1-26-2024

Elastin-like polypeptide as a model to study Intrinsically Disordered Proteins

Sadegh Majdi
Rowan University

ELASTIN-LIKE POLYPEPTIDE AS A MODEL TO STUDY INTRINSICALLY DISORDERED PROTEINS

by

Sadegh Majdi

A Thesis

Submitted to the
Department of Molecular and Cellular Biosciences
College of Science & Mathematics
In partial fulfillment of the requirement
For the degree of
Master of Science in Bioinformatics
at
Rowan University
August 05, 2021

Thesis Chair: Benjamin Carone, Ph.D., Associate Professor, Department of Biological & Biomedical Sciences

Committee Members:
Nathaniel Nucci, Ph.D., Associate Professor, Department of Physics & Astronomy
Chun Wu, Ph.D., Associate Professor, Department of Chemistry & Biochemistry

Acknowledgements

It is said that it takes a village to raise a child in an old proverb. With the same sentiment, I'll add that completing what I was able to do in this thesis needed an amazing group of people. As a result, I owe a debt of gratitude to a large number of people.

Scientists have played a significant role in the discovery and development of ELP including Dan W. Urry, Judith Ann Foster, Eveline Bruenger, William Gray, Lawrence Sandbergs, and Ashutosh Chilkoti.

Thank you, Ben, for teaching, advising, and friendship. Within the last two years, I grew up scientifically and personally by your company.

Thank you also to my committee members:

To Nathaniel Nucci for providing vital assistance in different steps of this research.

To Chun Wu, for introducing the I-TASSER platform to perform my research data.

To Behrooz Nazer and Gregory Caputo for helping me to join the bioinformatics master program.

And to all the members of Carone lab and Nucci lab, and MCB department staff for all their support.

And to all my friends and family for their constant love and support. I definitely could not have done this without you.

Abstract

Sadegh Majdi
ELASTIN LIKE-POLYPEPTIDE AS A MODEL TO STUDY INTRINSICALLY DISORDERED PROTEINS
2019-2021
Benjamin Carone, Ph.D.
Master of Science in Bioinformatics

The elastin-like polypeptide (ELP) is a well-studied structural protein that is easily amenable to amino acid (AA) sequence modifications and has the potential to yield a wide variety of uses in bioengineering and biomedical applications. One unique property of ELP is the inclusion of intrinsically disordered domains (IDP) within the structure that allow for its diversity of physical properties. While it is generally understood that amino acid sequence dictates protein folding arrangements, the contributions of specific amino acid sequences to the intrinsic disorder of ELP has yet to be fully resolved. Therefore, identifying the contributions of specific amino acid sequences to the final physical properties of ELP at the sequence level is an important step in understanding the basis of ELP's unique biophysical properties.

The results of this study resolve the role of amino acid sequence and repeat composition on the disordered structure of ELP. Experimentally, an enhanced synthetic ELP[AV-60] was developed, produced, and purified, with novel photocrosslinking and a transition temperature of 37°C. Computationally, the conserved AA sequence of valine-proline (VP) in the endogenous ELN sequence and its major function in the structure of ELP was identified. Finally, we identified a specific AA monomeric unit, VPGXAG with potential for application in IDPs research.

Table of Contents

Table of Contents (Continued)

Table of Contents (Continued)

List of Figures

List of Tables

Table Page

Chapter 1

Introduction

The Elastin-like polypeptide is a well-studied synthetic biopolymer inspired by the naturally occurring Elastin protein. Like all polypeptides, ELP is a chain of amino acids but it has unique sought after structural properties that are thought to result from a repetitive chain of five amino acids, Valine, Proline, Glycine, guest residue (could be any amino acids except proline), and Glycine. The iteration of these five amino acids, which are originally part of the ELN sequence, is thought to contribute to ELP's flexible structure. Unlike most other polypeptides, it does not have a well defined or stable tertiary structure, and this phenomenon is called an intrinsically disordered structure. Although the existence of IDPs has been known for decades, identifying the cause of their highly flexible structure has been challenging to resolve. In order to understand more about ELP structure as well as general IDP characteristics, we have chosen to perform in-depth characterization of ELP amino acid composition. ELP is a good choice for studying intrinsically disordered proteins because of its simple structure and highly repetitive AA sequence.

Elastin-like polypeptides (ELPs) are a type of stimuli-responsive biopolymers, polymers that respond to minor changes in environmental stimuli with noticeable physical or chemical alterations in their behavior, based on tropoelastin's intrinsically disordered domains, which are comprised of repetitions of the VPGXG pentapeptide motif, with X serving as a "guest residue". They go through a reversible, thermally triggered lower critical solution temperature (LCST) phase change that has been used in

protein purification, affinity capture, immunoassays, and drug delivery, among several other things. Although ELPs have been widely researched as protein polymers and biomaterials, much less is understood about their connection to other disordered proteins. ELPs have biophysical characteristics that are comparable to those of many intrinsically disordered proteins (IDP), which gives them their unique material behavior. ELPs are a fascinating "minimal" artificial IDP because of their low sequence complexity, phase behavior, and elastic characteristics, and studying them can give insights into the behavior of other more complicated IDPs. Inspired by the growing recognition of similarities between ELPs and IDPs. [1]

Elastin-Like Polypeptides

Elastin-like polypeptides (ELPs) are thermoresponsive polypeptides with a consensus repeat amino acid sequence of XPGZG, in which X is either valine (V) or isoleucine (I), and Z is an amino acid (along with non-canonical amino acids) besides proline[2]. The ELP sequence is one of the consensus repeat sequences derived from tropoelastin. Fifty years ago, Dr. Sandberg's research on elastin was the first insight for Elastin-like polypeptides. Seventeen peptides have been cut off from a tryptic digest of soluble elastin by ion-exchange chromatography and gel filtration. Molecular weights range from approximately 1,700 to 10,500. Around this same time, several different amino acid sequences, either complete or partial, were identified and made publicly available, identifying the primary sequence of elastin. From these datasets, several noteworthy observations were obtained including: 1) The sequences obtained expose an elastin primary structure quite different from collagen and possessing repeats of a tetrapeptide, Gly-Gly-Val-Pro, a pentapeptide, Pro-Gly-Val-Gly-Val, and a hexapeptide,

Pro-Gly-Val-Gly-Val-Ala. 2) Some of the peptides contain partial hydroxylation of proline residues, confirming hydroxyproline as a genuine elastin chain constituent [3].

ELPs undergo a lower critical solution temperature (LCST)-like transition in an aqueous solution in which they are soluble at lower temperatures and become insoluble when heated to a specific temperature, called the transition temperature (Tt). This transition to an insoluble state is assisted by a secondary structure change in the ELP from a random coil conformation when the ELP is soluble to a $\boldsymbol{\beta}$-turn conformation when the ELP is insoluble[4]. The ELP transition temperature is determined by several variables. The first of these is the overall hydrophobicity of the ELP sequence, which is tuned by the amino acid composition in the X and Z positions. Urry and coworkers demonstrated that the transition temperature of an ELP could be approximately expected by the average of the transition temperature contributions (on a molar basis) from variant fourth spot amino acids [2]. Later work by McDaniel et al. formulated a predictive model that also integrated the effects of ELP chain length and ELP concentration on ELP transition temperature in solution [2,5]. The other important factors that have an effect on the ELP transition temperature are the identities and concentrations of salts in the solution. Cho *et al*. analyzed the effect of Hofmeister anions on ELP transition temperature [2,5,6]. This study illustrated that chaotropic anions exhibited competing for salting-in and salting-out results on ELP solubility, whereas kosmotropic anions only exhibited salting-out effects. In addition, they revealed that the value of these salting-in and salting-out effects were related to a given ion's position in the Hofmeister series. The ability to predict solubility behavior of ELP accompanied by the highly flexible sequence has created ELPs an attractive protein material for a number of purposes, including protein-based hydrogels,

fibers, and underwater adhesives [7,8]. Using genetic engineering, ELP blocks can be fused to other protein genes of interest to produce fusion protein materials like the block copolymer or protein-polymer conjugates. The thermoresponsiveness of the ELP blocks can be used to drive fusion protein self-assembly in these systems.

ELP Applications

The magnificent control over the design of ELPs enables easy manipulation of their stimulus-responsive properties and other physical and functional characteristics. Their tunable physicochemical abilities and well biocompatibility resulted in extensive biological applications of ELPs as triggered molecular actuators for recombinant protein purification and drug delivery, and as stimuli-responsive materials for tissue engineering. The most popular applications are recombinant protein purification, drug delivery, tissue engineering, and actuators and transducers which are characterized next [9].

Recombinant Protein Purification

Around twenty years ago, Meyer and Chilkoti found that gene-level fusion of an ELP with a recombinant protein conveys stimulus responsive behavior (phase transition in response to temperature change) to the fusion protein. The stimulus-responsiveness of ELP fusion proteins ensures an easy and simple to use methodology for protein purification that has many positive aspects over current chromatographic techniques used to separate recombinant proteins.

The low cost of purification, technical simplicity (requires only a laboratory centrifuge or filter, all of which are readily available in most molecular biology and biochemistry laboratories), ease of multiplexing, and high yield are among these benefits.

The capability to stimulate the aggregation of ELP fusion proteins without affecting their function by temperature or ionic strength offers two orthogonal triggers that prepare a low cost and simple method to purify ELP fusions.

In this process for protein purification that we termed inverse transition cycling (ITC), the phase transition of an ELP fusion is typically triggered by the addition of heat or salt and the aggregated ELP is divided from the cell debris by centrifugation [10]. After throwing out the supernatant, the pellet of aggregated ELP is redissolved in the cold buffer. Centrifugation of solubilized ELP solution below the transition temperature allows for the removal of contaminating insoluble proteins that might have been taken down with the ELP pellet previously. This completes one round of ITC. Multiple ITC cycles improve the purity of the final ELP fusion protein, and after several cycles, highly purified protein is produced that is suitable for most biological applications.

Purification of an ELP fusion protein depends on the efficient aggregation of the ELP fusion in the complex milieu of cell lysate. Concentration affects the performance of this operation, such that at a low (<1 lM) concentration of the ELP fusion, the phase transition triggered aggregation is inefficient. In order to get around this problem, The effect of adding free ELP to cell lysate to increase the total concentration of ELP in solution was investigated and found the addition of free ELP to cell lysate allows the purification of small quantities of ELP fusion proteins from *E. coli* cultures; Ge and Filipe reported similar results [11,12].

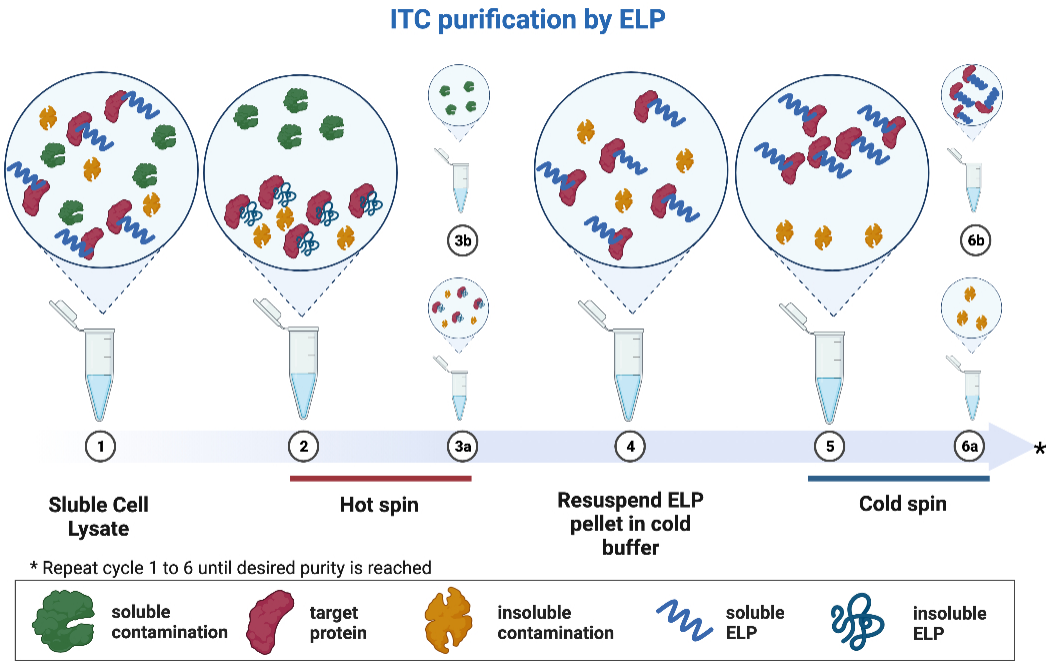
In the most widely used version of ITC, the ELP tag is fused to the target protein at the gene level. This direct method of purification with ELP tags has been used to purify

proteins such as thioredoxin (Trx), chloramphenicol acetyltransferase (CAT), calmodulin (CaM), green fluorescent protein (GFP), and blue fluorescent protein (BFP) [13]. It is possible to recover native proteins by combining self-cleaving inteins or peptide substrates of proteases such as thrombin or Factor Xa **(Figure 1.1)** between the protein and ELP [14]. Thus, the usage of ELP as a stimulus-responsive purification tag provides an economical and chromatography-free method to purify recombinant proteins.

The ELP fusion protein system is very robust, even though it is primarily used to purify fusion proteins formed in E. coli, as recombinant ELP fusion proteins have also been expressed in, and purified from plant expression systems such as tobacco leaves [15]. A different approach (indirect ITC) is to combine ITC with affinity chromatography methods utilizing an ELP chain that is linked to a protein that specifically binds to the target protein. Following binding of the target with the capture agent-ELP fusion in solution, purification of the complex is completed via ITC [16]. This indirect method avoids the need to cleave the ELP tag from the target protein and allows the capture agent-ELP fusion to be recycled. Using this strategy, a protein A-ELP fusion can be implemented to purify IgG and IgM antibodies. This indirect ITC approach can also be implemented in a different way, metal binding domains have been integrated into an ELP chain and bound to Ni2+. The Ni2+-ELP constructs were then applied to purify proteins tagged with an oligohistidine sequence by metal affinity capture [17].

Figure 1.1

Inverse Transition Cycling by ELP Purification Tag



Note. ELP is purified by means of its CST behavior from both soluble and insoluble contaminants. Starting with the ELP-rich lysate (1) the ELP transition is induced with the addition of heat and/or salt and the ELP is separated by centrifugation (hot spin) (2). The ELP pellet is retained (3a) while the supernatant containing soluble contaminants is discarded (3b). The ELP is resuspended in cold buffer (4) and centrifuged again at 4 °C (cold spin) (5). The pellet containing insoluble contaminants is discarded (6a) while the supernatant containing soluble EP is retained (6b). Cycles of alternating hot and cold spins are repeated until the desired purity is reached.

Drug Delivery

Having complete control over the size, sequence, and stimulus-responsive properties of ELPs provides a framework for the rational design of macromolecular carriers for use in drug delivery. The genetic engineering of ELPs enables the incorporation of targeting peptides, such as cell-penetrating domains or receptor ligands [18], presenting constructs with one or more copies of a targeting moiety—typically peptides that are ligands to cell-surface receptors to improve the uptake of conjugated therapeutics in pathological tissues.

At the gene level, control of ELP sequence also creates reactive sites along the ELP chain for chemical linking of drugs or fluorescent probes [19,20]; because the type, number, and location of these sites provided by chemically reactive peptide side chains is genetically encoded at the residue level of the ELP sequence, the degree of loading of the drug or imaging agent and its location on the ELP can be tuned at the molecular level. ELPs can also be designed to self-assemble into nanoscale constructs that are capable of circulating as assembled nanoparticles under physiologic conditions or can be triggered to self-assemble as a result of an external temperature stimulus.

ELPs' temperature responsive actions can be used to deliver therapeutics locally through the formation of a viscous coacervate, dense liquid droplets of macromolecules, triggered by physiological temperature. These ELP depots can release protein-ELP or peptide-ELP fusions or other nonpeptidic drugs that are chemically conjugated to the ELP. As ELPs are synthesized of amino acids, and the rate at which they degrade can be regulated by the ELP's design, As a result, nontoxic short peptide degradation products are produced [9].

Tissue Engineering

ELPs are a unique form of biopolymer that has a great deal of potential for the fabrication of highly customizable scaffolds for tissue engineering applications. Because ELPs are derived from an extracellular matrix (ECM) protein, they're especially well-suited for creating an ECM-like environment. that can deliver biologically relevant functional cues to cells and tissues. Furthermore, because ELPs are genetically encoded biopolymers, the gene-level design of an ELP allows for control over material parameters such as structure, chain length, polymer architecture, and the form, number, and location of crosslinking sites, all of which are essential in the design of tissue engineering constructs. The recombinant synthesis of ELPs also allows for the easy incorporation of bioactive peptide moieties, which can be inserted within the ELP or appended to the ELP's ends. Additionally, since the VPGXG motif is intrinsically biodegradable, the rate of degradation of the ELP scaffold can be tailored.

ELP-derived matrices have been used as vascular grafts, cell sheets, cartilage recovery, ocular surface scaffolds, soft-tissue replacement, and stem cell differentiation substrates because of their unique properties [9,31]. ELP scaffolds for tissue engineering can be generated by using three primary methods mentioned below: (1) coacervation of soluble ELP (2) physical crosslinking and (3) chemical cross-linking, which includes photocrosslinking, enzymatic crosslinking, and crosslinking by γ-irradiation. Next, We have a concise overview of the work that has been performed using these various techniques to create ELP scaffolds for tissue engineering [21,22].

ELP Coacervates. In the simplest technique for creating a tissue engineering scaffold utilizing ELPs, the shape of an ELP coacervate develops in response to a

temperature stimulation. Increasing the temperature of an ELP condensed solution above its Tt causes ELP chain aggregation and forms an insoluble coacervate phase. This gelation technique allows for easy Ex vivo cellular loading by simply adding heat to ELP-cell suspensions 23 and the cell-coacervate mixture can then be injected in vivo through a needle with a fine gauge, as the coacervate is a viscous fluid. In situ formation of the scaffold by injection of an ELP coacervate has the benefit of conforming to a defect site where the most structural and functional support is needed. We should point out, however, that ELP coacervates are not hydrogels, but they're better represented as viscous fluids, and that because of the coacervate's reduced structural flexibility and stiffness, this solution could be restricted in some applications 24.

Physical Crosslinking. Physical crosslinking via chain complexity and hydrophobic effects has been established with ELP triblock; an ELP triblock was constructed by capping a central hydrophilic domain with VPAVG-(IPAVG)4 hydrophobic blocks. The center hydrophilic domain of these triblock formed crosslinks that linked extensive networks of micellar nanoparticles. Engagement of the hydrophobic domains above their Tt by physical crosslinks between the hydrophobic domains of the triblock ELP. The mechanical features of these physically crosslinked hydrogels can be tuned with the choice of solvent in which the ELP gels. The hydrophobic content of the ELP triblock also helps to matrix stability, as larger hydrophobic domains in the ELP triblock created hydrogels with higher elastic modulus. While ELP modification allows for some mechanical property management, Some tissue engineering applications necessitate a high level of strength.

The strength of the hydrogel can be improved by forming chemical crosslinks between the ELP chains when greater mechanical integrity is required, as in load-bearing applications for cartilage repair [25,26].

Chemical Crosslinking of ELP Hydrogels. Serial lysine residues in ELPs were crosslinked into hydrogels by McMillan and Conticello using an electrophilic crosslinker, bis(sulfosuccinimidyl) suberate 7. Thermal cycling stimulated expansion or contraction of these matrices, retaining their temperature-sensitive properties. Temperature-induced structural changes, such as expansion and contraction, were visualized by cryo scanning electron microscopy (SEM) of rapidly vitrified ELP gels. Trabbic-Carlson similarly crosslinked ELPs containing periodic lysine residues with a trifunctional chemical crosslinker, tris-succinimidyl aminotriacetate to create hydrogels. It was discovered that the mechanical properties of these temperature-sensitive matrices are dependent on polypeptide concentration, the lysine repeat, and molecular weight of ELP. In ELP structures, the stiffness of crosslinked gels was maximized by high concentration, high molecular weight, and high lysine content 27,28.

Some crosslinkers are toxic to cells, while others cause the reaction to be carried out in an organic solvent, which is incompatible with the encapsulation of cells prior to crosslinking. In order to avoid these issues, Lim et al. demonstrated that ELPs containing periodic lysine residues can be crosslinked in aqueous conditions with hydroxymethylphosphines to develop hydrogels whose mechanical properties are regulated by the number of lysines incorporated in the ELP and the pH of the crosslinking reaction. Lim et al. created a library of single-segment ELPs of different molecular weights and lysine spacing for a comprehensive study of chemically

crosslinked ELP hydrogels. To examine the effect of these variables on the structural properties and morphology of the crosslinked ELP hydrogels, these ELPs were crosslinked at various ELP concentrations with an excess of a cytocompatible HMP, b-[tris(hydroxymethyl)phosphino]-propionic acid (THPP) [29].

The use of enzymes to construct crosslinked polymer scaffolds for tissue engineering is conceptually appealing because it is a bioinspired approach for creating usable tissue that mimics the normal crosslinking of proteins in vivo. McHale et al. is inspired by this idea and created two ELPs with periodic glutamine or lysine residues at the guest position (X) within the VPGXG pentapeptide motif. Tissue transglutaminase (tTG) can crosslink a mixture of ELPs by catalyzing the creation of a c-glutamyl-e-lysyl covalent bond between glutamine residues and primary amines. Inter-chain crosslinking between the glutamine residues and the amine of the lysine side chain occurred from mixing these ELPs in the presence of tTG enzymes. When compared to the ELP coacervate produced from a mixture of the two ELPs, enzymatic crosslinking of the ELPs substantially increased the shear stiffness of the ELP hydrogels. Since the tTG reaction is biocompatible, cells could be homogeneously combined in the ELP solution before crosslinking, allowing for cell encapsulation during in situ gelations [30].

Adjustable Tissue Engineering Features at Sequence Level. ELPs' genetically encoded ability enables other peptide sequences to be incorporated into them, which is an easy and convenient way to add functionality to an ELP for tissue engineering applications. Cell adhesion is essential in applications including engineered vascular grafts, where the creation of a confluent and functional endothelial cell layer is crucial to the graft's efficiency. ELP motifs can be adjusted to include peptide fragments from ECM

proteins, which can change how cells interact with the ELP scaffold. Implementing peptide sequences that mimic natural ligands found in ECM components to target the integrin class of cell surface receptors could facilitate binding by peptide sequences that mimic natural ligands found in ECM components. The CS5 domain of fibronectin, which contains the peptide REDV, has been shown to interact with α4$\boldsymbol{\beta}$1 integrin, encouraging endothelial cells to adhere and grow. In addition to surfaces with the CS5 domain and REDV sequence, RGD peptides interact with integrin targets such as α5$\boldsymbol{\beta}$1 to facilitate cell attachment and increase the intensity and speed with which cells can tether to a substrate 22,31–33.

Integrating protease cleavage sites into an ELP sequence is a convenient way to smooth ELP biodegradation. Rodriguez-Cabello and co-workers used this technique to facilitate the bioabsorption of ELP matrices in vivo, including the enzyme degradable sequence VGVAPG contained in natural elastin in the ELP design. These hexapeptide motifs are degraded by matrix metalloproteinases (MMPs), such as macrophage elastase (MMP-12), which decompose ELP hydrogels and promote the development of bioactive fragments ability to induce cell proliferation and neovascularization, which are both essential for tissue regeneration and healing [34,35].

Actuators and Transducers

The exceptional environmental sensitivity of ELPs is useful in the development of biomolecular transducers that go beyond the intrinsic temperature or ionic strength stimuli that all ELPs respond to. In this part, ELPs and their fusions that react to two different types of stimulus, light and biomolecular recognition, are addressed as examples

of ELPs that are engineered to respond to different stimuli. Sensors, controlled release systems, and bioanalytical instruments are expected to include these next-generation ELPs as molecular transducers and actuators [9].

Photoresponsive ELPs. Strzegowski et al. were the first to reveal modulation of the ELP transformation in reaction to light by incorporating an azobenzene chromophore into a random ELP copolymer. The ELP transformation was caused by the chromophore's cis-trans isomerization by UV irradiation within a narrow temperature range around 40°C. Alonso et al. conjugated a spiropyran chromophore to the glutamic acid side chains in random copolymers of VPGVG and VPGEG in a sort of similar approach. In PBS at pH 3.5, the spiropyran functionalized ELPs displayed photomodulation of the ELP phase transition over a 14°C temperature window. In the absence of UV light, spiropyran adopts a hydrophobic spiro conformation that assists the ELP transition by increasing the ELP's overall hydrophobicity, whereas UV irradiation converts the chromophore to its more hydrophilic merocyanine conformation, thereby raising the transition temperature and promotes dissolution of the ELP chains. The ability of ELPs to work as photoresponsive molecules was demonstrated in these experiments, as only a single spiropyran conjugated to 10 pentamers was required to induce this response 3,9,36,37.

Elastin-Like Polypeptides as Models of Intrinsically Disordered Proteins

Concept of Intrinsically Disordered Proteins

Intrinsically disordered proteins (IDPs), also known as intrinsically unstructured proteins, are globular proteins that lack a cooperatively folded structure under natural circumstances, in contrast to globular proteins that have organized and well-formed

domains. The conventional paradigm of protein structure, in which protein function is dependent on a fixed 3D structure, is challenged by IDPs [38].

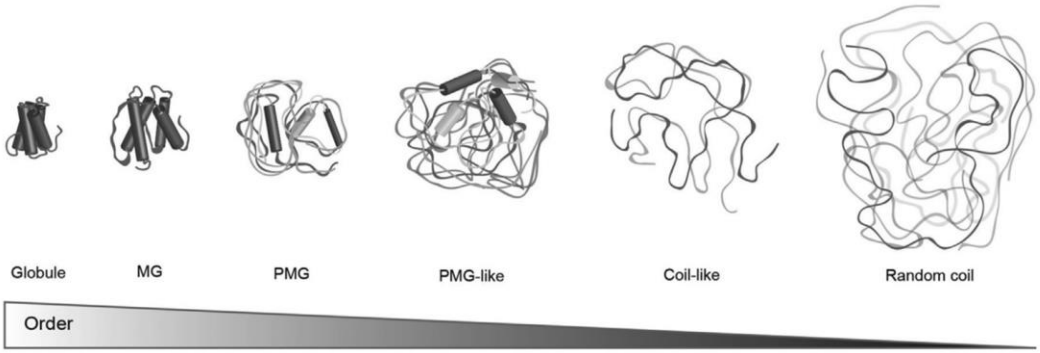
In eukaryotic proteins, disordered regions are relatively prevalent. Moreover, half of the eukaryotic proteome is made up of short sequences of 40 amino acids or more that are projected to be disordered under pathophysiological conditions. Until the mid-1990s, the frequency and functional relevance of IDPs were widely unknown. This can be explained in part by the absence of experimental and computational techniques for identifying inherently disordered regions (IDRs). Even now, characterization of disorder and assembly modeling remains a substantial difficulty that must be confirmed using a variety of methodologies. The structural characterization of IDPs is usually based on X-ray crystallography, when the interactions of IDPs with their physiological partners result in well-ordered structures, or nuclear magnetic resonance (NMR) spectroscopy, which has the additional advantage of providing insights into dynamic interactions, small-angle X-ray scattering (SAXS), single-molecule fluorescence energy transfer (smFRET), and computer simulations based on polymer physics [39].

