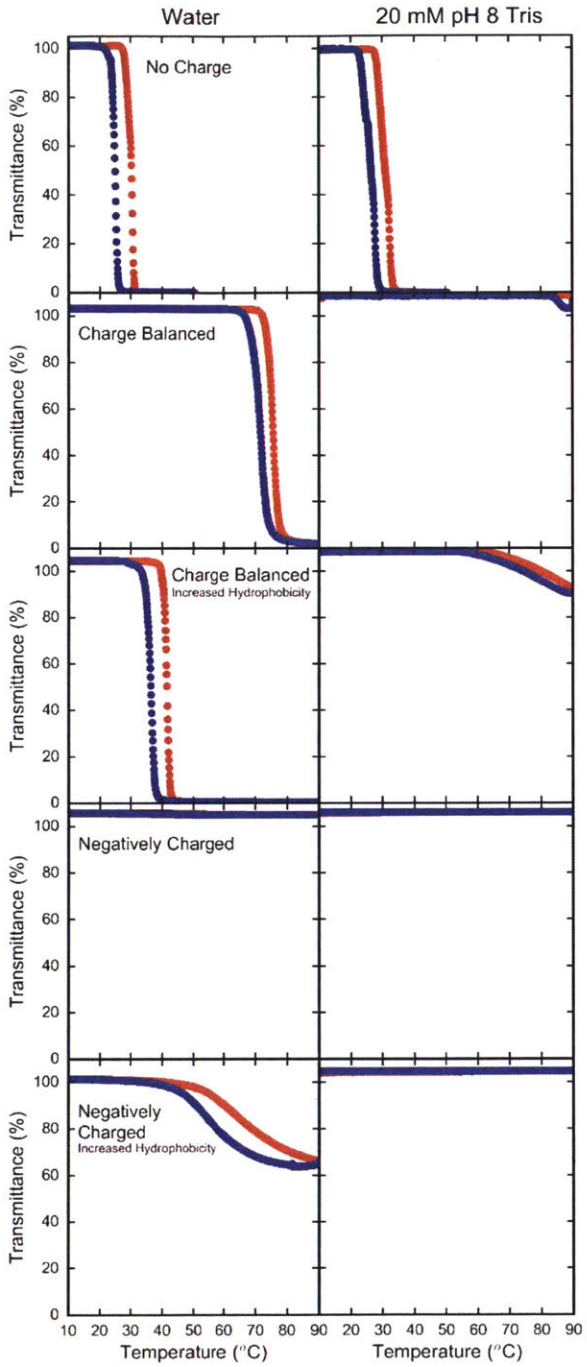


Figure B-18. Dilute solution (1 mg/mL) turbidimetry measurements on the different ELPs used in this study (uELP, bELP, bhELP, nELP, nhELP). Red data points correspond to the heating portion of the experiment, while blue data points correspond to the cooling portion of the experiment.

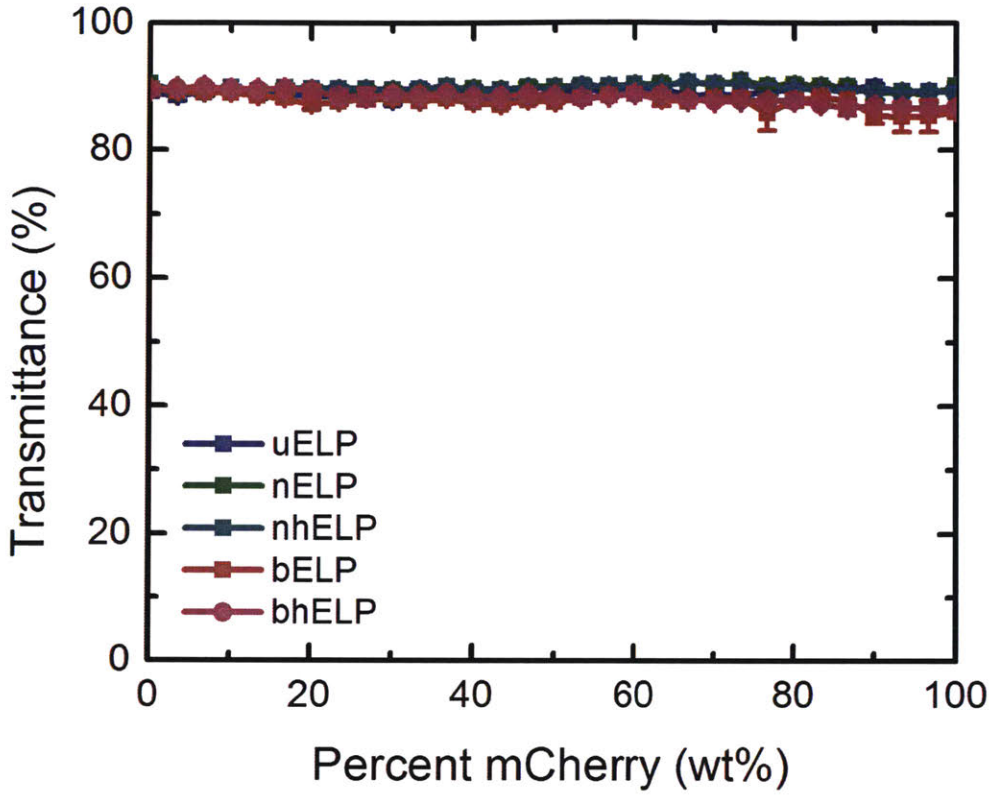


Figure B-19. Turbidimetry measurements on ELP/mCherry blends in 20 mM tris at pH 8.

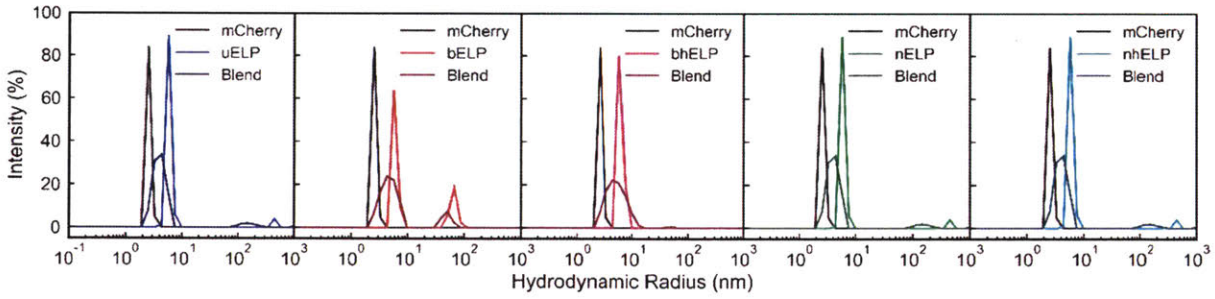


Figure B-20. Representative fits to DLS data on individual ELPs and mCherry as well as ELP/mCherry blends, using the CONTIN algorithm as implemented in the Wyatt Dynamics software.

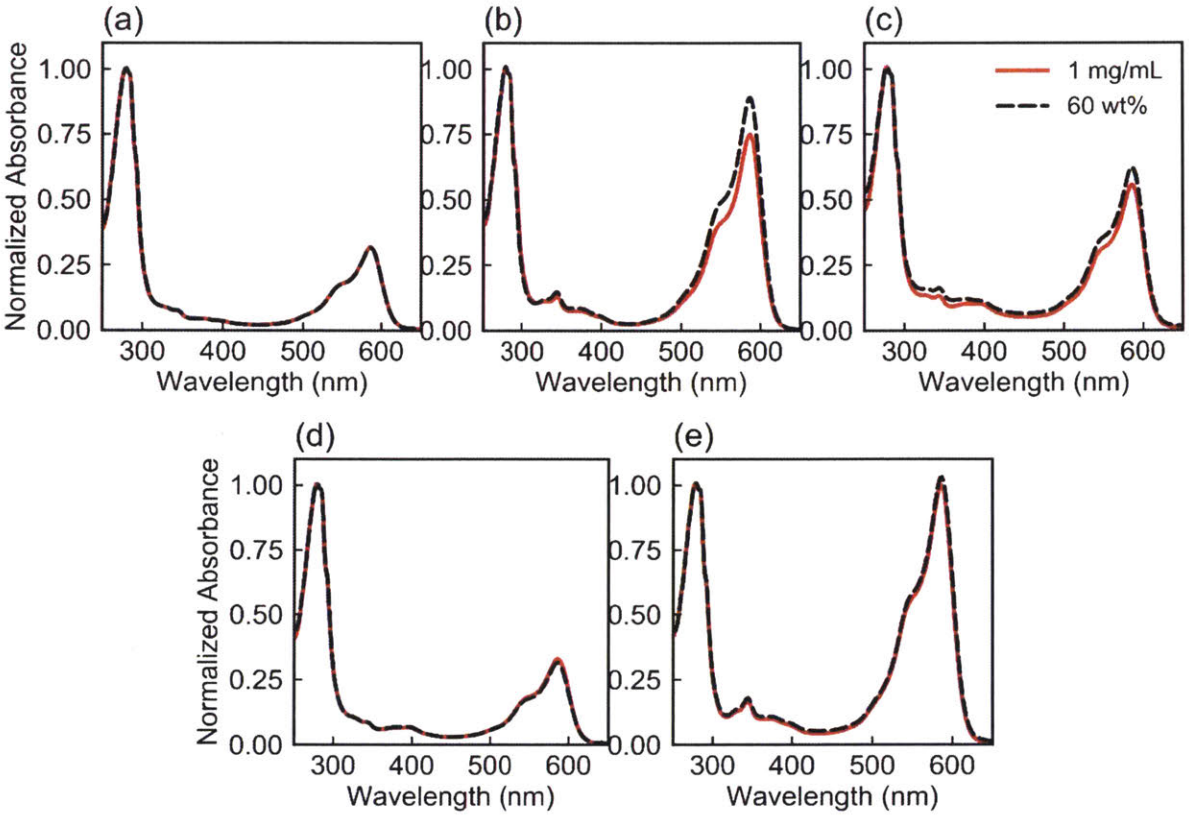


Figure B-21. Normalized absorbance spectra of (a) uELP-mCherry, (b) bELP-mCherry, (c) bhELP-mCherry, (d) nELP-mCherry, and (e) nhELP-mCherry comparing solutions that were prepared at 1 mg/mL initially (labeled 1 mg/mL) to solutions incubated at 60 wt% for 24 hours prior to being diluted down to the same 1 mg/mL concentration (labeled 60 wt%).

B1.3 References

1. Lam, C. N.; Chang, D.; Wang, M.; Chen, W.-R.; Olsen, B. D., The shape of protein–polymer conjugates in dilute solution. *J. Polym. Sci., Part A: Polym. Chem.* **2016,** *54* (2), 292-302.

2. Hammouda, B., SANS from homogeneous polymer mixtures: A unified overview. In *Polymer Characteristics*, Springer Berlin Heidelberg: Berlin, Heidelberg, 1993; pp 87-133.

3. Borue, V. Y.; Erukhimovich, I. Y., A statistical theory of weakly charged polyelectrolytes: fluctuations, equation of state and microphase separation. *Macromolecules* **1988,** *21* (11), 3240-3249.

4. Shibayama, M.; Tanaka, T., Small-angle neutron scattering study on weakly charged poly(N-isopropyl acrylamide-co-acrylic acid) copolymer solutions. *The Journal of Chemical Physics* **1995,** *102* (23), 9392-9400.

5. Wyman, J.; Ingalls, E. N., The Dielectric Constant of Deuterium Oxide. *J. Am. Chem. Soc.* **1938,** *60* (5), 1182-1184.

6. Scandola, M.; Pezzin, G., Density of Elastin-Water System. In *Water in Polymers*, AMERICAN CHEMICAL SOCIETY: 1980; Vol. 127, pp 225-234.

7. Ainavarapu, S. R. K.; Brujić, J.; Huang, H. H.; Wiita, A. P.; Lu, H.; Li, L.; Walther, K. A.; Carrion-Vazquez, M.; Li, H.; Fernandez, J. M., Contour Length and Refolding Rate of a Small Protein Controlled by Engineered Disulfide Bonds. *Biophys. J.* **2007,** *92* (1), 225-233.

8. Chung, J. A.; Wollack, J. W.; Hovlid, M. L.; Okesli, A.; Chen, Y.; Mueller, J. D.; Distefano, M. D.; Taton, T. A., Purification of prenylated proteins by affinity chromatography on cyclodextrin-modified agarose. *Anal. Biochem.* **2009,** *386* (1), 1-8.

9. Lee, B. C.; Zuckermann, R. N., Templated display of biomolecules and inorganic nanoparticles by metal ion-induced peptide nanofibers. *Chem. Commun.* **2010,** *46* (10), 1634-1636.

10. Shu, X. K.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J., Novel chromophores and buried charges control color in mFruits. *Biochemistry* **2006,** *45* (32), 9639-9647.

B2. MATLAB Code Used in SANS Fits

B2.1 Description

Fits to small-angle neutron scattering (SANS) data were performed in MATLAB. Fits were performed using the lsqnonlin() function in MatLab, part of the optimization toolbox. The non-linear least-squares algorithm was used to minimize the difference in the data and the predicted value divided by the standard deviation in the data. In the functions outlined below, any function name starting with "P*.m" denotes a form factor, and a function starting with "S*.m" denotes a structure factor. Functions named "*_ttlI.m" indicate calculation of total scattering function (including a combination of form factors). Functions titled "*_dy.m" were used to calculate the difference between the model prediction and the data. Functions titled "*_main.m" use the lsqnonlin() function to perform the fit. Functions titled "*_wrapper.m" execute the fit, make plots, and export the final fit results into a text file. These codes were based off those used by Dr. Christopher Lam for similar SANS fits. Included here is code used to attempt to fit SANS data taken on ELP-mCherry fusion proteins, assuming a worm-like chain. The form factor used in this analysis is based off a published model.[1] While this fit did not converge for the fusion proteins studied here, the code is included for completeness and to enable potential future work. Note that this fit in particular is quite slow as the code provided uses a triple-nested for loop to calculate the

solid angle integral involved in the calculation. Future use of this model would benefit from numerical optimization of this form factor calculation.

B2.2 Code for Swollen/Collapsed Gaussian Polymer Chain

B2.2.1 Pexcludedvol.m

```
function y = Pexcludedvol(q, Rg, m)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: June 20, 2017
% Updated: June 20, 2017
%
% This function calculates the structure factor of an excluded volume
% polymer
%
% Variables:
% INPUT:
%       q           scattering vector, array, 1 x length(q)      [=] 1/nm
%       Rg          radius of gyration                           [=] nm
%       m           Porod exponent                               [=] Unitless
%       Note: Rg^2 = a^2*N^(2/m)/((2/m+1)(2/m+2)), where a is statistical
%       segment length and N is degree of polymerization
%
% OUTPUTS:
%       y           Structure factor, array, 1 x length(q)    [=] unitless
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Calculate parameters
nu2 = 2/m;

% Define the function to be integrated

% Form factor calculation
y = zeros(length(q),1);

for i = 1:length(q)
    qi = q(i);
    fun = @(x)(2*(1-x).*exp(-qi.^2.*x.^nu2*((nu2+1)*(nu2+2)*Rg^2/6)));
    y(i) = integral(fun,0,1);
end

end
```

B2.2.2 Pexcludedvol_ttlI.m

```
function [Itot] = Pexcludedvol_ttlI(q,x)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: June 20, 2017
% Updated: June 20, 2017
%
% The objective of this function is to calculate the total I given a q and
% set of parameters, x
%
% Variables:
```

```
% INPUT:
%       x           Array of fit parameters:
%                       - x(1): Prefactor (scale * drho^2)      [=] 1/cm
%                       - x(2): Rg, radius of gyration          [=] nm
%                       - x(3): m, Porod exponent
%                       - x(4): background                      [=] 1/cm
%       q           Array of q values, length(q) x 1            [=] 1/nm
%
% OUTPUTS:
%       Itot        Array of the function intensity at all q    [=] 1/cm
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Itot = x(1).*Pexcludedvol(q,x(2),x(3)) + x(4); % [=] 1/cm

end
```

