2018

A Synthesis Platform for Temperature Responsive Star Polymers

Schmitt J. Richard Jr.
Cleveland State University

A SYNTHESIS PLATFORM FOR TEMPERATURE RESPONSIVE STAR POLYMERS

RICHARD J. SCHMITT

Bachelor of Science in Chemical Engineering

Cleveland State University

May 2016

Submitted in partial fulfillment of requirements for the degree

MASTER OF SCIENCE IN CHEMICAL ENGINEERING

at the

CLEVELAND STATE UNIVERSITY

May 2018

We hereby approve this thesis for

RICHARD J. SCHMITT

Candidate for the Master of Science in Chemical Engineering degree

for the Department of Chemical and Biomedical Engineering

and the CLEVELAND STATE UNIVERSITY'S

College of Graduate Studies by

_____

Thesis Committee Chairperson, Nolan B. Holland, PhD
Department of Chemical and Biomedical Engineering

_____

Date

_____

Thesis committee member, Kiril A. Streletzky, PhD
Department of Physics

_____

Date

_____

Thesis committee member, Christopher Wirth, PhD
Department of Chemical and Biomedical Engineering

_____

Date

Date of Defense: April 24, 2018

ACKNOWLEDGEMENTS

I would like to thank Dr. Nolan Holland for providing me with the opportunity to study in his laboratory. I appreciate his patience throughout my entire thesis research. His encouragement, questions, and insight have been extremely helpful in my research. I would like to thank my peers for their assistance throughout my time at Cleveland State University.

I appreciate the time my committee members have taken to assist me in creating this document. Thank you for your time.

I would also like to thank Becky Laird and Darlene Montgomery. They are both excellent at what they do and full of encouragement and advice. I appreciate all the help they have provided me during my time at Cleveland State University.

I appreciate the support I have received from both of my parents. They have and always will be encouraging into my endeavors. I appreciate their everlasting support and love.

Finally, I would like to thank my wife Alexandra. Without her encouragement this would not have been possible. Thank you for being there for me and supporting me.

A SYNTHESIS PLATFORM FOR TEMPERATURE RESPONSIVE STAR POLYMERS

RICHARD J. SCHMITT

ABSTRACT

Star polymers are a class of branched polymers comprised of several polymer chains extending from a central point. Star polymers have applications in biopharmaceuticals where they have been proposed to be suitable drug delivery vehicles. Star polymers have traditionally been synthesized through chemical synthesis with added functionality provided by grafting on the arms. This complex synthesis can be simplified by using a biosynthetic approach which enables precise control of molecular weight and composition. This approach is demonstrated using star polymers with arms composed of a temperature responsive protein-based polymer termed elastin-like polypeptide (ELP). Star polymers are characterized based on the number of arms, the length of the arms, and the arm functionality. Previously, an ELP star polymer was synthesized with three arms. Here this work is extended through the synthesis of a six-armed ELP star polymer. The controlled synthesis of a complex six-armed star polymer has not been performed through biosynthesis techniques. The method of recursive DNA assembly used to prepare the gene that codes for the polypeptide has been designed to allow the preparation of genes encoding a range of arm lengths and variable functional end groups. A star polymer with six 19 pentapeptide long arms has been created with this system. This six-armed star polymer temperature dependent phase behavior is similar to the three-armed polymer containing 37 pentapeptides each.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Biopharmaceuticals is a term used to distinguish recombinant proteins and other inherently biological products from traditional drugs. This field is the parent field of recombinant proteins that are enlisted for use in drug treatment. Recombinant proteins have a promising future in pharmaceuticals. In fact, there are currently over 125 approved recombinant proteins for use in the United States and the European Union for this purpose. Recombinant proteins are part of an industry which is responsible for billions of dollars per year [1].

The field of biopharmaceuticals is rapidly advancing. Specifically, technology advances to gene synthesis tools have led to more complex recombinant proteins. Recombinant proteins are favored due to the ease of synthesis and purification. Advances in genetic engineering methods like Golden Gate or Gibson Assembly have allowed complex DNA structures to be created through combination of smaller fragments. These DNA structures are taken into an organism in a process termed transformation, e.g. DNA plasmids can be transformed into chemically competent *Escherichia coli* (*E. coli*) cells. The organism then functions as a factory producing proteins (or polypeptides) of precise size and amino acid sequence that is encoded by the DNA codons. The synthesis of the

entire molecules is completely achieved in the bacteria. Purification of the product from the other bacterial components can be done by various robust methods [2]. These steps can be rather simple when compared to the synthesis and purification in chemical synthesis.

Recombinant proteins provide an alternative approach to multi-step chemical synthesis methods [3]. Chemical synthesis has the capability to produce complex structures, such as, a star polymer. Chemical synthesis methods used to create this structure are multi-step [4]. The arms and the core are combined in two separate reactions. This synthesis method can be simplified using recombinant proteins. The multi-block synthesis is carried out in one step when done through biosynthesis. However, recombinant proteins require the use of an organism for the protein synthesis. These organisms are controlled to synthesize the desired protein through transforming DNA into them. Another complication is achieving the required DNA which encodes for the multi-block structure. This is achieved through a set of DNA assembly techniques. Two of these techniques are Golden Gate and Gibson assembly.

Through Gibson assembly it is possible to join two or more strands of DNA in a single assembly reaction. The digestion step involves the addition of a restriction endonuclease which recognizes a string of nucleotides and digests based on that location. The digestion prepares the two DNA strands for assembly. After the digest, the assembly can occur in its own three step reaction. There is a Gibson exonuclease which chews on each piece of DNA from the 5’ end toward the 3’ end. The exonuclease produces a single stranded 3’ overhanging fragment. If there is a complementary overhang on another strand, they will anneal. The second step is for a DNA polymerase to fill in any gaps in the sequence with

the missing nucleotides. The third step is the ligase step. This step joins the DNA segments and removes any mismatches in the sequence. These three steps occur in the same reaction vessel [5]. Gibson et al. have assembled DNA structures of various sizes using the Gibson assembly technique.

Gibson assembly is possible with two or more strands of DNA. This enables the use of small simple DNA fragments to synthesize large complex DNA structures. The design of these complex structures is typically done with smaller pieces. Once assembled, the construct will contain the desired nucleotides. The Gibson assembly technique provides a basis for multiple overlapping of nucleotides which can be used to create a recursive system of DNA assembly [6].

Golden Gate provides another method for DNA assembly. In Golden Gate, type IIs restriction enzymes are used. The unique characteristic of type IIs enzymes is that the recognition site is not located at the digestion site. The enzyme will digest at a specific length away from the recognition site. The digestion location distance from the recognition site is enzyme specific. Golden Gate can be designed with specific enzymes to allow for appropriate fragments after digestion. The assembly of the DNA fragments is performed simultaneously with the digestion. The enzyme is not deactivated during the experiment. This requires the digestion to eliminate the recognition site from the fragment [7].

The assembly methods discussed allow for the synthesis of many DNA fragments in one reaction. In these reactions, DNA fragments are assembled together and translated into specific protein sequences. These protein sequences can form repetitive structures akin to synthetically created block copolymers. Block copolymers have been sought after

for their applications in drug delivery. A block copolymer must be designed to have amphiphilic characteristics necessary to form micellar systems. Drugs which are typically characterized by poor aqueous solubility will arrange themselves in the oily centers of the micelle systems [8]. To achieve amphiphilic characteristics the block structure would contain a hydrophobic functional group encapsulated by two hydrophilic groups [9].

A simple arrangement which can satisfy the requirements for drug delivery is a block copolymer system. A block copolymer consists of an arrangement of repetitive sequences titled blocks. These blocks are given single letter abbreviations, such as A or B; this letter stands for a specific molecular structure. The molecular blocks can be polymerized in various formations resembling: A – A – B – B, A – B – A, or even A – B – A – A – A – B – A. There are many possibilities in the arrangement of a block copolymer [10].

Block copolymers that form polymer micelles are particularly useful due to their low critical micelle concentration (CMC). This characteristic of block copolymers allows using lower concentrations of the copolymer to achieve larger concentrations of micelles in solution. The study of these solutions is also simplified by their low diffusion coefficient [11].

An example of an amphiphilic arrangement is the A – B – A arrangement, where block B is hydrophobic and block A is hydrophilic. When these polymers are placed in solution there are two arrangements the polymer can take. If the solution contains a copolymer concentration less than the CMC the copolymer will be arranged as block copolymer single unit arrangements in solution. If the solution contains a copolymer concentration greater than the CMC the single units will begin to form micelles.

These systems can contain chemically synthesized blocks or biosynthesized blocks. Biosynthesized blocks will be used due to the complexity of chemically synthesized blocks involving a series of synthesis and purification steps. In the further sections, the materials used in the blocks will be considered and discussed. A specific arrangement of interest involves a star-like structure.

1.1 Synthesis Structure

Star polymers are characterized by linear chains branching from a central point. Recently, star polymers have been an interesting platform in the biopharmaceutical field where they have been shown to be plausible drug encapsulation systems for drug delivery. Star polymers are synthesized to contain an amphiphilic nature. This is due to the hydrophobic branching polymeric arms from a hydrophilic point. Previously, star polymers utilized for drug delivery systems have been synthesized through chemical synthesis techniques [4].

The addition of ELP arms to the star polymer provide a means to space the functional groups from the core and explore ELP aggregation. A ELP experimentation has shown that aggregation occurs as the temperature is increased [12]. This is due to the hydrophobic tendencies of the side chains in the repeated amino acid structures. They drive the water which solubilizes them into a higher entropy state through aggregation above the transition temperature. More specifically, it has been proposed that they form a hydrophobic folded structure consisting of β turns [13]. This behavior is representative of materials which exhibit a lower critical solution temperature (LCST). This phenomenon has been shown to be suitable for drug delivery applications as the external environment can stimulate the response necessary to release pharmaceuticals [14].

Figure 1. Three-Armed Elastin-Like Polypeptide

Previously, a three-armed polymer which exhibits similar behavior in a non-symmetrical shape was synthesized [15]. This structure contains a natural trimer forming peptide sequence called Foldon. Foldon is a domain of the bacteriophage T4 fibritin protein used to stabilize a triple coil. It has been shown that the 27 amino acid Foldon sequence can be used as an artificial trimerization inducer [16]. Foldon causes three peptide chains to fold together. The round core (Figure 2) is the Foldon trimer and the arms represent their respective linear chains. The linear chains are made from an elstain-like polypeptide material. The trimerization of Foldon when bound to elastin-like polypeptide protein sequences was first reported by Ghoorchian et al. [12].

Elastin-like polypeptides (ELPs) are a class of biopolymers characterized by an amino acid sequence consisting of repetitions of (GαGβP) where α is any amino acid and β is any amino acid except proline. ELPs have a temperature dependent reversible phase transition occurring at a transition temperature [17]. Below this temperature the ELP is soluble and above this temperature the ELP forms a dense protein rich coacervate. Increased turbidity of the sample represents this phase transition. The ELP repetitions are arranged together and form sections denoted by $(G\alpha G\beta P)_X$. The X is any whole number of ELP repeats [13].