Despite the lack of structured domains, ‘intrinsic disorder' encompasses a wide range of states, from the completely unstructured or random coil-like to partially structured (pre-)molten globules. All of these phases are required for basic cellular operations such as transcription, translation, cellular signaling, and regulatory activities. Disordered regions are divided into six functional groups depending on their functionality, and a single protein might be considered in multiple functional classes or contain several disordered regions.

The following are the classes: (1) entropic chains, which operate without being structured but advantage from conformational disorder directly; (2) display sites, where the structural flexibility of disordered protein regions promote post-translational modifications (PTMs) by permitting convenient access and recognition by effectors that regulate downstream outcomes upon binding; (3) Chaperones, over 50% of IDRs can adapt and aid RNA and proteins in obtaining their functionally folded forms. (4) effectors, which bind to other proteins and regulate their activity through competitive interactions or allosteric modulation (this disorder-to-order transition is also known as coupled folding and binding); effectors, which bind to other proteins and regulate their activity through competitive interactions or allosteric modulation; (5) multiple binding partners bring disordered sequences together to facilitate the creation of higher-order protein complexes; assemblers, where multiple binding partners bring disordered sequences together to facilitate the production of higher-order protein complexes; and (6) scavengers, tiny ligands are stored and neutralized **(Figure 1.2)** [40–42].

Figure 1.2

Protein Conformational States Range



Note. Protein conformational states range from completely unstructured or random-coil-

like structures through partially structured premolten globule (PMG) and molten globule (MG) structures to organized forms. Any of these states can be transitioned by proteins. Staby et al [39] work was adapted with permission. CCC stands for Copyright Clearance Center, Inc [36] .

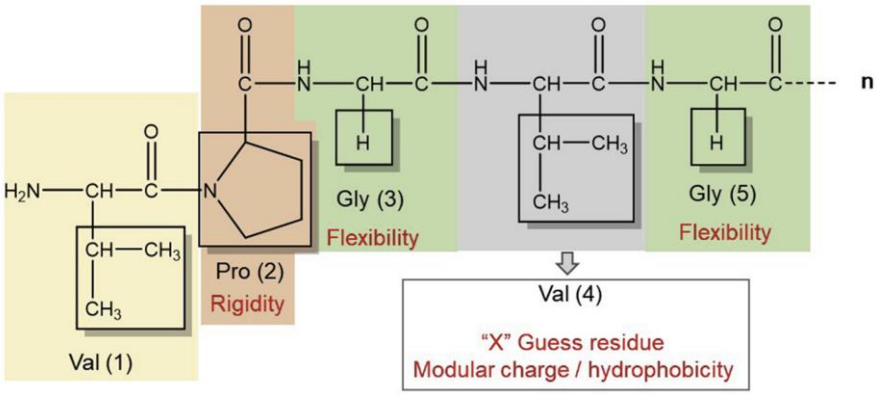
Elastin-Like Recombinamers (ELRs) as Intrinsically Disordered Proteins

Elastin is a crucial extracellular matrix (ECM) protein found in many vertebrate tissues that need elasticity and resilience, such as blood vessels (50 percent of dry weight), elastic ligaments (70 percent of dry weight), lungs (30 percent of dry weight), and skin (2–4% of dry weight). precursor. For many years, elastin's significant insolubility restricted its utilization to researchers looking for soluble derivatives with qualities comparable to those seen in the original protein. Tropoelastin derivatives (ELRs) derived from recombinantly synthesized tropoelastin demonstrated many of the features of the mature biopolymer, and thus provided a rich supply of adjustable building blocks for the production of biomaterials [4344].

The ‘guest residue' is the fourth amino acid valine (Val) in the consensus repeat unit (VPGVG)n. Except for proline, any amino acid can be used in this position. Indeed, changing the amino acid in that location might drastically change the physicochemical properties of the resultant recombinamer, making it a crucial parameter to consider during ELR design **(Figure 1.3)** [45].

Figure 1.3

ELP Monomeric Unit Structure



Note. The proline/glycine content, number of tandem (n) repetitions, and guest residue composition ('X') of the Elastin-like recombinamer (ELR) consensus repeat unit VPGXG and amino acid residues responsible for their primary physicochemical features [36] .

Elastin is an example of a protein whose function is dependent on the interaction of structure and dynamics, and it is intimately tied to disorder in its primary structure and solvate [46]. The capacity to contract reversibly after stretching due to a rise in entropy is a crucial functional property of elastin, [47] and it is commonly assumed that the hydrophobic domains of elastin are responsible for their elastic recoil [48]. After decades of divisive and seemingly contradicting structural investigations, an increasing agreement on the structural disorder of elastin and its derivatives has emerged.

The structure-function relationship of elastin has been a topic of continued interest and disagreement since the early hypotheses offered by Hoeve and Flory [49] and Aaron and Gosline, [50], which suggested random coil conformations as the cause of elastin's elasticity. Given the challenges of analyzing the structure of natural elastin, pioneering

investigations on structural characterisation have been conducted by Urry et al. and synthesized elastin-like peptides based on the most frequently repeating sequences inside the hydrophobic domains, such as (VPG)n, (VPGG)n, (VPGVG)n, and (VAPGVG)n. Urry et al. suggested that these sequences adopt a type II -turn structure, with neighboring proline and glycine residues creating the turn's corners and a hydrogen link between the main chain (C–O) of residue i and the (N–H) of residue i+3 stabilizing the turn. Non-polar side chains are arranged in this arrangement to keep water molecules out. Due to the regularity of these -turns in the peptides, the creation of an ordered -spiral consisting of sequential type II β-turns, as characterized by just under three VPGVG motifs per spiral turn, is predicted. However, structural study of comparable (GVPGV)7 elastomeric sequences and domains including VPGVG and VAPGVG repetitions revealed that the β-spiral model is unstable, with only isolated and fluctuating β-turns present [50–52].

The possibility to recombinantly create specific hydrophobic domains or higher ELRs, has allowed predictions of disorder and conformational flexibility to be proven, in which the formation of extended secondary structures is limited in favor of transitory and changing local motifs, such as β-turns, β-strands, and polyproline II (PPII) structures. [53,54].

Tandem repetitions with minimal hydrophobicity and sequence complexity are typical features of ELRs and IDPs. This concept is supported by the consensus repeat sequence unit (VPGXG), which is derived from the hydrophobic domains seen in tropoelastin (the monomeric precursor of natural elastin) [55]. The high proline and glycine content in the ELR backbone is thought to be a cause of the structural disorder, with a compositional threshold of $(2P+G) \geq 0.60$, meaning that the percentage of proline is a

larger predictor than the number of glycine residues. Both amino acids, for opposing reasons, help preserve the polypeptide backbone of elastomers disordered and hydrated. Because of its constrained φ dihedral angle, proline is too stiff to form secondary structures and provide structural stiffness on all length scales of ELRs, inhibiting hydrogen-bonded backbone self-interactions. As a result, proline is a typical extended secondary breaker (α-helix and β-sheet). In the same way, the lack of a side chain on glycine residues allows the ELR to sample a range of chain conformations in the presence of water, resulting in a greater backbone entropy, which also makes it difficult to form ordered assemblies [53,55].

When IDPs engage with one or more partners, the entropic force that causes conformational alterations happens as a result of ligand binding, scaffolding, or allosteric regulation, whereas various triggers can cause ELRs to adopt a new conformation set when swelling in water [55].

The ELR chains separate from the solvent when heated over a threshold LCST, benefiting from the increase in solvent entropy from the release of solvent molecules and also the gain in favorable interchain interactions. Changes in the parameters of the intensive solution (typically temperature, but also pH or pressure), and/or in colligative properties, can result in the simultaneous collapse of the chain into amorphous coacervates (‘fuzzy aggregates’) and break the hydrophobic solvation. With the rising temperature, the entropic cost of restoring ordered solvent molecules surrounding individual chains rises. As a result, the temperature dependency is influenced by the presence of charged residues and the solvation of proline-rich areas. Additional sequence

characteristics, such as the distribution of polymeric sequences into distinct block copolymer structures, modify the chain-solvent and chain-chain interactions [45,56].

In summary, a disorder in ELRs is studied at several levels, including sequence, aggregation, and phase behavior. At the sequence level, several factors are known to affect the state of chain disorder. Das and Pappu [57], for example, have demonstrated the significance of the amino acid tandem repeat mixing parameter. Wright and Conticello [58] found spherical micelles in a variety of co-polymers with the sequences (VPGX1G)n and (VPGX2G)n, where the guest residue X1 is considerably more hydrophobic than X2. When two guest residues (especially alanine and valine) are fully mixed in the polymer (VPG [X1/X2]G)n, however, solubility is maintained throughout the polymer chain [5,58].

Tools Utilized to Investigate Intrinsically Disordered Domains

There are many websites and different software packages commonly used to predict the 3D structure of proteins like IntFOLD, RaptorX, Biskit, ESyPred3D, FoldX, RaptorX, I-TASSER, ROBETTA, and Rosetta@home.

Predicting the structure of IDPs is a complicated but interesting field that every year a couple of new software emerges in this area. SEG was the first predictor in this field that was published in 1994. SPOT-Disorder2, Disprot, IUPred, and PONDR are the other predictors that were published later.

To have an accurate result for more than 200 ELP sequences, the combination of typical software for predicting the structure of proteins and specific tools to predict IDPs structure was used. Among protein structure prediction software I-TASSER was selected. I-TASSER is an Online server for protein modeling that works based on threading

fragment structure reassembly. I-TASSER used Ab initio quantum chemistry methods which are computational chemistry methods based on quantum chemistry. It is easy to use based on the amino acid sequence in FASTA file format. On the other hand, PONDR as a well-known disorder prediction software was selected because it is fast and easy to use.

The Iterative Threading Assembly Refinement (I-TASSER)

The iterative threading assembly refinement (I-TASSER) server is an integrated platform for automated protein structure and function prediction based on the sequence-to-structure-to-function paradigm. I-TASSER produces three-dimensional (3D) atomic models using numerous threading alignments and iterative structure assembly simulations, starting with an amino acid sequence. The function of the protein is then deduced by comparing the 3D models to other known proteins structurally. Full-length secondary and tertiary structure predictions, as well as functional annotations on ligand-binding sites, Enzyme Commission numbers, and Gene Ontology terms, are all included in the output of a normal server run. An estimate of the accuracy of the predictions is provided based on the confidence score of the modeling. Based on the modeling's confidence score, an estimate of prediction accuracy is presented. This protocol offers valuable insight and directions for the development of online server systems for current protein structure and function predictions. The server is available at http://zhanglab.ccmb.med.umich.edu/I-T[59].

In this study, a local Linux server was used to implement the I-TASSER program suite (with zhanglab permission) to calculate the C-score, which is the most important

indicator for ordered or disordered structure. C-score is in [-5,2] range and C-score > -1.5 indicates a model of correct global topology.

Prediction of Naturally Disordered Regions (PONDR)

Intrinsic protein disorder is a typical property of native proteins that is most likely encoded by the amino-acid sequence. As a result, the disorder should be considered a unique category of protein structure. A protein with the inherent disorder can bond to a partner indefinitely and hence persist as a structured protein for the rest of its existence. Proteins with the intrinsic disorder can go through disorder-to-order transitions when they bind to one or more partners, allowing them to be structured and unstructured at the same time. Finally, intrinsically disordered proteins can perform functions without ever being ordered, and hence stay disordered throughout their lives. A new paradigm for protein structure/function is required to account for all of these possibilities; as a result, The Protein Trinity has been considered as a potential model for 3D Protein structure, as depicted and presented in **(Figure 1.4)** [60].

Determination of Intrinsic Disorder

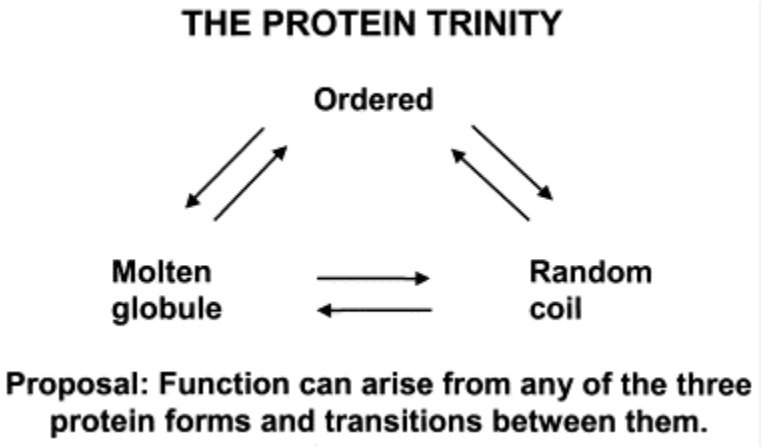
To define the intrinsic disorder in proteins, several approaches have been utilized, each with its own set of strengths and weaknesses, including X-ray crystallography, NMR spectroscopy, Circular dichroism (CD) spectroscopy, Protease digestion, and Stoke's radius determination.

Dunker and his collaborators created a number of predictors of natural disorder regions (PONDRs), which use amino acid sequences to predict order and disorder.

These predictors, in essence, compare a given sequence to a set of structurally described orderly and disordered sequences used in training.

Figure 1.4

An Alternate Structure/Function Theory for Proteins



Note. An alternate structure/function theory for proteins. The Protein Trinity proposes that native protein structure contains not only the ordered state, but also random coils and molten globules, as depicted in the diagram. Furthermore, this theory suggests that function can be performed by any of the three states or their transitions, not only the native state [60] .

ELN Gene and Elastin

ELN gene (Appendix A) produces one of the two proteins that make up elastic fibers. Elastic fibers are a kind of extracellular matrix that provide organs and tissues including the heart, skin, lungs, ligaments, and blood vessels with elasticity. Hydrophobic amino acids like glycine and proline are abundant in the encoded protein, forming mobile

hydrophobic regions bounded by crosslinks between lysine residues. The encoded protein's degradation products, known as elastin-derived peptides or elastokines, bind to the elastin receptor complex and other receptors, inducing monocytes and skin fibroblasts to migrate and proliferate. Elastokines can potentially hasten the growth of cancer. Supravalvular aortic stenosis (SVAS) and autosomal dominant cutis laxa are genetic diseases linked to deletions and mutations in this gene. [61]

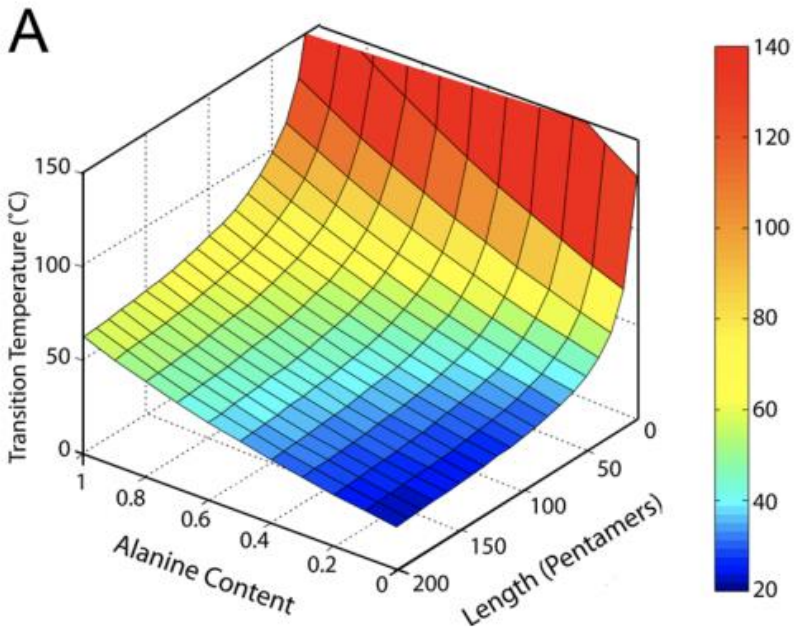
ELP [AV60-KCTS]

Chilkoti and Jonathan studied more than 25 sequences of ELP with different ratios of the guest residue of Alanine and Valine. They modeled the temperature transition of these sequences in different concentrations of buffer and reported that the ELP[AV-60] had the Tt at 37°C, which is the ideal temperature for in vivo studies. ELP[AV-60] has a 60 amino acids length containing 12 repeats of two pentapeptides. One of them has Alanine as a guest residue and another one has Valine as a guest residue **(Figure 1.5)** [5,62].

On the other hand, Ali and Nasim designed a photocrosslinkable hydrogel based on ELP. Their ELP sequence was [[VPGVG] 4 IPGVG]14. Also, they attached KCTS residue as a flanking side at both ends of the ELP sequences to enhance photo-crosslinkble properties to their polypeptide chains. UV light stimulates the bond between Cysteine residues of different chains that help to form a polymer network of ELP-KCTS chains [63].

Figure 1.5

A Three-Dimensional Plot of the Estimated Tt Landscape for ELPs



Note. (A) A three-dimensional plot of the estimated Tt landscape for ELPs related to the Ala and Val superfamilies in PBS at 25 M [5].

As previously stated, ELP has been utilized for a variety of applications, and the concept of applying ELP as a model for analyzing IDPs is not new. However, a rational explanation for the genesis of this physical property as resulting from the underlying amino acid sequence remains elusive. Because presently there is no rational method to tune the ELP characteristics to specific desired characteristics, thus far, ELPs have been utilized in many applications blindly.

Thesis Statement

The goal of this project is to take a bioinformatic approach to understand the genesis of ELP physical characteristics and natural intrinsic disorder by analyzing targeted amino acid modifications in the ELP protein. To achieve this objective, we began with a preliminary statistical analysis of the native ELN sequence, attempting to investigate the similarities between this sequence and ELP. Computationally, the degree of intrinsic disorder was calculated and then used as a physical property indicator for 220 distinct in silico ELP sequences in an attempt to identify a significant association between the monomeric unit structure, a guest residue amino acid, and the degree of IDP. Finally, a recombinant production process was developed and tested with a ELP[AV-60] for the experimental investigation of these sequences.

Understanding the scientific basis of intrinsically disorder in ELP and then tuning this feature depends on examining the amino acid sequence of this peptide and its consequence on the folding and physical structure of this peptide. This approach will enable researchers to precisely tune the ELP properties according to their biological application of interest and create bioengineered proteins with defined functionality.

Chapter 2

Investigation on Intrinsically Disordered Domains of Elastin

Investigation of the elastin sequence is fundamental for this research project since this protein originally inspired the synthesis of ELP. Hereupon, the amino acid sequence of ELN was examined to determine the origin of the physical properties of elastin. Finding the relationship between the physical properties of elastin and its amino acid sequence will enable us to design polypeptides with desired physical properties.

ELN Gene

The Elastin gene (Homo sapiens) has 786 amino acids length with 68398 Da molecular weight located on the 7q11.23 of the human chromosome. The ELN gene includes 34 exons and is responsible for the expression of elastin protein in humans [64]. The ELN gene encodes one main component of elastic fibers. Hydrophobic amino acids like glycine and proline are abundant in the encoded protein. The abundance of glycine and proline amino acids in the entire ELN sequence are 28.62% and 12.72% respectively.

This gene has many transcript variants that encode various isoforms. Tropoelastin is the soluble precursor of elastin. The entropy-driven process of elastic recoil is consistent with the characterization of the disorder. According to the findings, conformational disorder is required for the characterization of elastin structure and function [65–67].

Sandberg's [3] analysis on ELN genes revealed repetitive tetrapeptide, pentapeptide, and a hexapeptide sequence that was the first insight for developing elastin-like polypeptides as an elastic biopolymer for a variety of different applications. They

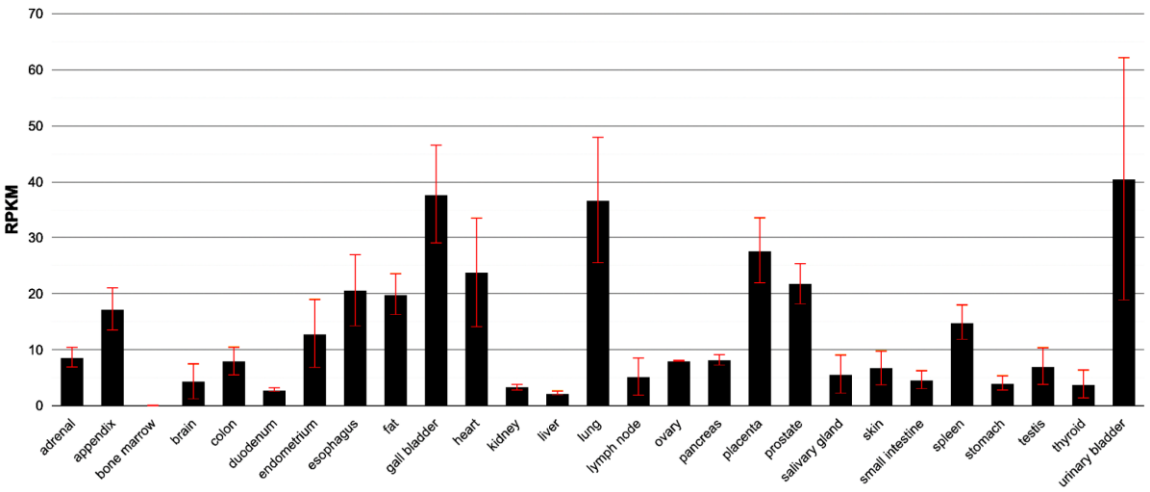
revealed that short tandem repeats of GGVP, PGVGV, and PGVGVA are repetitive sequences that may be candidates for the unique mechanical behavior of elastin, especially elasticity function.

Elastin

Elastin is an important extracellular matrix protein. It is very elastic and found in connective tissue that allows various tissues in the body to return to their original shape after being stretched or contracted. When skin is poked or pinched, elastin helps it return to its original shape. Elastin is a load-bearing tissue that is found in the bodies of vertebrates and is used in places where mechanical energy must be stored in human bodies [68]. Surprisingly, there is generally no additional turnover of elastin after initial deposition over the course of a lifetime [69].

Figure 2.1

The RPKM Listed for ELN Gene in 27 Different Organs



Note. The RPKM listed 27 organs. Top three organs with the highest read for ELN gene

are: Urinary bladder, gallbladder, and lung respectively. (Adopted from NCBI for ELN (homo sapiens) gene summary)

In order to represent the level of expression of Elastin in various tissues, next generation sequencing can be used to count the number of RNA molecules extracted from a specific tissue. Using this technique, Reads Per Kilobase of transcript, per Million, mapped reads (RPKM) can be counted and then used to create a normalized unit of transcript expression. Analysis of published RPKM values demonstrates that specific tissues have significantly increased Elastin expression **(figure 2.1)** [70]. In taking a closer look at this data it is clear that tissues which require the highest level of elasticity also exhibit the highest levels of ELN gene expression. For example, the greatest levels of RPKM are found in the urinary bladder, gallbladder, lung, placenta and heart respectively. In the heart, elastin is a key component of arterial tissues, the thoracic aorta being composed of up to 50% elastin by dry weight. For all these highly elastic tissues, the Elastin gene is also highly expressed.

By comparing the expression levels of the elastin gene in different tissues with the physical characteristics of that tissue's extracellular matrix, the relationship between the two may be better appreciated. As a consequence, we have a decent idea of how much ELP we'll require when building an ELP with specific physical features.

Naturally-Occurring ELN Gene Mutations

Remarkable findings are revealed regarding the mutations within the ELN gene and the associated physical consequences. These results shed light in revealing the genetic

origin of elastin's unique characteristics. Human ELN includes 118 mutations that 17 of them are non-synonymous based on genetic variation resources.

Worth of note, supravalvular aortic stenosis (SVAS), Williams-Beuren syndrome (WBS), and autosomal dominant cutis laxa (ADCL) are the diseases developed by major elastin gene abnormalities. Whether more subtle polymorphisms in the human elastin gene could affect the assembly and long-term durability of the elastic matrix, given the low turnover of elastin and the requirement for long-term durability of elastic fibers were investigated. While these mutations are unlikely to have an immediate physiological impact due to the limited turnover of the elastic matrix, they may affect susceptibility to degradative influences over the course of a person's lifetime, thus increasing the risk of later-onset cardiovascular disease [71].

John and David in their study on "Polymorphisms in the human tropoelastin gene … and mechanical properties of elastin-like polypeptides [72]" simulated the ELN mutations into ELP in order to evaluate their physical properties. Two of these mutations, G422S (GVP[S]VGG) and K463R (AAA[R]AAA), were introduced into elastin-like polypeptides as well as full-length tropoelastin, causing changes in their assembly and mechanical characteristics. G422S, which is prevalent in up to 40% of European populations, was discovered to improve various elastomeric qualities [72].

G422S is found in a repetitive VPG sequence in domain 20, which is an example of a conserved motif thought to be involved in the formation of β-turn structures in the protein's hydrophobic domains. K463R is a residue that aids in the formation of crosslinks that work to maintain the polymeric matrix stability. Coacervation studies

found that while G to S substitutions had no effect on assembly properties, K to R substitutions reduced the temperature of coacervation, exaggerating the effects of these relatively conservative substitutions [72].

On the opposite side, local changes in hydrophobicity caused by replacing the lysine residue with a more hydrophobic arginine residue appeared to be sufficient to result in a small but measurable decrease in coacervation temperature for the mutated ELP in the case of the K to R substitution [72].

Based on mechanical investigations, G to S and K to R substitutions in ELPs, as well as G to S substitutions in full-length tropoelastin, resulted in materials with different mechanical properties. While a single G to S substitution at position 422 reduced the percentage of energy loss and percentage of stress relaxation, a triple G to S substitution resulted in an increase in these properties in full-length tropoelastin and structural integrity loss in the ELP. According to the current models, when elastin is stretched, hydrophobic amino acids are exposed to the surrounding solvent, reducing entropy and providing a driving force for the return to the relaxed state. The observed beneficial effects for the single G to S substitution could thus be due to local structural changes in the domain that contribute to a larger decrease in entropy upon extension. The allele responsible for the G to S substitution (dbSNP reference: rs2071307) occurs with a frequency of up to 40% in certain human populations (e.g., those of European ancestry). Their findings suggest that people who have this allele should expect better performance from their elastin-bearing tissues, which could explain why this allele is so common. On the other hand, the catastrophic loss of structural integrity associated with the triple substitution in the ELP indicates a considerable change in local structure, apparently due

to considerable realignment of neighboring cross-linking domains. The existence of many extra domains in the ELP would partially rescue the right assembly architecture and crosslink creation, but the dramatic effect observed in the ELP would be mitigated by the existence of the triple substitution in the full-length protein [72].

John and David's [72] findings point to two critical elastin mutations that should be investigated further. The hypothesized influence of the G to S substitution on entropic changes during protein extension is of particular interest, and this hypothesis may be tested using well-established molecular dynamic models. These findings demonstrate that even minor changes in human elastin can have a profound influence on its assembly and mechanical capabilities. Given the protein's lengthy lifespan, such defects have the potential to change vulnerability to cardiovascular disease over the course of a lifetime. Because such diseases do not appear until later in life, evolutionary pressure against them would be low.

