Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature

_____ _____

Ye Tian Date

Biosynthesis and characterization of novel elastin-based biomaterials and theoretical calculation of proline analogues

By

Ye Tian

Master of Science
Chemistry

_____

Vincent P. Conticello
Advisor

_____

Elliot L. Chaikof
Committee member

_____

Fred M. Menger
Committee member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of Graduate School

_____

Date

Biosynthesis and characterization of fluorinated elastin-mimetic polypeptides and theoretical calculation of proline analogues

By

Ye Tian

B.S., Shanghai Jiao Tong University, 2005

Advisor: Vincent P. Conticello, Ph.D.

An Abstract of

A thesis submitted to the Faculty of the Graduate School of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science

in Chemistry

2008

Abstract

Biosynthesis and characterization of fluorinated elastin-mimetic polypeptides and theoretical calculation of proline analogues

By Ye Tian

Elastin is an important protein that provides connectivity and elasticity in the extracellular matrix. Elastin-mimetic polypeptides (EMPs) have been widely used as models for studying the structure and property of native elastin.

In this study, two fluorinated amino acids, 4, 4, 4-trifluorovaline and 3-fluorovaline were incorporated into the isoleucine and valine codons of an elastin-mimetic polypeptide sequence. Both the two fluorinated EMPs underwent inverse temperature transitions. Structurally, these two proteins self-assembled from random coil to type II β turn upon temperature increase. In addition, an EMP sequence comprised of repeating pentapeptide [Val-Pro-Gly-Ala-Gly] was designed to serve as the sequence for the incorporation of 4, 4-difluoroproline. The gene encoding this sequence was constructed by recursive directional ligation. Moreover, two plasmin-sensitive elastin-based triblock copolymers with different size in the mid-block were studied. It was revealed that these two proteins underwent inverse temperature transition upon a critical temperature of 19.8 ℃ and 20.0 ℃, respectively. This transition was accompanied by a structure change from random coil to an ordered structure, which was confirmed by the CD results. Biodegradation of both proteins was demonstrated to be complete within six hours. Furthermore, two oxidation products of N-acetyl-thioproline-methyl ester were geometry-optimized by DFT computation. The calculation results revealed that the *anti* product preferred a C^{γ}-*exo* pucker and a *trans* amide bond while the *syn* product preferred a C^{γ}-*endo* pucker and a *cis* amide bond. It was shown that both the gauche effect between the sulfoxide bond and the amide nitrogen as well as the n→π* interaction between the nonbonding oxygen and the carbonyl group of the ester contributed to the conformational preferences.

Biosynthesis and characterization of fluorinated elastin-mimetic polypeptides and theoretical calculation of proline analogues

By

Ye Tian

B.S., Shanghai Jiao Tong University, 2005

Advisor: Vincent P. Conticello, Ph.D.

A thesis submitted to the Faculty of the Graduate School of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Chemistry
2008

Acknowledgments

I would like to sincerely thank my advisor, Dr.Vince Conticello for his guidance, understanding, support and encouragement during my research. He not only taught me how to think independently and how to conduct experiment scientifically, but also showed me how to be a decent, responsible scientist. I would also thank my committee members Dr. Elliot Chaikof and Dr. Fred Menger for their guidance in my research.

I would express my deep gratitude to my lab mates, Dr. Wookhyun Kim, Dr. Jet Qi, Dr.Steve Dublin, Dr.John Shugart, Dr. Holly Carpenter, Yang Feng, Melissa Peterson, Weilin Peng, Yunyun Pei and I-lin Wu for their assistance and friendship.

I would thank Dr. Shaoxiong Wu and Dr.Bing Wang for their help in the acquisition of NMR data. I would also thank James Simmons for teaching me how to use the CD and DSC instruments. I wish to thank Seth Childers for helping me set up the automatic peptide synthesis. I also thank Andy Prussia for his assistance in DFT calculations. Actually there are so many people I appreciate that I can not list all their names here, but I want to say their help and friendship are really valuable wealth for me.

Finally I would like to thank my parents for their love and unselfish support all these years. I also want to express my appreciation to all my friends who shared friendship with me.

Table of contents

List of Figures

List of Tables

List of Schemes

Chapter 1 General introduction

As one of the most versatile materials in nature, elastin is an important protein derived from vertebrate tissues that provides connectivity in the extracellular matrix[1]. Located in the walls of arteries, lungs, intestines and even our skin, elastin has played a crucial role in human body and the dysfunction of elastin can cause a series of diseases and disorders. In addition, since elastin will not be produced after human being's puberty, we have to consume a specific amount of elastin in our whole life, that is why wrinkle appears and people look old when they are aged. As a result of elastin's uniqueness and importance, the study of its structure and function has attracted intensive attention.

Elastin forms from its soluble precursor, tropoelastin, which is composed of two domains: hydrophobic domain which is rich in hydrophobic residues such as Ala, Gly, Val as well as Leu and the more hydrophilic domain dominated by Lys and Ala[2]. Soon after the tropoelastin is synthesized by the cell, it will be cross-linked via lysine residues during its transport to the extracellular matrix, forming elastic fibers[3]. Due to the insoluble property of mature elastin fiber, the understanding of three dimensional structure of elastin is still not handy, thus making the structure-function relationship of this protein difficult for scientists to discover. In order to overcome this problem, elastin-mimetic polypeptides (EMPs), which borrow the repetitive sequences of native elastin but have no solubility problem, have been designed and used as models for the investigation of native elastin. The most common repeating sequences for elastin-mimetic polypeptides are the pentapeptides [(Val/Ile)-Pro-Xaa-Yaa-Gly], in which the fourth residue Yaa, which can be occupied by any amino acids except proline, determines the lower critical solution temperature (LCST) of the polypeptides in aqueous solution depending on the polarity of the amino acid chain, in a manner of affecting the polypeptide-solvent interaction. In addition, Xaa position amino acid will be determinant for the mechanical behavior of the corresponding polypeptides[4]. For example,

replacing Gly with Ala residue will change the mechanical response of the materials from elastomeric to plastic.

Elastin-mimetic polypeptides possess remarkable properties, which make them interesting topics for research. First, like native elastin, EMPs have substantial elasticity. They can be stretched and contracted to several times of their initial size and will return to the original shape without any damage. Secondly, since the sequences are borrowed from the native elastin, which are abundant in human body, good biocompatibility can be expected for EMP-based materials. Third and the most noteworthy property of EMPs is the thermo-sensitivity. EMPs undergo an inverse temperature transition in a narrow temperature range, which means there is a transition temperature T_t and when the ambient temperature is below its transition temperature (T_t), the polypeptides are soluble in water, but as the temperature is raised above T_t, the EMPs aggregate and ultimately precipitate from the solution[5, 6]. The transition temperature (T_t), which is one of the most important parameters of the thermo-responsible materials, can be influenced by factors including composition and molecular weight of the EMP, as well as the salt concentration of the solution[7-9]. Therefore, by appropriately adjusting these parameters, the inverse transition temperature can be modified to almost any value, which provides scientists an important means to create novel materials responsive to a wide variety of environmental stimuli.

Based on their thorough study of the model peptides poly (VPGVG), Dan Urry and coworkers not only suggested the secondary structural features of elastin-mimetic polypeptide[10, 11], but also proposed the molecular level mechanism for the LCST behavior[10, 11]. In their model, there is a type II β turn in each pentameric unit, in which Pro and Gly are responsible for the turn formation, due to their high turn-forming propensity[12]. This secondary structure is stabilized by the hydrogen boding between the C=O group of the first Val to the NH of the fourth Val. The repetition of type II β turn structure in each unit has made up of a helical arrangement, which they called β-spiral[11]. When the environment temperature is below the transition temperature T_t, the macromolecule is highly disordered and hydrated, mainly through hydrophobic hydrating. However, when the temperature is

raised above the critical temperature, the protein will fold into ordered structure, stabilized by hydrophobic interaction. At this point, the protein has lost entropy to display a β spiral conformation, while the water has been excluded to form bulky water.

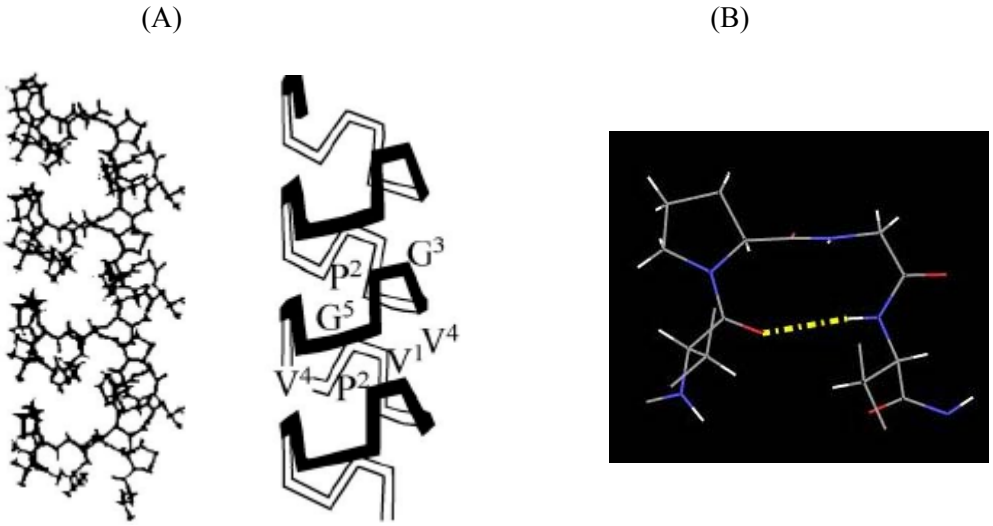
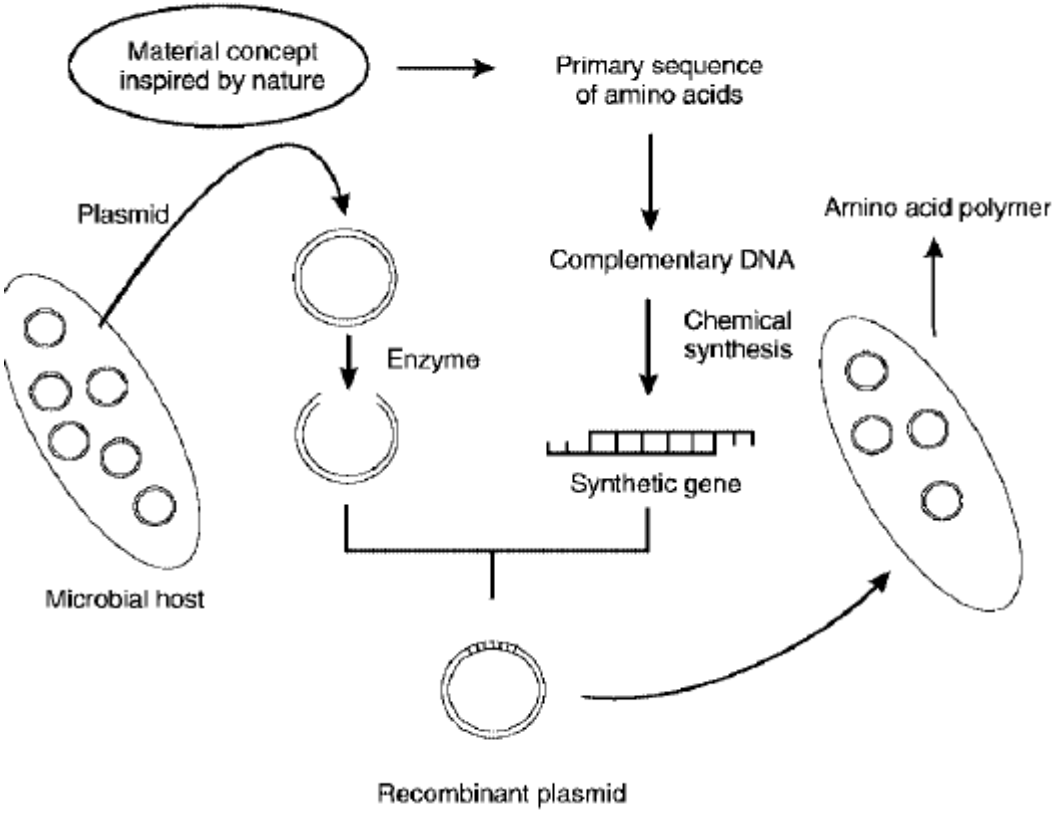


Figure 1. Secondary structure of elastin-mimetic polypeptide (A) Schematic representation of the helical structure, β spiral; (B) β turn structure[11]

Actually, a great many of scientists are fascinated with these protein-based materials not only because of their unique properties, but also because these EMPs can find potential applications in a wide range of fields, in which biomedical engineering and biotechnology are two major areas where elastin-mimetic polypeptide can play crucial roles. Protein engineering has been widely employed to produce large quantity of pure proteins, with nearly complete control of composition, sequence and length. Scheme 1 shows how the protein is synthesized by the biological machinery during the whole process.



Scheme 1. Overview of the protein engineering methodology[1]

Taking advantage of genetic engineering, Conticello *et al* have created an elastin-based amphiphilic block copolymer in which hydrophilic block contains polar residue Glu while hydrophobic domain includes hydrophobic amino acid Phe[13]. This copolymer undergoes reversible, temperature dependent phase separation in aqueous solution, producing biocompatible spherical shape nanoparticles that are highly potential for drug delivery and controlled release. The sequences for both the two blocks are shown below.

Hydrophobic block: $[Val\text{-}Pro\text{-}Gly\text{-}Phe\text{-}Gly\text{-}(Ile\text{-}Pro\text{-}Gly\text{-}Val\text{-}Gly)_4]_{16}$

Hydrophilic block： $[Val\text{-}Pro\text{-}Gly\text{-}Glu\text{-}Gly\text{-}(Ile\text{-}Pro\text{-}Gly\text{-}Ala\text{-}Gly)_4]_{14}$

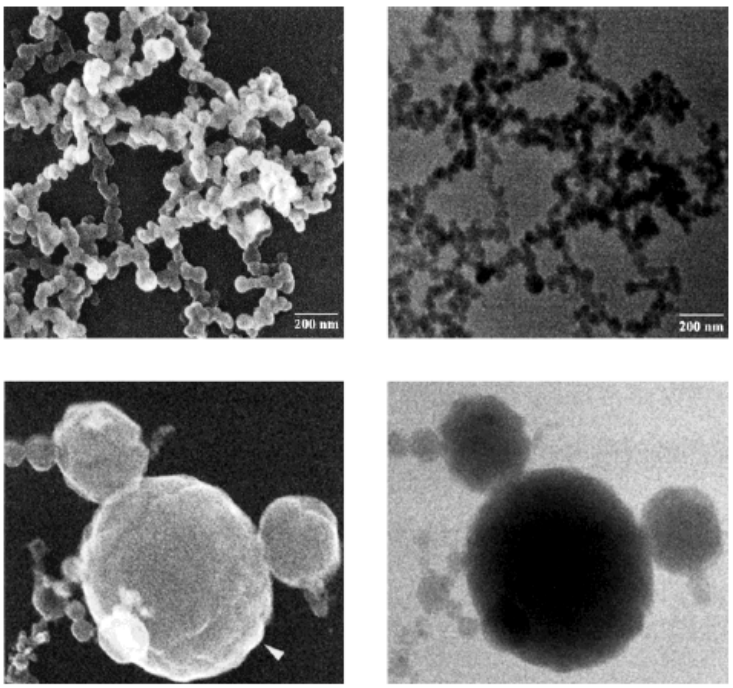
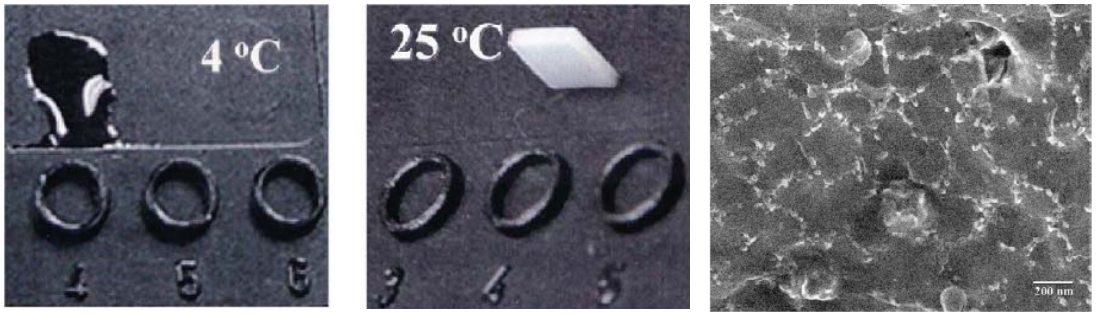
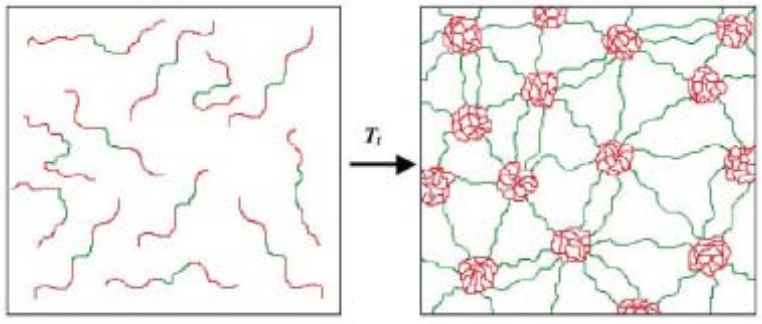


Figure 2. TEM images for the aggregate of the copolymer[13]

Another demonstration of elastin-based biomaterials also comes from the same group, who introduced a triblock copolymer BAB, in which end block B displays plastic mechanical response while central block A displays elastomeric behavior[14]. When the temperature was raised from 4℃ to 25 ℃, gelation was observed as a result of the temperature-dependent phase transition upon self-assembly. These materials have considerable promise as biomedical materials in drug delivery and tissue engineering. Further investigation of the gelation mechanism has revealed that the central block A serves as the bridge between end blocks. When the temperature is increased above T_t, plastic block B will aggregate and separate out from the solvent. However, elastic A blocks remain solvated and link B blocks together. In this way, a dispersed, cross-linked network formed, reflected macroscopically by the gelation process.



(A) (B)



(C)

Figure 3. Experimental and mechanistic study of the BAB copolymer (A) Gelation when temperature is increased from 4 ℃ to 25 ℃ (B) TEM image of the self-assembled copolymer (C) Schemetic represenation of the formation of the swollen network in which central A block serves as bridges between B blocks[4, 14]

Besides biomedical devices, elastin-based materials can be widely used in biotechnology sectors. One famous example is to facilitate protein purification relying on EMP's temperature-sensitivity. Wood group has fused elastin-mimetic polypeptide with a series of target proteins by introducing inteins[7]. After several cycles of hot-cold spin, only protein fusions with EMPs left in the solution while all the contaminants were washed away. After cleaving at the intein sites, target proteins with high purity and comparable yield were produced. The whole process is simple, efficient and more economical than traditional purification strategies. Figure 4 is a scheme for the EMP fusion purification described above.

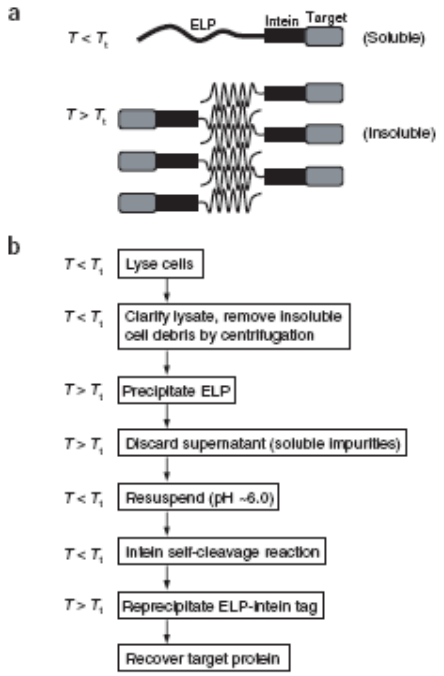


Figure 4. Purification of ELP-intein–tagged protein. (a) The EMP fusion protein complex will self-associate upon temperature increase above Tt, separated from contaminants. (b) Scheme of the target protein purification[7]

Because EMPs exhibit such unique and valuable properties and promise several useful applications, a deeper understanding of the EMP structure would facilitate the downstream manipulations and consequently, has been a research concern for a long time. We describe herein a combination of genetic engineering and analogues incorporation strategy to investigate the local and global structural features of elastin-mimetic polypeptides. In addition, elastin-based triblock copolymers with plasmin degradation sites are characterized to understand their primary and secondary structure, as well as the thermal transition. Moreover, theoretical calculation will also be employed to model and optimize the small molecule models, which will provide us insight of the conformational preference of proline analogues, thus facilitate our understanding of the structure and self-assembly process of the elastin-mimetic polypeptides.

