Protein Engineering for Biomedical Materials

Rachael N. Parker

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Chemistry

Tijana Z. Grove, Committee Chair

Felicia A. Etzkorn

Rich D. Gandour

Brian M. Tissue

Mark E. Van Dyke

February 16, 2017
Blacksburg, VA

Keywords: Protein engineering, biomaterials, leucine-rich repeat, keratin

Protein Engineering for Biomedical Materials

Rachael N. Parker

ABSTRACT

The inherent design freedom of protein engineering and recombinant protein production enables specific tailoring of protein structure, function, and properties. Two areas of research where protein engineering has allowed for many advances in biomedical materials include the design of novel protein scaffolds for molecular recognition, as well as the use of recombinant proteins for production of next generation biomaterials. The main focus of my dissertation was to develop new biomedical materials using protein engineering.

Chapters three and four discuss the engineering of repeat proteins as bio-recognition modules for biomedical sensing and imaging. Chapter three provides an overview of the most recent advances in engineering of repeat proteins in the aforementioned field. Chapter four discusses my contribution to this field. We have designed a *de novo* repeat protein scaffold based on the consensus sequence of the leucine rich repeat (LRR) domain of the NOD family of cytoplasmic innate immune system receptors. Innate immunity receptors have been described as pattern recognition receptors in that they recognize "global features" of a family of pathogens versus one specific antigen. In mammals, two main protein families of such receptors are: extracellular Toll-like receptors (TLRs) and cytoplasmic Nucletide-binding domain- and Leucine-rich Repeat-containing proteins (NLRs). NLRs are defined by their tripartite domain architecture that contains a C-terminal LRR (Leucine Rich Repeat) domain, the nucleotide-binding oligomerization (NACHT) domain, and the N-terminal effector domain. It is proposed that pathogen sensing in NLRs occurs through ligand binding by the LRR domain. Thus, we

hypothesized that LRRs would be suitable for the design of alternative binding scaffolds for use in molecular recognition.

The NOD protein family plays a very important role in innate immunity, and consequently serves as a promising scaffold for design of novel recognition motifs. However, engineering of *de novo* proteins based on the NOD family LRR domain has proven challenging due to problems arising from protein solubility and stability. Consensus sequence design is a protein design tool used to create novel proteins that capture sequence-structure relationships and interactions present in nature in order to create a stable protein scaffold. We implement a consensus sequence design approach to develop proteins based on the LRR domain of NLRs. Using a multiple sequence alignment we analyzed all individual LRRs found in mammalian NLRs. This design resulted in a consensus sequence protein containing two internal repeats and separate N- and C- capping repeats named CLRR2. Using biophysical characterization methods of size exclusion chromatography, circular dichroism, and fluorescence, CLRR2 was found to be a stable, monomeric, and cysteine free scaffold. Additionally, CLRR2, without any affinity maturation, displayed micromolar binding affinity for muramyl dipeptide (MDP), a bacterial cell wall fragment. To our knowledge, this is the first report of direct interaction of a NOD LRR with a physiologically relevant ligand. Furthermore, CLRR2 demonstrated selective recognition to the biologically active stereoisomer of MDP. Results of this study indicate that LRRs are indeed a useful scaffold for development of specific and selective proteins for molecular recognition, creating much potential for future engineering of alternative protein scaffolds for biomedical applications.

My second research interest focused on the development of proteins for novel biomaterials. In the past two decades, keratin biomaterials have shown impressive results as

scaffolds for tissue engineering, wound healing, and nerve regeneration. In addition to its intrinsic biocompatibility, keratin interacts with specific cell receptors eliciting beneficial biochemical cues, as well as participates in important regulatory functions such as cell migration and proliferation and protein signalling. The aforementioned properties along with keratins' inherent capacity for self-assembly poise it as a promising scaffold for regenerative medicine and tissue engineering applications. However, due to the extraction process used to obtain natural keratin proteins from natural sources, protein damage and formation of by-products that alter network self-assembly and bioactivity often occur as a result of the extensive processing conditions required. Furthermore, natural keratins require exogenous chemistry in order to modify their properties, which greatly limits sequence tunability.

Recombinant keratin proteins have the potential to overcome the limitations associated with the use of natural keratins while also maintaining their desired structural and chemical characteristics. Thus, we have used recombinant DNA technology for the production of human hair keratins, keratin 31 (K31) and keratin 81 (K81). The production of recombinant human hair keratins resulted in isolated proteins of the correct sequence and molecular weight determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis and mass spectrometry. Proteins with no unwanted sequence truncations, deletions, or mutations indicate recombinant DNA technology can be used to reliably generate full length keratin proteins. This allows for consistent starting materials with no observable impurities or undesired by-products, which combats a major challenge associated with natural keratins. Additionally, recombinant keratins must maintain the intrinsic propensity for self-assembly found in natural keratins. To test the propensity for self-assembly, we implemented size exclusion chromatography (SEC), dynamic light scattering (DLS), and transmission electron microscopy (TEM) to characterize K31, K81,

and an equimolar mixture of K31 and K81. The results of the recombinant protein characterization reveal novel homo-polymerization of K31 and K81, not previously reported, and formation of characteristic keratin fibers for the K31 and K81 mixture. Therefore, recombinant K31 and K81 retain the intrinsic biological activity (i.e. self-assembly) of natural keratin proteins. We have also conducted a comparative study of recombinant and extracted heteropolymer K31/K81. Through solution characterization and TEM analysis it was found that use of the recombinant heteropolymer allows for increased purity of starting material while also maintaining self-assembly properties necessary for functional use in biomaterials design. However, under the processing condition implemented, extracted keratins demonstrated increased efficiency of assembly. Through each study we conclude that recombinant keratin proteins provide a promising solution to overcome the challenges associated with natural protein materials and present an exceptional design platform for generation of new biomaterials for regenerative medicine and tissue engineering.

Protein Engineering for Biomedical Materials

Rachael N. Parker

GENERAL AUDIENCE ABSTRACT

Protein engineering and synthetic protein production enables the creation of new proteins that can perform specific tasks. Many advances in biomedical materials and medical diagnostic tools stem from the use of synthetic proteins. The main focus of my dissertation was to develop new biomedical materials using protein engineering.

In chapters three and four of the dissertation development of synthetic proteins for medical diagnostics is discussed. We have designed artificial protein sensors based on natural innate immunity proteins, which function in the body as the source for recognition of foreign pathogens, such as bacteria and viruses. Our goal was to create synthetic proteins with similar characteristics to the innate immunity receptors for the purpose of sensing bacteria and viruses in the form of a biosensors or medical diagnostic. Through our work we have developed an artificial protein scaffold that can selectively interact with a relevant biological target. This research provides the ground work for future development of proteins that can sense a wide variety of important pathogens and subsequently be manufactured into diagnostic devices.

Our research involving protein design for biomaterials is the focus of chapters five and six of the dissertation. Keratin is a ubiquitous protein found in the human body. It functions as a structural protein and helps create the complex network that makes up skin, hair, and epidermal appendages. We have created synthetic keratin proteins in an effort to fabricate biomaterials that can be used for regenerative medicine and tissue engineering applications. Our strategy allows for development of proteins that can be designed to have characteristics not afforded to naturally

occurring keratin proteins, and thus presents the opportunity to make materials with unique properties and characteristics that may make them more successful in our intended applications of tissue engineering. From our work we have shown that synthetic production of these proteins is possible and that the synthetically produced proteins retain the essential structural and functional properties associated with natural keratin proteins. Thus, this work highlights the potential for use of synthetic proteins for production of biomaterials with new and important features that cannot be obtained through use of natural proteins.

Acknowledgements

To say I had no idea what I was getting myself into when I entered graduate school five and a half years ago, would be the greatest understatement I could make. As I have reflected on my time at Virginia Tech, starting from picking a graduate school, to choosing an advisor, my committee, and finally a research project, I can honestly say I had little idea what I was doing. Thankfully, these decisions turned out to be some of the best ones I have made in my life two in particular include my decision to accept enrollment at Virginia Tech and most importantly to join the Grove research group. Once again, when I began working for Professor Tijana Grove I had little understanding of what I needed to do to be successful in the program. However, through her guidance and constant encouragement, I was able to find my way. I will always appreciate the incredible amount of time and effort Dr. Grove put into helping, teaching, and mentoring me throughout the past several years. As her first student to reach each graduate school milestone, we have experienced many firsts together, and each step of the way I felt her constant support and confidence in me. She has taught me to be a more confident and independent scientist, while at the same time being a wonderful mentor. Her dedication to her students is unmatched, and without her I would not be where I am today.

I am also incredibly thankful to all of my committee members, Professor Felicia Etzkorn, Professor Rich Gandour, Professor Brian Tissue, and Professor Mark Van Dyke. One of my first real experiences with my committee, outside of classes, was sitting in Dr. Gandour's office for two hours as we painstakingly went through my literature review. Although this was not the most pleasant experience, as I left his office that day I had learned a great deal in those two hours. I was also very surprised and grateful that someone would take that much time to help me and truly care about my success. This example exemplifies most interactions with my committee

throughout the years. I have always felt that each of my committee members go above and beyond in taking the time to listen, help, teach, and ensure that I learn and grow from the process. I am incredibly grateful for their support and mentorship throughout my time at Virginia Tech. I would especially like to thank Dr. Van Dyke for his guidance and support as Dr. Grove and I delved into the world of keratin biomaterials over the last two years. As we embarked on this collaboration, we had little idea what we were getting ourselves into. Dr. Van Dyke has not only been a great collaborator, but has also been a great mentor to me and for that I am very thankful.

I am very thankful to the Virginia Tech Chemistry Department for support throughout my graduate career. I am also appreciative of all the faculty and staff that keep our department running. In particular, a special thanks to Ms. Joli Huynh who has always been willing to help and always has the answers we need.

The Grove lab has been a wonderful environment to work in over the past five years. I am very thankful to all of my lab mates for their help, support, and friendship. I am especially thankful for Kristina Roth and her constant support and friendship throughout the years. Also I would like to thank Dr. Ana Mercedes-Camacho for teaching and guiding me throughout my first two years of graduate school. Her willingness to help, no matter how busy she may have been, is something I will always be grateful for and appreciate. Additionally, graduate school would not have been the same if it were not for the friendships I have cultivated along the way with Spencer Ahrenholtz, Eugene Camerino, Kristen Felice, Jeff Foster, Devon Hock, Bryce Kidd, Christina Kim, Evan Margaretta, Jen McCord, Scott Radzinski, Nicole Voelker, Guanyu Wang, and Yunnan Xu.

Without the fantastic faculty in the chemistry department at High Point University, graduate school would not have been possible. I am incredibly thankful to all of my professors from HPU for preparing me for graduate school, and investing their time in me. Professors B. Gray Bowman, Chris Fowler, Todd Knippenberg, Elizabeth McCorquodale, and Aaron Titus provided me with an enriching and personalized undergraduate experience, and I appreciate all of their help and support throughout my four years at HPU. I am especially thankful for Dr. Bowman, my first research advisor, who cultivated an educational and exciting lab experience that initiated my interest in research. During my time at HPU I was fortunate enough to participate in a summer research program (thanks to Dr. McCorquodale) at Wake Forest University under the guidance of Professor Christa Colyer. This experience and the mentorship and encouragement of Dr. Colyer is what ultimately led me to attend graduate school. I am very grateful to her for that opportunity, as well as her guidance and support throughout the summer.

Finally, and most importantly, I would like to thank my parents Cliff and Teresa Parker for all they have done for me. If not for them, none of this would be possible. They have supported and pushed me in everything I have done, and I would not be where I am today without them. I am also thankful for my sisters Rebekah and Elizabeth Parker for their love and support, particularly, Rebekah for all the time and guidance she has provided me with over the course of our lives.

Attributions

Chapters 3, 4, 5, and 6 of dissertation were written using published manuscripts (Chapters 3 and 4) or manuscripts in preparation for publication (Chapters 5 and 6) to which the author, Rachael N. Parker, majorly contributed.

Prof. Tijana Z. Grove

Professor of Chemistry at Virginia Tech and research advisor. Principle investigator and supervisor on all research presented in this dissertation. Provided technical insight during experimentation and data analysis, and assisted in writing and editing of all manuscripts.

Prof. Mark E. Van Dyke

Professor of Biomedical Engineering and Mechanics at Virginia Tech. Collaborated on Chapters 5 and 6 and provided technical insight during experimentation and data analysis, as well as assisted in writing and editing of manuscripts for Chapters 5 and 6.

Dr. Ana Mercedes-Camacho

Former research assistant in the Grove Lab who contributed to Chapter 4. Completed design and characterization of generation one LRR proteins.

Kristina L. Roth

Graduate student in the Grove Lab who contributed to Chapters 5 and 6. Assisted in DLS and TEM instrumentation for both chapters.

Alexis Trent

Graduate student in the Van Dyke lab who contributed to Chapter 6. Completed Western blot analysis.

Table of Contents

Chapter 1. Introduction

1.1 Dissertation overview

This dissertation focuses on using protein engineering tools to develop biomedically relevant materials for the design of molecular recognition agents for biosensors, and for the design of protein-based biomaterials for regenerative medicine and tissue engineering.

Chapter two reviews current strategies for generating new molecular recognition agents, as well as common design techniques employed. This chapter will also present an overview of current methods for engineering biomaterials for medical applications, and provide an introduction to the use of recombinant-protein based materials.

The engineering of molecular recognition agents for applications in medical imaging, diagnostic, and sensing applications will be described in chapter three, as well as techniques and strategies for engineering these scaffolds. Chapter four discusses the design of a molecular recognition agent based on leucine-rich repeat proteins. This chapter will detail our consensus design approach to develop molecular recognition modules for recognition of whole cell pathogens through selective binding of pathogen cell wall motifs.

Engineering of recombinant keratin proteins for development of biomaterials will be presented in chapters five and six. This work focuses on the design of human hair keratins for generation of biomaterials with tunable chemical and mechanical properties. Our production of recombinant human hair keratins 31 and 81 will be described, as well as their characterization and comparison to their naturally derived counterparts. Chapter 7 offers overall conclusions for this dissertation and proposed future work.

Chapter 2. Protein Engineering for Biomedical Sensors and Biomaterials

2.1 Abstract

Advances in the field of protein engineering over the last several decades have resulted in a vast array of new technologies for the design of tunable protein sequences with a wide variety of functions. Protein engineering allows for development of multifunctional and tailored systems, as well as creates the ability to expand our knowledge of protein sequence-structure-function relationships. These advantages result in improved designs for diagnostics and sensors, imaging tools, tissue engineering and regenerative medicine, and drug delivery applications. Genetic engineering techniques and recombinant DNA technology provide an avenue for design and engineering of proteins with controlled, specific properties. Consequently protein-based technologies are becoming increasingly used for development of biomedical materials. The work detailed here focuses on two examples of protein engineering, one for diagnostic and sensing applications and the second for tissue engineering and regenerative medicine materials design.

2.2. Introduction to Proteins for Molecular Recognition

Biosensors are beneficial tools in environmental, clinical, and chemical analysis.[1] They consist of three main components: a molecular recognition agent, a transducer, and a signal readout (Figure 2.1). The molecular recognition device interacts with a target analyte producing a response. The transducer then converts the response into a quantifiable signal.[1] Each component contributes to the overall biosensor's function; however, specificity and selectivity of biorecognition limits the effectiveness of the biosensor. This work focuses on molecular recognition modules because they are the critical component of a biosensor system. Modules for molecular recognition, as well as methods for their design will be discussed.

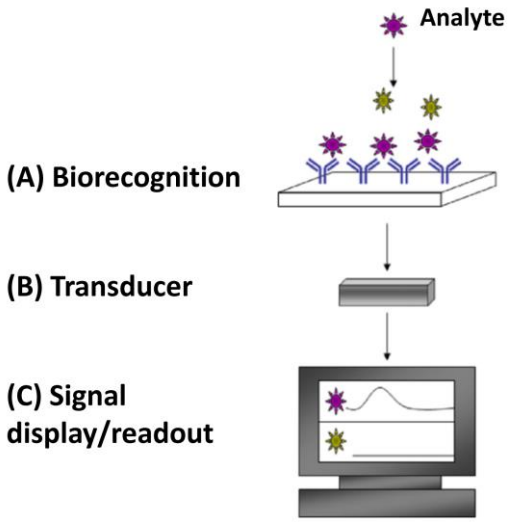


Figure 2.1. Three main components of biosensors: (A) molecular recognition agent; (B) signal transducer. (C) signal readout.[1] (Reprinted from Seminars in cell and developmental biology, 20/1, Conroy, P. J.; Hearty, S.; Leonard, P.; O'Kennedy, R. J., Antibody production, design and use for biosensor-based applications, 10-26, Copyright 2009, with permission from Elsevier)

2.2.1 Biosensors

Biosensors can be divided into three main classes: genetically encodable biosensors, enzymatic biosensors, and molecular recognition based biosensors.[2-11] Each of these types of sensors provides a unique approach to molecular detection. The following sections will provide background on each, with emphasis on molecular recognition-based sensors, which is the scope of the work presented in the following two chapters.

2.2.1.1 Genetically Encoded Biosensors

Genetically encodable biosensors provide valuable information about cell activity through monitoring of fluorescence.[6, 11] Plasmid DNA, encoding for the sensor, is incorporated into cells and subsequently expressed as a protein that is the working sensor within the cell.[6-7]

These proteins can be developed to target specific molecules and cellular locations with great spatial and temporal resolution.[7, 10] Examples of genetically encodable biosensors include lanthanide binding tags, tetra-cysteine tags, and fluorescent proteins (Figure 2.2 A-2.2 C).

Lanthanide-tagged proteins generate insight into cellular activity and function proving particularly useful for large proteins.[3, 12] Chemical coupling through cysteine residues allows for incorporation of lanthanide ions into proteins by chemical means. Alternatively, lanthanide binding tags (LBTs) can be incorporated into the protein on the gene level (Figure 2.2 A).[3, 13] LBTs provide the advantage of a long decay time of luminescence.[9]

The tetra-cysteine tag is incorporated into the target protein and produces a fluorescence signal upon binding to the biarsenical dye (Figure 2.2 B).[2] The comparatively small size of tetra-cysteine tags proves advantageous over the larger fluorescent proteins because they are less likely to interrupt the protein's function.[8] However, a potential disadvantage of this method is the toxicity of biarsenical dyes due to off target interactions.

Fluorescent proteins monitor ligand binding or cellular activity in three ways: (1) fluorescence resonant energy transfer (FRET) based on the principle of excited state energy being transferred between two proteins upon their interaction (Figure 2.2C (i)) (2) monitoring fluorescence from a single protein upon ligand binding (Figure 2.2 C (ii)) (3) visualization of ligand binding when the fluorescent protein is transported within the cell.[7, 14-15] Overcoming the disadvantage of low dynamic range or change in emission of these sensors will allow for improved spectral resolution and use in high throughput screens.[7, 16]

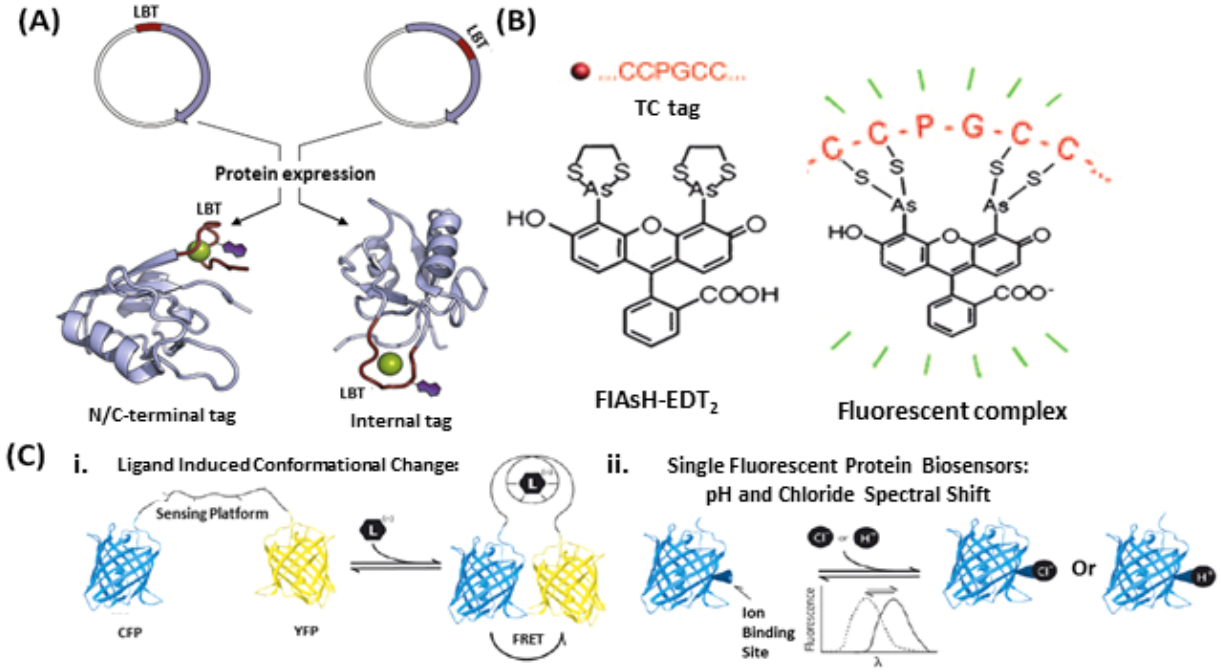


Figure 2.2. Genetically encodable biosensors: (A) Lanthanide tag (LBT) co-expressed for use as peptide binding tag;[3] (B) Example tetra-cysteine tag used in detection of pathogenic bacteria;[8] (C) Fluorescent protein biosensors: (i) FRET based interaction (ii) Single fluorescent protein for biosensor.[7] (Reprinted from Current Opinion in Chemical Biology, 14/2, Allen, K. N.; Imperiali, B., Lanthanide-tagged proteins--an illuminating partnership, 247-254, Copyright 2010, with permission from Elsevier-2.2A; Reproduced with permission from John Wiley & Sons, Inc.-2.2B; Reprinted from Current Opinion in Chemical Biology, 12/1, VanEngelenburg, S. B.; Palmer, A. E., Fluorescent biosensors of protein function, 60-65, Copyright 2008, with permission from Elsevier)

2.2.1.2 Enzymatic Biosensors

Enzymatic biosensors use enzymes to bind analytes, and subsequently catalyze a chemical reaction that is measured by a variety of transducers.[4] Reactions include electron transfer, hydrolysis, esterification, and bond cleavage.[4] The type of reaction determines the transducer that is used. For example, a potentiometric detector measures an electron transfer reaction by detecting a change in potential.[4] Glucose sensors, which monitor blood glucose levels in diabetics, are an example of enzymatic biosensors.[5] In this example glucose oxidase reacts with glucose through a redox reaction. Highly selective recognition of the substrate and

catalysis of a specific reaction enhances the biosensor effectiveness.[4] This high selectivity is a major advantage; however, available enzyme-substrate pairs limit the use of enzymes for biosensing applications.

2.2.1.3 Molecular Recognition Modules

One aim of the work presented here is the development of a new molecular recognition module for biosensing. Molecular recognition modules with binding constants corresponding to sensitivities in the nanomolar and picomolar range play an important role in the success of biosensing and bioimaging.[17-19] Coupling of molecular recognition modules to fluorophores or other visualization devices creates highly selective and specific probes with many applications. As the essential element in a biosensor, several molecular recognition scaffolds have been proposed over the past ten years. However, effective applications of these modules have been limited. This discussion focuses on successfully employed scaffolds that serve as comparative models for our research.

2.2.1.3.1 Antibodies as Molecular Recognition Modules

Antibodies have long been considered highly specific and sensitive antigen binders with nanomolar to picomolar affinities.[1] The production of antibodies, through immunization or by development of a library of affinity binders *in vitro*, allows for targeting a number of different molecules.[20] Yet, numerous difficulties arise with the use of antibodies. For example, the necessity for eukaryotic cell expression makes antibody production expensive and difficult.[1] Also their large size and complex structure lead to unwanted nonspecific interactions.[21] Interest in this area has increased over the past ten years as a result of the need to develop alternative protein scaffolds designs that surpass antibodies in their ability to function as molecular sensing and imaging agents.[19] However, perhaps the most astounding drawback of antibodies results

from their highly unselective nature. A 2008 study revealed that out of over 6,000 clinically used antibody products, less than half are selective to only their desired target.[22] These findings have led the push for standardization of antibody production and use, as well as the development of alternative scaffolds for molecular recognition.

2.2.1.3.2 Alternative Protein Scaffolds

Alternative scaffolds aim to surpass antibodies in biophysical properties such as size, stability, and an absence of cysteine residues (Figure 2.3).[17] Additionally, design of these scaffolds through recombinant or chemical synthetic means will reduce the cost and difficulties associated with antibody production and handling.[19, 23]

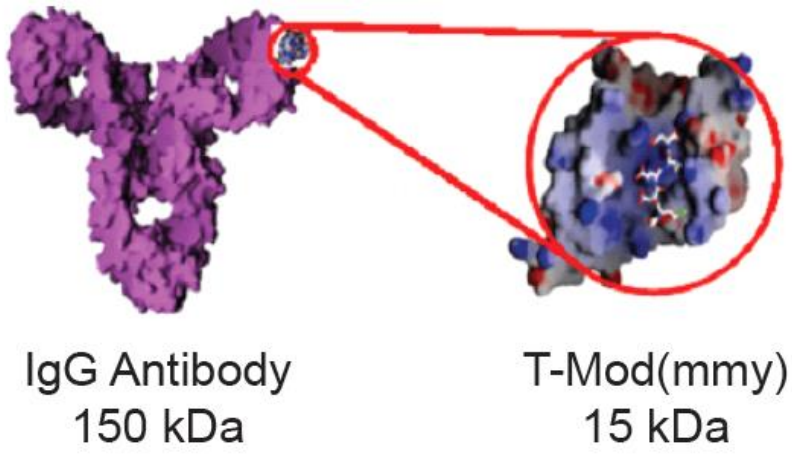


Figure 2.3. Relative size of an antibody and a binding scaffold T-Mod. The T-Mod scaffold is roughly one-tenth the size of an antibody; however, the ligand-binding site is of the same size.[24] (Reproduced with permission from John Wiley & Sons, Inc.)

The development of multifunctional and multispecific molecules is another advantage of alternative scaffolds.[19] Multispecific and multifunctional proteins facilitate several contacts to the target ligand and allow for potential interaction with multiple targets. Furthermore,

alternative scaffolds should be easily amended to enhance pharmacokinetic properties such as length of half-life and clearance of scaffolds used as imaging agents.[19]

A number of alternative protein scaffolds have been proposed over the past several years.[17-21, 25-34] Of these, significant progress has been made on adnectins, affibodies, anticalins, knottins, and several repeat protein scaffolds.

2.2.1.3.3 Proposed Scaffolds

Adnectins are based on the tenth extracellular domain of human fibronectin III, a conserved region of 100 amino acids found in human fibronectin.[19, 35] They have proven successful as alternative antibody scaffolds for use in a variety of therapeutic applications. The structure of adnectins, similar to the antibody Ig structure, consists of 94 amino acids and six loops in a beta sandwich (Figure 2.4A).[19, 26] Although adnectins contains disulfide bonds, folding and stability of these proteins are not dependent upon them.[19] Many targets selected against adnectins include ubiquitin, TNF-alpha, lysozyme, and Abelson kinase SH2 domain, with each having a binding affinity in the nanomolar or picomolar range.[19]

Engineered binding proteins known as affibodies are based on the B domain of staphylococcal protein A.[26] Affibodies are good scaffolds due to their small size of approximately 6 kDa, cysteine free sequence, ability for easy and fast folding, high stability, and efficient library generation through *in vitro* selection methods.[31] Affibodies consist of three helical bundles with 58 amino acids, and have been successfully used in therapeutics and molecular imaging applications (Figure 2.4B).[36-37]

Anticalins, engineered lipocalins, share many similarities with antibodies in their overall structure— a beta-barrel scaffold accompanied by four peptide loops (Figure 2.4C).[19]

Additionally, anticalins exhibit high stability and are monomeric. Their design enables them to target peptides and proteins in a number of different applications.[32, 38-39]

Another class of proteins that has generated significant interest as alternative scaffolds is cysteine knot peptides, which are also known as knottins. Knottins consist of approximately 20 to 60 amino acids (Figure 2.4D). A valuable characteristic of knottins includes their natural chemical, proteolytic, and thermal stability that results from three stabilizing disulfide bonds.[29] Additionally, the cysteine core of the knottins is the only highly conserved part of the structure.[28] As a result, knottins have high potential for engineering ligand binding because of their ability to introduce variation into the loop sequence without interrupting the overall protein structure.

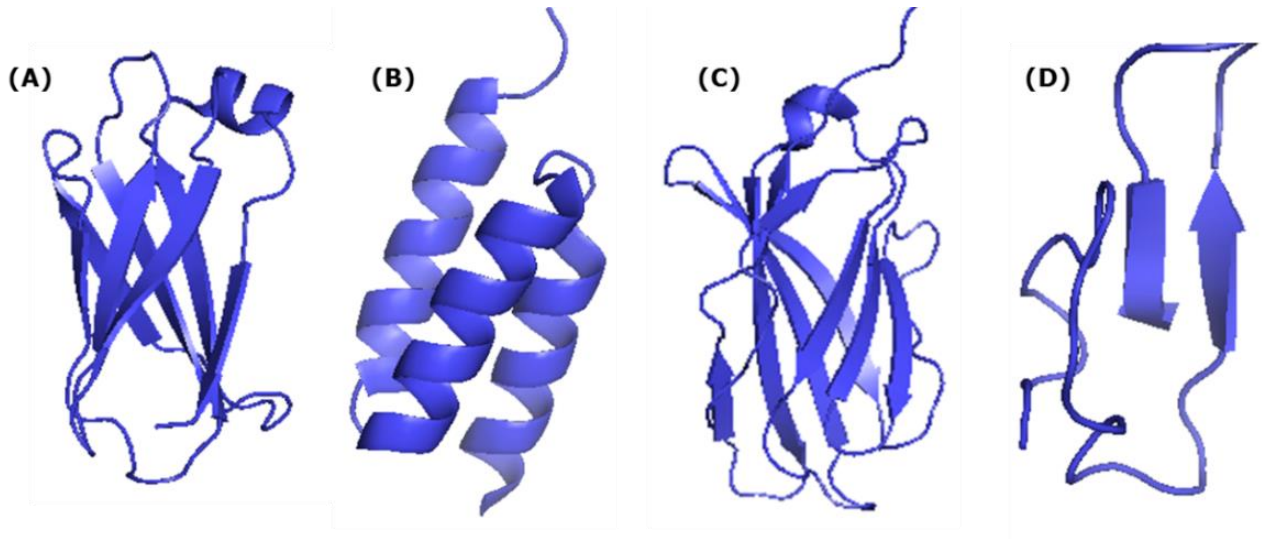


Figure 2.4. Structures of alternative proteins scaffolds: (A) Adnectin scaffold (PDB: 2OCF). (B) Affibody scaffold (PDB: 3MZW). (C) Anticalin scaffold (PDB: 3BX7). (D) Cysteine knot scaffold (PDB: 1MR0).

Naturally occurring repeat proteins are implicated in numerous protein-protein interactions.[17] These proteins consist of a repeating structural motif that varies in number. The

modular nature of repeat proteins proves advantageous when used in alternative scaffolding. Unlike globular proteins, repeat proteins form elongated structures resulting in a larger surface area for potential ligand interaction, as well as creating the possibility to design these proteins, through engineering repeating units and adding or removing repeats to have multispecific functions.[27] Repeat protein families previously studied as alternative scaffolding include: designed ankyrin repeat proteins (DARPins), tetratricopeptide repeats (TPRs), armadillo repeat proteins (ARMs), and leucine rich repeat proteins (LRRs) (Figure 2.5 A-2.5 D).

DARPins and TPRs bind target ligands through the concave face (Figure 2.5 B).[18] Ligand binding sites were engineered onto the DARPin and TPR scaffolds through directed evolution methods and binding site grafting.[27, 40] An important feature of TPRs is their rigid structure.[25] TPRs do not undergo a conformational change upon ligand binding, allowing for a simpler design; there is no need to account for structural change upon interaction with the target. ARM, known for their peptide binding affinities, bind one dipeptide of the ligand per repeating unit.[17, 27] A recent consensus-based analysis yielded an ARM design capable of binding the peptide neurotensin serving as one of the first successful designs of this scaffold.[33]

Several naturally occurring proteins contain LRRs. The proteins found in jawless fish are notable examples of LRR containing proteins. Here LRRs are found in the variable lymphocyte receptors (VLRs), the proteins that mediate adaptive immune response in the jawless fish.[30, 34] Crystal structures indicate that the concave face of rigid beta strands act as the ligand binding surface.[41]

LRRs from the nod-like receptor proteins (NLRs) serve as the basis for our scaffold design. As with previously discussed alternative scaffolds, LRRs gain their advantage as molecular recognition modules because of their biological role in nature. The innate ability of

NLRs to sense small molecules makes them the ideal scaffold for our molecular recognition device.