Analysis of Disordered Regions in The Human Elastin Gene (ELN)

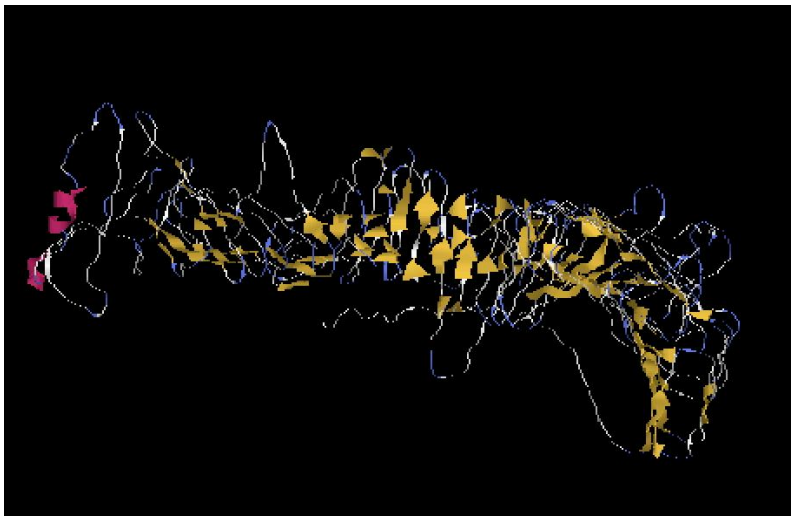
I-TASSER Result for ELN Gene

Linux code under the license of Zhang lab was used to obtain C-score for ELN amino acid sequence. As mentioned in the introduction, C-score is an indicator based on the I-TASSER platform to determine how predictable the protein structure or ordered structure. C-scores lower than -1.5 indicates protein structure prediction is hard and this is most probably because of the disordered structure of a protein. In this manner, C-score was obtained for the ELN sequence by -1.04.

Combining the C-score with PONDR results, which will be discussed later, it seems that although the structure of this gene does not seem disordered based on C-score, there are a lot of disordered regions in the ELN sequence. ELN sequence doesn't have a repetitive sequence like ELP, So C-score is not enough parameter to discuss the order or disorder structure of ELN.

Figure 2.2

ELN Amino Acid Structure Predicted by I-TASSER



Note. ELN amino acid ribbon structure predicted by I-TASSER

Statistical Analysis on PONDR Result

Identifying disordered regions in ELN sequences is very important to understanding their unique physical behavior. The statistical analysis of disordered regions and

comparing them with ordered regions will bring us very close to the goal of finding the relationship between amino acid sequences and their effect on the physical properties.

PONDR is one of the most popular platforms for finding disordered regions and suggests different algorithms for this prediction. In this statistical study, we used PONDR with the VL-XT predictor. As presented in **(figure 2.3)**, the PONDR score is shown in the vertical axis that indicates the highest probability of disorder, and zero indicates perfectly ordered regions. Also, X Axis indicates the position of each amino acid in the entire ELN amino acid sequence. This graph helps us to recognize ordered and disordered regions based on PONDR score. As an example, for the sequence of amino acids 75-99 scores are continuously zero which means this region is ordered and for the sequence of amino acids 1-11 scores are continuously one which means this region is disordered.

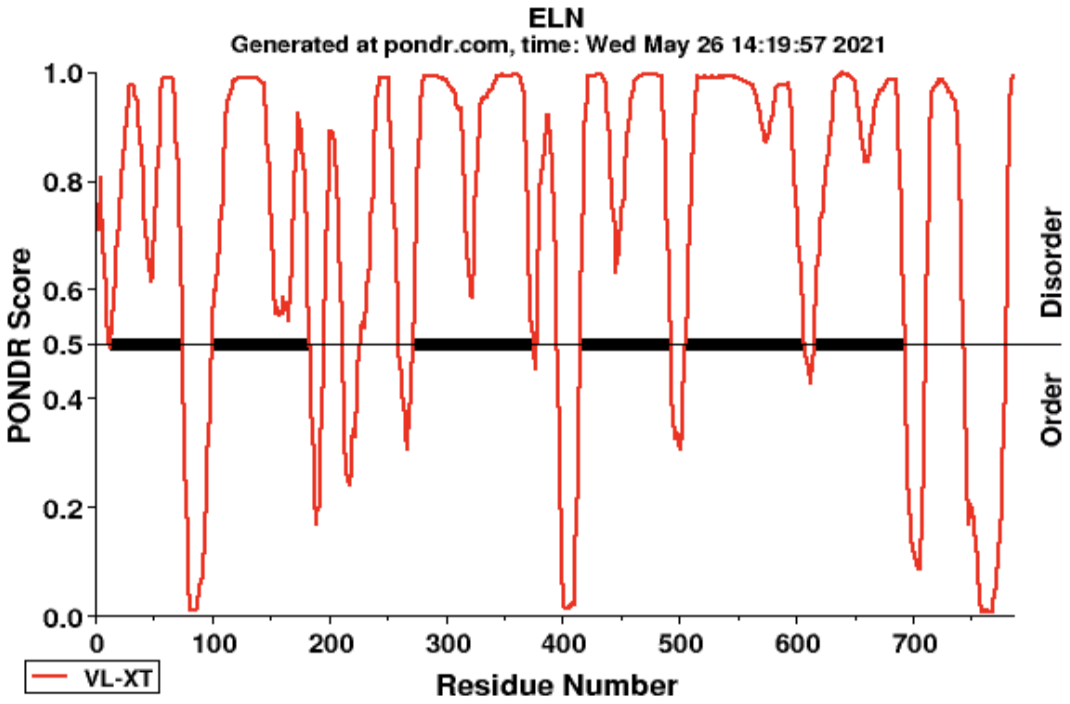
Twelve disorder regions with an average length of 51.75 amino acids and three order regions with an average length of 27 amino acids are recognized and mentioned in the table (Appendix B).

To have an attitude regarding the distribution of proline, glycine, and valine in disorder regions of ELN sequence a computational study was performed. This analysis helped us to find some relationship between ELP's most prevalent amino acids (proline, glycine, and valine) and the intrinsically disordered regions of ELN. Regarding this matter, individual amino acids of proline, valine and proline were counted on each disorder reigns and then divided by the length of that region to find the percentage of that amino acid contribution in a specific region. Moreover, as a curiosity, the number of VP, PG, VPG, and VPGXG , in the same order of amino acids as ELP, were counted and then

contribution percentages were calculated (Appendix C). As it is obvious in figure 2.3, there were three regions with lowest PONDR score which indicate the highest level of ordered regions. To have a better insight, the number of V, P, G, VP, PG, VPG and VPGXG were counted for these ordered regions to compare with disordered regions.

Figure 2.3

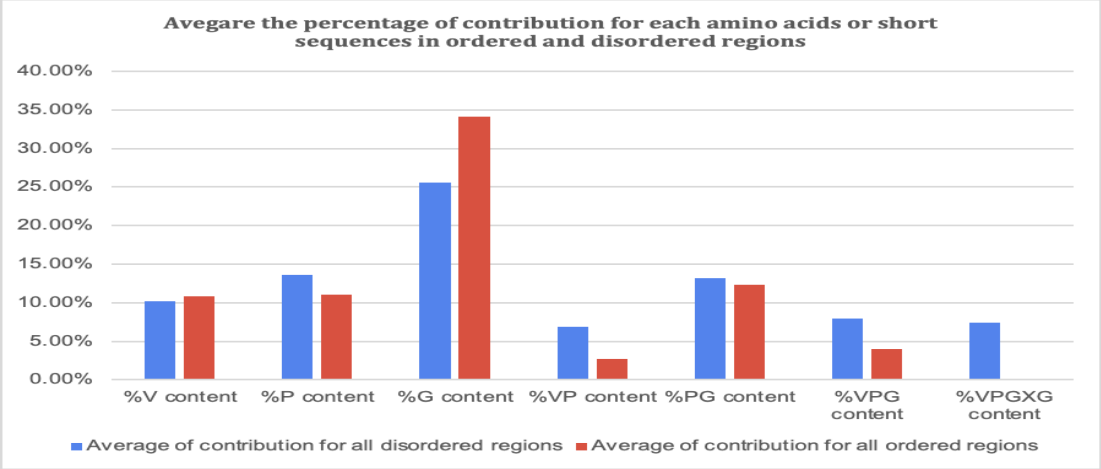
PONDR Score Result for the Entire ELN Gene



Note. PONDR score result for the Human ELN amino acid sequence. Results depicted 12 disorder regions with 1.0 score and 3 fully ordered regions with 0.0 score.

Figure 2.4

Average Contribution of Amino Acids for Disorder Regions and Order Regions



Note. Average contribution of V, P, G, VP, PG, VPG, and VPGXG for disorder regions (Blue) and order regions (Red) calculated by percentage. As an example the average contribution of valine in all disorder regions was 10%.

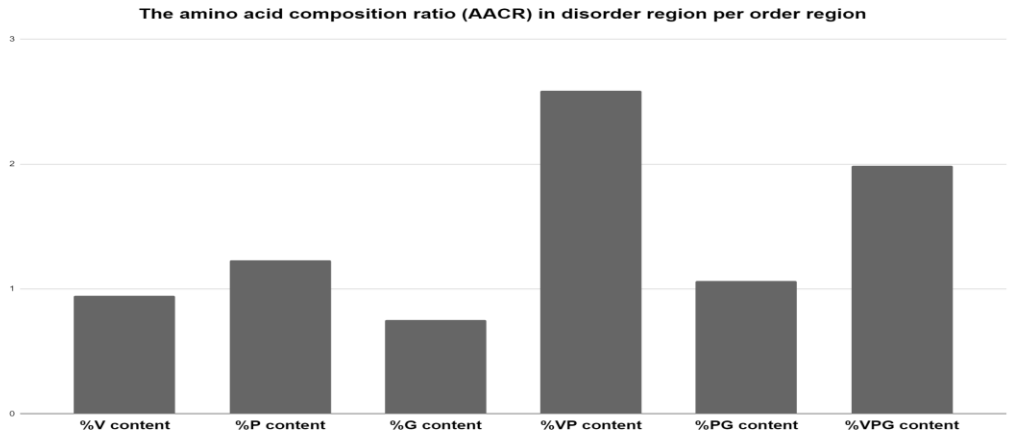
The amino acid composition ratio (AACR) in disorder region per order region is a valuable statistical indicator for examining the abundance of amino acids in disorder and order regions **(figure 2.5)**. This ratio was 0.94, 1.23, and 0.75 for valine, proline, and glycine respectively. These ratios indicate the greater presence of proline in disorder regions and glycine is more abundant in order regions and the frequency of valine is almost the same in both regions. These results are consistent with most previous studies that a large presence of proline can be a strong indication of disorder regions.

AACR were 2.59 for VP, 1.06 for PG, and 1.99 for VPG were calculated which means that the abundance of VP sequence in disordered regions is more than two times higher that ordered regions. These findings suggest that VP and VPG could be more

accurate markers than just proline for distinguishing disorder regions in human ELN sequence. Full detailed table presented in Appendix D.

Figure 2.5

The Amino Acid Composition Ratio (AACR) in Disorder Region Per Order Region in Human ELN Sequence



Note. The amino acid composition ratio (AACR) in disorder region per order region in Human ELN sequence.

In the human ELN sequence, the VPGXG (X might be any amino acid except proline) sequence can be found in 16 different locations. Only two of the sixteen VPGXG sequences had low PONDR scores: one in position 503-507 with a PONDR score of 0.3 and the other in position 606-610 with a PONDR score of 0.4. The rest are in completely disorder regions with the maximum PONDR score. These findings show that the ELPs formed by the repeating of this monomer unit have very disordered and unpredictably structured architectures.

Based on the information given in the previous chapter about the relationship between mutations and their physical effects, the PONDR score was examined for the rs2071307 dbSNP. In this mutation, sequence VPGVG is replaced by sequence VPSVG in position 422. VPSVG is located in a completely disordered region and the PONDR score is 0.8366 for (VPSVG)6 which is presenting almost the same graph score as (VPGVG)6. It illustrates this mutation does not have a significant structural effect.

Overall, these results suggest that the VP sequence has a key role for disorder regions and they are most likely the conserved sequences. Also, VPGXG sequences are markers for disorder regions where mutations happen very rarely. Interestingly, they still had disorder structure even after mutation.

Investigating The Conservation of Amino Acid Composition Using a Comparative Biology Approach

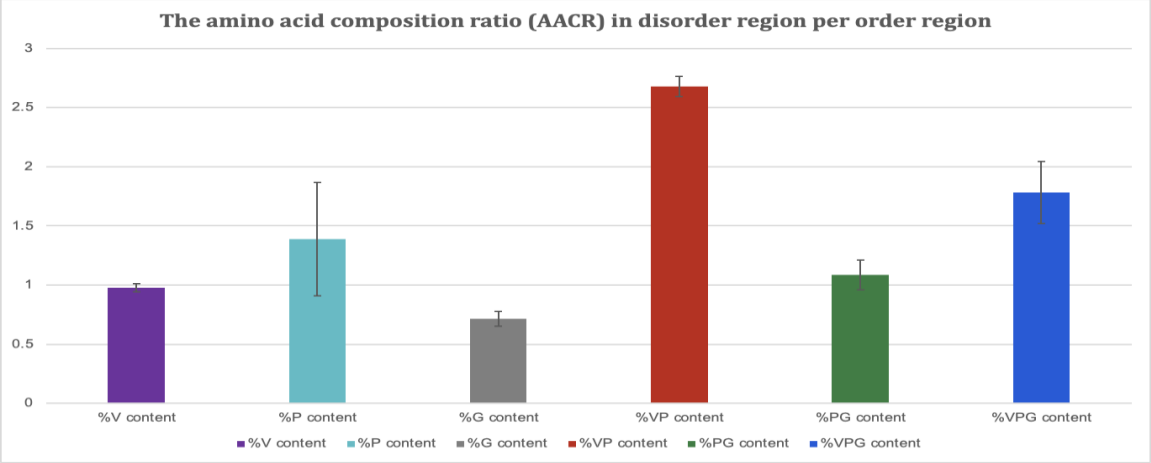
In order to investigate the potential function of enriched amino acid sequences in elastin genes ELN orthologs were identified using BLAST. Decent alignments have been found via Pan troglodites, Pan paniscus, Pongo Abelii, Gorilla and Mus musculus. These five ELN genes were uploaded on PONDR to achieve the primary results including order and disorder regions (Appendix A,B). Then, corresponding to what was done previously for the ELN Homo sapiens gene, statistical analysis was performed to determine the amino acid composition ratio (AACR) in the disorder region per order region (Appendix C).

Integrated results of ELN gene for all six species, robust the primary result of analysis in Homo sapiens ELN genes for VP amino acid composition ratio with

correction for Glycine ratio **(figure 2.6)**. AACR for VP indicates that the presence of this sequence in disordered regions is 2.5 times more than ordered regions in the average of all six species. These results show the key role of proline in disordered regions as expected with more emphasis on the VP and VPG sequence as unique markers to identify disordered regions.

Figure 2.6

AACR for V, P, G, VP, PG, VPG, and VPGXG



Note. AACR for V, P, G, VP, PG, VPG, and VPGXG represented order pre disorder regions. The highest AACR is for VP and VPG and the number of prolines in disordered regions is 1.39 times more than ordered regions.

Analyzing the guest residue in the VPGXG sequence in these six species is another notable feature of this computational analysis that is informative. In the VPGXG sequence between these six species, five amino acids of Alanine, Phenylalanine, Leucine,

Threonine, and Valine were detected as a guest residue, with the number of each provided in **table 2.1**.

VPGXG sequences include 8.47 percent of the total length of these six species on average. This percentage is substantially below average for *Mus musculus*, with only 5.35 percent of the length. With the exception of *Mus musculus*, the most frequent guest residues are Valine, Alanine, and Lucien respectively. All three of these most common guest residues are non-polar and have a hydrophobic side chain. Threonine is the only polar uncharged guest residue, and Phenylalanine is likewise a non-polar with a hydrophobic side chain.

Table 2.1

The Number of Specific Amino Acids Which Substitute as a Guest Residue in the VPGXG Sequence Was Counted for Six Species

Sequence	NCBI Reference Sequence	VPGAG	VPGFG	VPGLG	VPGTG	VPGVG	Total	number of order region	Length	VPGXG percentage from total length
Human	P155024	3	1	2	1	9	16	3	786	10.18%
Pantroglodytes	PNI925061	2	1	2	1	8	15	3	802	9.35%
Panpaniscus	XP0348072831	3	1	2	1	7	14	3	758	9.23%
Pongoabelii	PNJ857951	2	1	3	0	6	12	2	714	8.40%
Gorilla	XP_030868570.1	3	1	1	0	7	12	3	722	8.31%
Musmusculus	XP0112391571	2	5*	0	0	1	8	0	747	5.35%
								Average	754.83	8.47%

Note. The number of specific amino acids which substitute as a guest residue in VPGXG sequence was counted for six species. *All VPGFG sequences in Musmusculus are presented in the same region which is highly ordered.

Except for five VPGFG sequences, all other VPGXG sequences are in the disorder region, however the kind of guest residue can affect the level of disorder, as detailed in Chapter 3 and depicted in the heatmap. According to the heatmap in Chapter 3, sequences having a lower C-score are more likely to exhibit disorder structure. For sequences featuring valine, Alanine, or Leucine as a guest residue, the average C-score is -1.44, -1.54, and -1.28, respectively. These averages are even lower than the average C-score for other sequences. As a result, having valine, Alanine, or Leucine as a guest residue raises the chance of disordered regions.

Chapter 3

Statistical Study on Intrinsically Disordered Structure of ELP

Effect of Amino Acids Hydrophobicity on Intrinsically Disordered Structure of ELP

In terms of investigation about the structure of proteins, considering the effect of hydrophobicity on folding is valuable. Specifically, protein polymers such as ELPs occupy a niche that is midway between synthetic polymers and IDPs. IDPs owe their disorder to repetitive, low complexity sequences of limited hydrophobicity.[1]

Most importantly, an amino acids hydrophobicity has been shown to affect ELPs phase transition temperature. Urry utilized chemically produced ELPs to show that their Tt was related to the guest residue's mean hydrophobicity. He showed that the mean hydrophobicity of the guest residue composition impacted the critical transition temperature (Ttc). But later, Ashutosh Chilkoti, hypothesize that the Tt is not simply depend upon the mean hydrophobicity of the guest residue composition as previously proposed but is actually determined by the specific sequence of the guest residues and the difference in hydrophobicity between the guest residues. This hypothesis is currently under investigation. [5]

Based on the above, we decided to consider hydrophobicity as an influential parameter in the ELPs structure (Appendix F). [73,74]

Effect of Monomeric Unit Size And Guest Residue's Amino Acid on Disordered Structure of ELP

Valuable applications of ELP in various fields doubles the importance of its design to regulate physical properties. For this purpose, various sequences of ELP and evaluation of their physical properties were determined to generate helpful data that can lead to a clear understanding of the relationship between amino acid sequences and physical properties. This understanding is a new step towards achieving the goal of tuning an ELP of interest with specific physical properties.

According to the ELP monomeric unit, VPGXG, the number of amino acids per unit, guest residue's amino acid (X), and the order of amino acids in the unit are three main variables in tuning the structure of ELP. Based on these variables, 220 different sequences of ELP were designed with the priority of maintaining the structure of the VPGXG unit as much as possible.

To study these three variables, sequences of (VPGX)7, (VPGXG)6, (VPGXAG)5, (VPGAGXG)4, and (VPGAGAXG)4 were designed to determine the effect of the number of amino acids in a monomeric unit. Then twenty essential amino acids were substituted for X to consider their effect as guest residues. For example 20 essential amino acids are replaced with X in (VPGXG)6 to generate 20 different sequences with 30 amino acids length. Worth mentioning that, Alanine and Glycine are selected to expand the size of the monomeric unit from 5 to 8 amino acids. In order to eliminate the effect of hydrophilicity and electric charge changes.

C-score As an Indicator to Study the Level of Ordered/Disordered Structure

For an amino acid sequence, I-TASSER simulations generate a large ensemble of structural conformations, called decoys. To select the final models, I-TASSER uses the SPICKER program to cluster all the decoys based on the pairwise structure similarity and reports up to five models which correspond to the five largest structure clusters. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of higher value signifies a model with high confidence and vice-versa. Specifically in terms of intrinsically disordered structure, due to the lack of ordered structure I-TASSER can not predict confidence structure therefore C-score will be low.

Heatmap

The effect of the number of amino acids in a monomeric unit and guest residue composition at the same time on C-score could be an efficient way for initial classification of ELP sequences based on C-score. 20 different essential amino acids were replaced with X in (VPGAGXG)4 sequence, then the 20 generated sequences with 28 amino acids length were used to calculate the C-score. As it is shown in the heatmap, Most of the C-score for these monomeric units of 7 amino acids are pretty low. In the same way 20 sequences were generated by replacing 20 essential amino acids with X in (VPGAGAXG)4. C-score results for monomeric units of 8 amino acids, 32 amino acids length, are low as well.

As shown in the heatmap below, the C-score in ELP sequences with the (VPGAGXG)4 and (VPGAGAXG)4 are very low. That may indicate a high level of disordered structure.

(VPGX)7 and (VPGXG)6 sequences do not have low scores like seven and eight amino acid monomeric units but they still have low C-score that is not enough for reliable prediction. Interestingly, the C-score for (VPGAXG)5 sequences are a bit higher in the range of -0.17 to - 1.35. This range indicated interesting properties of structure that although it is disordered but not much. This could be a window of opportunity to find a repetitive sequence with tunable structure properties.

According to these findings, analyzes focused more on six amino acid monomer units with different arrangements. Therefore, (VPGXAG)5, (VPGGXG)5, (VPGXGG)5, (XPGAGG)5, (VXGAGG)5, (VPXAGG)5 sequences were generated and C-score calculated for them. C-score range of six amino acid monomeric units are -2.22 to 0.79 and surprisingly both maximum and minimum come from the VXGAGG unit which is emphasized on the rule of proline in the monomeric unit. (VPGXGG)5 sequences have the second-largest range (-1.98 to 0.51). This unit has just one more glycine in comparison to the typical ELP monomeric unit of VPGXG but a significant change in the C-score range. That may cause more rigidity added to this unit by glycine.

In comparison to VPGXG, adding just one alanine amino acid in the critical position of fourth substantially enhanced the average C-score. VPGXG had an average C-score of -1.62 for twenty distinct guest residue amino acids, whereas VPGAXG had average C-score of -0.7. The presence of alanine, a non-polar amino acid with the hydrophobicity of

-0.3, between glycine and the guest residue in this monomeric unit, renders the structure of this polypeptide more predictable.

Figure 3.1

C-Score Heatmap for Monomeric Units vs Guest Residue Amino Acids

Monomeric units

Amino acids	Average Hydrophobicity	$(VPGX)_7$	$(VPGXG)_6$	$(VPGAXG)_5$	$(VPGXAG)_5$	$(VPGGXG)_5$	$(VPGXGG)_5$	$(XPGAGG)_5$	$(VXGAGG)_5$	$(VPXAGG)_5$	$(VPGAGXG)_4$	$(VPGAGAXG)_4$	
P	1.5277	-2.03	-1.79	-0.22	-0.33	-0.44	-0.36	-1.13	-1.34	-1.34	-2.29	-2.49	Special
D	1.5143	-1.81	-1.71	-0.19	-0.8	-0.62	-1.71	-1.51	-2.08	-1.53	-2.36	-1.95	negative charged
S	1.4767	-1.9	-1.76	-0.42	-0.81	-0.71	-1.7	-1.59	-1.2	-1.44	-2.72	-2.42	Polar
K	1.2293	-1.68	-1.97	-0.41	-0.7	-0.57	-1.81	-1.63	-2.22	-1.4	-3.05	-1.76	Posetive Charged
Q	1.1617	-1.66	-1.74	-0.4	-0.67	-0.66	-1.27	-1.28	-1.05	-1.36	-1.91	-1.72	Polar
E	1.006	-1.93	-1.58	-0.17	-0.79	-0.6	-1.98	-1.54	-0.51	-1.44	-2.79	-1.96	negative charged
H	0.8037	-1.75	-1.63	-0.8	-0.78	-1.86	-1.48	-1.49	-0.76	-1.14	-2.16	-1.95	Posetive Charged
A	0.76	-2.32	-1.59	-0.86	-0.86	-1.45	-1.34	-1.62	-0.87	-1.69	-2.28	-2.07	Non-polar
F	0.5933	-1.46	-1.42	-1.35	-0.66	-1.52	-1.42	-1.34	-0.59	-1.55	-2.41	-1.97	Non-polar
T	0.462	-1.8	-1.8	-0.55	-0.76	-1.86	-1.54	-1.37	-1.2	-1.7	-3.02	-2.35	Polar
R	0.167	-1.49	-1.72	-0.4	-0.65	-0.99	-1.49	-1.56	-0.5	-1.55	-3.28	-1.73	Posetive Charged
L	-0.027	-1.93	-1.58	-0.17	-0.23	-1.31	0.51	-1.63	-1.1	-1.43	-2.9	-2.35	Non-polar
N	-0.304	-1.51	-1.73	-0.41	-1.44	-1.9	-1.66	-1.48	-1.18	-1.29	-3.42	-1.88	Polar
C	-0.546	-0.59	-1.66	-0.75	-0.56	-1.37	-1.36	-1.18	-1.01	-1.15	-2.9	-2.29	Special
W	-0.595	-1.66	-1.1	-1.12	-0.61	-1.26	-1.01	-1.17	-0.1	-1.33	-1.95	-1.23	Non-polar
I	-0.759	-1.57	-1.13	-1.14	-0.4	-1.65	-0.46	-1.42	0.79	-1.63	-2.89	-2.5	Non-polar
V	-0.877	-1.74	-2.02	-1.35	-0.46	-1.63	-0.62	-1.34	0.22	-1.64	-2.91	-2.37	Non-polar
M	-0.922	-1.63	-1.22	-0.88	-0.29	-1.61	-0.33	-0.69	-0.55	-1.62	-2.57	-2.49	Non-polar
Y	-1.02	-1.28	-1.53	-0.97	-0.58	-1.48	-1.53	-1.54	-1.07	-1.64	-1.78	-1.84	Non-polar
G	-1.096	-1.21	-1.69	-1.34	-1.45	-1.33	-1.33	-0.85	-0.5	-1.34	-3.09	-1.65	Special

Note. This heatmap magnitude of a C-score as color in two dimensions. Scores are classified from high, green, to low, red, by color intensity. For each column 20 amino acids were replaced with X in monomeric units and the C-score was calculated for each sequence. As an example C-score for (VPGVG)6 iS -2.02. Average Hydrophobicity was calculated for each amino acid based on three mentioned hypotheses.

Overall, the heatmap displays unpredictably monomeric unit sequences of seven and eight amino acids. The monomeric unit with four amino acids was almost similar to the original five amino acids monomeric unit in terms of C-score.

Because of the greater C-score range, the six amino acid monomeric unit structures were studied further, and the following conclusions are notable.

1. Substitution of proline with other amino acids makes the structure of these polypeptides more predictable.
2. Replacing the first valine with other amino acids increased the chances of disordered structure.
3. Most of the VPGAXG and VPGXAG sequences were promising within a high range of C-score.

Chapter 4

Build Up a Recombinant ELP Manufacturing Method

The novel sequences of ELP[AV-60] and ELP[V-60] (Appendix D) were developed and recombinantly synthesized to set up a procedure for the synthesis of various ELP sequences and investigate their disordered structure.

The transition temperatures for these two sequences were Tt = 34°C for ELP[V-60] and Tt = 37°C for ELP[AV-60]. These transition temperatures were within the range of human body temperature, making these ELP sequences suitable candidates for in vivo studies.