Chapter 2. Biosynthesis and characterization of fluorinated elastin-mimetic polypeptides

Background

Because elastin-mimetic polypeptides (EMPs) possess such unique and valuable properties and promise several useful applications, a deeper understanding of the EMP structure would facilitate downstream applications and consequently, has been a research concern for a long time. However, conventional methods for protein structural determination, such as x-ray diffraction and multidimensional solution NMR can not be easily applied to the study of EMP structure since phase separation occurs upon temperature increase[15]. Site-directed and random mutagenesis, in which specific amino acids in a protein are substituted for other natural amino acids, have been powerful tools to understand and modify the structure and function of proteins. However these studies are of limited use because there are intrinsic structural differences between the amino acids before and after mutations[16]. In order to overcome these limitations, chemosynthetic[2] and biosynthetic[17, 18] methods which enable the incorporation of non-canonical amino acids closely resembling their natural amino acid counterparts into the native proteins have been employed and developed. The research in the incorporation of unnatural amino acids not only provides us important methods to uncover the protein structures, but also create proteins with novel properties, unusual structural and biological characteristics.

Among all these non-canonical amino acids, fluorinated amino acids in which hydrogen atoms have been replaced by fluorine atoms have attracted special attention [16, 19-21]. Firstly, fluorine is usually considered to be isosteric with hydrogen. Therefore, a single atom exchange, from hydrogen (H) to fluorine (F), which has similar atom size with hydrogen, guarantees relatively low structural perturbation, thus maintaining the function and activity to the maximum extent. Secondly, fluorocarbons are usually considered to be more hydrophobic than the hydrocarbon of the corresponding size. The enhanced hydrophobic property of the fluorinated molecules is commonly utilized in drug delivery and in drug design to increase

biocompatibility. Because hydrophobic forces play important roles in protein folding and protein-protein recognition, it was suggested that fluorinated amino acids could be utilized in protein engineering to prepare peptides and proteins with enhanced structural stability. Tirrell *et al.* [22]and Kumar *et al* [23] ndependently demonstrated that replacement of leucine by 5, 5, 5-trifluoroleucine (5TFI) at the *d*-positions of the leucine zipper peptide GCN4-p1 increased the thermal stability of the coiled-coil structure. After 5TFI was incorporated by genetic engineering, the melting temperature T_m of the peptide was increased by 27 ℃ and $\Delta\Delta G_{unfold}$ in water was determined to be 2.7 kcal/mol in the chemical denaturation studies. Although remarkable difference was observed for the protein stability after fluorinated analogues incorporation, little impact was observed to the secondary and higher order structure of the peptide, and meanwhile physical properties such as DNA binding affinity and specificity were largely maintained. In addition, stereoelectronic effects arising from the fluorine substitution have been determined to contribute to the conformational stability of collagen-mimetic peptide[15], in which electronegative gauche effects stabilize both the C$^{\gamma}$-*exo* conformation of the pyrrolidine ring and the *trans* configuration of prolyl amide bond.

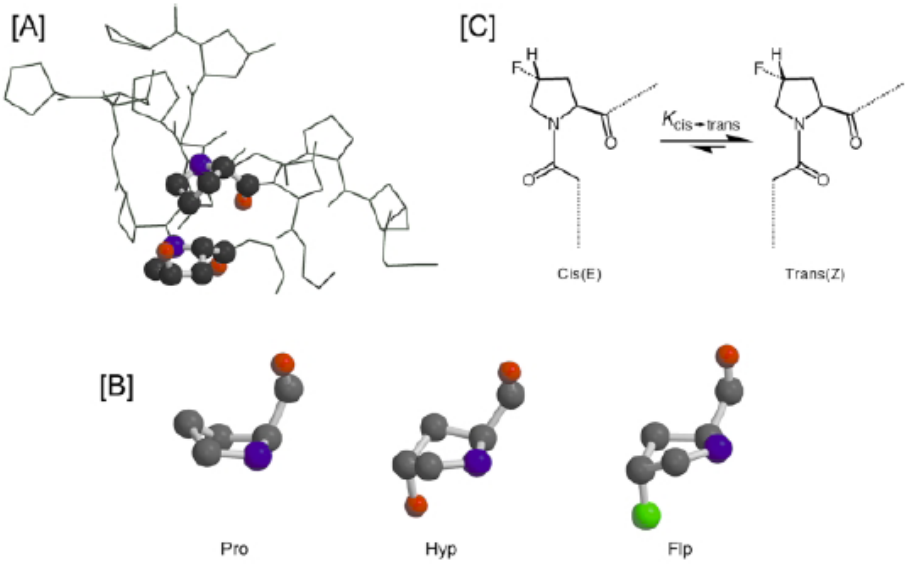


Figure 5. Stereoelectronic effect of fluorine substitution on proline rings and on collagen-mimetic peptides. [21][A] Section of the structure of a collagen mimetic peptide in which one praline and one hydroxyproline residue are shown in ball-in-stick form, showing the preferred conformation of each residue. [B] Alternate views of praline and hydroxyproline residues from structure in [A], as well as of

a 4-(R)-fluoroproline showing the similar *exo* conformation adopted by 4-(R)-substituted proline rings. [C]*Cis-trans* isomerism of the prolyl amide bond in a 4(R)-fluoroproline model compound.

Until now, a large variety of fluorinated amino acids (Figure 6) including mono-fluorinated phenylalanine[24], tryptophan, tyrosin[25] as well as trifluoroleucine[26], hexafluoroleucine[27], trifluoromethionine[28], trifluorovaline[29], 2-amino-5,5,5-trifluoro-3-methyl pentanoic acid[30], (2S,4S)-4-fluoroproline,(2S,4R)-4-fluoroproline, (2S)-4,4-difluoroproline[16] have been incorporated into target proteins, which greatly expended the diversity of protein engineering.

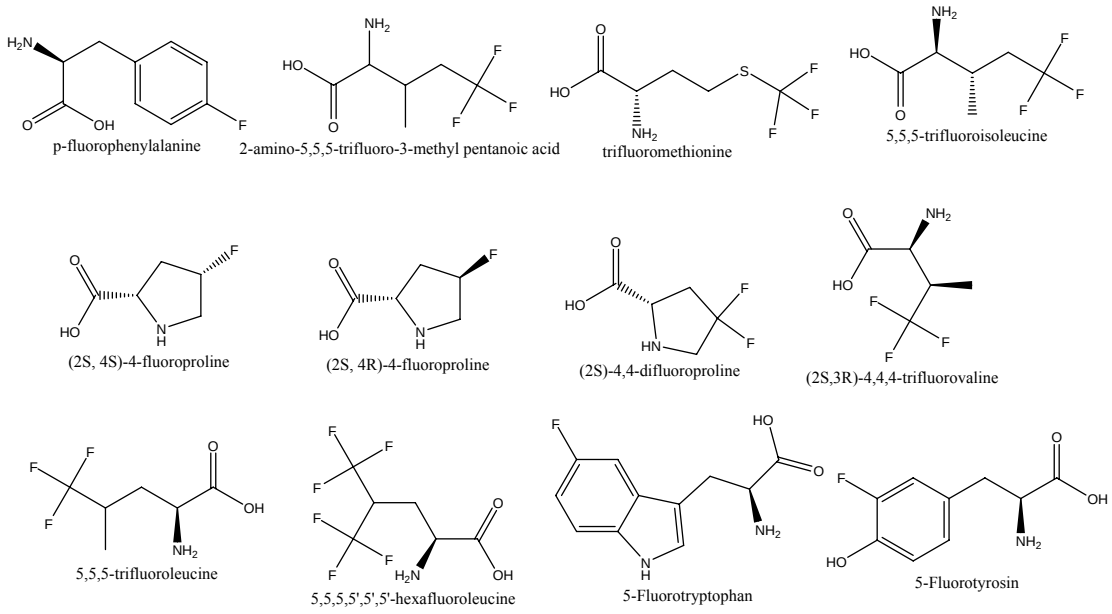


Figure 6. Fluorinated amino acids used for incorporation studies

The most straightforward way of synthesizing small fluorinated polypeptide is to use the automated peptide synthesis machine. This method allows almost any amino acid to be incorporated site-specifically into the polypeptides. However, this synthetic method is restricted to short peptides of less than 50 residues. The recent development of chemical ligation strategies, by Kent, Muir and coworkers[31, 32], which links several short peptides together under mild conditions, makes the chemical synthesis of proteins of containing up to 200 residues possible. Moreover, this technique can be employed to produce semi-synthetic

proteins by ligating the peptide containing unnatural amino acids to larger proteins produced by biosynthesis in *E. coli*.

For the synthesis of large and highly fluorinated proteins, the most efficient way is to use cell's own protein-making machinery to incorporate the non-canonical amino acids. This method relies on a series of specially engineered strains of bacteria called auxotrophes which lack the ability to biosynthesize one or more of the natural amino acids. When the strain is grown in the medium depleted of the desired natural amino acids but supplemented with the corresponding fluorinated analogues, the bacteria amino acid tRNA-synthetase (aaRS) will recognize the analogues and then catalyze the incorporation of these unnatural amino acids into the corresponding codons. When the endogenous aaRS is not active enough to activate the analogues, coexpression of the aaRS gene, which resides in the plasmid will be conducted in order to elevate the activity of the aaRS, or the active site or editing site of wild type aaRS will be modified, which allows it to charge rare unnatural amino acids and facilitate the incorporation of these amino acids into the protein. This method appears to be a really powerful tool for the production of novel proteins [21, 33].

Recently, Conticello *et al.* have successfully employed proline auxotrophic *E. coli.* strains to incorporate a structurally diverse series of proline analogues to the elastin-mimetic polypeptide sequence, in which proline accounts for one fifth of all encoded amino acids. By incorporating fluorinated analogues (2S, 4S)-4-fluoroproline and (2S, 4R)-4-fluoroproline into the target protein, they were able to identify a stereoelectronic effect on the EMP turn formation, which caused striking difference of the self-assembly upon phase transition of the corresponding EMPs. The experimental data, in combination with density functional theory (DFT) calculation led these researchers to this conclusion and provided further insight into the structural aspects of elastin protein assembly[16, 34].

On the basis of previous investigations, the structural roles of valine and isoleucine in the EMP repeat have become a topic of research interest, as nearly forty percent of EMP sequence is occupied by valine or isoleucine residues. In addition, the amino acid in the fourth position of the repeat unit, which is always valine or isoleucine, is an important determinant

of the EMP phase transition by intensively participating in the type II β turn. As a result, it is expected that the incorporation of fluorinated analogues into these positions would introduce stereoelectronic effect on the proteins, which would help discover the local structures of EMPs. Furthermore, the introduction of highly fluorinated amino acids may result in novel EMPs with 'super' properties. This chapter describes the incorporation of fluorinated valine and isoleucine analogues into the elastin-mimetic polypeptide sequence and the characterizations of the resulting EMPs.

Materials and Methods:

All chemical reagents were purchased from either Fisher Scientific (Pittsburgh, PA) or Sigma Chemical Corporation (St. Louis, MO) unless otherwise specified. Isoleucine analogue 2 (4, 4, 4-trifluorovaline) was purchased from SynQuest Laboratories (Alachua, FL), valine analogue 3 (3-fluoro-DL-valine) was obtained from matrix scientific (Columbia, SC). Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL) All restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs, Inc. (Beverly, MA). Platinum *pfx* DNA polymerase was purchased from Invitrogen Corp (Carlsbad, CA). Plasmid pPROTetE.133 was obtained from BD Biosciences, Inc. (Palo Alto, CA) and plasmid pPROLarA.231 was obtained from Professor Rik Myers of the University of Miami. The pPROTetLar vector was constructed from ligation of the *Avr* II/*Spe* I fragment of pPROTetE.133, containing the transcriptional /translational control elements, multiple cloning site, and chloramphenical resistance gene, to the corresponding fragment pPROLarA.231, which contains the p15A origin of replication. Plasmid pJET1 enconding ValRS gene was constructed by Jet Qi on the basis of pPROTetLar vector. Plasmid pAG2, which was constructed by Anna George, was employed as a source of the gene encoding elastin-mimetic polypeptide. The preparation of the plasmid pAG2 has been previously described[3]. Plasmid pZErO™-2 and *E. coli* strain TOP10F' were purchased from Invitrogen Corp (Carlsbad, CA). QIAGEN Plasmid Mini kit was purchased from QIAGEN Inc. (Chatsworth, CA). Synthetic oligonucleotides were purchased from Sigma-

Genosys, Inc (The Woodlands, TX) or Integrated DNA Technolgoies (Coralville, IA) and were used as received. NMM medium was prepared according to the protocol of Budisa *et al*. TALON metal affinity resin was obtained from BD Biosciences, Inc. His-tag AP Western Reagents were purchased from Novagen (Madison,WI). Procedures for the manipulation of DNA, the transformation of competent cells, and the growth and induction of bacterial cultures were adapted from the published literature or instructions supplied by manufactures. Reagents for the manipulation of DNA and bacterial culture were sterilized by either autoclave or passage through a 0.22 μm filter. Reactions with restriction endonucleases were performed in the reagent buffers supplied by the manufacture.

Physical and Analytical Measurements

DNA sequence analyses were performed at the Emory University Core DNA Sequencing Facility on a Perkin-Elmer ABI Prism model 377 DNA sequencer. Protein electrophoresis was performed either on 10%-15% gradient discontinuous SDS polyacrylamide gels on a phastSystem from GE Healthcare and was visualized via a silver staining procedure, or performed on 12% SDS polyacrylamide gels on a BioRad gel running system, and visualized via a commassiee blue procedure.

The molar masses of novel elastin-mimetic polypeptides were determined by MALDI-TOF MS on an Applied Biosystems Voyager System 428 mass spectrometer in the positive linear mode. The matrix, sinnapic acid, was used at a concentration of 15 mg/mL in a 50:50 mixture of water and acetonitrile. The protein solution (1 mg/ml in distilled water) was mixed with the matrix solution in a ratio of 1:1 and dried under vacuum or air. Bovine serum albumin was used as the standard for external calibration.

Solution NMR spectra were acquired with a Varian UNITY 600 instrument (564.044 MHz, ^{19}F) equipped with a 5-mm ^{19}F probe. Spectra were collected at 4℃ on specimens with a concentration of 10mg/ml dissolved in sterile H_2O/D_2O (70:30) in which the pH value was adjusted to 2.7 to retard amide proton exchange rates.

Circular dichroism (CD) spectra were recorded on a Jasco J-810 spectropolarimeter equipped with a PFD-425S Peltier temperature control unit in 1 mm sealed quartz cells at concentrations of 0.2 mg/ml (**elastin 2**) and 0.5 mg/ml (**elastin 3**) in distilled, deionized water. Temperature/wavelength CD-scans were performed within the temperature range from 5 °C to 65 °C with equilibration for 5 min at each temperature. Spectra were obtained from 260 to 190 nm at a resolution of 0.2 nm and at a scanning speed of 50 nm/min.

Plasmid construction

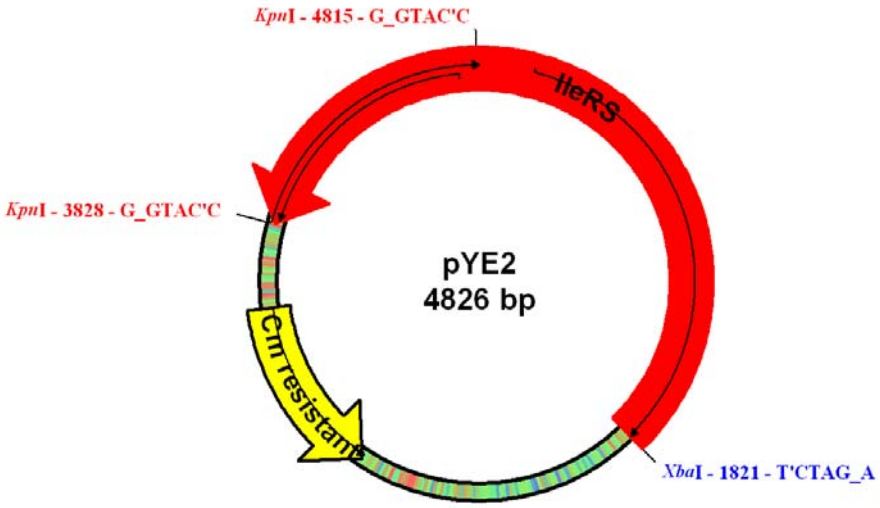


Figure 7. Plasmid map of pYE2 based on a pZErO 2 vector

The IleRS gene was amplified from the *E. coli* genomic DNA using polymerase chain reaction in which unique *Hind* III and *EcoR* I restriction sites were incorporated at the 5' and 3' termini, respectively. Cohesive ligation was performed between the double digested IleRS insert and pZerO2 vector via *Hind* III and *EcoR* I sites in the presence of T4 DNA ligase in 1X ligase buffer at 16 ˚C for 16 hrs. An aliquot of this ligation mixture (1µL) was used to transform competent cells of *E. coli* strain TOP10F' (40 µL). The transformed cells were resuspended in 1 mL SOC medium and incubated at 37 ˚C with agitation for 1 hour. The transformation mixture (100 µL) was cultured on LB/Kanamycin plates (LB: 1% tryptone, 0.5% yeast extract, 1% NaCl, pH 7.5; 1.5% Agar; 50 µg/mL kanamycin) at 37 ˚C overnight.

Six transformants were used to inoculate separate cultures of LB medium in the presence of kanamycin (50 μg/mL). The cultures were incubated at 37 ˚C with agitation overnight. Plasmid DNA was isolated from cultures by QIAGEN Plasmid Mini kit. The clones were screened by double digestion with *Hind* III and *EcoR* I. Recombinant plasmids, denoted as pYE1, were identified by analysis of the digested products with 1% agarose gel electrophoresis and confirmed by DNA sequence analysis.

Plasmid pYE1 was then digested by *Xba* I in the presence of NEBuffer 2 and BSA at 37℃ for 30 minutes and was followed by the enzymatic digestion by *Kpn* I for another 30 minutes. The two-step reaction mixture was analyzed by 1% agarose gel electrophoresis and the band with the correct size, 2.8kbp, was isolated from the gel and recovered using ZymoClean™ DNA recovery kit. Recovered DNA was then ligated to the double digested pPROTetLar vector through *Kpn* I and *Xba* I sites and generated pYE2, which was under control of the $P_L$tet promoter.

Gene expression and purification

The expression plasmid encoding elastin-mimetic polypeptide gene, pAG2, was co-transformed with either pJET1 or pYE2 into the valine and isoleucine double auxotrophic *E. coli* strain CAG18599 to generate the expression system. Single colonies of the expression strain was inoculated into sterile LB broth (25 mL) supplemented with the appropriate antibiotics (100 mg/ml ampicillin and 34 mg/ml chloramphenicol) as required for plasmid maintenance. The overnight culture was centrifuged at 4000 g for 20 min to isolate the cells, which were resuspended in sterile NMM medium (500ml) supplemented with the appropriate antibiotics. The valine or isoleucine concentration was adjusted to 50 mg/L from a sterile 100x stock solution. The culture was incubated with agitation (225 rpm) at 37℃ until the optical density at 600 nm (OD_{600}) reached between 0.8 and 1.0 absorbance units. The cells were then collected by centrifugation at 4000 g for 10 min. The cell pellet was washed with cold (4 ℃), sterile 0.9% aqueous NaCl twice (200 mL) and resuspended in sterile NMM medium containing antibiotics but without valine or isoleucine supplement. After incubation

at 37℃ for 30 min, the analogues were added to a final effective concentration of 1mM or 0.5 mM, respectively, from sterile 100X stock solutions. An aliquot of aqueous 1.0m IPTG was added to the cultures to a final concentration of 1 mM to induce synthesis of the elastin-mimetic protein. After a 4-hour induction period, the cells were harvested by centrifugation at 4000 g and 4℃ for 30 min. The cell pellet was resuspended in denaturing lysis buffer (50 mL, 50 mM sodium phosphate, 300 mm NaCl, 8M urea, pH 8.0) and shaken at room temperature overnight.

The mixture was centrifuged at 40 000 g at room temperature for 30 min. SDS PAGE analysis indicated that the majority of the target protein dissolved under these conditions. The soluble fraction was loaded onto cobalt-charged TALON resin (10 mL) that had been previously equilibrated with denaturing lysis buffer. The target protein was washed with denaturing lysis buffer (50 mL) containing 20 mm imidazole and eluted with denaturing lysis buffer (40 mL) containing 250 mm imidazole. The eluted target protein was dialyzed (molecular weight cut-off=10 kDa) against distilled water (4 L) for six times at 4℃. Lyopilization of the dialysate produced the elastin-mimetic polypeptides as white spongy solids.