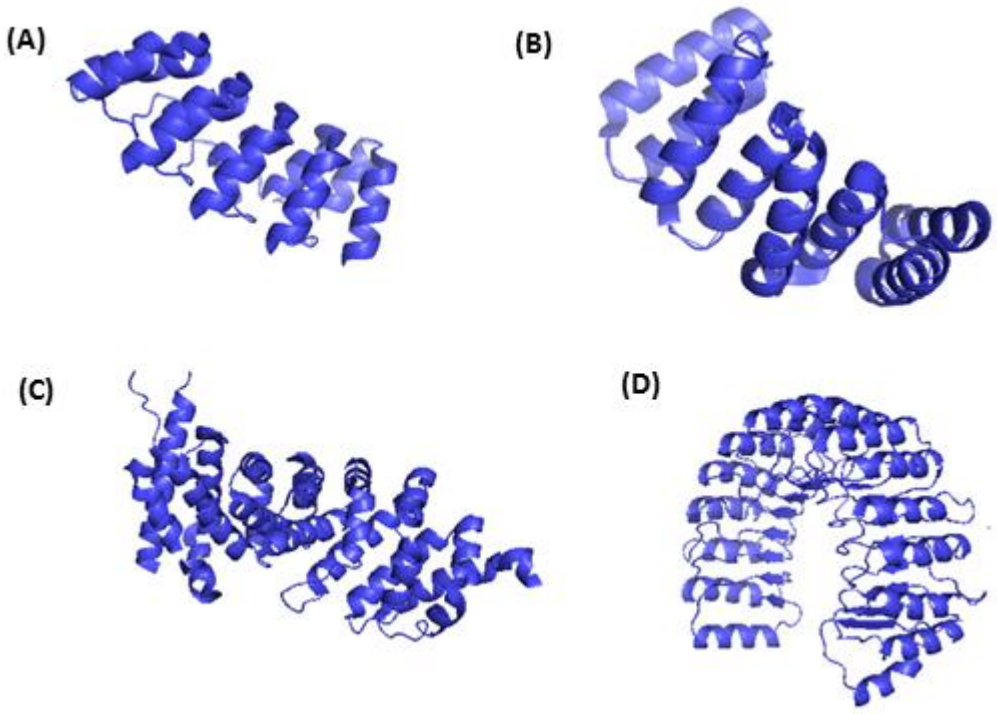


Figure 2.5. Structures of repeat proteins: (A) DARPin: 33 amino acids repeats that form a helix-turn-helix-β-hairpin structure (PDB: 2XEE). (B) TPR: 34 amino acid repeats, elongated alpha helical structure (PDB: 2FO7). (C) ARM: 42 amino acid repeats made up of three alpha helices PDB: 3NMW). (D) LRR: 28 amino acid repeating structural motif that forms a horseshoe-like scaffold consisting of beta-strand and helical segment linked by variable loops (PDB:1A4Y).

2.2.2 Methods for Design of Molecular Recognition Scaffolds

There are three primary strategies for design of molecular recognition scaffolds: high throughput screening, computational design, and consensus-based design. Principles of each technique, as well as advantages and disadvantages for each method will be discussed.

2.2.2.1 High-Throughput-Screening

In combinatorial protein engineering, a large library of potential target binders is developed by simultaneously changing multiple residues and screening for a desired function.[18] Step-by-step alteration, production, and analysis of sequences prove nearly impossible. Therefore, combinatorial protein engineering provides an efficient alternative method for selection of proteins with high binding affinity for the analyte. Numerous techniques exist for directed evolution and screening of protein libraries including: cell-dependent display systems, cell-free display systems, and non-display systems.[18] Provided below is a brief overview of these techniques.

Two commonly used cell-dependent methods include yeast surface display (YSD) and phage display. In YSD the proteins express as a fusion to the Aga2p mating agglutinin protein.[42] Aga2p is linked to Aga1p protein, and Aga1p covalently links to the yeast cell surface (Figure 2.6).[42] This allows the proteins of interest to interact with other molecules in solution. Use of cell sorting techniques, coupled with fluorescent tags that are incorporated into the library proteins and the target molecule, allow for isolation and detection of binding proteins.

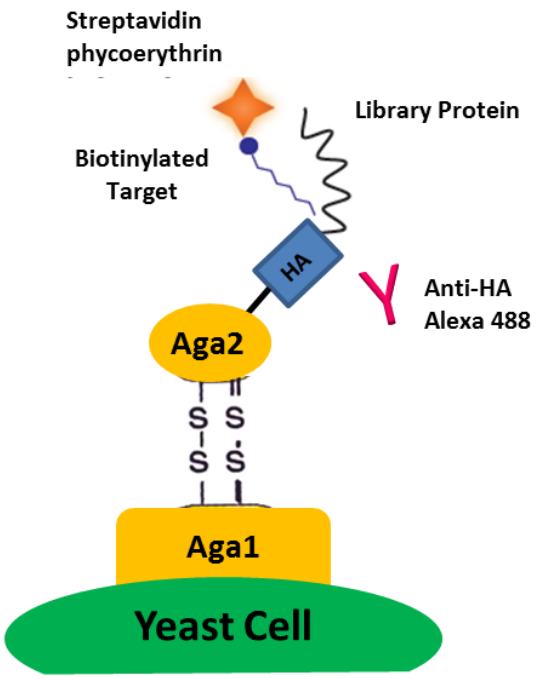


Figure 2.6. Schematic of yeast surface display: Protein library is expressed as a fusion to the Ag2p protein. Ag2p is linked to Ag1p through 2 disulfide bonds. Ag1p is covalently linked to the yeast. cell wall. HA and Anti-HA are epitope tags that can be used to quantify expression through fluorescence. The target ligand is biotinylated and fused to streptavidin. Adapted from reference 42.[42]

In phage display, proteins are expressed on the surface of bacteriophages by fusing the library of interest to the pIII or pVIII coat proteins.[18] Libraries, consisting of mutated copies of the gene, express up to approximately 10^8 mutants for YSD and 10^{10} for phage display.[1, 42] The efficiency of DNA transformation limits the library size.[43]

In an effort to overcome the limitations of cell-dependent systems, such as library size, cell-free selection systems have been explored. The most successful cell-free system is ribosomal display, where the target gets co-expressed with members of the generated mutant library.[18] An advantage of cell-free systems is that they can produce library sizes of approximately 10^{13} to 10^{14} mutants since they are not limited by transformation efficiency.[1, 44]

However, a limitation of this method is the requirement that the target protein must be expressed in its folded form.[18]

2.2.2.2 Computational Protein Design

Computational design aims at obtaining a better understanding of the relationship between protein sequence and function in order to produce proteins with ideal properties.[45] Previously, the implementation of such methods helped improve the stability of proteins and the specificity of their interactions to create novel designs with desired functionalities.[46] For example, the Rossetta suite has been used successfully for many macromolecular modeling applications, including design of novel protein interfaces.[47] While experimental methods do not always provide the same level of design control, computational methods have yet to consistently produce designs that match the affinity and specificities of experimental designs.[48] Nevertheless, computational methods consistently supply candidates for functional experimental design.[49]

2.2.2.3 Consensus-based Design

The consensus-based design approach to structural protein engineering analyzes the statistical occurrence of amino acids in each position of a sequence in order to generate a consensus design.[50] Sequences of proteins with related structure and function are analyzed in a multiple sequence alignment to determine amino acids with highest occurrence in each position. By selecting residues with the highest statistical frequencies, a sequence is constructed based on the idea that stabilizing amino acids will occur with the greatest frequency.[50] This analysis provides useful information regarding optimal sequence residues and their physical properties. Additional analyses, such as hyper-variability and covariation analyses, can also give better insight into potential ligand binding residues that can be used for later randomization in high

throughput assays.[51] A consensus sequence aims to generate an optimized structural motif with increased stability.[50]

2.2.4 Summary

Each method for development of alternative protein scaffolds successfully aids in the generation of new molecular recognition devices. Limitations of each method exist that are often overcome through use of two methods in tandem. High-throughput-screening assays limit the control over protein design. In contrast, computational or consensus approaches allow for specific sequence design, but usually need further experimental procedures to enhance the desired function of the scaffold. Consensus-based or computational design, along with combinatorial protein engineering, make it possible to develop new alternative scaffolds with great potential for many applications.[20, 28, 33, 50, 52] The ability to generate specific and enhanced designs through sequence analysis helps to create scaffolds with properties superior to those of antibodies. Additionally, directed evolution and selections methodologies significantly enhance the efficiency of selecting high affinity receptors for a wide range of targets.

2.3 Introduction to Biomaterials Design

Biomaterials engineering presents another field where recombinant protein technologies are emerging as important design tools. Research in the design of new biomaterials aims to develop alternatives to standard techniques used for regenerative medicine and tissue engineering applications. The field of regenerative medicine stems from translational research in tissue engineering and molecular biology. The goal of regenerative medicine and tissue engineering is to restore or create normal function in damaged or diseased tissues, cells, and organs through the use of biomaterials, cells, bioactive molecules, and the body's own ability for self-healing.[53-54] However, traditionally used scaffolds and synthetic materials for these

applications fall short in obtaining the characteristics necessary for successful clinical implementation. For example, autologous and allogenic grafts have long been the standard for tissue engineering applications. However, these techniques suffer from many drawbacks including donor site morbidity (autografts) and unwanted immune response, as well as possible risk of disease transfer (allografts).[55] For many years, despite the problems associated with their use, these procedures accounted for the vast majority of treatments for damaged tissues and organs. Alternatively, synthetic materials, such as polymers, metals, and ceramics have been used, but have also seen limited success in large part due to their lack of biocompatibility and the difficulty associated with modifying their chemical properties.[56] In an effort to overcome these challenges, research has turned toward the use of natural biomaterials, mainly protein-based materials, as well as recombinant protein-based materials as a promising alternatives to traditional methods.

When designing a scaffold for regenerative medicine and tissue engineering applications there are certain properties required.[57] First, the final material must provide good mechanical and structural support. The ability to modulate these parameters for specific applications also offers many advantages for development of tunable materials. Additionally, biomaterials should demonstrate excellent control over cell attachment, as well as migration, proliferation, and differentiation.[58] Many natural biopolymers contain cell adhesion sites within their primary sequence, which facilitate cellular attachment without additional modification.[59-63] Lastly, controlled degradation of the material is critical for its overall success. Materials with re-sorbable features allow for degradation of the scaffold in a precise manner that is complimentary to the rate of healing of the body, providing a temporary implant instead of a permanent structure.[64] Each of these features are essential to the effectiveness and success of the scaffold. While

traditional methods fall short of these characteristics, protein-based materials provide a promising solution to achieve the required biomaterials properties.[65]

Natural biopolymers have many advantages over previously implemented techniques including increased biocompatibility, ease of surface modification, and diverse chemical and physical properties.[66-67] Over the last three decades protein-based materials derived from extracted collagen, elastin, fibronectin, laminin, fibrin, silk, vitronectin, and keratin have achieved much success in regenerative medicine. However, while many of the materials produced from these proteins come close to achieving all of the desired characteristics, most scaffolds fail to achieve all the necessary qualifications needed for successful clinical use. Limited tunability of properties, need for functionalization or modification that may disrupt structure or function, and limited resource abundance demonstrate some of the problems associated with the use of natural materials. As a result, the use of recombinant protein-based materials as the next generation in biomaterials design has garnered much interest. Progress in protein design and genetic engineering techniques provides an avenue for creating scaffolds with essential properties through tailoring of the protein sequence, and thus function and materials properties. The following sections will discuss natural and recombinant protein-based biomaterials, as well as their use and effectiveness in regenerative medicine and tissue engineering applications.

2.3.1 Natural Protein-based Biomedical Materials

Natural protein-based materials provide many advantages to traditional methods and other synthetic systems. Collagen, elastin, fibronectin, laminin, fibrin, silk, vitronectin, and keratin provide some examples of the most widely used natural proteins for biomedical applications (Figure 2.7 A-D).[60, 68-76] Scaffolds engineered for regenerative medicine seek to

mimic the chemical and mechanical properties of the extra cellular matrix (ECM) or other important structural and mechanical properties of tissues. Consequently, as the natural proteins listed above are structural and/or regulatory components of tissues and the ECM, they are well poised for use in materials generation.

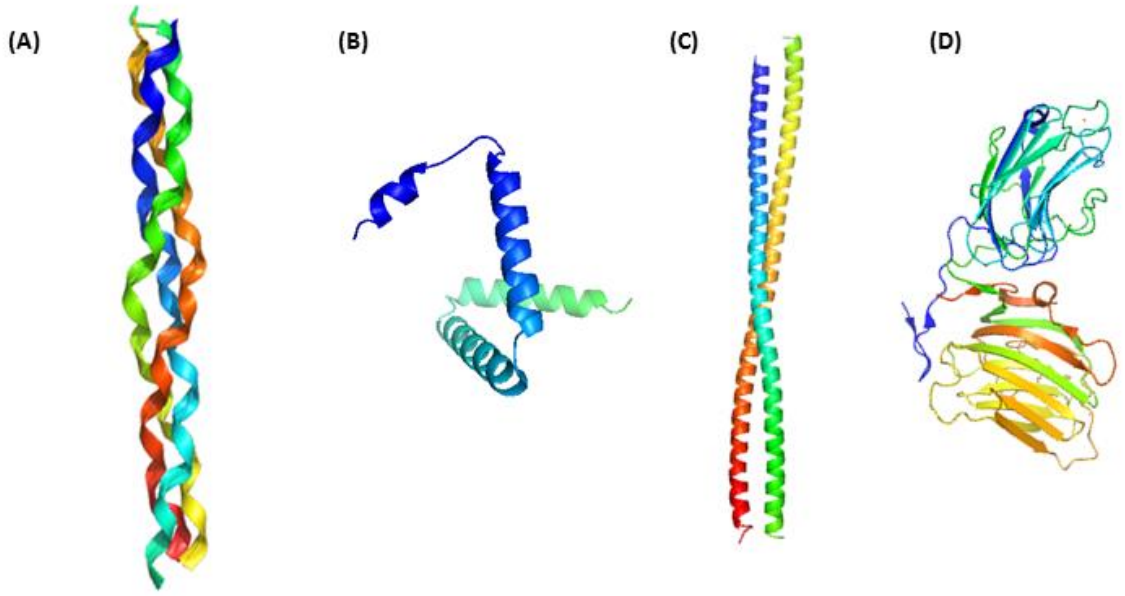


Figure 2.7. Crystal structure of (A) Collagen (B) Silk (C) Keratin and (D) Laminin. PDB codes: 1CAG, 5D2Q, 3TNU, and 5IK5.

2.3.1.1 Collagen

Collagen has been implemented in biomedical applications for several decades.[77-81] It is a structural protein that is found in many tissues, such as tendons, cartilage, bone, and skin. As the most abundant mammalian protein, it plays many important functional roles including regulating cell adhesion and migration, as well as participating in tissue repair.[67] Collagen provides advantages for medical applications such as, excellent biocompatibility and simple functionalization. Additionally, given its natural abundance, it is readily obtained from natural

sources, and its biodegradable quality affords it with another important characteristic. However, despite its success, collagen lacks the mechanical versatility needed for utilization in many applications, due to its rigid structural characteristics. Consequently, blending of collagens with other protein or polymer sources to improve mechanical properties has been completed with limited success.[82-83]

2.3.1.2 Elastin

Another component of the ECM, elastin, also offers many beneficial advantages over traditionally used scaffolds. Elastin is a structural protein mainly found in skin, lungs, arteries, and various connective tissues. Comprised of tropoelastin monomers, elastin provides mechanical flexibility to tissues (i.e. elasticity), as well as promotes cell adhesion and cell growth.[61] Another important feature of elastin comes from its integrin binding site for $\alpha_v\beta_3$, which improves its biocompatibility when used for materials development. Elastin is often used with other proteins and polymers to improve mechanical durability, notably extensive work on silk-elastin combinations.[84] The flexible mechanical features of elastin result from the hydrophobic domains in the protein sequence. However, crosslinking of the hydrophilic domains creates the insoluble fibrous structure, which in turn limits the use of elastin from natural sources, as its extraction and processing becomes difficult. Consequently, "elastin-like" polypeptides have been used as artificial alternatives to natural elastin due to improved solubility and tunability.[85-87]

2.3.1.3 Fibronectin

A glycoprotein, fibronectin demonstrates another example of a natural protein well suited for regenerative medicine and tissue engineering applications.[88-91] Similar to elastin and collagen, fibronectin provides structural support as an ECM component. It also plays a role in

cell adhesion and growth, and contains many functional binding motifs, including cell binding motifs and binding motifs for collagen, heparin, fibrin, and gelatin.[62] Due to its excellent biocompatibility and biological functions, modification of scaffolds with fibronectin to improve cellular attachment and proliferation has proved a successful method for improving performance of biomaterials.[69, 73, 91]

2.3.1.4 Vitronectin, Laminin, and Fibrin

Other ECM proteins suitable for tissue engineering and regenerative medicine include vitronectin, laminin, and fibrin. Vitronectin is a glycoprotein that binds collagen, glycosaminoglycans, and plasminogen.[92] In addition it contains the RGD cellular binding motif thus making it useful for improving cell attachment and growth.[74] Similar to vitronectin, laminin is also an ECM glycoprotein. Often used to coat polymer-based materials, laminin binds matrix proteins, as well as increases cellular adhesion and proliferation.[75, 93-94] Lastly, fibrin, a viscoelastic biopolymer, provides an example of an ECM protein only present during specific biological circumstances. Fibrin helps in blood clotting and thus is only present in the ECM when needed.[95] However, fibrin has found much clinical use in tissue engineering, wound and burn treatments, as well as with medical adhesives.[76, 96-99]

2.3.1.5 Silk

One of the most commonly used biopolymers for tissue engineering and regenerative medicine comes in the form of silk. Silk proteins have been widely used in large capacity for biomedical applications.[72, 100-104] One of the most astounding features of silk is derived from it excellent mechanical durability resulting from the hierarchical structure formed by the fibrous protein. Silk also has an exceptional biocompatibility and can be obtained in abundance from natural sources.[72] Yet, not all silk proteins are easily obtained from natural sources. For example,

spider silk production proves difficult, and has thus led to the engineering of spider silk-like proteins, demonstrating another example where protein engineering techniques can overcome the challenges associated with natural biopolymers.[105-106]

2.3.1.6 Keratin

Keratin proteins provide an additional example of natural proteins that have found much use in regenerative medicine and tissue engineering applications. The intrinsic capacity of keratin to self-assemble into mechanically robust fibers lends itself to many biomedical applications.[107-109] Keratin also provides excellent biocompatibility, and contains cellular binding motifs that have been shown to assist in cell adhesion and growth.[60] Consequently, keratin-based materials have been used in wound healing, tissue engineering, and nerve regeneration with great success.[110-120] Similar to previously presented examples, keratin proteins often need further modification following extraction from their natural sources thus lending them to improvement through protein engineering.

2.3.1.7 Summary of Natural Proteins

Natural biopolymers offer excellent alternatives to traditional tissue engineering methods and other synthetic materials. Their intrinsic biocompatibility and biological activities, as well as chemical and mechanical properties poise them as well-suited scaffolds for many biomedical applications. However, the need for further modification and functionalization in addition to limitations often encountered when obtaining these proteins from natural sources, create disadvantages with their use. Modification of the natural protein may lead to aggregation or denaturation and thus the loss of biological activity, as well as possible undesired immunogenic responses.[121] In an effort to improve on the beneficial properties of natural proteins and achieve each of the desired characteristics of the ideal scaffold, research has turned to protein

engineering where the use of recombinant DNA technology and genetic engineering methods may provide the solution to these problems.

2.3.2 Recombinant Proteins for Biomaterials Development

While natural protein-based scaffolds provide many advantages to traditional engineering methods, the use of recombinant protein engineering technology provides a possible solution to overcoming the limitations associated with these materials. Advances in recombinant DNA technology and genetic engineering present many benefits for the design of new biomaterials systems. The efficiency of recombinant DNA technology allows for facile synthetic gene construction, cloning, specific and selective sequence mutations, and enables subsequent protein production and isolation. The controlled and specific tuning of the protein sequence enables the precise control over the resulting functions and characteristics of the final material.

Recombinant protein engineering has several main advantages over extraction of proteins from their natural sources (Figure 2.8).[122] One such advantage is the efficient sequence tailoring accomplished through selective mutations, insertions, and deletions allowing for rational material design. The incorporation of non-canonical amino acids also expands the sequence scope and provides a handle for later conjugation or modification. Additionally, recombinant DNA technology allows for the creation of protein sequences from two or more non-related proteins. For example, the engineering of silk-elastin-like proteins has resulted in materials with improved mechanical properties through the combination of the elasticity of the elastin protein and the strength of silk.[84] The ability to apply combination design further expands and improves on previously used methods for blends of natural proteins through controlled design. Recombinant protein engineering methods also provide the ability to control processing conditions, thus removing problems associated with source variability. Specific sequence modifications provide

the means for incorporation of cellular adhesion motifs, as well as other protein binding motifs (e.g. integrin binding motif). This increases the biocompatibility and improves the biological activity of the scaffold. Analogous to incorporation of binding motifs is the incorporation of degradation sites into the protein sequence. The ability to control the number and type of degradation sites enables subsequent control over the rate of the scaffold degradation in a manner that is compatible with the bodies healing process. Lastly, recombinant protein production takes materials engineering from design of static systems into the next generation of multifunctional dynamic systems, such as the design of stimuli responsive materials. All of the advantages gained through recombinant protein engineering create the ability to develop materials with tunable chemical, mechanical, and physical properties. Proteins that have been successfully produced by recombinant protein methods include: silk, elastin, and collagen. Table 2.1 summarizes each proteins structure, their natural functions, and applications.

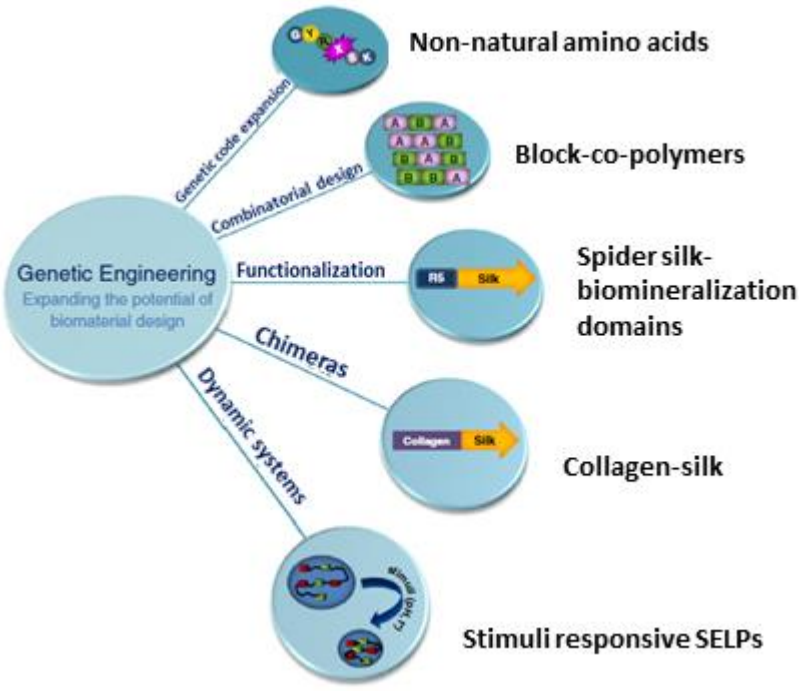


Figure 2.8. Chart demonstrating advantages of protein engineering techniques for the design of biomaterials for medical applications. (Reprinted from Current Opinion in Biotechnology, 39, Dinjaski, N. and Kaplan D.L., Recombinant protein blends: silk beyond natural design, 1-7, Copyright 2016, with permission from Elsevier).

2.3.2.1 Silk

Silk protein use in biomedical materials dates back several decades.[72] The robust mechanical properties of silk fibers provide an exquisite combination of elasticity and strength. Specifically, spider silk has garnered much interest for recombinant protein production due to the difficulties associated with procuring the material from natural sources.[106] Spiders produce and use a wide variety of silk proteins resulting from the need to fabricate silk for various purposes (e.g. building webs, catching prey, etc.). As a result, spider silks offer many possibilities for tailored materials design, so their recombinant production has been widely explored. Recombinant silk-based materials have been implemented in an assortment of applications including, tissue regeneration, drug and gene delivery, neural tissue engineering, delivery of

bioactive molecules, and bone regeneration.[101, 105, 123] Additionally, silk fusion proteins have been expansively investigated, most notably silk-elastin-like proteins (SELPs).[84, 122, 124] The main advantage provided by SELPs results from their stimuli-responsive characteristics. As previously noted, next generation biomaterials seek to develop dynamic systems, such as these stimuli-responsive SELPs.

A recent example exhibited in the work by Zhou et al. demonstrates the utility of the SELP scaffold for biomedical materials design.[125] In this work a rationally designed redox responsive injectable SELP hydrogel was developed for controlled drug delivery through a redox-sensitive release mechanism. The ratio of silk to elastin proved critical to optimization of the mechanical properties, as well as the redox-sensitive features of the system, once again demonstrating the advantage of recombinant protein production for precise design and control of the properties and features of the materials.

2.3.2.2 Elastin

The mechanical flexibility of elastin proteins makes it an attractable scaffold for biomaterials design. However, elastin suffers from difficulty in extracting the protein from natural sources as a result of the insolubility of its fibers.[85] Consequently, efforts have been made to produce elastin-like proteins that strive to mimic the mechanical and biological properties of native elastin proteins. Recombinant elastin proteins have been used in tissue engineering, drug delivery, and nanoparticle applications.[71, 86-87] However, while the elastic characteristic of protein fibers offers advantageous mechanical features, elastin suffers from a lack of mechanical durability. The need to improve mechanical resilience of elastin-based materials lead to its use in chimeric protein constructs. Elastin used in conjunction with silk is a strategy that has found much success in the design of tunable materials.[84, 124-126] Elastin represents an excellent example

of the value of protein engineering to improve upon natural protein-based scaffolds while retaining necessary properties and, through the use of combination designs, improving these qualities.

2.3.2.3 Collagen

Natural collagen proteins have many characteristics that make them useful for biomaterials design. However, limitations to their use have resulted from problems with batch variability, as well as lack of purity of the extracted material.[79] As with silk and elastin, recombinant protein engineering offers a solution to these shortcomings. Recombinant collagen proteins from bacterial collagen sequences have been expressed and purified with great success.[127-128] Human collagen proteins require post translational hydroxylation of proline in order to form the correct structures. Thus, their recombinant expression has proven difficult, leading to the use of bacterial collagen proteins that do not require these PTMs. Bacterial collagens provide the same mechanical properties as human collagen proteins.[128] Although bacterial collagens do not have the same biological components, such as specific binding motifs in their sequence, bacterial collagens are easily functionalized on the sequence level to contain the desired biological properties. Functionalized recombinant bacterial collagens have been used in bone regeneration, tissue engineering, wound healing, and drug delivery.[77, 80-81, 83, 129-133]

Table 2.1. Recombinant proteins for biomedical applications. (Adapted from references 58 and 109.)

Protein	Biological Functions	Recombinant Production	Recombinant Protein Applications	References
Collagen	Structural protein-tendon, skin, bone, cartilage, connective tissues	Bacterial collagen-like proteins	Bone regeneration, wound healing, drug and gene delivery	12, 60, 81
Elastin	Structural protein-skin, lungs, ligaments, connective tissue of arteries	Elastin-like proteins	Drug delivery, delivery of bioactive molecules, tissue engineering	62, 115, 116, 118
Silk	Building elements of many anthropod nests, cocoons, and prey traps	Silk-like proteins	Tissue engineering, drug and gene delivery, bone regeneration	73, 97, 109, 111
Keratin	Structural protein-epithelial cells and epidermal appendages	Epithelial keratins, Epidermal keratins	Mechanistic self-assembly studies, biomaterials generation	61, 104, 106, 108

2.3.2.4 Summary of Recombinant Proteins for Regenerative Medicine and Tissue Engineering

Silk, elastin, and collagen demonstrate the value of recombinant protein engineering tools to overcome the limitations of natural protein-based materials. Creating multidimensional materials in a controlled and specific manner facilitates the goal of developing a scaffold with each of the characteristics and features needed for a specific application. Recombinant proteins provide a promising avenue for the creation of next generation biomaterials with tailored sequences and tunable chemical, mechanical, and physical properties. Chapters five and six highlight our efforts to create keratin-based biomaterials using recombinantly produced human hair keratins.

2.4 References

1. Conroy, P. J.; Hearty, S.; Leonard, P.; O'Kennedy, R. J., Antibody production, design and use for biosensor-based applications. *Seminars in cell & developmental biology* **2009,** *20* (1), 10-26.
2. Adams, S. R.; Tsien, R. Y., Preparation of the membrane-permeant biarsenicals FlAsH-EDT2 and ReAsH-EDT2 for fluorescent labeling of tetracysteine-tagged proteins. *Nature protocols* **2008,** *3* (9), 1527-34.
3. Allen, K. N.; Imperiali, B., Lanthanide-tagged proteins--an illuminating partnership. *Current opinion in chemical biology* **2010,** *14* (2), 247-54.
4. Andreas Manz, N. P., Dimitri Lossifidis, *Bioanalytical Chemistry*. Imperial College Press: London, 2004; p 201.
5. Newman, J. D.; Setford, S. J., Enzymatic biosensors. *Molecular biotechnology* **2006,** *32* (3), 249-68.
6. Palmer, A. E.; Qin, Y.; Park, J. G.; McCombs, J. E., Design and application of genetically encoded biosensors. *Trends in biotechnology* **2011,** *29* (3), 144-52.
7. VanEngelenburg, S. B.; Palmer, A. E., Fluorescent biosensors of protein function. *Current opinion in chemical biology* **2008,** *12* (1), 60-5.
8. Wu, L.; Huang, T.; Yang, L.; Pan, J.; Zhu, S.; Yan, X., Sensitive and selective bacterial detection using tetracysteine-tagged phages in conjunction with biarsenical dye. *Angew Chem Int Ed Engl* **2011,** *50* (26), 5873-7.
9. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. 3 ed.; Springer: 2006.
10. Pertz, O.; Hahn, K. M., Designing biosensors for Rho family proteins--deciphering the dynamics of Rho family GTPase activation in living cells. *Journal of cell science* **2004,** *117* (Pt 8), 1313-8.
11. Pfleger, K. D.; Eidne, K. A., Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat Methods* **2006,** *3* (3), 165-74.
12. Su, X. C.; Huber, T.; Dixon, N. E.; Otting, G., Site-specific labelling of proteins with a rigid lanthanide-binding tag. *Chembiochem : a European journal of chemical biology* **2006,** *7* (10), 1599-604.
13. Franz, K. J.; Nitz, M.; Imperiali, B., Lanthanide-binding tags as versatile protein coexpression probes. *Chembiochem : a European journal of chemical biology* **2003,** *4* (4), 265-71.
14. Mank, M.; Reiff, D. F.; Heim, N.; Friedrich, M. W.; Borst, A.; Griesbeck, O., A FRET-based calcium biosensor with fast signal kinetics and high fluorescence change. *Biophysical journal* **2006,** *90* (5), 1790-6.
15. Nguyen, A. W.; Daugherty, P. S., Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nature biotechnology* **2005,** *23* (3), 355-60.
16. Nagai, T.; Miyawaki, A., A high-throughput method for development of FRET-based indicators for proteolysis. *Biochemical and biophysical research communications* **2004,** *319* (1), 72-7.
17. Boersma, Y. L.; Pluckthun, A., DARPins and other repeat protein scaffolds: advances in engineering and applications. *Current opinion in biotechnology* **2011,** *22* (6), 849-57.
18. Gronwall, C.; Stahl, S., Engineered affinity proteins--generation and applications. *Journal of biotechnology* **2009,** *140* (3-4), 254-69.
19. Lofblom, J.; Frejd, F. Y.; Stahl, S., Non-immunoglobulin based protein scaffolds. *Current opinion in biotechnology* **2011,** *22* (6), 843-8.