Initial Constructs

To start the process of producing two polypeptides of interest, pMAL-c5X-ELP[V-60], pMAL-c5X-ELP[AV-60], and pNIC28-Bsa4 plasmids were ordered from Addgene company. Their plasmid maps are shown in Appendix E.

The ELP[V-60] is an abbreviation for the 60 repeats of [GVGVP] monomeric units that were inserted into pMAL-c5X plasmid. Also, ELP[AV-60] was named for the 30 repeats of [GAGVPGVGVP] monomeric units that were inserted into the same plasmid. Furthermore, pNIC28-Bsa4 was an interesting vector containing a Bsa4 site that allowed us to use the golden gate cloning strategy to clone target genes and produce two ELP sequences.

Primers Design for Cloning of ELP[V-60] and ELP[AV-60] with KCTS Into pNIC28-Bsa4 by Golden Gate Strategy

pMAL-c5X-ELP[V-60], pMAL-c5X-ELP[AV-60], and pNIC28-Bsa4 plasmids were received frozen in the *E. coli* DH5-Alpha strain. To start the process of cloning, each pMAL-c5X-ELP[V-60] and pMAL-c5X-ELP[AV-60] samples were cultured on LB agar plates with Ampicillin antibiotic (50 μg/mL) overnight. Also, pNIC28-Bsa4 plasmids in the *E. coli* DH5-Alpha strain were cultured on LB agar plates with Kanamycin antibiotic (25 μg/mL) overnight.

The next day, a single colony of each sample was cultured in LB broth with the same concentration of antibiotic to achieve enough cell density for plasmid purification.

Cell culture mediums of ELP[V-60], ELP[AV-60], and pNIC28-Bsa4 were centrifuged to separate the cells from the medium. Then, the ZR Plasmid Miniprep-Classic kit was used to lyse the cells and purify the pMAL-c5X-ELP[V-60], pMAL-c5X-ELP[AV-60], and pNIC28-Bsa4 plasmids. At the end of this step, Purified plasmids are stored at -20°C. 62.45 ug/ml , 73.9 ug/ml and 58.25 ug/ml were the concentration of pMAL-c5X-ELP[V-60], pMAL-c5X-ELP[AV-60] and pNIC28-Bsa4 respectively by Qubit 2.0 Fluorometer.

Four primers were designed for cloning of ELP[V-60] and ELP[AV-60] genes by PCR. Due to the fact that we intend to use these two polypeptides for tissue engineering applications, the special sequence of KCTS (Lysine, Cysteine, Threonine, Serine) was included as a flanking side in the primers to generate this sequence in the target gene. These flanking sides were responsible for polymerization. Cysteine residue was

responsible for chain extension via disulfide bond formation. UV or visible light induced free cysteines to make disulfide bonds. Moreover, the Bsa site was also designed in the primers to clone the target gene in the Bsa site of pNIC28-Bsa4. Eventually, PCR amplified ELP[V-60] and ELP[AV-60] genes with KCTS codon (AAGTGCACATCA) and Bsa sites (GGTCTC).

V60_BSA_Forward:

GCAGAAGGTCTCGCCATGAAGTGCACATCATCTTCCGGTGGCGTTGGGGTAC
C

V60_BSA_Reverse:

TTCTGCGGTCTCGTACTTTATGATGTGCACTTTCCCGAGCCGGGGACTCCTAC
C

AV60_BSA_Forward:

GCAGAAGGTCTCGCCATGAAGTGCACATCATCTTCCGGTGGAGCTGGTGTGC
C

AV60_BSA_Reverse:

TTCTGCGGTCTCGTACTTTATGATGTGCACTTTCCCGAGCCCGGCACCCCGAC
G

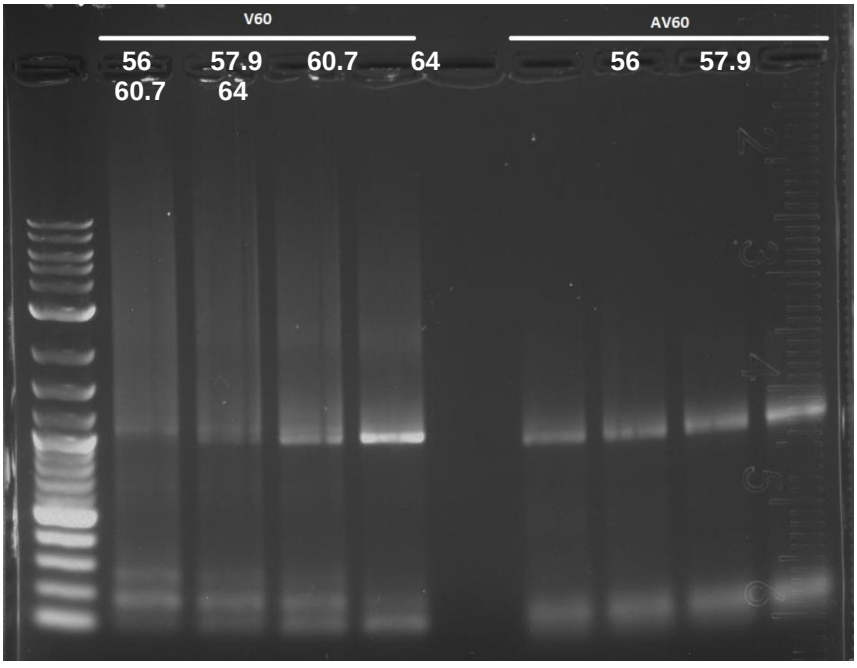
Afterward, PCR was performed by 5 ng/ml ELP[AV-60] or ELP[V-60] plasmids and 10uM forward and reverse primers in four different temperatures (56°C, 57.9°C, 60.7°C, 64°C) to find an optimum ligation temperature. Then the PCR products ran in 1% agarose

gel to find which ligation temperature works. As demonstrated in **(figure 4.1)**, ligation was easier at 64°C.

Then, the PCR products of 64°C for ELP[V-60] and ELP[AV-60] were purified by the ZYMO mini-prep kit and the final concentrations were 249 ng/ul for ELP[AV-60] and 510 ng/ul for ELP[V-60]. These two PCR products were cloned into the pNIC28-Bsa4 plasmid by the golden gate assembly protocol which is a high throughput process. The Bsa site in pNIC28-Bsa4 did not allow the digested fragment of itself to attach again into the Bsa site. Therefore, we only had a recombinant pNIC28-Bsa4 with ELP[AV-60] or ELP[V-60] gene at the end of the golden gate assembly PCR which was a perfect process for precise cloning.

Figure 4.1

1% Agarose Gel of ELP[V-60] and ELP[AV-60] PCR Products



Note. This is a 1% agarose gel of [V-60] ELP and [AV-60] ELP PCR products. The first lane on the left is 2-Log DNA Ladder. The second lane to fifth lane are [V-60] ELP PCR products in different temperatures and seventh to tenth lane are [AV-60] ELP PCR products in different temperatures. There is a band for each lane around 1000 kb that represents the amplified genes by PCR. The bands for 64C are strongest for both [V-60] and [AV-60].

Transformation and Selection of Host Cell Line

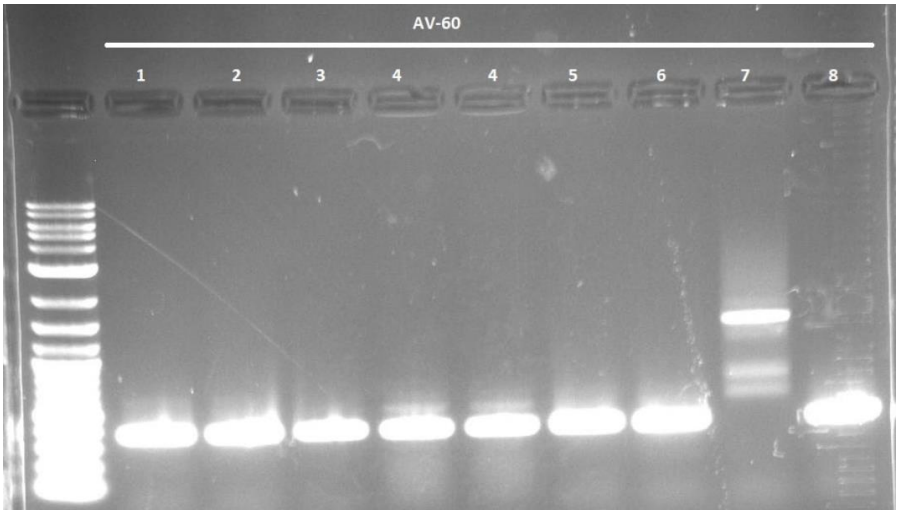
Both ELP[AV-60] and ELP[V-60] have a proline in their monomeric units. As a result, their sequence contains several rare CCC proline codons, and the majority of expression hosts lack tRNAs to supply this codon. To fulfill this requirement, research was conducted to find the optimum strain, and ultimately, the *E. coli* Rosetta strain was identified.

Rosetta host strains are BL21 derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*. These strains supply tRNAs for AGG, AGA, AUA, CUA, CCC, GGA codons on a compatible chloramphenicol-resistant plasmid. Thus the Rosetta strains provide for "universal" translation which is otherwise limited by the codon usage of *E. coli*. The tRNA genes are driven by their native promoters. In Rosetta(DE3)pLysS, the rare tRNA genes are present on the same plasmids that carry the T7 lysozyme gene. Therefore, the Rosetta competent cells were selected to transform by pNIC28-Bsa4-ELP[AV-60] and pNIC28-Bsa4-ELP[V-60].

After insertion of the pNIC28-Bsa4-ELP[AV-60] into Rosetta cells, they were cultured on agar plates with kanamycin and plain agar plates for positive selection of transformed Rosetta cells by pNIC28-Bsa4-ELP[AV-60] vectors (appendix E). Colony PCR was used to confirm the presence of the right gene of interest in pNIC28-Bsa4 for 8 single colonies. As shown in **(figure 4.2)**. colony number 7 is the colony with correct band size of ELP[AV-60] genes between 1000 and 1200 kb. Colony number 7 glycerol stock was prepared and stored at -20°C. Moreover, another agarose gel was performed to compare the right size of intact pNIC28-Bsa4 and ELP[AV-60] gene at figure 4.3.

Figure 4.2

1% Agarose Gel of Colony PCR for Rosetta pNIC28-Bsa4 ELP[AV-60]

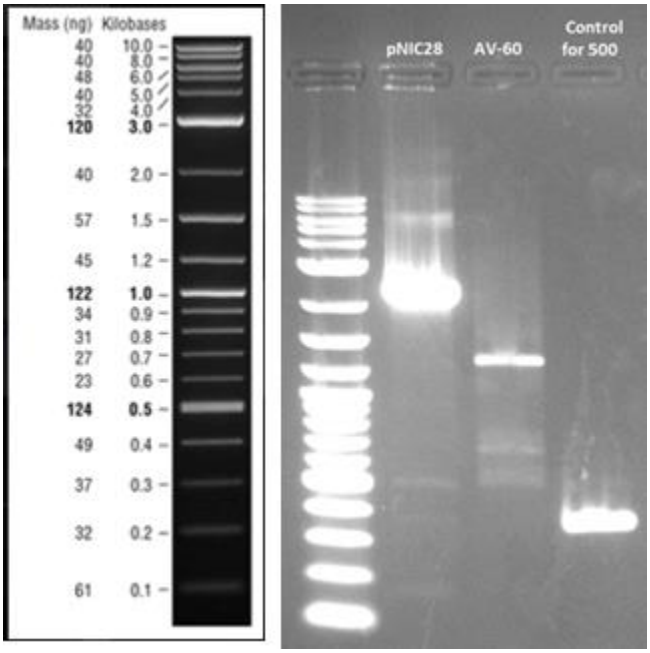


Note. 1% agarose gel of colony PCR for Rosetta pNIC28-Bsa4 [AV-60]ELP for 8 single colonies. Just colony number 7 has a band between 1000 kb and 1200 kb that is matched with [AV-60]ELP size ~ 1017 kb.

Colony number 7 was grown up in LB broth and then the pNIC28-Bsa4 ELP[AV-60] was purified by miniprep to perform a sequencing test. Sequencing results verified the presence of the exact ELP[AV-60] gene with proper flanking sites in the pNIC28-Bsa4.

Figure 4.3

2-Log DNA Ladder, and a 1% Agarose Gel of Colony PCR Result



Note. The left photo is 2-Log DNA Ladder, and the right photo is a 1% agarose gel of colony PCR result. First lane in lef is ladder, second lane in intact pNIC28-Bsa4, third is [AV-60] ELP and the last lane is a 500 kb control.

Expression of ELP[AV-60] in *E. coli* Rosetta Host Strain

Glycerol stock of Rosetta transformed by pNIC28-Bsa4 ELP[AV-60] was cultured on an kanamycin plate to achieve a single colony, also those cells were cultured on a plain plate of agar as a positive control. Just the transformed Rosetta cells had colonies on the kanamycin plate overnight. A single colony was picked to culture in 5ml LB broth with kanamycin. Those 5 ml were subcultured in 50 ml LB broth with kanamycin. The optical density of cultures was checked to find an optimum time of induction and temperature of growth for transformed Rosetta cells (table 4.1).

Furthermore, the concentration of isopropyl-D-1-thiogalactopyranoside (IPTG) was another parameter that was adjusted for ELP[AV-60] expression. Cells were induced at maximal OD (0.8-1) to produce a high cell density as well as a high level of ELP[AV-60] expression.

Transformed Rosetta cells grew faster at room temperature or less, as presented in table 4.1, which might indicate the existence of ELP in the cells that are expressed through leaky expression. In high temperatures, expressed ELP tends to aggregate, which might be the primary reason for cell growth inhibition. In addition, decreased IPTG concentrations resulted in greater ELP expression.

Table 4.1

Design of Experiments (DOE) to Find the Optimum Temperature and Concentration of IPTG

S.No	Sample	Temperature (°C)	IPTG (mM)	OD before IPTG	OD overnight
1	ELP	18	0.1	0.83	0.679
2	ELP	18	1	0.83	0.697
3	ELP	RT	0.1	0.83	0.687
4	ELP	RT	1	0.83	0.694
5	ELP	30	0.1	0.83	0.630
6	ELP	30	1	0.83	0.637
7	ELP	37	0.1	0.83	0.635
8	ELP	37	1	0.83	0.608

Note. Designed experiment to find the optimum temperature and concentration of IPTG.

ELP[AV-60] Purification by Inverse Transition Cycling

As stated earlier, the Inverse Transition Cycling (ITC) purification procedure for ELP is inexpensive and straightforward. The transition temperature, or Tt, is the most crucial factor in this process. As a result, developing an ELP with Tt in the body

temperature range is quite desirable. ELP[AV-60] is a unique polypeptide for in vivo study since it has a Tt of 37°C and photocrosslinking characteristics. The following were the processes in purifying ELP[AV-60].

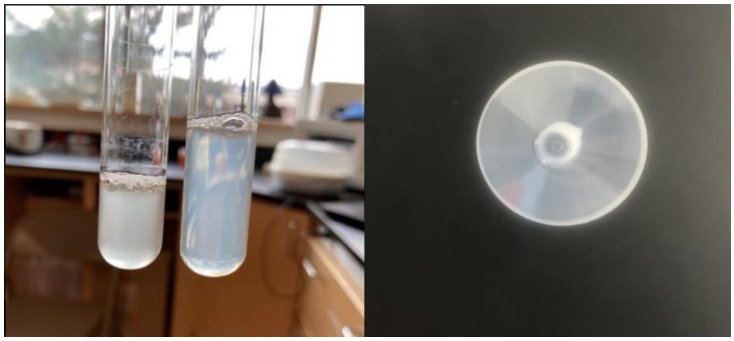
1. To get the cell pellet, the cell culture mixture centrifuged at 4°C for 10 min at 3000 x g.
2. The supernatant should be discarded, and the cell pellets should be resuspended in Phosphate-Buffered Saline (PBS). (If the cell pellet was insufficient, collect several sample pellets in a 50 ml falcon tube.)
3. The resuspended cell pellet was sonicated for 35 rounds and then rested on ice for 2 minutes for a total of 3 repetitions at an output power of 80 W while the cell suspension was kept on ice. (Because ELP is temperature sensitive, care should be taken not to overheat it by long sonication.)
4. The lysate was transferred to a 50 mL centrifuge tube with a round bottom.
5. For each liter of culture, we added 2 ml of 10% (w/v) polyethyleneimine (PEI) and mixed thoroughly.
6. Samples were centrifuged at 4 °C for 10 min at 16,000 x g.
7. Afterwards, the particle was removed and transferred the supernatant to a clean 50 mL round bottom centrifuge tube.
8. The samples were heated at 42°C for 5 minutes to denature protein impurities, then it cooled to 4°C or ice until the ELP was fully resolubilized (minimum of 4 hours). Accordingly, the samples were centrifuged at 16,000 x g for 10 minutes at 4 °C and the particle removed and transferred the supernatant to a clean 50 mL round bottom centrifuge tube.

9. With the addition of crystalline NaCl (1 M), the ELP transition may be induced.
10. The samples were centrifuged at 16,000 x g for 10 minutes at room temperature (RT) (hot spin). Worth of note: An ELP pellet should be seen, the size of which is determined by the ELP's expression yield. As the ELP pellet is purified and impurities are eliminated, it is expected to become more colorless **(figure 4.4)**.
11. Then, the supernatant was removed and the pellet was mixed in 1-5 mL of PBS buffer. The tubes were set to spin at 4°C to resuspend the pellet.
12. 1 ml aliquots of the resuspended ELP solution was transferred into clean 1.5-ml microcentrifuge tubes.
13. Centrifugation at 16,000 x g for 10 minutes at 4 °C (cold spin) was performed. It is expected to observe a tiny particle of insoluble contaminants visible.
14. The pellet was removed and supernatant was transferred to a clean tube.
15. To eliminate excess salt from the filtered ELP solution, the sample was dialyzed against ddH2O at 4 °C.
16. In the following the first and second rounds of purification by ITC, ELP samples were collected to run SDS-PAGE **(figure 4.5)**.

More steps of cold and hot spin could be performed to achieve more purity. Also, If desired, lyophilization could be done for the ELP for long-term storage by freezing it for 30 minutes at -80 °C before lyophilizing.

Figure 4.4

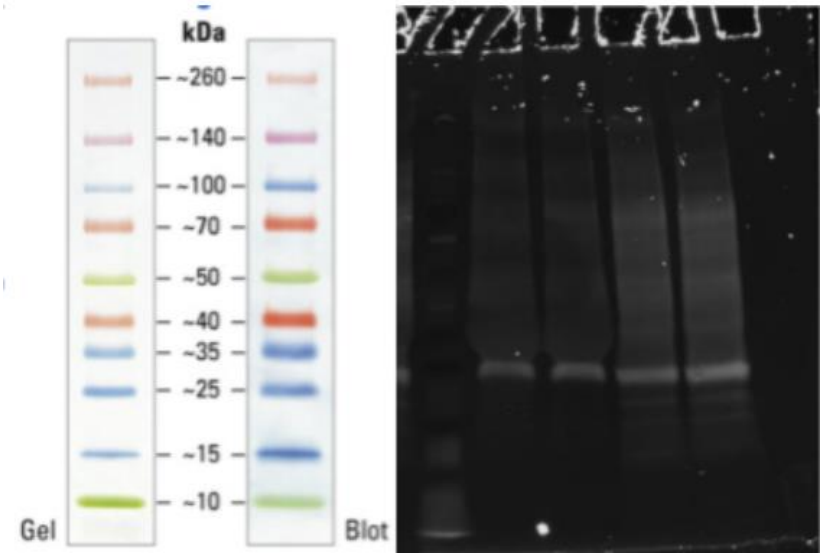
ELP Conservation and Pellet Inverse



Note. Left photo presents ELP in two different temperatures. Left glass tube is ELP after 45 C heat that is more opaque and the right one is ELP at room temperature. The photo in the right is the ELP pellet after hot spin centrifuge at 50ml conical tube.

Figure 4.5

SDS-PAGE Result of ELP[AV-60] After First and Second Round of ITC



Note. The first lane is the ladder, second and third lane are [AV-60]ELP after two rounds of ITC purification. Fourth and fifth lanes are after the first round of ITC purification.

[AV-60]ELP with all flanking sides has a 27.74 KDa molecular weight that is between 25 and 35 markers.

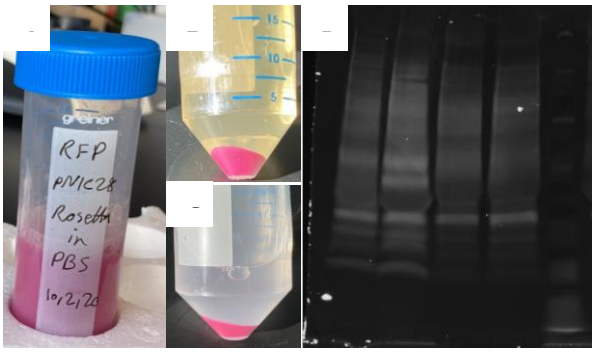
Although the major benefit of producing ELP is the simple and inexpensive purification procedure, the inherently disordered structure of ELP may cause irreversible polypeptide aggregation and impair the purification process. As a result, temperature and butter concentration optimization is critical in the purifying process.

Red Fluorescent Protein as an Indicator for Debugging

Red Fluorescent Protein (RFP) was cloned into the same pNIC28-Bsa4 vector at the Bsa site because its particular characteristic, which produces fluorescent red light, helps substantially to follow the target protein in expression and different phases of purification. As a result, we were able to debug the expression process more efficiently by expressing RFP concurrently with ELP.

Figure 4.6

RFP Pellets in Cell Culture and PBS



Note. A. Depict cell culture with RFP. Expression of RFP is obvious by the color change of the media to red. B. there is cell pellet of RFP and cell culture medium after sonicate and centrifuge. C.The pellet of RFP in PBS buffer. D. Lane1 and 2:RFP sample after cell lysis (RFP MW= 25KDa). Lane 3 and 4: [AV-60]ELP. Lane5: Ladder

Discussion

The primary goal of this thesis project is to understand which of ELP's inherent protein sequence features contribute to its unusual physical properties, such as reversible phase transition in response to temperature change. Previous studies have shown that temperature affects the folding and subsequent solubility of ELP in solution. When increasing the temperature over the lower critical solution temperature (LCST), ELP polypeptides tend to aggregate. This aggregation can be seen as a precipitate, cloudy structure in solution, which then returns to its solution state upon lowering the temperature. After observing that the unique amino acid sequence of ELP contains disordered regions we were motivated to look into the origin of this phenomenon, investigating ways to develop additional novel ELP with specific desirable properties.

To achieve this objective, this thesis endeavored to investigate the structure and underlying principles governing the unique properties of the ELP protein utilizing the following 3 main approaches:

1. The naturally occurring version of the synthetic peptide ELP is the ELN gene which, in its native state, maintains interesting and unique elastic properties as part of the extracellular matrix. We performed extensive comparative sequence analyses of ELN to represent a natural sample in order to identify and investigate specific disorder regions which may play key roles in maintaining ELP’s disordered state..

2. Targeted variants in the monomeric amino acid sequences were used to generate 220 distinct *in silico* ELP sequences. The bioinformatics tools I-TASSER and PONDER were applied to investigate the potential presence of a significant relationship between

monomeric unit structures and the structure of disorder regions. In other words, the effect of order and number of amino acids in monomeric units on disorder regions were studied.

3. Finally, a laboratory-scale approach for recombinant manufacture and purification of ELP's was developed and assessed.

In the disorder areas of the human ELN gene, the presence of VPGXG sequence, which is the main component of ELP, was clearly found. In the disorder regions, the presence of the VP is quite important. The presence of VPGXG in disorder regions, particularly the VP sequence, was later validated in the other five species with the most alignment similarity to the human ELN gene. These six ELN genes also support the idea that VP is a conserved sequence in disorder regions. As a result, the VP sequence might be applied to detect disorder regions in any amino acid sequence.

Different ELP sequences were generated based on preliminary ELP[AV-60] and ELP[V-60] results obtained from the I-TASSER and PONDR tools.

The followings are the three primary parameters that were considered while generating new sequences:

1. In the structure of (VPGX)7, (VPGXG)6, (VPGXAG)5, (VPGXAG)5, (VPGGXG)5, (VPGXGG)5, (XPGAGG)5, (VXGAGG)5, (VPXAGG)5, (VPGAGXG)4, (VPGAGAXG)4, the X is substituted with twenty different amino acids as a guest residue.

2. Changing the number of amino acids in the monomeric units to 4,5,6,7, and 8.

3. Changing the number of repeats per monomeric unit sequence to keep the sequence length in the range of ∼30 amino acids.

C-sore was used to arrange various sequences based on the degree of disorder in the structures. C-score is in [-5,2] range and C-score > -1.5 indicates a model of correct global topology. As a result, sequences with a C-score of less than -1.5 have a high level of disorder structure, which I-TASSER cannot accurately predict using global topology. On the other hand, sequences with a higher C-score in the [-1.5,2] range are likely to have some disorder structure but still be predictable by I-TASSER.

All sequences in the heatmap **(Figure 3.1)** have a C-score based on monomeric unit and guest residue. The C-score for sequences containing the (VPGXAG)5 monomeric unit is in the promising range, with an average of (-0.6915) This value indicates that sequences containing these monomeric units have the potential to be used to design sequences with tunable physical properties, such as LCST. These C-score ranges are not too low, indicating a high level of disorder structure and difficulty in predicting topology, nor are they too high, indicating that the structure is entirely organized and the polypeptide is no longer disorder. In this study, (VPGXAG)5 monomeric units with various guest residue amino acids (X) are recommended as a new member of the ELP family to be targeted to enable us to manufacture ELP with specific desirable physical properties.

ELPs have repetitive sequences that make cloning and expression challenging. Furthermore, their sequences contain several rare codons such as AGG and GGA, which ends in reducing the number of hosts that can support these codons. The Golden Gate Strategy was used to ensure an accurate cloning procedure for ELPs with high repetition

of short sequences to address the obstacle of cloning for ELPs with high repetition of short sequences. To overcome the challenge of expression for ELP sequence with rare codons, the Rosetta (DE3) cell line of E.coli was selected. Rosetta host strains are BL21 variants designed to promote the production of eukaryotic proteins that contain codons rarely utilized in E. coli. On a compatible chloramphenicol-resistant plasmid, these strains provide tRNAs for the AGG, AGA, AUA, CUA, CCC, and GGA codons. As a result, the Rosetta strains enable "universal" translation which is otherwise limited by the codon usage of E. coli.

Following the appropriate expression of ELPs in Rosetta, the cells were broken down using ultrasound methods to begin the purification process. Ultrasound paired with temperature control of cell debris on ice inhibits ELPs from irreversible aggregation. The ITC was a simple and inexpensive way to purify ELP back then. Based on the ELP sequence and the LCST results, the ITC process can be adjusted.

Lastly, successful models for ELP synthesis include the golden gate strategy for cloning, the Rosetta cell line as an expression host, and ITC purifications.