B2.2.3 Pexcludedvol_dy.m

```
function dy = Pexcludedvol_dy(x,q,I_data, I_std)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: June 20, 2017
% Updated: June 20, 2017
%
% The objective of this function is to calculate the difference between the
% scattering model I(q) and the data I(q), divided by the data standard
% deviation.
%
% Variables:
% INPUT:
% Variables:
% INPUT:
%       x           Array of fit parameters:
%                       - x(1): Prefactor (scale * drho^2)      [=] 1/cm
%                       - x(2): Rg, radius of gyration          [=] nm
%                       - x(3): m, Porod exponent
%                       - x(4): background                      [=] 1/cm
%       q           Array of q values, length(q) x 1            [=] 1/nm
%       I_data      Array of data I(q), length(q) x 1                   [=]
1/cm
%       I_std       Array of standard deviation data for I(q), length(q)x1
[=] 1/cm
%
% OUTPUTS:
%       dy          Array of the difference between the scattering model
%                   I(q) and the data I(q), divided by the data's standard
%                   deviation.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Calculate the total scattering intensity of the model
I_pred = Pexcludedvol_ttlI(q,x);

% Plot data and scattering intensity from fit parameters
% figure; hold off;
% errorbar(q,I_data,I_std,'ro');
% set(gca,'yscale','log');
% hold on;
```

```
% semilogy(q,I_pred,'b-');
% axis tight;
% pause(0.1);

% Calculate the difference divided by the standard deviation in the
% intensity
dy = (I_data - I_pred)./I_std;

end
```

B2.2.4 fit_Pexcludedvol_main.m

```
function [q,I_data,I_std,I_pred,Pf,Pf_err,chisq,chisqr] =
fit_Pexcludedvol_main(file_ID,P1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: June 20, 2017
% Updated: June 20, 2017
%
% The objective of this function is to fit the form factor of an excluded
% volume polymer
%
% Variables
% INPUTS:
%       file_ID         Name of a csv file containing data of the form:
%                           - column 1: q data
%                           - column 2: I data
%                           - column 3: I standard deviation data
%       P1              Vector of initial parameter guesses:
%                       - P1(1): Prefactor (scale * drho^2)         [=] 1/cm
%                       - P1(2): Rg, radius of gyration             [=] nm
%                       - P1(3): m, Porod exponent
%                       - P1(4): background                         [=] 1/cm
%
% OUTPUTS:
%       q               column vector of q data points from the input file
%       I_data          column vector of I data points from the input file
%       I_std           column vector of I std dev points from the input
%                       file
%       I_pred          column vector of I points predicted by the model
%                       fit
%       Pf              Vector of optimized parameters (same order as
%                       initial guess)
%       Pf_err          Error in optimized parameter fits corresponding to
%                       a 95% confidence interval
%       chisq           Chi squared value for the fit.
%       chisqr          Reduced chi squared value for the fit
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Load data from csv file
data = csvread(file_ID);

% Assign the data to separate arrays
q = data(:,1);
I_data = data(:,2);
I_std = data(:,3);
```

```
% Set fit options
options = optimset('MaxFunEvals',inf,'TolFun',1e-10);

% Set upper and lower bounds for the fit parameters
lb = [0;0;1.6;-1];
ub = [1000;500;10;10000];

% Fit the data
[Pf,~,residual,exitflag,~,~,jacobian] =
lsqnonlin('Pexcludedvol_dy',P1,lb,ub,options,q,I_data,I_std);

% Calculate the error in the fit using Matlab functions for 95% confidence
% interval
ci = nlparci(Pf,residual,'jacobian',jacobian);
Pf_err = Pf - ci(:,1);

% Calculate model I(q) over the q data range from the fit parameters
I_pred = Pexcludedvol_ttlI(q,Pf);

% Calculate the chi^2 and reduced chi^2
chisq = ((I_data- I_pred)./I_std).^2;
chisq = sum(chisq);
chisqr = chisq/(length(q)-length(Pf));

end
```

B2.2.5 fit_Pexcludedvol_wrapper.m

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: June 20, 2017
% Updated: June 20, 2017
%
% The purpose of this script is to perform a fit and make a plot to compare
% the fit to the original data.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% I. PERFORM THE FIT
file_ID = 'ELP0-D100-25mgmL-bkgsub-trim.csv';     % file name
P1 = [0.82;5.9;1.5;0.004];              % initial guesses

[q,I_data,I_std,I_pred,Pf,Pf_err,chisq,chisqr] =
fit_Pexcludedvol_main(file_ID,P1);

% Calculate the separate functions
I_ig = Pexcludedvol_ttlI(q,P1);

%% II. PLOT THE RESULTS
figure(1); clf
subplot(1,2,1);
errorbar(q,I_data,I_std,'bo');
hold on;
set(gca,'xscale','log','yscale','log');
loglog(q,I_pred,'k','linewidth',2);
xlabel('q (nm$^{-1}$)','interpreter','latex','fontsize',12);
```

```
ylabel('I(q) (cm$^{-1}$)','interpreter','latex','fontsize',12);
h_legend= legend('data','model');
xlim([0.1 2])
ylim([.002 1])
set(h_legend,'interpreter','latex','fontsize',10);
% axis square;
title('After Fit');

subplot(1,2,2);
errorbar(q,I_data,I_std,'bo');
hold on;
set(gca,'xscale','log','yscale','log');
loglog(q,I_ig,'k','linewidth',2);
xlabel('q (nm$^{-1}$)','interpreter','latex','fontsize',12);
ylabel('I(q) (cm$^{-1}$)','interpreter','latex','fontsize',12);
h_legend= legend('data','model');
set(h_legend,'interpreter','latex','fontsize',10);
% axis square;
title('Initial Guess');

%% III. WRITE THE RESULTS TO A TEXT FILE
filename_str = 'ELP0-trim_Pexcludedvol.txt';
fileID = fopen(filename_str,'wt');
fprintf(fileID,datestr(clock));
fprintf(fileID,'\n');
fprintf(fileID,'============================');
fprintf(fileID,'\n');
fprintf(fileID,'Model: Excluded volume polymer');
fprintf(fileID,'\n');
fprintf(fileID,'============================');
fprintf(fileID,'\n');
fprintf(fileID,'-----Input Parameters-----\n');
%write input parameters to file
fprintf(fileID,'%25s','P1 = ');
fprintf(fileID,'%15.5f',P1');
fprintf(fileID,'\n');
fprintf(fileID,'-----Results-----\n');
%write fit parameters to file
fprintf(fileID,'chi squared = %10.5f\n',chisq);
fprintf(fileID,'reduced chi squared = %10.5f\n',chisqr);
fprintf(fileID,'%25s','CF = ');
fprintf(fileID,'%10.4e +/- %10.4e',Pf(1),Pf_err(1));
fprintf(fileID,'\n');
fprintf(fileID,'%25s ','Radius of gyration = ');
fprintf(fileID,'%10.5f +/- %10.5f',Pf(2), Pf_err(2));
fprintf(fileID,'\n');
fprintf(fileID,'%25s ','Porod exponent = ');
fprintf(fileID,'%10.5f +/- %10.5f',Pf(3), Pf_err(3));
fprintf(fileID,'\n');
fprintf(fileID,'%25s','Bkg = ');
fprintf(fileID,'%10.6f +/- %10.6f',Pf(4),Pf_err(4));
fprintf(fileID,'\n');
fprintf(fileID,'\n');
fprintf(fileID,'\n');
matrixwrite = [q I_data I_std I_pred]';
```

```
fprintf(fileID,'%10s\t %10s\t %10s\t %10s\n','q (A^-1)', 'I(Q)', 'I(Q)
std.dev.', 'I(Q)_pred');
fprintf(fileID,'----------------------------------------------------------------
----------------------------------------------------------------');
fprintf(fileID,'\n');
fprintf(fileID,'%15.7f \t %15.5f \t %15.5f \t %15.5f\n',matrixwrite);
fclose(fileID);
% Write a data only file as well
filename_str = 'DATAONLY-ELP0-trim-Pexcludedvol.txt';
fileID = fopen(filename_str,'wt');
matrixwrite = [q I_data I_std I_pred]';
fprintf(fileID,'%10s\t %10s\t %10s\t %10s\n','q (A^-1)', 'I(Q)', 'I(Q)
std.dev.', 'I(Q)_pred');
fprintf(fileID,'\n');
fprintf(fileID,'%15.7f \t %15.5f \t %15.5f \t %15.5f\n',matrixwrite);
fclose(fileID);
```

B2.3 Code for Borue-Erukhimovich Model for Weakly-Charged Polyelectrolytes

B2.3.1 SBEPEmbspB.m

```
function y = SBEPEmbspB(q, a, alpha, chi)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: June 28, 2017
% Updated: June 28, 2017
%
% This function calculates the structure factor of a Borue-Erukhimovich
% polyelectrolyte.
%
% Variables:
% INPUT:
%       q           scattering vector, array, 1 x length(q)     [=]
1/angstroms
%       a           Segment length, scalar                      [=] angstroms
%       alpha       Degree of ionization, scalar                [=] fraction
%       chi         Flory interaction parameter                 [=] unitless
%
% OUTPUTS:
%       y           Structure factor, array, 1 x length(q)    [=] unitless
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Calculate parameters
zi = 1;  % Assume monovalent salt
LB = 0.704; % [=] nm, Bjerrum length of D2O at 10C
phiS = 0;   % Salt concentration is zero
phi = .020292; % Polymer volume fraction estimated from ELP density of 1.232
g/mL
B3 = 0;         % Set third virial coefficient to zero for fitting
tractability

% Debye length
kappa2 = 4*pi*(LB/alpha^3)*(phiS*zi + alpha*phi);       % [=] 1/angstroms^2

% Characteristic scale of screening in a saltless polyelectrolyte solution
r0 = a*(48*pi*LB*phi*alpha^2/a)^(-0.25);        % [=] angstroms
```

```
% Reduced temperature, representative of solvent quality (solven tquality
% is good when t > 0 and poor when t < 0
t = 12*((r0/a)^2)*(1-2*chi+3*B3*phi/a^3);       % [=] unitless

% Reduced charge concentration
s = kappa2*r0^2;    % [=] unitless

% Reduced scattering vector
x = r0.*q;       % [=] unitless

% Form factor calculation
y = (x.^2 + s)./((x.^2 + t).*(x.^2 + s) + 1);

end
```

B2.3.2 BEPEmbspB_ttlI.m

```
function [Itot] = BEPEmbspB_ttlI(q,x)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: June 28, 2017
% Updated: June 28, 2017
%
% The objective of this function is to calculate the total I given a q and
% set of parameters, x
%
% Variables:
% INPUT:
%       x           Array of fit parameters:
%                       - x(1): Prefactor (scale * drho^2)              [=]
1/cm
%                       - x(2): a, Segment length, scalar               [=]
angstroms
%                       - x(3): alpha, Degree of ionization, scalar     [=]
fraction
%                       - x(4): chi, Flory interaction parameter        [=]
unitless
%                       - x(5): Background                             [=]
1/cm
%       q           Array of q values, length(q) x 1                    [=]
1/angstroms
%
% OUTPUTS:
%       Itot        Array of the function intensity at all q [=] 1/cm
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Itot = x(1).*SBEPEmbspB(q,x(2),1./x(3),x(4))+x(5); % [=] 1/cm

end
```

B2.3.3 BEPEmbspB_dy.m

```
function dy = BEPEmbspB_dy(x,q,I_data, I_std)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: June 28, 2017
% Updated: June 28, 2017
```

```
%
% The objective of this function is to calculate the difference between the
% scattering model I(q) and the data I(q), divided by the data standard
% deviation.
%
% Variables:
% INPUT:
%       x           Array of fit parameters:
%                       - x(1):  Prefactor (scale * drho^2)             [=]
1/cm
%                       - x(2):  a, Segment length, scalar              [=]
angstroms
%                       - x(3):  alpha, Degree of ionization, scalar    [=]
fraction
%                       - x(4):  chi, Flory interaction parameter       [=]
unitless
%                       - x(5): Background                             [=]
1/cm
%       q           Array of q values, length(q) x 1                    [=]
1/angstroms
%       I_data      Array of data I(q), length(q) x 1                   [=]
1/cm
%       I_std       Array of standard deviation data for I(q), length(q)x1
[=] 1/cm
%
% OUTPUTS:
%       dy          Array of the difference between the scattering model
%                   I(q) and the data I(q), divided by the data's standard
%                   deviation.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Calculate the total scattering intensity of the model
I_pred = BEPEmbspB_ttlI(q,x);

% Plot data and scattering intensity from fit parameters
% figure; hold off;
% errorbar(q,I_data,I_std,'ro');
% set(gca,'yscale','log');
% hold on;
% semilogy(q,I_pred,'b-');
% axis tight;
% pause(0.1);

% Calculate the difference divided by the standard deviation in the
% intensity
dy = (I_data - I_pred)./I_std;

end
```

B2.3.4 fit_BEPEmbspB_main.m

```
function [q,I_data,I_std,I_pred,Pf,Pf_err,chisq,chisqr] =
fit_BEPEmbspB_main(file_ID,P1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
```

```
% Created: June 28, 2017
% Updated: June 28, 2017
%
% The objective of this function is to fit the structure factor of Borue
% Polyelectrolyte to SANS data
%
% Variables
% INPUTS:
%       file_ID         Name of a csv file containing data of the form:
%                           - column 1: q data
%                           - column 2: I data
%                           - column 3: I standard deviation data
%       P1              Vector of initial parameter guesses:
%                       - P1(1):  Prefactor (scale * drho^2)            [=]
1/cm
%                       - P1(2):  a, Segment length, scalar             [=]
angstroms
%                       - P1(3):  alpha, Degree of ionization, scalar
[=] fraction
%                       - P1(4):  chi, Flory interaction parameter      [=]
unitless
%                       - P1(5): Background                             [=]
1/cm
%
% OUTPUTS:
%       q               column vector of q data points from the input file
%       I_data          column vector of I data points from the input file
%       I_std           column vector of I std dev points from the input
%                       file
%       I_pred          column vector of I points predicted by the model
%                       fit
%       Pf              Vector of optimized parameters (same order as
%                       initial guess)
%       Pf_err          Error in optimized parameter fits corresponding to
%                       a 95% confidence interval
%       chisq           Chi squared value for the fit.
%       chisqr          Reduced chi squared value for the fit
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Load data from csv file
data = csvread(file_ID);

% Assign the data to separate arrays
q = data(:,1);
I_data = data(:,2);
I_std = data(:,3);

% Set fit options
options = optimset('MaxFunEvals',inf,'TolFun',1e-12,'MaxIter',1000);

% Set upper and lower bounds for the fit parameters
lb = [0;0;0;0;0];
ub = [200;100;100;0.5;100];

% Fit the data
```

```
[Pf,~,residual,exitflag,~,~,jacobian] =
lsqnonlin('BEPEmbspB_dy',P1,lb,ub,options,q,I_data,I_std);

% Calculate the error in the fit using Matlab functions for 95% confidence
% interval
ci = nlparci(Pf,residual,'jacobian',jacobian);
Pf_err = Pf - ci(:,1);

% Calculate model I(q) over the q data range from the fit parameters
I_pred = BEPEmbspB_ttlI(q,Pf);