ELPs have been extensively characterized by their transition temperature. Transition temperature is dependent on the sequence of the protein, the length of the protein, and the protein's concentration in solution. Ali Ghoorchian has studied the relationship of these properties as applied to a $(GVGVP)_x$ structure [17]. Where, G is glycine, V is valine, and P is proline. Ghoorchian studied ELP repeats of length 20, 40, and 60. These were arranged in monomer formations and trimeric formations incorporating the Foldon structure. He found that three linear chains could be brought together to form a trimer with the introduction of the Foldon molecule attached to the end of each chain. Major conclusions resulted which suggested that if the number of the ELP repeats is increased the transition temperature of the ELP decreases. Also, the same trend was found in a trimer arrangement of the protein. This study produced an important model for calculation of transition temperature of similar ELP structures. The model is summarized through the following equations:

$$T_t = T_{cr} - K_c \ln(C/C_{cr})$$

In this equation, K_c is a modeled parameter.

$$K_c = K_t/L$$

Where T_{cr} is the critical temperature, C_{cr} is the critical concentration, k_t is a constant with the units of °C, and L is the number of pentapeptide repeats. Parameters used in these equations are tabulated below.

	Linear (without Foldon)	**Trimer (with Foldon)**
K_t (°C)	114.5	190.8
C_{cr} (mM)	30.0	106.4
T_{cr} (°C)	20.8	18.5

Table 1. Critical Values for Transition Temperature Calculation

These parameters in conjunction with known concentration and length of the pentapeptide repeats will allow an accurate calculation of transition temperature. According to Goorchian, it is necessary to alter this model due to inconsistencies in T_{cr} for different molecular architectures [17]. The volume concentration model utilizes similar constants while eliminating differences in T_{cr}. The volume concentration is calculated in the following manner:

$$C_v = C_m L^n K'$$

$C_m - Molar\ Concentration$

$L^n - Pentapeptide\ Repeats\ to\ the\ nth\ power$

$K' - Fit\ Parameter, allows\ the\ right\ side\ of\ equation\ to\ be\ unitless$

where *n* is associated to the following conformations (Table 2). Using volume concentration, new fit parameters were found for monomer and trimer ELPs (Table 3).

***n* Value**	**Conformation**
1	Spheres
1.5-1.8	Random Coils
2.0	Stiff Coils
3.0	Rod-like

Table 2. Physical Meaning of *n* Value in Volume Concentration Analysis

	Linear (without Foldon)	**Trimer (with Foldon)**
K_t (°C)	119	179
n	1.51	2.30
T_{cr} (°C)	23.0	23.0
K'	3.3×10^{-4}	2.0×10^{-6}

Table 3. Fit Parameters for Volume Concentration Analysis

The three-armed structure was modified by the addition of a functional group to each arm of the ELP chain. A decorin motif was attached to the arms of the three-armed variant in previous work [18]. It is desirable to use this three-armed model and add an additional three arms to create a six-armed structure (Figure 2).

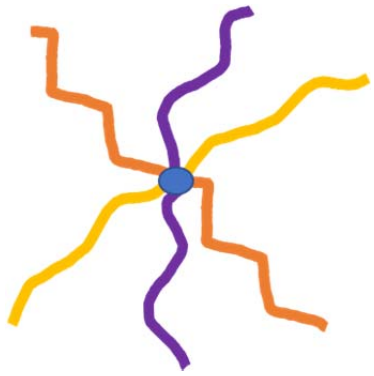


Figure 2. Six-Armed Star Polymer

A six-armed star polymer has advantages over its three-armed counterpart. The additional arms each contain a functional group located at the end. The six-arm variety will allow the functional groups to provide a larger avidity towards the desired bonding substance [19]. In the case of the decorin end group motif, it binds to the type I collagen molecule.

This complex structure will be synthesized in one bacterial expression step. Since it contains ELP, a simple inverse temperature cycling purification can be used to separate the protein from the other materials used in the synthesis. The design of the gene that codes for an elastin-like polypeptide (ELP) containing star polymer and the expression of it are done using standard recombinant methods. Star polymers are typically created using a variety of complex synthesis methods including cross-linking, polymerization, or grafting on arms through a chemical reaction [4]. Using recombinant means for synthesis will eliminate these complex steps and allow the synthesis to proceed in bacteria. The

addition of elastin like polypeptides supplies a temperature dependent phase transition allowing for easy purification of the protein from the cell lysate [19].

The decorin motif is located at the end of each arm. The center of the star polymer requires the foldon trimerizer. For the star polymer to function properly, a protein must be chosen which causes two other protein chains to fold together. The foldon motif follows this specification.

The connection of these two structures will be carried out using elastin-like polypeptide (ELP) arms. ELPs can be characterized by a temperature dependent phase transition. This transition is termed the transition temperature. Above this temperature the ELP forms a dense protein rich coacervate and below this temperature the protein is soluble in the aqueous phase. This temperature is dependent on the number of ELP repeats present in the protein and the concentration of the protein in the aqueous phase. The transition temperature will aid in the purification of the overall protein from the cell lysate through inverse temperature cycling around the transition temperature. The ELP arms will allow this structure to be easily purified and have shown properties suitable for drug delivery applications. An example of a drug delivery theory involving ELPs is the use of local hyperthermia to induce a drug release from a modified liposome which contains ELP [28].

Type I collagen is a highly prevalent protein which is associated with cellular matrix stability. A decorin motif has been shown to bind to type I collagen. Specifically, the motif contains the amino acid sequence of LRELHLNNN [20]. Using this decorin motif as the functional group protein will allow the star polymer to crosslink type I collagen.

1.2 Motivation

It is of interest to create a temperature responsive star polymer which can aid injectable drug delivery solutions by means of bonding with type I collagen. The star polymer will consist a protein-based polymer, namely ELP. It is desired to create the polymer using recombinant methods as discussed before. However, the synthesis of these proteins is dependent on providing the organism with the correct sequence of DNA. The DNA is then processed through transcription and translation for the bacteria to synthesize the desired protein. There are a variety of companies which can provide a desired DNA structure. However, there are also limitations as to what these companies can provide the end user with. Highly repetitive structures cause these companies to run into problems during the DNA synthesis. A simple DNA structure must be ordered from the company and then the assembly steps discussed will be used to expand this DNA structure into the desired sequence to code for translation into the protein of interest.

Elastin-like polypeptides have previously been created and expressed through the means of digestion and recursive directional ligation (RDL) [17]. However, RDL has limitations in the required composition of the proposed DNA for ligation to the parent. The digests require type II enzymes to allow for recursive assembly. Gibson assembly has been shown to be more robust and requires fewer steps to form the desired DNA product [18].

This thesis describes a platform for recursive assembly of a star polymer using Gibson assembly. This platform is designed such that different functional groups can be added to the ends of the ELP arms and the length of the arms can be systematically increased to the desired number of ELP repeats. The method is demonstrated with a six-armed ELP

with a collagen binding (decorin) functional group at the end of each arm. The protein synthesized is characterized using the model proposed by Ali Ghoorchian [17].

CHAPTER II

MATERIALS AND METHODS

The elastin-like polypeptide arms for the three-armed star polymers reported by Deyling were produced in various sizes, namely, 19, 37, and 73 pentapeptide repeats [18]. These three-armed structures exhibit phase transition characteristics of ELP while having the ability to bond or interact with additional substances through the functional groups located at the end of each arm. In this case, the functional group is decorin, which is a collagen binding protein sequence. The decorin motif was chosen due to their association with collagen which is abundant in the extracellular matrix [18]. To improve this system, additional arms would provide an increase in avidity for the functional group bonding sites. A method to synthesize a six-armed star polymer with functional groups attached for characterization is described here.

2.1 Recursive System for Star Polymer Synthesis

Acquiring custom DNA is a simple process through a large variety of corporations. These companies can create the DNA to code, in biological translation, for a star polymer. However, they are limited in the ability to create highly repetitive DNA sequences found within pentapeptide repeats of elastin-like polypeptides. Companies will charge a complexity fee and further indicate they are unable to create the desired

structure due to the GC rich nature of (GVGVP) pentapeptides. Therefore, it is necessary to produce a general platform for creating desired sizes of the star polymers with the desired functional groups. A general platform will allow an increase in the size of the protein without reaching a boundary of inaccessibility noted by the outside companies.

The system design contains ELP segments of a starting size of ten pentapeptide repeats illustrated by the term $(GVGVP)_{10}$. Therefore, the structure is:

Functional Group – $(GVGVP)_{10}$ – Foldon – $(GVGVP)_{10}$ – Functional Group

There are specific regions in the gene where a recursive process of replication occurs to allow an increase in size. The system operates through a series of digests by restriction enzymes (or endonucleases). This is followed by Gibson assembly to provide the recursive duplication. The platform is designed to allow it to be digested and reassembled such that it nearly doubles the length of the pentapeptide regions. The initial gene that codes for a starting length of ten ELP units on each side of the foldon was ordered from Invitrogen.

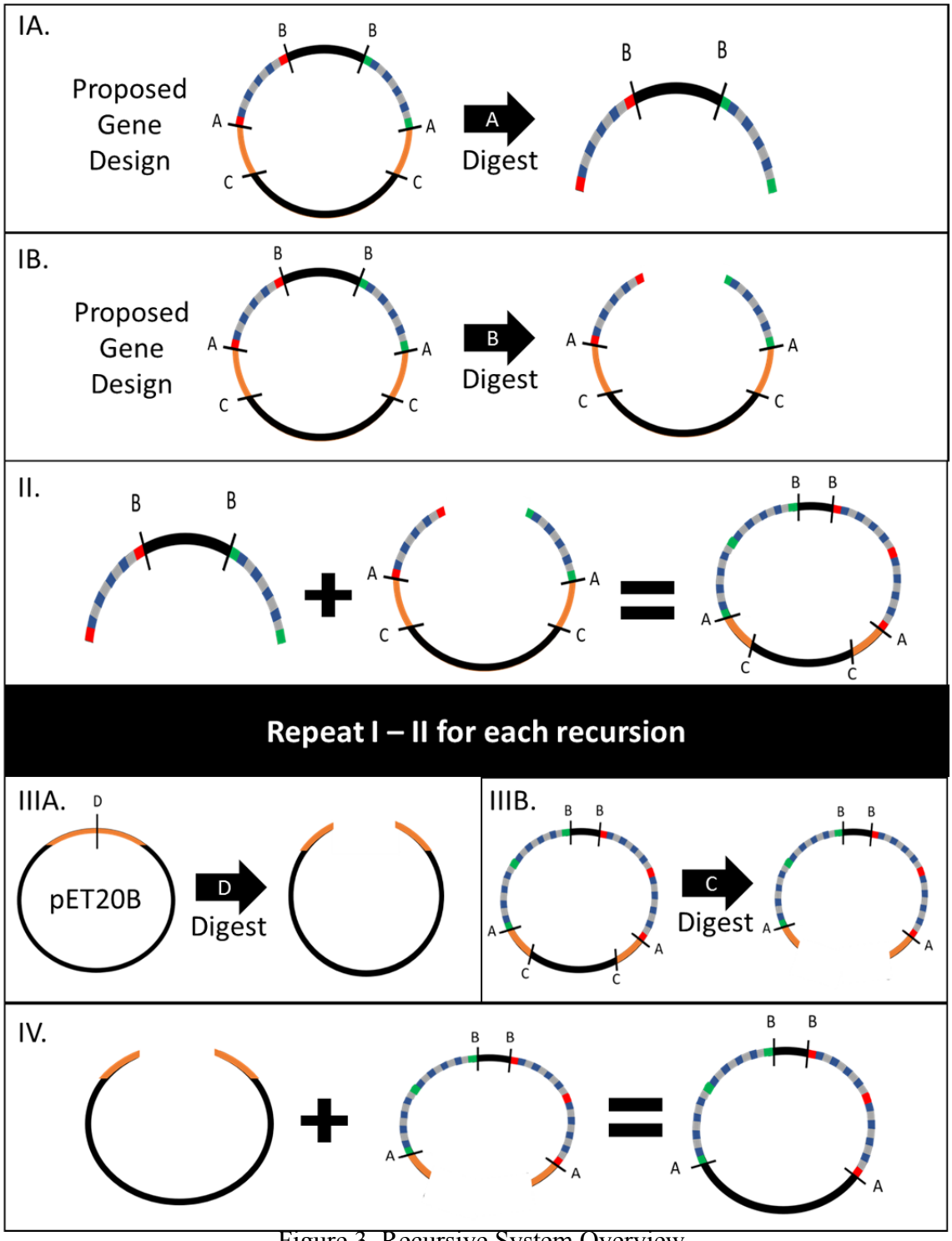


Figure 3. Recursive System Overview

The process of recursive DNA assembly (Figure 4) is used to produce the gene for star polymers of varying length. The initial gene was designed and ordered from Invitrogen in