20. Binz, H. K.; Amstutz, P.; Kohl, A.; Stumpp, M. T.; Briand, C.; Forrer, P.; Grutter, M. G.; Pluckthun, A., High-affinity binders selected from designed ankyrin repeat protein libraries. *Nature biotechnology* **2004,** *22* (5), 575-82.
21. Skerra, A., Alternative non-antibody scaffolds for molecular recognition. *Current opinion in biotechnology* **2007,** *18* (4), 295-304.
22. Berglund, L.; Bjoerling, E.; Oksvold, P.; Fagerberg, L.; Asplund, A.; Szigyarto, C. A.-K.; Persson, A.; Ottosson, J.; Wernerus, H.; Nilsson, P.; Lundberg, E.; Sivertsson, A.; Navani, S.; Wester, K.; Kampf, C.; Hober, S.; Ponten, F.; Uhlen, M., A Genecentric Human Protein Atlas for Expression Profiles Based on Antibodies. *Mol. Cell. Proteomics* **2008,** *7* (10), 2019-2027.
23. Werner, R. G., Economic aspects of commercial manufacture of biopharmaceuticals. *Journal of biotechnology* **2004,** *113* (1-3), 171-82.
24. Jackrel, M. E.; Valverde, R.; Regan, L., Redesign of a protein-peptide interaction: characterization and applications. *Protein science : a publication of the Protein Society* **2009,** *18* (4), 762-74.
25. Cortajarena, A. L.; Regan, L., Ligand binding by TPR domains. *Protein science : a publication of the Protein Society* **2006,** *15* (5), 1193-8.
26. Gebauer, M.; Skerra, A., Engineered protein scaffolds as next-generation antibody therapeutics. *Current opinion in chemical biology* **2009,** *13* (3), 245-55.
27. Grove, T. Z.; Cortajarena, A. L.; Regan, L., Ligand binding by repeat proteins: natural and designed. *Current opinion in structural biology* **2008,** *18* (4), 507-15.
28. Kimura, R. H.; Levin, A. M.; Cochran, F. V.; Cochran, J. R., Engineered cystine knot peptides that bind alphavbeta3, alphavbeta5, and alpha5beta1 integrins with low-nanomolar affinity. *Proteins* **2009,** *77* (2), 359-69.
29. Kolmar, H., Alternative binding proteins: biological activity and therapeutic potential of cystine-knot miniproteins. *The FEBS journal* **2008,** *275* (11), 2684-90.
30. Lee, S. C.; Park, K.; Han, J.; Lee, J. J.; Kim, H. J.; Hong, S.; Heu, W.; Kim, Y. J.; Ha, J. S.; Lee, S. G.; Cheong, H. K.; Jeon, Y. H.; Kim, D.; Kim, H. S., Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (9), 3299-304.
31. Nygren, P. A., Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold. *The FEBS journal* **2008,** *275* (11), 2668-76.
32. Skerra, A., Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *The FEBS journal* **2008,** *275* (11), 2677-83.
33. Varadamsetty, G.; Tremmel, D.; Hansen, S.; Parmeggiani, F.; Pluckthun, A., Designed Armadillo Repeat Proteins: Library Generation, Characterization and Selection of Peptide Binders with High Specificity. *Journal of molecular biology* **2012**.
34. Wezner-Ptasinska, M.; Krowarsch, D.; Otlewski, J., Design and characteristics of a stable protein scaffold for specific binding based on variable lymphocyte receptor sequences. *Biochimica et biophysica acta* **2011,** *1814* (9), 1140-5.
35. Lipovsek, D.; Lippow, S. M.; Hackel, B. J.; Gregson, M. W.; Cheng, P.; Kapila, A.; Wittrup, K. D., Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: molecular convergence with single-domain camelid and shark antibodies. *Journal of molecular biology* **2007,** *368* (4), 1024-41.

36. Gao, J.; Chen, K.; Miao, Z.; Ren, G.; Chen, X.; Gambhir, S. S.; Cheng, Z., Affibody-based nanoprobes for HER2-expressing cell and tumor imaging. *Biomaterials* **2011,** *32* (8), 2141-8.
37. Miao, Z.; Levi, J.; Cheng, Z., Protein scaffold-based molecular probes for cancer molecular imaging. *Amino acids* **2011,** *41* (5), 1037-47.
38. Schlehuber, S.; Beste, G.; Skerra, A., A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. *Journal of molecular biology* **2000,** *297* (5), 1105-20.
39. Vopel, S.; Muhlbach, H.; Skerra, A., Rational engineering of a fluorescein-binding anticalin for improved ligand affinity. *Biol Chem* **2005,** *386* (11), 1097-104.
40. Amstutz, P.; Binz, H. K.; Parizek, P.; Stumpp, M. T.; Kohl, A.; Grutter, M. G.; Forrer, P.; Pluckthun, A., Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. *The Journal of biological chemistry* **2005,** *280* (26), 24715-22.
41. Kim, H. M.; Oh, S. C.; Lim, K. J.; Kasamatsu, J.; Heo, J. Y.; Park, B. S.; Lee, H.; Yoo, O. J.; Kasahara, M.; Lee, J. O., Structural diversity of the hagfish variable lymphocyte receptors. *The Journal of biological chemistry* **2007,** *282* (9), 6726-32.
42. Colby, D. W.; Kellogg, B. A.; Graff, C. P.; Yeung, Y. A.; Swers, J. S.; Wittrup, K. D., Engineering antibody affinity by yeast surface display. *Methods Enzymol* **2004,** *388*, 348-58.
43. Rothe, A.; Hosse, R. J.; Power, B. E., In vitro display technologies reveal novel biopharmaceutics. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **2006,** *20* (10), 1599-610.
44. Lui, B. H.; Cochran, J. R.; Swartz, J. R., Discovery of improved EGF agonists using a novel in vitro screening platform. *Journal of molecular biology* **2011,** *413* (2), 406-15.
45. Chen, T. S.; Keating, A. E., Designing specific protein-protein interactions using computation, experimental library screening, or integrated methods. *Protein science : a publication of the Protein Society* **2012,** *21* (7), 949-63.
46. Sammond, D. W.; Bosch, D. E.; Butterfoss, G. L.; Purbeck, C.; Machius, M.; Siderovski, D. P.; Kuhlman, B., Computational design of the sequence and structure of a protein-binding peptide. *Journal of the American Chemical Society* **2011,** *133* (12), 4190-2.
47. Das, R.; Baker, D., Macromolecular modeling with rosetta. *Annual review of biochemistry* **2008,** *77*, 363-82.
48. Der, B. S.; Kuhlman, B., Biochemistry. From computational design to a protein that binds. *Science* **2011,** *332* (6031), 801-2.
49. Whitehead, T. A.; Chevalier, A.; Song, Y.; Dreyfus, C.; Fleishman, S. J.; De Mattos, C.; Myers, C. A.; Kamisetty, H.; Blair, P.; Wilson, I. A.; Baker, D., Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nature biotechnology* **2012,** *30* (6), 543-8.
50. Kajander, T.; Cortajarena, A. L.; Regan, L., Consensus design as a tool for engineering repeat proteins. *Methods Mol Biol* **2006,** *340*, 151-70.
51. Magliery, T. J.; Regan, L., Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif. *Journal of molecular biology* **2004,** *343* (3), 731-45.
52. Binz, H. K.; Stumpp, M. T.; Forrer, P.; Amstutz, P.; Pluckthun, A., Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *Journal of molecular biology* **2003,** *332* (2), 489-503.
53. Agrawal, C. M.; Ray, R. B., Biodegradable polymeric scaffolds for musculoskeletal tissue engineering. *J. Biomed. Mater. Res.* **2001,** *55* (2), 141-150.

54. Langer, R.; Vacanti, J. P., TISSUE ENGINEERING. *Science* **1993,** *260* (5110), 920-926.
55. Tsubouchi, M.; Matsui, S.; Banno, Y.; Kurokawa, K.; Kawakami, K., Overview of the clinical application of regenerative medicine products in Japan. *Health Policy* **2008,** *88* (1), 62-72.
56. von der Mark, K.; Park, J.; Bauer, S.; Schmuki, P., Nanoscale engineering of biomimetic surfaces: cues from the extracellular matrix. *Cell Tissue Res.* **2010,** *339* (1), 131-153.
57. Gomes, S.; Leonor, I. B.; Mano, J. F.; Reis, R. L.; Kaplan, D. L., Natural and genetically engineered proteins for tissue engineering. *Prog. Polym. Sci.* **2012,** *37* (1), 1-17.
58. Engler, A. J.; Sen, S.; Sweeney, H. L.; Discher, D. E., Matrix elasticity directs stem cell lineage specification. *Cell* **2006,** *126* (4), 677-689.
59. Brinckmann, J., Collagens at a glance. In *Collagen: Primer in Structure, Processing and Assembly*, Brinckmann, J.; Notbohm, H.; Muller, P. K., Eds. 2005; Vol. 247, pp 1-6.
60. Rouse, J. G.; Van Dyke, M. E., A Review of Keratin-Based Biomaterials for Biomedical Applications. *Materials* **2010,** *3* (2), 999-1014.
61. Vrhovski, B.; Weiss, A. S., Biochemistry of tropoelastin. *Eur. J. Biochem.* **1998,** *258* (1), 1-18.
62. Pankov, R.; Yamada, K. M., Fibronectin at a glance. *J. Cell Sci.* **2002,** *115* (20), 3861-3863.
63. McKeownLongo, P. J.; Panetti, T. S., Structure and function of vitronectin. *Trends in Glycoscience and Glycotechnology* **1996,** *8* (43), 327-340.
64. Kohn, J.; Welsh, W. J.; Knight, D., A new approach to the rationale discovery of polymeric biomaterials. *Biomaterials* **2007,** *28* (29), 4171-4177.
65. Williams, D. F., On the nature of biomaterials. *Biomaterials* **2009,** *30* (30), 5897-5909.
66. Silva, S. S.; Mano, J. F.; Reis, R. L., Potential applications of natural origin polymer-based systems in soft tissue regeneration. *Crit. Rev. Biotechnol.* **2010,** *30* (3), 200-221.
67. Mano, J. F.; Silva, G. A.; Azevedo, H. S.; Malafaya, P. B.; Sousa, R. A.; Silva, S. S.; Boesel, L. F.; Oliveira, J. M.; Santos, T. C.; Marques, A. P.; Neves, N. M.; Reis, R. L., Natural origin biodegradable systems in tissue engineering and regenerative medicine: present status and some moving trends. *Journal of the Royal Society Interface* **2007,** *4* (17), 999-1030.
68. Goldberg, M.; Langer, R.; Jia, X. Q., Nanostructured materials for applications in drug delivery and tissue engineering. *Journal of Biomaterials Science-Polymer Edition* **2007,** *18* (3), 241-268.
69. Nectow, A. R.; Marra, K. G.; Kaplan, D. L., Biomaterials for the Development of Peripheral Nerve Guidance Conduits. *Tissue Engineering Part B-Reviews* **2012,** *18* (1), 40-50.
70. Pakulska, M. M.; Ballios, B. G.; Shoichet, M. S., Injectable hydrogels for central nervous system therapy. *Biomedical Materials* **2012,** *7* (2).
71. Silva, R.; Fabry, B.; Boccaccini, A. R., Fibrous protein-based hydrogels for cell encapsulation. *Biomaterials* **2014,** *35* (25), 6727-6738.
72. Vepari, C.; Kaplan, D. L., Silk as a biomaterial. *Prog. Polym. Sci.* **2007,** *32* (8-9), 991-1007.
73. Wittmer, C. R.; Phelps, J. A.; Saltzman, W. M.; Van Tassel, P. R., Fibronectin terminated multilayer films: Protein adsorption and cell attachment studies. *Biomaterials* **2007,** *28* (5), 851-860.
74. Schleicher, I.; Parker, A.; Leavesley, D.; Crawford, R.; Upton, Z.; Xiao, Y., Surface modification by complexes of vitronectin and growth factors for serum-free culture of human osteoblasts. *Tissue Eng.* **2005,** *11* (11-12), 1688-1698.

75. Kuratomi, Y.; Nomizu, M.; Tanaka, K.; Ponce, M. L.; Komiyama, S.; Kleinman, H. K.; Yamada, Y., Laminin gamma 1 chain peptide, C-16 (KAFDITYVRLKF), promotes migration, MMP-9 secretion, and pulmonary metastasis of B16-F10 mouse melanoma cells. *Br. J. Cancer* **2002,** *86* (7), 1169-1173.
76. Park, S. H.; Park, S. R.; Chung, S. I.; Pai, K. S.; Min, B. H., Tissue-engineered cartilage using fibrin/hyaluronan composite gel and its in vivo implantation. *Artif. Organs* **2005,** *29* (10), 838-845.
77. Bauer, A. J. P.; Liu, J. Z.; Windsor, J.; Song, F. Y.; Li, B. B., Current Development of Collagen-Based Biomaterials for Tissue Repair and Regeneration. *Soft Materials* **2014,** *12* (4), 359-370.
78. O'Rourke, C.; Drake, R. A. L.; Cameron, G. W. W.; Loughlin, A. J.; Phillips, J. B., Optimising contraction and alignment of cellular collagen hydrogels to achieve reliable and consistent engineered anisotropic tissue. *J. Biomater. Appl.* **2015,** *30* (5), 599-607.
79. Parenteau-Bareil, R.; Gauvin, R.; Berthod, F., Collagen-Based Biomaterials for Tissue Engineering Applications. *Materials* **2010,** *3* (3), 1863-1887.
80. Chattopadhyay, S.; Raines, R. T., Collagen-Based Biomaterials for Wound Healing. *Biopolymers* **2014,** *101* (8), 821-833.
81. Ferreira, A. M.; Gentile, P.; Chiono, V.; Ciardelli, G., Collagen for bone tissue regeneration. *Acta Biomater.* **2012,** *8* (9), 3191-3200.
82. Kawazoe, N.; Inoue, C.; Tateishi, T.; Chen, G. P., A Cell Leakproof PLGA-Collagen Hybrid Scaffold for Cartilage Tissue Engineering. *Biotechnol. Progr.* **2010,** *26* (3), 819-826.
83. Villa, M. M.; Wang, L. P.; Huang, J. P.; Rowe, D. W.; Wei, M., Bone tissue engineering with a collagen-hydroxyapatite scaffold and culture expanded bone marrow stromal cells. *Journal of Biomedical Materials Research Part B-Applied Biomaterials* **2015,** *103* (2), 243-253.
84. Hu, X.; Wang, X.; Rnjak, J.; Weiss, A. S.; Kaplan, D. L., Biomaterials derived from silk–tropoelastin protein systems. *Biomaterials* **2010,** *31* (32), 8121-8131.
85. Almine, J. F.; Bax, D. V.; Mithieux, S. M.; Nivison-Smith, L.; Rnjak, J.; Waterhouse, A.; Wise, S. G.; Weiss, A. S., Elastin-based materials. *Chem. Soc. Rev.* **2010,** *39* (9), 3371-3379.
86. Smits, F. C. M.; Buddingh, B. C.; van Eldijk, M. B.; van Hest, J. C. M., Elastin-Like Polypeptide Based Nanoparticles: Design Rationale Toward Nanomedicine. *Macromol. Biosci.* **2015,** *15* (1), 36-51.
87. Rodriguez-Cabello, J. C.; Arias, F. J.; Rodrigo, M. A.; Girotti, A., Elastin-like polypeptides in drug delivery. *Adv. Drug Del. Rev.* **2016,** *97*, 85-100.
88. Baujard-Lamotte, L.; Noinville, S.; Goubard, F.; Marque, P.; Pauthe, E., Kinetics of conformational changes of fibronectin adsorbed onto model surfaces. *Colloids and Surfaces B-Biointerfaces* **2008,** *63* (1), 129-137.
89. Dolatshahi-Pirouz, A.; Jensen, T.; Foss, M.; Chevallier, J.; Besenbacher, F., Enhanced Surface Activation of Fibronectin upon Adsorption on Hydroxyapatite. *Langmuir* **2009,** *25* (5), 2971-2978.
90. Kadowaki, K.; Matsusaki, M.; Akashi, M., Control of Cell Surface and Functions by Layer-by-Layer Nanofilms. *Langmuir* **2010,** *26* (8), 5670-5678.
91. Li, G. C.; Yang, P.; Liao, Y. Z.; Huang, N., Tailoring of the Titanium Surface by Immobilization of Heparin/Fibronectin Complexes for Improving Blood Compatibility and Endothelialization: An in Vitro Study. *Biomacromolecules* **2011,** *12* (4), 1155-1168.
92. Schvartz, I.; Seger, D.; Shaltiel, S., Vitronectin. *Int. J. Biochem. Cell Biol.* **1999,** *31* (5), 539-544.

93. Hozumi, K.; Akizuki, T.; Yamada, Y.; Hara, T.; Urushibata, S.; Katagiri, F.; Kikkawa, Y.; Nomizu, M., Cell adhesive peptide screening of the mouse laminin alpha 1 chain G domain. *Arch. Biochem. Biophys.* **2010,** *503* (2), 213-222.
94. Kadoya, Y.; Nomizu, M.; Sorokin, L. M.; Yamashina, S.; Yamada, Y., Laminin alpha 1 chain G domain peptide, RKRLQVQLSIRT, inhibits epithelial branching morphogenesis of cultured embryonic mouse submandibular gland. *Dev. Dyn.* **1998,** *212* (3), 394-402.
95. Scheraga, H. A.; Laskowski, M., THE FIBRINOGEN-FIBRIN CONVERSION. *Adv. Protein Chem.* **1957,** *12*, 1-131.
96. Han, C. M.; Zhang, L. P.; Sun, J. Z.; Shi, H. F.; Zhou, J.; Gao, C. Y., Application of collagen-chitosan/fibrin glue asymmetric scaffolds in skin tissue engineering. *Journal of Zhejiang University-Science B* **2010,** *11* (7), 524-530.
97. Tschoeke, B.; Flanagan, T. C.; Cornelissen, A.; Koch, S.; Roehl, A.; Sriharwoko, M.; Sachweh, J. S.; Gries, T.; Schmitz-Rode, T.; Jockenhoevel, S., Development of a Composite Degradable/Nondegradable Tissue-engineered Vascular Graft. *Artif. Organs* **2008,** *32* (10), 800-809.
98. Vedakumari, W. S.; Sastry, T. P., Physiologically clotted fibrin - Preparation and characterization for tissue engineering and drug delivery applications. *Biologicals* **2014,** *42* (5), 277-284.
99. Spotnitz, W. D., Fibrin Sealant: Past, Present, and Future: A Brief Review. *World J. Surg.* **2010,** *34* (4), 632-634.
100. Zheng, Z. Z.; Liu, M.; Guo, S. Z.; Wu, J. B.; Lu, D. S.; Li, G.; Liu, S. S.; Wang, X. Q.; Kaplan, D. L., Incorporation of quantum dots into silk biomaterials for fluorescence imaging. *Journal of Materials Chemistry B* **2015,** *3* (31), 6509-6519.
101. Kapoor, S.; Kundu, S. C., Silk protein-based hydrogels: Promising advanced materials for biomedical applications. *Acta Biomater.* **2016,** *31*, 17-32.
102. Numata, K.; Kaplan, D. L., Silk-based delivery systems of bioactive molecules. *Adv. Drug Del. Rev.* **2010,** *62* (15), 1497-1508.
103. Wang, Y. Z.; Kim, H. J.; Vunjak-Novakovic, G.; Kaplan, D. L., Stem cell-based tissue engineering with silk biomaterials. *Biomaterials* **2006,** *27* (36), 6064-6082.
104. Kim, U. J.; Park, J.; Kim, H. J.; Wada, M.; Kaplan, D. L., Three-dimensional aqueous-derived biomaterial scaffolds from silk fibroin. *Biomaterials* **2005,** *26* (15), 2775-2785.
105. Spiess, K.; Lammel, A.; Scheibel, T., Recombinant Spider Silk Proteins for Applications in Biomaterials. *Macromol. Biosci.* **2010,** *10* (9), 998-1007.
106. Kluge, J. A.; Rabotyagova, U.; Leisk, G. G.; Kaplan, D. L., Spider silks and their applications. *Trends Biotechnol.* **2008,** *26* (5), 244-251.
107. Coulombe, P. A.; Fuchs, E., ELUCIDATING THE EARLY STAGES OF KERATIN FILAMENT ASSEMBLY. *J. Cell Biol.* **1990,** *111* (1), 153-169.
108. de Guzman, R. C.; Merrill, M. R.; Richter, J. R.; Hamzi, R. I.; Greengauz-Roberts, O. K.; Van Dyke, M. E., Mechanical and biological properties of keratose biomaterials. *Biomaterials* **2011,** *32* (32), 8205-8217.
109. Hill, P.; Brantley, H.; Van Dyke, M., Some properties of keratin biomaterials: Kerateines. *Biomaterials* **2010,** *31* (4), 585-593.
110. Wang, S.; Wang, Z. X.; Foo, S. E. M.; Tan, N. S.; Yuan, Y.; Lin, W. S.; Zhang, Z. Y.; Ng, K. W., Culturing Fibroblasts in 3D Human Hair Keratin Hydrogels. *Acs Applied Materials & Interfaces* **2015,** *7* (9), 5187-5198.

111. Saul, J. M.; Ellenburg, M. D.; de Guzman, R. C.; Van Dyke, M., Keratin hydrogels support the sustained release of bioactive ciprofloxacin. *Journal of Biomedical Materials Research Part A* **2011,** *98A* (4), 544-553.
112. Reichl, S.; Borrelli, M.; Geerling, G., Keratin films for ocular surface reconstruction. *Biomaterials* **2011,** *32* (13), 3375-3386.
113. Reichl, S., Films based on human hair keratin as substrates for cell culture and tissue engineering. *Biomaterials* **2009,** *30* (36), 6854-6866.
114. Nunez, F.; Trach, S.; Burnett, L.; Handa, R.; Van Dyke, M.; Callahan, M.; Smith, T., Vasoactive Properties of Keratin-Derived Compounds. *Microcirculation* **2011,** *18* (8), 663-669.
115. Lin, Y. C.; Ramadan, M.; Van Dyke, M.; Kokai, L. E.; Philips, B. J.; Rubin, J. P.; Marra, K. G., Keratin Gel Filler for Peripheral Nerve Repair in a Rodent Sciatic Nerve Injury Model. *Plast. Reconstr. Surg.* **2012,** *129* (1), 67-78.
116. Ham, T. R.; Lee, R. T.; Han, S.; Haque, S.; Vodovotz, Y.; Gu, J.; Burnett, L. R.; Tomblyn, S.; Saul, J. M., Tunable Keratin Hydrogels for Controlled Erosion and Growth Factor Delivery. *Biomacromolecules* **2016,** *17* (1), 225-236.
117. de Guzman, R. C.; Tsuda, S. M.; Ton, M. T. N.; Zhang, X.; Esker, A. R.; Van Dyke, M. E., Binding Interactions of Keratin-Based Hair Fiber Extract to Gold, Keratin, and BMP-2. *PLoS One* **2015,** *10* (8).
118. de Guzman, R. C.; Saul, J. M.; Ellenburg, M. D.; Merrill, M. R.; Coan, H. B.; Smith, T. L.; Van Dyke, M. E., Bone regeneration with BMP-2 delivered from keratose scaffolds. *Biomaterials* **2013,** *34* (6), 1644-1656.
119. Aboushwareb, T.; Eberli, D.; Ward, C.; Broda, C.; Holcomb, J.; Atala, A.; Van Dyke, M., A Keratin Biomaterial Gel Hemostat Derived from Human Hair: Evaluation in a Rabbit Model of Lethal Liver Injury. *Journal of Biomedical Materials Research Part B-Applied Biomaterials* **2009,** *90B* (1), 45-54.
120. Sierpinski, P.; Garrett, J.; Ma, J.; Apel, P.; Klorig, D.; Smith, T.; Koman, L. A.; Atala, A.; Van Dyke, M., The use of keratin biomaterials derived from human hair for the promotion of rapid regeneration of peripheral nerves. *Biomaterials* **2008,** *29* (1), 118-128.
121. Wu, F.; Jin, T., Polymer-Based Sustained-Release Dosage Forms for Protein Drugs, Challenges, and Recent Advances. *AAPS PharmSciTech* **2008,** *9* (4), 1218-1229.
122. Dinjaski, N.; Kaplan, D. L., Recombinant protein blends: silk beyond natural design. *Curr. Opin. Biotechnol.* **2016,** *39*, 1-7.
123. Harris, T. I.; Gaztambide, D. A.; Day, B. A.; Brock, C. L.; Ruben, A. L.; Jones, J. A.; Lewis, R. V., Sticky Situation: An Investigation of Robust Aqueous-Based Recombinant Spider Silk Protein Coatings and Adhesives. *Biomacromolecules* **2016,** *17* (11), 3761-3772.
124. Teng, W. B.; Cappello, J.; Wu, X. Y., Recombinant Silk-Elastinlike Protein Polymer Displays Elasticity Comparable to Elastin. *Biomacromolecules* **2009,** *10* (11), 3028-3036.
125. Zhou, M. L.; Qian, Z. G.; Chen, L.; Kaplan, D. L.; Xia, X. X., Rationally Designed Redox-Sensitive Protein Hydrogels with Tunable Mechanical Properties. *Biomacromolecules* **2016,** *17* (11), 3508-3515.
126. Jang, Y.; Champion, J. A., Self-Assembled Materials Made from Functional Recombinant Proteins. *Acc. Chem. Res.* **2016,** *49* (10), 2188-2198.
127. An, B.; Kaplan, D. L.; Brodsky, B., Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering. *Frontiers in Chemistry* **2014,** *2*.
128. Yu, Z. X.; An, B.; Ramshaw, J. A. M.; Brodsky, B., Bacterial collagen-like proteins that form triple-helical structures. *J. Struct. Biol.* **2014,** *186* (3), 451-461.

129. An, B.; DesRochers, T. M.; Qin, G. K.; Xia, X. X.; Thiagarajan, G.; Brodsky, B.; Kaplan, D. L., The influence of specific binding of collagen-silk chimeras to silk biomaterials on hMSC behavior. *Biomaterials* **2013,** *34* (2), 402-412.
130. An, B.; Abbonante, V.; Xu, H. F.; Gavriilidou, D.; Yoshizumi, A.; Bihan, D.; Farndale, R. W.; Kaplan, D. L.; Balduini, A.; Leitinger, B.; Brodsky, B., Recombinant Collagen Engineered to Bind to Discoidin Domain Receptor Functions as a Receptor Inhibitor. *J. Biol. Chem.* **2016,** *291* (9), 4343-4355.
131. An, B.; Lin, Y. S.; Brodsky, B., Collagen interactions: Drug design and delivery. *Adv. Drug Del. Rev.* **2016,** *97*, 69-84.
132. Khajavi, R.; Abbasipour, M.; Bahador, A., Electrospun biodegradable nanofibers scaffolds for bone tissue engineering. *J. Appl. Polym. Sci.* **2016,** *133* (3), 19.
133. Stoichevska, V.; An, B.; Peng, Y. Y.; Yigit, S.; Vashi, A. V.; Kaplan, D. L.; Werkmeister, J. A.; Dumsday, G. J.; Ramshaw, J. A. M., Formation of multimers of bacterial collagens through introduction of specific sites for oxidative crosslinking. *Journal of Biomedical Materials Research Part A* **2016,** *104* (9), 2369-2376.

Chapter 3. Designing Repeat Proteins for Biosensors and Medical Imaging

Rachael N. Parker, Tijana Z. Grove*

(Reproduced with permission from Portland Press: *Biochemical Society Transactions*, **2015**, *43 (5)*, 856-860)

Author Address: Department of Chemistry, Virginia Tech, Blacksburg, VA 24060

Key words: Biosensor, alternative protein scaffold, repeat proteins, molecular recognition, protein design, imaging agent

Abbreviations

DARPin, Designed ankyrin repeat protein; TPR, tetratricopeptide repeats; ARM, armadillo repeat proteins; LRR, leucine rich repeat proteins; GFP, green fluorescent protein; human epidermal growth factor receptor-2, HER2; Transcription activator-like effectors, TAL effectors; Pumilio and fem-3 binding factor proteins, Puf; Pentatricopeptide repeat, PPR; Variable lymphocyte receptor, VLR; Nod-like receptor, NLR; muramyl dipeptide, MDP

3.1 Abstract

Advances in protein engineering tools, both computational and experimental has afforded many new protein structures and functions. Here, we present a snapshot of repeat-protein engineering efforts towards new, versatile, alternative binding scaffolds for use in analytical sensors and as imaging agents. Analytical assays, sensors, and imaging agents based on the direct recognition of

certain analytes are increasingly important for biomedical research and diagnostics as well as areas in food safety and national security.

3.2 Repeat proteins as Alternative Scaffolds

The sensitivity and specificity of biosensors and imaging agents are decidedly dependent upon high-affinity, specific molecular recognition between the recognition element, i.e. affinity probe and desired target (Figure 3.1 A). Traditionally, the commonly used affinity probes are antibodies. However, a 2008 study by Berglund *et al*., showed that fewer than half of approximately 6,000 routinely used commercial antibodies were specific for only their intended target.[1-2] This lack of specificity is especially exaggerated for detection of analytes that are not highly immunogenic e.g., small molecules, peptidoglycans, or nucleic acids and for which high quality monoclonal antibodies are not available. As a result, many medically and environmentally important analytes still lack high affinity, specific recognition elements. However, the emergence of protein and peptide engineering techniques, such as synthetic libraries and selection and evolution technologies, has allowed substantial progress towards a new generation of affinity probes based on alternative protein scaffolds. Advantages of such probes are: (a) small size for ease of handling and reduced non-specific interactions; (b) the absence of aggregation, which is essential for sensitivity; (c) high chemical, thermal, and proteolytic stability; (d) ease of chemical coupling to transducer elements such as fluorophores, nanoparticles, and solid support; (e) cost-efficient production by chemical synthesis or recombinant means.[3] Additionally, alternative scaffolds can be easily amended to enhance pharmacokinetic properties such as length of half-life and clearance of probes used as imaging agents.[4] Furthermore, multifunctional scaffolds, i.e. multiple independent binding sites for the same or different targets, can be used to greatly enhance selectivity of sensors through avidity.

These advantages have been demonstrated in a number of alternative scaffold designs for therapeutics, diagnostics, and imagining devices (Figure 3.1 B).[3-8]

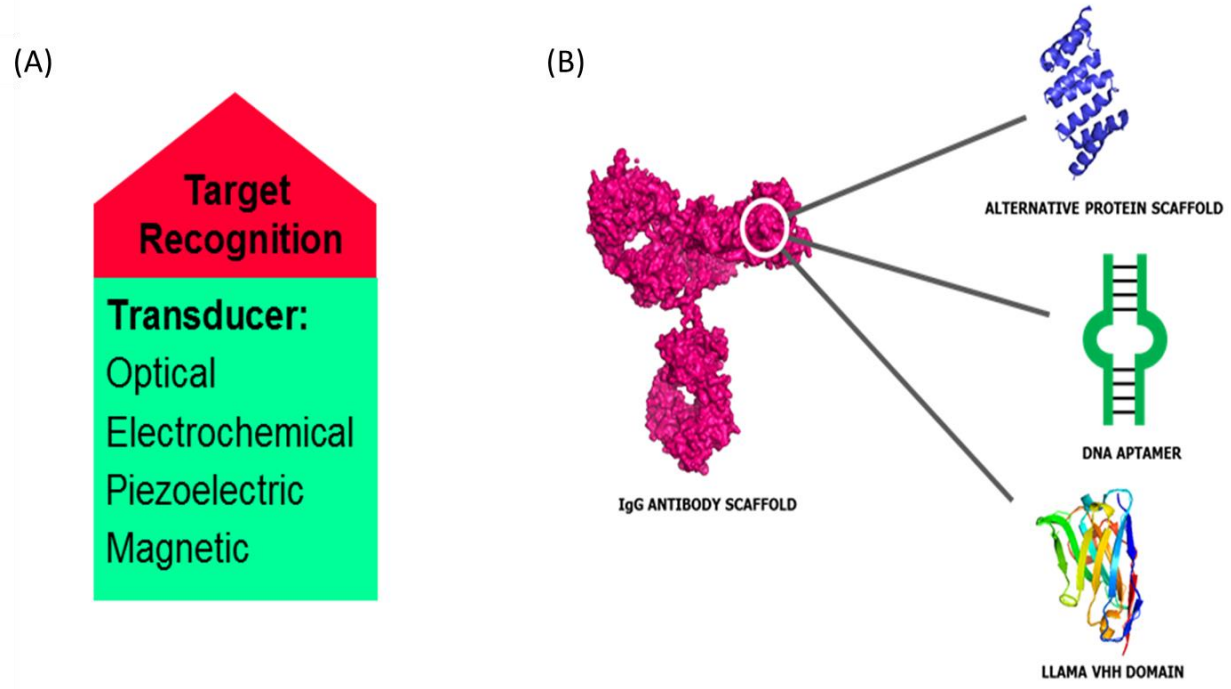


Figure 3.1. Biosensor architecture and recognition elements. **(**A) The architecture of a biosensor consists of two main components: 1. Target recognition that is required for specific and sensitive analyte detection and 2. A transducer element needed to convert target recognition into a readable signal. (B) Many alternatives to the IgG antibody scaffold have been proposed including protein scaffolds, DNA aptamers, and single chain antibody domains. The binding site of the IGG scaffold (PDB: 1HZH), indicated by the white circle is analogous to that of the CTPR3 alternative scaffold (PDB: 1NAO), DNA aptamer, and llama VHH domain (PDB: 1I3V).

The modular architecture, extended non-globular structure of repeat proteins positions them as increasingly successful alternative scaffolds.[3] Designed ankyrin repeat proteins (DARPins), tetratricopeptide repeats (TPRs), armadillo repeat proteins (ARMs), and leucine rich repeat proteins (LRRs) have all been used for development of high affinity binders.[9-13] For this reason designed repeat proteins that have been successfully used for imaging and/or biosensing applications will remain the focus of this review.