Results and discussion

Plasmid construction of pYE2

The IleRS gene was amplified by PCR from *E. coli* genomic DNA and the band corresponding to the desired PCR product with the size of 2.8k bp (Figure 8) was cut from the gel and recovered, which was followed by successful insertion to the pZErO-2 vector. The recombinant plasmid pYE1 was consecutively digested by *Xba* I and *Kpn* I and the reaction mixture was treated to the 1% agarose gel. The 2.8k bp band, corresponding to the whole IleRS gene, was cut and the DNA was recovered, which was then ligated to the pPROTetLar vector to afford the expression plasmid encoding IleRS--pYE2. The identity of pYE2 plasmids was checked by double digestion with *Xba* I and *Kpn* I completely. Three of the four digestion mixtures gave three bands in 1% agarose gel (Figure 9), with the size of 1.0k bp, 1.8 k bp and 2.0 k bp, respectively, corresponding to two parts of the IleRS gene insert and the vector itself. The whole plasmid was with the correct size of 4.8 k bp.

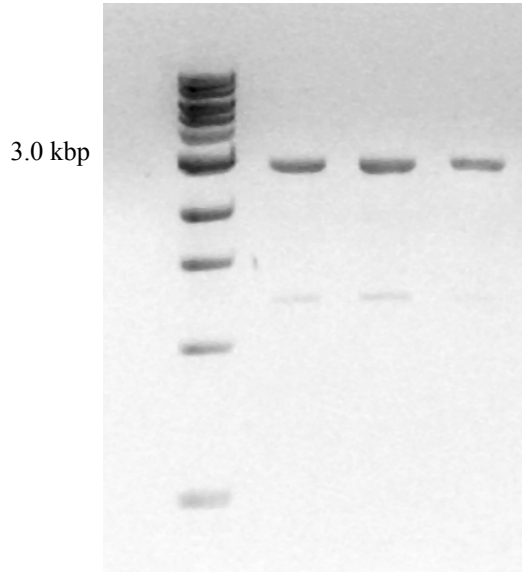


Figure 8. PCR product of IleRS from genomic DNA. Lane 1: NEB 1kb DNA ladder; Lane 2-4: PCR product of IleRS gene

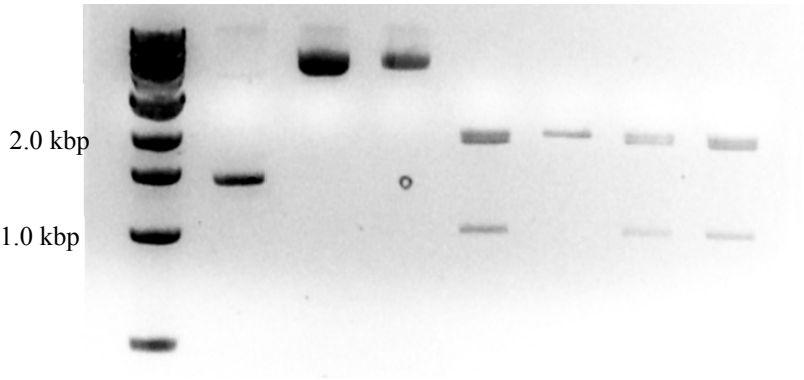
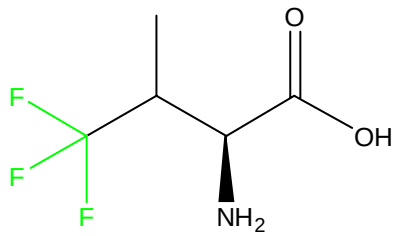


Figure 9. Double digestion check of pYE2 with *Xba* I and *Kpn* I. Lane 1: NEB 1kb DNA ladder; Lane3, 4: pYE2; Lane 5-8: Double digestion of pYE 2

Protein synthesis

Sequence for elastin-mimetic polypeptide: **$MGH_{10}S_2GHID_4KHM[(VPGVG)_4VPGIG]_{16}V$**

Using media shift protocol, **elastin 2** with 4, 4, 4-trifluorovaline (Figure 10) incorporated to all isoleucine codons was extracted from the CAG18599 strain. The expression was confirmed by the western blotting, in which a strong band at 37 kD (Figure 11) was observed, corresponding to **elastin 2**. The purification was performed on TALON cobalt resin and the pure product was observed as a single band above 35 kD (Figure 11). A yield of 25 mg was achieved for a 500ml culture and the final product was white spongy solid.



4,4,4-trifluorovaline

Figure 10. Chemical structure of the isoleucine analogue 4, 4, 4-trifluorovaline

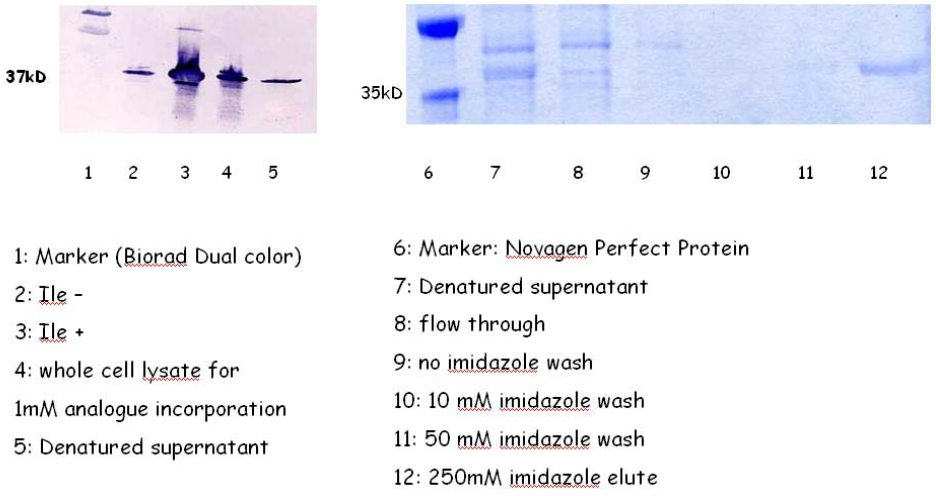
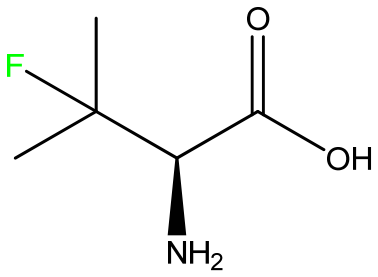


Figure 11. Expression and purification gels for **elastin 2**

Elastin 3 with 3-fluoro-DL-valine (Figure 12) incorporated to the valine codon was extracted from the same strain. The expression was confirmed by the western blotting, in which a strong band above 37 kD was observed, corresponding to elastin 2. The purification was performed on TALON cobalt resin and the pure product was observed as a single band above 35 kD (Figure 13). A yield of 15 mg was achieved for a 500 ml culture and the final product was white spongy solid.



3-fluorovaline

Figure 12. Chemical structure of valine analogue 3-fluorovaline

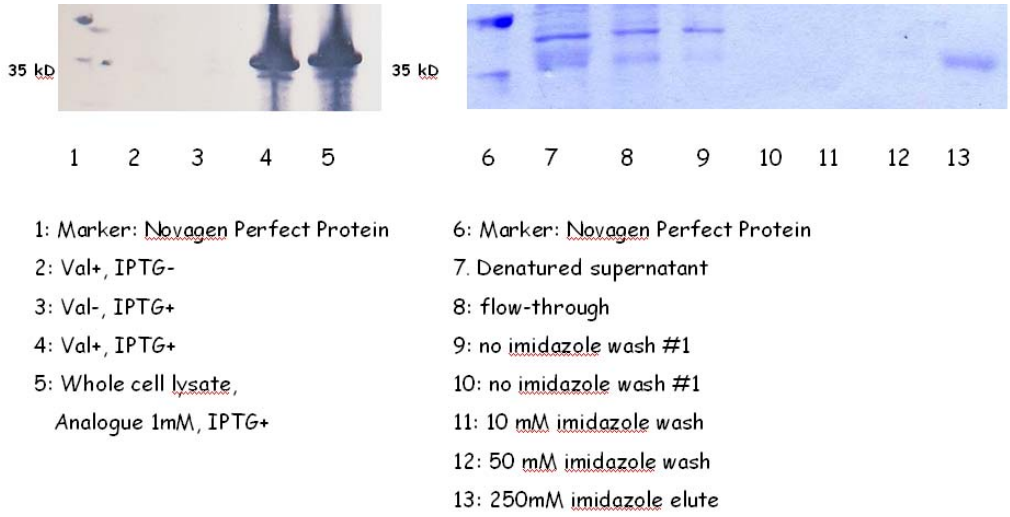


Figure 13. Expression and purification gels for **elastin 3**

MALDI-TOF mass spectrometry

Elastin 2

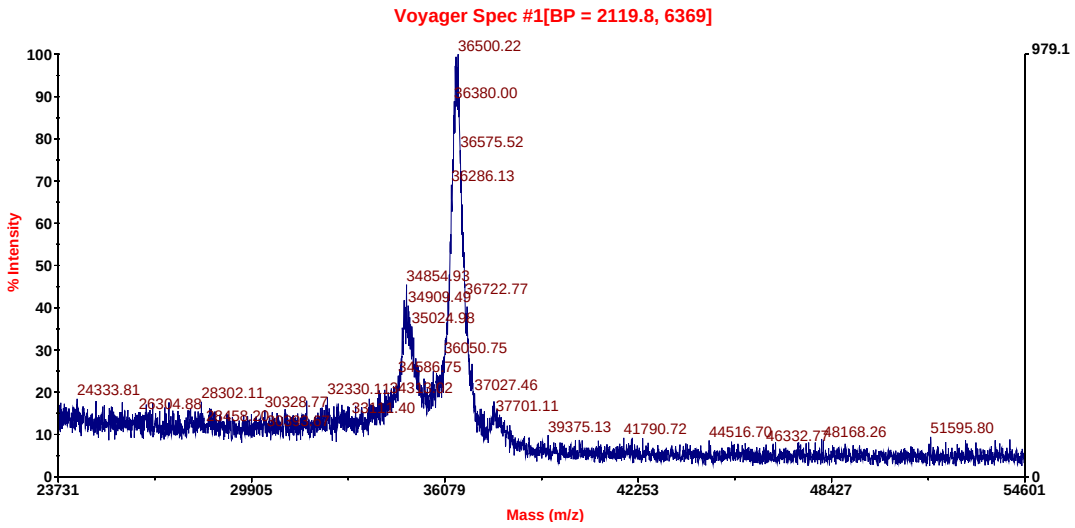


Figure 14. MALDI-TOF mass spectrum for **elastin 2**

Two peaks were observed in the MALDI-TOF mass spectrum, with the major peak showing a mass of 36500.22, which is close to the calculated molecular mass of **elastin 2**: 36534. The percentage error is 0.09%. The minor peak appears at a lower mass region, which may come from the degradation during expression or extraction. Also, it is possible that there is trace amount of isoleucine left after the medium shift, therefore not all isoleucine codons were replaced by the analogue.

Elastin 3

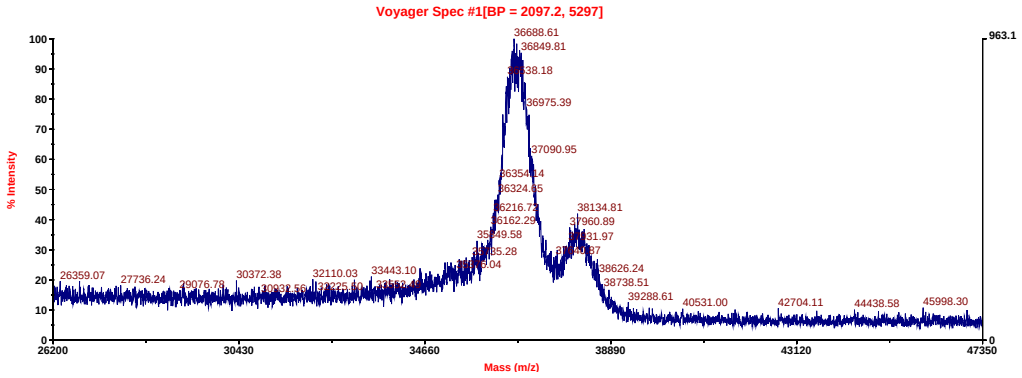


Figure 15. MALDI-TOF mass spectrum for elastin 3

Two peaks appear in the MS spectrum for **elastin 3**. The minor peak is at the position corresponding to a molecular weight of 38134.81, which matches the theoretical mass of **elastin 3**-8476, with a percentage error of 0.89%. The major peak shows the other species of mass 36688.61, which may result from degradation or deletion of the protein.

^{19}F NMR

^{19}F NMR is a powerful technique for detecting the incorporation of fluorinated analogues into proteins. On one hand, ^{19}F nucleus is very sensitive to the surrounding chemical environment and the sensitivity is only second to the proton; on the other hand, no protein in nature has been reported to contain fluorine atoms, as a result, zero background will be observed in ^{19}F NMR.

Both the two proteins (10 mg/ml) were used to conduct ^{19}F NMR study and a sharp peak at 68 ppm was observed for **elastin 2**, while two peaks were detected on the spectrum for **elastin 3,** with chemical shifts of -144 ppm and -146 ppm. The signals in the NMR spectra reflected that both of the two proteins were equipped with fluorine atoms, which confirmed the incorporation of fluorinated analogues in the elastin-mimetic polypeptides.

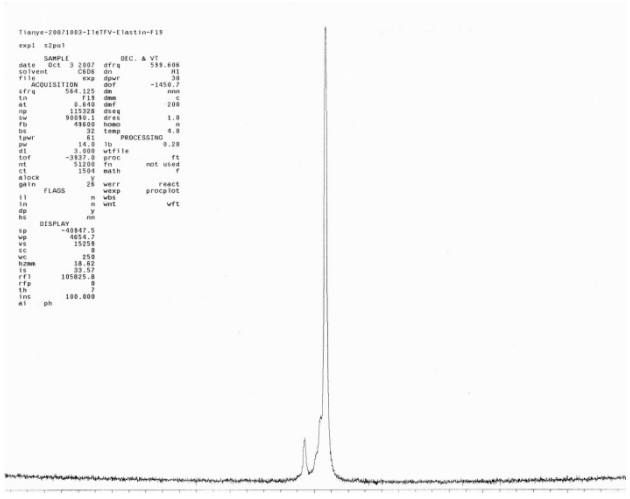


Figure 16. ^{19}F NMR spectrum for **elastin 2**

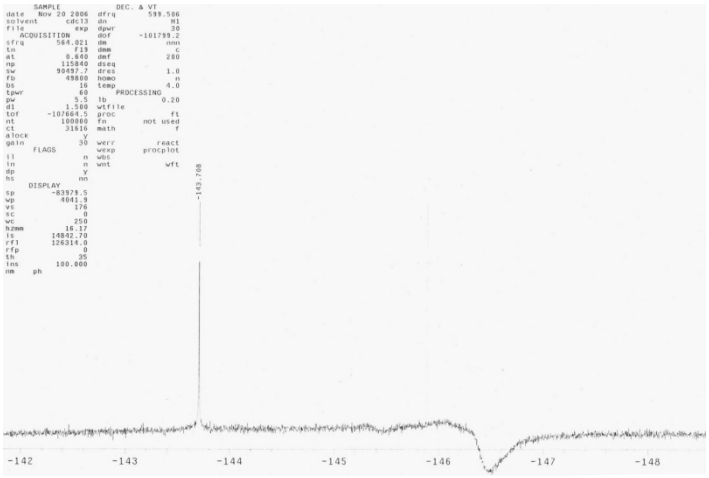


Figure 17. ^{19}F NMR spectrum for **elastin 3**

Circular Dichroism (CD)

Circular dichroism has been used to study the secondary structure of the elastin-mimetic polypeptides at different temperatures from 5℃ to 65℃ and the phase transition upon temperature increase has been detected. For both **elastin 2** and **elastin 3**, random coil was observed at low temperatures, characteristic at negative ellipticity at -198 nm. While at high temperatures, the protein self-assembled into ordered structure featured in type II β turn, featured by a shoulder at approximately 220 nm, which is similar to the native elastin-mimetic polypeptide[35-37]. These results confirmed that the introduction of fluoro group maintain protein structure, meanwhile the thermal transition patterns have not been changed.

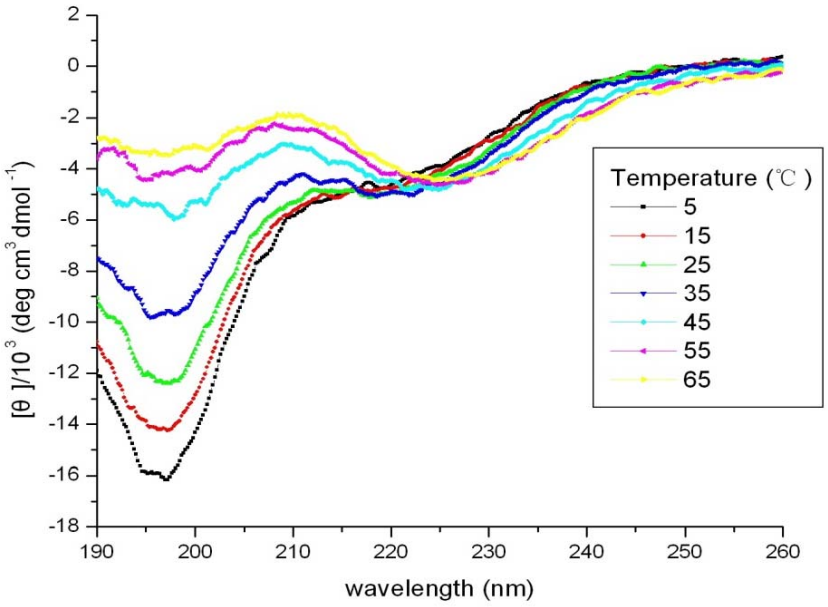


Figure 18. CD spectra for **elastin 2** from 5℃ to 65℃

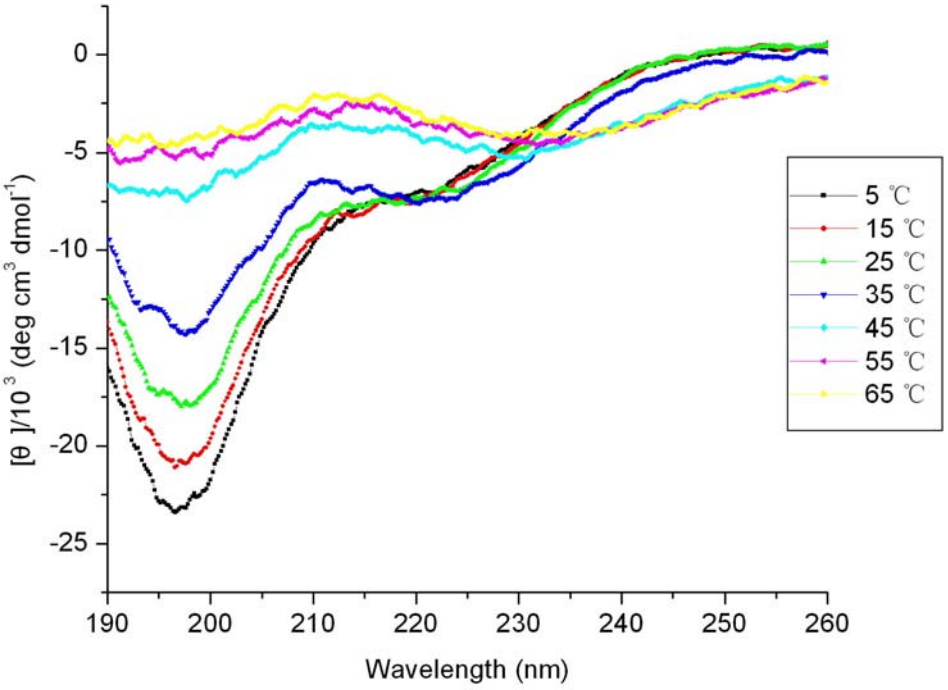


Figure 19. CD spectra for **elastin 3** from 5℃ to 65℃

Chapter 3. Design and construction of elastin-mimetic polypeptide gene suitable for the study of 4,4-difluoroproline incorporation

Background

Stereoelectronic effect on the protein conformational preference

Stereoelectronic effect has been shown to greatly influence the stability and conformational preference of proteins [15, 24, 38]. As the most abundant protein in mammals, collagen provides the human body with tissue connectivity and mechanical support. Consisting of a helix of three strands, collagen contains repeating units of X-Y-Gly, in which X positions are usually 2S-proline while Y residues are often equipped with 4R-hydroxyproline (4R-Hyp), via post-translational modification, which is widely accepted as the factor to stabilize the collagen structure. For a long time, the effect of 4R-Hyp on the triple helix stability was explained as a result of hydrogen bonding formation between hydroxyl groups. However, later investigation on the collagen-mimetic polypeptides, which incorporated stereochemical isomer 4S-Hyps into the corresponding positions, or 4R-Hyps were introduced into X positions destabilized the triple strands, thus suggesting the protein stability is highly dependent on the stereochemistry[24, 38]. That is why the stereoelectronic effect began to be considered as the contributing effect for collagen stability. Further studies have used fluorine, the most electronegative element as the probe for stereoelectronic effect, in which 4S-fluoroproline and 4R-fluoroproline were incorporated to collagen mimetic polypeptide sequence $[ProXaaGly]_7$[38]. The introduction of 4R-fluoroproline has produced hyperstable protein by enhancing melting temperature from 36℃ for 4-Hyp containing peptide to 45℃. No helix was observed for the peptide equipped with 4S-fluoroproline, like 4S-Hyp containing collagens.