Future Directions

The results of this study indicate potentially valuable contributions to the field of Intrinsically Disordered Proteins including a further understanding of how amino acid sequences in ELP can be modulated to directly affect IDPs as well as developing experimental techniques that are necessary to test these *in silico* predictions. Using these advances in our basic understanding of IDPs in ELP, results of this study can be used as a starting point for further experimental testing of amino acid sequence variants in ELP. For example, using PONDR, it is feasible to first identify the Intrinsically Disordered regions of the desired target IDP and then, using the same statistical model, assemble a polypeptide from short and repetitive sequences. Then, using the pNIC28 vector and Rosetta cells and experimental techniques developed in this thesis, synthesize the polypeptide to simulate the physical properties of the target IPD. The general potential structure and degree of disorder can be predicted using I-TASSER, and finally ELP proteins with a range of different intrinsic disorders can be generated and finally evaluated for tunable properties of both elasticity and temperature sensitive phase transitioning.

References

1. Roberts, S., Dzuricky, M. & Chilkoti, A. Elastin-like polypeptides as models of intrinsically disordered proteins. FEBS Lett. 589, 2477–2486 (2015).

2. Urry, D. W. et al. Hydrophobicity scale for proteins based on inverse temperature transitions. Biopolymers 32, 1243–1250 (1992).

3. Foster, J. A., Bruenger, E., Gray, W. R. & Sandberg, L. B. Isolation and amino acid sequences of tropoelastin peptides. J. Biol. Chem. 248, 2876–2879 (1973).

4. Urry, D. W. et al. The synthetic polypentapeptide of elastin coacervates and forms filamentous aggregates. Biochim. Biophys. Acta 371, 597–602 (1974).

5. McDaniel, J. R., Christopher Radford, D. & Chilkoti, A. A Unified Model for De Novo Design of Elastin-like Polypeptides with Tunable Inverse Transition Temperatures. Biomacromolecules vol. 14 2866–2872 (2013).

6. Cho, Y. et al. Effects of Hofmeister anions on the phase transition temperature of elastin-like polypeptides. J. Phys. Chem. B 112, 13765–13771 (2008).

7. Huang, L. et al. Generation of Synthetic Elastin-Mimetic Small Diameter Fibers and Fiber Networks. Macromolecules vol. 33 2989–2997 (2000).

8. Brennan, M. J., Kilbride, B. F., Wilker, J. J. & Liu, J. C. A bioinspired elastin-based protein for a cytocompatible underwater adhesive. Biomaterials 124, 116–125 (2017).

9. MacEwan, S. R. & Chilkoti, A. Elastin-like polypeptides: Biomedical applications of tunable biopolymers. Biopolymers vol. 94 60–77 (2010).

10. Meyer, D. E. & Chilkoti, A. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. Nature Biotechnology vol. 17 1112–1115 (1999).

11. Christensen, T., Trabbic-Carlson, K., Liu, W. & Chilkoti, A. Purification of recombinant proteins from Escherichia coli at low expression levels by inverse transition cycling. Analytical Biochemistry vol. 360 166–168 (2007).

12. Ge, X. & Filipe, C. D. M. Simultaneous Phase Transition of ELP Tagged Molecules and Free ELP: An Efficient and Reversible Capture System. Biomacromolecules vol. 7 2475–2478 (2006).

13. Lim, D. W., Trabbic-Carlson, K., Andrew MacKay, J. & Chilkoti, A. Improved Non-chromatographic Purification of a Recombinant Protein by Cationic Elastin-like Polypeptides. Biomacromolecules vol. 8 1417–1424 (2007).

14. Banki, M. R., Feng, L. & Wood, D. W. Simple bioseparations using self-cleaving elastin-like polypeptide tags. Nature Methods vol. 2 659–662 (2005).

15. Floss, D. M. et al. Influence of elastin-like peptide fusions on the quantity and quality of a tobacco-derived human immunodeficiency virus-neutralizing antibody. Plant Biotechnology Journal vol. 7 899–913 (2009).

16. Kim, J.-Y., O'Malley, S., Mulchandani, A. & Chen, W. Genetically Engineered Elastin-Protein A Fusion as a Universal Platform for Homogeneous, Phase-separation Immunoassay. Analytical Chemistry vol. 77 2318–2322 (2005).

17. Stiborova, H., Kostal, J., Mulchandani, A. & Chen, W. One-step metal-affinity purification of histidine-tagged proteins by temperature-triggered precipitation. Biotechnology and Bioengineering vol. 82 605–611 (2003).

18. Massodi, I., Bidwell, G. L. & Raucher, D. Evaluation of cell penetrating peptides fused to elastin-like polypeptide for drug delivery. Journal of Controlled Release vol. 108 396–408 (2005).

19. Furgeson, D. Y., Dreher, M. R. & Chilkoti, A. Structural optimization of a 'smart' doxorubicin–polypeptide conjugate for thermally targeted delivery to solid tumors. Journal of Controlled Release vol. 110 362–369 (2006).

20. Dreher, M. R., Liu, W., Michelich, C. R., Dewhirst, M. W. & Chilkoti, A. Thermal Cycling Enhances the Accumulation of a Temperature-Sensitive Biopolymer in Solid Tumors. Cancer Research vol. 67 4418–4424 (2007).

21. Liu, W. et al. Tumor accumulation, degradation and pharmacokinetics of elastin-like polypeptides in nude mice. Journal of Controlled Release vol. 116 170–178 (2006).

22. Girotti, A. et al. Design and bioproduction of a recombinant multi(bio)functional elastin-like protein polymer containing cell adhesion sequences for tissue engineering purposes. Journal of Materials Science: Materials in Medicine vol. 15 479–484 (2004).

23. Betre, H. et al. Chondrocytic differentiation of human adipose-derived adult stem cells in elastin-like polypeptide. Biomaterials vol. 27 91–99 (2006).

24. Betre, H., Setton, L. A., Meyer, D. E. & Chilkoti, A. Characterization of a Genetically Engineered Elastin-like Polypeptide for Cartilaginous Tissue Repair. Biomacromolecules vol. 3 910–916 (2002).

25. Wu, X. et al. Alterations in Physical Cross-Linking Modulate Mechanical Properties of Two-Phase Protein Polymer Networks. Biomacromolecules vol. 6 3037–3044 (2005).

26. Wu, X., Sallach, R. E., Caves, J. M., Conticello, V. P. & Chaikof, E. L. Deformation responses of a physically cross-linked high molecular weight elastin-like protein polymer. Biomacromolecules 9, 1787–1794 (2008).

27. McMillan, R. A., Andrew McMillan, R., Caran, K. L., Apkarian, R. P. & Conticello, V. P. High-Resolution Topographic Imaging of Environmentally Responsive, Elastin-Mimetic Hydrogels. Macromolecules vol. 32 9067–9070 (1999).

28. Trabbic-Carlson, K., Setton, L. A. & Chilkoti, A. Swelling and Mechanical Behaviors of Chemically Cross-Linked Hydrogels of Elastin-like Polypeptides. Biomacromolecules vol. 4 572–580 (2003).

29. Lim, D. W., Nettles, D. L., Setton, L. A. & Chilkoti, A. Rapid Cross-Linking of Elastin-like Polypeptides with (Hydroxymethyl)phosphines in Aqueous Solution. Biomacromolecules vol. 8 1463–1470 (2007).

30. McHale, M. K., Setton, L. A. & Chilkoti, A. Synthesis and in vitro evaluation of enzymatically cross-linked elastin-like polypeptide gels for cartilaginous tissue repair. Tissue Eng. 11, 1768–1779 (2005).

31. Nowatzki, P. J. & Tirrell, D. A. Physical properties of artificial extracellular matrix protein films prepared by isocyanate crosslinking. Biomaterials 25, 1261–1267 (2004).

32. Richman, G. P., Tirrell, D. A. & Asthagiri, A. R. Quantitatively distinct requirements for signaling-competent cell spreading on engineered versus natural adhesion ligands. J. Control. Release 101, 3–12 (2005).

33. Liu, J. C., Heilshorn, S. C. & Tirrell, D. A. Comparative cell response to artificial extracellular matrix proteins containing the RGD and CS5 cell-binding domains. Biomacromolecules 5, 497–504 (2004).

34. Martínez-Osorio, H. et al. Genetically engineered elastin-like polymer as a substratum to culture cells from the ocular surface. Curr. Eye Res. 34, 48–56 (2009).

35. Taddese, S., Weiss, A. S., Jahreis, G., Neubert, R. H. H. & Schmelzer, C. E. H. In vitro degradation of human tropoelastin by MMP-12 and the generation of matrikines from domain 24. Matrix Biology vol. 28 84–91 (2009).

36. Alonso, M., Reboto, V., Guiscardo, L., San Martín, A. & Rodríguez-Cabello, J. C. Spiropyran Derivative of an Elastin-like Bioelastic Polymer: Photoresponsive Molecular Machine to Convert Sunlight into Mechanical Work. Macromolecules vol. 33 9480–9482 (2000).

37. Strzegowski, L. A., Martinez, M. B., Channe Gowda, D., Urry, D. W. & Tirrell, D. A. Photomodulation of the inverse temperature transition of a modified elastin poly(pentapeptide). Journal of the American Chemical Society vol. 116 813–814 (1994).

38. Dyson, H. J., Jane Dyson, H. & Wright, P. E. Intrinsically unstructured proteins and their functions. Nature Reviews Molecular Cell Biology vol. 6 197–208 (2005).

39. Quintanilla-Sierra, L., García-Arévalo, C. & Rodriguez-Cabello, J. C. Self-assembly in elastin-like recombinamers: a mechanism to mimic natural complexity. Mater Today Bio 2, 100007 (2019).

40. Tompa, P. Intrinsically unstructured proteins. Trends in Biochemical Sciences vol. 27 527–533 (2002).

41. Lee, R. van der et al. Classification of Intrinsically Disordered Regions and Proteins. Chemical Reviews vol. 114 6589–6631 (2014).

42. Dunker, A. K. et al. Intrinsic Disorder and Protein Function†. Biochemistry vol. 41 6573–6582 (2002).

43. Mithieux, S. M. & Weiss, A. S. Elastin. Fibrous Proteins: Coiled-Coils, Collagen and Elastomers 437–461 (2005) doi:10.1016/s0065-3233(05)70013-9.

44. Zhang, Z., Ortiz, O., Goyal, R. & Kohn, J. Biodegradable Polymers. Principles of Tissue Engineering 441–473 (2014) doi:10.1016/b978-0-12-398358-9.00023-9.

45. Li, N. K., Quiroz, F. G., Hall, C. K., Chilkoti, A. & Yingling, Y. G. Molecular Description of the LCST Behavior of an Elastin-Like Polypeptide. Biomacromolecules vol. 15 3522–3530 (2014).

46. Lee, A. L. & Joshua Wand, A. Microscopic origins of entropy, heat capacity and the glass transition in proteins. Nature vol. 411 501–504 (2001).

47. Pometun, M. S., Chekmenev, E. Y. & Wittebort, R. J. Quantitative Observation of Backbone Disorder in Native Elastin. Journal of Biological Chemistry vol. 279 7982–7987 (2004).

48. Gronau, G. et al. A review of combined experimental and computational procedures for assessing biopolymer structure–process–property relationships. Biomaterials vol. 33 8240–8255 (2012).

49. Hoeve, C. A. J. & Flory, P. J. The elastic properties of elastin. Biopolymers vol. 13 677–686 (1974).

50. Aaron, B. B. & Gosline, J. M. Elastin as a random-network elastomer: A mechanical and optical analysis of single elastin fibers. Biopolymers vol. 20 1247–1260 (1981).

51. Tjin, M. S., Low, P. & Fong, E. Recombinant elastomeric protein biopolymers: progress and prospects. Polymer Journal vol. 46 444–451 (2014).

52. Urry, D. W., Long, M. M. & Gross, E. Conformations of the Repeat Peptides of Elastin in Solution: An Application of Proton and Carbon-13 Magnetic Resonance to the Determination of Polypeptide Secondary Structure. CRC Critical Reviews in Biochemistry vol. 4 1–45 (1976).

53. Biochemistry and Cell Biology: International Symposium on Recent Advances in Molecular, Clinical, and Social Medicine / Biochimie et biologie cellulaire : Symposium international sur les derniers développements en médecine moléculaire, clinique et sociale. Biochemistry and Cell Biology vol. 88 iii (2010).

54. Boulet-Audet, M., Terry, A. E., Vollrath, F. & Holland, C. 'Corrigendum to "Silk protein aggregation kinetics revealed by Rheo-IR" [Acta Biomater. 10 (2014) 776–784]'. Acta Biomaterialia vol. 10 2876 (2014).

55. Snowball, A. & Schorge, S. Erratum to 'Changing channels in pain and epilepsy: Exploiting ion channel gene therapy for disorders of neuronal hyperexcitability' [FEBS Lett. 589 (2015) 1620-1634]. FEBS Letters vol. 589 2429–2429 (2015).

56. Meng, Y., Ahuja, L. G., Kornev, A. P., Taylor, S. S. & Roux, B. Corrigendum to 'A Catalytically-Disabled Double Mutant of Src Tyrosine Kinase Can Be Stabilized into an Active-Like Conformation.' J. Mol. Biol. 430(6) (Mar 16 2018), 881–889. Journal of Molecular Biology vol. 430 4439 (2018).

57. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proceedings of the National Academy of Sciences vol. 110 13392–13397 (2013).

58. Wright, E. R. & Conticello, V. P. Self-assembly of block copolymers derived from elastin-mimetic polypeptide sequences. Advanced Drug Delivery Reviews vol. 54 1057–1073 (2002).

59. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738 (2010).

60. Dunker, A. K. et al. Intrinsically disordered protein. Journal of Molecular Graphics and Modeling vol. 19 26–59 (2001).

61. ELN elastin [Homo sapiens (human)] - Gene - NCBI. https://www.ncbi.nlm.nih.gov/gene/2006.

62. Meyer, D. E., Trabbic-Carlson, K. & Chilkoti, A. Protein purification by fusion with an environmentally responsive elastin-like polypeptide: effect of polypeptide length on the purification of thioredoxin. Biotechnol. Prog. 17, 720–728 (2001).

63. Zhang, Y.-N. et al. A Highly Elastic and Rapidly Crosslinkable Elastin-Like Polypeptide-Based Hydrogel for Biomedical Applications. Adv. Funct. Mater. 25, 4814–4826 (2015).

64. ELN elastin [Homo sapiens (human)] - Gene - NCBI. https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=2006.

65. Elastin. in Advances in Protein Chemistry vol. 70 437–461 (Academic Press, 2005).

66. Mark E. Curran, et al. The elastin gene is disrupted by a translocation associated with supravalvular aortic stenosis. Cell.73, 159-168 (1993)

67. ELN elastin [Homo sapiens (human)] - Gene - NCBI.

68. Lisa D. Muiznieks, et al. Structural disorder and dynamics of elastin. Biochemistry and Cell Biology, Review-Article (2010)

69. David He, et al. Polymorphisms in the human tropoelastin gene modify in vitro self-assembly and mechanical properties of elastin-like polypeptides. PLoS One. 7(9) (2012)

70. ELN elastin [Homo sapiens (human)] - Gene - NCBI. https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=2006.

71. Akagawa, H. et al. A haplotype spanning two genes, ELN and LIMK1, decreases their transcripts and confers susceptibility to intracranial aneurysms. Hum. Mol. Genet. 15, 1722–1734 (2006).

72. He, D. et al. Polymorphisms in the human tropoelastin gene modify in vitro self-assembly and mechanical properties of elastin-like polypeptides. PLoS One 7, e46130 (2012).

73. Black, S. D. & Mould, D. R. Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. Analytical Biochemistry vol. 193 72–82 (1991).

74. Cowan, R. & Whittaker, R. G. Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography. Pept. Res. 3, 75–80 (1990).

Appendix A

Statistics Result of Six Different ELN Amino Acids Sequence

- **P155024Human.fasta**

>sp|P15502.4|ELN_HUMAN RecName: Full=Elastin; AltName: Full=Tropoelastin; Flags: Precursor
MAGLTAAAPRPGVLLLLLSILHPSRPGGVPGAIPGGVPGGVFYPGAGLGALGGGALGPG
GKPLKPVPGGLAGAGLGAGLGAFPAVTFPGALVPGGVADAAAAAYKAAKAGAGLGGVP
GVGGLGVSAGAVVPQPGAGVKPGKVPGVGLPGVYPGGVLPGARFPGVGVLPGVPTGA
GVKPKAPGVGGAFAGIPGVGPFGGPQPGVPLGYPIKAPKLPGGYGLPYTTGKLPYGYGP
GGVAGAAGKAGYPTGTGVGPQAAAAAAAKAAAKFGAGAAGVLPGVGGAGVPGVPGA
IPGIGGIAGVGTPAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGAGVPGVGV
PGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARPGVGVGGIPTYGVGAGGF
PGFGVGVGGIPGVAGVPGVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGAAGAGVLG
GLVPGAPGAVPGVPGTGGVPGVGTPAAAAAKAAAKAAQFGLVPGVGVAPGVGVAPGV
GVAPGVGLAPGVGVAPGVGVAPGVGVAPGIGPGGVAAAAKSAAKVAAKAQLRAAAGL
GAGIPGLGVGVGVPGLGVGAGVPGLGVGAGVPGFGAGADEGVRRSLSPELREGDPSSSQ
HLPSTPSSPRVPGALAAAKAAKYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAA
KAAAKAAQFGLVGAAGLGGLGVGGLGVPGVGGLGGIPPAAAAKAAKYGAAGLGGVL
GGAGQFPLGGVAARPGFGLSPIFPGGACLGKACGRKRK

Seq.length: 786

'v'

13,29,37,41,66,85,92,96,115,118,123,128,129,136,141,144,149,154,163,165,169,175,182,192,20
2,235,250,273,277,282,285,298,323,333,334,336,341,344,346,353,354,364,367,368,389,391,399,
409,411,417,420,423,426,429,432,435,459,464,471,474,480,483,503,506,508,512,514,518,520,5
24,530,532,536,538,542,544,553,563,584,586,588,593,597,602,606,617,645,661,664,674,680,68
1,704,714,719,722,747,759
The number of query in the subject is: **98**

'p'

9,11,23,26,30,34,38,44,58,62,65,67,83,88,93,116,130,132,138,142,147,151,156,161,167,170,177,
180,190,194,198,200,203,207,211,214,220,227,232,245,252,275,283,286,290,301,324,327,331,3
37,342,347,352,355,360,365,370,387,395,405,415,421,427,433,439,465,468,472,475,481,486,50
4,510,516,522,528,534,540,546,550,580,589,598,607,623,630,637,640,643,646,662,677,685,720,
729,730,755,763,769,772
The number of query in the subject is: **100**

'g'

3,12,27,28,31,35,36,39,40,45,47,49,52,53,54,57,59,60,68,69,72,74,76,78,80,89,94,95,109,111,113,114,117,119,120,122,126,133,135,139,143,145,148,152,153,157,162,164,168,172,174,181,183,184,188,191,193,196,197,201,205,215,216,218,224,229,231,233,234,237,240,243,247,249,251,267,269,272,276,278,279,281,284,287,291,293,294,297,299,317,321,325,326,328,330,332,335,338,340,343,345,348,350,356,358,361,366,384,388,390,392,393,398,400,402,403,406,408,410,412,413,416,419,422,424,425,428,430,431,434,436,453,456,458,461,462,466,469,473,476,478,479,482,484,501,505,507,511,513,517,519,523,525,529,531,535,537,541,543,547,549,551,552,574,576,578,581,583,585,587,590,592,594,596,599,601,603,605,608,610,612,616,628,647,658,663,666,667,669,672,673,675,678,679,682,684,702,705,708,710,711,713,715,716,718,721,723,724,726,727,740,743,745,746,749,750,752,757,758,764,766,773,774,778,782
The number of query in the subject is: **225**

'vp'

29,37,66,92,115,129,141,169,202,282,285,323,336,341,346,354,364,420,426,432,464,471,474,480,503,588,597,606,645,661,719
The number of query in the subject is: **31**

'pg'

11,26,30,34,38,44,58,67,88,93,116,132,138,142,147,151,156,161,167,180,190,200,214,232,275,283,286,290,324,327,331,337,342,347,355,360,365,387,405,415,421,427,433,465,468,472,475,481,504,510,516,522,528,534,540,546,550,580,589,598,607,646,662,677,720,763,772
The number of query in the subject is: **67**

'vpg'

29,37,66,92,115,141,282,285,323,336,341,346,354,364,420,426,432,464,471,474,480,503,588,597,606,645,661,719
The number of query in the subject is: **28**

Disorder Regions:

==========VLXT NNP STATISTICS==========
Predicted residues: 786 Number Disordered Regions: 12
Number residues disordered: 621 Longest Disordered Region: 103
Overall percent disordered: 79.01 Average Prediction Score: 0.7345

1.Predicted disorder segment [1]-[11] Average Strength= 0.6729
MAGLTAAAPRP

2.Predicted disorder segment [13]-[74] Average Strength= 0.8220
VLLLLLSILHPSRPGGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPGGLAGAG

3.Predicted disorder segment [100]-[183] Average Strength= 0.8082
AAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGLPGVYPGGVLPGARFPG
VGVLPGVPTGAGVKPKAPGVG

4.Predicted disorder segment [196]-[211] Average Strength= 0.7511
GGPQPGVPLGYPIKAP

5.Predicted disorder segment [228]-[259] Average Strength= 0.7968
YGYGPGGVAGAAGKAGYPTGTGVGPQAAAAAA

6.Predicted disorder segment [273]-[375] Average Strength= 0.9031
VLPGVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVG
VPGAGVPGVGVPGAGIPVVPGAGIPGAAVPGVVSPEAAAK

7.Predicted disorder segment [378]-[395] Average Strength= 0.7828
AKAAKYGARPGVGVGGIP

8.Predicted disorder segment [416]-[492] Average Strength= 0.9002
GVAGVPGVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGAAGAGVLGGLVPGAPGAVPGVPGTG
GVPGVGTPAAAAAK

9.Predicted disorder segment [506]-[607] Average Strength= 0.9203
VGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAPGIGPGGVAAAAKSAAKVAAKA
QLRAAAGLGAGIPGLGVGVGVPGLGVGAGVPGLGVGAGVP

10.Predicted disorder segment [616]-[692] Average Strength= 0.8969
GVRRSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGVLGGLGALGGVGIPGGV
VGAGPAAAAAAA

11.Predicted disorder segment [712]-[743] Average Strength= 0.8974
LGVGGLGVPGVGGLGGIPPAAAAKAAKYGAAG

12.Predicted disorder segment [780]-[786] Average Strength= 0.8173
ACGRKRK

Order Regions:

1.Predicted order segment [75]-[99]
LGAGLGAFPAVTFPGALVPGGVADA

2.Predicted order segment [396]-[415]
TYGVGAGGFPGFGVGVGGIP

3.Predicted order segment [744]-[779]
LGGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGK

- **PNI925061Pantroglodytes.fasta**

>PNI92506.1 ELN isoform 3 [Pan troglodytes]
MAGLTAAAPRPGVLLLLLSILHPSRPGGVPGAIPGGVPGGVFYPGAGLGALGGGALAPG
VKPLKPVPGGLVGAGLGAGLGAFPAVTFPGALVPGGVADAAAAYKAAKAGAGLGGVP
GVGGLGVSAGAVVPQPGAGVKPGKVPGVGLPGVYPGGVLPGARFPGVGVLPGVPTGA
GVKPKAPGVGGAFAGIPGVGPFGGPQPGVPLGYPIKAPKLPGGYGLPYTTGKLPYGYGP
GGVAGVAGKAGYPTGTGVGPQAAAAAAAKAAAKLGAGAAGVLPGVGGAGVPGVPGA
IPGIGGIAGVGTPAAAAAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGAGVP
GVGVPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARPGVGVGGIPTYGVG
AGGFPGFGVGVGGIPGVAGVPSVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGAAGA
GVLGGLVPGAPGAVPGVPGTGGVPGVGTPAAAAAKAAAKAAQFGLVPGVGVAPGVGV
APGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGIGPGGVAAAAKS
AAKVAAKAQLRAAAGLGAGIPGLGVGVGVPGLGVGAGVPGLGVGAGVPGFGAGADEG
VRRSLSPELREGDPSSSQHLPSTPSPPRVPGALAAAKAAKYGAAVPGVLGGLGALGGVGI
PGGVVGAGPAAAAAAAKAAAKAAQFGLVGAAGLGGLGVGGLGVPGVGGLAGIPPAAA
AKAAKYGAAGLGGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGKACGRKRK

Seq.length: 802

'v'

13,29,37,41,60,66,71,85,92,96,115,118,123,128,129,136,141,144,149,154,163,165,169,175,182,1
92,202,235,238,250,273,277,282,285,298,327,337,338,340,345,348,350,357,358,368,371,372,39
3,395,403,413,415,421,424,427,430,433,436,439,463,468,475,478,484,487,507,510,512,516,518,
522,524,528,530,534,536,540,542,546,548,552,554,558,560,569,579,600,602,604,609,613,618,6
22,633,661,677,680,690,696,697,720,730,735,738,763,775
The number of query in the subject is: **106**

'p'

9,11,23,26,30,34,38,44,58,62,65,67,83,88,93,116,130,132,138,142,147,151,156,161,167,170,177,
180,190,194,198,200,203,207,211,214,220,227,232,245,252,275,283,286,290,301,328,331,335,3
41,346,351,356,359,364,369,374,391,399,409,419,425,431,437,443,469,472,476,479,485,490,50
8,514,520,526,532,538,544,550,556,562,566,596,605,614,623,639,646,653,656,658,659,662,678,
693,701,736,745,746,771,779,785,788
The number of query in the subject is: **103**

'g'

3,12,27,28,31,35,36,39,40,45,47,49,52,53,54,59,68,69,72,74,76,78,80,89,94,95,109,111,113,114,
117,119,120,122,126,133,135,139,143,145,148,152,153,157,162,164,168,172,174,181,183,184,1
88,191,193,196,197,201,205,215,216,218,224,229,231,233,234,237,240,243,247,249,251,267,26

9,272,276,278,279,281,284,287,291,293,294,297,299,321,325,329,330,332,334,336,339,342,344,
347,349,352,354,360,362,365,370,388,392,394,396,397,402,404,406,407,410,412,414,416,417,4
20,423,428,429,432,434,435,438,440,457,460,462,465,466,470,473,477,480,482,483,486,488,50
5,509,511,515,517,521,523,527,529,533,535,539,541,545,547,551,553,557,559,563,565,567,568,
590,592,594,597,599,601,603,606,608,610,612,615,617,619,621,624,626,628,632,644,663,674,6
79,682,683,685,688,689,691,694,695,698,700,718,721,724,726,727,729,731,732,734,737,739,74
0,743,756,759,761,762,765,766,768,773,774,780,782,789,790,794,798
The number of query in the subject is: **225**

'vp'

29,37,66,92,115,129,141,169,202,282,285,327,340,345,350,358,368,424,430,436,468,475,478,48
4,507,604,613,622,661,677,735
The number of query in the subject is: **31**

'pg'

11,26,30,34,38,44,58,67,88,93,116,132,138,142,147,151,156,161,167,180,190,200,214,232,275,2
83,286,290,328,331,335,341,346,351,359,364,369,391,409,419,431,437,469,472,476,479,485,50
8,514,520,526,532,538,544,550,556,562,566,596,605,614,623,662,678,693,736,779,788
The number of query in the subject is: **68**

'vpg'

29,37,66,92,115,141,282,285,327,340,345,350,358,368,430,436,468,475,478,484,507,604,613,62
2,661,677,735
The number of query in the subject is: **27**

Disorder Regions:

==========VLXT NNP STATISTICS==========
Predicted residues: 802 Number Disordered Regions: **11**
Number residues disordered: 645 Longest Disordered Region: 148
Overall percent disordered: 80.42 Average Prediction Score: 0.7501