3.3 Methods for Generation of New Protein Scaffolds

A defining feature of repeat protein structure is that both the individual repeats, and their positions relative to each other, are the same regardless of the protein in which they occur.[14] Upon ligand binding there is typically little, if any, conformational change. Thus repeat proteins are regarded as constant frameworks that are 'decorated' with a binding site. It is therefore not surprising that consensus sequence proteins are common starting points for the design of functional repeat proteins. Additionally, recent computational designs were used to create repeat protein scaffolds with specific shape complementary to that of the target ligand.[15-16] For instances where the desired analyte is another protein, this approach can potentially increase affinity and selectivity due to the shape complementarity of the binder and the target. Parmeggiani *et al*. demonstrated a general computational approach to designing repeat proteins that resulted in well expressed, stable scaffolds for Ankyrins, TPRs, LRRs, ARMs, HEAT repeats, and WD40 repeats.[17] The importance of shape complementarity has also been the focus of computational design efforts as highlighted in the work of Park *et al*. and Ramisch *et al*. where novel design approaches are used to create LRRs with a predetermined curvature.[15-16]

Experimentally, the design of new binders requires methods that can rapidly and robustly assemble and screen protein libraries that contain many millions of unique sequences. Recently Speltz *et al*. described a new cost-effective and efficient method for library creation.[18] This strategy utilizes the fluorescence from green fluorescent protein (GFP) to determine colonies that contain the desired insert after ligation. This "white and green screen" is not limited to GFP and can be easily implemented with any fluorescent protein.

Protein display methods, such as yeast, phage, and ribosome display, have been successfully employed for generation of high affinity binders. Libraries of LRRs, TPRs, and

DARPins have been screened to identify proteins with affinities for desired targets that are equivalent to and even surpassing that of antibodies.[19-22] Additionally, *in vivo* screening strategies, such as the split-GFP reassembly, have been used to screen TPR libraries.[23] TPR binders, T-mods, were then successfully used to modulate protein-protein interactions *in vivo*.[24]

3.4 Scaffolds for Recognition of Protein and Peptide Targets

DARPin G3 designed to specifically bind human epidermal growth factor receptor-2 (HER2) was used to image HER2 overexpression on breast cancer cells *in vivo* via selective attachment of ^{111}In-DOTA to a C-terminal cysteine residue (Figure 3.2 A).[25] This labeling strategy is generally applicable to other repeat-protein scaffolds and thus holds promise for future clinical use.

Design of binders through engineering repeating units and adding or removing repeats to create binding scaffolds increases their utility as alternative molecular recognition modules.[14] The modular recognition would allow for efficient generation of a virtually unlimited repertoire of binders and remove the need for evolution of binding affinity to individual targets. Plucktuhn's group used this strategy to design ARMs that selectively bind the peptide, Neurotensin.[11] This work illustrates the potential of modular peptide binders from ARM repeats based on the binding interaction of a repeat per dipeptide.

In an analogous strategy, design of TPR modules was used for detection of phosphopeptide-protein interaction in *E. coli* using split mCherry assembly as a readout.[26] The tetratricopeptide repeat affinity protein (TRAP) was designed to selectively recognize phosphorylated versus nonphosphorylated peptides in a sequence dependent manner. TRAP binding to the peptide is modular: amino acid "pockets" are designed to bind specific amino acid

sequences in the target peptide. Thus through mixing and matching of TRAP pockets virtually any peptide sequence can be recognized.

Another repeat protein scaffold, HEAT, has been used as a promising universal scaffold that can bind a variety of structurally and functionally diverse protein targets with micromolar to nanomolar affinities.[27] In this study, a library of HEAT proteins was created to bind four predefined protein targets with specific secondary structure demonstrating the ability to distinguish between proteins with differing characteristics.[27]

3.5 Scaffolds for Nucleic Acid Recognition

Recognition of nucleic acid targets has been an area of growing interest due to the potential for control of cellular transcription and translation.[28] Although not imaging and biosensing tools in the strict sense, these binding modules are important for systems biology research. The majority of DNA and RNA binding modules are based on the re-design of three classes of endogenous nucleic acid binders: Transcription activator-like effectors (TAL effectors)[29], Pumilio and fem-3 binding factor proteins (Puf)[28], and Pentatricopeptide repeats (PPRs)[30]. Vibrant research into better understanding of natural ligand recognition properties of these proteins has resulted in novel nucleic acid recognition modules.[31-35]

3.5.1 Transcription Activator-like Effectors

TAL effectors are 33-35 amino acids repeat containing proteins implicated in host gene expression. Two hypervariable amino acid residues in each repeat interact with one base pair in the target DNA. Several *de novo* TAL proteins were designed computationally and have demonstrated the ability to target specific DNA sequences resulting in selective control of gene transcription.[29, 36]

3.5.2 Pumilio and Fem-3 binding Factor Proteins

Puf proteins have drawn much interest due to their RNA recognition properties.[28] Eight consecutive Puf repeats comprise a 3-helix bundle that in turn contacts a single RNA base through "tripartite recognition motif" of 3 amino acids.[28] This modular recognition suggests great potential engineering affinity for specific RNA sequences. Yoshimura *et al*. demonstrate this in a recent example where they design a fluorescently tagged pumilio module for specific binding and imaging of β-actin mRNA *in vivo* (Figure 3.2 B).[37]

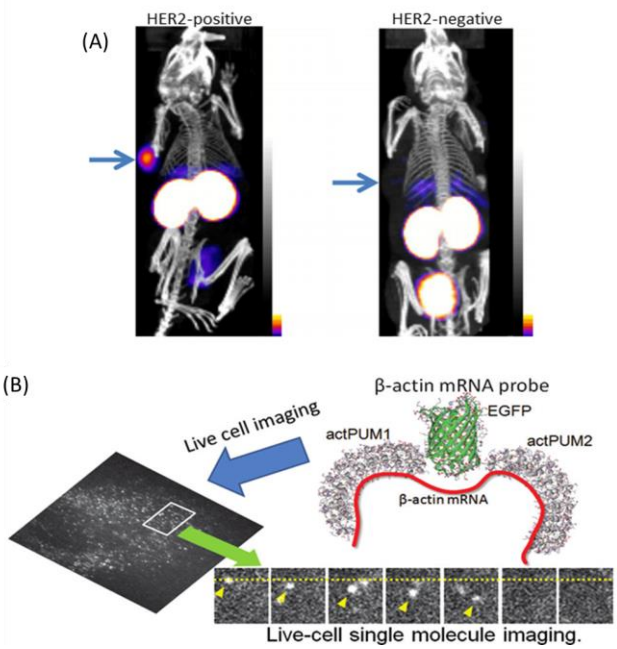


Figure 3.2. Examples of designed repeat protein binders for imaging applications (A) microSPECT/CT scan of radiolabeled DARPin G3 in SCID-beige mice bearing: a HER2-positive human breast tumor and HER2-negative human breast tumor (MDA-MB-468). Tumors, indicated by arrows, were analyzed at same sensitivity level. (Reprinted with permission from Goldstein *et al.* [25]). (B) A Pumilio homology domain (PUM-HD) was designed by Yoshimura et al. to recognize the sequence of β-actin mRNA and subsequently developed into an mRNA probe. The probe, consisting of two PUM-HD mutants flanking full-length enhanced green fluorescent protein (EGFP), successfully and specifically labeled β-actin mRNA in the cytosol as seen through fluorescence microscopy with the probe in living cells. Fluorescent spots from the probe were colocalized with microtubules and moved directionally in living cells indicating the potential for visualization of β-actin mRNA localization and dynamics in living cells. (Reprinted with permission from Yoshimura et al. [37])

3.5.3 Pentratricopeptide Repeat Proteins

PPRs are essential proteins for the regulation of many facets of RNA metabolism in eukaryotic cells; however, understanding of their natural function has been hindered by poor solubility and limited success of computational methods. Using a consensus based approach, Coquille *et al*. designed a synthetic, stable PPR scaffold for RNA binding.[30] This study confirmed that PPRs bind RNA according to the "PPR code" where amino acids at positions 4 and 34 determine specificity for certain nucleotides.[30] Advances such as this will enable further elucidation and exploitation of PPR-RNA recognition capabilities.

3.5.4 Designed Ankyrin Repeat Proteins

In parallel, the ever versatile DARPin framework was recently used for selective and specific DNA binding. Scholz *et al*. describe the redesign of DARPins for selective recognition of human telomere G-quadruplex DNA with nanomolar affinity.[38]

3.6 Other Biomolecular Targets

More than 100 different post translational protein modifications act as an on/off switch of function, activity, and stability. Detection of these events, such as phosphorylation and glycosylation, is thus biomedically important. Several carbohydrate binding lectins and mammalian antibodies have been investigated for specific binding to glycosylated proteins; however, these binders generally suffer from low affinity and broad specificity thus limiting their clinical utility.[39-42] Variable lymphocyte receptors (VLRs), which are leucine rich repeat proteins involved in the adaptive immunity of jawless vertebrates, have been successfully engineered for recognition of several glycosylated proteins.[12, 43-44] For example, Hong *et al*. describe a yeast surface display screen of VLRs from lamprey that enabled the selection of binders for a number of glycan targets with nanomolar affinities.[43]

Our group endeavors to engineer repeat protein modules for use in whole cell pathogen sensing. For that, we are drawing inspiration from the pattern recognition receptors, such as Nod-like receptors (NLRs) in the innate immune system. Analogous to manmade sensors, NLR proteins have an architecture where the recognition and transducer elements are easily identifiable.[45] Previous studies have shown that these proteins utilize LRR domains to bind a large repertoire of chemically diverse ligands including bacterial cell wall peptidoglycans, glycolipids, and bacterial RNA. We have recently used consensus sequence approach to create an artificial LRR, CLRR2, with micromolar binding affinity to muramyl dipeptide (MDP), a bacterial cell wall component.[13] Surprisingly, CLRR2 is a specific binder for the physiologically relevant isomer L-D MDP and does not bind D-D MDP or N-glycolyl MDP. This strict discrimination between similar ligands suggests that CLRR scaffolds are promising candidates for further design to specifically bind other biomolecules.

3.7 Summary

Here we have presented a snapshot of protein engineering efforts for design of modular binding scaffolds for analytical assays, sensors, and imaging applications. The approaches described here for design of scaffolds are applicable for a variety of ligands. Assays and sensors based on direct analyte recognition have important applications not only in medical research and diagnostics but also in food industry, environmental science, forensic science, and national security.

3.8 Acknowledgements

We acknowledge members of the Grove lab for critical reading of the manuscript and useful discussion.

3.9 References

1. Bradbury, A.; Plueckthun, A., Standardize antibodies used in research. *Nature* **2015,** *518* (7537), 27-29.
2. Berglund, L.; Bjoerling, E.; Oksvold, P.; Fagerberg, L.; Asplund, A.; Szigyarto, C. A.-K.; Persson, A.; Ottosson, J.; Wernerus, H.; Nilsson, P.; Lundberg, E.; Sivertsson, A.; Navani, S.; Wester, K.; Kampf, C.; Hober, S.; Ponten, F.; Uhlen, M., A Genecentric Human Protein Atlas for Expression Profiles Based on Antibodies. *Mol. Cell. Proteomics* **2008,** *7* (10), 2019-2027.
3. Boersma, Y. L.; Plueckthun, A., DARPins and other repeat protein scaffolds: advances in engineering and applications. *Curr. Opin. Biotechnol.* **2011,** *22* (6), 849-857.
4. Lofblom, J.; Frejd, F. Y.; Stahl, S., Non-immunoglobulin based protein scaffolds. *Curr. Opin. Biotechnol.* **2011,** *22* (6), 843-848.
5. Liu, S.; Liu, H.; Ren, G.; Kimura, R. H.; Cochran, J. R.; Cheng, Z., PET Imaging of Integrin Positive Tumors Using F-18 Labeled Knottin Peptides. *Theranostics* **2011,** *1*, 403-412.
6. Skerra, A., Alternative binding proteins: Anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *FEBS J.* **2008,** *275* (11), 2677-2683.
7. Gebauer, M.; Skerra, A., Engineered protein scaffolds as next-generation antibody therapeutics. *Curr. Opin. Chem. Biol.* **2009,** *13* (3), 245-255.
8. Nygren, P.-A., Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS J.* **2008,** *275* (11), 2668-2676.
9. Verdurmen, W. P. R.; Luginbuhl, M.; Honegger, A.; Pluckthun, A., Efficient cell-specific uptake of binding proteins into the cytoplasm through engineered modular transport systems. *J. Controlled Release* **2015,** *200*, 13-22.
10. Cortajarena, A. L.; Yi, F.; Regan, L., Designed TPR modules as novel anticancer agents. *ACS Chem. Biol.* **2008,** *3* (3), 161-166.
11. Varadamsetty, G.; Tremmel, D.; Hansen, S.; Parmeggiani, F.; Plueckthun, A., Designed Armadillo Repeat Proteins: Library Generation, Characterization and Selection of Peptide Binders with High Specificity. *J. Mol. Biol.* **2012,** *424* (1-2), 68-87.
12. Luo, M.; Velikovsky, C. A.; Yang, X.; Siddiqui, M. A.; Hong, X.; Barchi, J. J., Jr.; Gildersleeve, J. C.; Pancer, Z.; Mariuzza, R. A., Recognition of the Thomsen-Friedenreich Pancarcinoma Carbohydrate Antigen by a Lamprey Variable Lymphocyte Receptor. *J. Biol. Chem.* **2013,** *288* (32), 23597-23606.
13. Parker, R.; Mercedes-Camacho, A.; Grove, T. Z., Consensus design of a NOD receptor leucine rich repeat domain with binding affinity for a muramyl dipeptide, a bacterial cell wall fragment. *Protein Sci.* **2014,** *23* (6), 790-800.
14. Grove, T. Z.; Cortajarena, A. L.; Regan, L., Ligand binding by repeat proteins: natural and designed. *Curr. Opin. Struct. Biol.* **2008,** *18* (4), 507-515.
15. Park, K.; Shen, B. W.; Parmeggiani, F.; Huang, P.-S.; Stoddard, B. L.; Baker, D., Control of repeat-protein curvature by computational protein design. *Nat. Struct. Mol. Biol.* **2015,** *22* (2), 167-174.
16. Ramisch, S.; Weininger, U.; Martinsson, J.; Akke, M.; Andre, I., Computational design of a leucine-rich repeat protein with a predefined geometry. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (50), 17875-17880.

17. Parmeggiani, F.; Huang, P.-S.; Vorobiev, S.; Xiao, R.; Park, K.; Caprari, S.; Su, M.; Seetharaman, J.; Mao, L.; Janjua, H.; Montelione, G. T.; Hunt, J.; Baker, D., A General Computational Approach for Repeat Protein Design. *J. Mol. Biol.* **2015,** *427* (2), 563-575.
18. Speltz, E. B.; Regan, L., White and green screening with circular polymerase extension cloning for easy and reliable cloning. *Protein Sci.* **2013,** *22* (6), 859-864.
19. Xu, G.; Tasumi, S.; Pancer, Z., Yeast Surface Display of Lamprey Variable Lymphocyte Receptors. In *Immune Receptors: Methods and Protocols*, Rast, J. P.; Booth, J. W. D., Eds. 2011; Vol. 748, pp 21-33.
20. Kummer, L.; Parizek, P.; Rube, P.; Millgramm, B.; Prinz, A.; Mittl, P. R. E.; Kaufholz, M.; Zimmermann, B.; Herberg, F. W.; Plueckthun, A., Structural and functional analysis of phosphorylation-specific binders of the kinase ERK from designed ankyrin repeat protein libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (34), E2248-E2257.
21. Schilling, J.; Schoeppe, J.; Plueckthun, A., From DARPins to LoopDARPins: Novel LoopDARPin Design Allows the Selection of Low Picomolar Binders in a Single Round of Ribosome Display. *J. Mol. Biol.* **2014,** *426* (3), 691-721.
22. Petters, E.; Krowarsch, D.; Otlewski, J., Design, expression and characterization of a highly stable tetratricopeptide-based protein scaffold for phage display application. *Acta Biochim. Pol.* **2013,** *60* (4), 585-590.
23. Jackrel, M. E.; Cortajarena, A. L.; Liu, T. Y.; Regan, L., Screening Libraries To Identify Proteins with Desired Binding Activities Using a Split-GFP Reassembly Assay. *ACS Chem. Biol.* **2010,** *5* (6), 553-562.
24. Jackrel, M. E.; Cortajarena, A. L.; Liu, T. Y.; Regan, L., Screening Libraries To Identify Proteins with Desired Binding Activities Using a Split-GFP Reassembly Assay. *ACS Chem. Biol.* **2010,** *5* (6), 553-562.
25. Goldstein, R.; Sosabowski, J.; Livanos, M.; Leyton, J.; Vigor, K.; Bhavsar, G.; Nagy-Davidescu, G.; Rashid, M.; Miranda, E.; Yeung, J.; Tolner, B.; Plueckthun, A.; Mather, S.; Meyer, T.; Chester, K., Development of the designed ankyrin repeat protein (DARPin) G3 for HER2 molecular imaging. *Eur. J. Nucl. Med. Mol. Imaging* **2015,** *42* (2), 288-301.
26. Sawyer, N.; Gassaway, B. M.; Haimovich, A. D.; Isaacs, F. J.; Rinehart, J.; Regan, L., Designed Phosphoprotein Recognition in Escherichia coli. *ACS Chem. Biol.* **2014,** *9* (11), 2502-2507.
27. Guellouz, A.; Valerio-Lepiniec, M.; Urvoas, A.; Chevrel, A.; Graille, M.; Fourati-Kammoun, Z.; Desmadril, M.; van Tilbeurgh, H.; Minard, P., Selection of Specific Protein Binders for Pre-Defined Targets from an Optimized Library of Artificial Helicoidal Repeat Proteins (alphaRep). *PLoS One* **2013,** *8* (8).
28. Campbell, Z. T.; Valley, C. T.; Wickens, M., A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat. Struct. Mol. Biol.* **2014,** *21* (8), 732-738.
29. Boch, J.; Scholze, H.; Schornack, S.; Landgraf, A.; Hahn, S.; Kay, S.; Lahaye, T.; Nickstadt, A.; Bonas, U., Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **2009,** *326* (5959), 1509-1512.
30. Filipovska, A.; Razif, M. F. M.; Nygard, K. K. A.; Rackham, O., A universal code for RNA recognition by PUF proteins. *Nat. Chem. Biol.* **2011,** *7* (7), 425-427.
31. Jenkins, H. T.; Baker-Wilding, R.; Edwards, T. A., Structure and RNA binding of the mouse Pumilio-2 Puf domain. *J. Struct. Biol.* **2009,** *167* (3), 271-276.

32. Mak, A. N.-S.; Bradley, P.; Bogdanove, A. J.; Stoddard, B. L., TAL effectors: function, structure, engineering and applications. *Curr. Opin. Struct. Biol.* **2013,** *23* (1), 93-99.
33. Kubik, G.; Batke, S.; Summerer, D., Programmable Sensors of 5-Hydroxymethylcytosine. *J. Am. Chem. Soc.* **2015,** *137* (1), 2-5.
34. Garg, A.; Lohmueller, J. J.; Silver, P. A.; Armel, T. Z., Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res.* **2012,** *40* (15), 7584-7595.
35. Doyle, E. L.; Booher, N. J.; Standage, D. S.; Voytas, D. F.; Brendel, V. P.; VanDyk, J. K.; Bogdanove, A. J., TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* **2012,** *40* (W1), W117-W122.
36. Yoshimura, H.; Inaguma, A.; Yamada, T.; Ozawa, T., Fluorescent Probes for Imaging Endogenous beta-Actin mRNA in Living Cells Using Fluorescent Protein-Tagged Pumilio. *ACS Chem. Biol.* **2012,** *7* (6), 999-1005.
37. Coquille, S.; Rajappa, L.; Thore, S.; Filipovska, A.; Rackham, O.; Chia, T.; Lingford, J. P.; Razif, M. F. M., An artificial PPR scaffold for programmable RNA recognition. *Nat Commun* **2014,** *5*, 5729.
38. Scholz, O.; Hansen, S.; Plueckthun, A., G-quadruplexes are specifically recognized and distinguished by selected designed ankyrin repeat proteins. *Nucleic Acids Res.* **2014,** *42* (14), 9182-9194.
39. Almogren, A.; Abdullah, J.; Ghapure, K.; Ferguson, K.; Glinsky, V. V.; Rittenhouse-Olson, K., Anti-Thomsen-Friedenreich-Ag (anti-TF-Ag) potential for cancer therapy. *Frontiers in bioscience (Scholar edition)* **2012,** *4*, 840-63.
40. Fujimoto, Y. K.; Green, D. F., Carbohydrate Recognition by the Antiviral Lectin Cyanovirin-N. *J. Am. Chem. Soc.* **2012,** *134* (48), 19639-19651.
41. Li, Q.; Anver, M. R.; Li, Z. T.; Butcher, D. O.; Gildersleeve, J. C., GalNAc alpha 1-3Gal, a new prognostic marker for cervical cancer. *Int. J. Cancer* **2010,** *126* (2), 459-468.
42. Woodrum, B. W.; Maxwell, J. D.; Bolia, A.; Ozkan, S. B.; Ghirlanda, G., The antiviral lectin cyanovirin-N: probing multivalency and glycan recognition through experimental and computational approaches. *Biochem. Soc. Trans.* **2013,** *41*, 1170-1176.
43. Hong, X.; Ma, M. Z.; Gildersleeve, J. C.; Chowdhury, S.; Barchi, J. J., Jr.; Mariuzza, R. A.; Murphy, M. B.; Mao, L.; Pancer, Z., Sugar-Binding Proteins from Fish: Selection of High Affinity "Lambodies" That Recognize Biomedically Relevant Glycans. *ACS Chem. Biol.* **2013,** *8* (1), 152-160.
44. Lee, S.-C.; Park, K.; Han, J.; Lee, J.-J.; Kim, H. J.; Hong, S.; Heu, W.; Kim, Y. J.; Ha, J.-S.; Lee, S.-G.; Cheong, H.-K.; Jeon, Y. H.; Kim, D.; Kim, H.-S., Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (9), 3299-3304.
45. Martinon, F.; Mayor, A.; Tschopp, J., The Inflammasomes: Guardians of the Body. In *Annu. Rev. Immunol.*, 2009; Vol. 27, pp 229-265.

Chapter 4. Consensus Design of a NOD Receptor Leucine- rich Repeat Domain with Binding Affinity for a Muramyl Dipeptide (MDP), a Bacterial Cell Wall Fragment

Rachael Parker, Ana Mercedes-Camacho, Tijana Z. Grove*

(Reproduced with permission from John Wiley & Sons, Inc.

Protein Science, **2014**, *23 (6)*, 790-800)

Author Address: Department of Chemistry, Virginia Tech, Blacksburg, VA 24060

Keywords: Leucine rich repeat, NOD, MDP, consensus design, binding scaffold, repeat protein

4.1 Abstract

Repeat proteins have recently emerged as especially well-suited alternative binding scaffolds due to their modular architecture and biophysical properties. Here we present the design of a scaffold based on the consensus sequence of the leucine rich repeat (LRR) domain of the NOD family of cytoplasmic innate immune system receptors. Consensus sequence design has emerged as a protein design tool to create de novo proteins that capture sequence-structure relationships and interactions present in nature. The multiple sequence alignment of 311 individual LRRs, which are the putative ligand-recognition domain in NOD proteins, resulted in a consensus sequence protein containing two internal and N- and C- capping repeats named CLRR2. CLRR2

protein is a stable, monomeric, and cysteine free scaffold that without any affinity maturation displays micromolar binding to muramyl dipeptide, a bacterial cell wall fragment. To our knowledge, this is the first report of direct interaction of a NOD LRR with a physiologically relevant ligand.

4.2 Introduction

The advancement of protein engineering tools, such as synthetic libraries and evolution and selection technologies, has allowed substantial scientific progress towards the design of protein scaffolds with novel binding specificities.[46-47] However, most of the engineered protein scaffolds to date were directed against protein targets.[47] Other classes of biomolecules such as carbohydrates and nucleic acids have not been the focus of scaffold development until very recently.[34, 40, 42, 48-53] The need for alternative scaffolds with affinity and specificity for carbohydrates is especially pronounced since they are inefficient antigens for the adaptive immune system, and traditional antibodies are difficult, if not impossible, to produce.[54] Most readily available carbohydrate binding proteins, such as lectins and antibodies, typically display either broad specificity or low affinity for their antigen. Additionally, carbohydrate-binding proteins are available for only a small fraction of known glycans, in spite of their abundance and importance.[55]

Recently, repeat proteins emerged as especially well-suited binding scaffolds due to their modular structure that facilitates the binding of a variety of non-related protein and peptide ligands. Many repeat proteins, such as ankyrin repeats (ANK), tetratricopeptide repeats (TPR), leucine rich repeats (LRR), and armadillo (ARM) repeats have been successfully used for the development of novel binding scaffolds.[50, 56-62] The evolutionary advantage of a modular

architecture is the possibility to evolve the function through not only point mutations, but also by shuffling, deletion and/or insertion of repeats. This property makes repeat proteins a particularly attractive system for functional protein engineering.[63-70] For example, a specific set of tandem repeats can target a particular cellular compartment, while another set of tandem repeats binds specific endogenous ligands. These types of scaffolds would be beneficial for functional genomics, *in vivo* imaging, and drug delivery applications.

In contrast to the adaptive immune system, which uses the immunoglobulin scaffolds for ligand binding, the innate immune system relies primarily on LRR protein motifs for target recognition.[71] In mammals, two main protein families of such receptors have been identified: extracellular Toll-like receptors (TLRs) and cytoplasmic Nod-like receptors (NLRs). The common feature of both families is the presence of the LRR motif.[72] Co-crystal structures of TLR receptors with their ligand indicate that the LRR domain is the ligand binding site.[73] In analogy to TLRs, it is proposed that NLRs also bind ligands using their LRR motif.[72] Studies of cytoplasmic NLRs showed that these proteins bind a large repertoire of ligands including bacterial cell-wall peptidoglycans, bacterial RNA, uric crystals, and antiviral imidazoquinone.[72, 74] Thus, we hypothesized that LRR motifs from NLR proteins are especially well poised to function as a framework for development of glycan and nucleotide binding scaffolds since chemically similar types of molecules are within the repertoire of their natural ligands.

Here we describe the design of a peptidoglycan binding protein scaffold based on the LRR domain present in a NOD subgroup of NLR receptors of vertebrates.[75] The consensus sequence design resulted in a stable, monomeric, and cysteine free scaffold that without any affinity maturation, displays micromolar binding to the muramyl dipeptide, a bacterial cell wall fragment.

4.3 Results and Discussion

4.3.1 Repeat Protein Scaffolds

Repeat proteins are a ubiquitous class of proteins characterized by successive homology motifs that stack in tandem.[64, 69, 76] They are unique in the way that their well-defined three-dimensional structure is dominated by short-range, regularized intra- and inter-repeat hydrophobic interactions. For several classes of repeat proteins, analyses of amino acid variability at different positions within a single repeat have revealed that residues that compose the ligand binding site are significantly more variable than the other positions on the protein surface.[77-78] This sequence-function relationship is analogous to the complementarity determining regions (CDR) of antibodies[79] and is consistent with the notion that repeat proteins provide a constant framework that displays ligand-binding residues. This spatial separation of framework and ligand-binding function is important for the design of binding scaffolds so that the ligand-binding function does not compromise the overall structure and stability.

4.3.2 Consensus Sequence Design

Consensus sequence design has emerged as a protein design tool to create *de novo* proteins that capture sequence-structure relationships and interactions present in nature.[76, 80] Proteins created in this way are idealized structural motifs optimized for stability.[81-82] There are two motivations for using consensus design of repeat proteins as opposed to randomizing the surface of one particular family member. First, consensus design can markedly increase stability of engineered proteins. Secondly, full-consensus design in which all repeats are the same allows for addition, deletion, and shuffling of repeats.[46, 83] Additionally, the design of consensus sequences exposes principal features of the protein architecture, which is important for subsequent engineering and chemical coupling.

4.3.3 LRR Domains in NLR Proteins

In NLRs there is a striking correlation (not observed for analogous TLR proteins) between gene organization and the amino acid sequence of their LRR domain.[70, 72] Thus, based on their gene architecture, NLRs are divided into NLRP (α and β) and NOD subgroups.[72] Specifically, in a NLRP subgroup, LRR domains are formed by tandem repeats of exons where each exon encodes one central LRR repeat (β) and two halves of the neighboring LRRs (α). In a NOD subgroup, LRR domains are encoded by a single exon per repeat. This modular organization may have important structural and functional consequences and possibly allows extensive alternative splicing of the entire LRR domain. Furthermore, the remarkable gene structure of LRR domains from NLR proteins raises their potential as a modular scaffold for engineering multivalency and multispecificity.

4.3.4 Multiple Sequence Alignment

We have analyzed the multiple sequence alignment (MSA) of individual repeats in NLR proteins according to the previously published procedure for TPR proteins.[84] Briefly, confirmed human NLR proteins were identified in the HUGO Gene Nomenclature Committee (HGNC) database.[72] We have chosen this database because it provides the most complete information on both the DNA and protein level together with exon-intron gene structure. Of the 22 NLR proteins encoded in the human genome, 19 contain LRR domains and follow the typical exon pattern. Using the *blastp* algorithm from the National Center for Biotechnology Information (NCBI) all homologous mammalian proteins for these 19 proteins were determined. Only confirmed protein products were selected from this search. At this point, the database was curated to exclude alternatively spliced variants of the protein. Through manual searching of selected protein sequences, repeats were extracted and aligned in Microsoft Excel. This procedure resulted in

311 NOD, 255 NLRP-α, and 262 NLRP-β individual LRRs that were then included in the MSA. Using Excel counting functions the greatest percent of occurrence was determined for each position of the repeat. Comparing conserved and variable positions of individual repeats for LRR domains of NLRP and NOD proteins will indicate differences, if any, at the sequence level between the two subgroups.

4.3.5 Consensus Sequence of NOD Subgroup

MSAs for NOD, NLRP-α, and NLRP-β repeats resulted in three consensus sequences. The sequence conservation for each subgroup is shown by sequence logos (Figure 4.1a).[85] The sequence identities for the NOD and NLRP-β consensus sequences are 57%; however, the overall conservation in most positions in the NLRP-β sequence exceeds the NOD. For example, positions 6, 13, 18, 19, 21, 22, 25, 26, and 27 in the NLRP-β MSA all have significantly higher conservation than the NOD sequence (Figure 4.1b). The consensus sequence for the NLRP-α subgroup differs from both NOD and NLRP-β, sharing a sequence identity of 25% and 21% respectively. However, the sequence identity between the tandem α and β consensus sequence and the previously reported consensus protein based on the Ribonuclease Inhibitor (RI) is 61%.[70] This sequence similarity between NLRP and RI LRR repeats leads us to conclude that α and β repeat types do occur in tandem, i.e. repeating unit is 57 amino acids long. Hence we decided to pursue design based on the NOD subgroup that will allow for a shorter (28 amino acid) single repeat, a feature preferred for design of a binding scaffold.

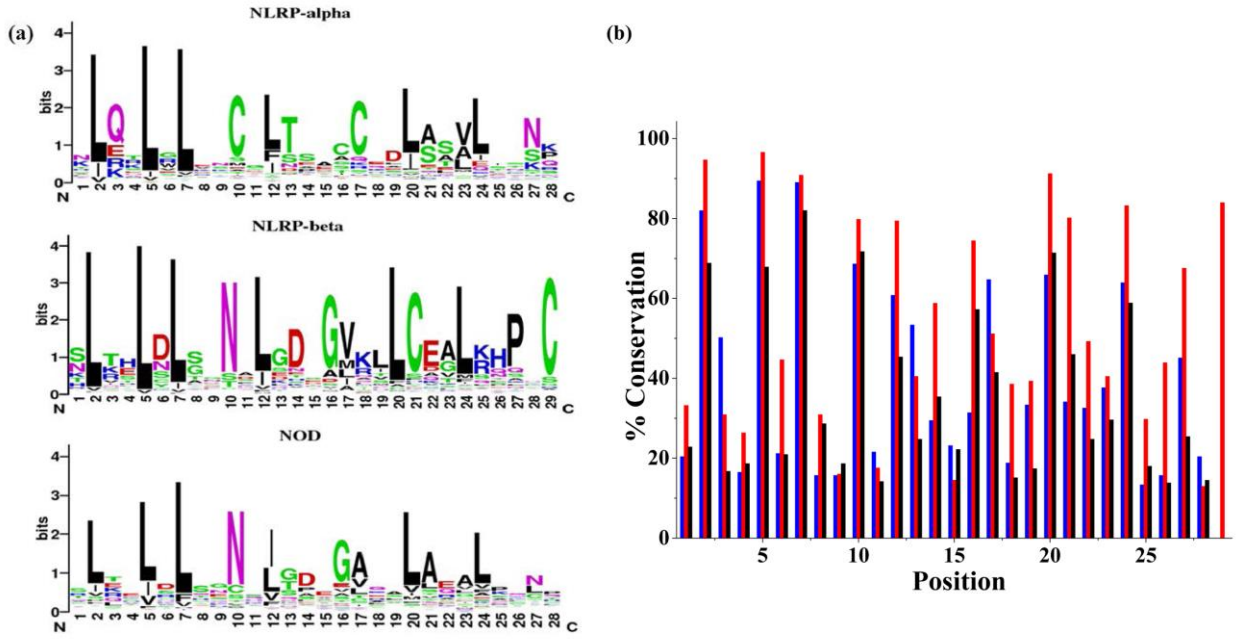


Figure 4.1. Sequence conservation of NLR subgroups (A) Sequence logos for each NLR subgroup. The height of individual letters indicates the frequency of occurrence of amino acid at specific position. (B) Percent conservation for each position in the LRR repeat sequence. NLRP-α (blue), NLRP-β (red), and NOD (black).