% Calculate the chi^2 and reduced chi^2
chisq = ((I_data- I_pred)./I_std).^2;
chisq = sum(chisq);
chisqr = chisq/(length(q)-length(Pf));

end
```

B2.3.5 fit_BEPEmbspB_wrapper.m

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: June 28, 2017
% Updated: September 13, 2017
%
% The purpose of this script is to perform a fit and make a plot to compare
% the fit to the original data.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% I. PERFORM THE FIT
file_ID = 'ELP0nc-D100-25mgmL-bkgsub-trim.csv';     % file name
P1 = [0.36;4.;0.3;0.41;0.02];              % initial guesses

[q,I_data,I_std,I_pred,Pf,Pf_err,chisq,chisqr] =
fit_BEPEmbspB_main(file_ID,P1);

% Calculate the separate functions
I_ig = BEPEmbspB_ttlI(q,P1);

%% II. PLOT THE RESULTS
figure(1); clf
subplot(1,2,1);
errorbar(q,I_data,I_std,'bo');
hold on;
set(gca,'xscale','log','yscale','log');
loglog(q,I_pred,'k','linewidth',2);
xlabel('q (nm$^{-1}$)','interpreter','latex','fontsize',12);
ylabel('I(q) (cm$^{-1}$)','interpreter','latex','fontsize',12);
h_legend= legend('data','model');
set(h_legend,'interpreter','latex','fontsize',10);
% axis square;
title('After Fit');

subplot(1,2,2);
errorbar(q,I_data,I_std,'bo');
```

```
hold on;
set(gca,'xscale','log','yscale','log');
loglog(q,I_ig,'k','linewidth',2);
xlabel('q (nm$^{-1}$)','interpreter','latex','fontsize',12);
ylabel('I(q) (cm$^{-1}$)','interpreter','latex','fontsize',12);
h_legend= legend('data','model');
set(h_legend,'interpreter','latex','fontsize',10);
% axis square;
title('Initial Guess');


% III. WRITE THE RESULTS TO A TEXT FILE
filename_str = 'ELP0nc-alphavary-chibounded_trim_BEPEmbspB.txt';
fileID = fopen(filename_str,'wt');
fprintf(fileID,datestr(clock));
fprintf(fileID,'\n');
fprintf(fileID,'==========================');
fprintf(fileID,'\n');
fprintf(fileID,'Model: Borue Erukhimovich Polyelectrolyte, Monovalent Salt,
Bjerrum Length = 0.704 nm, Salt concentration = 0, Polymer volume fraction =
0.020292, Third Virial Coeff = 0');
fprintf(fileID,'\n');
fprintf(fileID,'==========================');
fprintf(fileID,'\n');
fprintf(fileID,'-----Input Parameters-----\n');
%write input parameters to file
fprintf(fileID,'%25s','P1 = ');
fprintf(fileID,'%15.5f',P1');
fprintf(fileID,'\n');
fprintf(fileID,'-----Results-----\n');
%write fit parameters to file
fprintf(fileID,'chi squared = %10.5f\n',chisq);
fprintf(fileID,'reduced chi squared = %10.5f\n',chisqr);
fprintf(fileID,'%25s','CF = ');
fprintf(fileID,'%10.4e +/- %10.4e',Pf(1),Pf_err(1));
fprintf(fileID,'\n');
fprintf(fileID,'%25s ','Segment length = ');
fprintf(fileID,'%10.5f +/- %10.5f',Pf(2), Pf_err(2));
fprintf(fileID,'\n');
fprintf(fileID,'%25s ','m = ');
fprintf(fileID,'%10.5f +/- %10.5f',Pf(3), Pf_err(3));
fprintf(fileID,'\n');
fprintf(fileID,'%25s','Chi = ');
fprintf(fileID,'%10.6f +/- %10.6f',Pf(4),Pf_err(4));
fprintf(fileID,'\n');
fprintf(fileID,'%25s','Bkg = ');
fprintf(fileID,'%10.6f +/- %10.6f',Pf(5),Pf_err(5));
fprintf(fileID,'\n');
fprintf(fileID,'\n');
matrixwrite = [q I_data I_std I_pred]';
fprintf(fileID,'%10s\t %10s\t %10s\t %10s\n','q (A^-1)', 'I(Q)', 'I(Q)
std.dev.', 'I(Q)_pred');
fprintf(fileID,'----------------------------------------------------------------
----------------------------------------------------------------');
fprintf(fileID,'\n');
fprintf(fileID,'%15.7f \t %15.5f \t %15.5f \t %15.5f\n',matrixwrite);
```

```
fclose(fileID);
% Write a data only file as well
filename_str = 'DATAONLY-ELP0nc-alphavary-chibounded-trim-BEPEmbspB.txt';
fileID = fopen(filename_str,'wt');
matrixwrite = [q I_data I_std I_pred]';
fprintf(fileID,'%10s\t %10s\t %10s\t %10s\n','q (A^-1)', 'I(Q)', 'I(Q)
std.dev.', 'I(Q)_pred');
fprintf(fileID,'\n');
fprintf(fileID,'%15.7f \t %15.5f \t %15.5f \t %15.5f\n',matrixwrite);
fclose(fileID);
```

B2.4 Code for Excluded Volume Chain Fused to Cylindrical Colloid (Fit for mCherry-ELP Fusions)

B2.4.1 Pcyl.m

```
function y = Pcyl(q,L,R)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Christopher N. Lam
% Created: September 28, 2014
% Updated: October 20, 2014
%
% This function calculates the form factor of a homogeneous cylinder with
% length L and R.
%
% variables:
% input:
%       q           array, 1 x length(q)
%       L           length of cylinder
%       R           radius of cylinder
%
% output:
%       y           array, 1 x length(y)
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
alphavec = linspace(0.0001,pi/2,1000);

for ia = 1:length(alphavec)
    alphacurr = alphavec(ia);
    Pmatrix(ia,:) =
sin(alphacurr)*(2*besselj(1,q*R*sin(alphacurr))./(q*R*sin(alphacurr)).*(sin(q
*L*cos(alphacurr)/2)./(q*L*cos(alphacurr)/2))).^2;
end
y = ((pi/2-0.0001)/(1000-1))*trapz(Pmatrix);

end
```

B2.4.2 Pexcludedvol.m

```
function y = Pexcludedvol(q, Rg, m)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: June 20, 2017
% Updated: June 20, 2017
%
% This function calculates the structure factor of an excluded volume
% polymer
```

```
%
% Variables:
% INPUT:
%       q           scattering vector, array, 1 x length(q)      [=] 1/nm
%       Rg          radius of gyration                           [=] nm
%       m           Porod exponent                               [=] Unitless
%       Note: Rg^2 = a^2*N^(2/m)/((2/m+1)(2/m+2)), where a is statistical
%       segment length and N is degree of polymerization
%
% OUTPUTS:
%       y           Structure factor, array, 1 x length(q)     [=] unitless
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Calculate parameters
nu2 = 2/m;

% Define the function to be integrated


% Form factor calculation
y = zeros(length(q),1);

for i = 1:length(q)
    qi = q(i);
    fun = @(x)(2*(1-x).*exp(-qi.^2.*x.^nu2*((nu2+1)*(nu2+2)*Rg^2/6)));
    y(i) = integral(fun,0,1);
end

end
```

B2.4.3 Pcylev_conj.m

```
function Pconj = Pcylev_conj(q,b,L,mp,ak,rho,Lc,Rc,bkg)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: November 14,2017
% Updated: November 14, 2017
%
% This function calculates the form factor of an excluded volume polymer
% conjugated to a cylinder.
%
% Variables:
% INPUT:
%       q           scattering vector, array, 1 x length(q)      [=] 1/nm
%       b           scattering length of polymer                 [=] ?
%       L           contour length of polymer                    [=] nm
%       mp           Porod exponent of polymer                    [=]
%       unitless
%       ak          Kuhn segment length of polymer               [=] nm
%       rho         Cylinder scattering length density           [=] 1/nm^3
%       Lc          Cylinder length                              [=] nm
%       Rc          Cylinder radius                              [=] nm
%       bkg         Background
%
% OUTPUTS:
```

```
%       y           Structure factor, array, 1 x length(q)     [=] unitless
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Calculate polymer only form factor
nu2 = 2/mp; N = L/ak;
Rg = ak*(N^nu2 / ((nu2+1)*(nu2+2)))^0.5
Ppoly = (b*L/ak)^2*Pexcludedvol(q,Rg,mp); Ppoly = Ppoly(:);
N
ak
b
mp
rho
bkg

% Calculate cylinder form factor
V = pi*Lc*Rc^2;
Pcyli = (rho*V)^2*Pcyl(q,Lc,Rc);
Pcyli = Pcyli';

% Calculate the cross structure factor
n = 1:1:N;

% Define integration variables
dtheta=pi/23;
theta=0:dtheta:pi;
dalpha=dtheta/2;
alpha=0:dalpha:pi/2;

LHS3=zeros(length(q),1);
for i=1:length(n)
    Ln=n(i)*ak;
    dLn=ak/5;
    r=dLn:dLn:Ln;
    for j=1:length(r)
        LCY=zeros(length(q),1);
        solid_angle_theta=0;
        for k=1:length(theta)
            t=sqrt(r(j)^2+(Lc/2)^2+2*r(j)*(Lc/2)*cos(theta(k)));
            if t>0
                beta=acos(((Lc/2)^2+t^2-r(j)^2)/2/(Lc/2)/t);
                if t*sin(beta)>Rc || t*cos(beta)>Lc/2
                    omega=2*pi*sin(theta(k));
                    pnr=p_r(r(j),ak,n(i))/4/pi;
                    solid_angle_alpha=0;
                    RWW=zeros(length(q),1);
                    for m=1:length(alpha)

RWW=RWW+pnr*V*Fcyl(q,alpha(m),Rc,Lc).*cos(q*t*cos(theta(k)+alpha(m))).*sin(al
pha(m)).*dalpha;

solid_angle_alpha=solid_angle_alpha+sin(alpha(m))*dalpha;
                    end
                    LCY=LCY+RWW/solid_angle_alpha*omega*dLn*dtheta;
                    solid_angle_theta=solid_angle_theta+omega*dtheta;
                end
```

```
            end
        end
        LHS3=LHS3+LCY/solid_angle_theta*4*pi;
    end
end
Pcross=LHS3*2*b*rho;

Pconj = Pcross + Ppoly + Pcyli + bkg;

End
```

B2.4.4 p_r.m

```
function pr=p_r(r,ak,N1)
% This function calculates the probability of finding the end
% of an excluded volume chain of length N, with Kuhn segment
% length ak at distance r.

P=ak/2;

mu=0.588;
gamma=1.1619;
t=1/(1-mu);
theta=(gamma-1)/mu;
kappa=1.1040;

L=N1*ak;
rms=sqrt(2*P*L*(1-P/L*(1-exp(-L/P))));

pr=1/0.265545/rms*(r/rms).^(2+theta).*exp(-1*(kappa*(r/rms)).^t);

pr=pr(:);
```

B2.4.5 Pcylevconj_dy.m

```
function dy = Pcylevconj_dy(x,q,I_data, I_std)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: November 25, 2017
% Updated: November 25, 2017
%
% The objective of this function is to calculate the difference between the
% scattering model I(q) and the data I(q), divided by the data standard
% deviation.
%
% Variables:
% INPUT:
%       x           Array of fit parameters:
%                       - x(1): scattering length of kuhn seg   [=] 1/cm?
%                       [- L, polymer contour length       [=] nm] Do not
%                       fit, set
%                       - x(2): mp, porod exponent
%                       - x(3): ak, Kuhn segment length         [=] nm
%                       - x(4): rho, cylinder scattering length density   [=]
1/cm
%                       [Lc, Rc: Cylinder size, set by mCherry fit]
```

```
%                       - x(5): bkg
%       q           Array of q values, length(q) x 1            [=] 1/nm
%       I_data      Array of data I(q), length(q) x 1                   [=]
1/cm
%       I_std       Array of standard deviation data for I(q), length(q)x1
[=] 1/cm
%
% OUTPUTS:
%       dy          Array of the difference between the scattering model
%                   I(q) and the data I(q), divided by the data's standard
%                   deviation.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Set contour length of polymer
aa_length = 0.4;        % [=] nm, contour length of an amino acid
(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1697845/)
L = 518*aa_length;      % [=] nm, 518 amino acids per ELP!

% Set dimensions of cylinder
Lc = 4.73; Rc = 1.51; % [=] nm, CNL:
http://onlinelibrary.wiley.com/doi/10.1002/pola.27975/epdf

% Calculate the total scattering intensity of the model
I_pred = Pcylev_conj(q,x(1),L,x(2),x(3),x(4),Lc,Rc,x(5));

% Plot data and scattering intensity from fit parameters
% figure; hold off;
% errorbar(q,I_data,I_std,'ro');
% set(gca,'yscale','log');
% hold on;
% semilogy(q,I_pred,'b-');
% axis tight;
% pause(0.1);

% Calculate the difference divided by the standard deviation in the
% intensity
dy = (I_data - I_pred)./I_std

end
```

B2.4.6 fit_Pcylevconj_main.m

```
function [q,I_data,I_std,I_pred,Pf,Pf_err,chisq,chisqr] =
fit_Pcylevconj_main(file_ID,P1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: Nov 25, 2017
% Updated: Nov 25, 2017
%
% The objective of this function is to fit the form factor of a ev polymer
% conjugated to a cylinder
%
% Variables
% INPUTS:
%       file_ID         Name of a csv file containing data of the form:
%                           - column 1: q data
%                           - column 2: I data
```

```
%                           - column 3: I standard deviation data
%       P1              Vector of initial parameter guesses:
%                           - P1(1): prefactor/scattering length
%                           - P1(2): porod exponent
%                           - P1(3): Kuhn segment length [=] nm
%                           - P1(4): scattering length prefactor, cylinder
%                           - P1(5): Background          [=] 1/cm
%
% OUTPUTS:
%       q               column vector of q data points from the input file
%       I_data          column vector of I data points from the input file
%       I_std           column vector of I std dev points from the input
%                       file
%       I_pred          column vector of I points predicted by the model
%                       fit
%       Pf              Vector of optimized parameters (same order as
%                       initial guess)
%       Pf_err          Error in optimized parameter fits corresponding to
%                       a 95% confidence interval
%       chisq           Chi squared value for the fit.
%       chisqr          Reduced chi squared value for the fit
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Load data from csv file
data = csvread(file_ID);

% Assign the data to separate arrays
q = data(:,1);
I_data = data(:,2);
I_std = data(:,3);

% Set fit options
options = optimset('MaxFunEvals',1e4,'TolFun',1e-8);
%options = optimset('TolFun',1e-8);

% Set upper and lower bounds for the fit parameters
lb = [0;0;0;0;-1];
ub = [1000;500;1000;1000;1000];

% Fit the data
[Pf,~,residual,~,~,~,jacobian] =
lsqnonlin('Pcylevconj_dy',P1,lb,ub,options,q,I_data,I_std);

% Calculate the error in the fit using Matlab functions for 95% confidence
% interval
ci = nlparci(Pf,residual,'jacobian',jacobian);
Pf_err = Pf - ci(:,1);

% Set contour length of polymer
aa_length = 0.4;        % [=] nm, contour length of an amino acid
(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1697845/)
L = 518*aa_length;      % [=] nm, 518 amino acids per ELP!
% Set dimensions of cylinder
Lc = 4.73; Rc = 1.51; % [=] nm, CNL:
http://onlinelibrary.wiley.com/doi/10.1002/pola.27975/epdf
```

```
% Calculate model I(q) over the q data range from the fit parameters
I_pred = Pcylev_conj(q,Pf(1),L,Pf(2),Pf(3),Pf(4),Lc,Rc,Pf(5));

% Calculate the chi^2 and reduced chi^2
chisq = ((I_data- I_pred)./I_std).^2;
chisq = sum(chisq);
chisqr = chisq/(length(q)-length(Pf));

end
```

B2.4.7 fit_Pcylevconj_wrapper.m

```
function [] = fit_Pcylevconj_wrapper(filename)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Carolyn Mills
% Created: Nov 25, 2017
% Updated: Nov 25, 2017
%
% The purpose of this script is to perform a fit and make a plot to compare
% the fit to the original data.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% I. PERFORM THE FIT
close all;
file_ID = [filename,'-D100-25mgmL-10C-bkgsub-trim.csv'];    % file name
P1 = [0.001;2;1.0;0.01628;.000001];            % initial guesses
% Load data from csv file
data = csvread(file_ID);

% Assign the data to separate arrays
q = data(:,1);
I_data = data(:,2);
I_std = data(:,3);

[q,I_data,I_std,I_pred,Pf,Pf_err,chisq,chisqr] =
fit_Pcylevconj_main(file_ID,P1);

% Calculate the separate functions
% Set contour length of polymer
aa_length = 0.4;        % [=] nm, contour length of an amino acid
(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1697845/)
L = 518*aa_length;      % [=] nm, 518 amino acids per ELP!
% Set dimensions of cylinder
Lc = 4.73; Rc = 1.51; % [=] nm, CNL:
http://onlinelibrary.wiley.com/doi/10.1002/pola.27975/epdf
I_ig = Pcylev_conj(q,P1(1),L,P1(2),P1(3),P1(4),Lc,Rc,P1(5));

%% II. PLOT THE RESULTS
% figure(1); clf
% subplot(1,2,1);
% errorbar(q,I_data,I_std,'bo');
% hold on;
% set(gca,'xscale','log','yscale','log');
% loglog(q,I_pred,'k','linewidth',2);
```

```
% xlabel('q (nm$^{-1}$)','interpreter','latex','fontsize',12);
% ylabel('I(q) (cm$^{-1}$)','interpreter','latex','fontsize',12);
% h_legend= legend('data','model');
% xlim([0.1 2])
% ylim([.002 1])
% set(h_legend,'interpreter','latex','fontsize',10);
% % axis square;
% title('After Fit');
%
% subplot(1,2,2);
% errorbar(q,I_data,I_std,'bo');
% hold on;
% set(gca,'xscale','log','yscale','log');
% loglog(q,I_ig,'k','linewidth',2);
% xlabel('q (nm$^{-1}$)','interpreter','latex','fontsize',12);
% ylabel('I(q) (cm$^{-1}$)','interpreter','latex','fontsize',12);
% h_legend= legend('data','model');
% set(h_legend,'interpreter','latex','fontsize',10);
% axis square;
% title('Initial Guess');

%% III. WRITE THE RESULTS TO A TEXT FILE
 filename_str = [filename,'_Pcylevconj.txt'];
 fileID = fopen(filename_str,'wt');
 fprintf(fileID,datestr(clock));
 fprintf(fileID,'\n');
 fprintf(fileID,'==========================');
 fprintf(fileID,'\n');
 fprintf(fileID,'Model: Excluded Volume Polymer Conjugated to Cylinder');
 fprintf(fileID,'\n');
 fprintf(fileID,'==========================');
 fprintf(fileID,'\n');
 fprintf(fileID,'-----Input Parameters-----\n');
%write input parameters to file
 fprintf(fileID,'%25s','P1 = ');
 fprintf(fileID,'%15.5f',P1');
 fprintf(fileID,'\n');
 fprintf(fileID,'-----Results-----\n');
%write fit parameters to file
 fprintf(fileID,'chi squared = %10.5f\n',chisq);
 fprintf(fileID,'reduced chi squared = %10.5f\n',chisqr);
 fprintf(fileID,'%25s','CF = ');
 fprintf(fileID,'%10.4e +/- %10.4e',Pf(1),Pf_err(1));
 fprintf(fileID,'\n');
 fprintf(fileID,'%25s ','Contour length = ');
 fprintf(fileID,'%10.5f +/- %10.5f nm',L, 0);
 fprintf(fileID,'\n');
 fprintf(fileID,'%25s ','Porod Exponent = ');
 fprintf(fileID,'%10.5f +/- %10.5f',Pf(2), Pf_err(2));
 fprintf(fileID,'\n');
 fprintf(fileID,'%25s ','Kuhn Length = ');
 fprintf(fileID,'%10.5f +/- %10.5f nm',Pf(3), Pf_err(3));
 fprintf(fileID,'\n');
 fprintf(fileID,'%25s','Cylinder Contrast = ');
 fprintf(fileID,'%10.5f +/- %10.5f',Pf(4), Pf_err(4));
 fprintf(fileID,'%25s','Bkg = ');
 fprintf(fileID,'%10.6f +/- %10.6f',Pf(5),Pf_err(5));
```

```
 fprintf(fileID,'\n');
 fprintf(fileID,'\n');
 fprintf(fileID,'\n');
 matrixwrite = [q I_data I_std I_pred]';
 fprintf(fileID,'%10s\t %10s\t %10s\t %10s\n','q (nm^-1)', 'I(Q)', 'I(Q)
std.dev.', 'I(Q)_pred');
 fprintf(fileID,'----------------------------------------------------------------
----------------------------------------------------------------------');
 fprintf(fileID,'\n');
 fprintf(fileID,'%15.7f %15.5f %15.5f %15.5f\n',matrixwrite);
 fclose(fileID);
 % Write a data only file as well
 filename_str = ['DATAONLY-',filename,'_Pcylevconj.txt'];
 fileID = fopen(filename_str,'wt');
 matrixwrite = [q I_data I_std I_pred]';
 fprintf(fileID,'%10s\t %10s\t %10s\t %10s\n','q (nm^-1)', 'I(Q)', 'I(Q)
std.dev.', 'I(Q)_pred');
 fprintf(fileID,'\n');
 fprintf(fileID,'%15.7f %15.5f %15.5f %15.5f\n',matrixwrite);
 fclose(fileID);
end
```