1.Predicted disorder segment [1]-[11] Average Strength= 0.6729
MAGLTAAAPRP

2.Predicted disorder segment [13]-[73] Average Strength= 0.8073
VLLLLLSILHPSRPGGVPGAIPGGVPGGVFYPGAGLGALGGGALAPGVKPLKPVPGGLVGA

3.Predicted disorder segment [100]-[183] Average Strength= 0.8082

AAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGLPGVYPGGVLPGARFPG
VGVLPGVPTGAGVKPKAPGVG

4.Predicted disorder segment [196]-[211] Average Strength= 0.7511
GGPQPGVPLGYPIKAP

5.Predicted disorder segment [232]-[379] Average Strength= 0.9162
PGGVAGVAGKAGYPTGTGVGPQAAAAAAAKAAAKLGAGAAGVLPGVGGAGVPGVPGAIPGIGG
IAGVGTPAAAAAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGAGVPGVGVPGAGIPV
VPGAGIPGAAVPGVVSPEAAAK

6.Predicted disorder segment [382]-[399] Average Strength= 0.7828
AKAAKYGARPGVGVGGIP

7.Predicted disorder segment [420]-[496] Average Strength= 0.9017
GVAGVPSVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGAAGAGVLGGLVPGAPGAVPGVPGTG
GVPGVGTPAAAAAK

8.Predicted disorder segment [510]-[623] Average Strength= 0.9280
VGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGIGPGGVAA
AAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGVPGLGVGAGVPGLGVGAGVP

9.Predicted disorder segment [632]-[708] Average Strength= 0.8975
GVRRSLSPELREGDPSSSQHLPSTPSPPRVPGALAAAKAAKYGAAVPGVLGGLGALGGVGIPGGV
VGAGPAAAAAAA

10.Predicted disorder segment [728]-[759] Average Strength= 0.8952
LGVGGLGVPGVGGLAGIPPAAAAKAAKYGAAG

11.Predicted disorder segment [796]-[802] Average Strength= 0.8173
ACGRKRK

Order Regions:

1.Predicted order segment [74]-[99]
GLGAGLGAFPAVTFPGALVPGGVADA

2.Predicted order segment [400]-[419]
TYGVGAGGFPGFGVGVGGIP

3.Predicted order segment [760]-[795]
LGGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGK

● XP0348072831Panpaniscus.fasta

>XP_034807283.1 elastin isoform X13 [Pan paniscus]

MAGLTAAAPRPGVLLLLLSILHPSRPGGVPGAIPGGVPGGVFYPGAGLGALGGGALAPG
VKPLKPVPGGLVGAGLGAGLGAFPAVTFPGALVPGGVADAAAAAYKAAKAGAGLGGVP
GVGGLGVSAGAVVPQPGAGVKPGKVPGVGLPGVYPGGVLPGARFPGVGVLPGVPTGA
GVKPKAPGVGGAFAGIPGVGPFGGPQPGVPLGYPIKAPKLPGGYGLPYTTGKLPYGYGP
GGVAGVAGKAGYPTGTGVGPQAAAAAAAKAAAKLGAGAAGVLPGVGGAGVPGVPGA
IPGIGGIAGVGTPAAAAAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGAGVP
GVGVPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARPGVGVGGIPTYGVG
AGGFPGFGVGVGGIPGVAGVPSVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGAAGA
GVLGGLVPGAPGAVPGVPGTGGVPGVGTPAAAAAKAAAKAAQFVGVAPGVGVAPGVG
VAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGIGPGGVAAAAKSAAKVAAKAQL
RAAAGLGAGIPGLGVGVGVPGLGVGAGVPGLGVGAGVPGFGAVPGALAAAKAAKYGA
AVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAAKAAAKAAQFGLVGAAGLGGLGVG
GLGVPGVGGLAGIPPAAAAKAAKYGAAGLGGVLGGAGQFPLGGVAARPGFGLSPIFPG
GACLGKACGRKRK

Seq.length: 758

'v'

13,29,37,41,60,66,71,85,92,96,115,118,123,128,129,136,141,144,149,154,163,165,169,175,182,1
92,202,235,238,250,273,277,282,285,298,327,337,338,340,345,348,350,357,358,368,371,372,39
3,395,403,413,415,421,424,427,430,433,436,439,463,468,475,478,484,487,505,507,511,513,517,
519,523,525,529,531,535,537,541,543,547,549,558,568,589,591,593,598,602,607,611,617,633,6
36,646,652,653,676,686,691,694,719,731
The number of query in the subject is: **102**

'p'

9,11,23,26,30,34,38,44,58,62,65,67,83,88,93,116,130,132,138,142,147,151,156,161,167,170,177,
180,190,194,198,200,203,207,211,214,220,227,232,245,252,275,283,286,290,301,328,331,335,3
41,346,351,356,359,364,369,374,391,399,409,419,425,431,437,443,469,472,476,479,485,490,50
9,515,521,527,533,539,545,551,555,585,594,603,612,618,634,649,657,692,701,702,727,735,741,
744
The number of query in the subject is: **95**

'g'

3,12,27,28,31,35,36,39,40,45,47,49,52,53,54,59,68,69,72,74,76,78,80,89,94,95,109,111,113,114,
117,119,120,122,126,133,135,139,143,145,148,152,153,157,162,164,168,172,174,181,183,184,1
88,191,193,196,197,201,205,215,216,218,224,229,231,233,234,237,240,243,247,249,251,267,26
9,272,276,278,279,281,284,287,291,293,294,297,299,321,325,329,330,332,334,336,339,342,344,
347,349,352,354,360,362,365,370,388,392,394,396,397,402,404,406,407,410,412,414,416,417,4
20,423,428,429,432,434,435,438,440,457,460,462,465,466,470,473,477,480,482,483,486,488,50
6,510,512,516,518,522,524,528,530,534,536,540,542,546,548,552,554,556,557,579,581,583,586,

588,590,592,595,597,599,601,604,606,608,610,613,615,619,630,635,638,639,641,644,645,647,650,651,654,656,674,677,680,682,683,685,687,688,690,693,695,696,699,712,715,717,718,721,722,724,729,730,736,738,745,746,750,754
The number of query in the subject is: **218**

'vp'

29,37,66,92,115,129,141,169,202,282,285,327,340,345,350,358,368,424,430,436,468,475,478,484,593,602,611,617,633,691
The number of query in the subject is: **30**

'pg'

11,26,30,34,38,44,58,67,88,93,116,132,138,142,147,151,156,161,167,180,190,200,214,232,275,283,286,290,328,331,335,341,346,351,359,364,369,391,409,419,431,437,469,472,476,479,485,509,515,521,527,533,539,545,551,555,585,594,603,612,618,634,649,692,735,744
The number of query in the subject is: **66**

'vpg'

29,37,66,92,115,141,282,285,327,340,345,350,358,368,430,436,468,475,478,484,593,602,611,617,633,691
The number of query in the subject is: **26**

Disorder Regions:

==========VLXT NNP STATISTICS==========
Predicted residues: 758 Number Disordered Regions: **11**
Number residues disordered: 602 Longest Disordered Region: 148
Overall percent disordered: 79.42 Average Prediction Score: 0.7385

1.Predicted disorder segment [1]-[11] Average Strength= 0.6729
AGLTAAAPRP

2.Predicted disorder segment [13]-[73] Average Strength= 0.8073
VLLLLLSILHPSRPGGVPGAIPGGVPGGVFYPGAGLGLGGGALAPGVKPLKPVPGGLVGA

3.Predicted disorder segment [100]-[183] Average Strength= 0.8082
AAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGLPGVYPGGVLPGARFPGVGVLPGVP TGAGVKPKAPGVG

4.Predicted disorder segment [196]-[211] Average Strength= 0.7511
GGPQPGVPLGYPIKAP

5.Predicted disorder segment [232]-[379] Average Strength= 0.9162

PGGVAGVAGKAGYPTGTGVGPQAAAAAAAKAAAKLGAGAAGVLPGVGGAGVPGVPGAIPGIGG
IAGVGTPAAAAAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGAGVPGVGVPGAGIPV
VPGAGIPGAAVPG VVSPEAAAK

6.Predicted disorder segment [382]-[399] Average Strength= 0.7828
AKAAKYGARPGVGVGGIP

7.Predicted disorder segment [420]-[497] Average Strength= 0.8959
GVAGVPSVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGAAGAGVLGGLVPGAPGAVPGVPGTG
GVPGVGTPAAAAAKA

8.Predicted disorder segment [509]-[615] Average Strength= 0.9191
PGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGIGPGGVAAAAKSAAKVAA
KAQLRAAAGLGAGIPGLGVGVGVPGLGVGAGVPGLGVGAGVPGFG

9.Predicted disorder segment [625]-[664] Average Strength= 0.8822
KAAKYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAA

10.Predicted disorder segment [684]-[715] Average Strength= 0.8952
LGVGGLGVPGVGGLAGIPPAAAAKAAKYGAAG

11.Predicted disorder segment [752]-[758] Average Strength= 0.8173
ACGRKRK

Order Regions:

1.Predicted disorder segment [74]-[99]
GLGAGLGAFPAVTFPGALVPGGVADA

2.Predicted disorder segment [400]-[419]
TYGVGAGGFPGFGVGVGGIP

3.Predicted disorder segment [716]-[751]
LGGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGK

- **PNJ857951Pongoabelii.fasta**

>PNJ85795.1 ELN isoform 10 [Pongo abelii]
MAGLTAAAPRPGVLLLLLSILHPSRPGGVPGAIPGGVPGGVFYPVLGPGGKPLKPGLGTF
PAGTFPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGIGGLGVSAGAVVPQPGAGV
KPGKVPGVGLPGVYPGGVLPGAGARFPGVGVLPGVPTGAGVKPKAPGVGGAFAGIPGV
GPFAGPQPGVPLGYPIKAPKLPGGYGLPYTTGKLPYGYGPGGVAAAAGKAGYPTGTGV
GPQAAAAAAAKAAAKFGAGAAGVLPGVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAA
AAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGAAVPGVGVPGAGIPVVPGAGIPGAGVP
GVVSPEAAAKAAAKAAKYGARAGVGVGGIPTFGVGAGGFPGFGVGVGGIPGVAGVPSV

GGVPGVGGVPGAVISPEAQAAAAAKAAKYGVGTPPAAAAKAAAKAAQFGLVPGVGVA
PGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVAAAAK
SAAKAAAKAQLRAAAGLGAGVPGLGVGVGVPGLGVGVGVPGLGVGAGVPGFGAVPG
ALAAAKAAKYGAAVPGVLGGVGALGGVGIPGGVVGAGPAAAAAAAKAAAKAAQFGL
GGAAGLGGLGVGGLGVPGVGGLGGVPPAAAAKAAKYGAAGLGGVLGGAGQFPLGGV
AARPGFGLSPIFPGGACLGKACGRKRK

Seq.length: 714

'v'

13,29,37,41,45,70,74,93,96,104,109,110,117,122,125,130,135,146,148,152,158,165,175,185,218,
233,256,260,265,268,281,306,316,317,319,324,327,329,336,337,347,350,351,372,374,382,392,3
94,400,403,406,409,412,415,419,437,457,460,462,466,468,472,474,478,480,484,486,490,492,49
6,498,502,504,508,510,514,540,545,547,549,554,556,558,563,567,573,589,592,596,602,608,609,
642,647,650,656,675,687
The number of query in the subject is: **98**

'p'

9,11,23,26,30,34,38,44,48,52,55,61,66,71,94,111,113,119,123,128,132,137,144,150,153,160,163,
173,177,181,183,186,190,194,197,203,210,215,228,235,258,266,269,273,284,307,310,314,320,3
25,330,335,338,343,348,353,378,388,398,404,410,416,422,440,441,458,464,470,476,482,488,49
4,500,506,512,541,550,559,568,574,590,605,613,648,657,658,683,691,697,700
The number of query in the subject is: **90**

'g'

3,12,27,28,31,35,36,39,40,47,49,50,56,58,63,67,72,73,87,89,91,92,95,97,98,100,101,103,107,114
,116,120,124,126,129,133,134,138,140,145,147,151,155,157,164,166,167,171,174,176,180,184,1
88,198,199,201,207,212,214,216,217,223,226,230,232,234,250,252,255,259,261,262,264,267,27
0,274,276,277,280,282,300,304,308,309,311,313,315,318,321,326,328,331,333,339,341,344,346,
349,367,371,373,375,376,381,383,385,386,389,391,393,395,396,399,402,407,408,411,413,414,4
17,436,438,455,459,461,465,467,471,473,477,479,483,485,489,491,495,497,501,503,507,509,51
3,535,537,539,542,544,546,548,551,553,555,557,560,562,564,566,569,571,575,586,591,594,595,
597,600,601,603,606,607,610,612,630,632,633,636,638,639,641,643,644,646,649,651,652,654,6
55,668,671,673,674,677,678,680,685,686,692,694,701,702,706,710
The number of query in the subject is: **202**

'vp'

29,37,70,93,110,122,152,185,265,268,306,319,324,329,337,347,403,409,415,457,540,549,558,56
7,573,589,647,656
The number of query in the subject is: **28**

'pg'

11,26,30,34,38,48,55,66,71,94,113,119,123,128,132,137,144,150,163,173,183,197,215,258,266,269,273,307,310,314,320,325,330,338,343,348,388,398,410,416,458,464,470,476,482,488,494,500,506,512,541,550,559,568,574,590,605,648,691,700
The number of query in the subject is: **60**

'vpg'

29,37,70,93,122,265,268,306,319,324,329,337,347,409,415,457,540,549,558,567,573,589,647
The number of query in the subject is: **23**

Disorder Regions:

==========VLXT NNP STATISTICS==========
Predicted residues: 714 Number Disordered Regions: 12
Number residues disordered: 532 Longest Disordered Region: 112
Overall percent disordered: 74.51 Average Prediction Score: 0.7006

1.Predicted disorder segment [1]-[11] Average Strength= 0.6729
MAGLTAAAPRP

2.Predicted disorder segment [13]-[45] Average Strength= 0.7947
VLLLLLSILHPSRPGGVPGAIPGGVPGGVFYPV

3.Predicted disorder segment [78]-[166] Average Strength= 0.8241
AAAYKAAKAGAGLGGVPGVGGIGGLGVSAGAVVPQPGAGVKPGKVPGVGLPGVYPGGVLPGAGARFPGVGVLPGVPTGAGVKPKAPGVG

4.Predicted disorder segment [179]-[194] Average Strength= 0.7389
AGPQPGVPLGYPIKAP

5.Predicted disorder segment [211]-[242] Average Strength= 0.7878
YGYGPGGVAAAAGKAGYPTGTGVGPQAAAAAA

6.Predicted disorder segment [256]-[356] Average Strength= 0.9105
VLPGVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGAAVPGVGVPGAGIPVVPGAGIPGAGVPGVVSPEAA

7.Predicted disorder segment [362]-[369] Average Strength= 0.5943
KAAKYGAR

8.Predicted disorder segment [399]-[447] Average Strength= 0.8480
GVAGVPSVGGVPGVGGVPGAVISPEAQAAAAAKAAKYGVGTPPAAAAKA

9.Predicted disorder segment [460]-[571] Average Strength= 0.9198
VGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVAAAAKSA
AKAAAKAQLRAAAGLGAGVPGLGVGVGVPGLGVGVGVPGLGVGAGVPGFG

10.Predicted disorder segment [581]-[621] Average Strength= 0.8847
KAAKYGAAVPGVLGGVGALGGVGIPGGVVGAGPAAAAAAAK

11.Predicted disorder segment [639]-[671] Average Strength= 0.8927
GLGVGGLGVPGVGGLGGVPPAAAAKAAKYGAAG

12.Predicted disorder segment [708]-[714] Average Strength= 0.8173
ACGRKRK

Order Regions:

1.Predicted disorder segment [370]-[398]
AGVGVGGIPTFGVGAGGFPGFGVGVGGIP

2.Predicted disorder segment [672]-[707]
LGGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGK

● XP011239157.fasta

>XP_030868570.1 elastin isoform X2 [Gorilla gorilla gorilla]
MAGLTAAPRPGVLLLLLSILHPSRPGGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGV
KPLKPVPGGLAGAGLGAGLGAFPAVTFPGALVPGGVADAAAAAYKAAKAGAGLGGVPG
VGGLGVSAGAVVPQPGAGVKPGKVPGVGLPGVYPGGVLPGARFPGVGVLPGVPTGAG
VKPKAPGVGGAFAGIPGVGPFGGPQPGVPLGYPIKAPKLPGGYGLPYTTGKLPYGYGPG
GVAGAAGKAGYPTGTGVGPQAAAAAAAAKAAAKFGAGAAGVLPGVGGAGVPGVPGAI
PGIAGVGTPAAAAAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGAGVPGVG
VPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARPGVGVGGIPTYGVGAGG
FPGFGVGVGGIPGVAGVPSVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGLVPGVGVA
PGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPG
VGVAPGIGIVPGGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGVPGLGVGAG
VPGFGAVPGALAAAKAAKYGAAVPGVLGGLGALGGVGVPGGVVGAAPAAAAAAAAKA
AAKAAQFGLVGAAGLGGLGVGGLGVPGVGALGGVPPAAAAKAAKYGAAGLGGVLGG
AGQFPLGGVAARPGFGLSPIFPGGACLGKACGRKRK

Seq.length: 722

'v'

12,28,36,40,59,65,84,91,95,114,117,122,127,128,135,140,143,148,153,162,164,168,174,181,191,
201,234,249,272,276,281,284,294,323,333,334,336,341,344,346,353,354,364,367,368,389,391,3
99,409,411,417,420,423,426,429,432,435,455,458,460,464,466,470,472,476,478,482,484,488,49

0,494,496,500,502,506,508,512,514,518,520,527,531,541,562,564,566,571,575,581,597,600,610,612,616,617,640,650,655,658,664,683,695
The number of query in the subject is: **102**

'p'

8,10,22,25,29,33,37,43,57,61,64,66,82,87,92,115,129,131,137,141,146,150,155,160,166,169,176,179,189,193,197,199,202,206,210,213,219,226,231,244,251,274,282,285,289,297,324,327,331,337,342,347,352,355,360,365,370,387,395,405,415,421,427,433,439,456,462,468,474,480,486,492,498,504,510,516,522,528,558,567,576,582,598,613,621,656,665,666,691,699,705,708
The number of query in the subject is: **92**

'g'

3,11,26,27,30,34,35,38,39,44,46,48,51,52,53,56,58,67,68,71,73,75,77,79,88,93,94,108,110,112,113,116,118,119,121,125,132,134,138,142,144,147,151,152,156,161,163,167,171,173,180,182,183,187,190,192,195,196,200,204,214,215,217,223,228,230,232,233,236,239,242,246,248,250,266,268,271,275,277,278,280,283,286,290,293,295,317,321,325,326,328,330,332,335,338,340,343,345,348,350,356,358,361,366,384,388,390,392,393,398,400,402,403,406,408,410,412,413,416,419,424,425,428,430,431,434,436,453,457,459,463,465,469,471,475,477,481,483,487,489,493,495,499,501,505,507,511,513,517,519,523,525,529,530,552,554,556,559,561,563,565,568,570,572,574,577,579,583,594,599,602,603,605,608,609,611,614,615,618,638,641,644,646,647,649,651,652,654,657,659,662,663,676,679,681,682,685,686,688,693,694,700,702,709,710,714,718
The number of query in the subject is: **207**

'vp'

28,36,65,91,114,128,140,168,201,281,284,323,336,341,346,354,364,420,426,432,455,527,566,575,581,597,612,655,664
The number of query in the subject is: **29**

'pg'

10,25,29,33,37,43,57,66,87,92,115,131,137,141,146,150,155,160,166,179,189,199,213,231,274,282,285,289,324,327,331,337,342,347,355,360,365,387,405,415,427,433,456,462,468,474,480,486,492,498,504,510,516,522,528,558,567,576,582,598,613,656,699,708
The number of query in the subject is: **64**

'vpg'

28,36,65,91,114,140,281,284,323,336,341,346,354,364,426,432,455,527,566,575,581,597,612,655
The number of query in the subject is: **24**

Disorder Regions:

==========VLXT NNP STATISTICS==========
Predicted residues: 722 Number Disordered Regions: 11
Number residues disordered: 567 Longest Disordered Region: 164
Overall percent disordered: 78.53 Average Prediction Score: 0.7270

1.Predicted disorder segment [1]-[9] Average Strength= 0.7037
MAGLTAAPR

2.Predicted disorder segment [13]-[73] Average Strength= 0.8159
LLLLLSILHPSRPGGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGVKPLKPVPGGLAGAG

3.Predicted disorder segment [99]-[182] Average Strength= 0.8082
AAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGLPGVYPGGVLPGARFPG
VGVLPGVPTGAGVKPKAPGVG

4.Predicted disorder segment [195]-[210] Average Strength= 0.7511
GGPQPGVPLGYPIKAP

5.Predicted disorder segment [227]-[258] Average Strength= 0.7968
YGYGPGGVAGAAGKAGYPTGTGVGPQAAAAAA

6.Predicted disorder segment [272]-[375] Average Strength= 0.9033
VLPGVGGAGVPGVPGAIPGIAGVGTPAAAAAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVV
GVPGAGVPGVGVPGAGIPVVPGAGIPGAAVPGVVSPEAAAK

7.Predicted disorder segment [378]-[395] Average Strength= 0.7828
AKAAKYGARPGVGVGGI

8.Predicted disorder segment [416]-[579] Average Strength= 0.9158
GVAGVPSVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGLVPGVGVAPGVGVAPGVGVAPGVG
VAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGIGIVPGGVAAAAKSAAK
VAAKAQLRAAAGLGAGIPGLGVGVGVPGLGVGAGVPGFG

9.Predicted disorder segment [589]-[628] Average Strength= 0.8817
KAAKYGAAVPGVLGGLGALGGVGVPGGVVGAAPAAAAAAA

10.Predicted disorder segment [648]-[679] Average Strength= 0.8982
LGVGGLGVPGVGALGGVPPAAAAKAAKYGAAG

11.Predicted disorder segment [716]-[722] Average Strength= 0.8173
ACGRKRK

Order Regions:

1.Predicted disorder segment [46]-[77]
LGAGLGAFPAVTFPGALVPGGVADA

2.Predicted disorder segment [370]-[398]
TYGVGAGGFPGFGVGVGGIP

3.Predicted disorder segment [672]-[707]
GGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGK

● XP0112391571Musmusculus.fasta

>XP_011239157.1 elastin isoform X1 [Mus musculus]
MAGLTAVVPQPGVLLILLLNLLHPAQPGGVPGAVPGGLPGGVPGGVYYPGAGIGGLGG
GGGALGPGGKPPKPGAGLLGTFGAGPGGLGGAGPGAGLGAFPAGTFPGAGALVPGGAA
GAAAAYKAAAKAGAGLGGVGGVPGGVGVGGVPGGVGVGGVPGGVGVGGVPGGVGG
IGGIGGLGVSTGAVVPQVGAGIGAGGKPGKVPGVGLPGVYPGGVLPGTGARFPGVGVLP
GVPTGTGVKAKAPGGGGAFAGIPGVGPFGGQQPGVPLGYPIKAPKLPGGYGLPYTNGKL
PYGVAGAGGKAGYPTGTGVGSQAAAAAAKAAKYGAGGAGVLPGVGGGGIPGGAGAIP
GIGGIAGAGTPAAAAAAKAAAKAAKYGAGGAGALGGLVPGAVPGALPGAVPAVPGAG
GVPGAGTPAAAAAAAAAKAAAKAGLGPGVGGVPGGVGVGGIPGGVGVGGVPGGVGP
GGVTGIGAGPGGLGGAGSPAAAKSAAKAAAKAQYRAAAGLGAGVPGFGAGAGVPGFG
AGAGVPGFGAGAGVPGFGAGAGVPGFGAGAVPGSLAASKAAKYGAAGGLGGPGGLGG
PGGLGGPGGLGGAGVPGRVAGAAPPAAAAAAAKAAAKAAQYGLGGAGGLGAGGLGA
GGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGGVSPAAAAKAAKYGAAGLGGVLGA
RPFPGGGVAARPGFGLSPIYPGGGAGGLGVGGKPPKPYGGALGALGYQGGGCFGKSCG
RKRK

Seq.length: 747

'v'

7,8,13,30,34,42,46,111,135,138,142,144,147,151,153,156,160,162,165,169,180,185,186,189,202,
205,210,215,226,228,232,238,255,265,293,308,329,333,384,388,396,399,405,432,435,439,441,4
48,450,453,457,462,503,512,521,530,539,547,588,592,663,682,693,715
The number of query in the subject is: **64**

'p'

9,11,24,27,31,35,39,43,49,65,69,70,72,84,92,100,105,112,139,148,157,166,187,199,203,208,212,
217,224,230,233,243,253,257,263,266,270,274,277,283,290,303,331,339,346,357,385,389,393,3
97,400,406,411,430,436,445,454,459,469,478,504,513,522,531,540,548,568,574,580,589,597,59
8,665,687,689,697,703,706,719,720,722
The number of query in the subject is: **81**

'g'

3,12,28,29,32,36,37,40,41,44,45,50,52,54,55,57,58,59,60,61,64,66,67,73,75,78,81,83,85,86,88,89
,91,93,95,97,102,106,108,113,114,117,129,131,133,134,136,137,140,141,143,145,146,149,150,1
52,154,155,158,159,161,163,164,167,168,170,171,173,174,176,177,179,183,190,192,194,196,19
7,200,204,206,209,213,214,218,220,225,227,231,235,237,244,245,246,247,251,254,256,259,260,
264,268,278,279,281,287,292,295,297,298,301,305,307,309,323,325,326,328,332,334,335,336,3
37,340,341,343,347,349,350,353,355,373,375,376,378,381,382,386,390,394,401,403,404,407,40
9,427,429,431,433,434,437,438,440,442,443,446,447,449,451,452,455,456,458,460,461,464,466,
468,470,471,473,474,476,498,500,502,505,507,509,511,514,516,518,520,523,525,527,529,532,5
34,536,538,541,543,545,549,560,563,564,566,567,569,570,572,573,575,576,578,579,581,582,58
4,585,587,590,594,615,617,618,620,621,623,625,626,628,630,631,633,635,636,638,640,641,643,
645,646,648,650,651,653,655,656,658,660,661,662,675,678,680,681,684,690,691,692,698,700,7
07,708,709,711,712,714,716,717,724,725,728,731,734,735,736,739,743
The number of query in the subject is: **273**

'vp'

8,30,34,42,111,138,147,156,165,186,202,232,265,384,388,396,399,405,435,453,503,512,521,530
,539,547,588
The number of query in the subject is: **27**

'pg'

11,27,31,35,39,43,49,65,72,84,92,105,112,139,148,157,166,199,203,208,212,217,224,230,243,25
3,263,277,331,339,346,385,389,393,400,406,430,436,445,454,459,469,504,513,522,531,540,548,
568,574,580,589,689,697,706
The number of query in the subject is: **55**

'vpg'

30,34,42,111,138,147,156,165,202,384,388,399,405,435,453,503,512,521,530,539,547,588
The number of query in the subject is: **22**

Disorder Regions:

==========VLXT NNP STATISTICS==========
Predicted residues: 747 Number Disordered Regions: 9
Number residues disordered: 596 Longest Disordered Region: 201
Overall percent disordered: 79.79 Average Prediction Score: 0.7759