Molecular modeling and high resolution structures of collagen triple helix $[ProProGly]_{10}$ have suggested that Pro in the X position adopts a C^{γ}-*endo* conformation while

C^{γ}-*exo* pucker is dominant for the proline residues in Y positions[39]. As we have known, the electron-withdrawing hydroxyl or fluoro group would orient itself in the gauche position to the amide nitrogen in the pyrrolidine ring, called 'gauche effect'. As a result, 4R-Hyp and 4R-fluoroproline prefer C^{γ}-*exo* conformations while 4S-Hyp and 4S-fluoroproline prefer C^{γ}-*endo* conformations. In contrast, although C^{γ}-*endo* conformation in a pyrrolidine ring in 2S-proline is slightly preferred, both *endo* and *exo* conformations are populated at room temperature[40]. As a result, if Y positions of the repeats are replaced by 4R-Hyp, which occurred in nature, or 4R-fluoroproline, the C^{γ}-*exo* preference would be satisfied, thus stabilizing the triple-helix structure. However, if the 4S diastereomers are incorporated, or X positions are occupied by 4S diastereomers, the resulting proteins would be highly disfavored energetically, reflected by the failure to form triple helix experimentally. Above is the mechanistic explanation of the stereoelectronic effect.

Elastin is another important protein in extracellular matrix that cooperates with collagen to support the body's tissues. This combination of collagen and elastin is very essential in many parts of the body, including the lungs, bones, and tendons. Even the blood vessels heavily rely on both of them. One of the common features between these two proteins is that both of them contain proline-rich domains. For example, in one of the common repeating sequence of elastin-mimetic polypeptide, VPGVG, proline has made up of 20% of all the residues. Therefore, inspired by the stereoelectronic studies on collagen-mimetic polypeptides, the same effect have been investigated regarding the proline residues of elastin-mimetic polypeptide, in order to understand the structural features of elastin. Conticello *et al* has successfully incorporated a series of fluorinated proline analogues into the elastin mimetic polypeptide sequence. These analogues include 4R-fluoroproline, 4S-fluoroproline, 3R-fluoroproline, 3S-fluoroproline and 4, 4-difluoroproline (Figure 20). Experimental data including CD, DSC, and two-dimensional NMR have revealed that elastin-mimetic polypeptide incorporated with 4R-fluoroproline was easier to self-assemble into ordered structure upon temperature increase, reflected by the lower transition temperature than native protein. This assembly facilitation has been accompanied by the increase of type II β turn

percentage above the transition temperature. However, the diastereomer 4S-fluoroproline has the opposite effect on both the protein self-assembly and the secondary structure, with increased transition temperature and the destabilization of type II β turn. Computational studies have pointed out that C^{γ}-*exo* pucker of the pyrrolidine ring in 4R-fluoroproline energetically stabilizes the type II β turn while C^{γ}-*endo* conformation of the 4S diastereomer has the destabilization effect. This result also offered evidence that type II β turn do play important roles in the self-assembly of the elastin-mimetic polypeptide. The 3-fluoroproline diastereomers have also been incorporated to the same sequence by Conticello group, and the investigation of resulting elastin-mimetic polypeptides, together with the analysis of small molecule models have suggested similar results as that of the 4-fluoroproline pair [15, 16, 41].

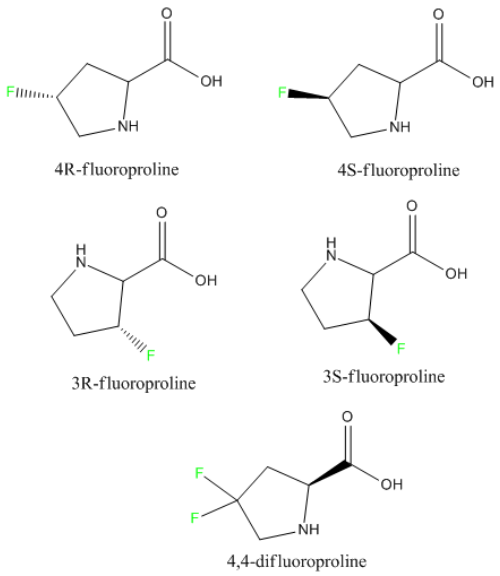


Figure 20. Fluorinated proline analogues incorporated to the elastin mimetic polypeptide

4, 4-difluoroproline has been proved to be good surrogate for proline and this non-canonical amino acid has been incorporated into the elastin-mimetic polypeptide with almost 100% replacement[16]. Due to the striking stereoelectronic effect of 4-fluorproline and 3-fluorproline pairs on the self-assembly of the corresponding elastin-mimetic polypeptides, together with the fact that fluorine atom has very special electronic properties, it will be worthy to explore how this perfluoro-molecule influences the structure and self-assembly of

the resulting elastin-mimetic polypeptide. CD spectra by Wookhyun Kim at different temperatures shown that even at temperature as low as 5 ℃, the protein has self-associated to ordered structure in which type II β turn is the major structural feature [Unpublished data]. This result suggested that the incorporation of 4.4-difluoroproline facilitates the assembly process, but in the other word, this facilitation has prevented the study of this highly fluorinated protein in terms of its phase transition. In order to conduct research on the effect of this amino acid on protein self-assembly, a new sequence suitable for 4,4-difluoroproline incorporation should be designed, with the criteria that the new protein closely mimic the native elastin and the lower critical solution temperature of the protein is much lower than the original one.

Design of new elastin mimetic polypeptide sequence

Transition temperature (T_t) of an elastin-mimetic polypeptide can be modulated by a series of factors including protein composition, in which the fourth residue of the pentapeptide repeat is determinant. Dan Urry[11] has conclusively discussed the effect of composition on the transition temperature by mutating the fourth position into a wide variety of amino acids including hydroxyproline. The study demonstrated that by replacing 99.7% 4th position valine with alanine, the resulting protein undergoes inverse temperature transition at the temperature that is 21℃ lower than that of the original poly(VPGVG). Inspired by Urry's result, VPGAG is proposed as the repeating sequence for the new elastin-mimetic polypeptide. It is expected that when 4, 4-difluoroproline is incorporated to the new protein sequence, the phase transition would be detected within experimental range. The results showing how mutations of the 4th position valine influence the transition temperature (T_t) of the corresponding elastin-mimetic polypeptides are shown below.

Amino acid X	T_t [°C]	Correlation coefficient
Lys(NMeN, red.) [a]	−130	1.000
Trp (W)	−90	0.993
Tyr (Y)	−55	0.999
Phe (F)	−30	0.999
His (pH 8) (H)	−10	1.000
Pro (P)	(−8)	[b]
Leu (L)	5	0.999
Ile (I)	10	0.999
Met (M)	20	0.996
Val (V)	24	[c]
Glu($COOCH_3$) (E^m)	25	1.000
Glu(COOH) (E)	30	1.000
Cys (C)	30	1.000
His (pH 4) (H^+)	30	1.000
Lys(NH_3) (K)	35	0.936
Asp(COOH) (D)	45	0.994
Ala (A)	45	0.997
HyP	50	0.998
Asn (N)	50	0.997
Ser (S)	50	0.997
Thr (T)	50	0.999
Gly (G)	55	0.999
Arg (R)	60	1.000
Gln (Q)	60	0.999
Lys(NH_3^+) (K^+)	120	0.999
Tyr(Ph-O$^-$) (Y$^-$)	120	0.996
Lys(NMeN, ox.) [a]	120	1.000
Asp(COO$^-$) (D$^-$)	170	0.999
Glu(COO$^-$) (E$^-$)	250	1.000
Ser(PO_4^{2-})	1000	1.000

Table 1. Temperature of the inverse temperature transition, T_t for poly[f_v(VPGVG)f_x(VPGXG)] [11]

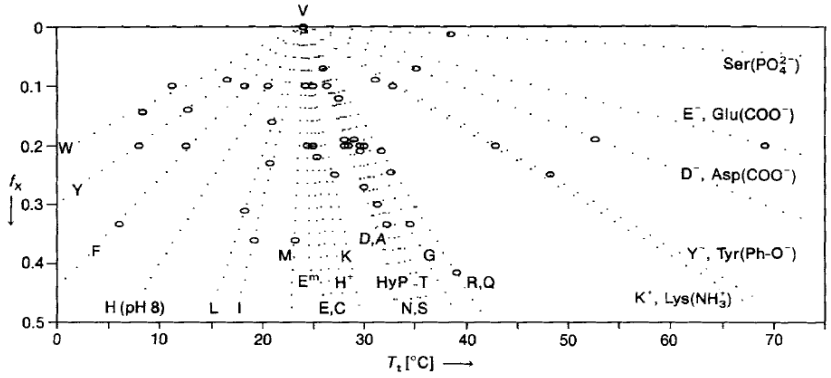


Figure 21. Plots of T_t for poly[f_v(VPGVG)f_x(VPGXG)] in PBS as a function of f_x, at small f_v. At such low X content the plots are essentially linear, and the relative values can be compared for a reference composition [11]

Materials

All chemical reagents were purchased from either Fisher Scientific (Pittsburgh, PA) or Sigma Chemical Corporation (St. Louis, MO) unless otherwise specified. All restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs, Inc. (Beverly, MA). Plasmid pZErO™-2 and *E. coli* strain TOP10F' were purchased from Invitrogen Corp (Carlsbad, CA). QIAGEN Plasmid Mini kit and Qiafilter Plasmid Maxi kit were purchased from QIAGEN Inc. (Chatsworth, CA). Synthetic oligonucleotides were purchased from Sigma-Genosys, Inc (The Woodlands, TX) and were used as received. Procedures for the manipulation of DNA, the transformation of competent cells, and the growth and induction of bacterial cultures were adapted from the published literature or instructions supplied by manufactures.

General Methods

Reagents for the manipulation of DNA and bacterial culture were sterilized by either autoclave or passage through a 0.22 μm filter. Reactions with restriction endonucleases were performed in the reagent buffers supplied by the manufacture. DNA sequence analyses were performed at the Emory University Core DNA Sequencing Facility on a Perkin-Elmer ABI Prism model 377 DNA sequencer.

Gene construction

We designed an elastin-mimetic polypeptide monomer sequence [denoted as $(VPGAG)_5$]. Purified oligonucleotides corresponding to the sense and anti-sense strands of the DNA monomer were dissolved in EB buffer (10 mM Tris-HCl, pH 8.5) to a final concentration of 0.5 μg/μL. The oligonucleotides (20 μL each) were mixed with 4M NaCl (5 μL), 1M MgSO4 (4 μL), and sterile water (151 μL), and heated to 99 ˚C to melt residual secondary structure. The mixture was gradually cooled to 30 ˚C over a period of 4 hours and incubated at 4 ˚C to anneal the two strands together. The mixture was analyzed by preparative agarose gel electrophoresis (2% NuSieveGTG agarose, 1X TBE buffer), which indicated a single band at the correct position versus the DNA standards. The synthetic duplex was isolated by ethanol

precipitation and dried *in vacuo*. Then the duplex was phosphorylated by T4 polynucleotide kinase (50 U) in 1X T4 DNA ligation buffer (150 µL) at 37 ˚C for 1 hour. The reaction mixture was extracted with phenol/chloroform/isoamyl alcohol (25:24:1) to remove the enzyme, and the phosphorylated DNA was isolated by ethanol precipitation, dried *in vacuo*, and dissolved in 50 µL EB buffer.

Cohesive end ligation was performed between the synthetic DNA cassette and the double-digested pZErO-2 vector by *Hind* III and *BamH* I in the presence of T4 DNA ligase in 1X ligation buffer at 16 ˚C for 16 hours. An aliquot of this ligation mixture (1 µL) was used to transform competent cells of *E. coli* strain TOP10F' (40 µL). The transformed cells were resuspended in 1 mL SOC medium and incubated at 37 ˚C with agitation for 1 hour. The transformation mixture (100 µL) was cultured on LB/Kanamycin plates at 37 ˚C for 16 hours. Six transformants were used to inoculate separate cultures of LB medium in the presence of kanamycin (50 µg/mL). The cultures were incubated at 37 ˚C with agitation for 16 hours. Plasmid DNA was isolated from cultures by QIAGEN Plasmid Mini kit. The clones were screened by double digestion with *Hind* III and *BamH* I. Recombinant plasmids were identified by analysis of the digested products with 2% NuSieveGTG agarose gel electrophoresis and confirmed by DNA sequencing. The new plasmid encoding the monomer was denoted as pTIAN5.

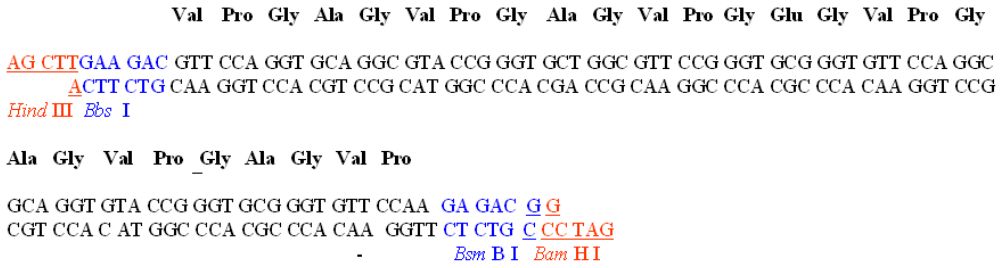


Figure 22. Sequence for the monomer $(VPGAG)_5$

A single colony containing the correct recombinant plasmid was used to inoculate 500 mL of LB medium in the presence of kanamycin (50 µg/mL). The plasmid was isolated from cell cultures via QIAfilter Plasmid Maxi kit and was double digested by *Bbs*

I and *BsmB* I to liberate the $(VPGAG)_5$ monomers, which were separated from other fragments by preparative agarose gel electrophoresis (2% NuSieveGTG agarose, 1X TBE buffer). A 102 bp band, which corresponded to the expected size of the DNA monomer, was excised from the gel, macerated and extracted by EB buffer. The agraose gel was removed by filtration through a 0.45 μm microcentrifuge filter, and the DNA monomer was recovered from the filtrate via ethanol precipitation.

The plasmid pTIAN5 was digested by *Bbs* I in the corresponding buffer and the purified linear DNA was incubated with digested pTIAN5 in the presence of T4 DNA ligase in 1X ligation buffer at 16 ˚C for 16 hours. An aliquot of this ligation mixture (1μL) was used to transform competent cells of *E. coli* strain TOP10F' (40 μL). After incubated in 1ml SOC medium at 37℃ for 1 hour, the transformation mixture (100 μL) was cultured on LB/Kanamycin plates at 37 ˚C for 16 hrs. Six transformants were used to inoculate separate cultures of LB medium in the presence of kanamycin (50 μg/mL). The cultures were incubated at 37 ˚C with agitation for 16 hours. Plasmid DNA was isolated from cultures by QIAGEN Plasmid Mini kit. The clones were screened by double digestion with *Hind* III and *BamH* I, which indicated that plasmids with a variety of multimer inserts with different sizes ranging from 500 bp to 900 bp were produced. The plasmid with 600 bp long insert, which was named pYT600, was double digested with *BsmB* I and *Nco* I and a 2300 bp segment was recovered from the 1% agarose gel. Another aliquot of the plasmid was double digested with *Bbs* I and *Nco* I and a 2100 bp segment was recovered from the 1% agarose gel. These two segments were ligated together with T4 DNA ligase. A multimer with a size of 1200 bp was created within the isolated plasmid, pYT1200 and the identity of pYT1200 was confirmed by DNA sequencing.

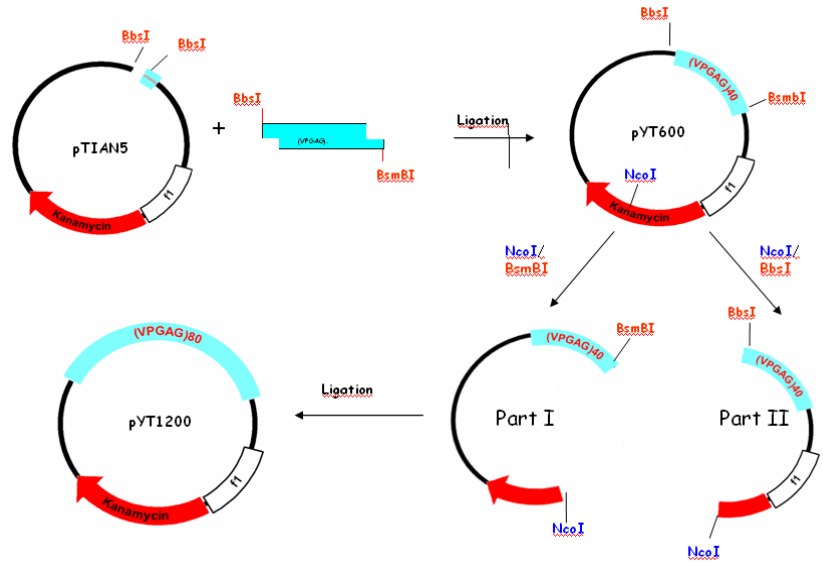


Figure 23. Scheme for the creation of multimer gene

Results and discussion

The elastin-mimetic polypeptide monomer sequence $(VPGAG)_5$ was successfully inserted into the pZErO-2 vector and the recombinant plasmid pTIAN5 was extracted using QIAfilter Plasmid Maxi kit. The plasmid was double digested with *Bbs* I and *BsmB* I to release the $(VPGAG)_5$ monomer, which was separated by electrophoresis on 2% NuSieveGTG agarose gel (Figure 24). The 100 bp band was cut from the gel and the monomer was isolated from gel fragments and purified by ethanol precipitation.

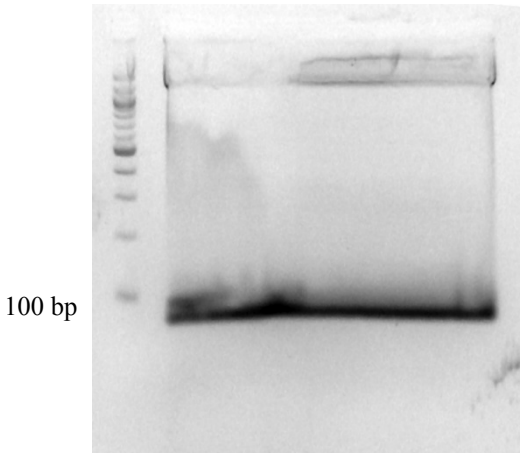


Figure 24. $(VPGAG)_5$ monomer on 2% NuSieveGTG agarose gel

The recombinant plasmid was digested by *Bbs I* and then dephosphorylated by *Antarctic Phosphatase* to afford the acceptor plasmid. Then the recovered monomer $(VPGAG)_5$ was ligated with the acceptor and the ligation mixture was used to transform TOP10F' competent cell. Recombinant plasmids were isolated from the screened colonies using the QIAGEN plasmid mini kit and were checked by double digestion by *Hind* III and *BamH* I. Digestion results revealed that a series of plasmids with inserts of different size varying from 400 bp to 900 bp were produced (Figure 25). Since elastin-mimetic polypeptide gene with size between 1000bp to 1500 bp usually produce polypeptides with good solubility, plasmids with 600bp insert and 700 bp insert, which were denoted as pYT600 and pYT700, were selected as the starting materials for the next step. Plasmid pYT600 was digested with *Bbs* I/*Nco* I restriction enzyme pair and *BsmB* I/*Nco* I pair, respectively. Each digestion mixture was separated on a 1% agarose gel and the upper bands, with the size of 2300 bp for *Bbs* I/*Nco* I pair and 2100 bp for *BsmB* I/*Nco* I pair were cut from the gel, extracted and purified by ZymoClean DNA recovery kit. The two purified DNA fragments were ligated with the presence of T4 DNA ligation buffer at 16℃ for 16 hours, and the ligation mixture was used to transform TOP10F'. Recombinant plasmids, named pYT1200 were isolated from 4 screened transformants and were then double digested by *Hind* III and *BamH* I. The four digestion mixtures were analyzed on 1% agarose gel and the result showed that all the plasmids were equipped with 1.2 kbp insert (Figure 27), suggesting that the elastin-mimetic polypeptide gene with desired size was created. The identity of the multimer was confirmed by DNA sequencing analysis. Using the same strategy, pYT1400 with 1.4 kbp insert and pYT1300 with 1.3 kbp insert were constructed separately from pYT600 and pYT700.