The first generation consensus sequence design proteins were based on the most common amino acid in each of the 28 positions. Proteins consisting of 4, 5, 6 and 10 identical repeats were well expressed in *E. coli* and soluble, but lacked secondary structure as indicated by the minima at 195 nm in the circular dichroism (CD) spectra (SI Figure 1). This was not surprising because repeat proteins commonly contain N- and C-terminal repeats that act as capping domains to shield the inner hydrophobic core from solvent and aid in solubility and folding and all previous LRR designs contained capping repeats.[61-62, 86-87] Therefore, we separately constructed MSAs of individual N- and C-terminal repeats. Interestingly, whereas in RI LRRs capping repeats differ from internal repeats in both amino acid composition and length, N- and C-capping repeats for NOD LRRs have exactly the same length as internal repeats, but higher overall conservation (Figure 4.2).

Additionally, in the MSAs of internal and terminal repeats for NOD (Figure 4.2, top row in each table), only half of the positions are defined with 30% or higher frequency of a specific amino acid. Stumpp et al.[70] observed a similar phenomenon for the RI LRR. However, among many positions with less than 40% frequency of one specific amino acid, the position was readily defined by the type, i.e. physical properties of the amino acid in that position. Thus we reanalyzed the MSA of the terminal and internal repeats on the basis of conservation of the amino acid's physical properties by considering the hydrophobicity, polarity, size, and charge for the five most common amino acids in each position of the alignment (Figure 4.2). The bottom row of each table shows the second generation consensus sequence based on the re-analyzed MSA. The top row of each panel in Figure 4.2 is color coded for percent conservation of each amino acid while the bottom row is color-coded for conservation of physical properties from the top five consensus sequences. Results indicate that even with amino acid conservation appearing low in many positions, physical properties remain highly conserved. As a result of these findings, the second-generation sequence took into consideration the preferred physical properties in each position of the MSA. The sequence in the top row of each table in Figure 4.2 was modified in positions where conservation of identity was less than 50%. For example, in position three of the internal repeat (Figure 4.2 a), the frequency of occurrence is less than 30%. However, the Glu, Lys, and Arg residues present are highly favored. Although the frequency of a specific amino acid is low, this position is readily defined with a positively charged residue so Lys was selected for this position. Lys was selected over Glu based on the combined percent of occurrence for Lys and Arg being greater than that of the percent occurrence for Glu. We therefore chose the positively charged Lys over the negatively charged Glu. On the contrary, in position 22 of the internal repeat Lys, Arg, and Glu are also highly favored. However, in this

case the percent occurrence of Glu exceeded that of both Lys and Arg together resulting in the selection of Glu at this position.

Additionally, tryptophan residues are in position 22 of the N-terminal repeat, position 4 of the internal repeat, and position 6 of the C-terminal repeat. In the N-terminal repeat, the highest conserved residue of position 22 is a phenylalanine. This preference for an aromatic residue justified the tryptophan substitution. Position 4 in the internal repeat had very low conservation in physical properties, so we selected tryptophan. In position 6 of the C-terminal repeat tryptophan and tyrosine are preferred, each with a conservation of 24%. It is important to note that both position 4 and 6 are within LRR canonical motif that commonly assumes a beta-strand structure. Since tryptophan has a high propensity for beta-sheet formation, we hypothesized that these substitution will not disrupt the overall fold of the repeat, but will allow us to use fluorescence spectroscopy for further protein analysis.

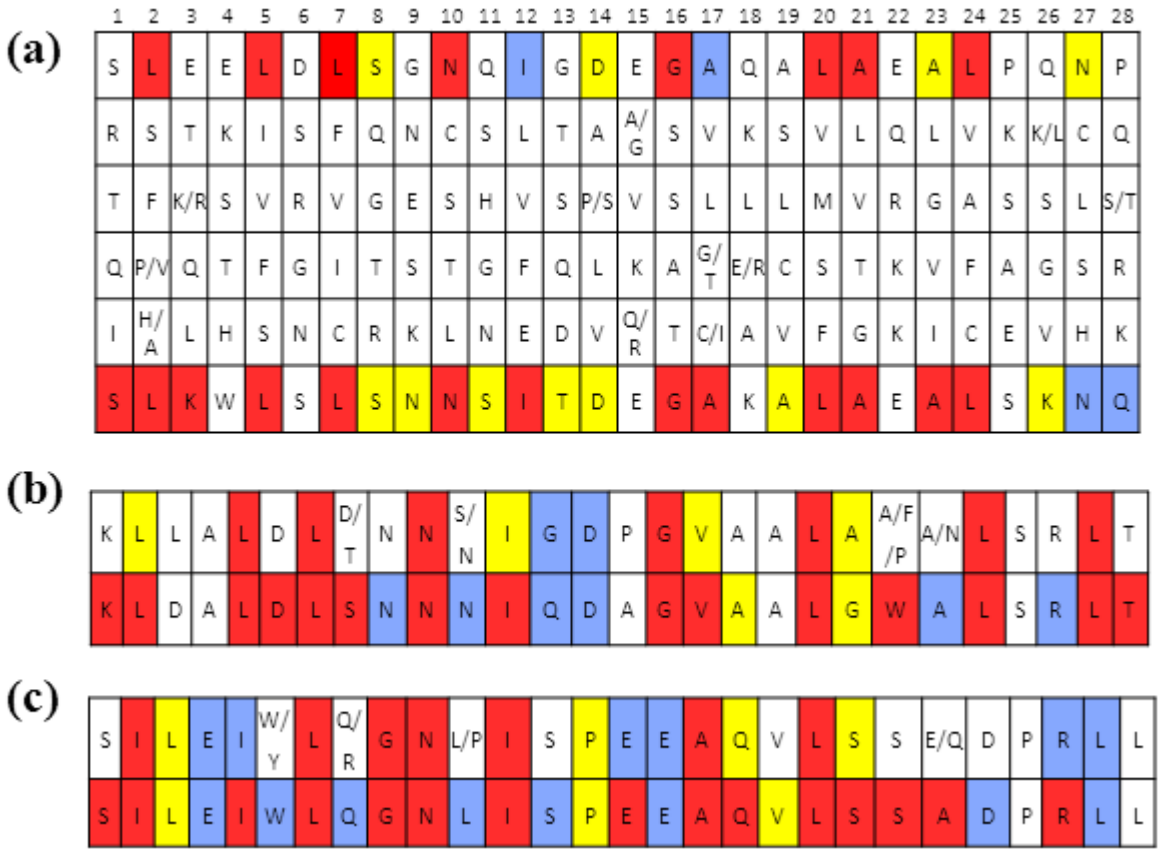


Figure 4.2. The most frequent amino acid residues found in the statistical analysis of NOD LRRs. Top and bottom row of each table are color coded: conservation of 50% or greater is red, 40% blue, and 30% yellow and all position less than 30% are white. (A) Internal repeat; rows 2-5 correspond to the second, third, fourth, and fifth most commonly preferred amino acid for each position; (B) N-terminal repeat (C) C-terminal repeat. The top row in each table shows the frequency of occurrence of amino acid identity. The bottom row is the consensus sequence reanalyzed and corrected for preference of amino acid physical properties. Thus, the bottom row of each table is the amino acid sequence of repeats in CLLR2 protein used in this study. CLRR2 protein contains one N-terminal, two internal, and one C-terminal repeat.

The second-generation consensus sequence protein consisted of three types of consensus sequence repeats: N-terminal repeat, internal LRR repeat, and C-terminal repeat. We will refer to this consensus protein as CLRRx, where C stands for consensus sequence and x is equal to the number of internal repeats in the protein. For example, CLRR2 protein consists of two internal, and N- and C- terminal consensus LRR repeats. The actual sequence of repeats is shown as the bottom row of each panel in Figure 4.2.

4.3.6 Biophysical Characterization

For further biophysical characterization, we chose the smallest of the designed CLRRs, CLLR2, since smaller proteins are preferred for application as binding scaffolds. CLRR2 was expressed at 37 ˚C in BL-21 cells and purified under denaturing conditions following standard procedure.[88] After on-column refolding, CLRR2 elutes from SEC Superdex 75 column in a single symmetrical peak (SI Figure 2). The apparent molecular mass value estimated from molecular mass standards was 17 kD and is in agreement with the expected molecular mass of 15,977 Da as determined by MALDI, indicating that CLRR2 is monomeric in solution.

4.3.6.1 Secondary Structure

CD spectroscopy was used to measure the secondary structure content, showing two minima at 207 nm and 220 nm. The relative intensities of the 207 nm and 220 nm peaks are characteristic of alpha helical and beta sheet secondary structure (Figure 4.3a). The CDPro Software's SELCON program predicts a structure comprised of 24.8% alpha helix and 21.5% beta sheet.[89-90] The alpha helical content is about 10% lower than that estimated for designed RI LRRs.[70] This is not surprising because it has been proposed that while the signature LRR motif contributes to much of the beta structure in LRR proteins, the helical regions may contribute to variability between different LRR families.[91] Additionally, we used MUlti-Sources ThreadER (MUSTER) protein threading algorithm to obtain a predicted structural model of CLRR2 (Figure 4.4b).[92] The homology modeling was based on protein NLRX1 that shares 24.1 % sequence identity with CLLR2. NLRX1, also known as NOD9, is a mitochondrial protein thought to be involved in an antiviral immune response against viral Ribonucleic Acid (RNA).[93] Hong et al. showed that cNLRX directly binds to single and double stranded RNA.[93] The alternating alpha-helical and beta-sheet segments of each repeat in the homology model are consistent with the CD

data and CDPro structure prediction confirming that the consensus design retains a structure typical of natural leucine rich repeats.

4.3.6.2 Thermal Denaturation

The thermal denaturation of the CLRR2 was followed by both tryptophan fluorescence and CD spectroscopy (Figure 4.3b-c). Here, it should be taken into consideration that Trp residues are located in both capping repeats and internal repeats and that fluorescence reports on the chemical environment of the indole fluorophore, whereas the CD signal is indicative of the overall amount of the secondary structure. Thermal denaturation of CLRR2 was accompanied by a reduction of fluorescent signal intensity and a shift of the emission maxima to higher wavelengths. The plot of normalized signal change as a function of temperature results in a curve indicative of two transitions (Figure 4.3b). To estimate the transition temperatures, we treated each of the transitions as a single two state transition. Although an approximation, this fitting allowed us to estimate transition temperatures of 32˚C and 57˚C. Similarly, thermal unfolding monitored by the change in CD signal at 217 nm shows two transitions at 30 ˚C and 68 ˚C (Figure 4.3c). After incubating CLLR2 for 10 minutes at 50˚C and then repeating the thermal unfolding experiment, qualitatively different melting curves are observed for fluorescence and CD. The overall shape of the melting curve observed in the fluorescence measurement does not change after incubation, but the melting curve observed in the CD experiment now shows a single sharp transition with a new transition temperature of 60 ˚C (Figure 4.3c).

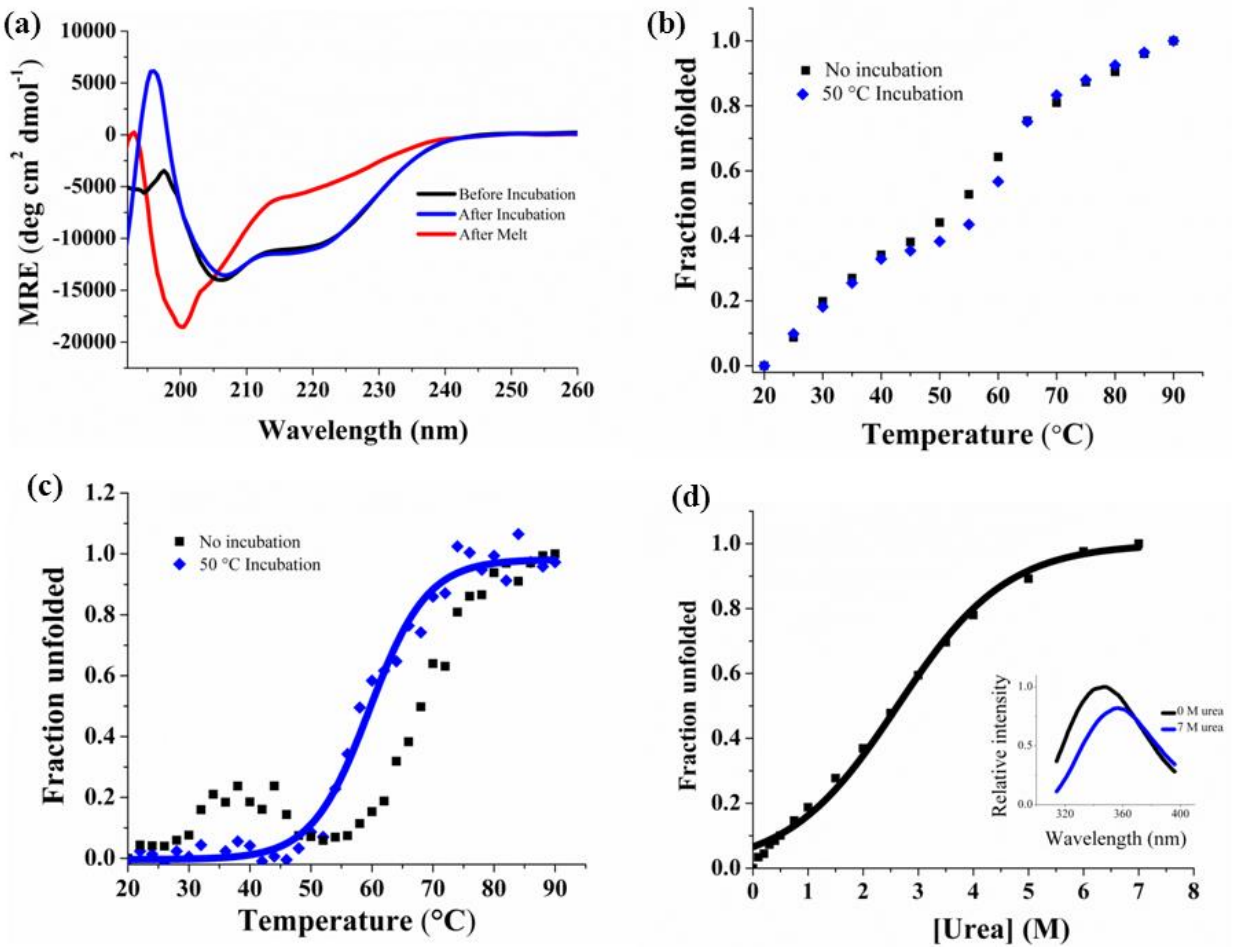


Figure 4.3. Biophysical characterization of CLRR2, before (black) and after (blue) incubation at 50˚C and irreversible unfolding after heating (red). (A) CD spectrum of CLRR2 normalized to units of mean residue ellipticity indicates a mixed secondary structure of α-helix and β-sheet and irreversible unfolding after denaturation. (B) Thermal denaturation following changes in Trp fluorescence. (C) Thermal denaturation following CD signal at 217 nm. (D) Chemical denaturation of CLRR2 monitoring changes in Trp fluorescence with increasing urea concentration. Inset shows fluorescence spectra at 0 M urea (black) and 7 M urea (blue).

Intriguingly, the temperature of the first transition in the CD, which disappears after incubation at 50˚C, coincides to the first transition observed in fluorescence experiments. Similarly, the temperature of the second transition in the CD experiment is 10 degrees higher than the T_m for the second transition in fluorescence experiment before incubation, but only 3 degrees higher after incubation. The overall CD signal and fluorescence intensity are the same before and after incubation of CLLR2 indicating no change in the overall amount of secondary

structure (Figure 4.3a). Spectra taken before and after the thermal melt with or without incubation suggest irreversible unfolding (Figure 4.3a).

Although the thermal stability observed in fluorescence and CD experiments is qualitatively similar, the curves from these two experiments are not superimposable. This leads us to conclude that unfolding does not follow a simple two-state mechanism and that stable intermediates are formed during thermal unfolding. Additionally, the difference in fluorescence and CD curves indicates that the capping repeats and internal repeats may not unfold in a concerted manner. Due to the presence of Trp in each repeat, fluorescence unfolding curves provide a more complete picture of the unfolding process.

4.3.6.3 Chemical Denaturation

The equilibrium denaturation of the designed CLRR2 was followed by observing the change in Trp fluorescence with increasing concentration of urea. The denaturing curve shows a broad transition around 3 M urea (Figure 4.3d). Fitting this curve to a two state model results in a midpoint of denaturation of 2.6 M urea, but cooperativity of the protein is only 2.5 kJ/mol M. This value is less than expected for highly cooperative unfolding proteins of this size, confirming the conclusions from the thermal denaturation. CLRR2 displays similar cooperativity to the designed RI LRR but is much less cooperative than previously studied YopM LRR (22.6 kJ/mol M) and Internalin B LRR (10.25 kJ/mol M).[70, 94-95]

4.3.6.4 Binding Affinity of CLLR2

Proteins of the innate immune system differ from antibodies of the adaptive immune system in that they detect pathogen associated molecular patterns (PAMPS) instead of one specific antigen.[96] In other words, NLRs recognize the global features of a family of pathogens. This characteristic of the NLR family led us to hypothesize that the consensus design method

may result in a protein that retains similar recognition properties. To test this hypothesis, we measured the affinity of CLLR2 for a muramyl dipeptide (MDP). MDP is a known ligand for the NOD2 protein family of NLRs. Recently, Grimes et al. showed that human NOD2 binds directly to MDP.[97] They analyzed the binding of full-length human NOD2 to MDP using Surface Plasmon Resonance and determined an affinity of 51 ± 18 nM.[97] However, there is no clear evidence that the LRR domain mediates this interaction. Here we used fluorescence quenching and fluorescence anisotropy to investigate the affinity of designed CLRR2 for the biologically relevant L-D isomer of MDP.[98]

We investigated the ligand binding properties of CLLR2 by observing quenching of the fluorescence signal at 340 nm in the presence of MDP. The binding curve in Figure 4a is obtained by plotting the fraction of the bound ligand as a function of MDP concentration. Fitting of this curve to the single-site binding isotherm resulted in a K_d of 2.0 ± 0.4 μM for an average of two trials. Closer inspection of the homology model of CLRR2 structure shows that the Trp residue from the two internal repeats and C-terminal repeat form a cluster on the concave face of the protein (Figure 4.4b). Similar carbohydrate recognition through tryptophan residues was seen by Luo et al. in recognition of Thomsen-Friedenreich antigen by a VLR protein.[60] Although our observed affinities are an order of magnitude lower than reported for the full-length NOD2 protein[97] it is important to note that CLRR2 is representative of the entire NOD family and shares only 70 % sequence identity with NOD2.

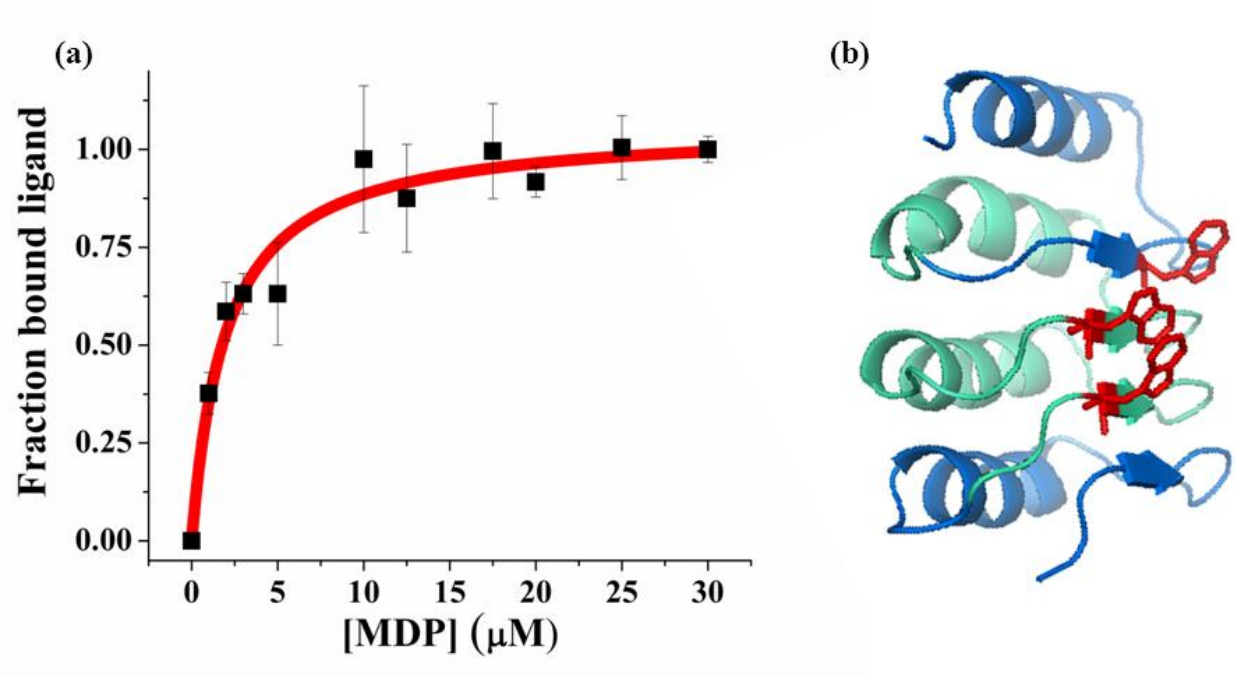


Figure 4.4. (A) Fluorescence quenching of Trp residues in CLRR2 is shown as a function of MDP concentration. Error bars are the standard deviation from three measurements. The K_d was determined to be 2.0 ± 0.4 μM. (B) The MUSTER model of CLRR2 showing location of Trp residues (red) on the concave face of the protein that are the proposed binding site of MDP. The terminal repeats are shown in blue and the internal repeats are shown in cyan.

To confirm that the observed interaction is not an artifact of collisional tryptophan quenching, we performed a fluorescence anisotropy experiment, where we now observed the change in the signal originating on FITC labeled MDP. Fluorescence anisotropy change was measured as a function of increasing concentration of CLRR2 titrated into a solution of FITC-MDP. Fitting of the binding curve to a single-site binding isotherm resulted in a K_d = 20 ± 10 μM (SI Figure 3). This difference in the observed K_d values is not unexpected when comparing two techniques. Additionally, fluorescein is attached to MDP through a flexible linker on the N-acetylmuramic acid and the overall change in the anisotropy signal is low thus affecting the overall signal-to-noise ratio for the anisotropy experiment.

As a control experiment, we investigated the binding of CLRR2 to the D-D isomer of MDP, which is unable to stimulate NOD2 in vivo.[98] While results indicate that tryptophans of CLLR2 are quenched in the presence of D-D MDP, the data is representative of random collisional quenching and does not follow the expected trend of a single-site binding isotherm (SI Figure 4). This control experiment leads us to conclude that under the concentration regime tested, CLLR2 is a specific binder for physiologically relevant L-D isomer of MDP2.

Thus, through the consensus sequence design, and without the evolution and selection step, we have developed a scaffold with micromolar affinity for a glycopeptide. This result indicates that consensus sequence LRRs based on the NOD protein family are a good starting point for the design of glycopeptide binding scaffolds. Moreover, this is the first report, to our knowledge, of direct interaction of NOD LRR with a physiologically relevant ligand.

4.4 Conclusions

Through consensus-based design, we have developed a novel leucine rich repeat protein CLRR2. Structural and physical analysis of this protein indicate that it preserves important features of natural LRRs as well as the desired biophysical properties needed for use as a binding scaffold. We have shown that the consensus-based design resulted in a construct that retains binding information of natural NOD proteins.

The mechanism of pathogen sensing by NLRs is still largely unknown. By creating a LRR protein based on structurally and functionally homologous NODs we have developed a protein that can serve as a model structure for this class of NLRs. In the future, analysis of variable residues in the sequence alignment will allow for engineering diverse binding affinities for the development of new binding scaffolds as well as further elucidation of the role of LRR in NLR pathogen recognition.

4.5 Material and Methods

4.5.1 Consensus Design and Multiple Sequence Alignment

The 22 genes known to encode for NLR proteins in humans were retrieved from the HUGO Gene Nomenclature Committee (HGNC) database. Each confirmed gene sequence was translated into its corresponding protein sequence and input into the National Center for Biotechnology Information (NCBI) basic local alignment search tool (BLAST). Alignment of repeats that followed the canonical LRR motif of $XLXXLXLXXNXL(X)_nL$ was accomplished using Microsoft Excel. Repeats were extracted by manually searching the LRR domain of the selected NLR sequences. We aligned repeats following a pattern of $XLXXLXLXXNXL(X)_nL(X)_{8.}$ In the canonical motif L is defined as Leu, Ile, Val, or Phe, and X is any other naturally occurring amino acid.[72] N can also be defined as Arg, Cys, Ser, or Thr, and n is equal to 7. Repeats were aligned using Microsoft Excel and the consensus sequence was obtained by determining the amino acid with the highest frequency of occurrence in each position of the repeat by using the Microsoft Excel counting function. Consensus amino acid sequences were found for the top 5 most preferred amino acid in each position.

4.5.2 Cloning

For cloning of the LRR protein, a gene was designed consisting of an N-terminal repeat, internal repeat, and C-terminal repeat. Synthetic genes were synthesized by GENEWIZ Inc. and cloned into plasmid pProExHtam by ligation of restrictions sites BamHI and HindIII. Gene identity was confirmed by sequencing at the Virginia Tech Bioinformatics Institute.

4.5.3 Protein Expression and Purification

Overnight cultures of BL21 (DE3) cells were diluted 1:100 in 1 L of Terrific Broth media at 37°C, with shaking at 250 rpm, and were grown to an OD_{600} of 0.5–0.8. Expression was induced with 1 mM IPTG, followed by 4 hours of expression at 37˚C. The cells were harvested

by centrifugation at 5,000 rpm for 15 min and the pellets were frozen at −20°C until purification. To purify proteins, the cell pellet was resuspended in lysis buffer consisting of 50 mM Tris, 300 mM sodium chloride, 0.1% Tween 20, and 8 M urea. After one minute sonication at 30% power using a microtip and Mison sonicator, lysed cells were centrifuged at 16,000 rpm for 30 min and the protein supernatant was collected. Proteins were purified under denaturing conditions using standard Ni-NTA affinity purification protocol and eluted with 300 mM imidazole in lysis buffer with 8 M urea. Proteins were then refolded on the size-exclusion column in 50 mM sodium phosphate buffer pH 8 with 150 mM sodium chloride. Protein identity was confirmed with MALDI indicating a molecular weight of 15,977 Da. Proteins were quantified by absorption at 280 nm using an extinction coefficient of 27, 960 M^{-1} cm^{-1}, calculated from the amino acid sequence using the Expasy Protparam tool.[99]

4.5.4 Size Exclusion Chromatography

Akta Prime Plus FPLC was used for size exclusion chromatography. Refolding of denatured proteins after affinity purification was completed on the Superdex 75 16/600 Prep Grade column in 150 mM sodium chloride and 50 mM sodium phosphate buffer pH 8 at a flow rate of 0.5 milliliters per minute. The Superdex 75 10/300 analytical column was used for analysis of molecular weights under the same conditions. A comparison to known standards (Bio-Rad) allowed for determination of the molecular weights and oligomeric states of each LRR protein.

4.5.5 Circular Dichroism

Circular dichroism (CD) spectra were acquired using 5 to 10 μM protein samples in 10 mM phosphate buffer pH 7.4 with 10 mM NaCl using a Jasco J-815 CD spectrometer. Far-UV CD (190–260 nm) spectra were recorded at 25°C to assess the secondary structure of CLRR2.

Each sample was recorded three times, from 190 to 260 nm in a 2 mm pathlength cuvette, and averaged. Data collected using a 1 nm bandwidth, 2 nm data pitch, and a data integration time of 1 second, was normalized to units of mean residue ellipticity for all samples. Thermal denaturation curves were recorded by monitoring molar ellipticity at 217 nm while heating from 20 to 90°C in 2°C increments with an equilibration time of 10 min at each temperature.

4.5.6 Urea Denaturation

Tryptophan fluorescence was monitored using a Cary Eclipse Fluorometer. Excitation of samples occurred at 295 nm and spectra were recorded from 310 to 380 nm with the excitation and emission slits equal to 5 nm. A 10 M urea in 10 µM protein stock solution was titrated into a 10 µM protein sample. After a ten min equilibration period for each addition, 3 scans were collected and averaged.

4.5.7 Fluorescence Anisotropy

To determine the binding affinity, increasing amounts of protein CLLR2, were titrated to a FITC-labeled MDP (Purchased from Invivogen) in 10 mM Na_2HPO_4 pH 7.4 and 10 mM NaCl, buffer. Binding was performed at 10 nM peptide concentration in a 10 mm path-length cuvette at 25°C, and the fluorescence anisotropy was recorded after a 30 min equilibration period. Fluorescence anisotropy experiments were recorded in a Cary Eclipse Fluorometer equipped with excitation and emission polarizers. Excitation was achieved with a 5 nm slit-width at 495 nm and the emission recorded at 515 nm with a slit-width of 5 nm. For excitation at the vertical orientation (0°) the anisotropy (r) is:

$$r = \frac{(I_{VV} - I_{B,VV}) - G(I_{VH} - I_{B,VH})}{(I_{VV} - I_{B,VV}) + 2G(I_{VH} - I_{VH} - I_{B,VH})} \quad (1)$$

where G is the G-factor, I_{VV} and I_{VH} are the vertical and horizontal emission of the sample, respectively, and $I_{B,VV}$ and $I_{B,VH}$ are the intensity of the emission of the blank with emission polarizer at vertical and horizontal orientation, respectively. The G-factor corrections were calculated using the equation: $G = (I_{HV} - I_{B,HV})/(I_{HH} - I_{B,HH})$, where I_{HV} is the vertical emission (0°) of a standard solution with excitation in horizontal orientation (90°), I_{HH} is the horizontal emission (90°) of a standard solution with excitation in vertical orientation (0°), $I_{B,HV}$ is the vertical emission (0°) of a blank solution with excitation in horizontal orientation (90°) and $I_{B,HH}$ is the horizontal emission (90°) of a blank solution with excitation in vertical orientation (0°) using phosphate buffer as a blank solution and a 10 nM FITC labeled MDP as a standard solution. The fraction of ligand bound at each point in the binding curve was calculated by the equation,

$$\textbf{Fraction bound} = \frac{r - r_f}{r_b - r_f} \qquad (2)$$

where r is the observed anisotropy of the peptide at any protein concentration, r_f is the anisotropy of the free peptide and r_b is the anisotropy of the ligand in the plateau region of the binding curve. The data was fit using nonlinear regression analysis with Origin software based on the equation,

$$r = \frac{r_b[P]}{K_d + [P]} \quad (3)$$

where r is equal to the fraction of bound peptide, r_b is equal to the maximum signal, K_d is the dissociation constant, and [P] is the concentration of protein in the sample.

4.5.8 Fluorescence Quenching

Fluorescence quenching experiments were completed using a Cary Eclipse Fluorometer. Samples of 10 µM protein with MDP (purchased from Invivogen) from 0 to 30 µM were incubated in a 96 well plate for 30 min. Spectra were recorded with an excitation wavelength of 295 nm and excitation and emission slits equal to 10 and 20 nm. Each sample was measured from 310-380 nm. Three spectra were recorded and averaged for each trial. The binding curve was obtained by plotting fluorescence signal against MDP concentration where the percent of bound ligand was calculated using the equation,

$$Fraction\ bound = \frac{F_o - F}{F_o - F_{min}} \quad (4)$$

where F_o is the fluorescent signal without MDP, F is the signal at any ligand concentration, and F_{min} is the fluorescent signal at saturation. The data was fit using nonlinear regression analysis with Origin Software using the equation,

$$r = \frac{r_b[P]}{K_d + [P]} \quad (5)$$

where r is equal to the fraction of bound ligand, r_b is equal to the maximum signal, K_d is the dissociation constant, and [P] is the concentration of peptide in the sample.

4.6 Supporting Information

Sequence and CD spectra of first generation LRRs; Size exclusion chromatography of CLRR2; Fluorescence anisotropy of CLRR2; MDP control- fluorescence quenching.

4.7 Acknowledgments

The authors would like to thank Virginia Tech and the VT Chemistry Department for funding. The authors are also grateful to Dr. Rich Helm and Keith Ray of the VT Biochemistry Department for MALDI analysis. The authors are also thankful to Prof. Webster Santos, Prof. Pablo Sobrado, Dr. Aitziber Cortajarena, and members of the Grove lab for discussion and careful reading of the manuscript.