1. Predicted disorder segment [4]-[7] Average Strength= 0.5132
LTAV

2. Predicted disorder segment [15]-[91] Average Strength= 0.8914

LILLLNLLHPAQPGGVPGAVPGGLPGGVPGGVYYPGAGIGGLGGGGGALGPGGKPPKPGAGLLGT
FGAGPGGLGGAG

3. Predicted disorder segment [111]-[226] Average Strength= 0.9159
VPGGAAGAAAAYKAAAKAGAGLGGVGGVPGGVGVGGVPGGVGVGGVPGGVGVGGVPGGVGG
IGGIGGLGVSTGAVVPQVGAGIGAGGKPGKVPGVGLPGVYPGGVLPGTGARFPGV

4. Predicted disorder segment [230]-[245] Average Strength= 0.7250
PGVPTGTGVKAKAPGG

5. Predicted disorder segment [260]-[274] Average Strength= 0.6993
GQQPGVPLGYPIKAP

6. Predicted disorder segment [295]-[495] Average Strength= 0.9580
GAGGKAGYPTGTGVGSQAAAAAAKAAKYGAGGAGVLPGVGGGGIPGGAGAIPGIGGIAGAGTP
AAAAAAKAAAKAAKYGAGGAGALGGLVPGAVPGALPGAVPAVPGAGGVPGAGTPAAAAAAAAA
AKAAAKAGLGPGVGGVPGGVGVGGIPGGVGVGGVPGGVGPGGVTGIGAGPGGLGGAGSPAAAK
SAAKAAAKAQYRA

7. Predicted disorder segment [545]-[677] Average Strength= 0.9350
GAVPGSLAASKAAKYGAAGGLGGPGGLGGPGGLGGPGGLGGAGVPGRVAGAAPPAAAAAAAK
AAAKAAQYGLGGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGGVSPAA
AAKAAKYGAA

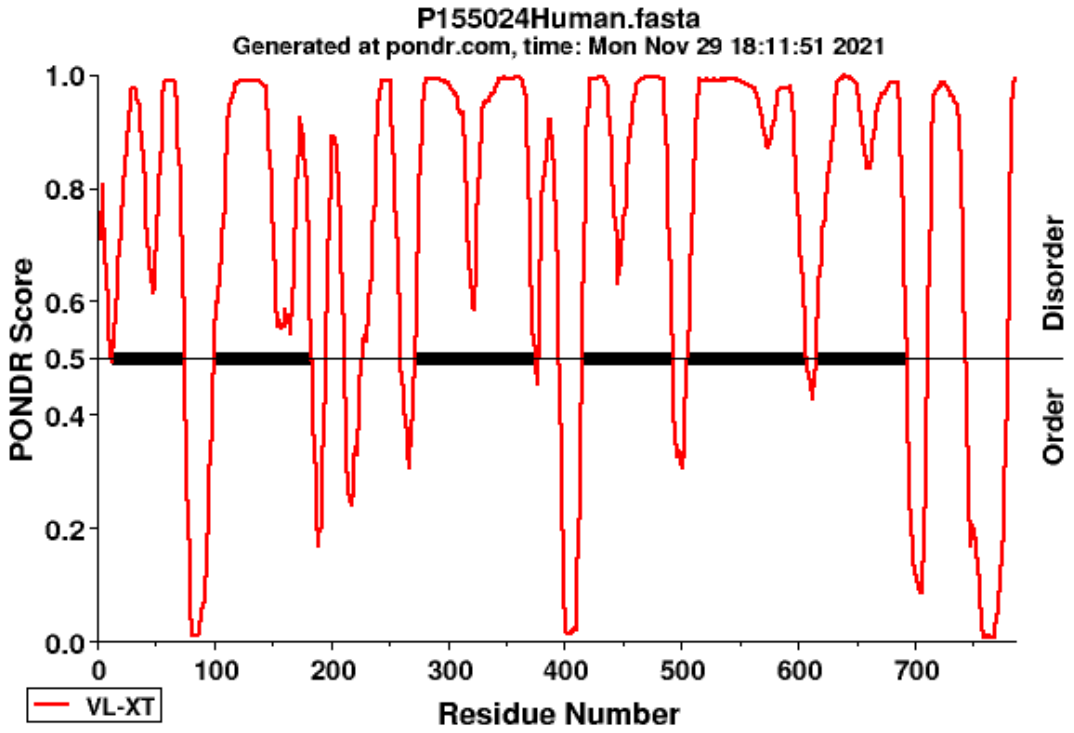
8. Predicted disorder segment [700]-[725] Average Strength= 0.8201
GLSPIYPGGGAGGLGVGGKPPKPYGG

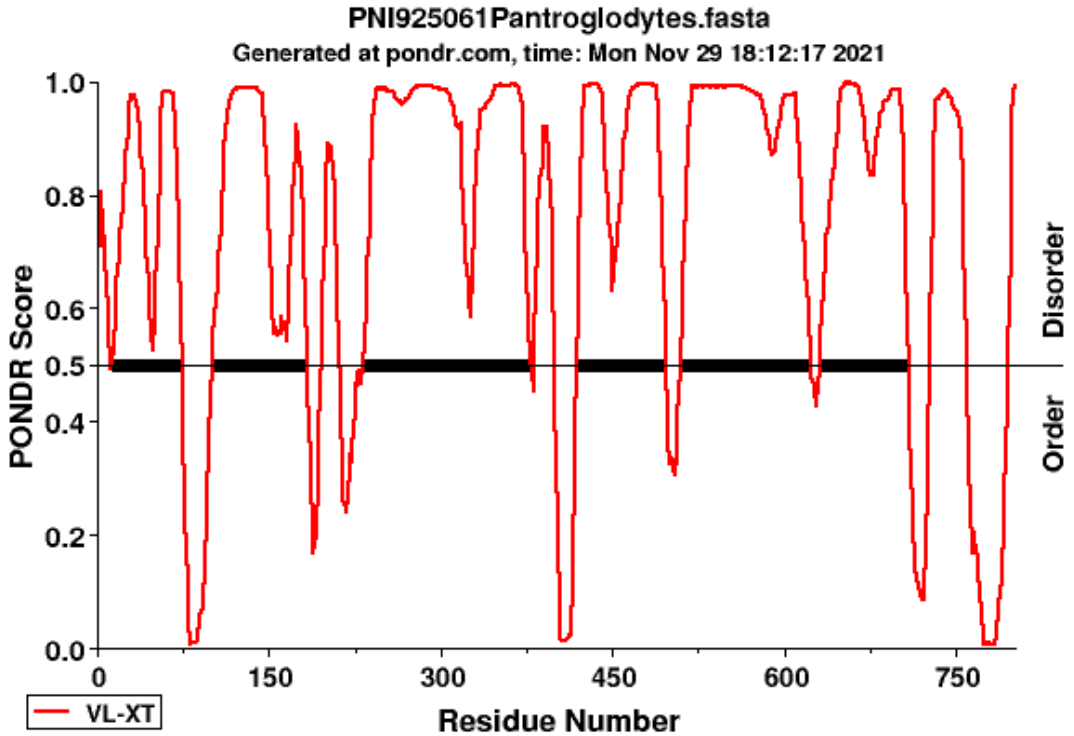
9. Predicted disorder segment [741]-[747] Average Strength= 0.8466
SCGRKRK

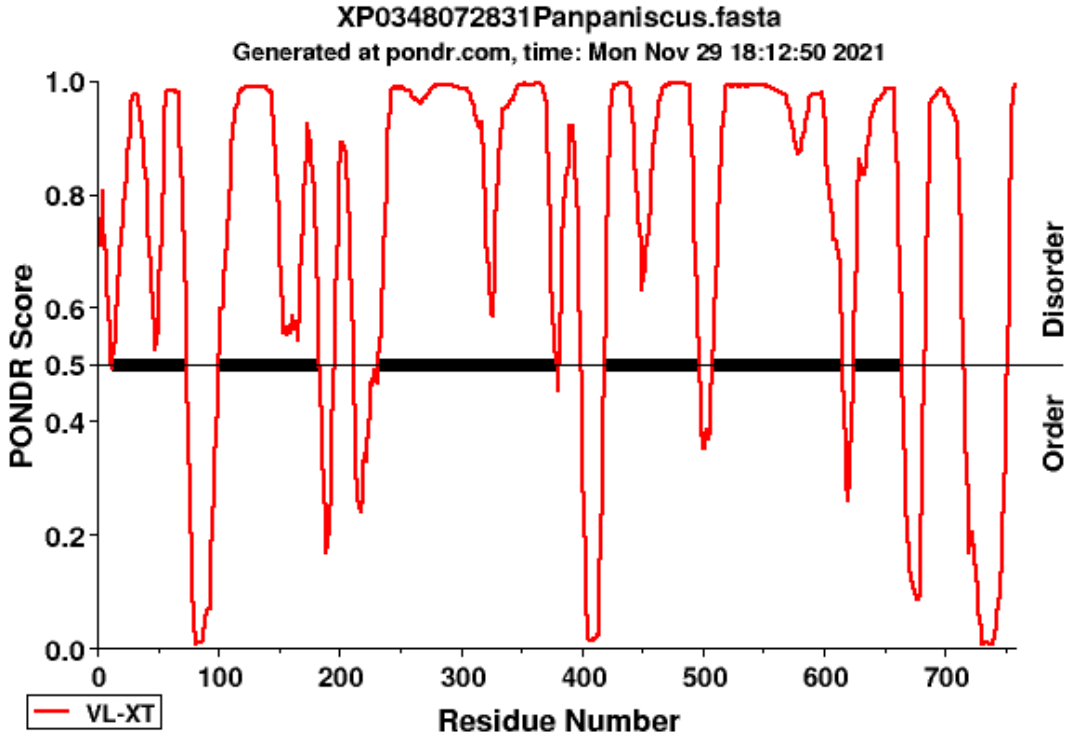
Appendix B

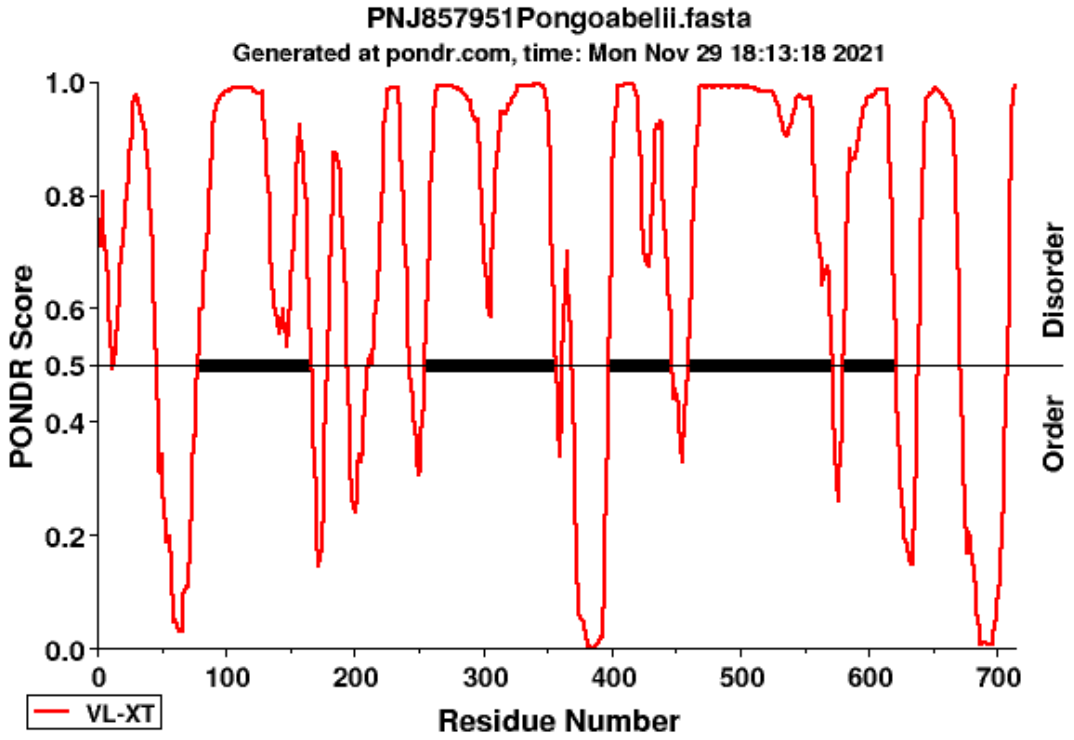
PONDR Result by VL-TX Predictor for Human ELN Disorder Regions

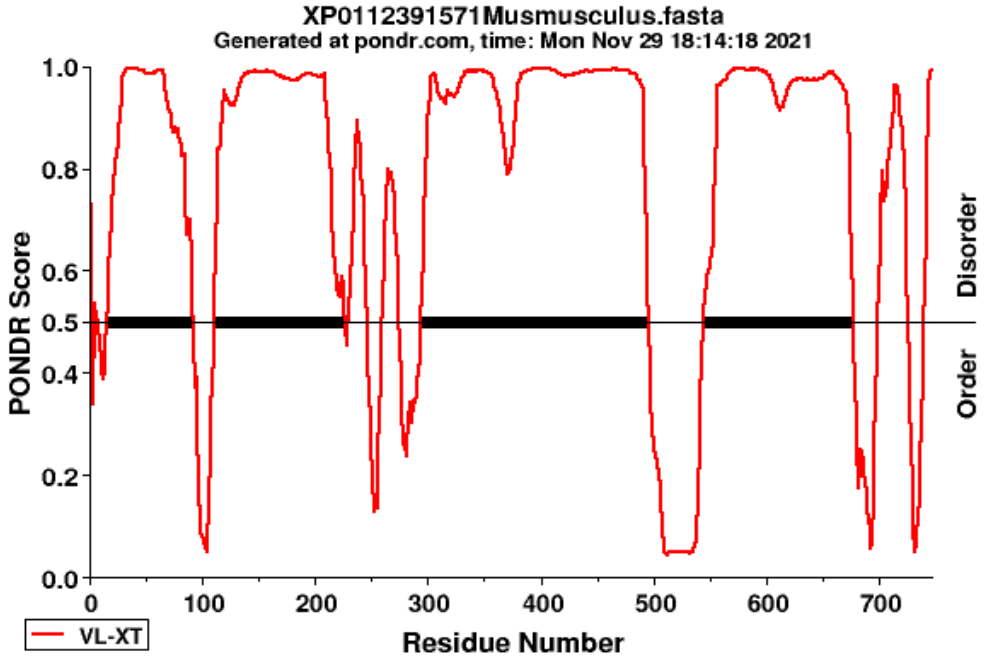
Region	Predicted disorder segment	Average Strength
1	1-11	0.6729
2	13-74	0.822
3	100-183	0.8082
4	196-211	0.7511
5	228-259	0.7968
6	273-375	0.9031
7	378-395	0.7828
8	416-492	0.9002
9	506-607	0.9203
10	616-692	0.8969
11	712-743	0.8974
12	780-786	0.8173
Predicted residues: 786	Number Disordered Regions: 12	Number residues disordered: 621
Longest Disordered Region: 103	Overall percent disordered: 79.01	Average Prediction Score: 0.7345

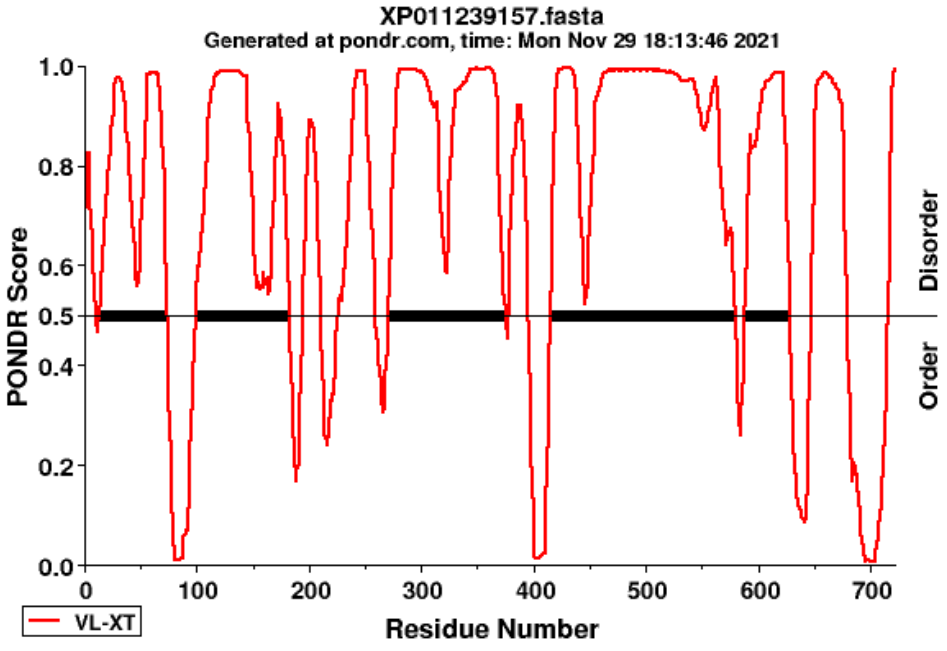

P155024Human.fasta
Generated at pondr.com, time: Mon Nov 29 18:11:51 2021
PONDR Score
1.0
0.8
0.6
0.5
0.4
0.2
0.0
Disorder
Order
0
100
200
300
400
500
600
700
Residue Number
VL-XT


PNI925061Pantroglodytes.fasta
Generated at pondr.com, time: Mon Nov 29 18:12:17 2021
PONDR Score
1.0
0.8
0.6
0.5
0.4
0.2
0.0
Disorder
Order
0
150
300
450
600
750
Residue Number
VL-XT


XP0348072831Panpaniscus.fasta
Generated at pondr.com, time: Mon Nov 29 18:12:50 2021
PONDR Score
Residue Number
Disorder
Order
VL-XT
1.0
0.8
0.6
0.5
0.4
0.2
0.0
0
100
200
300
400
500
600
700


PNJ857951Pongoabelii.fasta
Generated at pondr.com, time: Mon Nov 29 18:13:18 2021
PONDR Score
Residue Number
Disorder
Order
VL-XT
1.0
0.8
0.6
0.5
0.4
0.2
0.0
0
100
200
300
400
500
600
700


XP0112391571Musmusculus.fasta
Generated at pondr.com, time: Mon Nov 29 18:14:18 2021
PONDR Score
Disorder
Order
Residue Number
VL-XT


XP011239157.fasta
Generated at pondr.com, time: Mon Nov 29 18:13:46 2021
PONDR Score
Disorder
Order
Residue Number
VL-XT

Appendix C
Statistical Analysis of ELN Amino Acids Sequence in Disorder/Order Regions

Homo sapiens	Sequence	region	Length	V count	%V content	P count	%P content	G count	%G content	VP count	%VP content	PG count	%PG content	VPG count	%VPG content	VPGXG count	%VPGXG content
	ELN	1-786	786	98	12.47%	100	12.72%	225	28.63%	29	7.38%	63	16.03%	25	9.54%	16	10.18%
Disorder Regions	1	1-11	11	0	0.00%	2	18.18%	1	9.09%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
	2	13-74	62	5	8.06%	10	16.13%	20	32.26%	3	9.68%	7	22.58%	3	14.52%	0	0.00%
	3	100-183	84	15	17.86%	13	15.48%	25	29.76%	4	9.52%	10	23.81%	2	7.14%	2	11.90%
	4	196-211	16	1	6.25%	5	31.25%	4	25.00%	1	12.50%	1	12.50%	0	0.00%	0	0.00%
	5	228-259	32	2	6.25%	3	9.38%	10	31.25%	0	0.00%	1	6.25%	0	0.00%	0	0.00%
	6	273-375	103	17	16.50%	16	15.53%	29	28.16%	8	15.53%	13	25.24%	8	23.30%	4	19.42%
	7	378-395	18	2	11.11%	2	11.11%	4	22.22%	0	0.00%	1	11.11%	0	0.00%	0	0.00%
	8	416-492	77	13	16.88%	10	12.99%	23	29.87%	7	18.18%	8	20.78%	7	27.27%	5	32.47%
	9	506-607	102	22	21.57%	12	11.76%	32	31.37%	3	5.88%	11	21.57%	2	5.88%	2	9.80%
	10	616-692	77	7	9.09%	9	11.69%	15	19.48%	2	5.19%	3	7.79%	2	7.79%	0	0.00%
	11	712-743	32	3	9.38%	3	9.38%	11	34.38%	1	6.25%	1	6.25%	1	9.38%	1	15.63%
	12	780-786	7	0	0.00%	0	0.00%	1	14.29%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
	Mean		51.75	7.25	10%	7.08	14%	14.58	26%	2.41	7%	4.66	13%	2.08	8%	1.16	7%
	Std Dev		36.22	7.55	0.06	5.21	0.07	11.06	0.07	2.74	0.06	4.81	0.09	2.74	0.09	1.74	0.10

Ordered Regions	1	75-99	25	3	12.00%	3	12.00%	6	24.00%	1	8.00%	2	16.00%	1	12.00%	0	0.00%
	2	396-415	20	3	15.00%	2	10.00%	9	45.00%	0	0.00%	1	10.00%	0	0.00%	0	0.00%
	3	744-779	36	2	5.56%	4	11.11%	12	33.33%	0	0.00%	2	11.11%	0	0.00%	0	0.00%
	Mean		27	2.66	11%	3	11%	9	34%	0.33	3%	1.66	12%	0.33	4%	0	0%
	Std Dev		8.18	0.57	0.04	1	0.01	3	0.10	0.57	0.04	0.57	0.03	0.57	0.06	0	0
		Disorder per Order	1.91	2.71	0.94	2.36	1.23	1.62	0.75	7.25	2.59	2.8	1.06	6.25	1.99	--	--

Pantroglodytes	**Sequence**	**region**	**Length**	**V count**	**%V content**	**P count**	**%P content**	**G count**	**%G content**	**VP count**	**%VP content**	**PG count**	**%PG content**	**VPG count**	**%VPG content**	**VPGXG count**	**%VPGXG content**
	ELN	1-802	802.00	106.00	0.13	103.00	0.13	225.00	0.28	31.00	0.08	68.00	0.17	27.00	0.10	15.00	0.09
Disorder Regions	1	44207.00	11.00	0.00	0.00	2.00	0.18	1.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	13-73	61.00	7.00	0.11	10.00	0.16	17.00	0.28	3.00	0.10	7.00	0.23	2.00	0.10	0.00	0.00
	3	100-183	84.00	15.00	0.18	13.00	0.15	25.00	0.30	4.00	0.10	10.00	0.24	2.00	0.07	2.00	0.12
	4	196-211	15.00	1.00	0.07	5.00	0.33	4.00	0.27	1.00	0.13	1.00	0.13	0.00	0.00	0.00	0.00
	5	232-379	147.00	20.00	0.14	3.00	0.02	40.00	0.27	8.00	0.11	14.00	0.19	8.00	0.16	4.00	0.14
	6	382-399	18.00	2.00	0.11	16.00	0.89	5.00	0.28	0.00	0.00	1.00	0.11	0.00	0.00	0.00	0.00
	7	420-496	77.00	13.00	0.17	2.00	0.03	22.00	0.29	7.00	0.18	7.00	0.18	6.00	0.23	4.00	0.26
	8	510-623	114.00	27.00	0.24	10.00	0.09	36.00	0.32	3.00	0.05	13.00	0.23	2.00	0.05	2.00	0.09
	9	632-708	77.00	7.00	0.09	12.00	0.16	15.00	0.19	2.00	0.05	3.00	0.08	2.00	0.08	0.00	0.00

	10	728-759	32.00	3.00	0.09	9.00	0.28	10.00	0.31	1.00	0.06	1.00	0.06	1.00	0.09	1.00	0.16
	11	796-802	7.00	0.00	0.00	3.00	0.43	1.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean		58.45	8.64	0.11	7.73	0.25	16.00	0.25	2.64	0.07	5.18	0.13	2.09	0.07	1.18	0.07
	Std Dev		46.26	9.03	0.07	4.94	0.25	13.59	0.07	2.77	0.06	5.29	0.09	2.63	0.08	1.60	0.09
Ordered Regions	1	74-99	26.00	3.00	0.12	3.00	0.12	7.00	0.24	1.00	0.08	2.00	0.16	1.00	0.12	0.00	0.00
	2	400-419	20.00	3.00	0.15	2.00	0.10	9.00	0.45	0.00	0.00	1.00	0.10	0.00	0.00	0.00	0.00
	3	760-795	36.00	2.00	0.06	4.00	0.11	12.00	0.33	0.00	0.00	2.00	0.11	0.00	0.00	0.00	0.00
	Mean		27.33	2.67	0.11	3.00	0.11	9.33	0.34	0.33	0.03	1.67	0.12	0.33	0.04	0.00	0.00
	Std Dev		8.08	0.58	0.05	1.00	0.01	2.52	0.11	0.58	0.05	0.58	0.03	0.58	0.07	0.00	0.00
		Disorder per Order	2.14	3.24	1.00	2.58	2.24	1.71	0.73	7.91	2.68	3.11	1.07	6.27	1.80	--	--

Panpanis cus	**Sequence**	**region**	**Length**	**V count**	**%V content**	**P count**	**%P content**	**G count**	**%G content**	**VP count**	**%VP content**	**PG count**	**%PG content**	**VPG count**	**%VPG content**	**VPGXG count**	**%VPGXG content**
	ELN	1-758	758.00	102.00	0.13	95.00	0.13	218.00	0.29	30.00	0.08	66.00	0.17	26.00	0.10	14.00	0.09
Disorder Regions	1	44207.00	11.00	0.00	0.00	2.00	0.18	1.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	13-73	61.00	7.00	0.11	10.00	0.16	17.00	0.28	3.00	0.10	7.00	0.23	3.00	0.15	0.00	0.00
	3	100-183	84.00	15.00	0.18	13.00	0.15	25.00	0.30	4.00	0.10	10.00	0.24	2.00	0.07	2.00	0.12
	4	196-211	16.00	1.00	0.06	5.00	0.31	4.00	0.25	1.00	0.13	1.00	0.13	0.00	0.00	0.00	0.00

	5	232-379	148.00	20.00	0.14	19.00	0.13	40.00	0.27	9.00	0.12	14.00	0.19	8.00	0.16	4.00	0.14
	6	382-399	18.00	2.00	0.11	2.00	0.11	5.00	0.28	0.00	0.00	1.00	0.11	0.00	0.00	0.00	0.00
	7	420-497	78.00	13.00	0.17	10.00	0.13	22.00	0.28	7.00	0.18	7.00	0.18	6.00	0.23	4.00	0.26
	8	509-615	107.00	23.00	0.21	13.00	0.12	35.00	0.33	3.00	0.06	13.00	0.24	3.00	0.08	3.00	0.14
	9	625-664	40.00	5.00	0.13	3.00	0.08	12.00	0.30	1.00	0.05	2.00	0.10	1.00	0.08	0.00	0.00
	10	684-715	32.00	3.00	0.09	3.00	0.09	10.00	0.31	1.00	0.06	1.00	0.06	1.00	0.09	1.00	0.16
	11	752-758	7.00	0.00	0.00	0.00	0.00	1.00	0.14	0.00	0.00	1.00	0.29	0.00	0.00	0.00	0.00
	Mean		54.73	8.09	0.11	7.27	0.13	15.64	0.26	2.64	0.07	5.18	0.16	2.18	0.08	1.27	0.07
	Std Dev		45.43	8.31	0.07	6.07	0.08	13.49	0.07	3.01	0.06	5.25	0.09	2.68	0.08	1.68	0.09
Ordered Regions	1	74-99	26.00	3.00	0.12	3.00	0.12	7.00	0.27	1.00	0.08	2.00	0.16	1.00	0.12	0.00	0.00
	2	400-419	20.00	3.00	0.15	2.00	0.10	9.00	0.45	0.00	0.00	1.00	0.10	0.00	0.00	0.00	0.00
	3	716-751	36.00	2.00	0.06	4.00	0.11	12.00	0.33	0.00	0.00	2.00	0.11	0.00	0.00	0.00	0.00
	Mean		27.33	2.67	0.11	3.00	0.11	9.33	0.35	0.33	0.03	1.67	0.12	0.33	0.04	0.00	0.00
	Std Dev		8.08	0.58	0.05	1.00	0.01	2.52	0.09	0.58	0.04	0.58	0.03	0.58	0.07	0.00	0.00
		Disorder per Order	2.00	3.03	1.02	2.42	1.23	1.68	0.73	7.91	2.79	3.11	1.30	6.55	1.97	--	--