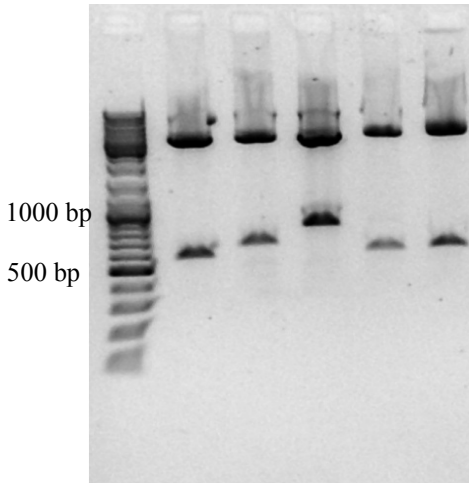


Figure 25. 1% agarose gel for the recombinant plasmids with different inserts

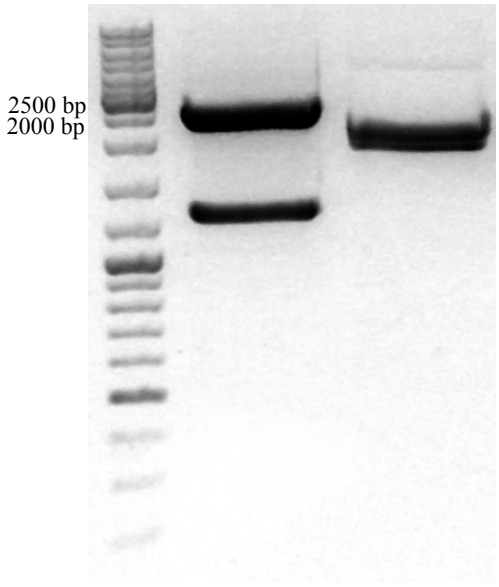


Figure 26. 1% agarose gel for double digestion of pYT600 with BbsI/NcoI (Lane 2) and BsmbI/NcoI (Lane 3) pairs.

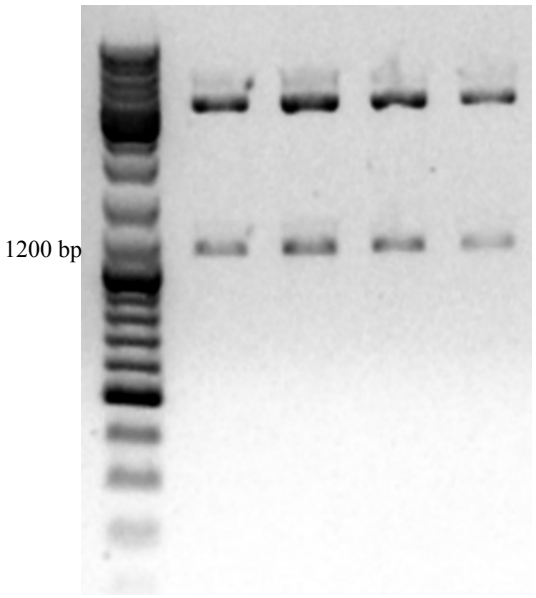


Figure 27. 1% agarose gel for double digestion check for pYT1200, all the four plasmids were equipped with the desired insert of 1.2kbp size.

Chapter 4. Characterization of biodegradable materials based on elastin-mimetic polypeptide

Background

Magnificent developments in biomaterial engineering have been achieved during the past several decades, in which biomaterials have been successfully used in clinical practice and have profoundly improved people's health and life quality. The goal of early biomaterials research was to create appropriate physical properties to replace host tissues[42]. This requires that these biomaterials are biologically inert to the human body in order to minimize the repulsive response from the human immune system. On the basis of previous research, new generations of biomaterials have been designed to stimulate specific cellular response at the molecular level[42], which requires the precise control on the molecular architecture and macromolecular properties of the materials. In this aspect, genetic engineering based biosynthesis of the proteins provides a powerful method to produce target with high yield and desired purity, as well as nearly complete control of the composition, sequence and molecular weight[43-45]. Till now, a large variety of protein based materials have been obtained using this strategy, some with non-canonical amino acids, with enhanced or unusual physical and biological properties [16, 20, 26, 27, 46].

Elastin-mimetic polypeptides (EMPs) are one of the families commonly used as building blocks for biomaterials because of their biocompatibility and controlled thermo-sensitivity. In a common repeating sequence for elastin-mimetic polypeptide, Val-Pro-Xaa-Yaa-Gly, third position residue Xaa is determinant of the mechanical behavior of the corresponding protein[13, 34] . For example, a substitution of Xaa from Gly to Ala will convert the protein from elastomeric to plastic. Recently, EMP based block copolymers, in which all the blocks are derived from EMP sequences, but with strikingly varied mechanical property, have been constructed and characterized. At Conticello group, a BAB type triblock copolymer was synthesized and characterized to investigate the potential of this material for biomedical applications[4, 14]. The sequence is shown below, in which end block B displays

plastic mechanical behavior while the central block A exhibits elastomeric mechanical response. The whole polymer is expected to display behavior that is similar to that of synthetic thermoplastic elastomers.

B=VPAVG[(IPAVG)$_4$(VPAVG)]$_{16}$IPAVG

A=VPGVG[(VPGVG)$_2$VPGEG(VPGVG)$_2$]$_{30}$VPGVG

It was observed that this protein underwent a reversible so-gel transition over a narrow temperature range. The phase transition of this protein was studied by DSC, which displayed an endothermic transition at 23 ^{o}C, reversible and virtually independent of the pH of the solutions. NMR spectroscopic study, morphology analysis by SEM, as well as the dynamic rheological study suggested the formation of a cross-linked network, in which selective microscopic phase separation occurred to the end plastic blocks, connected via the bridging midblock elastic domains. It is expected that materials derived from this protein may be used for in-vivo applications including tissue regeneration.

In order to make a material good candidate for biomedical uses, the biodegradable property has to be taken into consideration. This criterion refers that after the biomaterials are implanted into the human body and have finished their functions, they can be degraded by the components in the human system, into small pieces which can be metabolized and excreted. In order to make the EMP-based triblock copolymer more ideal for biomedical uses, plasmin-degradable sites (VRN) were incorporated into the central elastomeric blocks. Two BAB triblock copolymers, in which end blocks (S3) are plastic and central blocks (S1VRN) are elastomeric and sensitive to plasmin were designed and produced by Yang Feng successfully. [Unpublished data] The sequences and theoretical molecular weight for the two protein based polymers are shown below.

S3 (58kD) S1VRN 800/1600 (21/42 kD) S3

Polymer-800: (IPAVG)$_{133}$ -[(VPGAG)$_2$GVRNGG(VPGAG)$_2$]$_{10}$-(IPAVG)$_{133}$

Molecular weight: 137 kD

Polymer-1600: (IPAVG)$_{133}$ -[(VPGAG)$_2$GVRNGG(VPGAG)$_2$]$_{20}$-(IPAVG)$_{133}$

Molecular weight: 158 kD

The polymers were hyper-expressed in *E. coli* strain BL21(DE3) with a C-terminal decahistidine tag and were purified by a hot-cold spin method due to the phase transition property. Expression and purification were monitored by SDS-PAGE electrophoresis.

This chapter describes the characterizations of these two protein-based polymers in terms of their chemical composition, secondary structure, thermal transition profiles, as well as the enzymatic degradation by plasmin.

Materials and methods

Sodium chloride was purchased from EMD Chemicals Inc. (Gibbstown, NJ), tris base was obtained from Fisher Scientific (Pittsburgh, PA). Other chemicals including human plasmin, deuterium oxide, and sodium azide were purchased from Sigma Chemical Corporation (St. Louis, MO). Reagents for the manipulation of proteins were sterilized by either autoclave or passage through a 0.22 μm filter.

Physical and Analytical Measurements

Solution NMR spectra were acquired with a Varian INOVA 600 instrument (599.742 MHz, 1H) equipped with a 5-mm $^{1}H/^{13}C$ probe. Spectra were collected at 4℃ on specimens with a concentration of 44.4 mg/ml (**polymer-800**), and 12.5 mg/ml (**polymer-1600**) dissolved in sterile H_2O/D_2O mixture (30: 70 for **polymer-800** and 70:30 for **polymer-1600**) in which the pH value was adjusted to around 2.7 to retard amide proton exchange rates. Standard solvent suppression techniques were employed to reduce signal due to the residue water protons in the proton NMR solutions. Chemical shifts for ^{1}H NMR spectra were referenced and reported relative to the water peak (4.80 ppm). Two dimensional ^{1}H-^{1}H COSY NMR spectra were acquired at a spectral width of 5568.3 Hz, with 512 t_1 increments and 2048 complex data points with 48 scans. ^{1}H-^{1}H NOESY NMR spectra were acquired in phase-sensitive mode using the hypercomplex method with a mixing time of 500 ms at a spectral width of 6000.2 Hz. Spectra were collected with 512 t_1 increments and 2048 complex data points with 16 scans. The NMR data were processed with the program NutsPro (Acorn NMR, Inc.)

Circular dichroism (CD) spectra were recorded on a Jasco J-810 spectropolarimeter equipped with a PFD-425S Peltier temperature control unit in 1 mm sealed quartz cells at concentrations of 0.68 mg/ml (**polymer-800,** 12.4 μM), and 0.64 mg/ml (**polymer-1600,** 10.1 μM) in distilled, deionized water. Temperature/wavelength CD-scans were performed within the temperature range from 5 °C to 45 °C with equilibration for 5 min at each temperature.

Spectra were obtained from 260 to 190 nm at a resolution of 0.2 nm and at a scanning speed of 50 nm/min.

The inverse temperature transitions of the tri-block copolymers were monitored as a function of temperature using an ultra-sensitive differential scanning calorimeter (VP-DSC MicroCal, LLC, Northampton, MA). Protein samples were dissolved in distilled, deionized water at 4 °C in concentrations of 1.7 mg/ml (**polymer-800**), and 1.6 mg/ml (**polymer-1600**), degassed under dynamic vacuum and scanned from 4 to 40 °C at a rate of 40 °C /h. Reversibility was tested by cooling and rescan of samples *in situ*. DSC data were processed using the program Origin (MicroCal, LLC, Northampton, MA), and T_t, ΔH, and ΔC_p values were calculated by curve fitting using the Levenberg/Marquardt nonlinear least-squares method.

Results and Discussion

Nuclear magnetic resonance spectroscopy (NMR)

^{1}HNMR of the proteins in pure D_2O and D_2O/H_2O mixture, together with two-dimensional NMR spectroscopy including COSY and NOESY provide data that support the primary sequence of the two proteins. The shift assignments were made according to previous reported data on other similar polypeptide sequences. The tables for chemical shift assignments for all the major residues and the spectra are shown below

Residue	NH (ppm)	αH (ppm)	βH (ppm)	Other (ppm)
Val	7.99	3.82	1.77	δCH_3 0.67
Pro	----	4.10	2.01, 1.68	δCH_2 3.60, 3.39 γCH_2 1.68, 1.68
Gly	8.27	3.69	----	----
Ile	7.90	4.19	1.59	δCH_2 1.21, 0.87 γCH_3 0.67 δCH_3 0.57
Ala	8.37	4.03	1.10	-----

Table 2. Assignments for the ^{1}HNMR for **polymer-800**. Chemical shifts are referenced and reported relative the water peak.

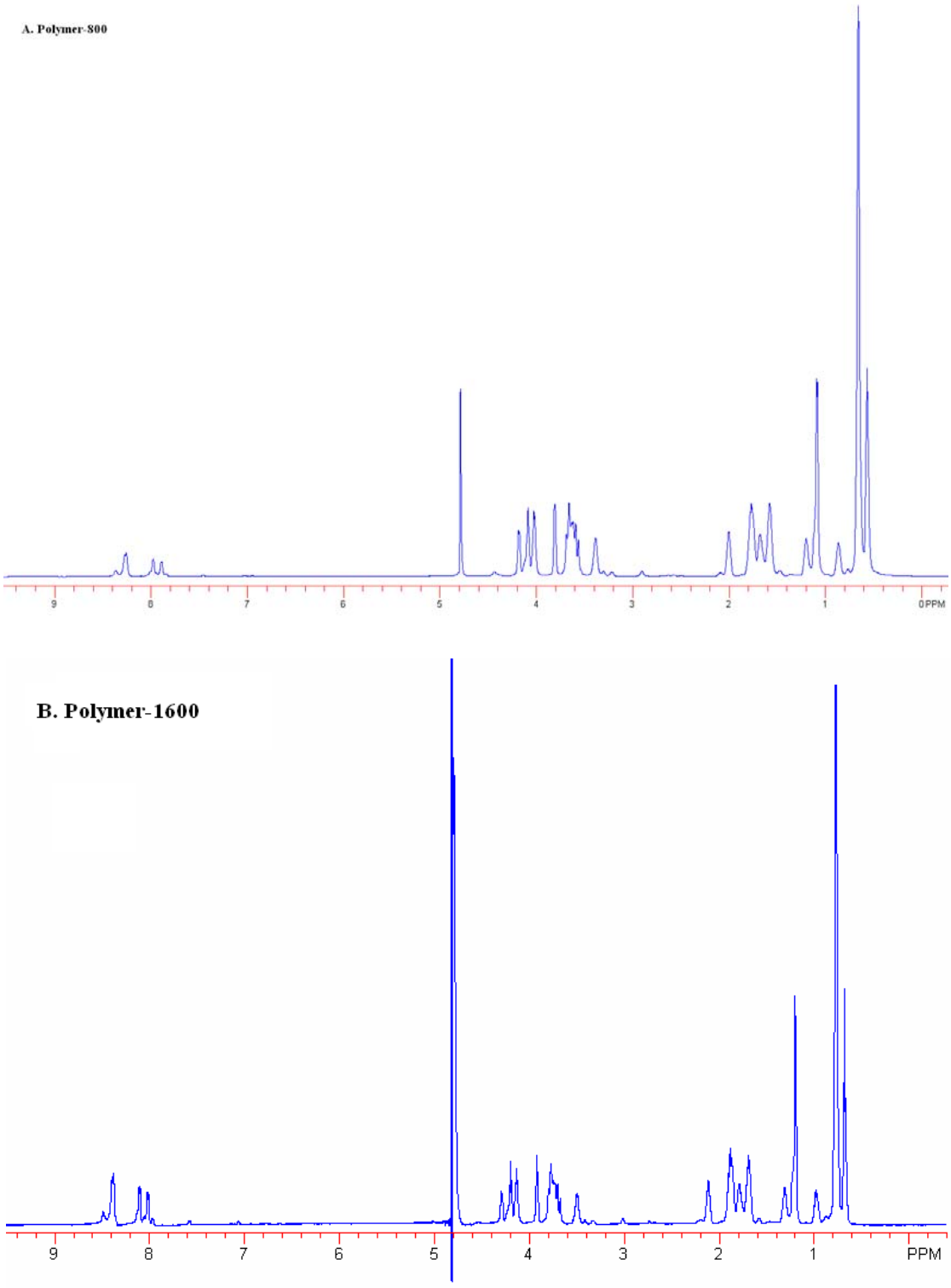


Figure 28. [1]HNMR for **polymer-800** (A) and **polymer-1600** (B). The water peak was used as internal reference and was set to be 4.80 ppm.

Conformational analysis by Circular Dichroism (CD)

Circular dichroism (CD) spectroscopy was employed to examine the secondary structure development of the triblock copolymers during the thermal transition in dilute aqueous solutions. The CD studies indicated that conformational rearrangements occurred when the temperature was increased above the transition temperature, corresponding to a secondary structure transition from a random coil conformation to a more ordered structure. The temperature-dependent transitions for both **polymer-800** and **polymer-1600** were detected and monitored by CD spectra as a function of temperature, in which the random coil (featured by negative ellipticity near 198 nm) is gradually replaced with an ordered secondary structure, (featured by a shoulder at approximately 220 nm), as the temperature was raised through the critical point. Interestingly, the ordered structure observed for these two proteins is different from type II β turn, the secondary structure observed for the central elastic block at high temperatures. Instead, this secondary structural signature is more like type I β turn [15, 41]. Moreover, the thermal transition curves for the disappearance of the random coil conformation, reflected by the shift in mean residue ellipticity at the wavelength associated with the random coil conformation, $[\theta]_{198}$, suggest that a two-state transition took place for both the two copolymers as the temperature increase [[15]. Mathematical fit of the thermal transition curves, combined with the differentiation calculation, determines the transition temperature of 20.0 ℃ for **polymer-800** and 19.8 ℃ for **polymer-1600**, respectively.

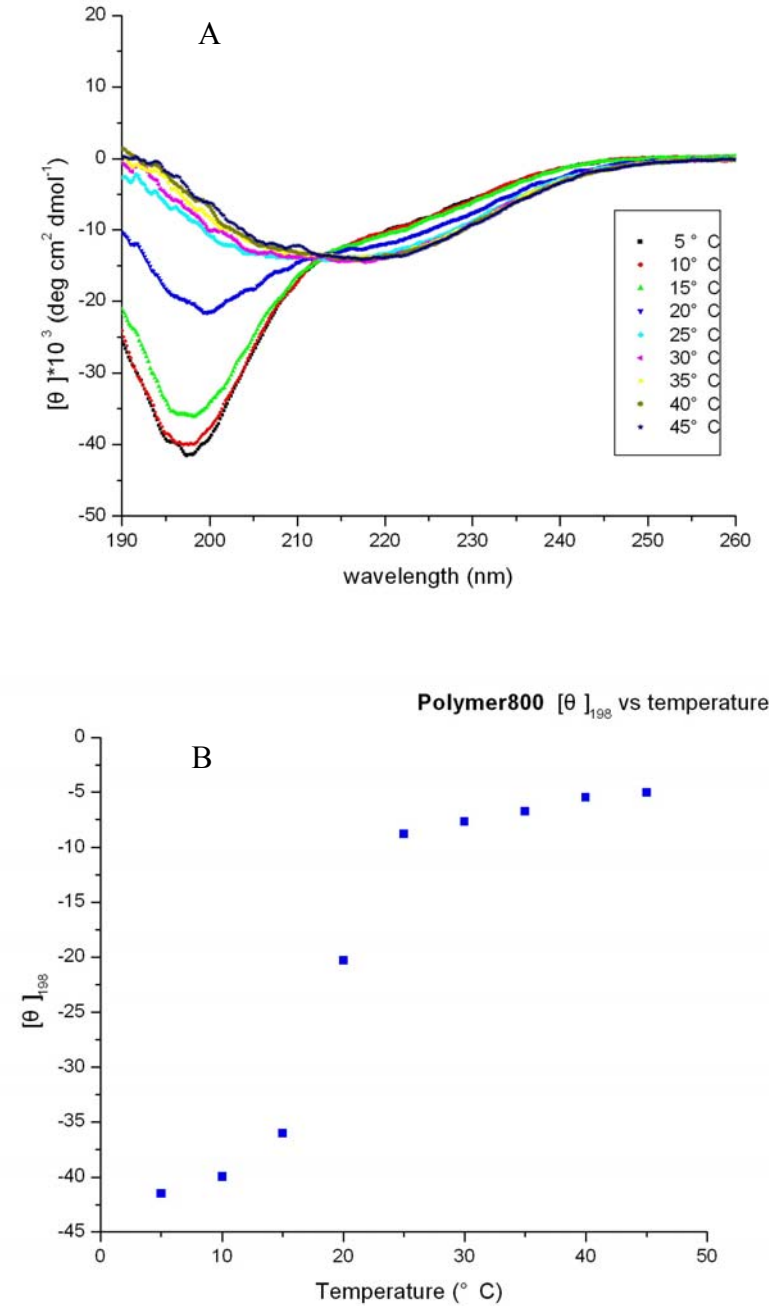


Figure 29 .CD results for **polymer-800**. (A) CD spectra depicting the thermal dependent conformational transition for **polymer-800** (B) Thermal transition profiles monitoring the disappearance of the random coil conformation $[\theta]_{198}$ in the spectra of **polymer-800**