pMX, a vector commonly used by Invitrogen. These are cloning vectors lacking the promoters present in expression vectors. The original vector received from Invitrogen contained restriction endonuclease recognition sites indicated by A, B, and C (Figure 4). The C represents the digestion required to cut the gene sequence from the pMX vector. This digestion is the final digestion used to create a fragment for assembly with an expression plasmid. The required nucleotides that overlap with the pET expression vector for Gibson assembly reside between the A and the C cut sites. The DNA fragment is assembled with a digested pET-20b (Figure 4-IIIA). The sections between A and B are the 10 ELP pentapeptide repeats. However, the DNA located in the red sections and the green sections are the same to their respective colors. This is important for the Gibson assembly process. The Gibson assembly relies on matching sequences to assemble after the digestion. The section between the B digestion sites codes for the foldon.

The first digestion uses endonuclease A cutting the plasmid into two fragments (Figure 4 IA). The section including the plasmid ring (black ring) is inactive in the subsequent stages because its ends do not overlap with any other fragments. The important segment to consider is the section coding for the ELP segments and the foldon, which is the insert in the subsequent assembly.

A different digestion of the original vector is used to generate a template fragment that will have the insert assembled into it (Figure 4 IB). This template is one of the two fragments resulting from the digestion of the vector with endonuclease B. The unused fragment that encodes only the foldon again does not have common overlap and so is not active in the Gibson assembly. The larger segment consists of the pMX, the pentapeptide

repeats, and the pET-20b overlaps. The larger chain is displayed because it is the necessary fragment for use in further steps.

The template DNA fragment and the insert DNA fragment are assembled to form a circular plasmid. The assembly is made possible through the green and red sections of the template ELP being the same nucleotide sequence as the incoming insert DNA sequence. Assembly requires overlap of matching ends of the DNA, therefore, this is the only assembly which can occur (Figure 4-II). After this assembly there are 19 pentapeptide repeats including three green sites and three red sites. However, only two of the red sites and two of the green sites are available for assembly in the next recursive step. These sites are available due to their proximity to the enzyme recognition sitesrequired to open the ends for Gibson assembly. The second step following this approach will produce a gene with 37 pentapeptide repeats.

These digests and subsequent assemblies can be repeated multiple times to achieve a desired number of pentapeptide repeats. The number of pentapeptides (*L)* of the assembled gene will be defined by the number of recursions (*n)* and the number of pentapeptides in the original gene (*N)*.

$$L = 2^n N - 2^n + 1$$

For our system with N equal to 10, the length of the arms pertaining to 0 to 3 recursions are given in Table 4.

Recursion	Pentapeptide Repeats
0	10
1	19
2	37
3	73

Table 4. Pentapeptide Repeats Based on Recursion Count

The pMX region is shown by the black section between the A's (Figure 4). The section for Gibson assembly with the expression vector pET-20b is highlighted again for consideration. Section IIIA, section IIIB, and section IV are essential to moving the gene into the pET vector. As a side note, the decorin functional group is located between the orange section and the A digestion sites. It has not been shown due to having no effect on the overall system assembly.

A new template-insert assembly system is defined for the insertion of the desired gene into the pET-20b vector (Figure 4 IIIA). This vector is digested with D to expose the overlaps for assembly with the insert. The insert DNA segment is created by digesting the pMX vector with enzyme C (Figure 4 IIIB). The overlap (orange) of these two fragments allows for their assembly. The assembled vector contains the final gene sequence ready for protein expression (Figure IV).

This gene contains the digestion sites A and B. It is still plausible to use the recursive system in the pET-20b vector. However, pMX is a cloning vector which has a larger plasmid copy number. Therefore, it is best to reserve the pET-20b only for the assembly of the final gene for expression.

2.2 Genetic Design Details

The system described above gives a general description for the design of the system. Here we describe the specific system used to demonstrate that the system can work. It is important to note that the encoded protein is bounded by a start codon and a stop codon. The common sequences referred to in this work are provided in single letter amino acid abbreviations (Table 5). The linear ELP forms the star polymer arms, the foldon forms a trimer, and the decorin is a collagen binding peptide.

Name	Amino Acid Composition
Linear ELP	$(GVGVP)_x$, where x is a number 10-73.
Foldon	GYIPEAPRDGQAYVRKDGEWVLLSTFL
Decorin	MPHLRELHLNNNKL

Table 5. Important Amino Acid Sequences

To allow for the digestion of the plasmids for recursive Gibson assembly, regions of DNA between the coding sequences need to contain recognition regions for the endonucleases to digest. As these sites are added to the gene, the DNA is adjusted such that the codons remain in frame; sometimes requiring the addition of bases leading to additional amino acids.

Decorin $(GVGVP)_{10}$ Foldon $(GVGVP)_{10}$ Decorin

The gene is first constructed without the endonuclease recognition sites. The recognition sites corresponding to the digest locations are added after the desired gene is formulated. Each enzyme digests at a specified recognition site. These enzymes are positioned as indicated by their respective names:

Decorin (AvaI) $(GVGVP)_{10}$ (BseYI) Foldon (BseYI) $(GVGVP)_{10}$ (AvaI) Decorin

The recursive functionality of the arm expansion is driven by the recognition site locations. However, the gene is required to contain this functionality combined with the ability to assemble into the pET-20b vector because the gene is delivered in a pUC variant. This means an additional design step must be taken. A 15-25 nucleotide overlap is added at the beginning and end to allow for the Gibson assembly into the pET plasmid:

(AgeI) pET Overlap Decorin (AvaI) $(GVGVP)_{10}$ (BseYI) Foldon (BseYI) $(GVGVP)_{10}$ (AvaI) Decorin pET Overlap (AgeI)

The protein contains large pentapeptide repeats represented by $(GVGVP)_x$. The (GVGVP) is rich in GC content which many DNA production companies have trouble

creating. The best way around this is to design this segment to start at a smaller 10 pentapeptide repeats. Repetition also adds to the complexity of the gene, so varying the codons used for the amino acids to make each segment unique with respect to the other segments. The goal of the DNA design is to create something unique while maintaining a low GC content. The completed DNA sequence and the respective amino acids are shown. The highlighted areas indicate the endonuclease recognition sites. The underlined areas are the overlaps for Gibson assembly.

```
ACCGGTTTAAGAAGGAGATATACATATGCCTCATTTACGTGAATTACACTTGAATAATA
ATAAACTCGGGGTAGGGGTGCCGGGAGTCGGGGTACCAGGTGTCGGTGTTCCTGGTGTC
GGTGTACCAGGTGTCGGTGTACCAGGTGTGGGTGTACCAGGTGTAGGTGTACCAGGTGT
GGGTGTACCAGGTGTTGGTGTTCCTGGGGTAGGGGTGCCGGGAGTCGGCTGGGGGTACA
TCCCGGAAGCTCCGCGTGACGGTCAAGCTTACGTTCGTAAAGACGGTGAATGGGTTCTG
CTGTCTACCTTCCTGGGCTGGGGTGTTGGTGTCCCCGGTGTCGGTGTCCCAGGTGTCGG
TGTTCCTGGTGTCGGTGTGCCAGGTGTCGGTGTGCCAGGTGTGGGTGTGCCAGGTGTAG
GTGTCCCAGGTGTGGGTGTACCAGGTGTAGGGGTACCGGGTGTTGGTGTCCCCCCGGGT
CCTCATTTACGTGAATTACACTTGAATAATAATAAACTATAAGGCCGTGTAGGCTGAAA
TACCTGCTGCCGACCGGT
```

```
                                    M  P  H  L  R  E  L  H  L  N  N
N  K  L  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V
G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V
G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  W  G  Y  I
P  E  A  P  R  D  G  Q  A  Y  V  R  K  D  G  E  W  V  L  L
S  T  F  L  G  W  G  V  G  V  P  G  V  G  V  P  G  V  G  V
P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V
P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  P  G  P  H
L  R  E  L  H  L  N  N  N  K  L  X
```

Figure 4. Designed Gene, DNA (Top) Amino Acids (Bottom)

Creating longer ELP segments is possible by first digesting the foldon coding region out of the gene with BseYI (enzyme B in Figure 4). The digestion occurs in 10 µL reactions. The reaction volume includes 3 µL supplied NEB buffer, 1 µL template DNA (30 ng/µL), and 0.25 µL enzyme. These are placed in a PCR tube and a thermocycler incubates the vessel for an hour at 37 °C followed by deactivation of the enzyme at for 20

minutes at 80 °C. BseYI recognizes the site CCCAGC which has been placed to digest out the region coding for the foldon, leaving exposed ends of the ELP segments of the template gene. The overlaps designed for the assemblies (Figure 5) are denoted by an underline on the segment. The ellipsis contained in the DNA fragment indicates a break for brevity in the sequence.

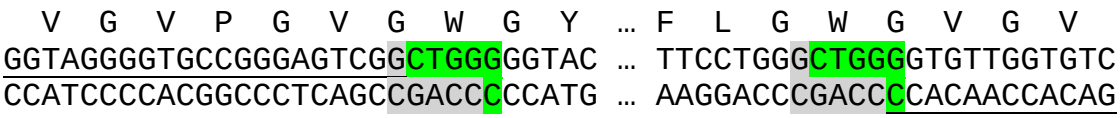


Figure 5. Gene Template Displaying BseYI Recognition Site

An insert that extends the ELP regions and replaces the foldon and restriction enzyme recognition sites is obtained by taking a second identical gene and digesting it with AvaI (enzyme A in Figure 4) for an hour at 37 °C. AvaI digests a recognition site of CYCGRG where Y is T or C and R is A or G. The highlighted regions correspond to an AvaI digest occurring between the different colored boxes as shown by the spaces (Figure 6). This type of cut leaves behind a 5' overhang. This overhang is useful because the Gibson exonuclease removes DNA from the 5' end, eliminating all but one nucleotide of the enzyme recognition sequence.