B2.5 References

(1) Li, X.; Lam, C. N.; Sanchez-Diaz, L. E.; Smith, G. S.; Olsen, B. D.; Chen, W. R., Scattering from Colloid-Polymer Conjugates with Excluded Volume Effect. *Acs Macro Lett* **2015,** *4* (2), 165-170.

Appendix C. Supporting Information for Chapter 4

```
MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSASGLVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVG
IPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVG
IPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVG
IPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVG
IPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVG
IPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVG
IPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVG
IPGVGVPGVGVPGVGETTS
```

Figure C-1. Amino acid sequence of ELP used in this study

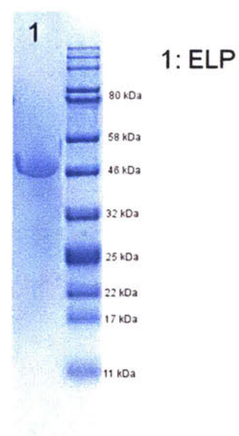


Figure C-2. SDS-PAGE gel of ELP used in this study. In this gel, the ELP runs at ~46 kDa, which is expected based on the molecular weight calculated from the protein sequence (45.69 kDa).

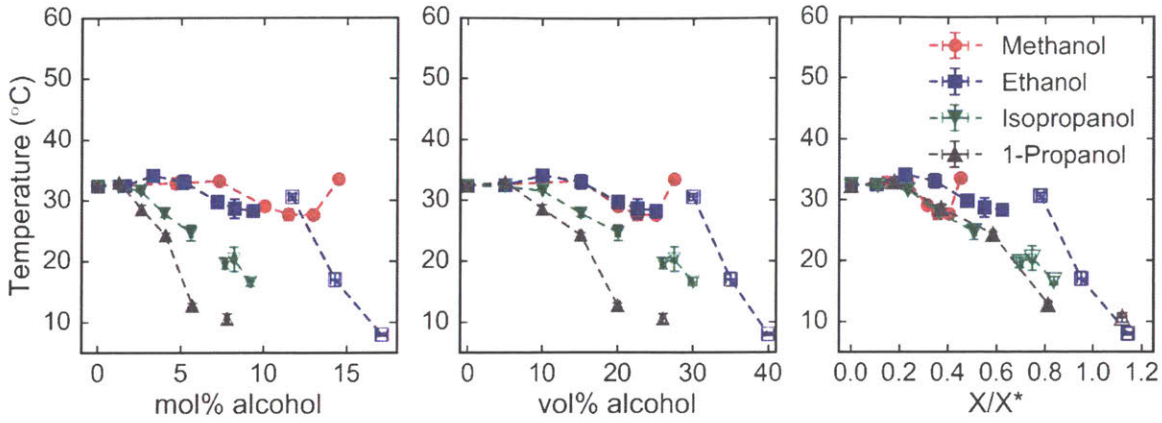


Figure C-3. Transition temperatures upon heating in different solvents plotted against mole percent alcohol, mole fraction solvent divided by mole fraction alcohol at which water-alcohol mixing enthalpy is minimized (x*), and volume percent alcohol.

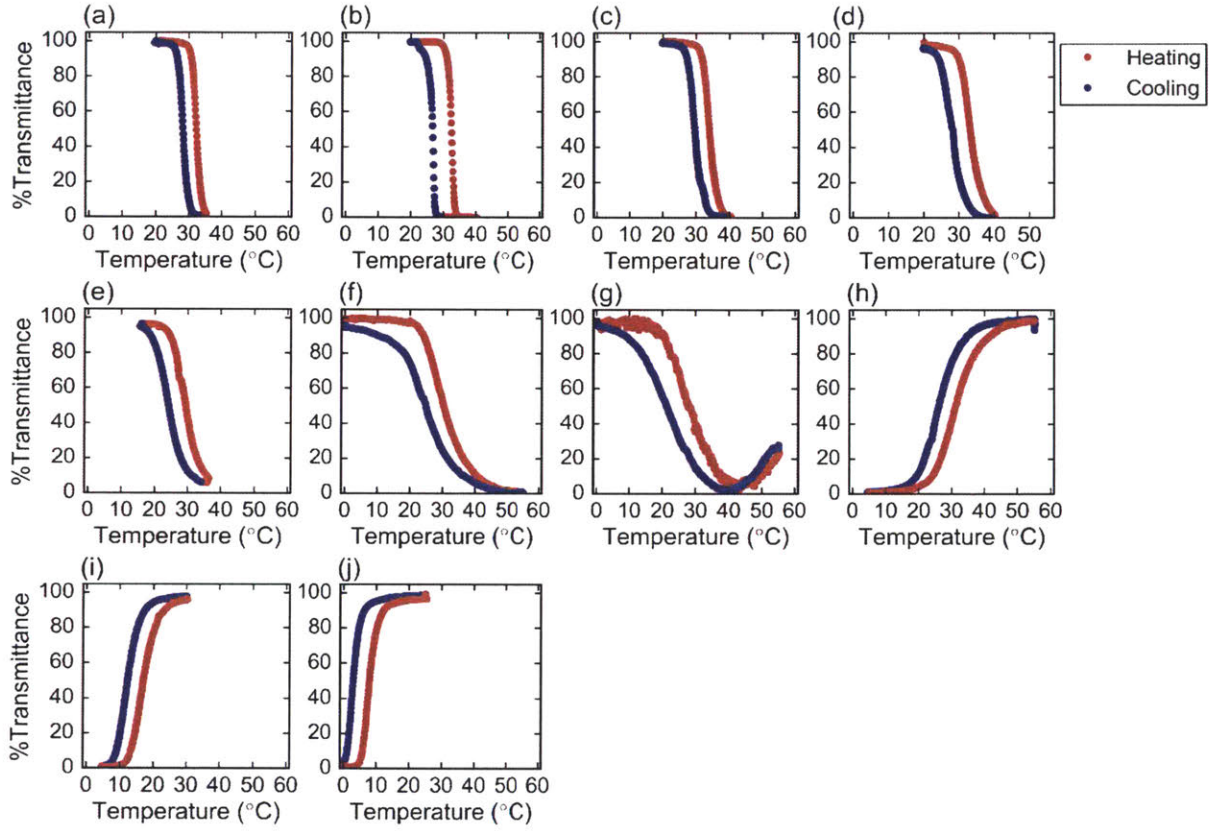


Figure C-4. Examples of raw transmittance data taken on the ELP in various water/ethanol blends. (a) 0 mol% ethanol, (b) 1.6 mol% ethanol, (c) 3.3 mol% ethanol, (d) 5.2 mol% ethanol, (e) 7.2 mol% ethanol, (f) 8.2 mol% ethanol, (g) 9.4 mol% ethanol, (h) 11.7 mol% ethanol, (i) 14.3 mol% ethanol, and (j) 17.1 mol% ethanol.

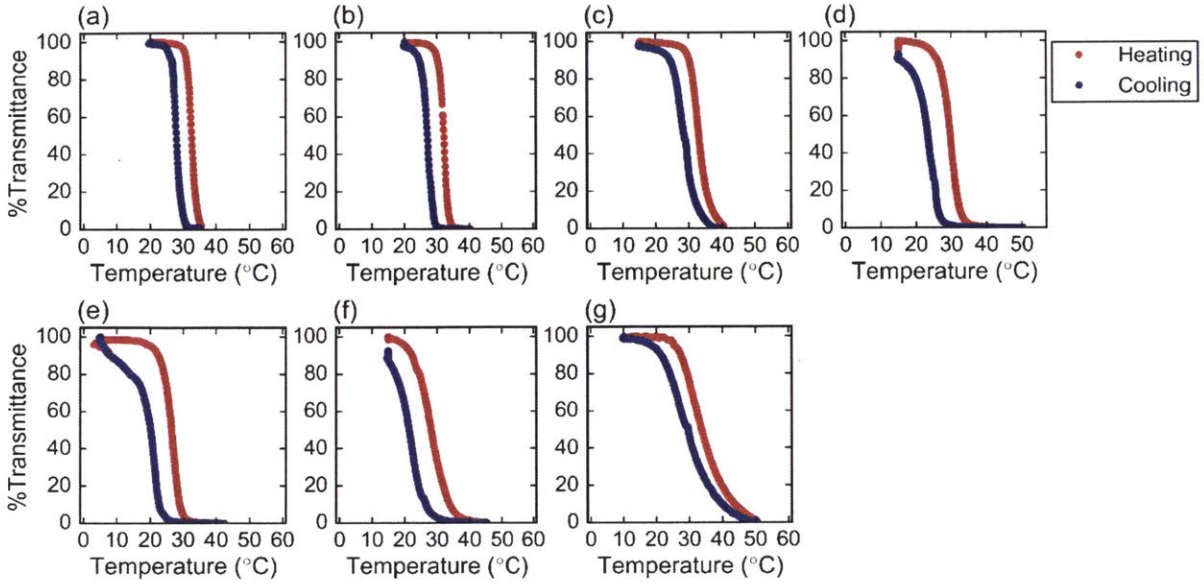


Figure C-5. Raw transmittance data taken on the ELP in various water/methanol blends. (a) 0 mol% methanol, (b) 4.7 mol% methanol, (c) 7.3 mol% methanol, (d) 10.0 mol% methanol, (e) 11.5 mol% methanol, (f) 13.0 mol% methanol, and (g) 14.5 mol% methanol.

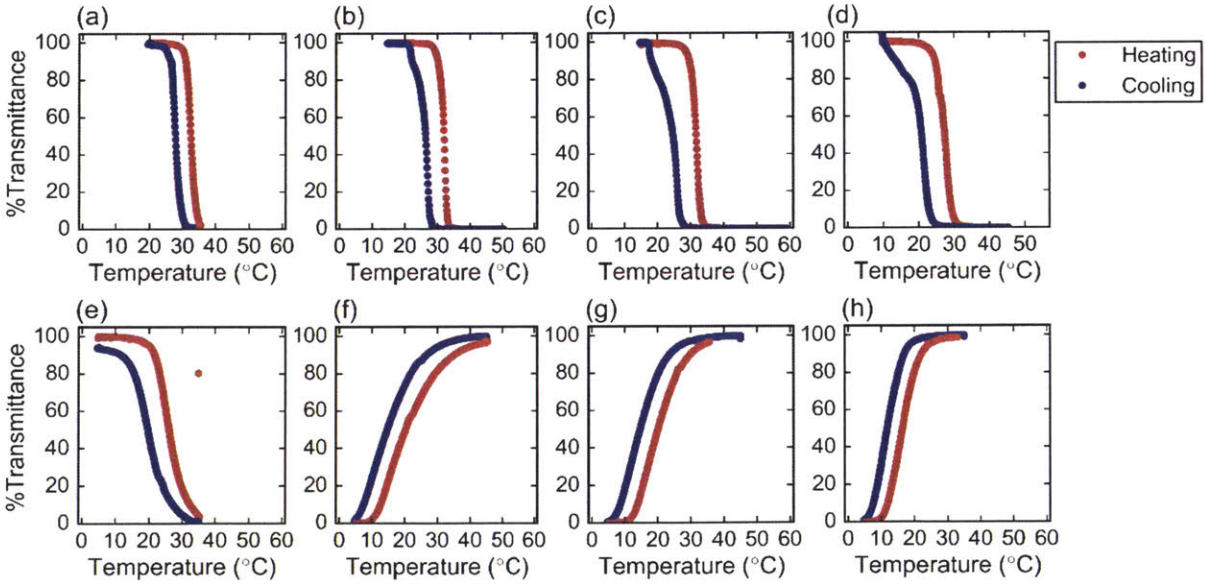


Figure C-6. Raw transmittance data taken on the ELP in various water/isopropanol blends. (a) 0 mol% isopropanol, (b) 1.2 mol% isopropanol, (c) 2.6 mol% isopropanol, (d) 4.0 mol%

isopropanol, (e) 5.6 mol% isopropanol, (f) 7.7 mol% isopropanol, (g) 8.2 mol% isopropanol, and (h) 9.2 mol% isopropanol.

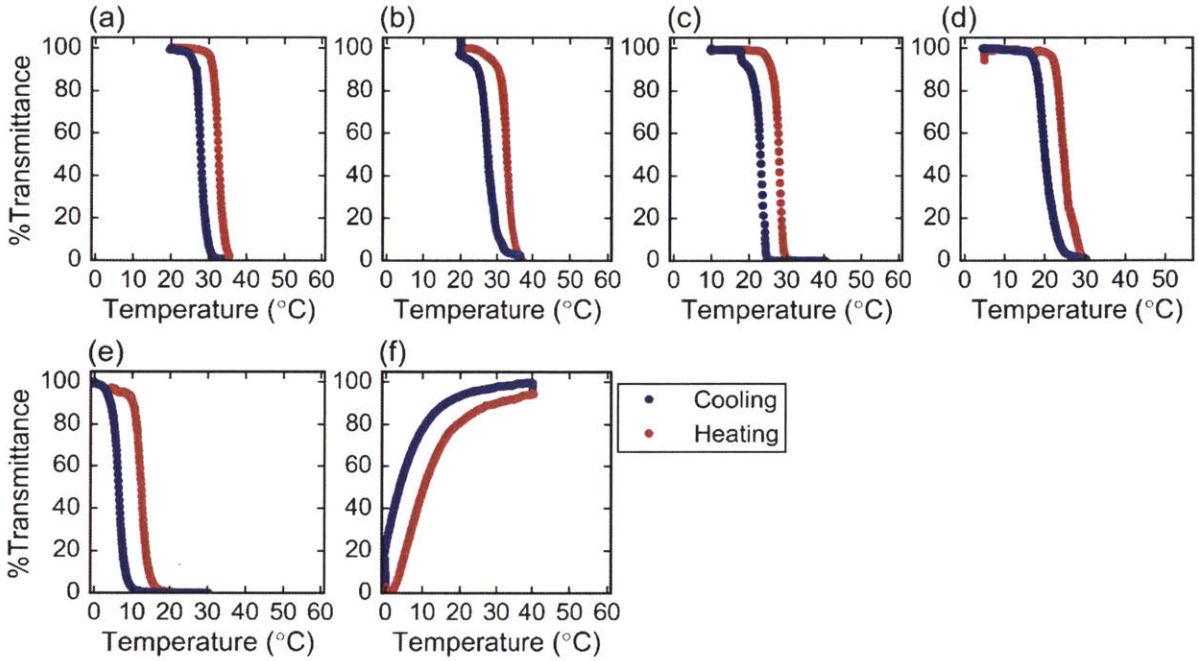


Figure C-7. Raw transmittance data taken on The ELP in various water/1-propanol blends. (a) 0 mol% 1-propanol, (b) 1.3 mol% 1-propanol, (c) 2.6 mol% 1-propanol, (d) 4.1 mol% 1-propanol, (e) 5.7 mol% 1-propanol, and (f) 7.8 mol% 1-propanol.