4.8 References

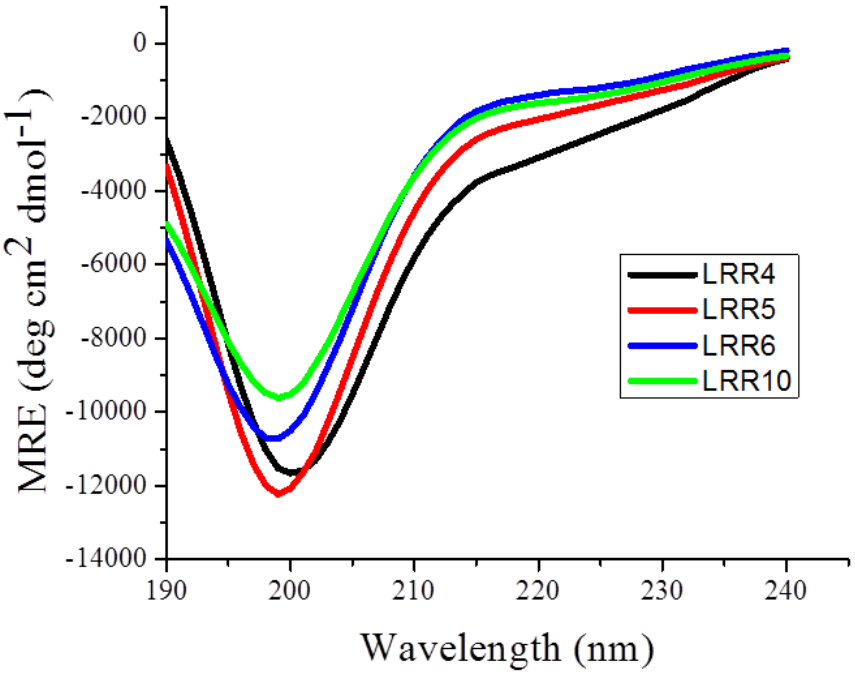
1. Boersma, Y. L.; Pluckthun, A., DARPins and other repeat protein scaffolds: advances in engineering and applications. *Curr. Opin. Biotechnol.* **2011,** *22* (6), 849-57.
2. Lofblom, J.; Frejd, F. Y.; Stahl, S., Non-immunoglobulin based protein scaffolds. *Curr. Opin. Biotechnol.* **2011,** *22* (6), 843-8.
3. Yoshimura, H.; Inaguma, A.; Yamada, T.; Ozawa, T., Fluorescent probes for imaging endogenous beta-actin mRNA in living cells using fluorescent protein-tagged pumilio. *ACS Chem. Biol.* **2012,** *7* (6), 999-1005.
4. Filipovska, A.; Razif, M. F.; Nygard, K. K.; Rackham, O., A universal code for RNA recognition by PUF proteins. *Nat Chem Biol* **2011,** *7* (7), 425-7.
5. Hong, X.; Ma, M. Z.; Gildersleeve, J. C.; Chowdhury, S.; Barchi, J. J., Jr.; Mariuzza, R. A.; Murphy, M. B.; Mao, L.; Pancer, Z., Sugar-binding proteins from fish: selection of high

affinity "lambodies" that recognize biomedically relevant glycans. *ACS Chem. Biol.* **2013,** *8* (1), 152-60.
6. Garg, A.; Lohmueller, J. J.; Silver, P. A.; Armel, T. Z., Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res* **2012,** *40* (15), 7584-95.
7. Fujimoto, Y. K.; Green, D. F., Carbohydrate Recognition by the Antiviral Lectin Cyanovirin-N. *J. Am. Chem. Soc.* **2012,** *134* (48), 19639-19651.
8. Woodrum, B. W.; Maxwell, J. D.; Bolia, A.; Ozkan, S. B.; Ghirlanda, G., The antiviral lectin cyanovirin-N: probing multivalency and glycan recognition through experimental and computational approaches. *Biochem. Soc. Trans.* **2013,** *41*, 1170-1176.
9. Zheng, H.; Wang, F.; Wang, Q.; Gao, J. M., Cofactor-Free Detection of Phosphatidylserine with Cyclic Peptides Mimicking Lactadherin. *J. Am. Chem. Soc.* **2011,** *133* (39), 15280-15283.
10. Mak, A. N. S.; Bradley, P.; Cernadas, R. A.; Bogdanove, A. J.; Stoddard, B. L., The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. *Science* **2012,** *335* (6069), 716-719.
11. Mak, A. N.-S.; Bradley, P.; Bogdanove, A. J.; Stoddard, B. L., TAL effectors: function, structure, engineering and applications. *Curr. Opin. Struct. Biol.* **2013,** *23* (1), 93-99.
12. Gebauer, M.; Skerra, A., Engineered protein scaffolds as next-generation antibody therapeutics. *Curr Opin Chem Biol* **2009,** *13* (3), 245-55.
13. Cummings, R. D., The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **2009,** *5* (10), 1087-1104.
14. Binz, H. K.; Amstutz, P.; Kohl, A.; Stumpp, M. T.; Briand, C.; Forrer, P.; Grutter, M. G.; Pluckthun, A., High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat Biotechnol* **2004,** *22* (5), 575-82.
15. Cortajarena, A. L.; Kajander, T.; Pan, W.; Cocco, M. J.; Regan, L., Protein design to understand peptide ligand recognition by tetratricopeptide repeat proteins. *Protein Eng Des Sel* **2004,** *17* (4), 399-409.
16. Varadamsetty, G.; Tremmel, D.; Hansen, S.; Parmeggiani, F.; Pluckthun, A., Designed Armadillo repeat proteins: library generation, characterization and selection of peptide binders with high specificity. *J. Mol. Biol.* **2012,** *424* (1-2), 68-87.
17. Seeger, M. A.; Zbinden, R.; Flutsch, A.; Gutte, P. G. M.; Engeler, S.; Roschitzki-Voser, H.; Grutter, M. G., Design, construction, and characterization of a second-generation DARPin library with reduced hydrophobicity. *Protein Sci.* **2013,** *22* (9), 1239-1257.
18. Luo, M.; Velikovsky, C. A.; Yang, X.; Siddiqui, M. A.; Hong, X.; Barchi, J. J., Jr.; Gildersleeve, J. C.; Pancer, Z.; Mariuzza, R. A., Recognition of the Thomsen-Friedenreich pancarcinoma carbohydrate antigen by a lamprey variable lymphocyte receptor. *J Biol Chem* **2013,** *288* (32), 23597-606.
19. Wezner-Ptasinska, M.; Krowarsch, D.; Otlewski, J., Design and characteristics of a stable protein scaffold for specific binding based on variable lymphocyte receptor sequences. *Biochim Biophys Acta* **2011,** *1814* (9), 1140-5.
20. Lee, S. C.; Park, K.; Han, J.; Lee, J. J.; Kim, H. J.; Hong, S.; Heu, W.; Kim, Y. J.; Ha, J. S.; Lee, S. G.; Cheong, H. K.; Jeon, Y. H.; Kim, D.; Kim, H. S., Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. *Proc Natl Acad Sci U S A* **2012,** *109* (9), 3299-304.
21. Cortajarena, A. L.; Liu, T. Y.; Hochstrasser, M.; Regan, L., Designed proteins to modulate cellular networks. *ACS Chem Biol* **2010,** *5* (6), 545-52.

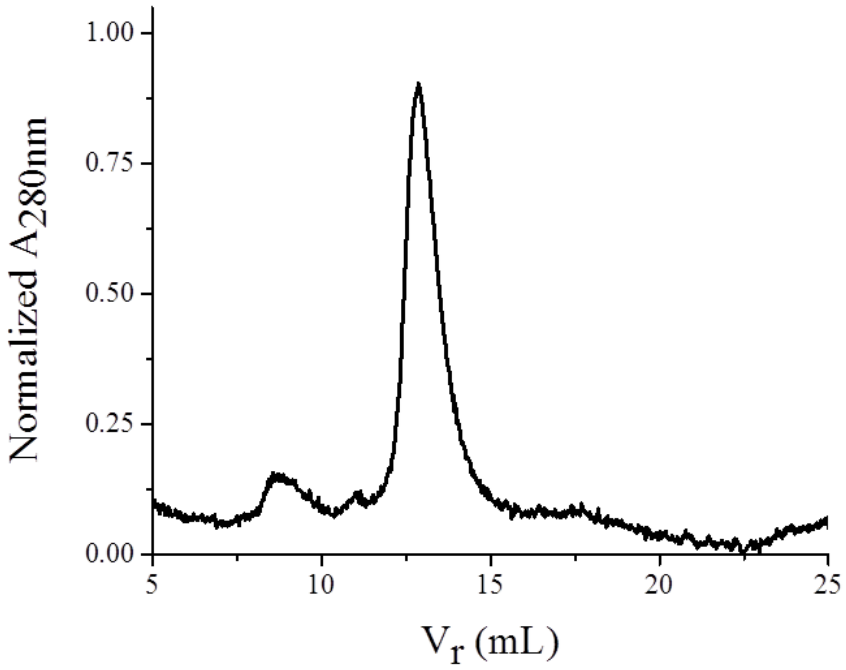
22. Grove, T. Z.; Cortajarena, A. L.; Regan, L., Ligand binding by repeat proteins: natural and designed. *Curr Opin Struct Biol* **2008,** *18* (4), 507-15.
23. Grove, T. Z.; Forster, J.; Pimienta, G.; Dufresne, E.; Regan, L., A modular approach to the design of protein-based smart gels. *Biopolymers* **2012**.
24. Grove, T. Z.; Hands, M.; Regan, L., Creating novel proteins by combining design and selection. *Protein engineering, design & selection : PEDS* **2010,** *23* (6), 449-55.
25. Grove, T. Z.; Osuji, C. O.; Forster, J. D.; Dufresne, E. R.; Regan, L., Stimuli-responsive smart gels realized via modular protein design. *J Am Chem Soc* **2010,** *132* (40), 14024-6.
26. Jackrel, M. E.; Valverde, R.; Regan, L., Redesign of a protein-peptide interaction: characterization and applications. *Protein science* **2009,** *18* (4), 762-74.
27. Mosavi, L. K.; Cammett, T. J.; Desrosiers, D. C.; Peng, Z. Y., The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **2004,** *13* (6), 1435-48.
28. Stumpp, M. T.; Forrer, P.; Binz, H. K.; Pluckthun, A., Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. *J. Mol. Biol.* **2003,** *332* (2), 471-87.
29. Kumar, H.; Kawai, T.; Akira, S., Pathogen recognition in the innate immune response. *Biochem. J.* **2009,** *420*, 1-16.
30. Istomin, A. Y.; Godzik, A., Understanding diversity of human innate immunity receptors: analysis of surface features of leucine-rich repeat domains in NLRs and TLRs. *BMC Immunol* **2009,** *10*, 48.
31. Wang, Y.; Liu, L.; Davies, D. R.; Segal, D. M., Dimerization of Toll-like receptor 3 (TLR3) is required for ligand binding. *Journal of biological chemistry* **2010,** *285* (47), 36836-41.
32. Mariathasan, S.; Monack, D. M., Inflammasome adaptors and sensors: intracellular regulators of infection and inflammation. *Nat Rev Immunol* **2007,** *7* (1), 31-40.
33. Kobe, B.; Deisenhofer, J., The leucine-rich repeat: a versatile binding motif. *Trends Biochem Sci* **1994,** *19* (10), 415-21.
34. Kajander, T.; Cortajarena, A. L.; Regan, L., Consensus design as a tool for engineering repeat proteins. *Methods. Mol. Biol.* **2006,** *340*, 151-70.
35. Magliery, T. J.; Regan, L., Sequence variation in ligand binding sites in proteins. *BMC Bioinformatics* **2005,** *6*, 240.
36. Magliery, T. J.; Regan, L., Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif. *J Mol Biol* **2004,** *343* (3), 731-45.
37. Johnson, G.; Wu, T. T., Kabat Database and its applications: future directions. *Nucleic Acids Res* **2001,** *29* (1), 205-206.
38. Forrer, P.; Binz, H. K.; Stumpp, M. T.; Pluckthun, A., Consensus design of repeat proteins. *Chembiochem* **2004,** *5* (2), 183-9.
39. Magliery, T. J.; Lavinder, J. J.; Sullivan, B. J., Protein stability by number: high-throughput and statistical approaches to one of protein science's most difficult problems. *Curr Opin Chem Biol* **2011,** *15* (3), 443-51.
40. Sullivan, B. J.; Durani, V.; Magliery, T. J., Triosephosphate isomerase by consensus design: dramatic differences in physical properties and activity of related variants. *Journal of molecular biology* **2011,** *413* (1), 195-208.
41. Jackrei, M. E.; Valverde, R.; Regan, L., Redesign of a protein-peptide interaction: Characterization and applications. *Protein Sci.* **2009,** *18* (4), 762-774.
42. Kajander, T.; Cortajarena, A. L.; Regan, L., Consensus design as a tool for engineering repeat proteins. *Methods Mol Biol* **2006,** *340*, 151-70.

43. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E., WebLogo: A sequence logo generator. *Genome Res.* **2004,** *14* (6), 1188-1190.
44. Kloss, E.; Courtemanche, N.; Barrick, D., Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. *Arch Biochem Biophys* **2008,** *469* (1), 83-99.
45. Pancer, Z.; Amemiya, C. T.; Ehrhardt, G. R.; Ceitlin, J.; Gartland, G. L.; Cooper, M. D., Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* **2004,** *430* (6996), 174-80.
46. Wingfield, P. T.; Palmer, I.; Liang, S.-M., Folding and Purification of Insoluble (Inclusion Body) Proteins from Escherichia coli. In *Current Protocols in Protein Science*, John Wiley & Sons, Inc.: 2001.
47. Sreerama, N.; Venyaminov, S. Y.; Woody, R. W., Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci.* **1999,** *8* (2), 370-80.
48. Sreerama, N.; Woody, R. W., Computation and analysis of protein circular dichroism spectra. *Methods Enzymol* **2004,** *383*, 318-51.
49. Kobe, B.; Kajava, A. V., The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* **2001,** *11* (6), 725-32.
50. Wu, S.; Zhang, Y., MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins: Struct., Funct., Bioinf.* **2008,** *72* (2), 547-56.
51. Hong, M.; Yoon, S. I.; Wilson, I. A., Structure and functional characterization of the RNA-binding element of the NLRX1 innate immune modulator. *Immunity* **2012,** *36* (3), 337-47.
52. Courtemanche, N.; Barrick, D., Folding thermodynamics and kinetics of the leucine-rich repeat domain of the virulence factor Internalin B. *Protein Sci.* **2008,** *17* (1), 43-53.
53. Kloss, E.; Barrick, D., Thermodynamics, kinetics, and salt dependence of folding of YopM, a large leucine-rich repeat protein. *J. Mol. Biol.* **2008,** *383* (5), 1195-209.
54. Akira, S.; Uematsu, S.; Takeuchi, O., Pathogen recognition and innate immunity. *Cell* **2006,** *124* (4), 783-801.
55. Grimes, C. L.; Ariyananda Lde, Z.; Melnyk, J. E.; O'Shea, E. K., The innate immune protein Nod2 binds directly to MDP, a bacterial cell wall fragment. *J Am Chem Soc* **2012,** *134* (33), 13535-7.
56. Inohara, N.; Ogura, Y.; Fontalba, A.; Gutierrez, O.; Pons, F.; Crespo, J.; Fukase, K.; Inamura, S.; Kusumoto, S.; Hashimoto, M.; Foster, S. J.; Moran, A. P.; Fernandez-Luna, J. L.; Nunez, G., Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J Biol Chem* **2003,** *278* (8), 5509-12.
57. Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A, *The Proteomics Protocols Handbook*. 2005.

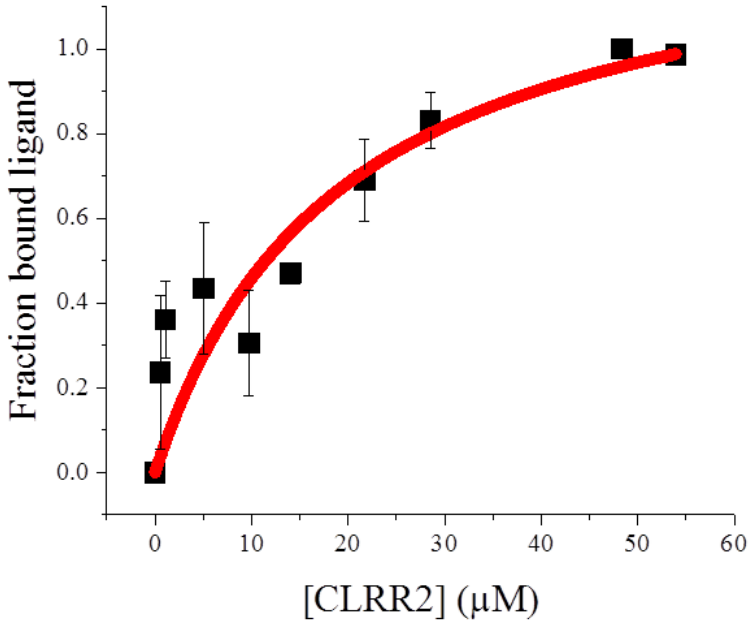
4.9 Supplemental Figures



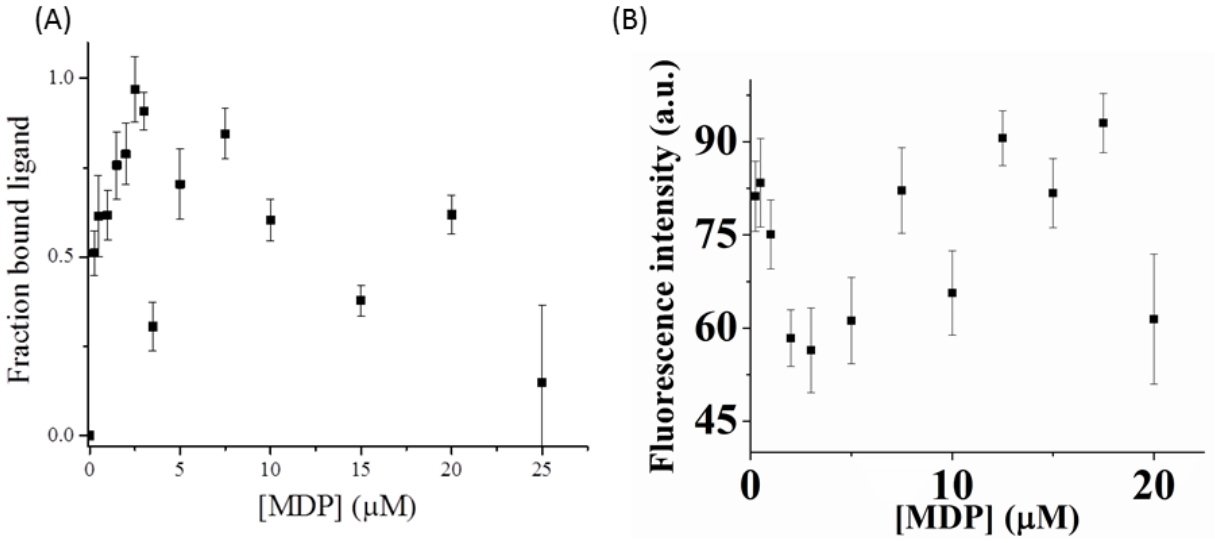
SI Figure 1. CD spectra of first generation LRRs. Spectrum of proteins containing 4 (black), 5 (red), 6 (blue), and 10 (green) repeats are shown. Each protein consisted of repeats determined by a single consensus sequence with amino acid composition: S L T E L D L S G N Q I G D E G A K A L A E A L P Q N P.



SI Figure 2. Size exclusion chromatography of CLRR2. Analysis of oligomeric state of CLRR2 by SEC shows a monomeric protein. Proteins of 44 kD (Ovalbumin), 17 kD (Myoglobin), and 1.35 kD (Vitamin B12) were used as molecular weight standards. The void volume of the column is approximately 8 mLs. CLLR2 (15,977 kD) elutes at 12.8 mLs.



SI Figure 3. Fluorescence anisotropy of CLRR2 and FITC labeled muramyl dipeptide. Graph shows the fraction of bound peptide at varying protein concentrations. The data was fit using nonlinear regression analysis to obtain a K_d of 20 ± 10 μM. Error bars are standard deviation of three measurements.



SI Figure 4. Fluorescence quenching of CLRR2 with muramyl dipeptide synthetic diastereomer: MDP-(L). Graph shows the fraction of bound peptide at varying protein concentrations. Error bars are standard deviation of three measurements. (A) Normalized graph (B) Un-normalized graph

Chapter 5. Investigations of homo- and heteropolymer formation of recombinant human hair keratins

*Rachael N. Parker[1], Kristina L. Roth[1], Mark E. Van Dyke[2], Tijana Z. Grove[1]**

1. Department of Chemistry, Virginia Tech, Blacksburg, VA 24060

2. Department of Biomedical Engineering and Mechanics, Virginia Tech, Blacksburg, VA 24060

(S*ubmitted for publication)*

*Corresponding author email: tijana.grove@vt.edu

Keywords: human hair keratin, recombinant protein, fibers, self-assembly, intermediate filaments, SEC, DLS, TEM

Abbreviations: K31- keratin 31, K81- keratin 81, IF- intermediate filament, SDS-PAGE- sodium dodecyl sulfate polyacrylamide gel electrophoresis, SEC- size exclusion chromatography, DLS- dynamic light scattering, TEM- transmission electron microscopy, DTT- dithiolthreitol, βME- β-Mercaptoethanol

5.1 Abstract

In the past two decades, keratin biomaterials have shown impressive results as scaffolds for tissue engineering, wound healing, and nerve regeneration. In addition to its intrinsic biocompatibility, keratin interacts with specific cell receptors eliciting beneficial biochemical cues. However, during extraction from natural sources such as hair and wool fibers, natural

keratins are subject to extensive processing conditions that lead to formation of unwanted by-products. Additionally, natural keratins suffer from limited sequence tunability. Recombinant keratin proteins can overcome these drawbacks while maintaining the desired chemical and physical characteristics of natural keratins. Herein, we present the bacterial expression, purification, and solution characterization of human hair keratins K31 and K81. The obligate heterodimerization of the K31/K81 pair that results in formation of intermediate filaments is maintained in the recombinant proteins. Surprisingly, we have for the first time observed new zero- and one-dimensional nanostructures from homooligomerization of K81 and K31, respectively. Further analysis of the self-assembly mechanism highlights the importance of disulfide crosslinking in keratin self-assembly.

5.2 Introduction

Protein-based biomaterials have been extensively used for biomedical applications from tissue engineering and regenerative medicine to drug delivery and medical devices.[1-7] Their intrinsic biocompatibility, as well as the variety of chemical and structural properties positions them as superior scaffolds compared to traditional synthetic polymer-based materials. Collagen, elastin[8-10], silk[5, 11-12], and keratin[3, 13-15] provide a few exquisite examples of functional biopolymers that have been implemented for materials design. Among these, keratin biopolymers have some of the most unique properties that make them especially well-suited for the generation of biomaterials.[14, 16-24] First, as with silk and collagen proteins, keratin dimers retain their inherent propensity for self-assembly following extraction from natural sources and subsequent reconstitution into functional materials.[25-27] Conservation of the biological activity (i.e. self-assembly) of reconstituted keratins networks allows for creation of materials with desirable chemical and structural properties. Furthermore, cell binding motifs[28-29] found in keratin protein

sequences aid in cellular attachment and thus promote cell proliferation, a quality that improves the utility of keratin-based scaffolds for regenerative medicine applications. Additionally, proposed regulatory functions of keratin proteins, such as regulation of protein synthesis and cell growth and proliferation, further enhance their biological compatibility when used in biomaterials.[30-35] Due to these desirable characteristics, many keratin materials have been developed over the past two decades toward a wide variety of applications including wound healing, tissue engineering, and nerve regeneration.[15, 21-23, 36-43] For example, hydrogels derived from human hair keratin proteins[15] increased cellular activity and gene expression *in vitro* that enabled successful nerve recovery *in vivo*. Another notable example of keratin materials can be seen in the work by Tomblyn et al. where keratin-based hydrogels were used to deliver growth factors and muscle progenitor cells *in vivo*, demonstrating their potential use in tissue repair systems.[44]

Traditional keratin biomaterials are produced from proteins obtained through extraction from natural sources such as wool and human hair fibers. This process involves considerable sample treatment and the use of harsh solvents. Consequently, and despite the numerous advantages of keratin materials, the extensive processing needed for extraction presents a major challenge to keratins use in some biomedical applications. Extraction procedures can result in protein damage and unwanted by-products that may elicit an undesired immune response[45], as well as change the network self-assembly capacity of the final material. Moreover, the properties of extracted materials are critically dependent on the raw material source. Extracted keratin materials are limited in the tunability of their nanostructure and chemical and biological properties as sequence modifications can only be achieved via exogenous chemistry. The use of recombinant proteins provides a means for overcoming all of the aforementioned limitations while

maintaining the desired characteristics. Recombinant DNA technology allows for specific design of protein sequence, structure, and function, which creates the ability to develop tunable biomaterials with tailored chemical and mechanical properties.[46-47] Indeed, recombinant silk, collagen, resilin, and elastin proteins have all been successfully used as biomaterials.[9-10, 48-59]

Keratin proteins belong to a class of cytoskeletal proteins known as intermediate filaments (IFs).[60] Known for their specific self-assembly capabilities, keratin proteins are ubiquitous throughout epithelial and epidermal cells, and function as structural support. Although all keratins contain highly conserved secondary structural features, they are divided into two main groups based on differences in their primary sequence composition.[61] Keratin proteins found in epithelial cells are known as epithelial keratins, or "soft" keratins. Conversely, keratins found in epidermal appendages such as hair, nails, wool, and hooves make up trichocytic or "hard" keratins. Soft keratins self-assemble into loose bundles while keratins found in epidermal appendages form tougher, more rigid structures. This difference in structural properties can be attributed to the differences in the primary sequence of hard and soft keratins, specifically the high cysteine content of hard keratins. The increased number of cysteine residues allows for crosslinking through the formation of intermolecular disulfide bonds, and thus creates more tightly packed and mechanically rigid structures.[62] These interesting sequence-structure-function differences open up a breadth of possibility for the use of keratins for the tailored design of new biomaterials, as the differences in their primary amino acid sequences can be exploited to create materials with tunable properties.

The self-assembly process of keratin IFs has been extensively studied.[27, 63] It is known that the precursor to IF formation is the dimerization of one type I (acidic) and one type II (basic) keratin protein to create a heterodimer.[63] These obligate heterodimers then align in an antiparallel,

staggered conformation to form a tetramer. Parallel, head to tail stacking of tetramers creates protofilaments, which then assemble into the characteristic 10 nm IFs.[63-64] Herein, we describe the heterologous expression and purification of two human hair keratins, type I keratin, K31 and type II keratin, K81. While the extraction of natural keratin proteins only allows for isolation of keratin heterodimers, through recombinant protein production we were able to express and purify each heterodimer component individually. Investigation into their self-assembly process and solution behavior revealed novel self-assembly properties for each biopolymer. The characteristic heteropolymer formation (i.e. K31 and K81 pair) is maintained in recombinant proteins, indicating that the recombinant protein pair retains the desired functionality of its naturally produced counterpart. This work further shows the utility of recombinant proteins for use in next generation biomaterials design. Furthermore, the ability to modulate each biopolymer individually expands the potential design platform.

5.3 Materials and Methods

5.3.1 Gene design and cloning

Gene sequences corresponding to K31 and K81 were synthesized by GeneWiz Inc (Plainfield, NJ). The amino acid sequence for each protein was reverse translated to its corresponding DNA sequence and optimized for *E. coli* codon usage. Each sequence included restriction sites BamHI at the 5' end and HindIII at the 3' end. The desired gene was digested from plasmid puc57 using BamHI and HindIII enzymes. Following isolation of the DNA fragment, each gene was subsequently ligated into plasmid pProExHtam, which contains an N-terminal histidine affinity tag and an ampicillin resistance gene. All enzymes were purchased from New England Biolabs (Ipswich, MA). Gene sequencing, completed by the Virginia Bioinformatics Institute (Blacksburg, VA), confirmed that the plasmids contained the gene for K31 and K81 proteins. Protein sequences are provided in Supporting Information.

5.3.2 Protein Expression and Purification

K31 and K81 were expressed and purified following the same protocols. The proteins were first expressed in BL21 (DE3) *E. coli* cells. Cell cultures were grown for 16 hours overnight in Luria Broth (LB) media at 37℃ with shaking at 250 rpm. Cells were then diluted 1:100 in LB media and grown to an optical density of 0.6-0.8 at which time protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Protein expression was conducted at 37℃ for 4 hours. Following expression, cells were harvested through centrifugation at 5,000 rpm for 15 minutes and the cell pellet was resuspended in lysis buffer, pH 8, containing 50 mM Tris HCl, 300 mM sodium chloride, and 1% Tween 20, and then stored at -80℃ until purification. The desired protein was purified from inclusion bodies under denaturing conditions following a procedure adapted from Honda et al.[65] The resuspended cell pellet was first thawed in a 37℃ water bath. Following this step, 10 mg/mL of lysozyme was added to the sample and incubated on ice for 30 minutes. Subsequently, 10 mM $MgCl_2$, 1 mM $MnCl_2$, and 10 µg/mL of DNase were added to the mixture and incubated on ice for 30 minutes. Following incubation, 25 mL of detergent buffer, pH 8, consisting of 20 mM Tris HCl, 200 mM NaCl, 1% Triton X-100, and 2 mM EDTA was added and mixed with the protein sample. The sample was then centrifuged at 5,000 rpm for 15 minutes, and the supernatant removed. This step was repeated until a tight pellet of inclusion bodies was formed. After obtaining the desired inclusion body pellet, 30 mL of extraction buffer, pH 8, containing 10 mM Tris HCl, 2 mM EDTA, 8 M urea, 10 mM βME, and 1 protease inhibitor cocktail tablet was added to resuspend the pellet. The sample was then centrifuged at 16,000 rpm for 1 hour. The supernatant was collected for purification using a standard Ni-NTA affinity purification protocol and eluted with 300 mM imidazole in lysis buffer with 8 M urea.

5.3.3 Gel Electrophoresis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to estimate molecular weight of the purified protein. Samples were prepared in a 1:1 ratio of SDS buffer to protein and analyzed on a 10% acrylamide gel. A broad range protein marker from New England Biolabs (Ipswich, MA) was used as a standard and contained proteins from 212 kDa to 6.5 kDa.

5.3.4 Mass Spectrometry

Mass spectrometry and peptide fragmentation analysis of K31 and K81 were performed at the Fralin Life Sciences Institute at Virginia Tech following a previously published procedure.[17]

5.3.5 Dialysis

Following affinity purification and molecular weight verification by SDS-PAGE and MS analysis, K31 and K81 were individually dialyzed out of elution buffer, pH 8, containing 300 mM NaCl, 50 mM Tris HCl, 300 mM imidazole, 10 mM βME, and 8 M urea. In the first step of dialysis the protein was dialyzed against buffer, pH 8, with 10 mM Na_2HPO_4, 75 mM NaCl, 5 mM DTT, and 8 M urea. Four additional dialysis steps were completed with decreasing amounts of urea equal to 6, 4, 2, and 0 M. Each of the steps were completed at 3 hour intervals except for the last step, which was allowed to equilibrate overnight.

5.3.6 Size Exclusion Chromatography

Size exclusion chromatography (SEC) was completed using a Dionex chromatography system. Protein fractions were detected using the Ultimate 3000 UV/Vis detector at 280 nm and analyzed using Chromeleon v6.8 chromatography software. Samples were removed from the dialysis cassette after each 3 hour equilibration period and filtered through a 0.22 µm filter prior to

injection. SEC analysis was completed with a 0.5 mL/min flow rate, and a mobile phase corresponding to the relevant dialysis buffer.

5.3.7 Dynamic Light Scattering

The average particle size and size distribution of proteins in solution was measured using a Malvern Zetasizer Nano-ZS. All dynamic light scattering (DLS) measurements were performed at 25 °C. All solutions were passed through a 0.22 μm filter prior to measurement. Samples correspond to each dialysis step with varying concentrations of urea, as was described for the SEC analysis.

5.3.8 Transmission Electron Microscopy

Transition electron microscopy (TEM) analysis was performed on a Philips EM420 microscope with an accelerating voltage of 120 kV. The TEM samples were prepared on 300 mesh carbon-coated copper grids (Electron Microscopy Science, Hatfield, PA). The keratin sample was deposited on the grid for 1 minute and excess sample was removed. Two water washes were performed to remove salt. The samples were stained with 2% uranyl acetate for 30 seconds, excess stain was removed and the sample allowed to air-dry for 24 hours before observation under the microscope.

5.4 Results and Discussion

5.4.1 Recombinant expression and purification of keratin proteins

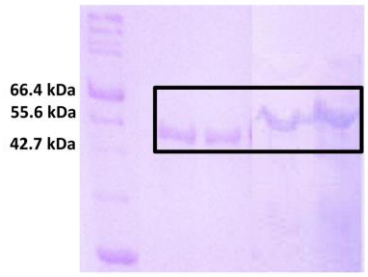
Recombinant hard and soft keratins have been previously produced and their assembly properties studied.[64-68] Human hair keratins are hard keratins found in epidermal appendages of most mammals.[62] Their utility as scaffolds for the development of nanostructured biomaterials has been demonstrated in a variety of applications over the past two decades.[15-17, 69] In an effort to develop tailored and tunable materials, we have recombinantly produced and characterized

two human hair keratins that are known to self-assemble and form durable, fibrous structures in nature.