Pongoabelii	Sequence	region	Length	V count	%V content	P count	%P content	G count	%G content	VP count	%VP content	PG count	%PG content	VPG count	%VPG content	VPGXG count	%VPGXG content
	ELN	1-714	714.00	98.00	0.14	90.00	0.13	202.00	0.28	28.00	0.08	60.00	0.17	23.00	0.10	12.00	0.08
Disorder Regions	1	44207.00	11.00	0.00	0.00	2.00	0.18	1.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	13-45	33.00	5.00	0.15	6.00	0.18	7.00	0.21	2.00	0.12	4.00	0.24	2.00	0.18	0.00	0.00
	3	78-166	89.00	15.00	0.17	13.00	0.15	28.00	0.31	4.00	0.09	10.00	0.22	2.00	0.07	2.00	0.11
	4	179-194	16.00	1.00	0.06	5.00	0.31	3.00	0.19	1.00	0.13	1.00	0.13	0.00	0.00	0.00	0.00
	5	211-242	32.00	2.00	0.06	3.00	0.09	9.00	0.28	0.00	0.00	1.00	0.06	0.00	0.00	4.00	0.63
	6	256-356	101.00	17.00	0.17	16.00	0.16	29.00	0.29	8.00	0.16	13.00	0.26	8.00	0.24	0.00	0.00
	7	362-369	7.00	0.00	0.00	0.00	0.00	1.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	4.00	2.86
	8	399-447	49.00	8.00	0.16	6.00	0.12	10.00	0.20	3.00	0.12	2.00	0.08	2.00	0.12	3.00	0.31
	9	460-571	112.00	28.00	0.25	13.00	0.12	35.00	0.31	4.00	0.07	13.00	0.23	4.00	0.11	0.00	0.00
	10	581-621	41.00	6.00	0.15	3.00	0.07	12.00	0.29	1.00	0.05	2.00	0.10	1.00	0.07	1.00	0.12
	11	639-671	33.00	4.00	0.12	3.00	0.09	12.00	0.36	2.00	0.12	1.00	0.06	1.00	0.09	0.00	0.00
	12	708-714	7.00	0.00	0.00	0.00	0.00	1.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00		
	Mean		44.25	7.17	0.11	5.83	0.12	12.33	0.24	2.08	0.08	4.27	0.13	1.82	0.08	1.27	0.37
	Std Dev		36.84	8.67	0.08	5.34	0.08	11.89	0.08	2.41	0.06	5.14	0.10	2.40	0.08	1.68	0.85
Ordered Regions	1	370-398	29.00	5.00	0.17	3.00	0.10	13.00	0.45	0.00	0.00	1.00	0.16	o	0.12	0.00	0.00
	2	672-707	36.00	2.00	0.06	4.00	0.11	12.00	0.33	0.00	0.00	2.00	0.10	0.00	0.00	0.00	0.00

	Mean		32.50	3.50	0.11	3.50	0.11	28.00	0.39	0.00	0.00	1.50	0.13	0.00	0.06	0.00	0.00
	Std Dev		4.95	2.12	0.08	0.71	0.01	0.71	0.08	0.00	0.00	0.71	0.04	#DIV/0!	0.08	0.00	0.00
		Disorder per Order	1.36	2.05	0.95	1.67	1.15	0.44	0.60	#DIV/0!	#DIV/0!	2.85	0.97	#DIV/0!	1.33	--	--

Gorilla	**Sequence**	**region**	**Length**	**V count**	**%V content**	**P count**	**%P content**	**G count**	**%G content**	**VP count**	**%VP content**	**PG count**	**%PG content**	**VPG count**	**%VPG content**	**VPGXG count**	**%VPGXG content**
	ELN	1-722	722.00	102.0	0.14	92.00	0.13	207.00	0.29	29.00	0.08	64.00	0.18	24.00	0.10	12.00	0.08
Disorder Regions	1	44205.00	9.00	0.00	0.00	1.00	0.11	1.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	13-73	61.00	5.00	0.08	10.00	0.16	19.00	0.31	3.00	0.10	7.00	0.23	3.00	0.15	0.00	0.00
	3	99-182	84.00	15.00	0.18	13.00	0.15	25.00	0.30	4.00	0.10	10.00	0.24	2.00	0.07	2.00	0.12
	4	195-210	16.00	1.00	0.06	5.00	0.31	4.00	0.25	1.00	0.13	1.00	0.13	0.00	0.00	0.00	0.00
	5	227-258	32.00	2.00	0.06	3.00	0.09	10.00	0.31	0.00	0.00	1.00	0.06	0.00	0.00	0.00	0.00
	6	272-375	104.00	17.00	0.16	16.00	0.15	27.00	0.26	8.00	0.15	13.00	0.25	8.00	0.23	4.00	0.19
	7	378-395	18.00	2.00	0.11	1.00	0.06	5.00	0.28	0.00	0.00	1.00	0.11	0.00	0.00	0.00	0.00
	8	416-579	164.00	38.00	0.23	20.00	0.12	49.00	0.30	7.00	0.09	18.00	0.22	6.00	0.11	5.00	0.15
	9	589-628	40.00	6.00	0.15	3.00	0.08	11.00	0.28	2.00	0.10	2.00	0.10	2.00	0.15	0.00	0.00
	10	648-679	32.00	4.00	0.13	3.00	0.09	10.00	0.31	2.00	0.13	1.00	0.06	1.00	0.09	1.00	0.16

	11	716-722	7.00	0.00	0.00	0.00	0.00	1.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean		51.55	8.18	0.11	6.82	0.12	14.73	0.26	2.45	0.07	4.91	0.13	2.00	0.07	1.09	0.06
	Std Dev		48.60	11.44	0.07	6.84	0.08	14.47	0.07	2.84	0.06	6.20	0.09	2.72	0.08	1.81	0.08
Ordered Regions	1	74-98	25.00	3.00	0.12	3.00	0.12	6.00	0.24	1.00	0.08	2.00	0.16	1.00	0.12	0.00	0.00
	2	396-415	20.00	3.00	0.15	2.00	0.10	9.00	0.45	0.00	0.00	1.00	0.10	0.00	0.00	0.00	0.00
	3	680-715	36.00	2.00	0.06	4.00	0.11	12.00	0.33	0.00	0.00	2.00	0.11	0.00	0.00	0.00	0.00
	Mean		27.00	2.67	0.11	3.00	0.11	9.00	0.34	0.33	0.03	1.67	0.12	0.33	0.04	0.00	0.00
	Std Dev		8.19	0.58	0.05	1.00	0.01	3.00	0.11	0.58	0.05	0.58	0.03	0.58	0.07	0.00	0.00
		Disorder per Order	1.91	3.07	0.98	2.27	1.10	1.64	0.76	7.36	2.67	2.95	1.03	6.00	1.83	--	--

Musmusculus	**Sequence**	**region**	**Length**	**V count**	**%V content**	**P count**	**%P content**	**G count**	**%G content**	**VP count**	**%VP content**	**PG count**	**%PG content**	**VPG count**	**%VPG content**	**VPGXG count**	**%VPGXG content**
	ELN	1-747	747.00	64.00	0.09	81.00	0.11	273.00	0.37	27.00	0.07	55.00	0.15	22.00	0.09	8.00	0.05

Disorder Region

	1	44293.00	4.00	1.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	15-91	77.00	4.00	0.05	12.00	0.16	31.00	0.40	3.00	0.08	9.00	0.23	3.00	0.12	0.00	0.00
	3	111-226	116.00	22.00	0.19	12.00	0.10	48.00	0.41	7.00	0.12	11.00	0.19	6.00	0.16	1.00	0.04
	4	230-245	16.00	2.00	0.13	3.00	0.19	5.00	0.31	1.00	0.13	2.00	0.25	0.00	0.00	0.00	0.00
	5	260-274	15.00	1.00	0.07	4.00	0.27	3.00	0.20	1.00	0.13	1.00	0.13	0.00	0.00	0.00	0.00
	6	295-495	201.00	17.00	0.08	19.00	0.09	66.00	0.33	7.00	0.07	14.00	0.14	6.00	0.09	2.00	0.05
	7	545-677	133.00	4.00	0.03	8.00	0.06	53.00	0.40	2.00	0.03	5.00	0.08	2.00	0.05	0.00	0.00
	8	700-725	26.00	1.00	0.04	5.00	0.19	11.00	0.42	0.00	0.00	1.00	0.08	0.00	0.00	3.00	0.58
	9	741-747	7.00	0.00	0.00	0.00	0.00	1.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean		66.11	5.78	0.09	7.00	0.12	24.22	0.29	2.33	0.06	4.78	0.12	1.89	0.05	0.67	0.07
	Std Dev		70.14	8.00	0.08	6.34	0.09	25.75	0.15	2.83	0.06	5.29	0.09	2.57	0.06	1.12	0.19

Appendix D
AACR Summary Table for Six Species

Sequence	NCBI Reference Sequence	Length	V AACR*	P AACR*	G AACR*	VP AACR*	PG AACR*	VPG AACR*	VPGXG count
Human	P155024	786.00	0.94	1.23	0.75	2.59	1.06	1.99	16
Pantroglodytes	PNI925061	802.00	1.00	2.24	0.73	2.68	1.07	1.80	15
Panpaniscus	XP0348072831	758.00	1.02	1.23	0.73	2.79	1.30	1.97	14
Pongoabelii	PNJ857951	714.00	0.95	1.15	0.60	∞	0.97	1.33	12
Gorilla	XP_030868570.1	722.00	0.98	1.10	0.76	2.67	1.03	1.83	12
Musmusculus	XP0112391571	747.00	NA	NA	NA	NA	NA	NA	8
Average		**754.83**	**0.98**	**1.39**	**0.72**	**2.68**	**1.08**	**1.78**	12.83
standard deviation		34.68	0.03	0.48	0.06	0.09	0.12	0.26	2.86

*** Average AACR of all disordered regions per average AACR of all ordered regions in each species**
Average reflected in figure 2.6

Appendix E
ELP[AV-60] and ELP[V-60] Amino Acids Sequences

ELP[AV-60]:

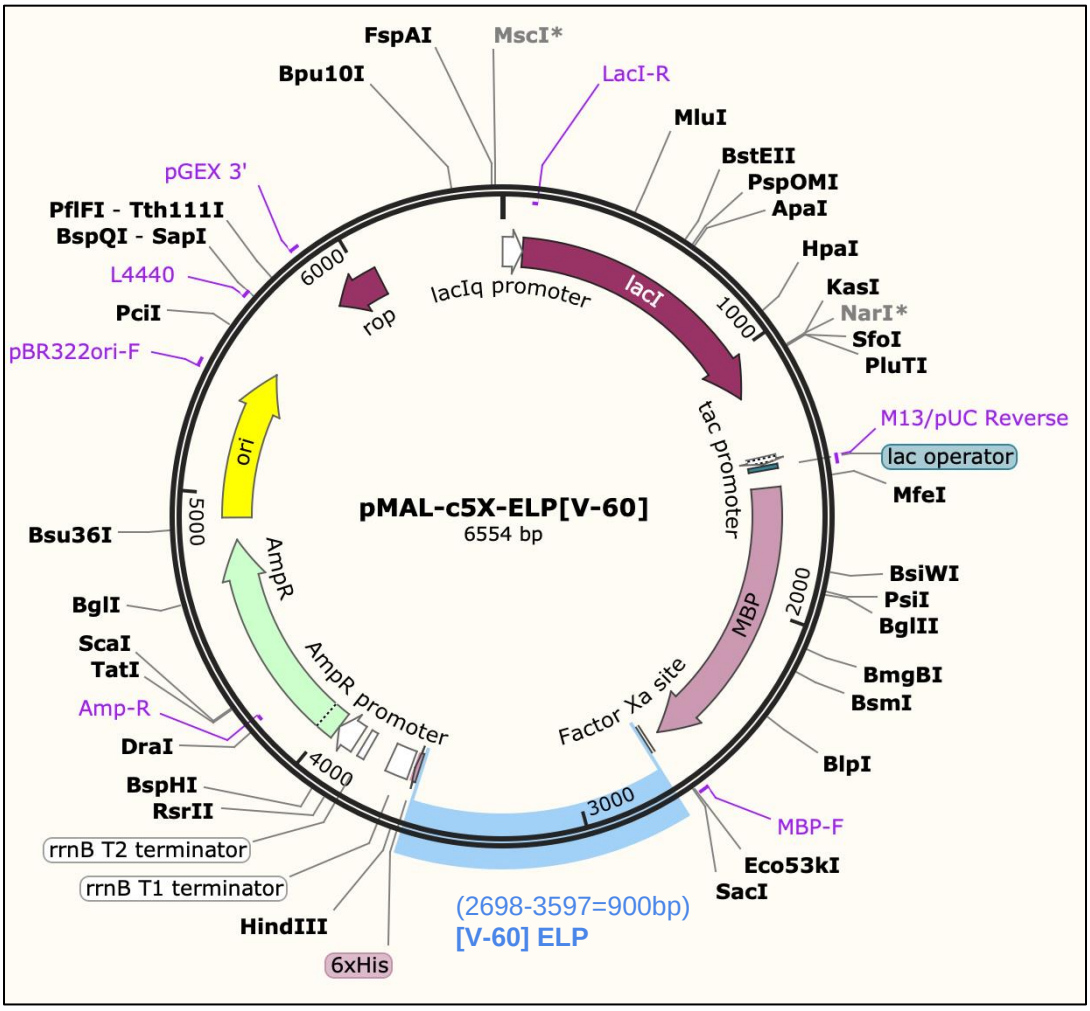
VPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVP
GAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPG
VGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGA
GVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVG
VPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVPGVGVPGAGVP
GVGVPGAGVPGVGVPGAG

ELP[V-60]:

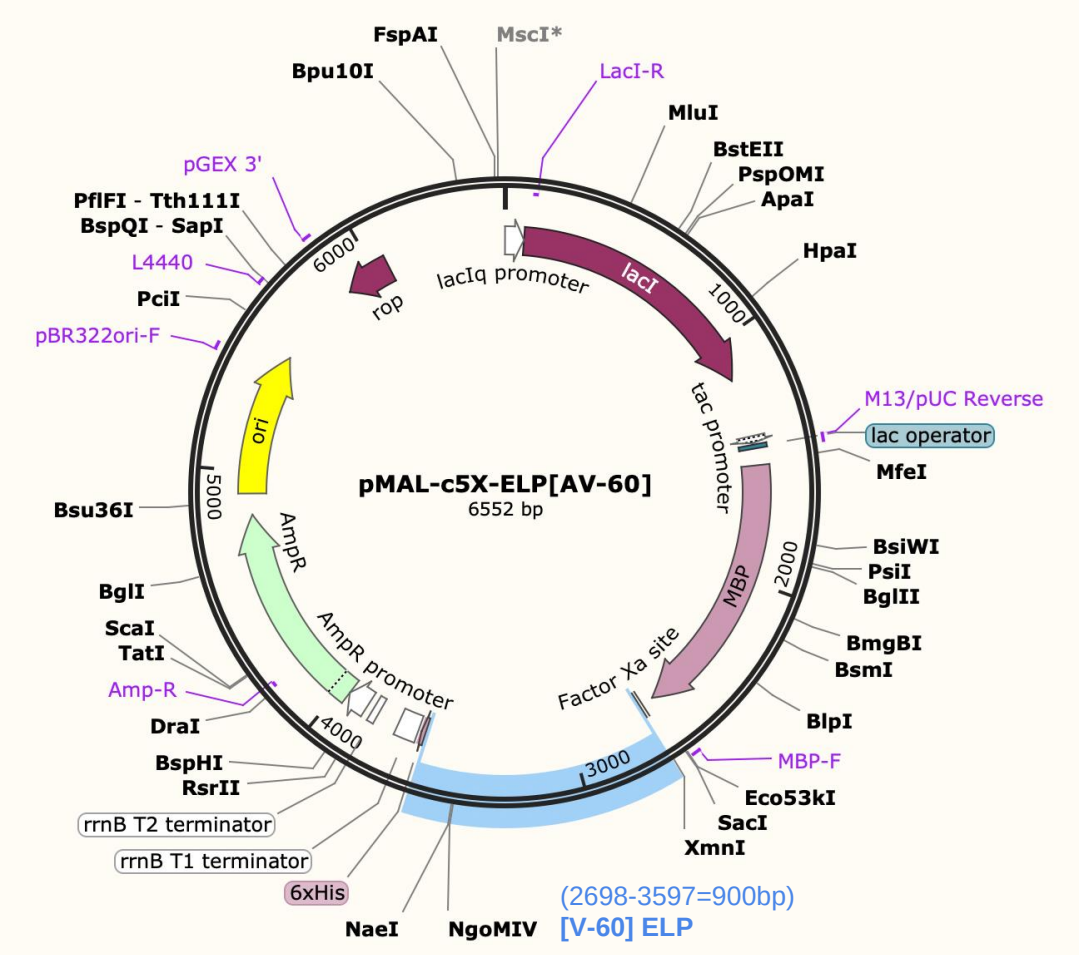
VPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVP
GVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPG
VGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGV
GVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVG
VPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVP
GVGVPGVGVPGVGVPGVG

Appendix F
Plasmids Map

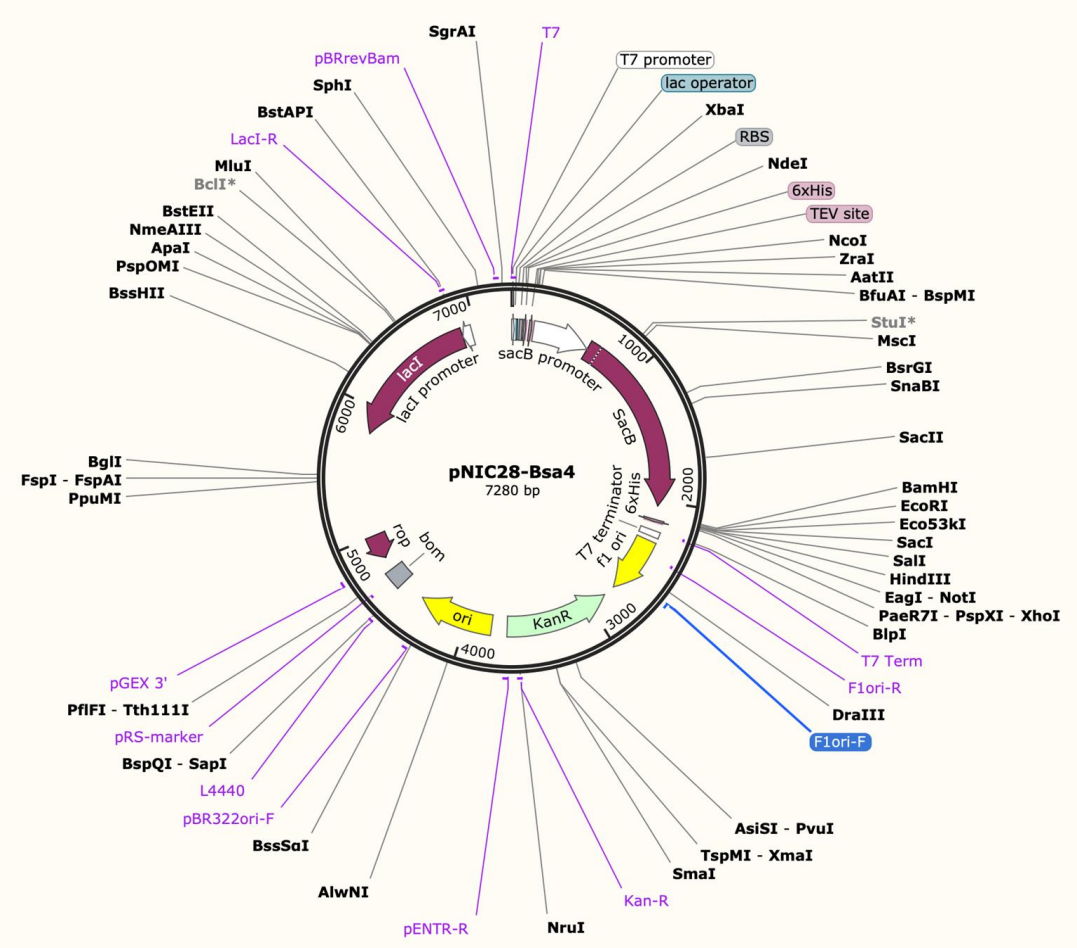
pMAL-c5X-ELP[V-60]



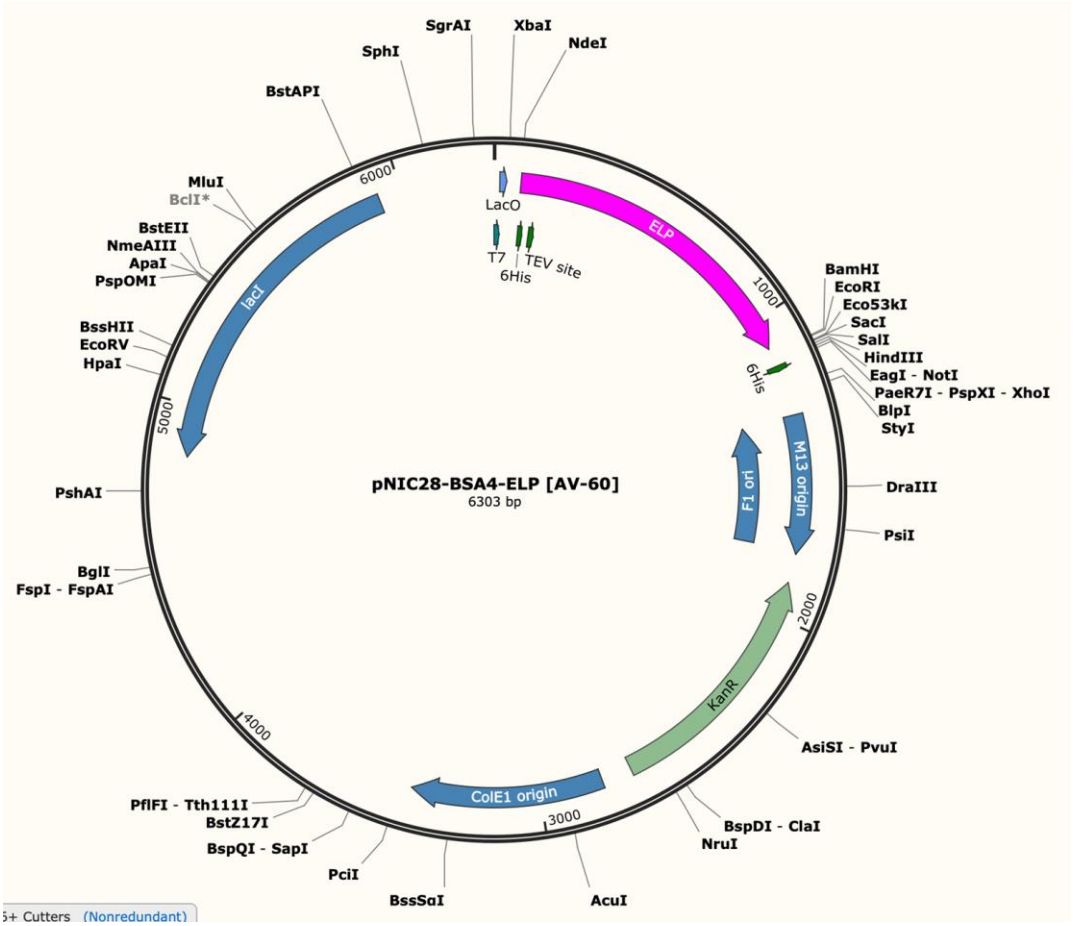
pMAL-c5X-ELP[AV-60]:



pNIC28-Bsa4:


pNIC28-Bsa4
7280 bp
SgrAI
T7
pBRrevBam
T7 promoter
SphI
lac operator
BstAPI
XbaI
LacI-R
RBS
MluI
NdeI
BclI*
6xHis
BstEII
TEV site
NmeAIII
NcoI
ApaI
ZraI
PspOMI
AatII
BssHII
BfuAI - BspMI
StuI*
MscI
BsrGI
SnaBI
SacII
BglI
FspI - FspAI
PpuMI
BamHI
EcoRI
Eco53kI
SacI
SalI
HindIII
EagI - NotI
PaeR7I - PspXI - XhoI
BlpI
T7 Term
F1ori-R
DraIII
F1ori-F
AsiSI - PvuI
TspMI - XmaI
SmaI
Kan-R
NruI
pENTR-R
AlwNI
BssSαI
pBR322ori-F
L4440
BspQI - SapI
pRS-marker
PflFI - Tth111I
pGEX 3'
lacI
lacI promoter
sacB promoter
SacB
6xHis
T7 terminator
f1 ori
rop
bom
ori
KanR
1000
2000
3000
4000
5000
6000
7000

pNIC28-BSA4-ELP[AV-60]:



Appendix G
Average Hydrophobicity of Three Most Recent Hypothesis

Amino $Acid	Amino acid scale: Hydrophobicity of physiological L-alpha amino acids(1991)	Amino acid scale: Hydrophobicity indices at ph 7.5 determined by HPLC.(1990)	Amino acid scale: Hydrophobicity indices at ph 3.4 determined by HPLC.(1990)	Average
Arg	0.616	0.35	0.42	0.462
Asp	0	-1.5	-1.56	-1.02
Glu	0.236	-0.99	-1.03	-0.5946666667
Lys	0.028	-2.15	-0.51	-0.8773333333
Asn	0.68	0.76	0.84	0.76
His	0.251	-0.93	-0.96	-0.5463333333
Gln	0.043	-1.95	-0.37	-0.759
Thr	0.501	0	0	0.167
Ser	0.165	-0.65	-2.28	-0.9216666667
Gly	0.943	1.83	1.81	1.527666667
Tyr	0.943	1.8	1.8	1.514333333
Pro	0.283	-1.54	-2.03	-1.095666667
Trp	0.738	1.1	1.18	1.006
Met	1	1.69	1.74	1.476666667
Cys	0.711	0.84	0.86	0.8036666667

Ala	0.359	-0.63	-0.64	-0.3036666667
Phe	0.45	-0.27	-0.26	-0.02666666667
Val	0.878	1.35	1.46	1.229333333
Leu	0.88	0.39	0.51	0.5933333333
Ile	0.825	1.32	1.34	1.161666667

Table. Average hydrophobicity of three most recent hypothesis