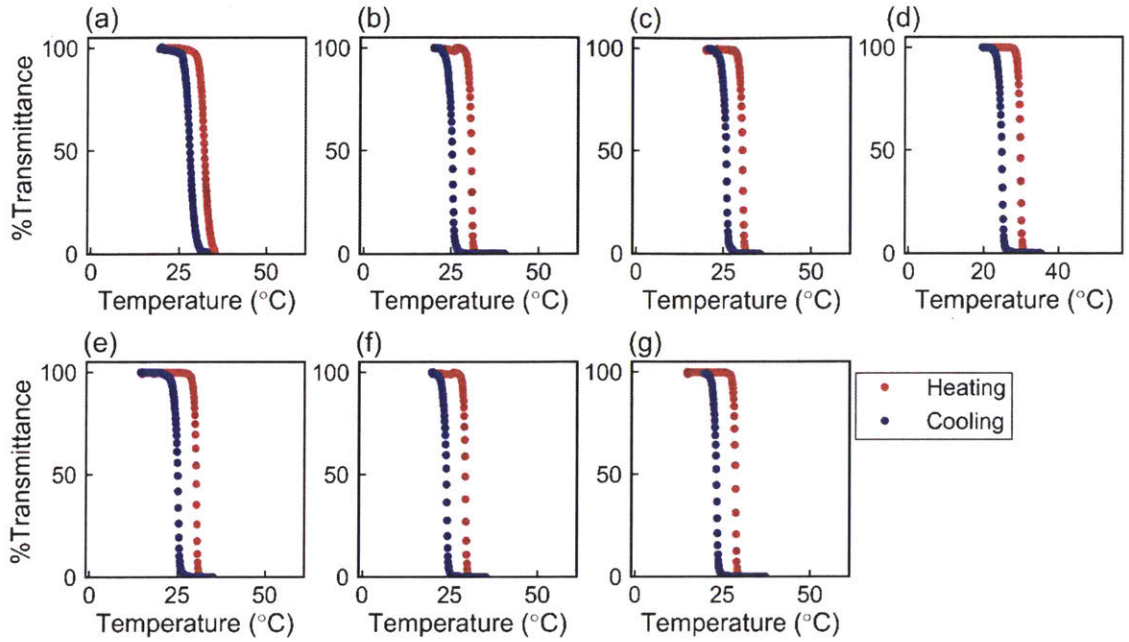


Figure C-8. Raw transmittance data taken on the ELP in water with (a) 0 mM NaCl, (b) 25 mM NaCl, (c) 50 mM NaCl, (d) 75 mM NaCl, (e) 100 mM NaCl, and (f) 200 mM NaCl

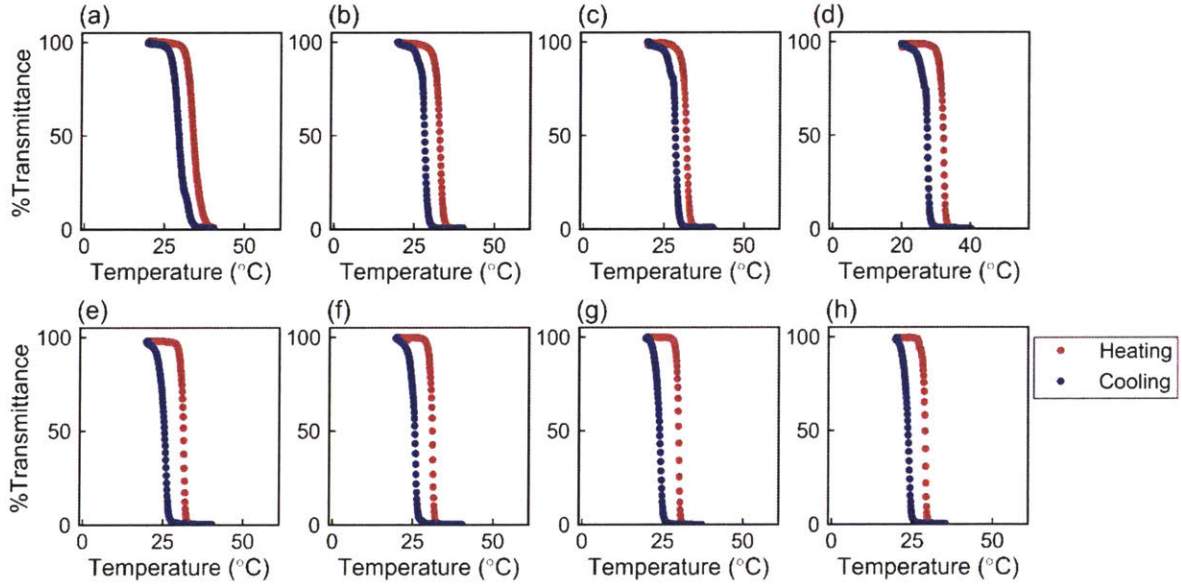


Figure C-9. Raw transmittance data taken on the ELP in 3.3 mol% ethanol, 96.7 mol% water with (a) 0 mM NaCl, (b) 12.5 mM NaCl, (c) 25 mM NaCl, (d) 50 mM NaCl, (e) 75 mM NaCl, (f) 100 mM NaCl, (g) 150 mM NaCl, and (h) 200 mM NaCl.

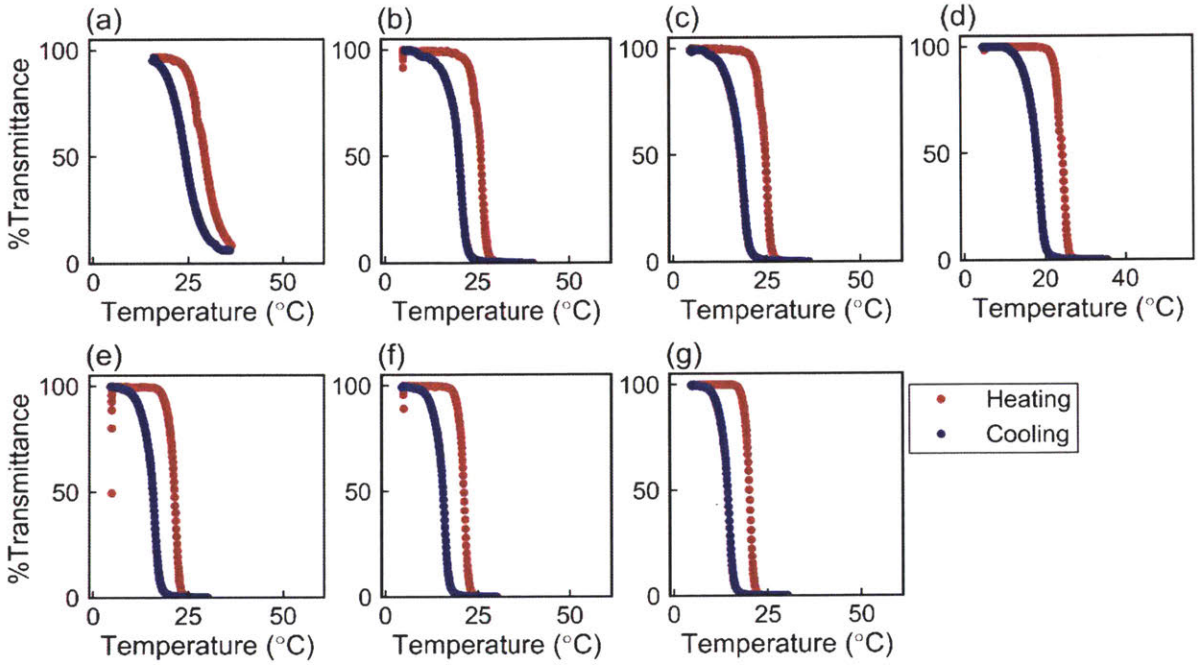


Figure C-10. Raw transmittance data taken on the ELP in 7.2 mol% ethanol, 92.8 mol% water with (a) 0 mM NaCl, (b) 25 mM NaCl, (c) 50 mM NaCl, (d) 75 mM NaCl, (e) 100 mM NaCl, (f) 150 mM NaCl, and (g) 200 mM NaCl.

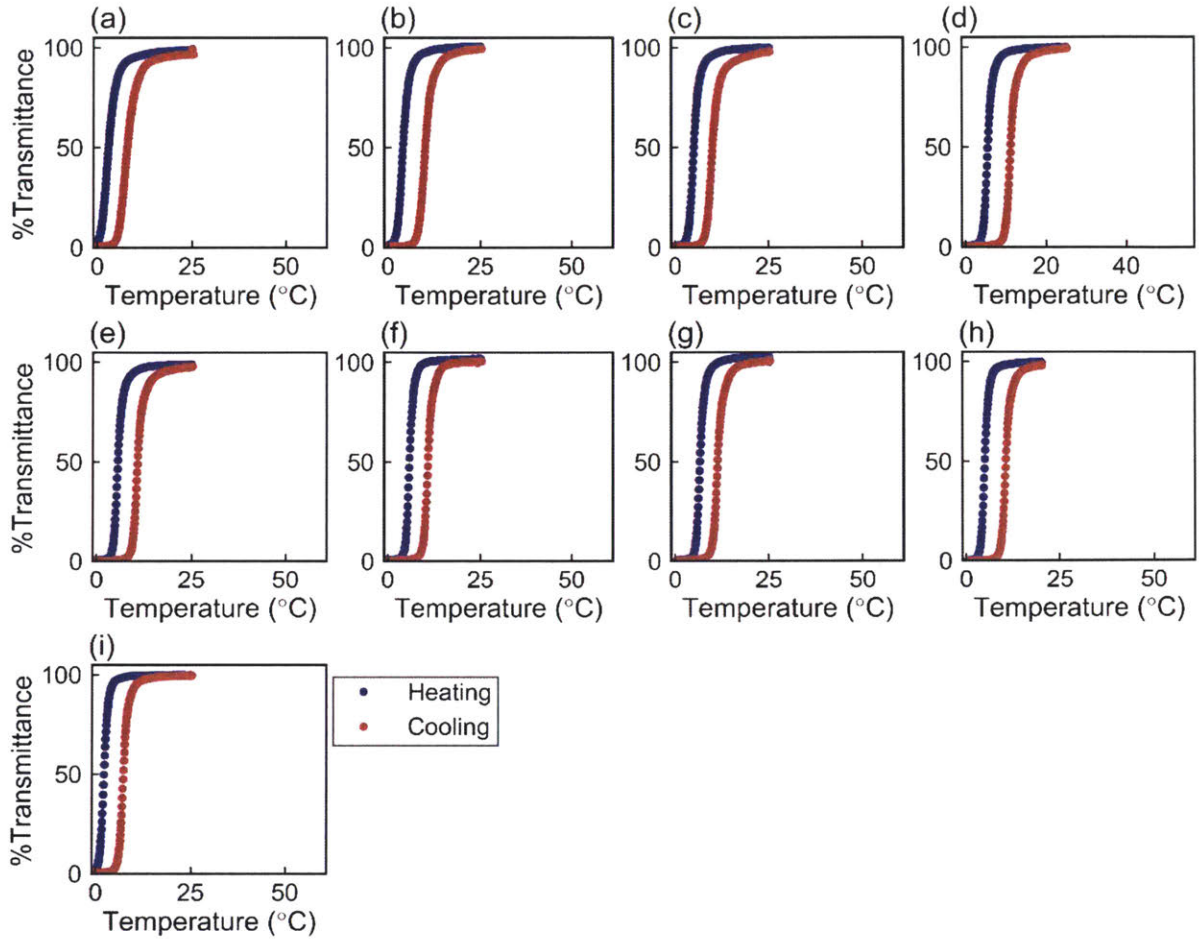


Figure C-11. Raw transmittance data taken on the ELP in 17.1 mol% ethanol, 82.9 mol% water with (a) 0 mM NaCl, (b) 5 mM NaCl, (c) 12.5 mM NaCl, (d) 25 mM NaCl, (e) 50 mM NaCl, (f) 75 mM NaCl, (g) 100 mM NaCl, (h) 150 mM NaCl, and (i) 200 mM NaCl.

C.1 Propagation of Pipetting Error

Errors were propagated through summation using the following formula

$$err_{\text{add}} = \left(\sum_i err_i \right)^{0.5} \tag{C-1}$$

where err_{add} is the error on the final sum, err_i is the error on each component being summed, and i sums over all components that contribute to the funal summed value.

Errors were propagated through divisions and multiplications using the following formula

$$err_{\text{multdiv}} = val_{\text{multdiv}} \left(\sum_i \left(\frac{err_i}{val_i} \right) \right)^{0.5} \quad \text{(C-2)}$$

where err_{multdiv} is the error in the final value produced through multiplication/division, val_{multdiv} is the final value produced through multiplication/division, val_i are the values being multiplied/divided, err_i are the errors in each of these values being multiplied/divided, and i sums over all values being multiplied/divided.

Errors in volumes mixed were derived from the pipetting errors supplied by the manufacturer (VWR Signature™ Ergonomic High Performance Single-Channel Variable Volume Pipettors). For volumes between 50 and 200 μL, the pipettor (VWR# 89079-970) was assumed to have an error in accuracy of 1.2 vol%. Volumes added that fell between 200 and 1000 μL were assumed to have an error in accuracy of 1.6 vol% (VWR# 89079-974).

Volume percent alcohol was calculated assuming to change in volume upon mixing using the following formula

$$vol\%_{\text{alcohol}} = \frac{vol_{\text{alcohol}}}{vol_{\text{total}}} \quad \text{(C-3)}$$

where vol_{alcohol} is the volume of alcohol added to a given solution and vol_{total} is the total volume of the final solution. To propagate error through this expression, the additive error propagation formula was first used to determine the error propagated into vol_{total} by the different volumes added to the solution (aqueous and alcohols). The formula for propagating error through division steps was then used (with the error in total volume produced by the calculation from the previous sentence) to calculate the error in the final calculated volume percent alcohol reported. Mole percent alcohol was calculated by first converting the volumes and respective volumetric pipetting errors added to solution into moles using the density and molecular weight of water and a given alcohol. After this conversion, the propagation of error into the calculated *mol*% alcohol reported in this manuscript was identical to the error propagation above for *vol*% alcohol.

Final sodium chloride concentration in solutions was adjusted by adding 500 mM NaCl in water to the final NaCl concentrations reported in the manuscript. To propagate pipetting error into this reported [NaCl], we considered the following formula

$$C_{\text{final}} = C_{\text{stock}} \frac{v_{\text{stock}}}{v_{\text{total}}} \tag{C-4}$$

where C_{final} is the final NaCl concentration in solution, C_{stock} is 500 mM, v_{stock} is the volume of 500 mM NaCl added to the solution to achieve the final NaCl concentration and v_{total} is the final total volume of the prepared solution. The error propagation for this calculated NaCl concentration, then, was identical to the procedure for that used in the vol% calculation, except here the error in pipetting of the 500 mM NaCl stock was used in the propagation of error through division instead of the error in alcohol pipetting. In this case, the final error was calculated by multiplying the vol% error in 500 mM NaCl pipetting by 500 mM, as multiplicative constants apply to errors the same as the values being multiplied.