Notice the difference in the 5' end of the DNA from Figure 6 to Figure 7. The nucleotides on the 5' end from before have now been removed. The ellipsis again indicates a break in the sequence; in this break there is a continuation of GVGVP codons on both ends of a foldon sequence. The overlaps in both instances of the gene fragment are displayed with the underline. The red overlap from Figure 4 is indicated by the left most underlined sequence. The green overlap is indicated by the rightmost overlap. The

underlined sequences must be complimentary such that the overlap of the insert matches up with the template (Figure 5).

```
      G  V  G  V  P  G  V  G  …  V  G  V  P  G  V
C  TCGGGGTAGGGGTGCCGGGAGTCGGGG…GGTAGGGGTGCCGGGAGTCGG  CTGGG
GAGCC  CCATCCCCACGGCCCTCAGCCCC…CCATCCCCACGGCCCTCAGCCGACC  C
```

Figure 6. ELP-Foldon-ELP Showing All Nucleotides after Digest

Following this digest, the recursive assembly occurs to insert the template through a Gibson assembly. The Gibson assembly begins with the Gibson exonuclease chewing back on both the 5' ends of the template and insert (Figure 6). The reaction occurs in a PCR tube. The reaction volume is 25 µL. Gibson assembly master mix is 12.5 µL, 2 µL insert DNA, and 1 µL template DNA. DNA concentration are approximately 3 ng/µL for the assembly. The reaction occurs for 15 minutes in a thermocycler set to 50 °C.

```
5′                                GG…GGTAGGGGTGCCGGGAGTCGG    5′
3′       CCATCCCCACGGCCCTCAGCCCC…                             3′
```

Figure 7. ELP-Foldon-ELP After Gibson Exonuclease Step

The template DNA fragment has complimentary overlaps to the insert fragment. These overlaps allow assembly into the open structure of the template. The single overlap of (GVGVP) creates an addition of nine pentapeptides $(GVGVP)_9$ to the template. This is the first recursion that extends the star polymer to $(GVGVP)_{19}$ on both sides of the foldon.

The recursive steps occur in a cloning plasmid (pMX). Unfortunately, pMX plasmids are useful for cloning and not protein synthesis. A pET plasmid is used for expressing the protein. The DNA is removed from the pMX plasmid and placed into a pET plasmid. There is a recognition site outside of the gene that removes the gene from the pMX and leaves behind overlap for a pET assembly. This recognition site is AgeI or C in Figure 4. The AgeI recognition site recognizes the string sequence ACCGGT, which is placed such

that it leaves a sequence which overlaps with the pET vector digested with NdeI (Figure 8).

```
     5′(pMX)ACCGGTTTAAGAAGGAGATATA 3′
     3′(pMX)TGGCCAAATTCTTCCTCTATAT 5′
5′ (pMX)A      5′ CCGGTTTAAGAAGGAGATATA 3′
3′ (pMX)TGGCC  3′      AAATTCTTCCTCTATAT 5′
```

Figure 8. 5' End of Gene Before (Top) and After (Bottom) AgeI Digest

In this separate reaction, the enzyme separates the pUC plasmid from the desired gene. Once released, the gene undergoes the Gibson exonuclease step to appear as in Figure 9.

```
5′           GATATA 3′
3′AAATTCTTCCTCTATAT 5′
```

Figure 9. 5' End of Plasmid after Gibson Exonuclease

The pET plasmid is digested with NdeI, D in Figure 4. This enzyme opens the circular plasmid into a linear structure leaving overlaps corresponding to the outside of the assembled gene after the AgeI digest. A similar reaction happens to the opposite side of the gene to prepare for insertion into the pET plasmid.

2.3 Synthesis

Once the gene is in the expression vector pET-20b, it is necessary to transform the DNA into BL21* *E. Coli* bacteria for expression. BL21* cells (Fisher Scientific) were used for transformation. The transformation step is detailed elsewhere [18]. After the transformation step, the bacteria are grown overnight on an agar plate containing lysogeny broth with ampicillin. The plate is incubated at 37°C in an incubator. Four to six colonies are picked to inoculate autoclaved lysogeny broth containing ampicillin. The lysogeny broth media is created by mixing 10 g peptone, 5 g NaCl, and 5 g yeast extract with 1 L of water and autoclaving the mixture and containing vessel. The lysogeny broth

is mixed with 100 mg/μL ampicillin and the resulting solution is termed LB Amp. The media aliquoted in 10 mL amounts for growth. The growth occurs in a shaking incubator at 37°C and 200 RPM overnight. The DNA is then purified through a QIAprep Spin miniprep (Quiagen). The purified DNA is amplified through polymerase chain reaction (PCR) and analyzed through agarose gel electrophoresis to verify the number of base pairs with the expected number of base pairs. Q5® High-Fidelity 2X Master Mix, T7 promoter, and T7 terminator primers were used in the PCR reaction. The PCR master mix has a standard protocol incorporating a custom annealing temperature based on the primers used. The annealing temperature used for the T7 primers was 58° C. These conditions were used along with an Invitrogen E-Gel® iBase® system using a 2% agarose E-gel. The DNA was compared to a standard. A calculated number of base pairs was used as a reference point. If the number of base pairs from the gel match the number of base pairs calculated, the sample will be sent for sequencing (Eurofins). Once verification has been completed, a frozen stock of the *E. coli* is made for permanent storage. The frozen stock is created by mixing a 500 μL aliquot of an overnight culture with 500 μL of 50% glycerol and flash freezing in a dry ice bath. This frozen stock is used to inoculate future cultures of the bacteria containing this DNA.

A 2 L Erlenmeyer flask is filled to 1 L with LB and it is autoclaved. Ampicillin will be added upon cooling to below 55°C. Protein synthesis begins with pipetting 5 mL of LB amp into two 10 mL culture tubes. Two culture tubes are inoculated with a sterile pipette tip gently scraped against the frozen stock. These culture tubes are placed in an incubator shaker where it is incubated at 37°C and shaken at 200 RPM overnight.

The next day, the contents of the two culture tubes are poured into the Erlenmeyer flask. The flask is placed into the incubator shaker under the same conditions; the rotation speed was increased to 250 RPM to produce quicker growth. The incubation continues until an optical density of 0.8 is reached.

At an optical density of 0.8 the culture is induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) to cause expression of the gene. After induction, the flask remains in the incubator at 37°C and 250 RPM. After a 5 h wait, the flask is removed from the incubator shaker. The contents of the flask are poured into 500 mL centrifuge tubes. These tubes are centrifuged at 5000 x g for 45 min. Following centrifugation, the bacterial pellet is separated from the supernatant, which contains bacterial secretions and the remaining media. After this separation, the pellet is frozen overnight for storage. Phosphate-buffered saline (PBS) is used to resuspend the cells. Approximately 40 mL of phosphate buffered saline is used to resuspend the cells. The mixture is transferred to a 50 mL centrifuge tube. The centrifuge tube is positioned in ice water using a ring stand and utility clamp to prepare for sonication. The solution is sonicated using pulse sonication provided by a Fisher Scientific 550 Sonic Dismembrator set to a power level of 10. There are ten cycles of 10 s on and 10 s off to allow temperature equilibration. The sonication process breaks the cell membrane releasing the cell contents, including the desired protein.

The resulting solution is processed through inverse temperature cycling to purify the protein. The solution is first chilled on ice to facilitate the phase transition and allow the protein to solubilize into the water. A cold centrifuge cycle is used to separate the dense bacterial waste from the water containing the protein. The cold centrifuge cycle proceeds

at 14000xg and 10 °C for 30 minutes. The supernatant is decanted into a new 50 mL centrifuge tube. The new centrifuge tube is placed into an incubation oven set to 60 °C. After 30 minutes, the falcon tube becomes turbid signifying the phase transition of the ELP. If the solution is limpid, 10x PBS may be added in one mL increments to force the aggregation of the protein. This addition of PBS is necessary for the 6-armed star structures. The ELP will cause the entire protein to form a dense coacervate. The centrifuge is heated up to its max temperature of 40 °C and the centrifuge tube is centrifuged for 5 minutes at 14000xg. The supernatant is discarded from the system. Water is added to the coacervate to create a final volume of 15 mL; For this second instance a smaller volume is used to increase the overall concentration of protein. The tube is placed on ice and the solution is vortexed until there are no, or minimal visible solids located in the solution. This completes one cycle of inverse temperature transitioning of the protein. This cycle is repeated once more to complete purification. Additionally, a cold cycle is used to check for and eliminate any dense waste components still present. These cycles can continue if more concentrated protein is required or if the solution contains components causing coloration of the limpid purified protein. The purified protein is pushed through a syringe filter to sterilize and eliminate remaining particulates.

2.4 Characterization

Samples may now be prepared for characterization steps. To characterize the protein, it is important to determine the concentration of the protein solutions. An absorbance measurement is used to quantify protein concentrations. The concentration of the protein is found through the Beer-Lambert relationship. To determine the protein concentration, a

2000 µL sample of PBS is placed into a cuvette. This sample is used to create a zero-absorbance reference point for 280 nm wavelength light in a Biomate3 system. After referencing the system, the cuvette is emptied and a protein is diluted 20x by mixing 100 µL protein with 1900 µL PBS is placed into the same cuvette. The absorbance of the protein solution is measured and converted to protein concentration based on the Beer-Lambert Law:

$$A_{\lambda 280} = \epsilon l c$$

where,

$A_{\lambda 280} - Absorbance\ at\ 280\ nm\ Wavelength$

$\epsilon - Extinction\ Coefficient$

$l - Solution\ Path\ Length$

$c - Concentration\ of\ Solution$

The extinction coefficient is a calculation based on the sequence of the protein used in the study. It is dependent on the number of three specific amino acids present in the protein [21]. These amino acids are tryptophan, tyrosine, and cystine. They have an extinction coefficient of 5500, 1490, and 125 M^{-1} cm^{-1}, respectively [23]. These amino acids are only present in and surrounding the foldon sequence. They are independent of the size of the arms. The extinction coefficient will always be calculated from three tryptophans, two tyrosines, and zero cysteines. This equates to 19,480 M^{-1} cm^{-1}. The cuvettes used are a 1 cm solution length for the laser to go through. With these values, the measured absorbance can be used to calculate the concentration in of the solution.

The value found is in molarity; it is used to calculate the volume of protein solution required to load two ng/well of protein on an acrylamide gel for SDS-PAGE. To verify the protein's molecular weight, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is performed. It is desired to load 2 ng/well of protein on the electrophoresis gel. This protein mass will produce distinguishable bands. The molecular weight can be predicted through an online calculator. This molecular weight will be compared to the molecular weight received through SDS-PAGE. The SDS-PAGE used a Fisher Brand SeeBlue^{TM} Plus2 Pre-stained Protein Standard for molecular weight comparison. A detailed description of the marker standard for the SDS-PAGE experiment can be found online. NuPAGE MES was the running buffer for the experiment.

A characteristic feature of ELP systems is their inverse temperature phase transition, which can be observed by measuring turbidity as a function of temperature. The turbidity can be observed through a scattering experiment with a UV-1800 (Shimadzu) equipped with temperature control and software to create temperature-adsorption curves. These curves are observed at a fixed 350 nm wavelength to minimize absorption and identify the aggregation point of the ELP molecules.

Protein solutions were prepared over a range of concentrations, namely, 1 μM, 15 μM, 25 μM, 50 μM, and 75 μM in phosphate buffered saline (PBS). Temperature is the independent variable for these samples and absorbance is the dependent variable. In physical terms, the light scattered by the turbid sample is used to indicate that the ELP underwent a phase transition forming small coacervate droplets. These systems contain a relatively rapid increase in light scattering as the temperature reaches the transition

temperature. Once, the temperature exceeds the transition temperature the test is completed.