5.4.2 Molecular weight and sequence analysis

Human hair keratins, K31 (type I) and K81 (type II) were expressed using an *E. coli* expression system with yields of 20 mg/L. Synthetic genes used to produce the proteins were sequences homologous to those found in nature optimized for *E. coli* codon usage. Purification from inclusion bodies resulted in the isolation of pure, full length proteins. Expression of each protein with a six-histidine affinity tag on the N-terminus allowed for further Ni-NTA affinity purification. Following affinity purification, analysis of the protein fractions by SDS-PAGE verified the isolation of proteins with the molecular weights expected for K31 and K81 (Figure 5.1A). Comparison of the purified fractions to the molecular weight standards (New England Biolabs) shows that the single protein band corresponds to the expected molecular weights of K31 and K81 estimated from their amino acid sequence using the Expasy Protparam tool, which was approximately 50 kDa for K31 and 58 kDa for K81. Additionally, SDS-PAGE analysis indicates that the proteins were expressed and purified with little to no impurities, as the desired protein bands are the only bands observed in the gel. Purification of keratin proteins without detrimental by-products has proven challenging when dealing with extracted materials. The ability to consistently produce keratins with high purity will further enable consistent self-assembly and network structures. Moreover, biomaterial scaffolds developed from starting materials of high purity have the potential to reduce or even eliminate biological responses associated with the low molecular weight components of keratin extracts.[20, 70]

A)



B)

Accession Number (NCBI)	Protein/Gene	Score	Matches	Molecular Weight (Da)
NP_002268	Keratin 31, type I cuticular Ha1 [Homo sapiens]/KRT31	111	54	47, 207
Q14533	Keratin 81, type II cuticular Hb1 [Homo sapiens]/KRT81	1670	33	54,936

Figure 5.1. Molecular weight analysis of recombinantly expressed K31and K81. (a) SDS-PAGE analysis of purified fractions analyzed on a 10% acrylamide gel stained with coomassie blue. Bands outlined in black box were excised for MS peptide fragmentation analysis (b) Results of MS peptide fragmentation analysis of protein bands highlighted in the black box.

Identity of the expressed proteins was verified using peptide fragmentation analysis and mass spectrometry (MS). Following the procedure reported by de Guzman *et al*., protein bands from the SDS-PAGE analysis were excised from the gel for extraction and subsequent sequence analysis. Results from this procedure indicate that recombinant, full length K31 and K81 were successfully expressed and purified. The expected molecular weight of the recombinant K31 was estimated at 50,163 Da using the Expasy Protparam tool. This estimation agrees well with the results from SDS-PAGE and the MS data (Figure 5.1B). Data obtained from the peptide fragmentation analysis matched the protein to K31 type 1 keratin Ha1 with a molecular weight of 47,207 Da. The discrepancy in molecular weights of the recombinant K31 and the protein matched during sequencing results from the additional N-terminal His-tag region of the recombinant protein that has a molecular weight of approximately 2,944 Da. Using the same procedures, the K81 sequence, with an expected molecular weight of 57,894 Da, was matched to K81 type II keratin HbI with a molecular weight of 54, 936 Da (Figure 5.1B). Analogous to K31, the difference in the estimated and calculated molecular weight is attributed to the N-terminal sequence added to the recombinant proteins. Thus, full-length human hair keratins with no truncations, deletions, or unwanted modifications to the sequence were expressed.

5.4.3 Solution characterization of purified keratin proteins

5.4.3.1 Size exclusion chromatography

Keratin proteins form obligate heterodimers from one acidic and one basic keratin, even under strong denaturing conditions.[63] This is consistent with keratins extracted from natural raw materials as these are also in their heterooligomeric state, and those derived from hair or wool fibers have not, to date, been reported to have been isolated in their monomeric form. Thus, by expressing individually acidic, K31, and basic, K81, keratin we are in a unique position to independently study each protein, especially their propensity for homooligomerization. Although not observed in nature, homooligomers are relevant in the realm of biomaterials design. Moreover, these nanostructures are accessible only through recombinant protein production. To investigate the solution phase behavior of, K31 and K81, as well as their equimolar mixture we have used SEC. Following affinity purification, samples were dialyzed into buffer containing 8 M urea, which is a strong chaotropic agent. As mentioned previously, hard keratins have high thiol content, 6% and 6.7% cysteine in the protein sequences for K31 and K81 respectively. Thus, in order to prevent random disulfide bond formation in the protein denatured state, the buffer system also included a strong reducing agent, 5 mM DTT. Under these conditions, proteins will typically be in a denatured state, and the expected result of the SEC analysis would be protein in its monomeric form. However, initial studies showed that both K31 and K81 are in multiple oligomeric states even in 8 M urea and 5 mM DTT (Figure S1). This interesting result suggests that K31 and K81 may self-oligomerize to form higher-order structures. Fractions corresponding to the peaks in the chromatograms were collected and analyzed via SDS-PAGE (Figure S2). Only one protein band, corresponding to either K31 or K81, was present in all fractions, indicating that shorter elution time peaks indeed correspond to homooligomers in

multiple oligomerization states and not to sample impurities. Using BioRad protein standards, we can estimate that the K31 peaks observed in the SEC correspond to oligomeric states of approximately a monomer, tetramer, and 9-mer for the peaks labeled one, two, and three (Figure S1). Peak four was determined to be an aberrant buffer artifact as it appears in all sample and buffer only chromatograms. It is important to note that the calibration curve used to estimate oligomeric states (Figure S3) provides a general approximation of the structures present; however, further analysis is needed to determine exact solution components formed during self-assembly.

The SEC analysis of the equimolar K31 and K81 mixture in 8 M urea also resulted in chromatograms indicative of multiple oligomeric states (Figure S1). SDS-PAGE analysis of each fraction again confirmed that all fractions contained only K31 and K81 proteins (Figure S2). Thus, the SEC chromatograms are consistent with multiple oligomeric states, suggesting existence of higher order structures. Table S1 provides a complete list of retention times and estimated oligomeric states for each peak in K31, K81, and K31/K81 chromatograms. Interestingly in the chromatogram of K31/K81 there is one newly observed peak (peak 3), not present in K31 or K81 chromatograms (Figure S1), which corresponds to the estimated dimer of K31 and K81. However, in all fractions observed in the SDS-PAGE analysis it appears the concentration of K31 is higher. This is likely due to remaining homooligomers of K31 that do not preferentially heterodimerize with K81 due to the strong disulfide bonds already present. Therefore, under these solution processing conditions we do not observe complete heterooligomerization and would need to process the samples under stronger reducing conditions in order to obtain more efficient heterooligomerization.

As previously stated, K31 and K81 proteins are hard keratins, and therefore contain high cysteine content in their primary amino acid sequences. This sequence feature increases the potential for intermolecular disulfide crosslinking and thus facilitates the formation of higher order structures, even under the described solution conditions. Although heterodimerization under denaturing conditions was previously reported for a pair of soft keratins, higher order oligomers were not observed in those studies.[63] However, soft keratin proteins have a lower cysteine residue content, less than 1%,[71] and perhaps further oligomerization under denatured conditions is not feasible.

5.4.3.2 TEM

Nano scale structures of K31 and K81 homooligomers, and K31/K81 heterooligomers were investigated by TEM. Samples were prepared through step-wise dialysis of each protein solution against buffer containing 10 mM Na_2HPO_4, 75 mM NaCl, and 5 mM DTT with decreasing concentrations of urea. Slow removal of the denaturant enables the return of the proteins to their native conformations. Following removal of urea, K31, K81, and K31/K81 were imaged using TEM. All samples for TEM were prepared at approximately 1 mg/mL or greater, which is above the reported critical assembly concentration of 0.4 mg/mL.[68, 72]

During step wise dialysis one-dimensional nanostructures formed in K31 and K31/K81 samples (Figure 5.2 A, C, D). The homooligomerization of K31 results in tightly packed structures with an average width of 150 nm and length of 1 µm (Figure 5.2 A). Heterooligomerization of the K31/K81 pair resulted in fibers that formed bundles ranging from 50 to 300 nm wide and that were several µm long (Figure 5.2 C). The typical fiber entanglement that enables hydrogel formation in bulk samples was also present in these samples (Figure 5.2 D). The K81 formed regular, spherical zero dimensional (i.e. all dimensions at nanoscale, x, y,

and z, d ≤ 100 nm) structures with an average diameter of 80 nm (Figure 5.2 B). Although some packing of these structures is evident, further fiber formation did not occur.

As the K31/K81 heteropolymer pair represents a typical human hair keratin combination, the formation of fibrous structures from these proteins is expected. The structures observed for the K31/K81 heterooligomer closely resemble the typical keratin IFs. Due to the high cysteine composition of both K31 and K81, it is likely that these structures form extensive intermolecular disulfide bonds leading to subsequent bundling of individual fibers. Interestingly, K31 homooligomerized into fibers even in the absence of its obligate dimerization partner, K81. These fibers do not have the typical IF morphology, but are relatively thicker and shorter. This difference can be attributed to the number and pattern of disulfide bonds in homooligomers versus obligate heterooligomers. It has been previously shown that disulfide bonding between keratin pairs may result in fibers of shortened lengths *in vitro* when both heterodimer components are involved in intermolecular disulfide formation.[73] It has also been shown that control over disulfide bond formation through reducing agent of the solution plays an important role in the resulting IF length.[67] This is the first time, to our knowledge, that homooligomeric keratin fibers have been observed. Although these fibers are not observed in nature, and are probably aberrant structures *in vivo*, they are nevertheless relevant for future synthetic material design. The range of zero and one dimensional nanostructures observed in recombinant, pure keratin samples, indicate the potential to create a breadth of tailored biomaterials through both homo- and heterooligomerization of keratin proteins. Additionally, these structures were not accessible by previously existing keratin biopolymer extraction and processing methods. The ability to recombinantly produce these proteins in a pure and functional state will allow for future design and engineering of each individual component.

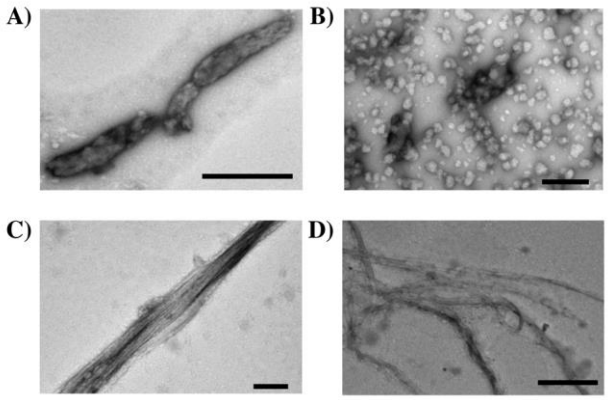


Figure 5.2. TEM images of keratin proteins (a) K31 (b) K81 (c) K31/K81 and (d) K31/K81. All images are stained with 2% uranyl acetate. Scale bars are 500 nm (a) 300 nm (b and c) and 500 nm (d).

5.4.4 Mechanism of K31 Self-assembly

SEC and TEM data for K31 suggest the proteins ability not to only self-oligomerize, but further self-assemble in higher order structures. These newly observed structures do not resemble standard IFs. To gain further insight into the mechanism of K31 self-assembly, we implement a combination of solution characterization techniques and TEM analysis.

5.4.4.1 Size Exclusion Chromatography and Dynamic Light Scattering

SEC and DLS were performed on K31 samples at each step of the urea removal. We hypothesized that higher order structure formation is associated with K31 folding into its native structure. As seen in Figure 5.3, chromatograms corresponding to 8M (black curve), 4M (blue curve), and 0M (red curve) urea solution contain several peaks indicative of higher molecular weight species. We have shown that these species are larger oligomeric states of K31 (Figure S1). Even in 8 M urea, K31 is not only a monomer, but in an equilibrium between several oligomeric states. When the concentration of urea decreased from 8M to 4M, the fraction of the K31 eluting at shorter retention times increased, indicating that equilibrium was shifted towards

higher oligomeric states. However, following complete removal of urea from the sample, the void volume peak previously observed is absent and remaining peaks shift towards longer retention times indicating that only lower MW species are present in the solution. At the same time, the overall intensity of the signal decreased. Together, these observations prompted us to consider if, upon urea removal, most of K31 is in nanostructures that are too large to enter the SEC column. All samples were filtered through a 0.22 μm filter, thus any structures in the sample that are larger than 0.22 μm will remain on the filter and not enter the SEC column.

Subsequent examination of the sample by DLS further confirms these observations. Figure 5.3 B depicts DLS data obtained from K31 under equivalent sample conditions to SEC. It is important to note that larger particles scatter more light and therefore the percent intensity does not correspond to the relative abundance of particles in the solution. DLS performed on filtered K31 samples in decreasing urea concentration show a shift in K31 oligomer size following the trend observed in SEC. The 8 M urea sample has two populations with sizes of approximately 18 and 255 nm. The population at 225 nm then shifts to a larger size (~315 nm) when the urea concentration is decreased to 4 M. Following complete urea removal and sample filtration, only ~ 50 nm oligomers remain in the K31 solution. As in SEC, the higher-order oligomers and structures are filtered out during the sample preparation process. To verify this observation, we compared unfiltered and filtered samples (inset Figure 5.3 B). It is evident from this data that a size population corresponding to larger structures in the sample is removed by filtering. Since all samples are also filtered before running SEC, it can be concluded that the final traces obtained by solution analysis methods only represent sample components that are not incorporated into higher-order structures. From the change in intensity of the SEC signal, an estimated 70% of the protein sample forms these higher order structures, while only 30% of the sample remains in the

solution phase. Additional TEM investigation allowed for further insight into the structures formed that correspond to the changes observed in SEC and DLS.

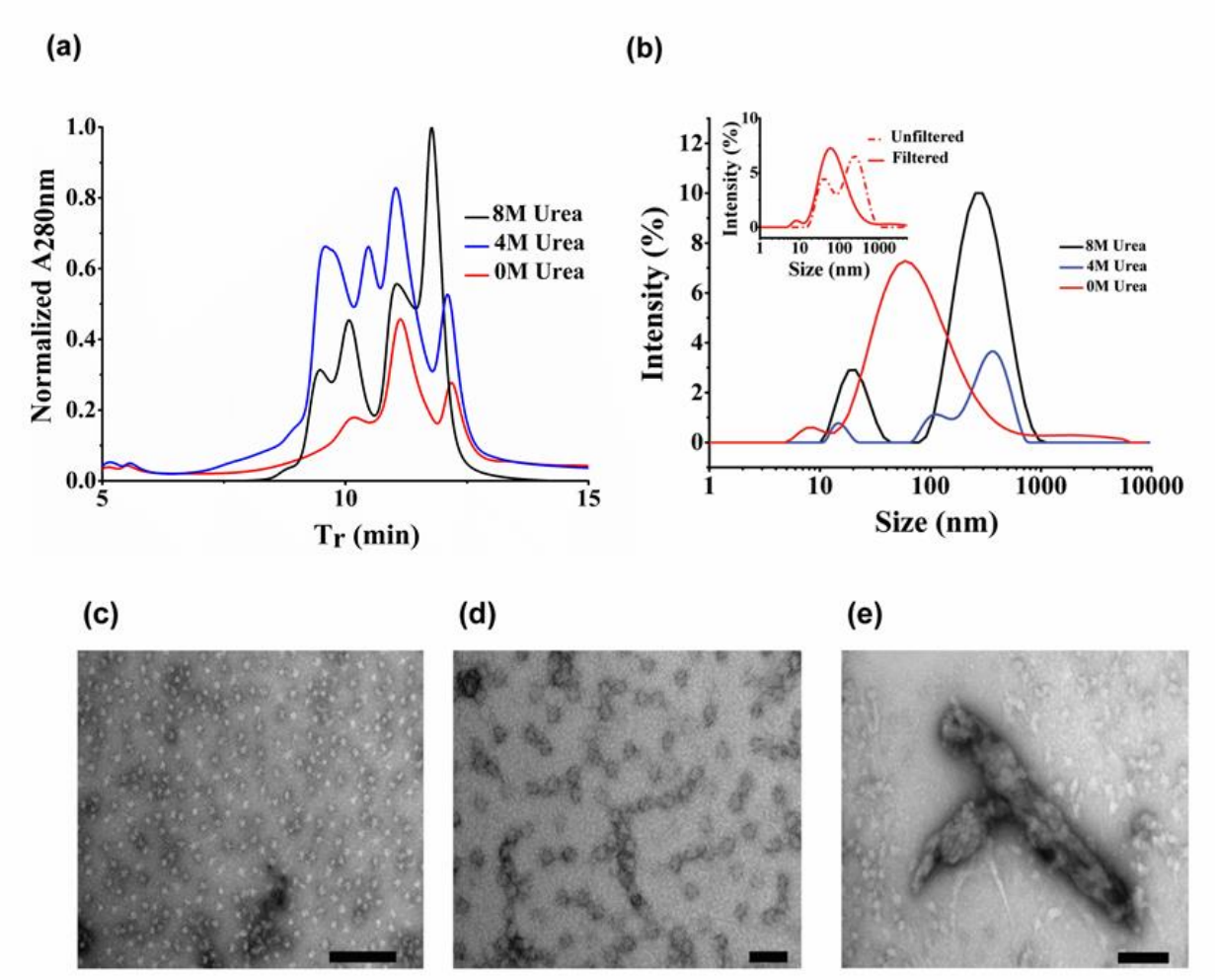


Figure 5.3. Characterization of homopolymer K31. (a) Normalized SEC and (b) DLS analysis of K31 in varying concentrations of urea: 8 M (black), 4 M (blue), 0 M (red). The black arrow in (a) indicates the column void volume. The inset figure shows K31 in 0 M urea before (dashed line) and after (solid line) filtering. (c-e) TEM of K31 in buffer containing no urea. Scale bars are 300 nm (c) 100 nm (d) and 200 nm (e).

5.4.4.2 TEM

As evidenced in the SEC and DLS experiments, K31 self-assembly proceeds as the urea concentration in the buffer system decreases. Initial formation of larger structures lead to a final

system comprised of largely K31 fibers and a small percentage of monomer and smaller oligomeric components once all denaturant has been removed. TEM analysis of K31 verifies these results. Structures observed in Figures 3 C-E coexist in the same urea-free K31 sample and can be observed throughout the TEM grid. As seen in Figure 5.3 C, uniform 30 nm structures are observed in the sample. These structures align and elongate to form the larger components also seen in the solution (Figure 5.3 D). Lastly, side-to-side stacking and bundling of these larger components leads to the final fiber formation (Figure 5.3 E). The bundled fibers represent unobservable portions of the sample from filtered SEC and DLS samples, as they are well above the 0.22 μm filter cutoff. Scheme 5.1 shows the overall proposed mechanism of assembly. Even under high denaturant concentrations monomeric keratin is in equilibrium with the several oligomeric structures formed in the first phases of self-assembly. However, as the denaturant is further reduced, K31 assembles into fibers through association of the oligomers. As the denaturant is removed, this equilibrium shifts towards higher oligomeric structures and finally fibers. However, the final step of assembly, fiber formation, is irreversible.

Keratin heterooligomerization and IF formation through hydrophobic and electrostatic interactions is a well-studied process.[63] Additionally, the role of cysteine residues found in the non-helical tail domain of epithelial keratins has been shown to promote bundling through formation of disulfide crosslinks.[74] However, unlike epithelial keratins, which contain less than 2% cysteine in their tail domain, hair keratins contain up to 14% cysteine content in their tail domain.[71] This striking difference in cysteine content likely explains the propensity toward homooligomerization and high degree of bundling observed in K31 self-assembled structures. In order to test this hypothesis, we completed the self-assembly procedure as previously described, but varied the strength of the reducing agent used. Representative TEM images of samples with

no reducing agent (Figure 5.4 A), 10 mM βME (Figure 5.4 B), and 5 mM DTT (Figure 5.4 C), show the effect of reducing agent strength on the resulting K31 fiber morphology. Under all solution conditions analyzed, K31 maintains its ability to assemble into fibers, but interestingly, as the reducing agent strength is increased, bundling and branching of fibers is decreased. K31 prepared with no reducing agent resulted in branched fibers that were more elongated than fibers originally observed in samples prepared with DTT. However, the large bundles observed were composed of individual fibers of approximately 10 nm, which corresponds to the typical IF structure. Increasing the reducing agent strength from no reducing agent to 10 mM βME resulted in fibers with less branching, but increased lateral association, where bundles were on the order of 1 μm in width. As previously shown, fibers assembled with DTT saw significantly less lateral association compared to βME and also less elongation than each of the other samples. Despite the differences in the final fiber morphology, spherical 30 nm structures were present in all samples. These results suggest that the precursor to self-assembly remains constant (first equilibrium step in Scheme 5.1) despite the reducing agent's effect on disulfide formation, and points to the importance of intermolecular disulfide bonds on the elongation and bundling of these structures into larger fibers. The ability to control crosslinking density and thus final fiber morphology through manipulation of solution processing conditions provides the opportunity to potentially control the resulting chemical and mechanical properties of materials through simple changes to solution conditions. Overall we conclude that the ability of K31 to homopolymerize into fibers is highly dependent on its ability to form disulfide bonds, and that this unique feature of hard keratin proteins can be further exploited for development of biomaterials with tunable properties.

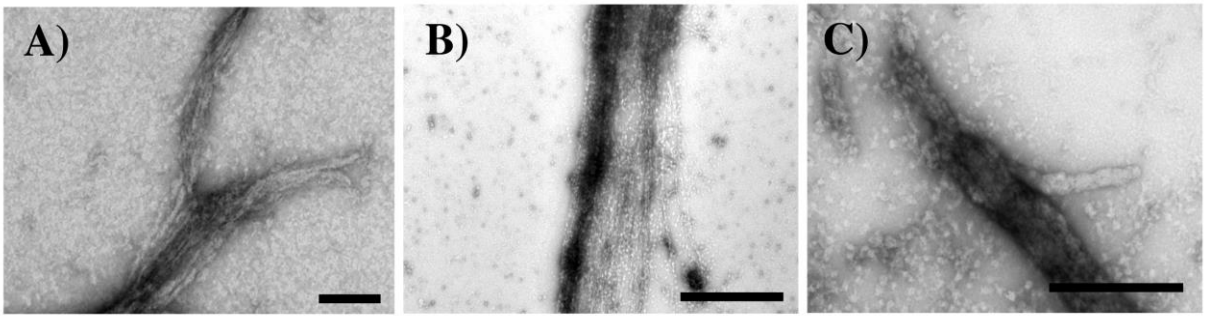


Figure 5.4. TEM images of K31 fibers with varying reducing agents. a) No reducing agent b) βME- 10 mM c) DTT- 5 mM. All images are stained with 2% uranyl acetate. Scales bars are 100 nm (a), 1 µm (b), and 500 nm (c).

5.5 Conclusions

The *E. coli* production of recombinant human hair keratins resulted in isolated proteins of the correct sequence and molecular weight. Hard keratins, K31 and K81, have high propensity for oligomerization in solution, which is attributed to their high cysteine content. IF formation is maintained in the heterooligomer K31/K81. Surprisingly, K31 also homooligomerized into fibers of varying morphologies, while self-assembly of homopolymer K81 resulted in zero-dimensional nanostructures. Thus, we have shown for the first time that obligate heterodimers are not required for keratin assembly into higher order structures. Although homooligomeric structures have not been observed in nature and are probably aberrant structures in natural materials, their *in vitro* existence expands the design scope for potential new biomaterials. Moreover, these structures are accessible only through recombinant production of monomeric proteins. Results of this work will enable further understanding of the IF formation in hard keratins, as well as new material for biomaterials development. The use of recombinant DNA technology and protein engineering techniques in the future will allow for sequence control and subsequent tailoring of material properties.

5.6 Supporting Information

K31 and K81 Protein sequences; Size exclusion chromatography of K31, K81, and K31/K81; SDS-PAGE of K31, K81, and K31/K81 SEC fractions; Calibration curve; SEC data table.

5.7 Acknowledgements

The authors would like to thank Virginia Tech, the VT Chemistry Department, and the Department of Biomedical Engineering and Mechanics for funding. The authors are also grateful to Dr. Rich Helm and Keith Ray of the VT Life Sciences Institute for MS analysis. The authors are also thankful to the Grove lab for discussion and careful reading of the manuscript.

5.8 References

1. Nectow, A. R.; Marra, K. G.; Kaplan, D. L., Biomaterials for the Development of Peripheral Nerve Guidance Conduits. *Tissue Engineering Part B-Reviews* **2012,** *18* (1), 40-50.
2. Pakulska, M. M.; Ballios, B. G.; Shoichet, M. S., Injectable hydrogels for central nervous system therapy. *Biomedical Materials* **2012,** *7* (2).
3. Rouse, J. G.; Van Dyke, M. E., A Review of Keratin-Based Biomaterials for Biomedical Applications. *Materials* **2010,** *3* (2), 999-1014.
4. Stoppel, W. L.; Ghezzi, C. E.; McNamara, S. L.; Black, L. D.; Kaplan, D. L., Clinical Applications of Naturally Derived Biopolymer-Based Scaffolds for Regenerative Medicine. *Ann. Biomed. Eng.* **2015,** *43* (3), 657-680.
5. Vepari, C.; Kaplan, D. L., Silk as a biomaterial. *Prog. Polym. Sci.* **2007,** *32* (8-9), 991-1007.
6. Goldberg, M.; Langer, R.; Jia, X. Q., Nanostructured materials for applications in drug delivery and tissue engineering. *Journal of Biomaterials Science-Polymer Edition* **2007,** *18* (3), 241-268.
7. Silva, R.; Fabry, B.; Boccaccini, A. R., Fibrous protein-based hydrogels for cell encapsulation. *Biomaterials* **2014,** *35* (25), 6727-6738.
8. Almine, J. F.; Bax, D. V.; Mithieux, S. M.; Nivison-Smith, L.; Rnjak, J.; Waterhouse, A.; Wise, S. G.; Weiss, A. S., Elastin-based materials. *Chem. Soc. Rev.* **2010,** *39* (9), 3371-3379.
9. MacEwan, S. R.; Chilkoti, A., Elastin-Like Polypeptides: Biomedical Applications of Tunable Biopolymers. *Biopolymers* **2010,** *94* (1), 60-77.
10. Rodriguez-Cabello, J. C.; Arias, F. J.; Rodrigo, M. A.; Girotti, A., Elastin-like polypeptides in drug delivery. *Adv. Drug Del. Rev.* **2016,** *97*, 85-100.
11. Wang, Y. Z.; Kim, H. J.; Vunjak-Novakovic, G.; Kaplan, D. L., Stem cell-based tissue engineering with silk biomaterials. *Biomaterials* **2006,** *27* (36), 6064-6082.
12. Kapoor, S.; Kundu, S. C., Silk protein-based hydrogels: Promising advanced materials for biomedical applications. *Acta Biomater.* **2016,** *31*, 17-32.
13. Apel, P. J.; Garrett, J. P.; Sierpinski, P.; Ma, J. J.; Atala, A.; Smith, T. L.; Koman, A.; Van Dyke, M. E., Peripheral Nerve Regeneration Using a Keratin-Based Scaffold: Long-Term

Functional and Histological Outcomes in a Mouse Model. *Journal of Hand Surgery-American Volume* **2008,** *33A* (9), 1541-1547.
14. Ham, T. R.; Lee, R. T.; Han, S.; Haque, S.; Vodovotz, Y.; Gu, J.; Burnett, L. R.; Tomblyn, S.; Saul, J. M., Tunable Keratin Hydrogels for Controlled Erosion and Growth Factor Delivery. *Biomacromolecules* **2016,** *17* (1), 225-236.
15. Sierpinski, P.; Garrett, J.; Ma, J.; Apel, P.; Klorig, D.; Smith, T.; Koman, L. A.; Atala, A.; Van Dyke, M., The use of keratin biomaterials derived from human hair for the promotion of rapid regeneration of peripheral nerves. *Biomaterials* **2008,** *29* (1), 118-128.
16. Aboushwareb, T.; Eberli, D.; Ward, C.; Broda, C.; Holcomb, J.; Atala, A.; Van Dyke, M., A Keratin Biomaterial Gel Hemostat Derived from Human Hair: Evaluation in a Rabbit Model of Lethal Liver Injury. *Journal of Biomedical Materials Research Part B-Applied Biomaterials* **2009,** *90B* (1), 45-54.
17. de Guzman, R. C.; Saul, J. M.; Ellenburg, M. D.; Merrill, M. R.; Coan, H. B.; Smith, T. L.; Van Dyke, M. E., Bone regeneration with BMP-2 delivered from keratose scaffolds. *Biomaterials* **2013,** *34* (6), 1644-1656.
18. de Guzman, R. C.; Tsuda, S. M.; Ton, M. T. N.; Zhang, X.; Esker, A. R.; Van Dyke, M. E., Binding Interactions of Keratin-Based Hair Fiber Extract to Gold, Keratin, and BMP-2. *PLoS One* **2015,** *10* (8).
19. Lin, Y. C.; Ramadan, M.; Van Dyke, M.; Kokai, L. E.; Philips, B. J.; Rubin, J. P.; Marra, K. G., Keratin Gel Filler for Peripheral Nerve Repair in a Rodent Sciatic Nerve Injury Model. *Plast. Reconstr. Surg.* **2012,** *129* (1), 67-78.
20. Nunez, F.; Trach, S.; Burnett, L.; Handa, R.; Van Dyke, M.; Callahan, M.; Smith, T., Vasoactive Properties of Keratin-Derived Compounds. *Microcirculation* **2011,** *18* (8), 663-669.
21. Reichl, S., Films based on human hair keratin as substrates for cell culture and tissue engineering. *Biomaterials* **2009,** *30* (36), 6854-6866.
22. Reichl, S.; Borrelli, M.; Geerling, G., Keratin films for ocular surface reconstruction. *Biomaterials* **2011,** *32* (13), 3375-3386.
23. Saul, J. M.; Ellenburg, M. D.; de Guzman, R. C.; Van Dyke, M., Keratin hydrogels support the sustained release of bioactive ciprofloxacin. *Journal of Biomedical Materials Research Part A* **2011,** *98A* (4), 544-553.
24. Wang, S.; Wang, Z. X.; Foo, S. E. M.; Tan, N. S.; Yuan, Y.; Lin, W. S.; Zhang, Z. Y.; Ng, K. W., Culturing Fibroblasts in 3D Human Hair Keratin Hydrogels. *Acs Applied Materials & Interfaces* **2015,** *7* (9), 5187-5198.
25. Thomas, H.; Conrads, A.; Phan, K. H.; Vandelocht, M.; Zahn, H., INVITRO RECONSTITUTION OF WOOL INTERMEDIATE FILAMENTS. *Int. J. Biol. Macromol.* **1986,** *8* (5), 258-264.
26. Steinert, P. M.; Gullino, M. I., BOVINE EPIDERMAL KERATIN FILAMENT ASSEMBLY INVITRO. *Biochem. Biophys. Res. Commun.* **1976,** *70* (1), 221-227.
27. Steinert, P. M.; Steven, A. C.; Roop, D. R., STRUCTURAL FEATURES OF EPIDERMAL KERATIN FILAMENTS REASSEMBLED INVITRO. *J. Invest. Dermatol.* **1983,** *81* (1), S86-S90.
28. Tachibana, A.; Furuta, Y.; Takeshima, H.; Tanabe, T.; Yamauchi, K., Fabrication of wool keratin sponge scaffolds for long-term cell cultivation. *J. Biotechnol.* **2002,** *93* (2), 165-170.
29. Verma, V.; Verma, P.; Ray, P.; Ray, A. R., Preparation of scaffolds from human hair proteins for tissue-engineering applications. *Biomedical Materials* **2008,** *3* (2).