Appendix D. Supporting Information for Chapter 5

```
MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSMVSSGLVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPG
VGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPG
VGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPG
VGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPG
VGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPG
VGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPG
VGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPG
VGIPGVGVPGVGVPGVGETTSGSACELMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFI
CTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNR
IELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK*
```

Figure D-1. Amino acid sequence of ELP-sfGFP used in this paper.

```
ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGACT
GGTGGACAGCAAATGGGTCGCGGATCCATGGTGTCTAGCGGTCTCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGTACCTGGTGTTGGCGTCCCGGGT
GTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGAGACCACTAGTGGATCCGCCTGC
GAACTGATGGTGAGCAAAGGTGAAGAACTGTTTACTGGTGTAGTTCCGATCCTGGTAGAACTGGACGGCGATGTT
AACGGCCACAAGTTCAGCGTGCGTGGCGAAGGTGAAGGTGATGCAACCAACGGCAAACTGACCCTGAAATTCATC
TGTACTACCGGCAAACTGCCGGTGCCTTGGCCGACTCTGGTCACCACGCTGACTTACGGTGTTCAGTGTTTCTCT
CGTTACCCGGATCACATGAAACAGCACGATTTCTTCAAATCTGCTATGCCGGAAGGTTATGTGCAAGAGCGTACC
ATCTCCTTCAAAGATGATGGCACCTACAAAACCCGTGCGGAAGTTAAATTCGAAGGTGACACGCTGGTCAACCGC
ATTGAACTGAAAGGCATCGATTTCAAAGAAGATGGCAACATCCTGGGCCACAAGCTGGAATACAACTTTAATTCC
CACAACGTATATATCACTGCGGATAAACAGAAAAACGGCATTAAAGCTAACTTCAAGATCCGCCACAACGTTGAA
GATGGTTCCGTTCAACTGGCGGACCACTACCAGCAGAACACTCCAATCGGTGACGGCCCAGTTCTGCTGCCAGAT
AACCACTACCTGTCTACCCAAAGCGCGCTGTCTAAAGACCCGAACGAAAAACGTGACCATATGGTTCTGCTGGAA
[illegible]
```

Figure D-2. DNA sequence of ELP-sfGFP used in this paper. Gene was cloned into pET28a between the BamHI and HindIII restriction sites. The bolded underlined sequence corresponds to the BamHI restriction site used for cloning this construct into the pET28a vector.

D.1 Selection of Technique for Phase Diagram Construction

A BCA assay was chosen for determine protein concentration based on preliminary experiments constructing similar phase diagrams. In these experiments, three methods for measuring protein concentration were compared—BCA assay, Bradford assay, and sfGFP fluorescence. Methods were evaluated based on their ability to reproducibly close the mass balance; that is, whether or

not the protein concentrations measured in the soluble and insoluble fractions summed to the same value across the different ethanol and sodium chloride concentrations tested. Of the three methods tested, the BCA assay showed the best closure of the mass balance, so it was chosen for performing the final protein concentration measurements. BCA assay mass balances across the conditions studied in this work can be found in **Figure D-3**, and the respective standard deviations in these mass balances across three replicate experiments can be found in **Figure D-4**. Standard deviations on the protein concentrations measured in the soluble and insoluble fractions at each temperature can be found plotted in **Figure D-5**. The protein concentration data before normalization are reported in **Table D-1**, and the standard deviations across three replicate experiments, also before normalization, are reported in **Table D-2**. To generate these data, standard curves of α-amylase were used to convert BCA absorbance to protein concentration at various solvent conditions; these standard curves can be found in **Figure D-6**, **Figure D-7**, and **Figure D-8**.

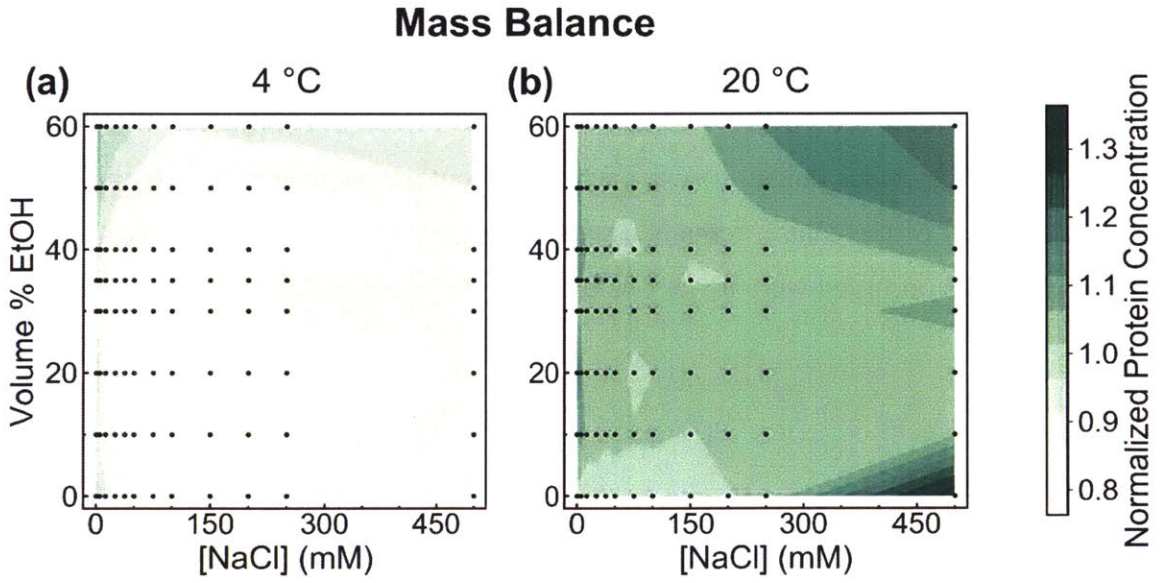


Figure D-3. Mass balance of the ELP-sfGFP phase diagram data generated at (a) 4 ˚C and (b) 20 ˚C. Values reported are the sum of the soluble and insoluble average protein concentration at each solvent condition, normalized to the sum of the soluble and insoluble average protein concentration at 0 vol% EtOH and 0 mM NaCl. Solvent conditions where data were taken are represented by black points, between which the contours were generated by bilinear interpolation. Mass balance values greater than 1 suggest evaporative effects.

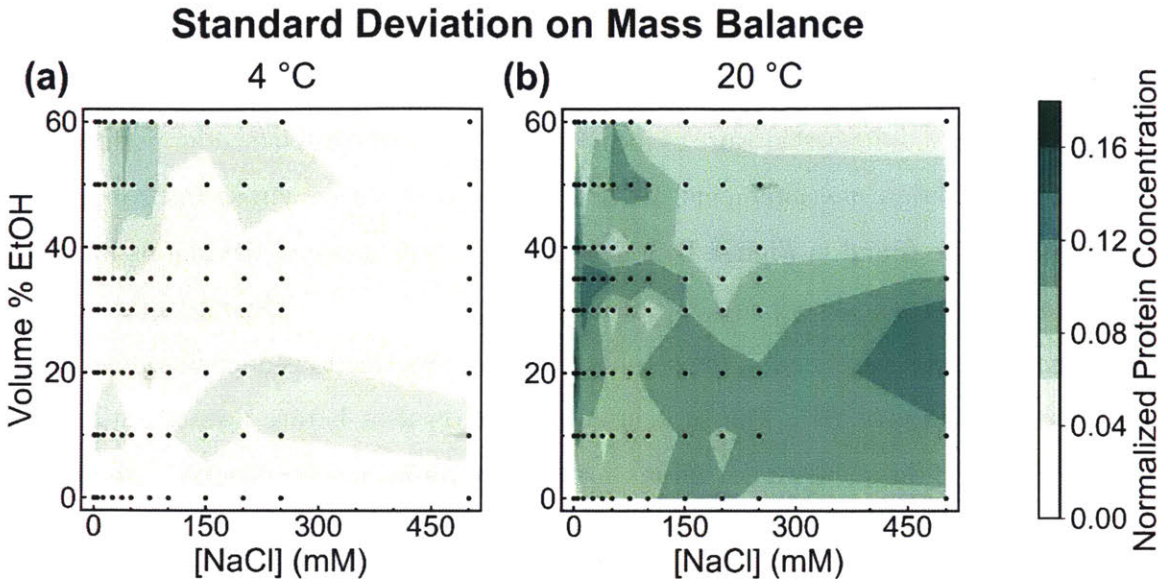


Figure D-4. Standard deviation of the mass balance values at (a) 4 ˚C and (b) 20 ˚C. Values reported are standard deviation, over three replicates, of the sum of the soluble and insoluble protein concentration values at each data point, normalized to the sum of the average soluble and insoluble protein concentration at 0 vol% EtOH and 0 mM NaCl. Solvent conditions where data were taken are represented by black points, between which the contours were generated by bilinear interpolation. The majority of the dataset remained within 15% error (standard deviation < 0.15) though standard deviations observed were lower overall when the experiment was performed at 4 ˚C.

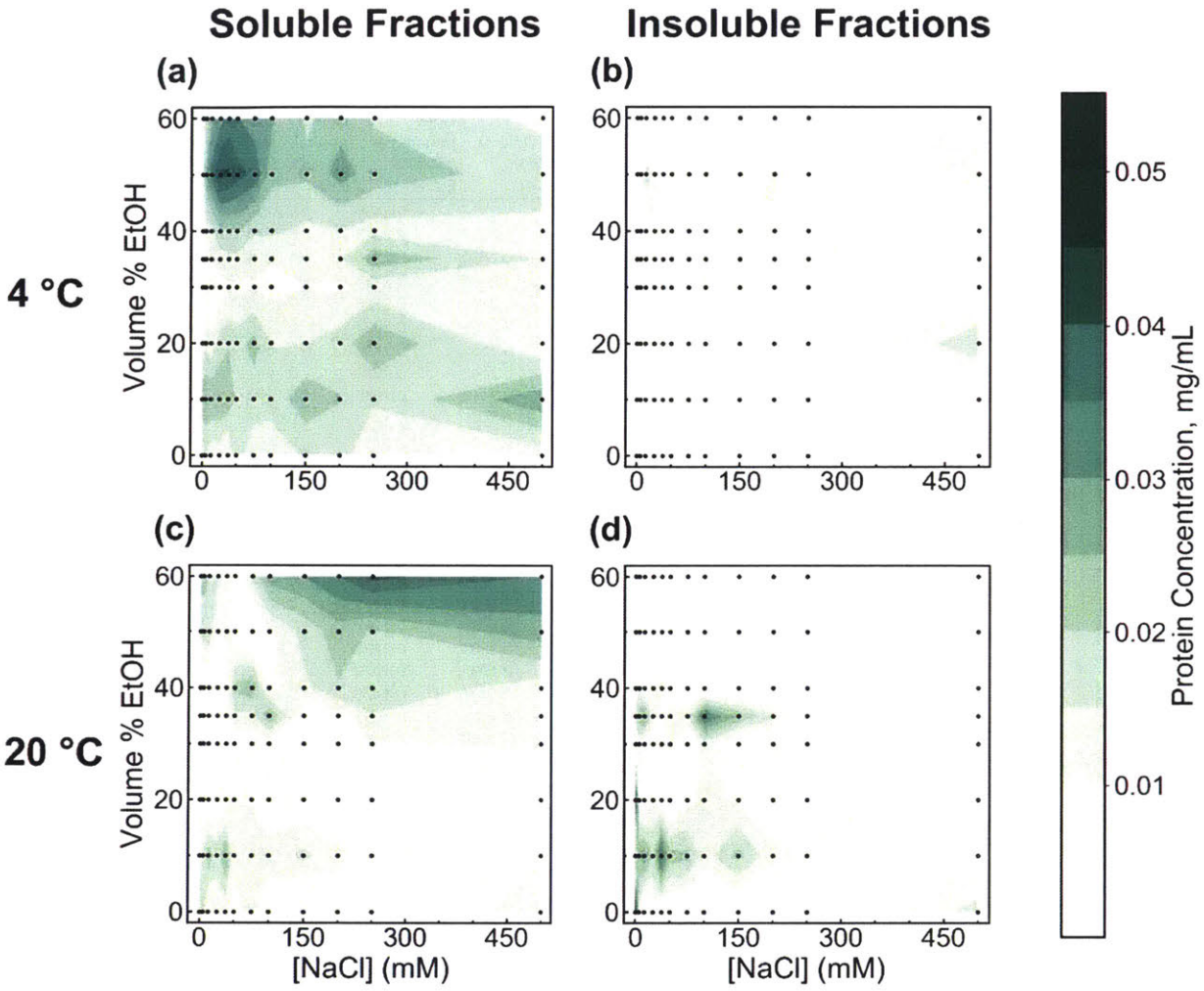


Figure D-5. Standard deviations across three replicate experiments of the measured ELP-sfGFP concentration in each of the (a, c) soluble and (b, d) insoluble fractions at (a, b) 4 ˚C and (c, d) 20 ˚C.

Table D-1. Data used to generate ELP-sfGFP phase diagrams*.

Temperature (°C)	Soluble or Insoluble Fraction	Concen-tration of NaCl (mM)	Protein Conc. (mg/mL) at 0 vol% EtOH	Protein Conc. (mg/mL) at 10 vol% EtOH	Protein Conc. (mg/mL) at 20 vol% EtOH	Protein Conc. (mg/mL) at 30 vol% EtOH	Protein Conc. (mg/mL) at 35 vol% EtOH	Protein Conc. (mg/mL) at 40 vol% EtOH	Protein Conc. (mg/mL) at 50 vol% EtOH	Protein Conc. (mg/mL) at 60 vol% EtOH
4	Insoluble	0	0.0669	0.0677	0.1672	0.2522	0.2495	0.2259	0.1681	0.0715
		5	0.0651	0.0654	0.1322	0.2344	0.2343	0.2127	0.1721	0.0693
		12.5	0.0653	0.0649	0.1205	0.2254	0.2250	0.2018	0.1414	0.0707
		25	0.0650	0.0667	0.0903	0.2185	0.2163	0.1904	0.1044	0.0691
		37.5	0.0636	0.0642	0.0826	0.2161	0.2125	0.1829	0.0906	0.0670
		50	0.0650	0.0651	0.0803	0.2125	0.2094	0.1716	0.0887	0.0676
		75	0.0653	0.0662	0.0746	0.2151	0.2109	0.1590	0.0819	0.0680
		100	0.0655	0.0660	0.0756	0.2142	0.2036	0.1482	0.0741	0.0692
		150	0.0647	0.0681	0.0725	0.2118	0.1970	0.1147	0.0728	0.0700
		200	0.0652	0.0689	0.0751	0.2145	0.1853	0.0937	0.0723	0.0682
		250	0.0678	0.0698	0.0768	0.2198	0.1763	0.0850	0.0664	0.0705
		500	0.0709	0.0687	0.1411	0.2239	0.1371	0.0796	0.0764	0.0740
4	Soluble	0	0.3232	0.3025	0.2103	0.1079	0.1135	0.1477	0.2322	0.3370
		5	0.3036	0.2850	0.2246	0.1081	0.1114	0.1437	0.2150	0.3246
		12.5	0.2965	0.2770	0.2210	0.1071	0.1114	0.1464	0.2308	0.3203
		25	0.2882	0.2718	0.2366	0.1072	0.1146	0.1542	0.2573	0.3164
		37.5	0.2844	0.2649	0.2372	0.1076	0.1175	0.1640	0.2636	0.3150
		50	0.2808	0.2628	0.2341	0.1071	0.1181	0.1689	0.2617	0.3049
		75	0.2705	0.2530	0.2294	0.1080	0.1187	0.1773	0.2588	0.2976
		100	0.2663	0.2474	0.2255	0.1060	0.1192	0.1831	0.2551	0.2889
		150	0.2691	0.2554	0.2252	0.1092	0.1329	0.2076	0.2544	0.2889
		200	0.2724	0.2567	0.2271	0.1137	0.1476	0.2284	0.2596	0.2970
		250	0.2722	0.2617	0.2354	0.1181	0.1651	0.2429	0.2679	0.2930
		500	0.2787	0.2770	0.1988	0.1245	0.2052	0.2639	0.2792	0.3023
20	Insoluble	0	0.0944	0.1920	0.3088	0.2851	0.2721	0.2502	0.1235	0.0613
		5	0.0719	0.1913	0.2865	0.2842	0.2722	0.2406	0.0729	0.0670
		12.5	0.0570	0.1548	0.2637	0.2793	0.2589	0.2075	0.0578	0.0592
		25	0.0622	0.1385	0.2667	0.2724	0.2518	0.1597	0.0592	0.0605
		37.5	0.0602	0.1191	0.2664	0.2707	0.2341	0.1096	0.0590	0.0607
		50	0.0594	0.1268	0.2592	0.2608	0.2166	0.0767	0.0580	0.0587
		75	0.0595	0.1049	0.2468	0.2583	0.1955	0.0700	0.0576	0.0589
		100	0.0589	0.0796	0.2460	0.2579	0.1674	0.0683	0.0576	0.0617
		150	0.0587	0.0833	0.2447	0.2441	0.1065	0.0649	0.0589	0.0596
		200	0.0575	0.0916	0.2417	0.2309	0.0780	0.0649	0.0583	0.0606
		250	0.0576	0.1415	0.2426	0.2222	0.0653	0.0648	0.0589	0.0593
		500	0.2063	0.2299	0.2842	0.2371	0.0647	0.0656	0.0582	0.0607
20	Soluble	0	0.2985	0.2280	0.1516	0.1287	0.1438	0.1814	0.3135	0.3515
		5	0.3106	0.2348	0.1371	0.1285	0.1461	0.1841	0.3360	0.3429
		12.5	0.3238	0.2583	0.1371	0.1289	0.1530	0.2102	0.3442	0.3434
		25	0.3222	0.2703	0.1389	0.1359	0.1677	0.2532	0.3385	0.3497
		37.5	0.3191	0.2849	0.1416	0.1344	0.1827	0.3007	0.3403	0.3500
		50	0.3235	0.2792	0.1478	0.1409	0.2001	0.3126	0.3488	0.3498
		75	0.3287	0.2950	0.1458	0.1485	0.2165	0.3227	0.3462	0.3438
		100	0.3294	0.3213	0.1519	0.1563	0.2425	0.3389	0.3404	0.3496
		150	0.3391	0.3137	0.1624	0.1717	0.2874	0.3354	0.3475	0.3513
		200	0.3395	0.3112	0.1617	0.1815	0.3195	0.3386	0.3562	0.3693
		250	0.3453	0.2743	0.1599	0.1881	0.3379	0.3385	0.3697	0.3790
		500	0.3303	0.1850	0.1217	0.1856	0.3480	0.3613	0.4002	0.4086

* These data represent the protein concentration values before normalization to the sum of soluble and insoluble fractions at 0mM NaCl and 0 vol% EtOH (data at different temperatures would be normalized separately).