The transition temperature of these systems is determined from the onset of turbidity, which is determined by the X-intercept of the maximum slope of the transition curve (Figure 10). The region of maximum slope is denoted by the black line through the curve. A three-point method was used to determine the region of maximum slope [22]. The circle indicates the maximum X-intercept, which corresponds to the transition temperature.

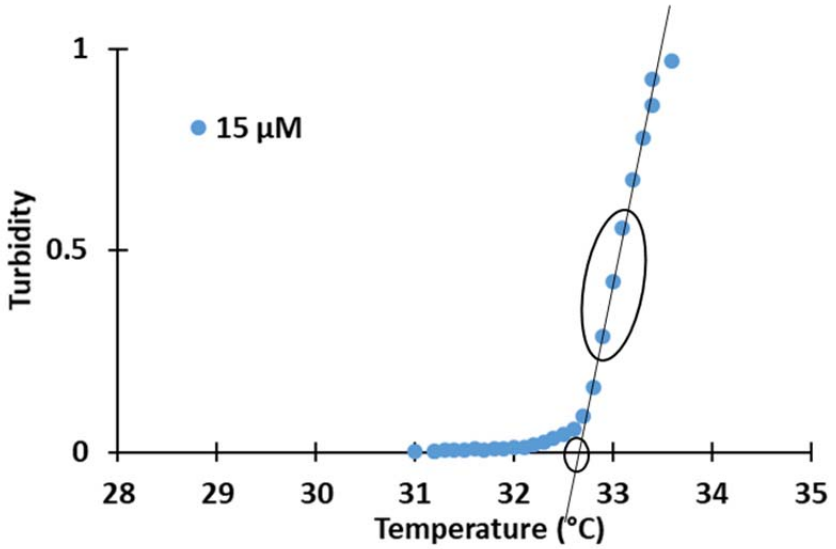


Figure 10. Example of Transition Temperature Measurement

CHAPTER III

RESULTS AND DISCUSSIONS

A recursive method was previously described for generating elastin-like polypeptides (ELPs) of varying lengths using Gibson assembly [18]. The goal of this work was to demonstrate that a six-armed star polymer could be constructed using the same approach to extend within a single gene two different encoded ELP chains simultaneous. A single recursion of such an assembly was performed to create the gene which encodes a six-armed star polymer with a 19 pentapeptide arm length. The final gene assembly was verified by DNA sequencing. The system was shown to express protein which was successfully purified. SDS-PAGE results indicated that the protein product was the appropriate molecular weight but was not able to show that it formed a trimer. However, transition temperature measurements indicate that the product trimerized to form a six-armed star polymer.

3.1 Genetic Design Identity Verification

The procedure in previous sections was used to initiate the creation of the required DNA sequence. The resulting DNA was purified through a mini-prep and an agarose gel electrophoresis was performed to determine the number of base pairs. The agarose gels

require a DNA ladder to be prepared in the marker lane. The ladder used is an E-gel 1 kb plus DNA ladder.

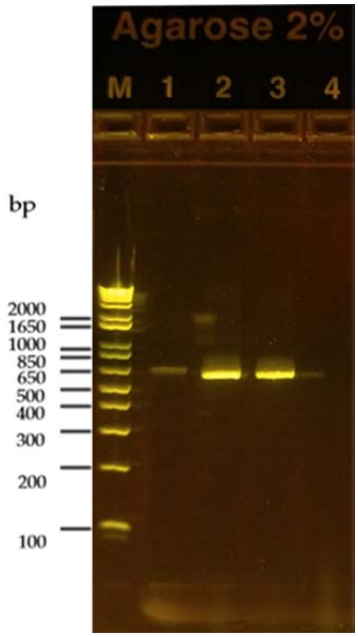


Figure 11. Agarose Gel Electrophoresis of Resulting DNA

The marker is in well M. The marker consists of DNA with base pair counts configured in an exponential spacing pattern. Wells 1 through 3 contain the PCR amplified DNA of the assembled DNA construct.

The electrophoresis results displayed exhibit the number of base pairs of the DNA assembly. The base pair count is approximately matched to the marker line which corresponds to 650 base pairs (bp). The number of theoretical base pairs can be calculated by reviewing the design specifications of the assembly. The projected base pair count for the star polymer containing 19 pentapeptides on each side of the foldon is 800 bp. This differs from projected by 150 bp. However, the next marker is 200 bp away from the 650 bp marker. It was decided to proceed to DNA sequencing due to the band being relatively close to the location of the expected band. The DNA was sent out for

sequencing where a forward sequence and a reverse sequence were primed with primers which primed slightly outside of the pET-20b overlaps. The forward sequence was able to verify the decorin motif at the front of the DNA (Figure 13). However, the quality of the sequence was inadequate to show the remainder of the sequence.

```
     M  P  H  L  R  E  L  H  L  N  N  N
CATATGCCGCATCTACGCGAGTTACATCTGAATAATAAC
     K  L  G  V  G  V  P  G  V  G  V  P
    AAACTCGGTGTTGGTGTTCCTGGAGTCGGTGTCCCA
```

Figure 12. Verification of Decorin Using Forward Sequence

There are 17 remaining pentapeptide repeats in the section before the foldon. The reverse sequence was sufficient to verify the remainder of the DNA. The reverse compliment of this reverse sequence was taken to view the DNA in the correct perspective. The pentapeptide repeats located before the foldon are verified through the sequence information.

```
 G  V  G  V  P  G  V  G  V  P  G  V  G  X  P  G  V  G  V
GGAGTCGGTGTCCCAGGTGTNGGAGTGCCGGGAGTAGGAGNCCCCGGCGTTGGTGTG
 P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G
CCAGGGGTGGGAGTACCTGGAGTTGGGGTTCCAGGCGTAGGCGTTCCGGGGGTCGGA
 V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V
GTTCCGGGTGTTGGTGTTCCTGGAGTCGGTGTCCCAGGTGTTGGAGTGCCGGGAGTA
 G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G
GGAGTCCCCGGCGTTGGTGTGCCAGGGGTGGGAGTACCTGGAGTTGGGGTTCCAGGC
 V  G  V  P  G  V  G  V  P  L  G  Y  I  P  E  A
GTAGGCGTTCCGGGGGTCGGAGTTCCGCTGGGGTACATTCCAGAAGCG
```

Figure 13. Sequenced Pentapeptide Repeats overlapping Foldon.

The prolines (P) can be counted to determine the repetitions of the pentapeptide. Figure 14 contains 17 of the repeats and Figure 13 contains two of the pentapeptide repeats. The presence of the region of overlap allows the two sequences to verify the

DNA assembly of the desired structure. Figure 14 contains the remaining nucleotides and the amino acid representatives.

```
 G  Y  I  P  E  A  P  R  D  G  Q  A  Y  V  R  K  D  G  E
GGGTACATTCCAGAAGCGCCTCGTGATGGTCAGGCCTATGTCCGCAAAGACGGCGAA
 W  V  L  L  S  T  F  L  G  W  G  V  G  V  P  G  V  G  V
TGGGTCCTCCTGAGCACCTTTCTGGGCTGGGGAGTGGGGGTACCCGGTGTCGGCGTA
 P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G
CCTGGTGTGGGAGTCCCAGGGGTCGGCGTGCCAGGCGTCGGGGTCCCTGGGGTAGGT
 V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V
GTACCGGGAGTGGGTGTGCCTGGCGTTGGCGTCCCCGGTGTAGGGGTTCCCGGAGTG
 G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G
GGGGTACCCGGTGTCGGCGTACCTGGTGTGGGAGTCCCAGGGGTCGGCGTGCCAGGC
 V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P
GTCGGGGTCCCTGGGGTAGGTGTACCGGGAGTGGGTGTGCCTGGCGTTGGCGTCCCC
 G  V  G  V  P  G  V  G  V  P  P  G  P  H  L  R  E  L  H
GGTGTAGGGGTTCCCGGAGTGGGGGTACCCCCGGGTCCGCATTTGCGTGAACTGCAC
 L  N  N  N  K  L  STOP(x2)
CTGAATAACAATAAGCTGTAATAA
```

Figure 14. Sequenced Pentapeptide Repeats for Foldon through Stop

For this first recursion the platform was found to be ineffective due to a second location of the recognition site. By using enzyme B alone, the backbone of the plasmid is digested causing reassembly to be impossible. The actual progression used to create the six-armed 19 pentapeptide star polymer is displayed (Figure 15).

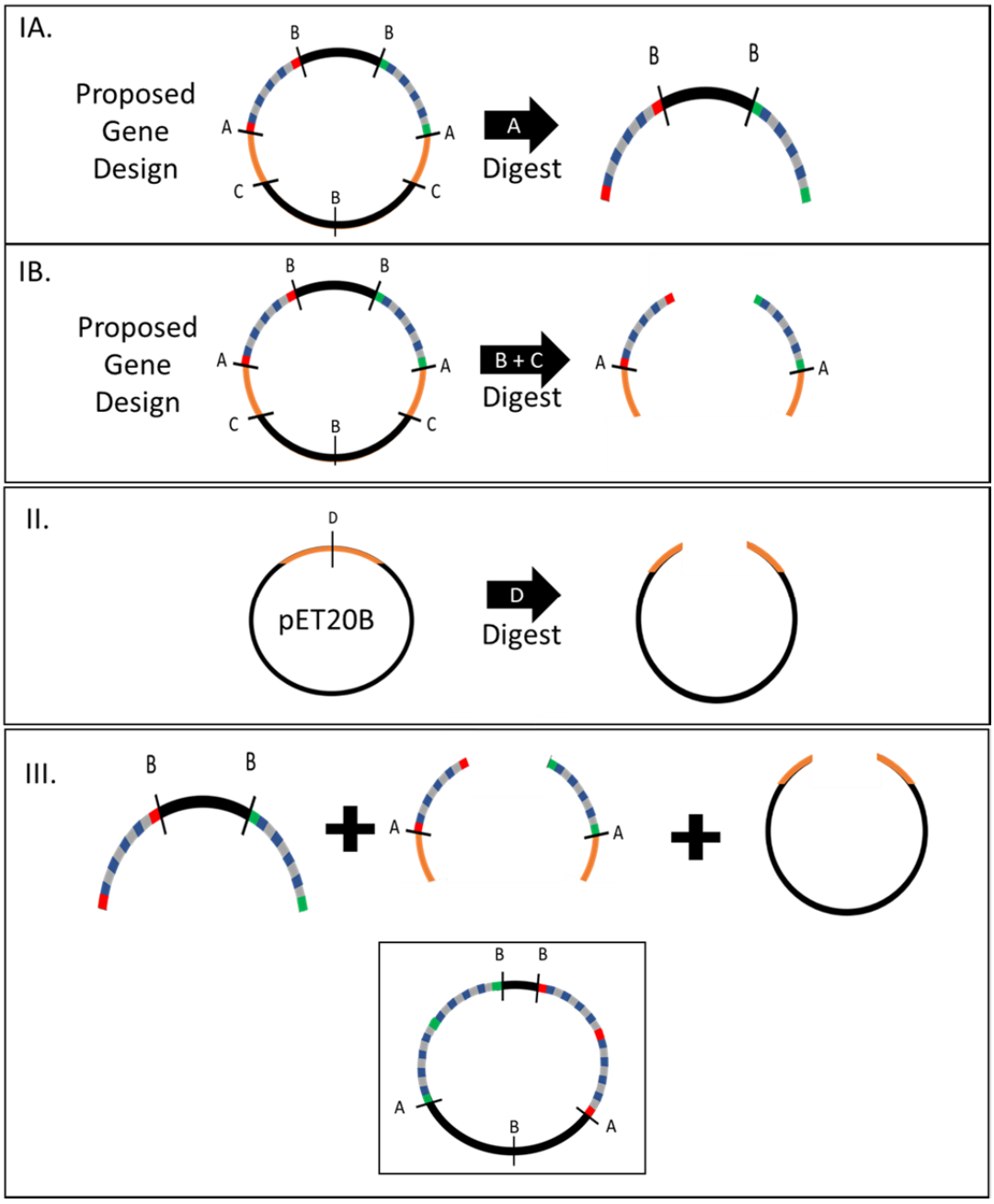


Figure 15. Digestion and Assembly of six-armed 19 Pentapeptide

A second recursion of the original design was unable to operate as designed. This recursion would have created a 37 pentapeptide system on each side of the foldon in the star polymer. The previous recursion model was used to propose new enzymes at locations A, B, and C.