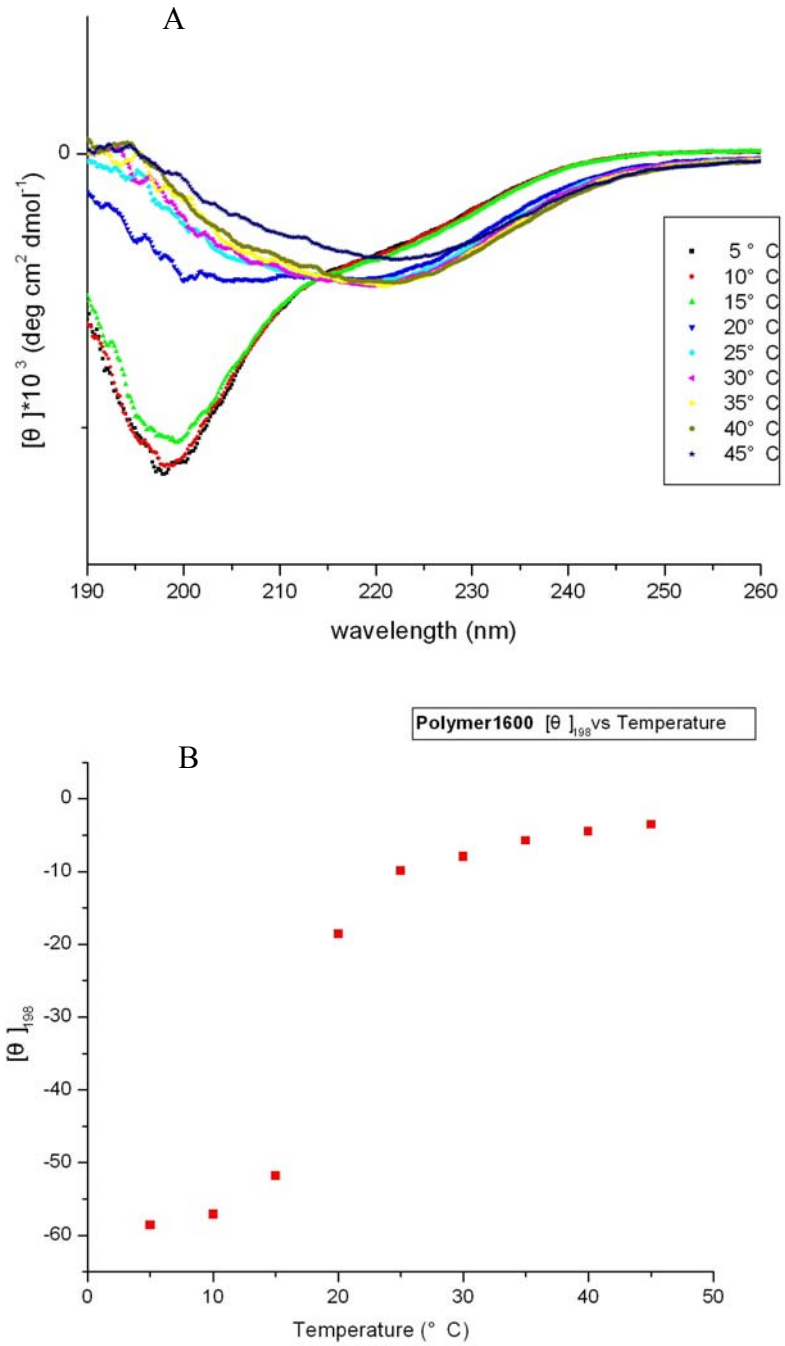


Figure 30. CD results for **polymer-1600.** (A) CD spectra depicting the thermal dependent conformational transition for **polymer-1600** (B) Thermal transition profiles monitoring the disappearance of the random coil conformation $[\theta]_{198}$ in the spectra of **polymer-1600**

Calorimetric Measurement of the copolymer phase transition

Thermodynamic profiles of the two copolymers during self-assembly from aqueous solutions can be easily determined by differential scanning calorimetry (DSC)[48, 49]. DSC measurements of diluted solutions of **polymer-800** and **polymer-1600** demonstrated that both the two proteins undergo a sharp endothermic transition, at 20.1 ℃ and 19.8 ℃, respectively. The transition temperatures are in good agreement with the results determined by CD (20 ℃ and 19.8 ℃). The transitions are reversible upon cooling and rescan, reflected by the overlapping of the rescanned curves.

The DSC endotherms for the two proteins were fit to a standard two-state transition, with a molar enthalpy of ΔH=436 kcal/mol for **polymer-800** and ΔH=686.2 kcal/mol for **polymer-1600**. These results suggested an entropy driven (positive ΔS) process that is in agreement with hydrophobic or other salvation related interactions [14, 15]. Both the two proteins display a measurable decrease in apparent heat capacity above the transition temperature, evident in the lower DSC baseline. The decrease of apparent heat capacity ΔC_p was determined to be -39.4 kcal mol^{-1} K^{-1} for **polymer-800** and -61.7 kcal mol^{-1} K^{-1} for **polymer-1600**. The heat capacity decrement of the two proteins above the transition temperature is characteristic of the phase conversion of the proteins from an aqueous environment to non-polar condensed aggregates[14, 15].

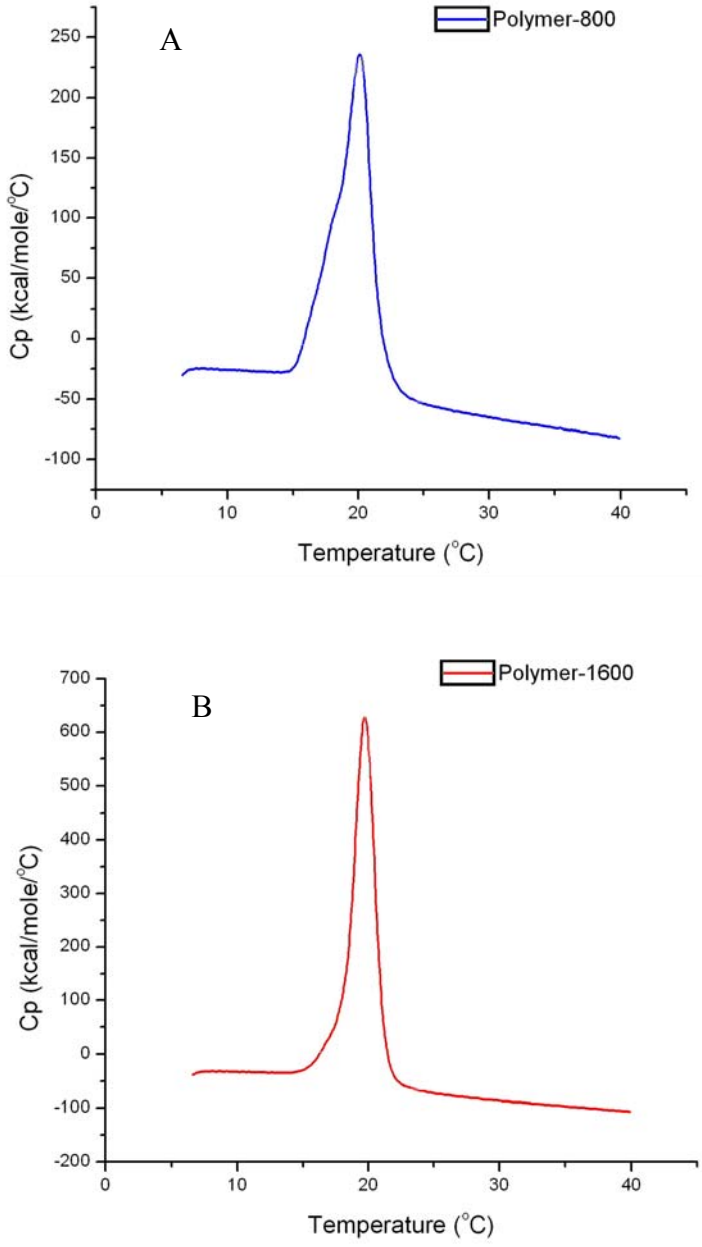


Figure 31. DSC results for **polymer-800** and **polymer-1600.** (A) DSC measurement of **polymer-800**, thermal transition occurs at a transition point of 20.1 ℃. (B) DSC measurement of **polymer-1600**, thermal transition occurs at a transition point of 19.8 ℃.

Degradation by plasmin in solution

To test the degradation of **polymer-800** and **polymer-1600** by plasmin in solution, two tri-block proteins (1 mg/mL) were mixed with plasmin (0.5 U/mL) in HBS (pH 7.4) /NaN3 (0.2 mg/mL). NaN_3 was added to prevent bacteria growth. Reaction mixture was incubated at 18 °C (**polymer-1600**) and 37 °C (**polymer-800**), respectively. An aliquot of 10 μL of the reaction mixture was transferred into a clean vial at 0, 6, 25, 48, 72 and 90 hours. The aliquots were then analyzed by western blot. The results (Fig. 31) indicated that both the two proteins were degraded well by plasmin and the digestions were nearly complete within 6 hours. The degradation products were detected on the gel as single bands at around 75 kDa (theoretically 58kDa), corresponding to the plastic endblocks.

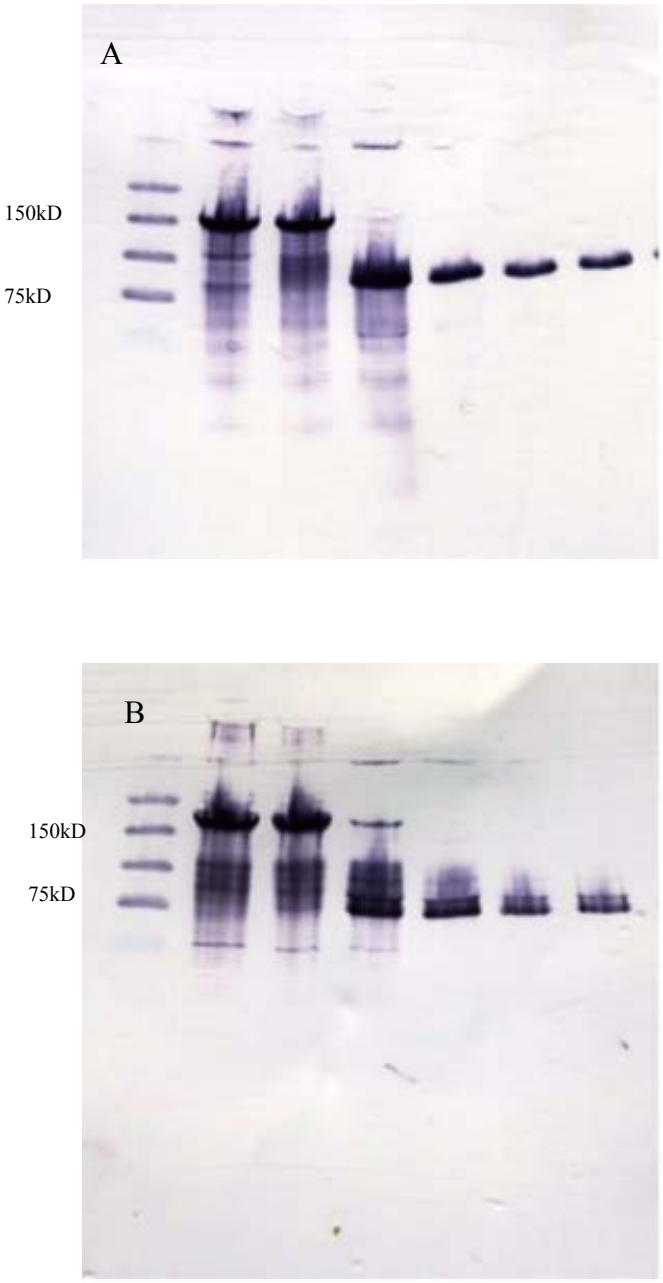


Figure 32. Western blot for the plasmin degradation of (A) **polymer-800** and (B) **polymer-1600** at different time length. Lane 1: Protein standard (Biorad dual color); Lane 2: Polymer without plasmin; Lane 3: Polymer with plasmin at 0 hour; Lane 4: Digestion after 6 hours; Lane 5: Digestion after 24 hours; Lane 6: Digestion after 48 hours; Lane 7: Digestion after 72 hours

Chapter 5. DFT calculation of the conformational preference of thioproline derivatives

Background

Stereoelectronic effect has been widely accepted to affect the conformation of organic molecules and has recently been intensively studied on its influence on the protein conformational preference and stability. This effect was also considered as an effective means to confine the local conformation for the polypeptides and to modulate the thermodynamic properties and higher order structures of the proteins [15, 24, 34, 41, 47-51].

Proline is the element intensively involved in the stereoelectronic effect in which the pucker of the pyrrolidine ring can be influenced markedly by the electronegative substituents, (X=N, O, F). The influence depends on the configuration and electron-withdrawing property of the substituent, that is why the effect is named as stereoelectronic effect. In addition, n→π* interaction has been shown to contribute to the population of a particular amide bond configuration [52-54].

Collagen mimetic polypeptides and elastin mimetic polypeptides, both of which are rich in proline residues can be remarkablely influenced by the stereoelectronic effect in terms of protein stability, conformation and thermodynamic properties. Besides structural determination studies, such as CD, x-ray crystallography and multidimensional NMR that are used to investigate the global structure of the proteins, small molecules and segments are also frequently used as models to study the influence of stereoelectronic effect on the local structural change, especially on the pyrrolidine rings [41, 55-58]. Both experimental (x-ray crystallography, NMR) and theoretical methods have played important roles in uncovering the structural features of the substituted pyrrolidine rings, which facilitated understanding protein stability and self-assembly.

Till now, proline derivatives, N-acetyl-A-methyl esters have served as the research model, in which A can be 2S-proline[53], (2S,3S)-3-fluoroproline, (2S,3R)-3-fluoroproline[41], (2S,4S)-4-fluoroproline and (2S,4R)-4-fluoroproline[15]. Crystal structures have been obtained

and analyzed, multi-dimensional NMR experiments were carried out and computational studies were conducted to uncover the structures. The chemical structures and the analysis results for the fluorinated derivatives are described below.

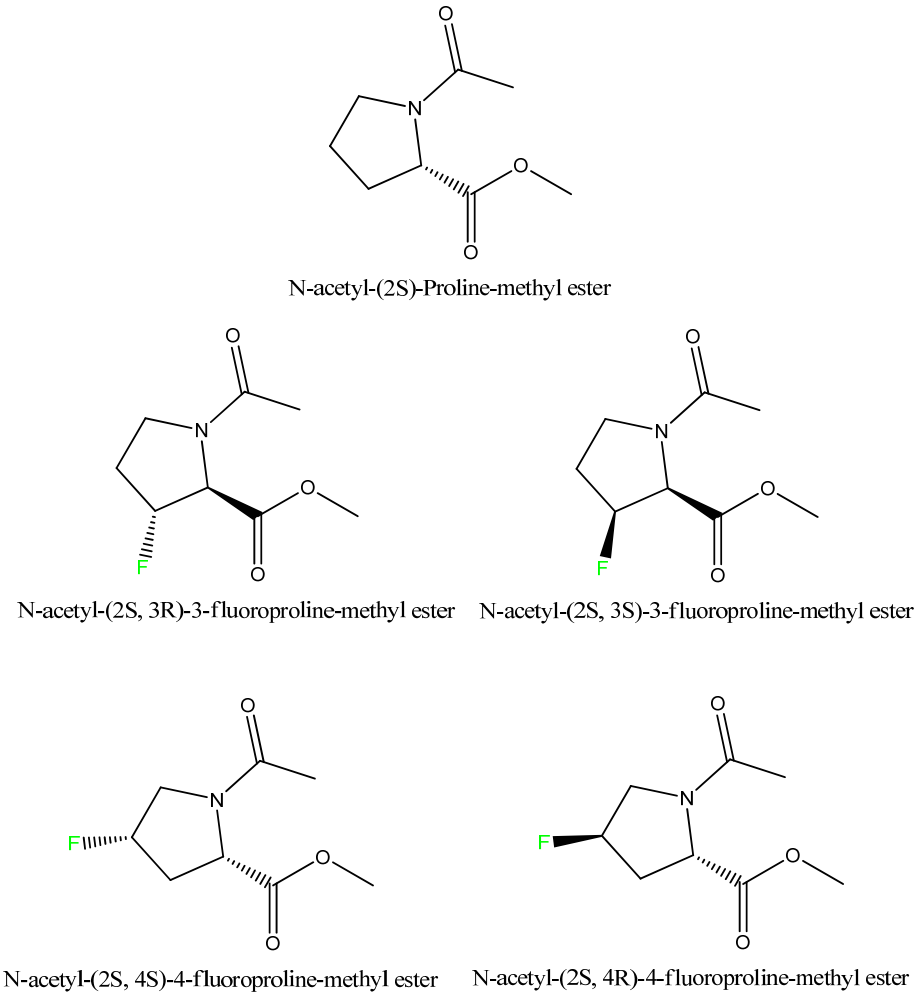


Figure 33. Structures of proline derivatives used as research models

Ac-FPro-OMe	$K_{trans/cis}$	$E_{exo/endo}$ (kcal/mole)	(°)	(°)
(3*R*) (**1**)	8.9	-0.48	-56.4	151.5
(3*S*) (**2**)	4.3	2.13	-71.2	158.4
(4*S*) (**3**)	2.5	0.61	-76.4	172.0
(4*R*) (**4**)	6.7	-0.85	-55.1	140.5

Table 3. Thermodynamic data and main-chain torsion angles for the *N*-acetyl-fluoroproline methyl ester derivatives.[41]

Is has been shown that 3R and 4R derivatives have a strong preference of C$^{\gamma}$-*exo* pucker for the pyrrolidine ring, while the corresponding diastereomers, 3S and 4S derivatives favor the *endo* pucker [15, 40, 41]. However, unsubstituted proline derivative only exhibits a slight preference for C$^{\gamma}$-*endo* pucker, with two conformations nearly equal at room temperature[24]. How can a conformational preference be elicited upon introduction of a fluorine atom? Later studies suggested that the gauche effect may be the answer. In an X-C-C-Y structure, in which X and Y are two electron-withdrawing substituents, the gauche conformers are preferred over the corresponding anti conformers as a result of increased hyperconjugative delocalization [59, 60]. The gauche relationship between the amide and fluorine substituents forces the selection of each ring pucker as the dominant conformation. The crystal structures and the Newman projections for the 3R and 3S derivatives are shown below.

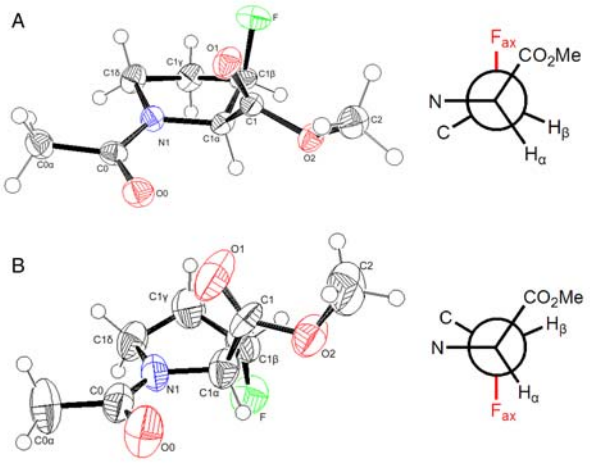
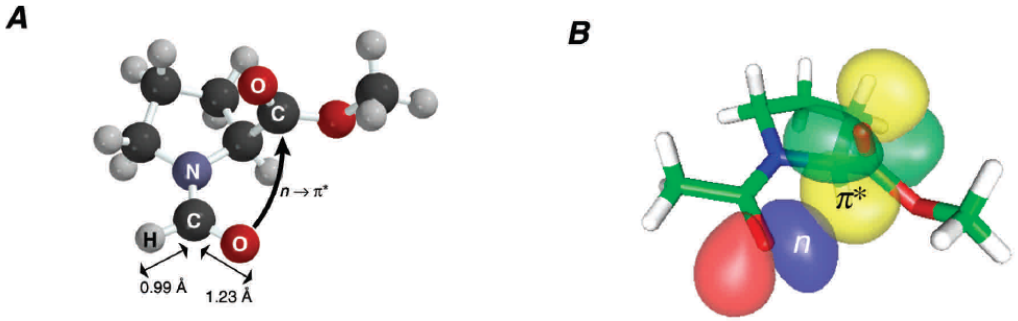


Figure 34. Crystallographically determined structures of *N*-acetyl-(2*S*,3*R*)-3-fluoroproline methyl ester (**1**), **A**, and *N*-acetyl-(2*S*,3*S*)-3-fluoroproline methyl ester (**2**), **B**, in conjunction with the Newman projections along the C -C bond vector depicting the gauche relation between the amide and fluorine substituents [41]

Current knowledge has revealed that all the above structures favor the *trans* prolyl amide bond, although with different values for the *trans/cis* ratio. This preference has been attributed to an n→π* interaction, in which the amide oxygen approaches the carbonyl carbon in the methyl ester in a manner similar to the Burgi-Dunitz trajectory [24, 38, 48, 61]. The n→π* interaction is usually defined as the one in which the O-C distance is less than 3.2 Å and the angle O--C=O is between 99° to 119°, within 10° of the conventional Burgi-Dunitz trajectory. This interaction was used to explain the conformational preference of fluorinated proline derivatives, as well as the structure preference of other species such as azido proline derivatives[62].