Table D-2. Standard deviations on data used to generate ELP-sfGFP phase diagrams*.

Temperature (°C)	Soluble or Insoluble Fraction	Concen-tration of NaCl (mM)	Standard Deviation (mg/mL) at 0 vol% EtOH	Standard Deviation (mg/mL) at 10 vol% EtOH	Standard Deviation (mg/mL) at 20 vol% EtOH	Standard Deviation (mg/mL) at 30 vol% EtOH	Standard Deviation (mg/mL) at 35 vol% EtOH	Standard Deviation (mg/mL) at 40 vol% EtOH	Standard Deviation (mg/mL) at 50 vol% EtOH	Standard Deviation (mg/mL) at 60 vol% EtOH
4	Insoluble	0	0.0016	0.0018	0.0078	0.0062	0.0071	0.0064	0.0066	0.0022
		5	0.0017	0.0024	0.0027	0.0030	0.0100	0.0057	0.0084	0.0001
		12.5	0.0010	0.0020	0.0114	0.0014	0.0069	0.0036	0.0177	0.0019
		25	0.0011	0.0005	0.0047	0.0012	0.0060	0.0104	0.0145	0.0015
		37.5	0.0018	0.0020	0.0067	0.0031	0.0038	0.0056	0.0024	0.0007
		50	0.0034	0.0005	0.0021	0.0021	0.0056	0.0051	0.0068	0.0020
		75	0.0034	0.0007	0.0023	0.0019	0.0043	0.0096	0.0042	0.0038
		100	0.0029	0.0007	0.0004	0.0067	0.0047	0.0029	0.0116	0.0030
		150	0.0033	0.0019	0.0019	0.0073	0.0049	0.0071	0.0102	0.0027
		200	0.0038	0.0033	0.0017	0.0047	0.0028	0.0089	0.0120	0.0041
		250	0.0036	0.0015	0.0009	0.0034	0.0080	0.0017	0.0038	0.0071
		500	0.0028	0.0034	0.0198	0.0054	0.0102	0.0020	0.0125	0.0124
4	Soluble	0	0.0122	0.0244	0.0120	0.0051	0.0093	0.0080	0.0153	0.0138
		5	0.0149	0.0249	0.0139	0.0058	0.0087	0.0116	0.0261	0.0172
		12.5	0.0119	0.0250	0.0109	0.0057	0.0104	0.0140	0.0325	0.0247
		25	0.0119	0.0253	0.0169	0.0067	0.0123	0.0197	0.0412	0.0285
		37.5	0.0074	0.0243	0.0197	0.0095	0.0146	0.0189	0.0410	0.0320
		50	0.0187	0.0189	0.0176	0.0090	0.0137	0.0203	0.0369	0.0305
		75	0.0116	0.0180	0.0219	0.0096	0.0117	0.0157	0.0340	0.0269
		100	0.0125	0.0168	0.0150	0.0086	0.0104	0.0162	0.0213	0.0258
		150	0.0157	0.0234	0.0150	0.0099	0.0125	0.0139	0.0211	0.0193
		200	0.0168	0.0204	0.0175	0.0092	0.0145	0.0148	0.0272	0.0232
		250	0.0120	0.0157	0.0241	0.0118	0.0236	0.0130	0.0227	0.0195
		500	0.0121	0.0289	0.0098	0.0110	0.0149	0.0138	0.0178	0.0139
20	Insoluble	0	0.0544	0.0126	0.0451	0.0034	0.0183	0.0087	0.0064	0.0011
		5	0.0135	0.0239	0.0087	0.0110	0.0244	0.0110	0.0062	0.0095
		12.5	0.0016	0.0292	0.0034	0.0067	0.0275	0.0065	0.0008	0.0019
		25	0.0063	0.0199	0.0096	0.0047	0.0153	0.0085	0.0010	0.0024
		37.5	0.0042	0.0381	0.0073	0.0083	0.0062	0.0159	0.0016	0.0006
		50	0.0035	0.0226	0.0019	0.0113	0.0144	0.0023	0.0017	0.0014
		75	0.0030	0.0218	0.0075	0.0008	0.0141	0.0020	0.0018	0.0038
		100	0.0023	0.0123	0.0118	0.0126	0.0363	0.0001	0.0009	0.0030
		150	0.0018	0.0220	0.0093	0.0030	0.0245	0.0004	0.0018	0.0016
		200	0.0002	0.0116	0.0070	0.0065	0.0164	0.0012	0.0007	0.0017
		250	0.0012	0.0097	0.0035	0.0024	0.0032	0.0008	0.0014	0.0010
		500	0.0173	0.0082	0.0087	0.0048	0.0015	0.0055	0.0031	0.0013
20	Soluble	0	0.0231	0.0092	0.0153	0.0066	0.0042	0.0090	0.0158	0.0225
		5	0.0176	0.0147	0.0045	0.0031	0.0056	0.0032	0.0180	0.0213
		12.5	0.0074	0.0237	0.0038	0.0017	0.0045	0.0041	0.0195	0.0195
		25	0.0108	0.0176	0.0056	0.0010	0.0068	0.0075	0.0071	0.0159
		37.5	0.0110	0.0261	0.0069	0.0081	0.0090	0.0064	0.0128	0.0129
		50	0.0080	0.0119	0.0101	0.0066	0.0143	0.0212	0.0047	0.0094
		75	0.0119	0.0140	0.0096	0.0074	0.0135	0.0246	0.0025	0.0166
		100	0.0081	0.0114	0.0077	0.0072	0.0241	0.0082	0.0090	0.0282
		150	0.0024	0.0164	0.0064	0.0049	0.0066	0.0139	0.0167	0.0304
		200	0.0054	0.0123	0.0011	0.0071	0.0081	0.0197	0.0230	0.0351
		250	0.0066	0.0061	0.0008	0.0097	0.0147	0.0190	0.0195	0.0416
		500	0.0118	0.0099	0.0029	0.0109	0.0101	0.0124	0.0280	0.0348

* These data are presented visually in Figure S5 – standard deviations across three replicate experiments of the protein concentration. Values presented here are protein concentration standard deviations before normalization to the sum of soluble and insoluble fractions at 0mM NaCl and 0 vol% EtOH.

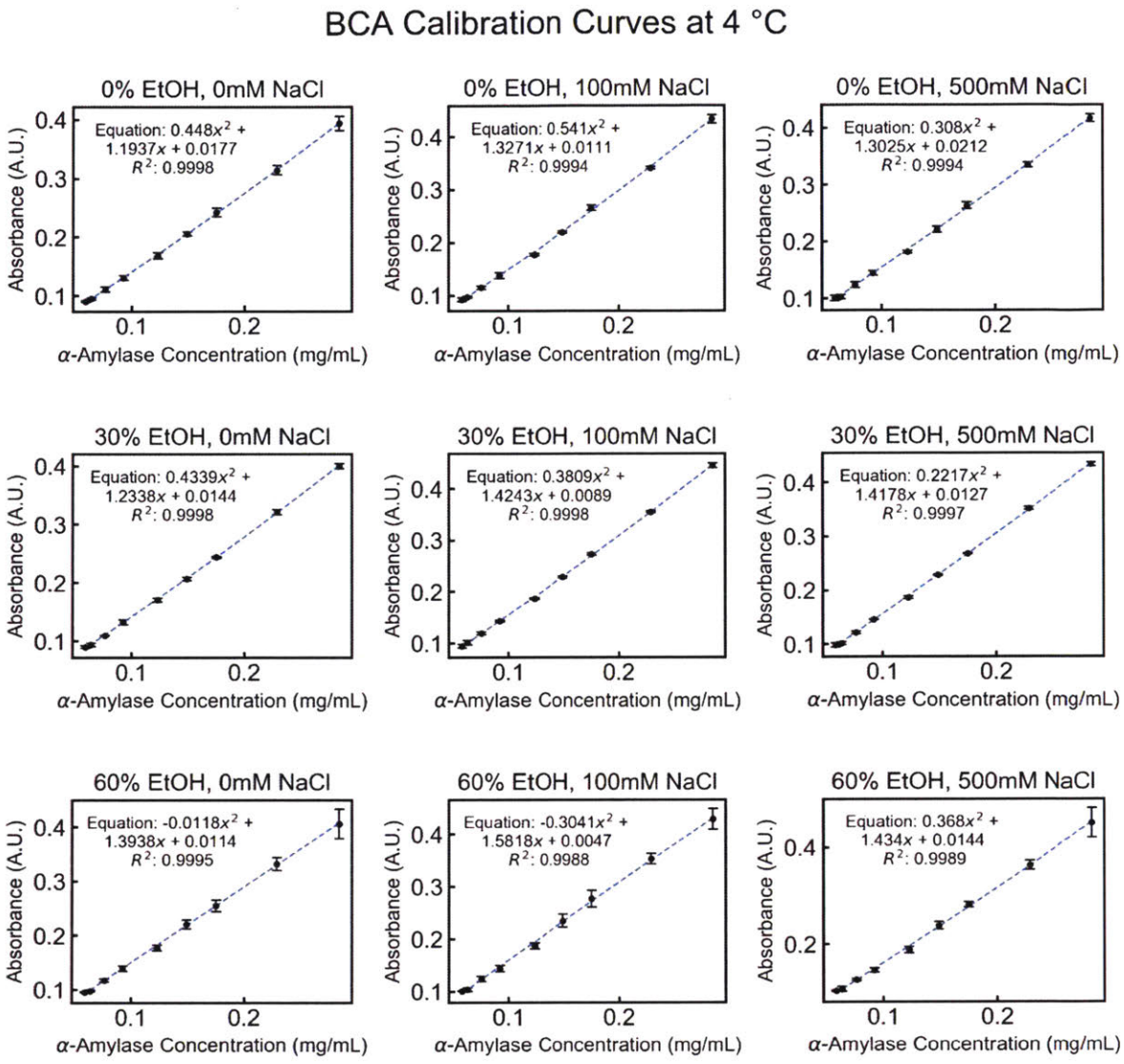


Figure D-6. Quadratic curves used to relate absorbance to protein concentration for the generation of the ELP-sfGFP phase diagram at 4 ˚C.

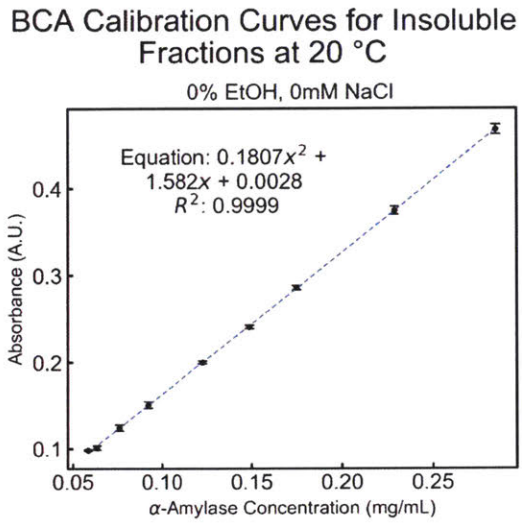


Figure D-7. Quadratic curve used to relate absorbance to protein concentration for the generation of the ELP-sfGFP phase diagram for the 20 ˚C experiment insoluble fraction. Because all insoluble fractions were resuspended in water, only one standard curve needed to be generated for these samples.

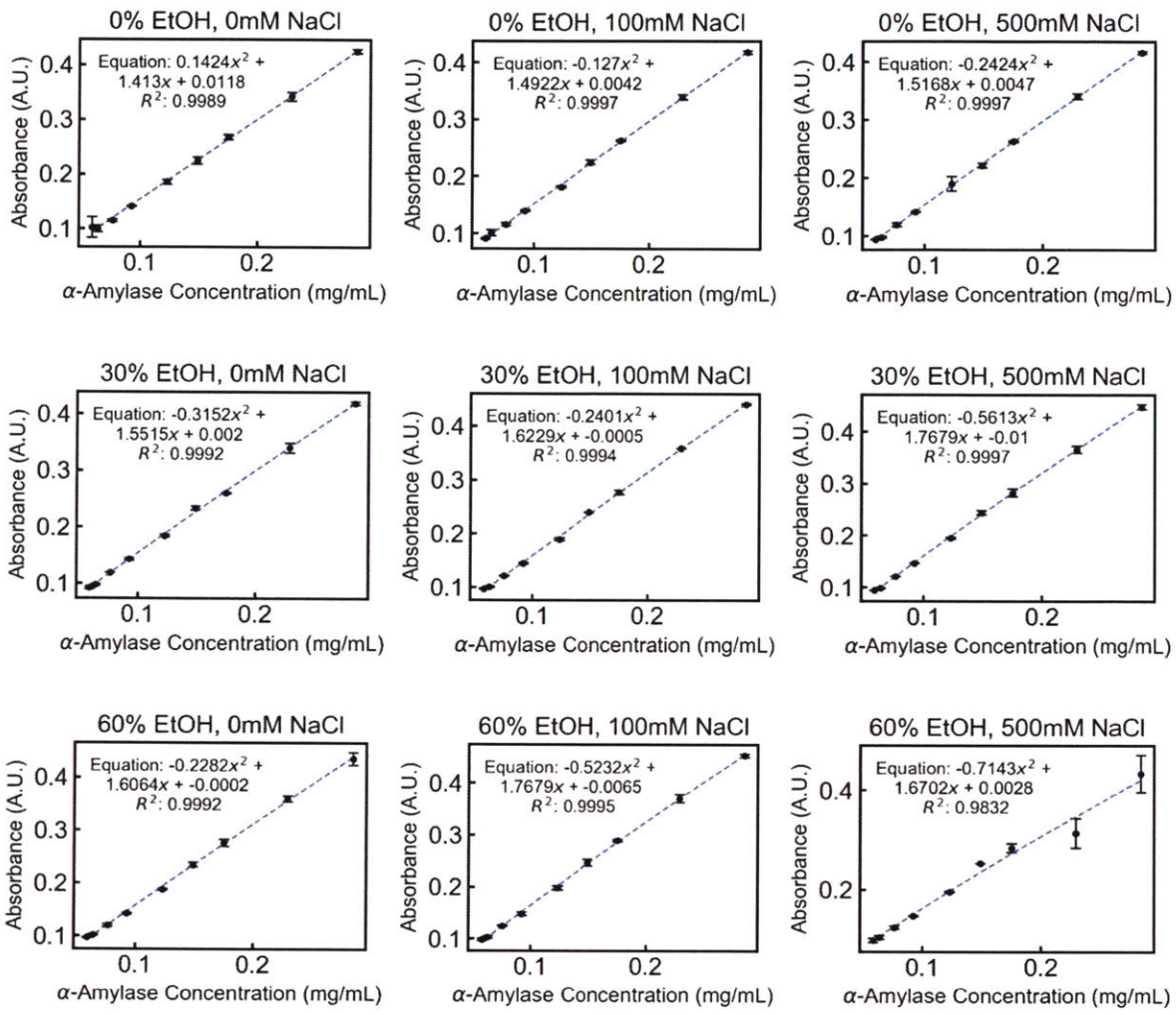


Figure D-8. Quadratic curves used to relate absorbance to protein concentration for the generation of the ELP-sfGFP phase diagram for the 20 ˚C experiment supernatant soluble fractions.