The 19 pentapeptide recursion was proved to have assembled correctly. Protein expression was performed next. The protein expression step uses BL21* bacteria. The verified gene is transformed into the bacteria and protein expression and purification are carried out. To verify the correct protein has been expressed an SDS-PAGE was run to obtain the molecular weight of the protein (Figure 16).

The bands are indicated by the purple lines displayed at different vertical locations from left to right across the gel. The first well is the marker. Wells 2-5 contain protein which has been boiled. Boiling has been shown to disrupt the folding of foldon containing proteins [12]. This allows the protein to be shown in a single chain form or colloquially referred to as a monomer of the protein. Well 2 contains the 19 pentapeptide six-armed star polymer after boiling. Wells 3, 4, and 5 contain the 19, 37, and 73 pentapeptide three-armed variants, respectively. Well 6 contains the 19 pentapeptide six-armed star polymer without the boiling step. Wells 7, 8, and 9 contain the three-armed variants, in respect to before, without the boiling step. At this point, it is important to note the expected sizes of each protein (Table 6).

Protein Description	Protein Size Expectation (kDa)
19 – Pentapeptides, 6 – Armed, Boiled	23
19 – Pentapeptides, 3 – Armed, Boiled	13
37 – Pentapeptides, 3 – Armed, Boiled	21
73 – Pentapeptides, 3 – Armed, Boiled	35
19 – Pentapeptides, 6 – Armed, Unboiled	69
19 – Pentapeptides, 3 – Armed, Unboiled	40
37 – Pentapeptides, 3 – Armed, Unboiled	62
73 – Pentapeptides, 3 – Armed, Unboiled	106

Table 6. Expected Size of Each Protein

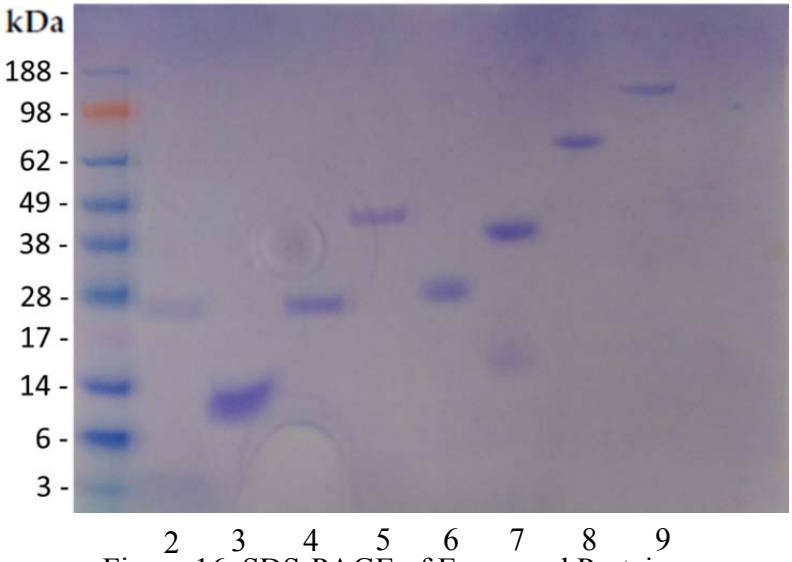


Figure 16. SDS-PAGE of Expressed Protein

The three-armed polymers show bands which are relatively close to expectations. Wells seven, eight, and nine are approximately three times as large as three, four, and five, respectively. Further, the boiling step is denaturing the three-armed polymers, but the denaturing is independent of the boiling step in the six-armed polymer. This is contrary to what was expected; the six-armed protein was expected to form a trimer in the unboiled state and a monomer in the boiled state. A hypothesis for this involves the procedure for SDS-PAGE. Sodium dodecyl sulfate (SDS) is a chemical used in the electrophoresis process to denature the protein. The six-armed protein is the only protein which contains large ELP chains on both ends of the foldon trimer. The trimer stability of foldon may be decreased due to the complexity of the chain involving two ELP segments encapsulating the foldon.

3.2 Star Polymer Characterization by Turbidity Measurements

A new protein has been expressed and verified to be the desired six-armed star polymer by relative molecular weight. Whereas, the SDS-PAGE shows that when the

three-armed polypeptide samples are not boiled they show a molecular weight consistent with a trimer, the six-armed polypeptide under the same conditions are at a weight consistent with the monomer. However, it is important to determine whether the six-armed polypeptide forms a trimer under normal solution conditions like the three-armed polymers. Since the phase transition temperature of ELPs, as measured by solution turbidity, depends on whether or not the protein is a trimer or a monomer, a comparison of the transition temperatures of the three- and six-armed polypeptides can be used to answer this question.

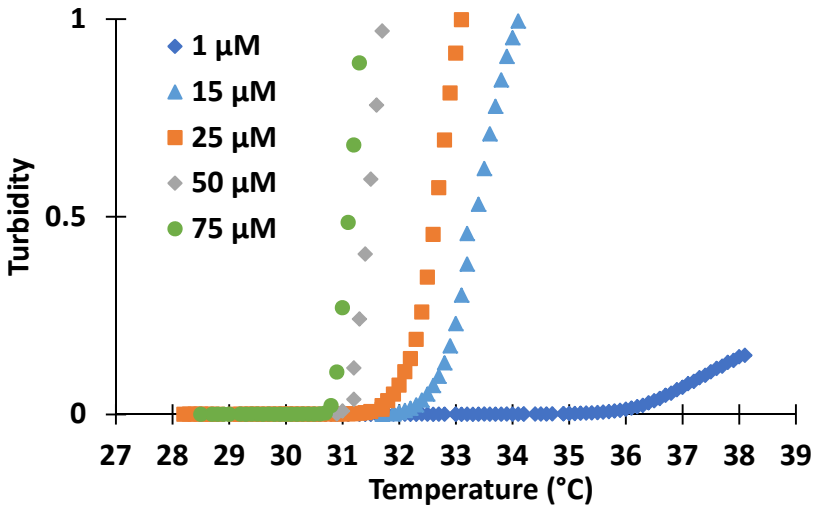


Figure 17. UV-VIS Data for six-armed Star Polymer

Concentration	T_t (C°)
1 µM	36.2
15 µM	32.7
25 µM	32.2
50 µM	31.2
75 µM	30.9

Table 7. Transition Temperature (T_t) of 19 Pentapeptide Star Polymer

These values will be compared to the three-armed version with differing arm lengths. They will be fit to a model for ELP structures containing foldon as a trimerizing agent [17].This comparison will allow a determination into understanding the effects of the addition of three arms and to conclude whether or not the trimer is forming. The transition temperature experiment was performed for all of the three-armed polymers.

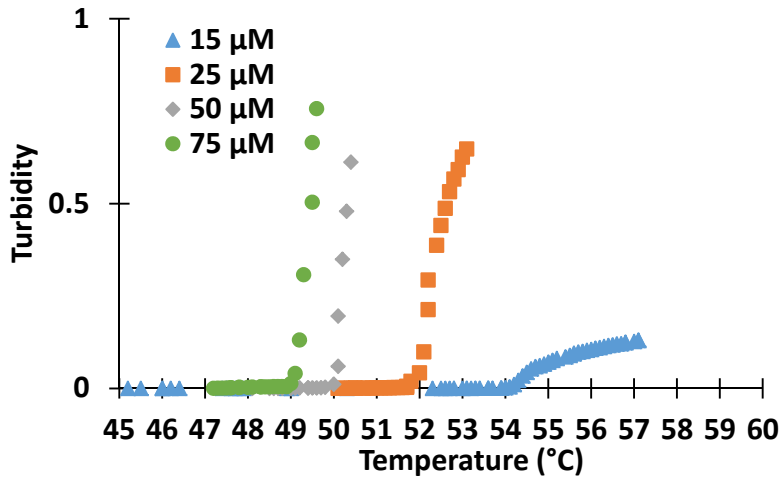


Figure 18. UV-VIS Data for Three-armed 19 Pentapeptide Polymer

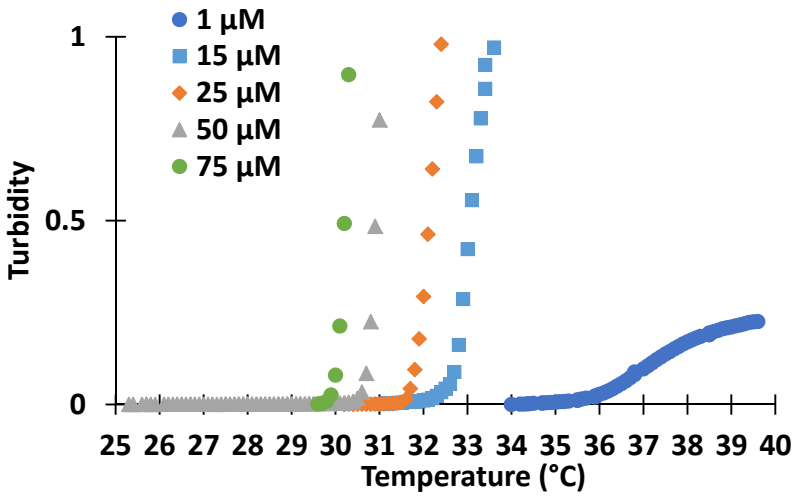


Figure 19. UV-VIS Data for Three-armed 37 Pentapeptide Polymer

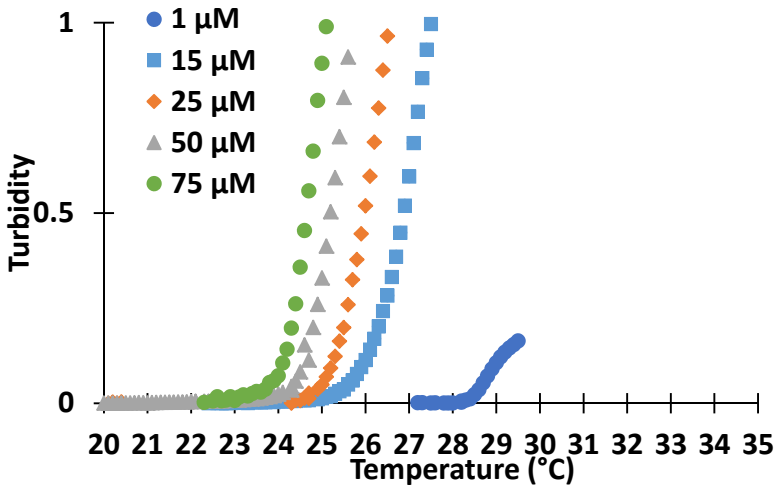


Figure 20. UV-VIS Data for Three-armed 73 Pentapeptide Polymer

Concentration (µM)	19-Pentapeptides (°C)	37-Pentapeptides (°C)	73-Pentapeptides (°C)
1	N/A	36.1	28.4
15	54.1	32.7	26.3
25	52.0	31.8	25.5
50	50.0	30.7	24.7
75	49.1	30.0	24.2