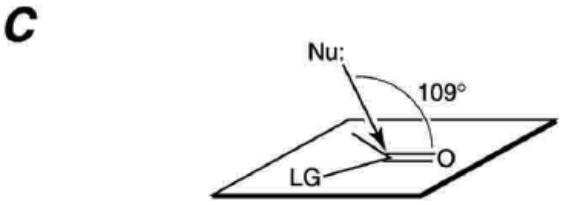
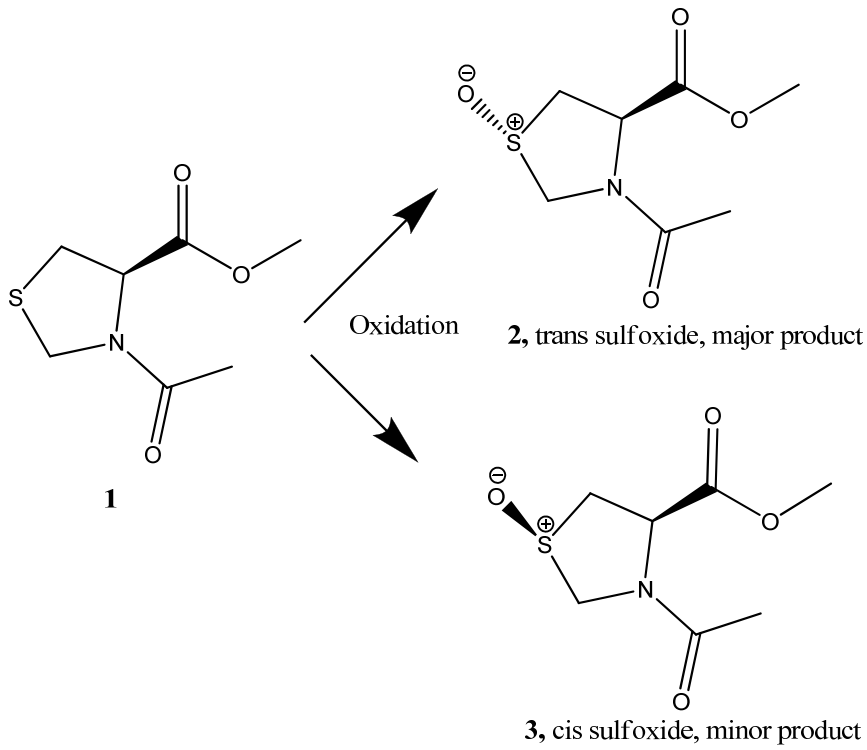


Figure 35. Representation of the n→π* interaction (A)Structure of the *trans* isomer of *N*-formyl-L-proline methyl ester (B) Depiction of the n and π* orbitals in the *trans* isomer of *N*-formyl-L-proline methyl ester (C) Bürgi–Dunitz trajectory for the attack of a nucleophile on a carbonyl group to displace a leaving group [53]

Recently, a novel proline analogue, thioproline has been successfully incorporated into the elastin-mimetic polypeptide, using the medium switch protocol described in chapter 2[16] . Although no stereochemistry exists in the sulfur position, the oxidation of this analogue will yield a mixture of two diastereomers-the *trans* and *cis* sulfoxide derivatives. The resulting pair of diastereomers is considered as a good probe for the stereoelectronic effect within the elastin-mimetic polypeptide in a manner similar to fluorinated proline derivatives, which will be potentially used to study the local and global conformations of the protein. The oxidation of N-acetyl-thioproline-methyl ester has been finished by Conticello group to afford two sulfoxide products, with *trans* and *cis* relationship between the sulfoxide group and the methyl ester, respectively[63]. *Trans*-sulfoxide derivative was produced as the major product and *cis* diastereomer was the minor product. Both the two structures were crystallized and studied by x-ray crystallography and NMR.

The crystal structure for the *trans* isomer revealed that the five-membered ring displays a C^{γ}-*exo* pucker and a *trans* peptide bond, while the C^{γ}-*endo* pucker and the *cis* amide bond are characteristic for the *cis* product. It is interesting to observe this strong correlation of the ring pucker and amide bond configuration, which may result from the stereoelectronic influence of the endocyclic sulfoxide group. In this chapter, DFT based calculations were performed to optimize both two oxidation products and the original thioproline derivative. The energies of

different conformations were compared and the structural preferences were rationalized in term of the gauche effect, steric repulsion and the n→π* interaction.



Scheme 2. Oxidation of thioproline derivative **1** to afford two diastereomers of thiazolidine sulfoxide derivative **2** and **3**

Calculation procedure:

Three model compounds **1, 2** and **3** were constructed using Maestro 8.0 (Schrödinger, Inc). Each compound constructed was subjected to a conformational search by Macromodel 9.1 (Schrödinger, Inc) using the following parameters: OPLS_2005 force field, Monte Carlo method, gas phase and 1000 steps. Four conformations were generated for each compound, in which they differed in C^{γ} ring pucker orientation (*exo/endo*) and the amide bond configuration (*trans / cis*). Structures with lowest energies for each conformation were used as the starting structures and the corresponding Cartesian coordinates were input into Gaussian 03 program for density functional theory (DFT) calculations. The geometries of all conformations were first optimized using Becke3LYP/6-31G* method. More accurate energies of these optimized geometries were obtained with optimization using a larger basis set in Becke3LYP/6-311+G* calculations. Relative energies for all the conformations of these three compounds were calculated based on units of kcal/mol and the energies at different levels were compared.

Results and discussion

*Oxidation product **2** and **3** (N-acetyl-thiazolidine sulfoxide-methyl esters)*

The calculation results for all conformations of **2** and **3**, using different methods were listed in the following table. The total energy, relative energy as well as the data sets for n-π* stabilization are included in the table. In red are the structures with lowest energy from the calculations. The conformations for the two crystal structures were labeled using underlines

	Total E (a.u.) 6-31 G* [a]	Relative energy (kcal/mol)	6-311+G* [b]	Relative energy(kcal/mol)	n- π * stabilization [c]
2 (*Trans* sulfoxide)					[$O_{(C=O)}$-C distance and $O_{(C=O)}$-C=O angle]
trans-exo	-1027.16742845	0.000000e+000	-1027.37218748	0.000000e+000	**3.017, 102.1**
trans-endo	-1027.16606294	0.856870	-1027.36939747	1.750756	2.938, 97.5
cis-exo	-1027.16544823	1.242606	-1027.36987266	1.452570	4.389, 84.9
cis-endo [d]	**-1027.16282554**	**2.888367**			4.304, 72.8 **(6-31 G*)**
Crystal structure *(trans-exo)*					3.022, 93.0
3 (*Cis* sulfoxide)					[$O_{(S=O)}$-C distance and $O_{(S=O)}$-C=O angle]
trans-exo	-1027.16703573	0.000000e+000	-1027.37091111	0.000000e+000	4.918, 97.3
trans-endo[e]	-1027.16697907	0.035555	-1027.37063528	0.173086	3.680, 97.19
cis-exo	-1027.16454449	1.563276	-1027.36813537	1.741802	4.878, 103.9
cis-endo	-1027.16651494	0.326800	-1027.36922393	1.058721	**3.085, 104.6**
Crystal structure *(cis-endo)*					2.910, 97.8

Table 4. Calculated values for energy difference of all confomations of **2** and **3**

[a] Beck3LYP/6-31G*// Beck3LYP/6-31G*

[b] Beck3LYP/6-311+G*// Beck3LYP/6-31G*

[c] The n-π* stabilization effect is examined by $O_{(C=O)}$-C distance and $O_{(C=O)}$-C=O angle for compound **2** and $O_{(S=O)}$-C distance and $O_{(S=O)}$-C=O angle for compound **3.** All measurements are from calculation result using Beck3LYP/6-311+G(d, p)//PCM except mentioned otherwise.

[d] Calculation with method Beck3LYP/6-31G*// Beck3LYP/6-31G* produced one correct C^{γ}-*endo* pucker while all the puckers have been reversed to *exo* orientation in Beck3LYP/6-311+G* level calculation. The data set for n-π* stabilization comes from the result at Beck3LYP/6-31G* level calculation.

[e] The pucker has shifted to C^{α} position, instead of C^{γ} position

Single crystal x-ray crystallography has been used to determine the structure of both **2** and **3**. As the major product, **2** displays a C^{γ}-*exo* pucker and a *trans* amide bond, which is in line with the calculation results at different levels. The figure below shows the overlap of the optimized structure of **2** with *trans-exo* conformation and the crystal structure of **2.** We can notice that the two five-membered rings well overlap to each other; the only difference comes from the methyl ester branch.

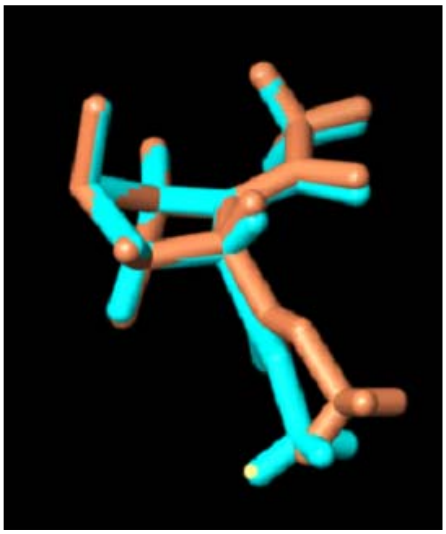


Figure 36. Supercomposition of the crystal structure of **2** (green) and the optimized structure of **2** with *trans-exo* conformation (orange)

The stability of *trans* amide bond of **2** can be attributed to the n→ π* interaction. The oxygen of the acetyl group is at the angle of 102.1° toward the carbonyl group of the methyl ester, comparable to 93.0°, an angle observed in the crystal structure. Meanwhile, the distance between the oxygen of the acetyl and the carbonyl carbon of the *exo-trans* structure is 3.017 Å, which is in accord with what the crystal structure displays, 3.022 Å. In a word, both the O--C=O angle and the O--C distance for **2** with *trans* amide bond are typical for the Burgi-Dunitz trajectory in which the nucleophile approaches a carbonyl group. However, this n→ π* interaction is only observed in the *trans*-conformations. In compound **2** with *cis* amide bond, the O-C=O distance is usually too large (> 4 Å) and the angle usually deviate far away from the reasonable range, so no contribution can be made to the overall molecular stability.

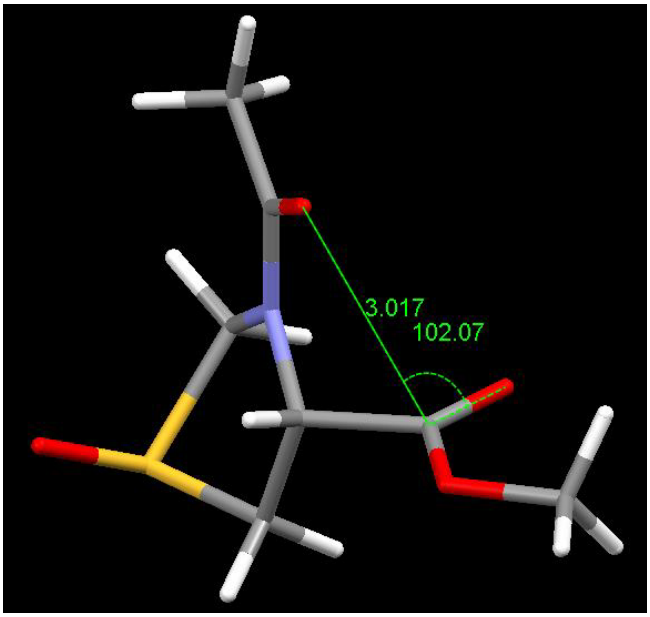


Figure 37. n→π* interaction stabilizes the *trans* amide bond of **2**

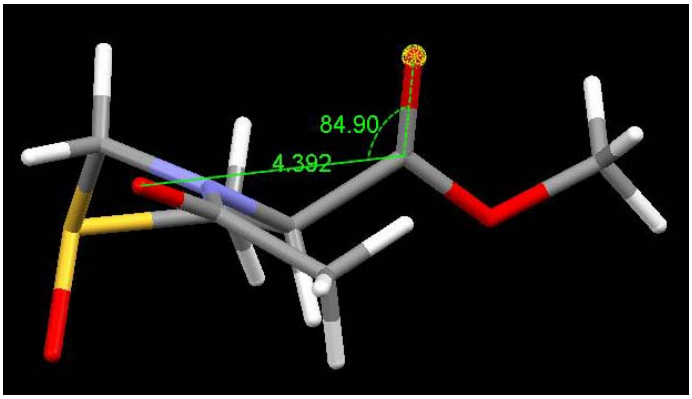


Figure 38. No n→π* interaction is observed in **2** with *cis* amide bond

Why C$^{\gamma}$-*exo* pucker is preferred than *endo* pucker in the major product **2**? Gauche effect is not the answer since the dihedral angle within N-C-C-S of **2** with *exo* pucker is only 24.27º, far away from the ideal value 60 º for a gauche orientation. Later investigation revealed that the steric repulsion between the sulfur and acetyl oxygen may eliminate the conformer with *endo* pucker in which O-S distance can be 3.751 Å. Although the distance does not exceed the sum of the van der waals radii of sulfur and oxygen (3.37 Å) [64], it is short enough to result in a steric repulsion between the nonbonding lone pairs from each atom. However, much longer distance (4.521 Å) has been observed in compound **2** with *exo* pucker, which makes this conformation energetically favored.

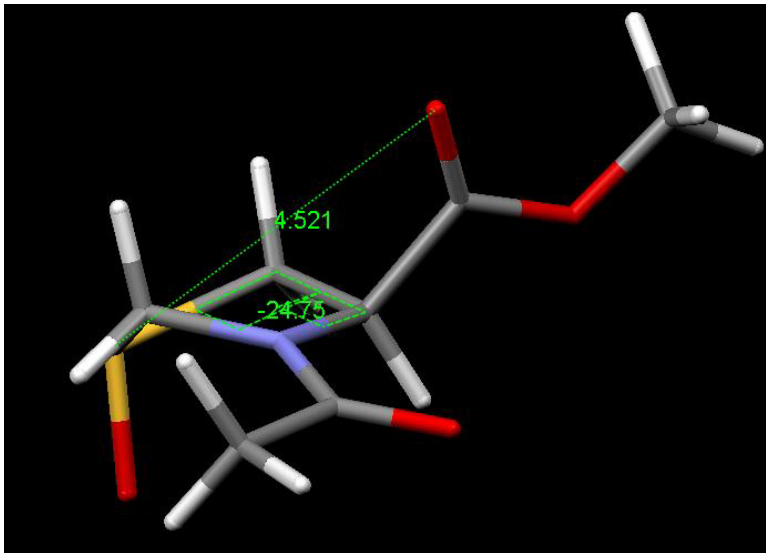


Figure 39. Compound **2** with favored C$^{\gamma}$-*exo* pucker.

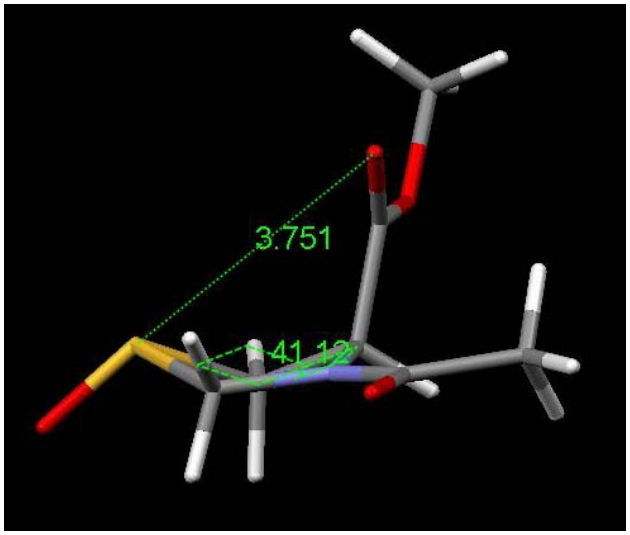


Figure 40. Compound **2** with disfavored C$^{\gamma}$-*endo* pucker *(2-cis-endo)*. Steric repulsion can be observed between the sulfur and the oxygen atom, whose distance is 3.751 Å

The calculation results for the minor product **3** seem to be more complicated. Although the calculation at both levels favored the C$^{\gamma}$-*exo* conformation as the lowest energy conformation, it was not in accord with the x-ray crystallography result, which preferred C$^{\gamma}$-*endo* conformation. Firstly, gauche effect can be used to explain why *endo* pucker is favored experimentally in this case. The compound **3** with C$^{\gamma}$-*exo* pucker and *cis* amide bond displays a N-C-C-S dihedral angle of -17.21°, which is far away from the reasonable range for gauche orientation. The C$^{\gamma}$-*endo* pucker containing structure, however, has a dihedral angle of 43.36°,

which is within the reasonable range for gauche effect and is close to what is observed in the crystal structure of **3** (42.84°). Secondly, C^{γ}-*exo* conformation can be well stabilized by the n→π* interaction within the molecule. Specifically, two possible n→π* interactions may play important roles in this stabilization, one from oxygen of the acetyl to carbonyl of the methyl ester, the other from thio oxygen to the carbonyl of methyl ester. In the crystal structure of **3**, $O_{S=O}$—C distance is 2.910 Å and the $O_{S=O}$--C=O angle is 99°, both of which are typical values for the n→π* interaction. A similar set of value can be found in **3**-*cis-endo* conformation, which are 3.085 Å and 104.6° and close to what are observed for the crystal structure. However, for **3** with C^{γ}-*exo* pucker, the distance is so large (~5 Å) that this stabilization can not be achieved. Although the other n→π* interaction has the opposite pucker preference, the $O_{S=O}$--C=O interaction may dominate the overall preference and as a result, **3** with *cis-endo* conformation was shown to be the favored structure

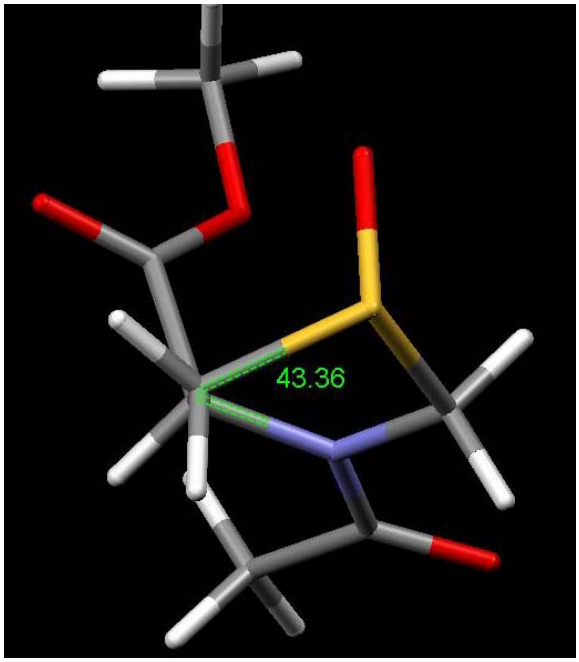


Figure 41. Gauche effect within the favored conformation of **3** (*cis-endo)*. The N-C-C-S dihedral angle is 43.36°

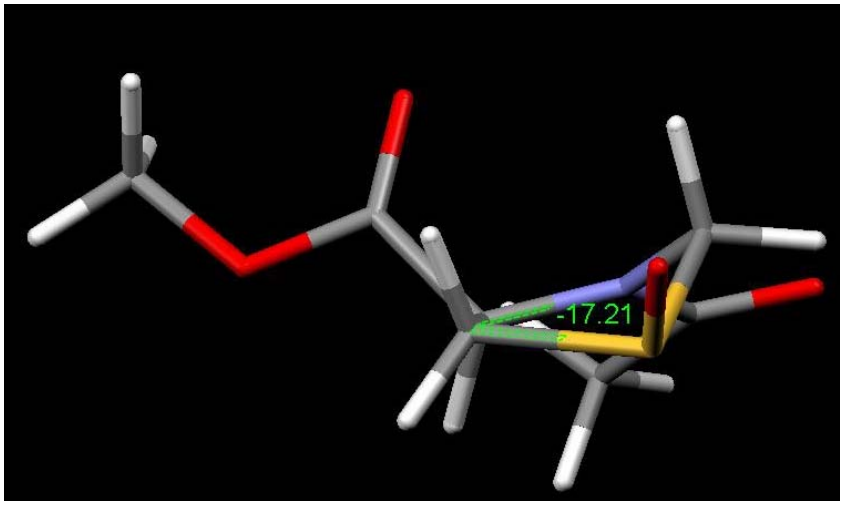


Figure 42. No gauche effect in **3** with *cis-exo* conformation. The N-C-C-S dihedral angle is -17.21°