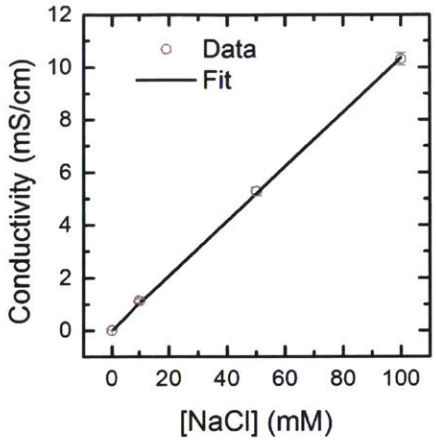


Figure D-9. Conductivity of standards prepared with 0, 10, 50, and 100 mM sodium chloride. Error bars on data points represent standard deviation of measured conductivity over three individually prepared standards at the given concentration. The black line represents the fit used to convert conductivity measurements into equivalents of NaCl in the manuscript.

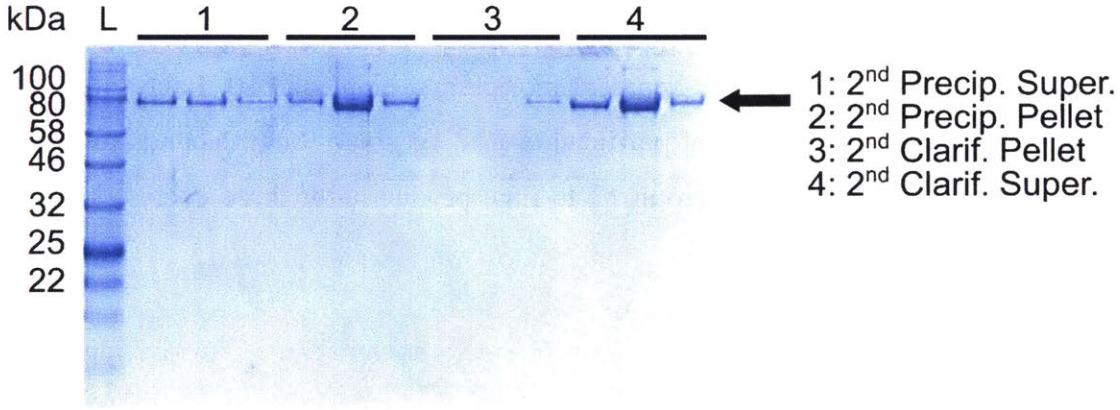


Figure D-10. Coomassie-stained SDS-PAGE gel showing product loss and well-to-well variability in the second consecutive NaCl-dependent purification cycle performed on well-plate expressions of ELP-sfGFP. The arrow indicates the apparent molecular weight of ELP-sfGFP.

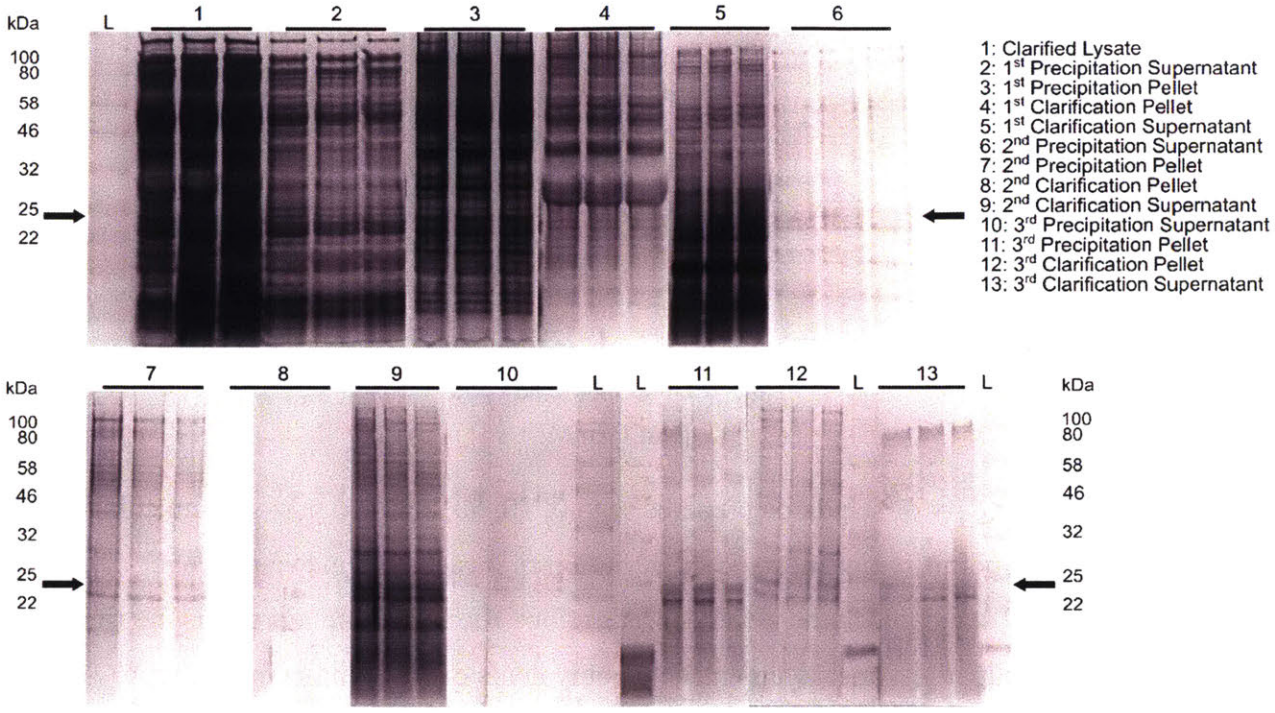


Figure D-11. Silver-stained SDS-PAGE gels showing purification fractions of P10 purification at the well-plate scale. The arrow indicates the apparent molecular weight of P10. Lane group 5 corresponds to final product after one cycle of purification (NaCl); group 9 corresponds to final product after two cycles (NaCl-NaCl); and group 13 to final product after three cycles (NaCl-NaCl-EtOH).

MGWGSASGLVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGV
GVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGV
GVPGVGVPGVGETTSSGSIGTGDRINTVRGPITISEAGFTLTHEHICGSSAGFLRAWPEFFGSRKALAEKAVRGL
RRARAAGVRTIVDVSTFDIGRDVSLLAEVSRAADVHIVAATGLWFDPPLSMRLRSVEELTQFFLREIQYGIEDTG
IRAGIIKVATTGKATPFQELVLRAAARASLATGVPVTTHTAASQRGGEQQAAIFESEGLSPSRVCIGHSDDTDDL
SYLTALAARGYLIGLDRIPFSAIGLEDNASASALLGNRSWQTRALLIKALIDQGYMKQILVSNDWLFGFSSYVTN
IMDVMDSVNPDGMAFIPLRVIPFLREKGVPQETLAGITVTNPARFLSPTLRASLQGSTSGLVGVPGVGVPGVGIP
GVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIP
GVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGETTS

Figure D-12. Amino acid sequence of ELP-OPH-ELP used in this paper.

D.2 Well-plate purification of ELP-tagged organophosphate hydrolase (ELP-OPH-ELP)

Expression and purification of ELP-OPH-ELP. Protein was expressed in Tuner(DE3) *E. coli* cells using the pET28b vector. Starter cultures for protein production were inoculated from a single

colony on an agar plate. A 10 mL starter culture was prepared in Luria broth (LB) supplemented with 50 μg/mL kanamycin and grown to confluence overnight at 37 °C in an orbital shaker. Expressions (252 biological replicates) were conducted in a 96-well plate prepared with a sterilized glass bead in each well (to improve mixing) and containing 1480 μL of terrific broth (TB) supplemented with 50 μg/mL kanamycin. Expressions were inoculated with 20 μL of starter culture. Cells were grown at 30 °C to an average OD_{600} of ~0.5 and then put on ice for 20 minutes prior to induction with 0.5 mM β-D-1-thiogalactopyranoside (IPTG). Expressions proceeded overnight at 24 °C in an orbital shaker at 300 rpm. Cells were collected via centrifugation (4,000 xg, 15 minutes) and subsequent aspiration of media. Cells were then frozen for at least 16 h at -20 °C. Cells were resuspended in 140 μL of buffer containing 3 mM $MgCl_2$, 1 mM EDTA, 100 mM NaCl and 10 mM trizma at pH 7.5 (MENT buffer) supplemented with 1 mg/mL lysozyme and 0.1 mg/mL each of DNase I and RNase A. Resuspended cells were incubated overnight at 4 °C to allow for lysis and degradation of RNA and DNA. Lysate was subsequently refrozen at -80 °C for at least 18 h. Lysates were then thawed at 4 °C and clarified by centrifugation at 6,000 xg at 4 °C for 1 h. Protein purification was performed using one NaCl-induced precipitation followed by one EtOH-induced precipitation. NaCl-induced precipitation was performed using 2.5 M NaCl at 20 °C for 16 h. Higher salt content and lower temperatures were used in this precipitation to avoid thermal denaturation of the OPH based on previous reports on purifying ELP-tagged OPH. [1] EtOH-induced precipitations were performed using 30 vol% EtOH at 4 °C for 1 h.

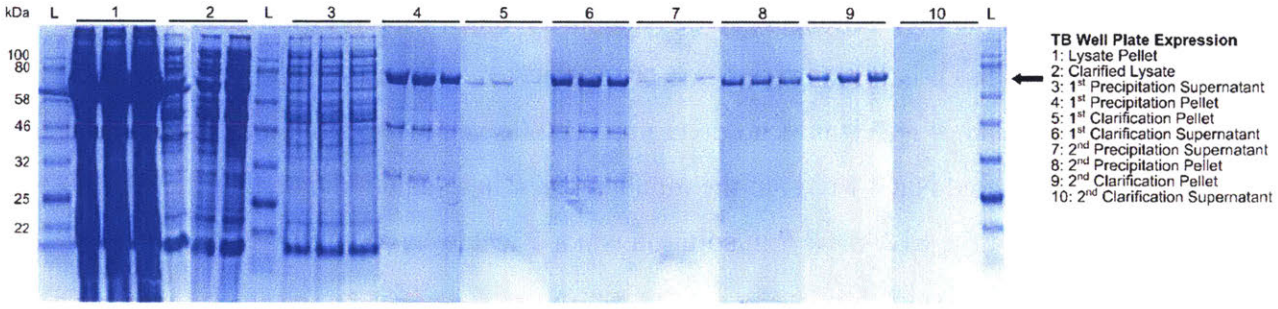


Figure D-13. Coomassie-stained SDS-PAGE gels showing various fractions from two consecutive cycles of NaCl- then EtOH-induced precipitation (NaCl-EtOH method) using organophosphorus hydrolase (OPH) as the ELP-tagged protein of interest. The arrow indicates the apparent molecular weight of ELP-OPH-ELP. Most product loss was observed in the lysate pellet and 2nd clarification pellet. The final product fraction (fraction 10) contains only a faint band at ~70 kDa, indicating that the protein was denatured or otherwise unable to be resuspended after the precipitation with

ethanol. Product loss was confirmed by BCA assay on the final product reporting negligible amounts of protein present (data not shown).

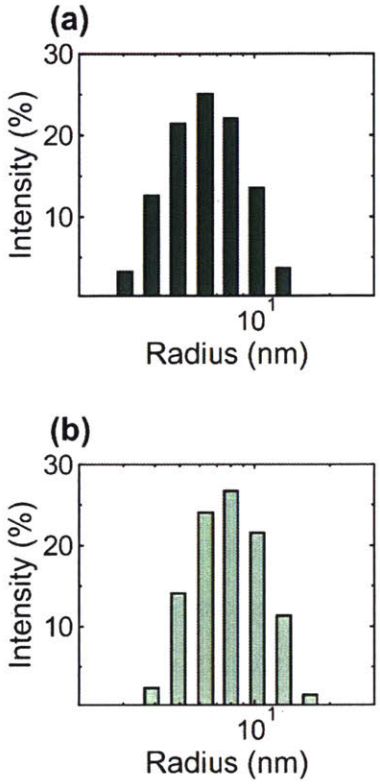


Figure D-14. Dynamic light scattering (DLS) data taken on ELP-sfGFP purified by (a) two rounds of NaCl-induced precipitation (labeled NaCl-NaCl in the manuscript) and ELP-sfGFP purified by (b) one round of NaCl-induced precipitation followed by one round of EtOH-induced precipitation (labeled NaCl-EtOH in the manuscript). These data indicate that the final aggregation state of the ELP-sfGFP in solution is not affected by the purification process.

D.3 References

(1) Shimazu, M.; Mulchandani, A.; Chen, W., Thermally triggered purification and immobilization of elastin–OPH fusions. *Biotechnol. Bioeng.* **2003,** *81* (1), 74-79.

Appendix E. Supporting Information for Chapter 6

E.1 DNA and Amino Acid Sequences

Plasmid map for these constructs can be found in **Figure A-3**.

ELP-rcSso7d.SA-ZE (Engineered rcSso7d)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGTTGGAAGCGGG
AAGACTGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCC
AGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAG
GCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGC
GTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGT
TGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTG
GTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGC
GTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGT
ACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTAC
CTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCT
GGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGAAACCGA
ATTTGTAGAACGGCCGGGTGGTGGTGGTAGCGGTGGTGGCGGATCCATGGCAACCGTGAAATTCACATACCAAGGCG
AAGAAAAACAGGTGGATATTAGCAAAATCAAGATCGTGGCTCGTGACGGCCAGTACATTGACTTTAAATATGATGAA
GGTGGTGGTGCCTATGGTTATGGTTGGGTGAGCGAAAAAGATGCACCGAAAGAACTGCTGCAGATGCTGGAAAAGCA
AGAATTCGGTGGTAGCGGTGGTGGCGGTTCACTGGAGATCGAAGCGGCGGCGCTGGAGCAGGAAAACACCGCGCTGG
AAACCGAGGTGGCGGAGCTGGAACAGGAAGTGCAACGTCTGGAAAACATTGTGAGCCAATACCGTACCCGTTATGGT
CCGCTGTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (412 AA, 38030.5859 Da, Estimated pI = 5.22)

MGSSHHHHHHSSGLVPRGSHMVGSGKTVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPG
VGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVG
VPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVP
GVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGETEFVERPGGGGSGGGGSMATVKFTYQGE
EKQVDISKIKIVARDGQYIDFKYDEGGGAYGYGWVSEKDAPKELLQMLEKQEFGGSGGGGSLEIEAAALEQENTALE
TEVAELEQEVQRLENIVSQYRTRYGPL*

ELP-rcSso7d.neg-ZE (Naïve rcSso7d)

Vector: pET28b; Cloning Site: NcoI/XhoI

DNA Sequence:

CC**ATG**GGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGTTGGAAGCGGG
AAGACTGTTGGTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCC
AGGCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAG
GCGTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGC
GTTGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGT
TGGCGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTG
GTGTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGC
GTACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGT
ACCTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTAC
CTGGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGCGTACCT
GGTGTTGGCGTCCCGGGTGTAGGTATCCCAGGCGTTGGTGTACCGGGTGTAGGCGTTCCAGGCGTTGGTGAAACCGA
ATTTGTAGAACGGCCGGGTGGTGGTGGTAGCGGTGGTGGCGGATCCATGGCAACCGTGAAATTCACATACCAAGGCG
AAGAAAAACAGGTGGATATTAGCAAAATCAAGCATGTGCGTCGTATCGGCCAGTCTATTGCTTTTATCTATGATGAA
GGTGGTGGTGCCCATGGTGGCGGTAACGTGAGCGAAAAAGATGCACCGAAAGAACTGCTGCAGATGCTGGAAAAGCA
AGAATTCGGTGGTAGCGGTGGTGGCGGTTCACTGGAGATCGAAGCGGCGGCGCTGGAGCAGGAAAACACCGCGCTGG
AAACCGAGGTGGCGGAGCTGGAACAGGAAGTGCAACGTCTGGAAAACATTGTGAGCCAATACCGTACCCGTTATGGT
CCGCTGTAACTCGAG

NcoI/XhoI sites; **Start codon**

Amino Acid Sequence: (412 AA, 37798.3594 Da, Estimated pI = 5.88)

MGSSHHHHHHSSGLVPRGSHMVGSGKTVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPG
VGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPG
VPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGVP
GVGVPGVGIPGVGVPGVGVPGVGVPGVGVPGVGIPGVGVPGVGVPGVGETEFVERPGGGGSGGGGSMATVKFTYQGE
EKQVDISKIKHVRRIGQSIAFIYDEGGGAHGGGNVSEKDAPKELLQMLEKQEFGGSGGGGSLEIEAAALEQENTALE
TEVAELEQEVQRLENIVSQYRTRYGPL*