Table 8. Transition Temperature of three-Armed Polymers with Different Arm Lengths

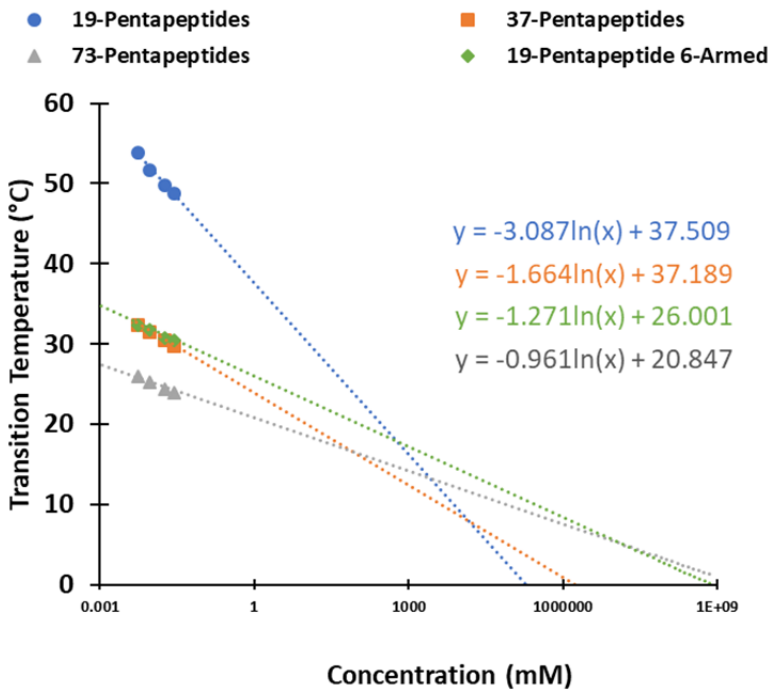


Figure 21. Transition Temperature as a Function of Concentration

The six-armed star polymer behaves most closely to the three-armed 37 pentapeptide (Table 6). The transition temperatures (Figure 21) can be fit to a model developed by Ali Ghorchian for trimer ELPs. The reported fit parameters (T_{cr}, K_t, and C_{cr}) are for ELP monomer and trimer structures without the decorin motif [17]. These values were used to calculate transition temperatures according to the model outlined in literature. The method of least squares was used to establish new fit the parameters to the data in molar concentrations.

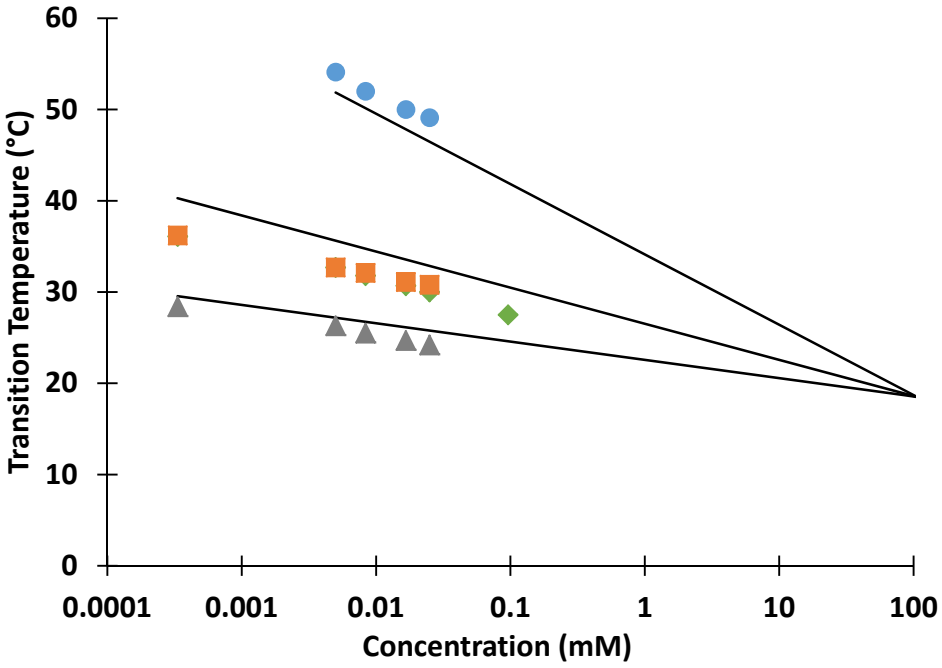


Figure 22. Transition Temperatures on the Ghorchian Model in Trimer Fit

Parameter	Linear (without Foldon)	Trimer (with Foldon)
K_t (°C)	114.5	190.8
C_{cr} (mM)	30.0	106.4
T_{cr} (°C)	20.8	18.5

Table 9. Fit Parameters for ELP Transition Temperatures

The fit parameters from the Ghoorchian model (Table 9) fit the data for the decorin containing trimers quite well (Figure 22). The 37-pentapeptide three-armed star polymer and the 19-pentapeptide six-armed star polymer contains a slightly lower transition temperature than expected. Whereas, the 19-pentapeptide three-armed star polymer and the 73-pentapeptide three-armed star polymer were more closely fit by the model.

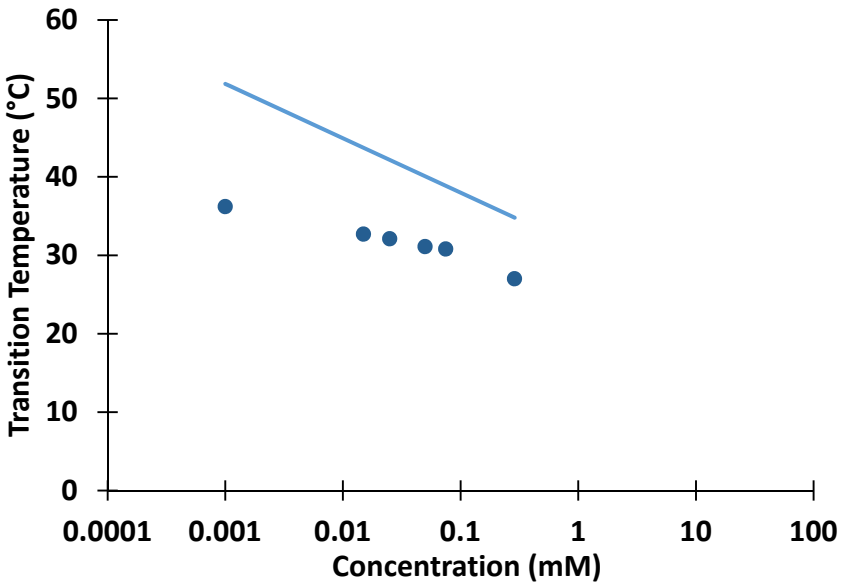


Figure 23. Six-Armed 19 Pentapeptide Star Polymer Fit to Monomer 38 Pentapeptide Parameters

Since the SDS-PAGE results suggested that the six-armed 19-pentapeptide may not form a trimer, it was compared to the monomer model found in the literature. The model is fit for a 38-pentapeptide ELP monomer (Figure 23). The six-armed 19 pentapeptide star polymer is displayed by the circle symbols on the plot. It is apparent that the data model does not fit the six-armed 19-pentapeptide star polymer since the transition temperature is significantly lower than the model would predict. The lower T_t indicates that the ELP is larger than the monomer and suggest that it does fold as a trimer. Ghoorchian proposed using volume concentration to fit the data more accurately to physical phenomena, including the critical temperature for large molecular weight ELPs of 23 °C and the scaling parameter *n* [17]. The curves were converted to a volume concentration (i.e. volume fraction) basis. Each sample represents the fraction of total polymer coil volume to the solution volume. Using the reported volume concentration

model parameters for ELP trimers (Table 10), the model reasonably fit the data (Figure 24).

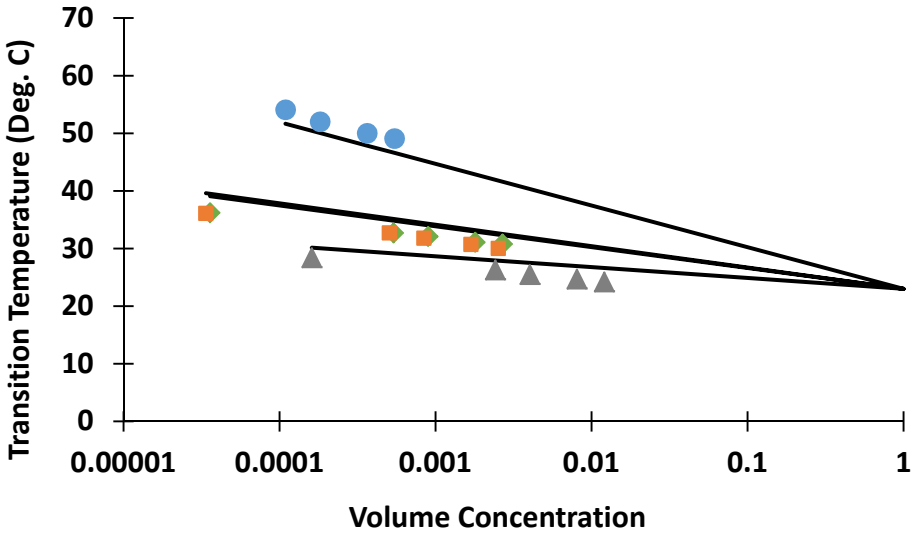


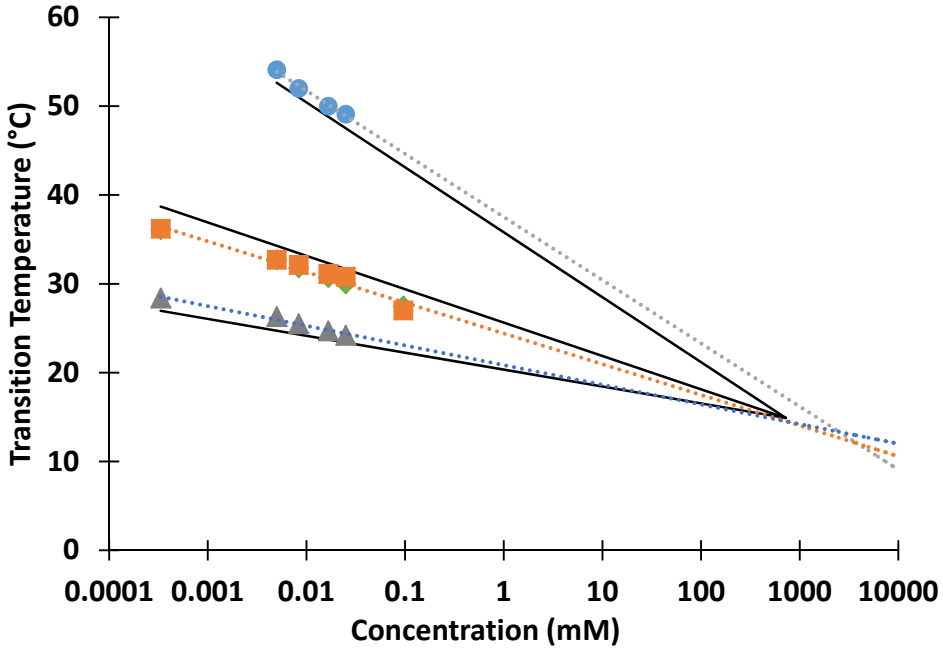
Figure 24. Volume Concentration Fit

Parameter Name	Parameter Value
K_t (Deg. C)	179
K'	2.00e-06
T_{cr} (Deg. C)	23
n	2.3

Table 10. Volume Concentration Fit Parameters

The molar and volume models were fit using custom parameters in combination with the method of least squares to determine the most ideal fit (Figure 25). The values found were different from the literature values (Table 11). In the molar concentration fit, the critical temperature (T_t) was slightly lower than the literature value. The constant (K_t) slightly increases. The change in the critical concentration (C_{cr}) is the most significant changing parameter. This value is increased considerably from the literature value. This

change suggests that the point at which the data becomes independent of protein molecular weight is at much larger concentrations than prior. The molar concentration model does not accurately indicate the physical behavior of the protein polymer chains. To correct for the physical behavior of the polymer chains the volume concentration model was applied with custom parameters (Figure 25). The volume concentration fit had minor differences in the fit parameters from that in literature (Table 12).