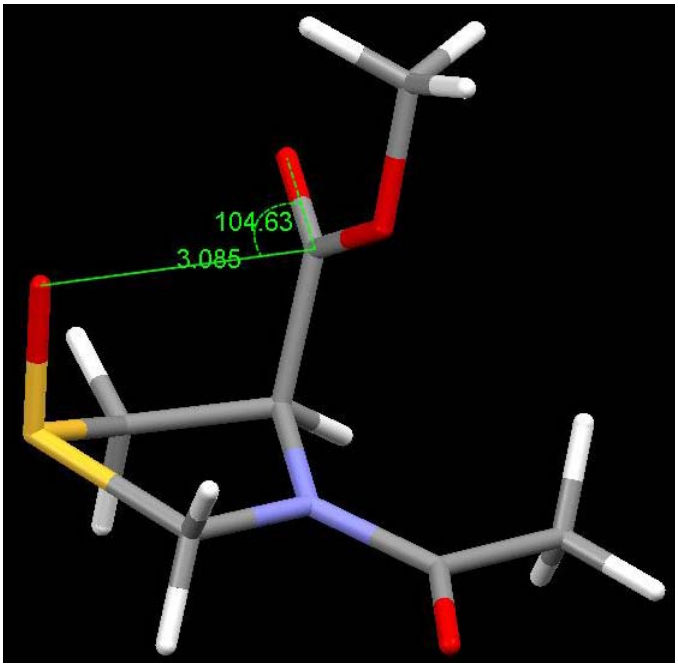


Figure 43. $O_{(S=O)}$--C=O n→π* interaction in favored conformation of **3**, *cis-endo* ($O_{(S=O)}$—C distance 3.085 Å; $O_{(S=O)}$--C=O angle 104.63 º)

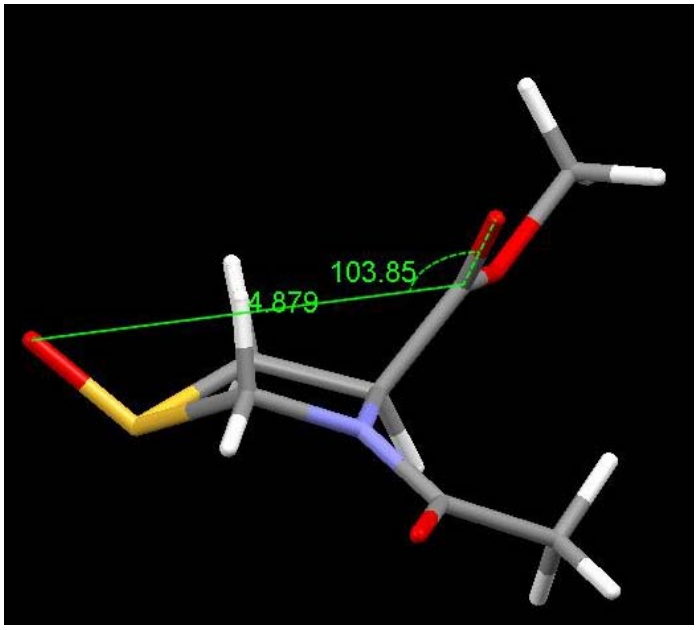


Figure 44. The $O_{S=O}$-$C_{C=O}$ distance within **3** with *cis-exo* conformation is too large for the n→π* interaction (4.879 Å)

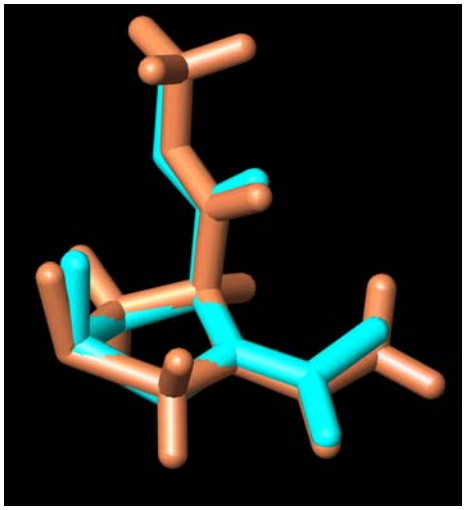


Figure 45. Supercomposition of crystal structure (green) and the optimized **3** with *cis-endo* conformation (orange).

*Compound **1** (N-acetyl-thioproline-methyl ester)*

The calculation results for all conformations of **1**, using different methods were listed in the following table. The total energy, relative energy as well as the data sets for n-π* stabilization are included in the table. In red are the structures with lowest energy from the calculations.

	6-31G*,total energy (a.u.)	Relative energy (kcal/mol)	6-311+G*, total energy (a.u.)	Relative energy (kcal/mol)	n- π * stabilization [O-C distance and O-C=O angle]
Endo-trans	-951.99474805	0.000000e+000	-952.17164131	0.000000e+000	3.453, 92.2
Endo-cis	-951.99351709	0.772438	-952.17026440	0.864026	4.432, 168.9
Exo-trans	-951.99394004	0.507030	-952.17158794	0.033491	**2.917, 99.8**
Exo-cis	-951.99170114	1.911962	-952.16882935	1.764529	4.400, 140.8

Table 5. Calculated values of energy difference between different conformations of N-acetyl-thioproline-methyl ester

From the table, we are able to find out that at becke3LYP/6-31G* level calculation, structure with endo-trans conformation was determined to be have the lowest energy. However, when the calculation was conducted at becke3LYP/6-311+G* level, these two conformations are nearly equally populated, with energy gap of only 0.033 kcal/mol. *Trans* amide bond is favored since the structures (*endo-trans* & *exo-trans*) are stabilized by the n→π* interaction from the acetyl oxygen to the carbonyl group.

According to the table, compound **1** with *exo-trans* conformation seems to more ideal measurement of O-C distance and O-C=O angle comparable to that in a Burgi-Dunitz trajector

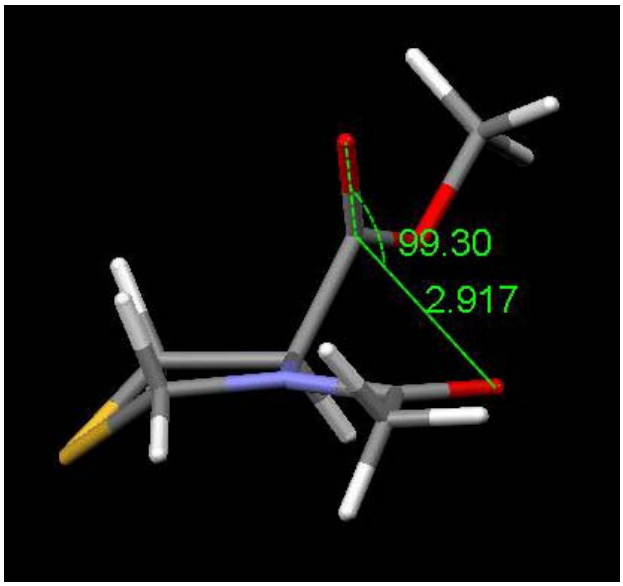


Figure 46. Compound **1** with *exo-trans* conformation. O-$C_{C=O}$ distance is 2.917 Åand O-C=O angle is 99.30º, both of which are within the reasonable range of the n-π* stabilization

Chapter 6. Conclusion

Employing genetic engineering and medium switch protocol, two fluorinated amino acids were successfully incorporated into the elastin-mimetic polypeptide sequence. **Elastin 2** containing 4, 4, 4-trifluorovaline in all isoleucine codons was expressed in isoleucine auxotrophic *E.coli.* strain CAG 18599 and the expression was confirmed by a single band at 37 kD in western blotting. The protein was purified using metal affinity chromatography with a yield of 25 mg/500ml culture and was checked by SDS PAGE electrophoresis. The peak in MALDI-TOF MS spectrum indicated a mass of 36500.22 D, in accord with the theoretical mass of 36534 D. The small percentage error of 0.09% guaranteed the complete incorporation of 4, 4, 4-trifluorovaline. In addition, the strong peak in ^{19}F NMR at -68 ppm confirmed the identity of fluorinated **elastin 2**. Structural study by circular dichroism (CD) demonstrated a transition from random coil to type II β turn when the temperature is raised up, a self-assembly process similar to native elastin. Using the same strategy, **elastin 3** equipped with 3-fluorovaline was produced with a yield of 15 mg/500 ml culture. The expression and the purification were monitored by western blotting and SDS PAGE electrophoresis and the single bands above 35 kD confirmed the identity of **elastin 3.** Two peaks were observed in MALDI TOF MS spectrum with the minor peak corresponding to the desired elastin 3, while the major peak indicating the other species with a lower mass, which may come from the degradation or deletion of **elastin 3.** CD spectrum of **elastin 3** showed a thermal transition from random coil to type II β turn as the temperature increase.

In Chapter 3, [Val-Pro-Gly-Ala-Gly] was designed as the elastin-mimetic polypeptide sequence suitable for the study of 4, 4-difluoroproline incorporation, in which the resulting elastin is expected to undergo the inverse temperature transition within an experimental measurable range. PCR was used to anneal the two strands DNA to afford the monomer $(VPGAG)_5$ and recursive directional ligation was employed to construct plasmids with varied inserts from 500 bp to 900 bp. The plasmid pYT600 with insert of 600 bp long was digested by two restrictions enzyme pairs and the resulting compatible parts were ligated together to

generate the desired plasmid pYT1200, with an insert of 1200 long. In the same manner, pYT1300 and pYT1400, which contain 1300 bp insert and 1400 bp insert, were constructed. The identities of the genes were confirmed by gel electrophoresis and DNA sequencing analysis.

In chapter 4, the identities of **polymer-800** and **polymer-1600** were confirmed by proton NMR spectroscopic studies, in which protons for all major residues were assigned with the help of two-dimensional NMR. Similar structural transition was detected by circular dichroism (CD) for both the two polypeptides in which an ordered structure with negative shoulder at 220 nm was developed to replace the random coil signature when the temperature was raise up. The transition point was determined by mathematically fitting for the mean residue ellipticity at 198 nm versus temperature, which indicated a transition temperature of 20.0 ℃ for **polymer-800** and 19.8 ℃ for **polymer-1600**. The endothermic transition was detected by DSC, showing sharp peaks at the position corresponding to thermal transition. The transition point was determined to be 20.1 ℃ for **polymer-800** and 19.8 ℃ for **polymer-1600**, respectively, This result is very close to what is generated from CD. For both the two triblock proteins, plasmin digestion tests were conducted and it was shown that the degradations were complete within 6 hours at appropriate temperatures, producing residues corresponding to plastic endblocks.

In chapter 5, DFT theory was utilized to optimize the two oxidation products of N-acetyl-thioproline-methyl ester: **2** and **3.** The computation results in gas phase at both 6-31G* and 6-311+G* indicated that **2** preferred C^{γ}-*exo* pucker and *trans* amide bond, which is in line with the crystal structure. Detailed investigation revealed that the although no *gauche effect* was between the sulfur and the amide nitrogen, for neither *exo* pucker conformations, nor *endo* pucker conformations, the small distance between sulfur and acetyl oxygen within the *endo* pucker containing conformation destabilizes this structure, thus making *exo* pucker the preferred pucker orientation. The *trans* preference within amide bond was selected as a result of the $n \rightarrow \pi^*$ interaction between the amide oxygen and the carbonyl group of the methyl ester.

However, the calculation results for the **3** were not in accord with the corresponding crystal structure, in which the C^{γ}-*endo* pucker and *cis* amide bond were favored. Detailed structural analysis showed that both the *gauche effect* as well as the n→π* interaction between the sulfoxide oxygen and the carbonyl group of the methyl ester (unique only in **3**) favored the C^{γ}-*endo* pucker. Moreover, DFT calculation for N-acetyl-thioproline-methyl ester (**1**) suggested a preference for *trans* amide bond. This preference can be explained by the n→π* interaction from the acetyl oxygen to the carbonyl group. At the 6-311+G* level, both *endo* and *exo* C^{γ} pucker are nearlly equally populated at room temperature.

In summary, genetic engineering and theoretical calculation have been employed to study the potential influence of non-canonical amino acids incorporation on the elastin-mimetic polypeptide and this study has provided insights about the local and global conformations of the elastin-mimetic polypeptide.

Reference

[1] J. C. van Hest, D. A. Tirrell, *Chem Commun (Camb)* **2001**, 1897.

[2] X. L. Yao, V. P. Conticello, M. Hong, *Magn Reson Chem* **2004**, *42*, 267.

[3] M. Hong, D. Isailovic, R. A. McMillan, V. P. Conticello, *Biopolymers* **2003**, *70*, 158.

[4] E. R. Wright, V. P. Conticello, *Adv Drug Deliv Rev* **2002**, *54*, 1057.

[5] D. E. Meyer, A. Chilkoti, *Nat Biotechnol* **1999**, *17*, 1112.

[6] K. Trabbic-Carlson, D. E. Meyer, L. Liu, R. Piervincenzi, N. Nath, T. LaBean, A. Chilkoti, *Protein Eng Des Sel* **2004**, *17*, 57.

[7] M. R. Banki, L. Feng, D. W. Wood, *Nat Methods* **2005**, *2*, 659.

[8] D. E. Meyer, A. Chilkoti, *Biomacromolecules* **2004**, *5*, 846.

[9] D. W. Urry, *J Protein Chem* **1988**, *7*, 1.

[10] W. J. Cook, T. L. Trapane, K. U. Prasad, *Int J Pept Protein Res* **1985**, *25*, 481.

[11] D. W. Urry, *Angew Chem Int Ed Engl* **1993**, *32*, 819.

[12] E. G. Hutchinson, J. M. Thornton, *Protein Sci* **1994**, *3*, 2207.

[13] T. A. Lee, A. Cooper, R. P. Apkarian, V. P. Conticello, *Advanced Materials* **2000**, *12*, 1105.

[14] E. R. Wright, R. A. McMillan, A. Cooper, R. P. Apkarian, V. P. Conticello, *Advanced functional materials* **2002**, *12*, 149.

[15] W. Kim, R. A. McMillan, J. P. Snyder, V. P. Conticello, *J Am Chem Soc* **2005**, *127*, 18121.

[16] W. Kim, A. George, M. Evans, V. P. Conticello, *Chembiochem* **2004**, *5*, 928.

[17] A. J. Link, M. L. Mock, D. A. Tirrell, *Curr Opin Biotechnol* **2003**, *14*, 603.

[18] L. Wang, P. G. Schultz, *Chem Commun (Camb)* **2002**, 1.

[19] N. G. Marsh, *Chemistry & Biology* **2000**, *7*, R153.

[20] P. Wang, Y. Tang, D. A. Tirrell, *J Am Chem Soc* **2003**, *125*, 6900.

[21] N. C. Yoder, K. Kumar, *Chem Soc Rev* **2002**, *31*, 335.

[22] Y. Tang, G. Ghirlanda, N. Vaidehi, J. Kua, D. T. Mainz, I. W. Goddard, W. F. DeGrado, D. A. Tirrell, *Biochemistry* **2001**, *40*, 2790.

[23] B. Bilgicer, A. Fichera, K. Kumar, *J Am Chem Soc* **2001**, *123*, 4393.

[24] M. L. DeRider, S. J. Wilkens, M. J. Waddell, L. E. Bretscher, F. Weinhold, R. T. Raines, J. L. Markley, *J Am Chem Soc* **2002**, *124*, 2497.

[25] C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber, L. Moroder, *Angew Chem Int Ed Engl* **2001**, *40*, 923.

[26] Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado, D. A. Tirrell, *Angew Chem Int Ed Engl* **2001**, *40*, 1494.

[27] Y. Tang, D. A. Tirrell, *J Am Chem Soc* **2001**, *123*, 11089.

[28] H. Duewel, E. Daub, V. Robinson, J. F. Honek, *Biochemistry* **1997**, *36*, 3404.

[29] P. Wang, A. Fichera, K. Kumar, D. A. Tirrell, *Angew Chem Int Ed Engl* **2004**, *43*, 3664.

[30] S. Son, I. C. Tanrikulu, D. A. Tirrell, *Chembiochem* **2006**, *7*, 1251.

[31] G. J. Cotton, T. W. Muir, *Chem Biol* **1999**, *6*, R247.

[32] J. Wilken, S. B. Kent, *Curr Opin Biotechnol* **1998**, *9*, 412.

[33] A. J. Link, D. A. Tirrell, *Methods* **2005**, *36*, 291.

[34] W. Kim, V. P. Conticello, *Journal of Macromolecular Science, Part C: Polymer Reviews* **2007**, *47*, 93.

[35] H. Reiersen, A. R. Clarke, A. R. Rees, *J Mol Biol* **1998**, *283*, 255.

[36] D. W. Urry, R. G. Shaw, K. U. Prasad, *Biochem Biophys Res Commun* **1985**, *130*, 50.

[37] T. Yamaoka, T. Tamura, Y. Seto, T. Tada, S. Kunugi, D. A. Tirrell, *Biomacromolecules* **2003**, *4*, 1680.

[38] J. A. Hodges, R. T. Raines, *J Am Chem Soc* **2003**, *125*, 9262.

[39] R. Improta, F. Mele, O. Crescenzi, C. Benzi, V. Barone, *J Am Chem Soc* **2002**, *124*, 7857.

[40] J. A. Hodges, R. T. Raines, *J Am Chem Soc* **2005**, *127*, 15923.

[41] W. Kim, K. I. Hardcastle, V. P. Conticello, *Angew Chem Int Ed Engl* **2006**, *45*, 8141.

[42] L. L. Hench, J. M. Polak, *Science* **2002**, *295*, 1014.

[43] S. C. Heilshorn, J. C. Liu, D. A. Tirrell, *Biomacromolecules* **2005**, *6*, 318.

[44] J. C. Liu, S. C. Heilshorn, D. A. Tirrell, *Biomacromolecules* **2004**, *5*, 497.

[45] Y. Zhou, S. Wu, V. P. Conticello, *Biomacromolecules* **2001**, *2*, 111.

[46] N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann, R. Huber, *Eur J Biochem* **1995**, *230*, 788.

[47] D. Barth, A. G. Milbradt, C. Renner, L. Moroder, *Chembiochem* **2004**, *5*, 79.

[48] L. E. Bretscher, C. L. Jenkins, K. M. Taylor, M. L. DeRider, R. T. Raines, *J Am Chem Soc* **2001**, *123*, 777.

[49] M. Doi, Y. Nishi, S. Uchiyama, Y. Nishiuchi, T. Nakazawa, T. Ohkubo, Y. Kobayashi, *J Am Chem Soc* **2003**, *125*, 9922.

[50] S. K. Holmgren, K. M. Taylor, L. E. Bretscher, R. T. Raines, *Nature* **1998**, *392*, 666.

[51] A. V. Persikov, J. A. Ramshaw, A. Kirkpatrick, B. Brodsky, *J Am Chem Soc* **2003**, *125*, 11500.

[52] B. C. Gorske, B. L. Bastian, G. D. Geske, H. E. Blackwell, *J Am Chem Soc* **2007**, *129*, 8928.

[53] M. P. Hinderaker, R. T. Raines, *Protein Sci* **2003**, *12*, 1188.

[54] J. A. Hodges, R. T. Raines, *Org Lett* **2006**, *8*, 4695.

[55] C. L. Jenkins, L. E. Bretscher, I. A. Guzei, R. T. Raines, *J Am Chem Soc* **2003**, *125*, 6422.

[56] C. L. Jenkins, A. I. McCloskey, I. A. Guzei, E. S. Eberhardt, R. T. Raines, *Biopolymers* **2005**, *80*, 1.

[57] F. W. Kotch, I. A. Guzei, R. T. Raines, *J Am Chem Soc* **2008**, *130*, 2952.

[58] M. D. Shoulders, I. A. Guzei, R. T. Raines, *Biopolymers* **2008**, *89*, 443.

[59] S. D. Mooney, P. A. Kollman, T. E. Klein, *Biopolymers* **2002**, *64*, 63.

[60] C. M. Taylor, R. Hardre, P. J. Edwards, *J Org Chem* **2005**, *70*, 1306.

[61] J. M. Lehn, G. Wipff, *J Am Chem Soc* **1974**, *96*, 4048.

[62] L. S. Sonntag, S. Schweizer, C. Ochsenfeld, H. Wennemers, *J Am Chem Soc* **2006**, *128*, 14697.

[63] K. Kanai, B. Podanyi, S. Bokotey, F. Hajdu, I. Hermecz, *Tetrahedron Asymmetry* **2002**, *13*, 491.

[64] *http://en.wikipedia.org/wiki/Van_der_Waals_radius*.