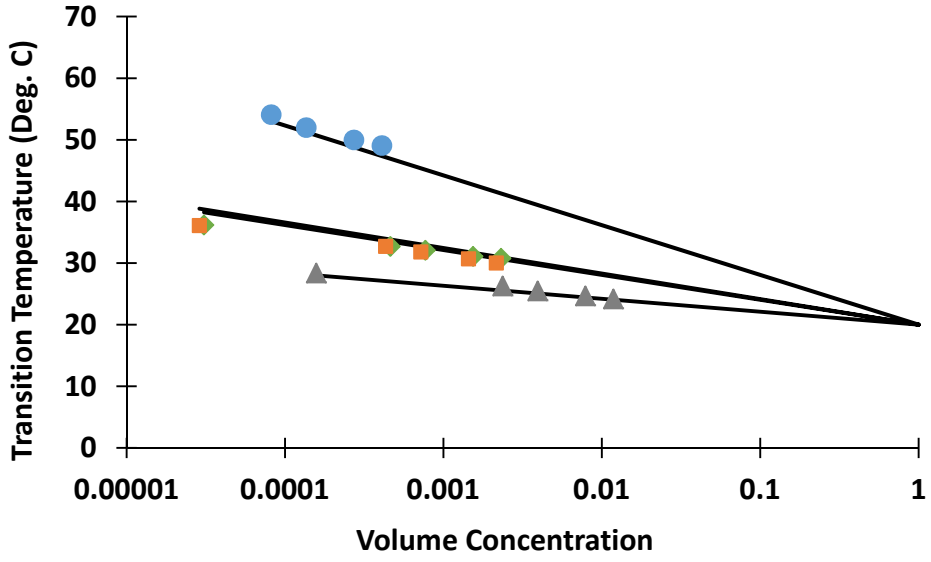


Figure 25. Molar (Top) and Volume (Bottom) Concentration Fit with Custom Parameters

Parameter	Linear (without Foldon)
K_t (°C)	181
C_{cr} (mM)	722
T_{cr} (°C)	14.9

Table 11. Custom Molar Concentration Fit Parameters

Parameter Name	Parameter Value
K_t (Deg. C)	200
K’	2.00e-06
T_{cr} (Deg. C)	20
n	2.5

Table 12. Custom Volume Concentration Fit Parameters

3.4 Future Considerations

This work utilized the 19 pentapeptide six-armed star polymer. A platform was designed for recursion to create 37 and 73 pentapeptide repeats; The platform was ineffective for a second recursion. It would be best to use a new design to explore the possibility of larger pentapeptide repeats on the arms of the star polymer. These different lengths would help to explore the behavior of the polymer.

It would be ideal to use other analytical methods to determine the folding properties of the six-armed star polymer. It is important to understand if similar folding is occurring in the six-armed star polymer as in the three-armed star polymer.

CHAPTER IV

CONCLUSION

In these experiments, a recursive design was explored for the synthesis of a biological based star polymer. Gibson Assembly was successfully performed to produce one recursion on the synthesis platform. This recursion created the star polymer containing 19 pentapeptide repeats on each of the six arms. Attempts were made to create the further recursions. These attempts had failed due to the restriction endonucleases used in the design. The endonucleases digested the plasmid backbone that the gene resides in. The design could be changed to include endonucleases which do not digest the backbone of the gene. The proposed redesign could contain recognition sites for the enzymes BseYI, AgeI, and AscI for the C, A, and B locations in Figure 4, respectively.

The trimerization of the six-armed 19 pentapeptide star polymers were not fully verified. The SDS-PAGE results displayed only the denatured monomer state. However, the transition temperature results indicated that this polymer behaves like the 37 pentapeptide three-armed polymer which did exhibit folding behavior during the SDS-PAGE experiment. Other analytical methods could be pursued to verify the trimerization of the six-armed polymer.

Ali Ghoorchian had produced a model that the three-armed and six-armed polymers were compared to [17]. The molar concentration model provided a fit to the star polymers. The molar concentration model provided a fit to the transition temperature curves. The volume concentration model was used as a second attempt to fit the data. However, these models were found to predict an increase in the transition temperature found within the three-armed 37 pentapeptide and the six-armed 19 pentapeptide star polymers. The molar and volumetric concentration fits were relatively close to the data.

BIBLIOGRAPHY

[1] R. A. Rader, "(Re)defining biopharmaceutical," *Nature Biotechnology,* vol. 26, no. 7, pp. 743-751, 2008.

[2] T. Christensen, K. Trabbic-Carlson, W. Liu and A. Chilkoti, "Purification of recombinant proteins from Escherichia coli at low expression levels by inverse transition cycling," *Analytical Biochemistry,* vol. 369, no. 1, pp. 166-168, 2006.

[3] D. McPherson, J. Xu and D. Urry, "Product purification by reversible phase transition following Escherichia coli expression of genes encoding up to 251 repeats of the elastomeric pentapeptide GVGVP," *Protein Expression and Purification,* vol. 7, no. 1, pp. 51-57, 1996.

[4] J. M. Ren, T. G. McKenzie, Q. Fu, E. H. H. Wong, J. Xu, Z. An, S. Shanmugam, T. P. Davis, C. Boyer and G. G. Qiao, "Star Polymers," *Chemical Reviews,* vol. 116, no. 12, pp. 6743-6836, 2016.

[5] D. G. Gibson, L. Young, C. Ray-Yuan, J. C. Venter, C. A. Hutchison and H. O. Smith, "Enzymatic Assembly of DNA Molecules Up to Several Hundred Kilobases," *Nature Methods,* vol. 6, no. 5, pp. 343-U41, 2009.

[6] J. K. Deyling, E. M. Turk and N. B. Holland, "Preparing Genes for Repetitive Elastin-like Polypeptides Using Gibson Assembly," in *Protein Scaffolds: Design, Synthesis, and Applications*, vol. 1798, A. K. Udit, Ed., Humana Press, 2018.

[7] C. Engler, R. Kandzia and S. Marillonnet, "A One Pot, One Step, Precision Cloning Method with High Throughput Capability," *Plos One,* vol. 3, no. 11, 2008.

[8] M. Adams, A. Lavasanifar and G. Kwon, "Amphiphilic block copolymers for drug delivery," *Journal of Pharmaceutical Sciences,* vol. 92, no. 7, pp. 1343-1355, 2003.

[9] G. Riess, "Micellization of Block Copolymers," *Progress in Polymer Science,* vol. 28, no. 7, pp. 1107-1170, 2003.

[10] L. H. Sperling, Physical Polymer Science, Hoboken, New Jersey: John Wiley & Sons, Inc., 2006, pp. 51-53.

[11] J. C. Berg, An Introduction to Interfaces & Colloids, Hackensack, New Jersey: World Scientific Publishing Company, 2010, pp. 148-160.

[12] A. Ghoorchian, J. T. Cole and N. B. Holland, "Thermoreversible Micelle Formation Using a Three-Armed Star Elastin-like Polypeptide," *Macromolecules,* vol. 43, no. 9, pp. 4340-4345, 2010.

[13] D. Urry, "Physical chemistry of biological free energy transduction as demonstrated by elastic protein-based polymers," *Journal of Physical Chemistry B,* vol. 101, no. 51, pp. 11007-11028, 1997.

[14] K. Chaturvedi, K. Ganguly, M. N. Nadagouda and T. M. Aminabhavi, "Polymeric hydrogels for oral insulin delivery," *Journal of Controlled Release,* vol. 165, no. 2, pp. 129-138, 2013.

[15] A. Ghoorchian, K. Vandemark, K. Freeman, S. Kambow, N. B. Holland and K. A. Streletzky, "Size and Shape Characterization of Thermoreversible Micelles of Three-Armed Star Elastin-Like Polypeptides," *Journal of Physical Chemistry B,* vol. 117, no. 29, pp. 8865-8874, 2013.

[16] S. Meier, S. Guthe, T. Kiefhaber and S. Grzesiek, "Foldon, the natural trimerization

domain of T4 fibritin, dissociates into a monomeric A-state form containing a stable beta-hairpin: Atomic details of trimer dissociation and local beta-hairpin stability from residual dipolar couplings," *Journal of Molecular Biology,* vol. 344, no. 4, pp. 1051-1069, 2004.

[17] A. Ghoorchian and N. B. Holland, "Molecular Architecture Influences the Thermally Induced Aggregation Behavior of Elastin-like Polypeptides," *Biomacromolecules,* vol. 12, no. 11, pp. 4022-4029, 2011.

[18] J. K. Deyling, Design and Production of a Hydrogel Forming Polypeptide: Engaging High School Students in Protein Design, Cleveland: Cleveland State University, 2016.

[19] O. Can and B. N. Holland, "Utilizing Avidity to Improve Antifreeze Protein Activity: Type III Antifreeze Protein Trimer Exhibits Increased Thermal Hysteresis," *Biochemistry,* vol. 52, pp. 8745-8752, 2013.

[20] T. Christensen, K. Trabbic-Carlson, W. Liu and A. Chilkoti, "Purification of recombinant proteins from E. coli at low expression levels by inverse transition cycling," *Analytical Biochemistry,* vol. 360, no. 1, pp. 166-168, 2007.

[21] S. Frederico, B. F. Pierce, S. Piluso, C. Wischke, A. Lendlein and A. T. Neffe, "Design of Decorin-Based Peptides That Bind to CollagenI and their Potential as Adhesion Moieties in Biomaterials," *Angewandte Chemie-International Edition,* vol. 54, no. 37, pp. 10980-10984, 2015.

[22] S. C. Gill and P. H. von Hippel, "Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data," *Analytical Biochemistry,* vol. 182, no. 1, pp.

319-326, 1989.

[23] N. C. Pace, F. Vajdos, L. Fee, G. Grimsley and T. Gray, "How to measure and predict the molar absorption coefficient of a protein," *Protein Science,* vol. 4, pp. 2411-2423, 1995.

[24] K. M. Otto, Characterization of Highly Concentrated Elastin-Like Polypeptide Solutions: Rheometric Properties and Phase Separation Analysis, Cleveland State University, 2013.