30. Kellner, J. C.; Coulombe, P. A., Keratins and protein synthesis: the plot thickens. *J. Cell Biol.* **2009,** *187* (2), 157-159.
31. Kim, S.; Wong, P.; Coulombe, P. A., A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. *Nature* **2006,** *441* (7091), 362-365.
32. Coulombe, P. A.; Wong, P., Cytoplasmic intermediate filaments revealed as dynamic and multipurpose scaffolds. *Nat. Cell Biol.* **2004,** *6* (8), 699-706.
33. Izawa, I.; Inagaki, M., Regulatory mechanisms and functions of intermediate filaments: A study using site- and phosphorylation state-specific antibodies. *Cancer Sci.* **2006,** *97* (3), 167-174.
34. Magin, T. M.; Vijayaraj, P.; Leube, R. E., Structural and regulatory functions of keratins. *Exp. Cell Res.* **2007,** *313* (10), 2021-2032.
35. Kirfel, J.; Magin, T. M.; Reichelt, J., Keratins: a structural scaffold with emerging functions. *Cell. Mol. Life Sci.* **2003,** *60* (1), 56-71.
36. Hill, P. S.; Apel, P. J.; Barnwell, J.; Smith, T.; Koman, L. A.; Atala, A.; Van Dyke, M., Repair of Peripheral Nerve Defects in Rabbits Using Keratin Hydrogel Scaffolds. *Tissue Engineering Part A* **2011,** *17* (11-12), 1499-1505.
37. Yamauchi, K.; Maniwa, M.; Mori, T., Cultivation of fibroblast cells on keratin-coated substrata. *Journal of Biomaterials Science-Polymer Edition* **1998,** *9* (3), 259-270.
38. Lee, K. Y.; Kong, S. J.; Park, W. H.; Ha, W. S.; Kwon, I. C., Effect of surface properties on the antithrombogenicity of silk fibroin/S-carboxymethyl kerateine blend films. *Journal of Biomaterials Science-Polymer Edition* **1998,** *9* (9), 905-914.
39. Tachibana, A.; Kaneko, S.; Tanabe, T.; Yamauchi, K., Rapid fabrication of keratin-hydroxyapatite hybrid sponges toward osteoblast cultivation and differentiation. *Biomaterials* **2005,** *26* (3), 297-302.
40. Thilagar, S.; Jothi, N. A.; Omar, A. R. S.; Kamaruddin, M. Y.; Ganabadi, S., Effect of Keratin-Gelatin and bFGF-Gelatin Composite Film as a Sandwich Layer for Full-Thickness Skin Mesh Graft in Experimental Dogs. *Journal of Biomedical Materials Research Part B-Applied Biomaterials* **2009,** *88B* (1), 12-16.
41. Belcarz, A.; Ginalska, G.; Zalewska, J.; Rzeski, W.; Slosarczyk, A.; Kowalczulk, D.; Godlewski, P.; Niedzwiadek, J., Covalent Coating of Hydroxyapatite by Keratin Stabilizes Gentamicin Release. *Journal of Biomedical Materials Research Part B-Applied Biomaterials* **2009,** *89B* (1), 102-113.
42. Dias, G. J.; Mahoney, P.; Swain, M.; Kelly, R. J.; Smith, R. A.; Ali, M. A., Keratin-hydroxyapatite composites: Biocompatibility, osseointegration, and physical properties in an ovine model. *Journal of Biomedical Materials Research Part A* **2010,** *95A* (4), 1084-1095.
43. Shen, D. L.; Wang, X. F.; Zhang, L.; Zhao, X. Y.; Li, J. Y.; Cheng, K.; Zhang, J. Y., The amelioration of cardiac dysfunction after myocardial infarction by the injection of keratin biomaterials derived from human hair. *Biomaterials* **2011,** *32* (35), 9290-9299.
44. Tomblyn, S.; Kneller, E. L. P.; Walker, S. J.; Ellenburg, M. D.; Kowalczewski, C. J.; Van Dyke, M.; Burnett, L.; Saul, J. M., Keratin hydrogel carrier system for simultaneous delivery of exogenous growth factors and muscle progenitor cells. *Journal of Biomedical Materials Research Part B-Applied Biomaterials* **2016,** *104* (5), 864-879.
45. Anderson, J. M.; Rodriguez, A.; Chang, D. T., Foreign body reaction to biomaterials. *Semin. Immunol.* **2008,** *20* (2), 86-100.
46. Lin, C. Y.; Liu, J. C., Modular protein domains: an engineering approach toward functional biomaterials. *Curr. Opin. Biotechnol.* **2016,** *40*, 56-63.

47. Dinjaski, N.; Kaplan, D. L., Recombinant protein blends: silk beyond natural design. *Curr. Opin. Biotechnol.* **2016,** *39*, 1-7.
48. Foo, C. W. P.; Kaplan, D. L., Genetic engineering of fibrous proteins: spider dragline silk and collagen. *Adv. Drug Del. Rev.* **2002,** *54* (8), 1131-1143.
49. Reguera, J.; Fahmi, A.; Moriarty, P.; Girotti, A.; Rodriguez-Cabello, J. C., Nanopore formation by self-assembly of the model genetically engineered elastin-like polymer (VPGVG)(2)(VPGEG)(VPGVG)(2) (15). *J. Am. Chem. Soc.* **2004,** *126* (41), 13212-13213.
50. Stark, M.; Grip, S.; Rising, A.; Hedhammar, M.; Engstrom, W.; Hjalm, G.; Johansson, J., Macroscopic fibers self-assembled from recombinant miniature spider silk proteins. *Biomacromolecules* **2007,** *8* (5), 1695-1701.
51. Teng, W. B.; Cappello, J.; Wu, X. Y., Recombinant Silk-Elastinlike Protein Polymer Displays Elasticity Comparable to Elastin. *Biomacromolecules* **2009,** *10* (11), 3028-3036.
52. Hu, X.; Wang, X.; Rnjak, J.; Weiss, A. S.; Kaplan, D. L., Biomaterials derived from silk–tropoelastin protein systems. *Biomaterials* **2010,** *31* (32), 8121-8131.
53. Spiess, K.; Lammel, A.; Scheibel, T., Recombinant Spider Silk Proteins for Applications in Biomaterials. *Macromol. Biosci.* **2010,** *10* (9), 998-1007.
54. An, B.; Kaplan, D. L.; Brodsky, B., Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering. *Frontiers in Chemistry* **2014,** *2*.
55. Harris, T. I.; Gaztambide, D. A.; Day, B. A.; Brock, C. L.; Ruben, A. L.; Jones, J. A.; Lewis, R. V., Sticky Situation: An Investigation of Robust Aqueous-Based Recombinant Spider Silk Protein Coatings and Adhesives. *Biomacromolecules* **2016,** *17* (11), 3761-3772.
56. Jang, Y.; Champion, J. A., Self-Assembled Materials Made from Functional Recombinant Proteins. *Acc. Chem. Res.* **2016,** *49* (10), 2188-2198.
57. An, B.; DesRochers, T. M.; Qin, G. K.; Xia, X. X.; Thiagarajan, G.; Brodsky, B.; Kaplan, D. L., The influence of specific binding of collagen-silk chimeras to silk biomaterials on hMSC behavior. *Biomaterials* **2013,** *34* (2), 402-412.
58. Qin, G. K.; Lapidot, S.; Numata, K.; Hu, X.; Meirovitch, S.; Dekel, M.; Podoler, I.; Shoseyov, O.; Kaplan, D. L., Expression, Cross-Linking, and Characterization of Recombinant Chitin Binding Resilin. *Biomacromolecules* **2009,** *10* (12), 3227-3234.
59. Elvin, C. M.; Carr, A. G.; Huson, M. G.; Maxwell, J. M.; Pearson, R. D.; Vuocolo, T.; Liyou, N. E.; Wong, D. C. C.; Merritt, D. J.; Dixon, N. E., Synthesis and properties of crosslinked recombinant pro-resilin. *Nature* **2005,** *437* (7061), 999-1002.
60. Steinert, P. M.; Steven, A. C.; Roop, D. R., THE MOLECULAR-BIOLOGY OF INTERMEDIATE FILAMENTS. *Cell* **1985,** *42* (2), 411-419.
61. Coulombe, P. A.; Omary, M. B., 'Hard' and 'soft' principles defining the structure, function and regulation of keratin intermediate filaments. *Curr. Opin. Cell Biol.* **2002,** *14* (1), 110-122.
62. Yu, J. L.; Yu, D. W.; Checkla, D. M.; Freedberg, I. M.; Bertolino, A. P., HUMAN HAIR KERATINS. *J. Invest. Dermatol.* **1993,** *101* (1), S56-S59.
63. Coulombe, P. A.; Fuchs, E., ELUCIDATING THE EARLY STAGES OF KERATIN FILAMENT ASSEMBLY. *J. Cell Biol.* **1990,** *111* (1), 153-169.
64. Hatzfeld, M.; Weber, K., The coiled coil of in vitro assembled keratin filaments is a heterodimer of type I and II keratins: use of site-specific mutagenesis and recombinant protein expression. *The Journal of Cell Biology* **1990,** *110* (4), 1199-1210.
65. Honda, Y.; Koike, K.; Kubo, Y.; Masuko, S.; Arakawa, Y.; Ando, S., In vitro Assembly Properties of Human Type I and II Hair Keratins. *Cell Struct. Funct.* **2014,** *39* (1), 31-43.

66. Herrmann, H.; Wedig, T.; Porter, R. M.; Lane, E. B.; Aebi, U., Characterization of early assembly intermediates of recombinant human keratins. *J. Struct. Biol.* **2002,** *137* (1-2), 82-96.
67. Wang, H.; Parry, D. A. D.; Jones, L. N.; Idler, W. W.; Marekov, L. N.; Steinert, P. M., In vitro assembly and structure of trichocyte keratin intermediate filaments: A novel role for stabilization by disulfide bonding. *J. Cell Biol.* **2000,** *151* (7), 1459-1468.
68. Wu, K. C.; Bryan, J. T.; Morasso, M. I.; Jang, S. I.; Lee, J. H.; Yang, J. M.; Marekov, L. N.; Parry, D. A.; Steinert, P. M., Coiled-coil trigger motifs in the 1B and 2B rod domain segments are required for the stability of keratin intermediate filaments. *Mol. Biol. Cell* **2000,** *11* (10), 3539-58.
69. Pace, L. A.; Plate, J. F.; Smith, T. L.; Van Dyke, M. E., The effect of human hair keratin hydrogel on early cellular response to sciatic nerve injury in a rat model. *Biomaterials* **2013,** *34* (24), 5907-5914.
70. Poranki, D. R.; Van Dyke, M. E., The effect of gamma keratose on cell viability in vitro after thermal stress and the regulation of cell death pathway-specific gene expression. *Biomaterials* **2014,** *35* (16), 4646-4655.
71. Strnad, P.; Usachov, V.; Debes, C.; Grater, F.; Parry, D. A. D.; Omary, M. B., Unique amino acid signatures that are evolutionarily conserved distinguish simple-type, epidermal and hair keratins. *J. Cell Sci.* **2011,** *124* (24), 4221-4232.
72. Steinert, P. M., Organization of coiled-coil molecules in native mouse keratin 1/keratin 10 intermediate filaments: evidence for alternating rows of antiparallel in-register and antiparallel staggered molecules. *J. Struct. Biol.* **1991,** *107* (2), 157-74.
73. Feng, X.; Coulombe, P. A., A role for disulfide bonding in keratin intermediate filament organization and dynamics in skin keratinocytes. *J. Cell Biol.* **2015,** *209* (1), 59-72.
74. Lee, C. H.; Kim, M. S.; Chung, B. M.; Leahy, D. J.; Coulombe, P. A., Structural basis for heteromeric assembly and perinuclear organization of keratin filaments. *Nat. Struct. Mol. Biol.* **2012,** *19* (7), 707-+.

5.9 Supplemental Figures

Protein Sequences

K31

mpynfclpslscrtscssrpcvppschsctlpgacnipanvsncnwfcegsfngseketmqflndrlasylekvrqlerdnaelenlirer
sqqqepllcpsyqsyfktieelqqkilctksenarlvvqidnaklaaddfrtkyqtelslrqlvesdinglrrildeltlcksdleaqveslkee
llclksnheqevntlrcqlgdrlnvevdaaptvdlnrvlnetrsqyealvetnrreveqwfttqteelnkqvvssseqlqsyqaeiielrrtv
naleielqaqhnlrdslentltesearyssqlsqvqslitnvesqlaeirsdlerqnqeyqvlldvrarleceintyrsllesedcnlpsnpcatt
nacskpigpclsnpctscvppapctpcaprprcgpcnsfvr

K81

mtcgsgfggrafscisacgprpgrccitaapyrgiscyrgltggfgshsvcggfragscgrsfgyrsggvcgpsppcittvsvneslltpln
leidpnaqcvkqeekeqikslnsrfaafidkvrfleqqnklletklqfyqnreccqsnleplfegyietlrreaecveadsgrlaselnhvqe
vlegykkkyeeevslrataenefvalkkdvdcaylrksdleanvealiqeidflrrlyeeeirilqshisdtsvvvkldnsrdlnmdciiaei
kaqyddivtrsraeaeswyrskceemkatvirhgetlrrtkeeinelnrmiqrltaevenakcqnskleaavaqseqqgeaalsdarckl
aelegalqkakqdmaclireyqevmnsklgldieiatyrrllegeeqrlcegigavnvcvsssrggvvcgdlcvsgsrpvtgsvcsapc
ngnvavstglcapcgqlnttcgggscgvgscgisslgvgscgsscrkc

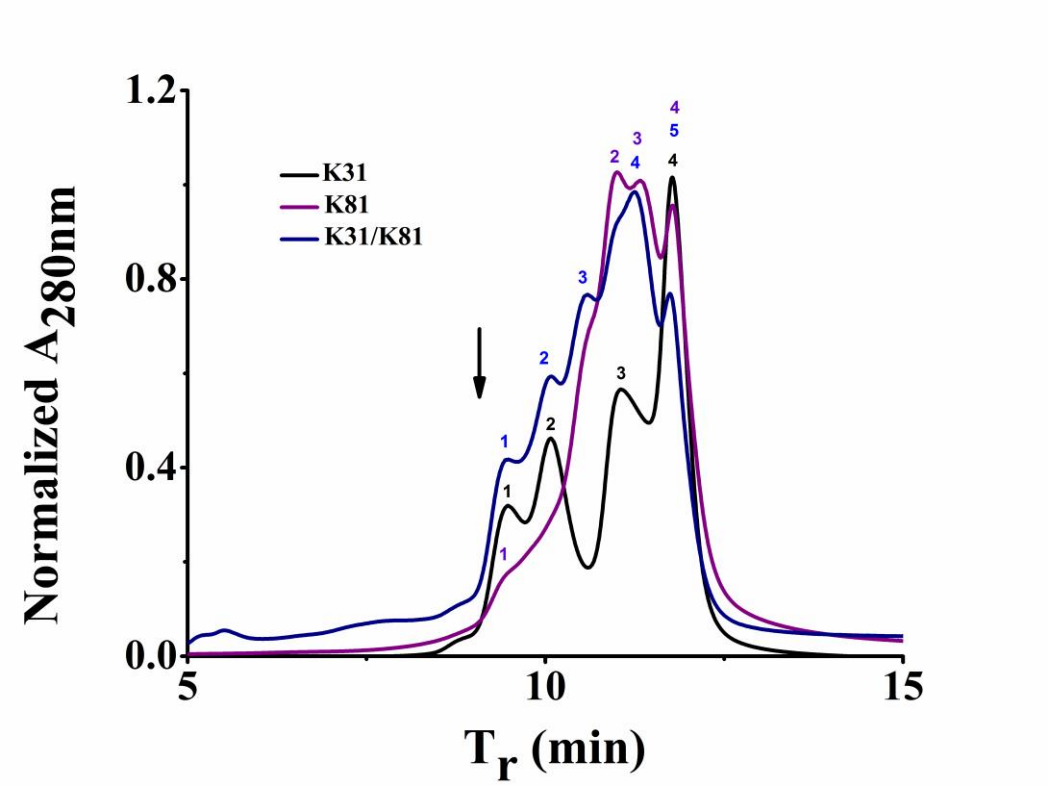


Figure S1. SEC traces of K31 (black), K81 (purple), and K31/K81 (blue) samples in 8 M urea. The black arrow indicates the void volume and the numbers indicate peak labels for K31 (black numbers), K81 (purple numbers) and K31/K81 (blue numbers). Absorbance is normalized to 1 for all samples.

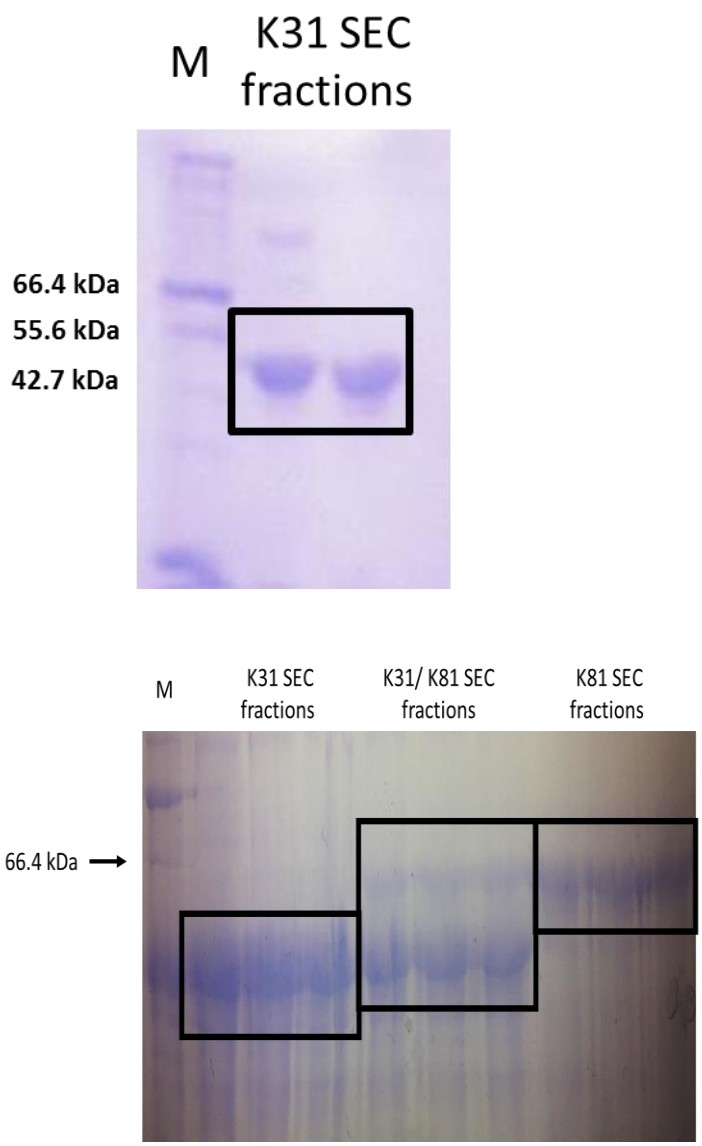


Figure S2. SDS-PAGE analysis of SEC traces of K31, K81, and K31/K81 samples in 8 M urea. The black boxes highlight protein bands for each sample.

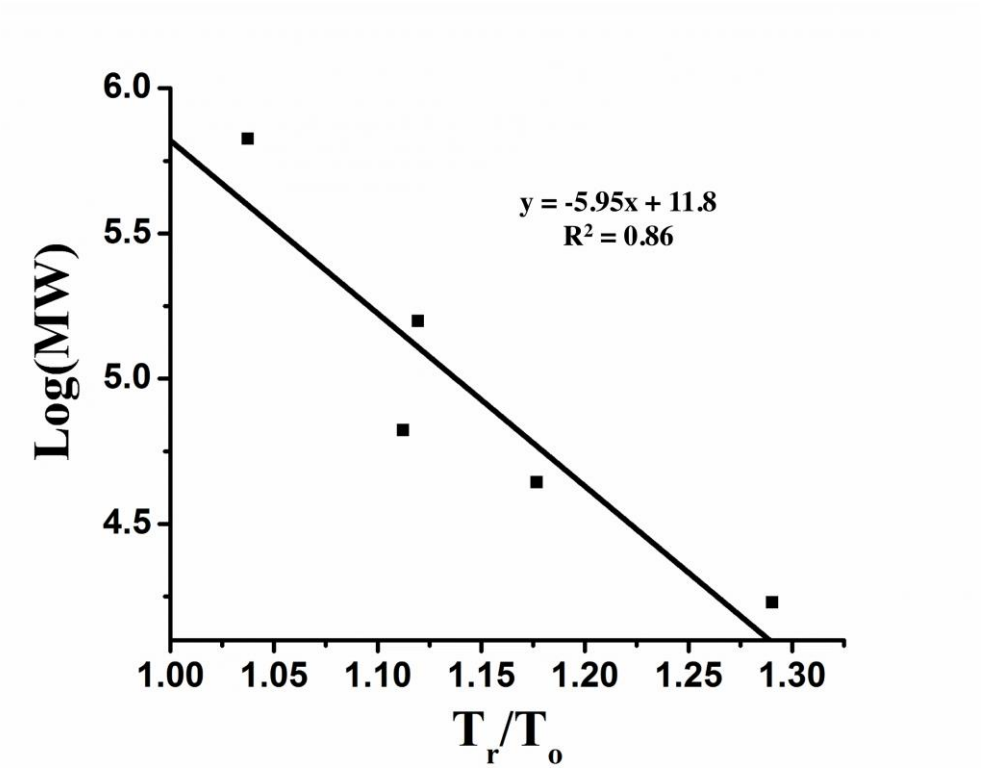


Figure S3. Calibration curve from protein standards used to estimate oligomeric states of peaks in the SEC for K31, K81, and K31/K81.

Table S1. Peak labels and estimated oligomeric states from SEC

Protein	Sample	Peak #	Retention time (T_r)	Estimated Oligomeric State
K31	8 M	1	9.46	9
K31	8 M	2	10.07	4
K31	8 M	3	11.03	1
K31	8 M	4	---	----
K31	4 M	1	9.58	8
K31	4 M	2	10.47	2
K31	4 M	3	11.03	1
K31	4 M	4	----	----
K31	0 M	1	10.18	3
K31	0 M	2	11.12	1
K31	0 M	3	----	----
K81	8 M	1	9.43	8
K81	8 M	2	11.00	1
K81	8 M	3	----	----
K31/K81	8 M	1	9.55	4
K31/K81	8 M	2	10.16	2
K31/K81	8 M	3	10.67	1
K31/K81	8 M	4	11.00	monomer
K31/K81	8 M	5	----	----

Chapter 6. A Comparative Study of Recombinant and Extracted Human Hair Keratins

*Rachael N. Parker[1], Kristina L. Roth[1], Alexis Trent[2], Mark E. Van Dyke[3], Tijana Z. Grove[1]**

1. Department of Chemistry, Virginia Tech, Blacksburg, VA 24060

2. Department of Materials Science and Engineering, Virginia Tech, Blacksburg, VA 24060

3. Department of Biomedical Engineering and Mechanics, Virginia Tech, Blacksburg, VA 24060

(*In preparation for submission)*

*Corresponding author email: tijana.grove@vt.edu

KEYWORDS: human hair keratin, biomaterials, recombinant protein, self-assembly, SEC, DLS, TEM

Abbreviations: K31- keratin 31, K81- keratin 81, IF- intermediate filament, SDS-PAGE- sodium dodecyl sulfate polyacrylamide gel electrophoresis, SEC- size exclusion chromatography, DLS- dynamic light scattering, TEM- transmission electron microscopy, DTT- dithiolthreitol

6.1 Abstract

Natural biopolymer-based materials have found much success in tissue engineering and regenerative medicine applications. Their intrinsic biocompatibility and biological activity make them well suited for biomaterials development. Specifically keratin-based biomaterials have demonstrated much utility for regenerative medicine applications including bone regeneration, wound healing, and nerve regeneration. However, studies of structure-function properties of keratin biomaterials have been hindered by the lack of homogeneous biopolymer preparations. The use of recombinant DNA technology presents a potential solution to overcome challenges associated with extraction of natural biopolymers. Furthermore, recombinant proteins offer a tunable design platform that allows for tailoring of the chemical and mechanical properties of materials while maintaining the desired inherent traits of natural biopolymers. Here we present a side-by-side comparison of natural and recombinant human hair keratin proteins K31 and K81. Trichocytic keratin biopolymers K31 and K81 were individually expressed in *E. coli* and purified to homogeneity using a combination of metal affinity chromatography and size-exclusion chromatography. When combined, K31 and K81 assemble into characteristic intermediate filament-like fibers. In contrast, the extracted keratin biopolymers are extracted as obligate dimers. In conclusion, homogenous preparation of recombinant human hair keratins result in proteins that maintain the self-assembly properties of natural keratins. These biopolymers will facilitate the study of structure-function properties of keratin biomaterials, as well as enable the development of future tailored biomaterials.

6.2 Introduction

Biopolymer-based scaffolds for tissue engineering and regenerative medicine have garnered much interest over the past few decades.[1-11] Natural biopolymers offer many advantages to traditional implant materials and synthetic polymer-based scaffolds. Although synthetic polymer-based scaffolds provide excellent structural support, they lack the intrinsic biological activity of natural biopolymers[12] and often illicit an unwanted foreign body response.[13] Many natural biopolymers have inherent self-assembly properties, functional cell binding motifs, and other important regulatory functions, such as control over cell migration and proliferation that provide them with intrinsic biological activity.[6, 10, 14-18] Furthermore, natural biopolymer-based scaffolds are often constructed from proteins found in the extracellular matrix (ECM) and other native tissues, which contributes to their biocompatibility, and minimizes undesirable immune responses. Conversely, synthetic polymer-based scaffolds often require chemical modifications in order to improve their biocompatibility and impart biological functions for subsequent use as biomaterials. Functionalization of these scaffolds has proven difficult and often requires the use of harsh solvents.[19-20] Additionally, it is difficult to control the degree of functionalization and the spatial organization of functional groups.[21-22] Therefore, due to the inherent characteristics of natural biopolymers (e.g. self-assembly, cell binding), they provide promising alternatives as biomaterials for regenerative medicine.[23]

When designing biomaterials for regenerative medicine and tissue engineering, certain criteria should be considered.[24] First, biomaterials need to provide adequate mechanical and structural support. Systems that allow for specific tailoring of mechanical and structural properties create an additional advantage. Moreover, the ability to facilitate and control cellular adhesion, migration, proliferation, and differentiation is essential for successful biomaterials use.[25] Lastly, materials that degrade over time and resorb into the body allow for temporary

implants. The ability to control the rate of degradation further enhances the utility of the material as the degradation rate can be tailored to the body's healing process.[26] Natural biopolymers possess many of these qualities, thus making them attractive replacements to traditional synthetic scaffolds used for the design of biomaterials. Biopolymers such as elastin[18, 27], collagen[14], keratin[28], and silk[29] are characterized by their hierarchical structures and exquisite mechanical properties. The aforementioned biopolymers can reassemble into their native structures following extraction from natural sources, thus retaining biological function and mechanical properties necessary for development of biomaterials.[17, 27, 30-31] Furthermore, biopolymers that have been implemented for materials design not only have important structural properties, but also important biological functions, including the ability to promote cellular attachment through specific cell-binding motifs in their primary amino acid sequences, induction of cell proliferation, as well as regulation of cellular differentiation and protein synthesis.[27, 32-34]

Of the natural biopolymers used for design of biomaterials, keratin has demonstrated much promise as a suitable scaffold for tissue engineering and regenerative medicine.[35-47] The inherent self-assembly of keratins into fibrous nanostructures is advantageous over designing synthetic polymer assemblies and allows for processing keratin into materials with excellent mechanical properties. Furthermore, keratins' biological and regulatory functions enhance its biocompatibility and provide useful bioactivity. For example, it has been proposed that keratin biopolymers contain cell binding motifs that participate in regulation of protein synthesis, cell growth, and proliferation.[48-49] Consequently, keratin has been employed in bone regeneration[50], wound healing[51], and nerve regeneration[52] applications. A notable example of keratin-based materials demonstrated its utility as a scaffold for bone regeneration through the delivery of recombinant human bone morphogenetic protein 2 (rhBMP-2) via a keratin hydrogel system.[50]

Another recent example of the benefits of keratin-based scaffolds can be seen in the work by Ham et al.[53] In this study, human hair keratins were used to fabricate hydrogels with controlled degradation profiles to allow for delivery of recombinant human insulin-like growth factor 1. However, despite the utility of keratin biomaterials, structure-property relationship investigations of keratin biomaterials have been mired by the lack of homogeneous biopolymer preparations. During extraction from natural sources such as hair and wool fibers, natural keratins are subject to extensive processing conditions that can lead to biopolymer damage. Additionally, the quality of the materials is highly source dependent. The harsh processing methods required for keratin extraction may result in protein damage leading to alterations in network assembly and undesirable immune response despite the biocompatibility of the biopolymer. Furthermore, major biopolymer components, K31 and K81, co-purify with low molecular weight additives such as melanin and other keratin associated proteins.

We have recently reported cloning, expression, and purification of recombinant human hair keratin 31 (K31) and keratin 81 (K81). K31 and K81 have been identified as the major components of extracted hair keratin materials,[54] which fall into the category of "hard" keratins. Hard keratins are found in epidermal appendages, such as hair, skin, and nails.[55] The defining feature of hair keratins comes from their high cysteine content, with hair keratins containing 5% or more cysteine residues.[56] This is in stark contrast to epithelial or "soft" keratins, which contain less than 1% of cysteines.[56] Consequently, hard keratins form more rigid structures, compared to the loose bundles formed by soft keratins, which result from extensive intermolecular disulfide bonds formed during assembly. Keratins extracted from natural sources are obtained either through oxidative extraction, keratose (KOS) or through reductive extraction, kerateine (KTN). In KOS samples disulfide bonds are not formed due to the conversion of the cysteine thiol

groups to sulfonic acid. KTN cysteines do contain thiol groups and readily form disulfide crosslinks. As a result of this chemical difference KOS materials are generally less stable than KTN materials as there are no covalent bonds formed in the material.[53] Herein, we present the side-by-side comparison of the solution characterization of recombinant human hair keratins K31 and K81 (rhK31 and rhK81) to KTN.

We compare the sample homogeneity, solution behavior, and self-assembly characteristics of rhK31 and rhK81 with the keratins obtained through reductive extraction from hair. Full-length rhK31 and rhK81 with no sequence deletions, truncations, or mutations were expressed in E.coli. Combination of affinity purification methods and size exclusion chromatography yielded homogeneous recombinant protein samples without unwanted byproducts or impurities. In contrast, extracted keratins are obtained as a heterogeneous mixture of K31/K81 dimer and additional hair protein components. Further solution characterization by SEC and DLS elucidated the initial stages of keratin fiber assembly. TEM images of the nanostructures formed during self-assembly of extracted and recombinant keratins allowed for comparison of fiber size and morphology. Recombinant keratins assemble into typical intermediate filaments (IF) that subsequently associate into large bundles. With homogeneous biopolymers in hand, we are now in a position to perform structure-property studies of keratin biomaterials.

6.3 Materials and Methods

6.3.1 Gene Design and Cloning of Recombinant K31 and K81

Amino acid sequences for K31 and K81 were reverse translated to the resultant DNA sequence and optimized for *E. coli* codon usage. Synthetic genes corresponding to each protein sequence were synthesized by GeneWiz Inc. The gene sequence contained restriction sites for subsequent cloning into the expression vector pProExHtam. BamHI and HindIII, at the 5’ and 3’

ends, were employed to ligate the gene into the plasmid, following digestion and isolation of the gene from the commercial plasmid puc57. Enzymes were purchased from New England Biolabs (Ipswich, MA). In order to confirm successful cloning of each gene, sequencing was completed by the Virginia Tech Bioinformatics Institute, which confirmed the correct gene sequences were contained in the plasmids. The same procedure was followed for cloning of the K31 and K81 genes. Plasmid pProExHtam contains an N-terminal histidine affinity tag to be used for protein purification.

6.3.2 Protein Expression and Purification of Recombinant Proteins

Recombinant K31 and K81 were expressed using an *E. coli* expression system. The same procedures were followed for both proteins. First, the proteins were expressed in BL21 (DE3) *E. coli* cells. Luria Broth (LB) was used for cell cultures. Cells were grown overnight for 16 hours in 50 mL of media at 37℃ with shaking at 250 rpm. LB media was then used to dilute the cells in a 1:100 ratio and cells were grown to an optical density (OD) of 0.6-0.8. Once OD had been reached, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was used to induce protein expression, which was completed at 37℃ for 4 hours. Cells were then harvested by centrifugation at 5,000 rpm for 15 minutes and subsequently resuspended in lysis buffer pH 8 containing 50 mM Tris HCl, 300 mM sodium chloride, and 1% Tween 20 and then stored at -80℃ until purification. An inclusion body purification procedure adapted from Honda et al.was used to extract and purify K31 and K81. Cell pellets were thawed in a a 37℃ water bath followed by a 30 minute incubation with 10 mg/ml of lysozyme. Following this step, 10 mM $MgCl_2$, 1 mM $MnCl_2$, and 10 µg/mL of DNase were each added to the protein samples and incubated for an additional 30 minutes. Detergent buffer pH 8 consisting of 20 mM Tris HCl, 200 mM NaCl, 1% Triton X-100, and 2 mM EDTA was then added at an equivalent volume to the sample volume and mixed well before centrifuging for 15 at 5,000 rpm. Following removal

of the supernatant, an additional 25 mL of detergent buffer was added to each sample, and the samples were again centrifuged. This procedure was repeated until a tight pellet of inclusion bodies was formed at which time 25 mL of extraction buffer was added. Extraction buffer consists of 10 mM Tris HCl, 2 mM EDTA, 8 M urea, 10 mM βME, and 1 protease inhibitor cocktail tablet at a pH of 8, and was used to resuspend the inclusion body pellet. The samples were then centrifuged for 1 hour at 16, 000 rpm. The resultant supernatant containing the extracted keratin proteins was collected for further purification. K31 and K81 containing an N-terminal histidine affinity tag were purified using a standard Ni-NTA affinity purification procedure. All buffers used for the purification process also contained 8 M urea to keep proteins in their denatured form until further dialysis.

6.3.3 Extraction of Natural Keratin Proteins

Natural keratins used for this study were extracted as previously described.[57] A sample of human Chinese hair was obtained from a commercial vendor and used as received. Keratin extraction was accomplished through a multistep reductive process described as follows: 100 grams of hair was placed into a 2 L solution of 0.5 M thioglycolic acid (TGA) adjusted to a pH of 10.5 and shaken at 100 rpm for 15 hours at 37 ^{0}C. The hair was recovered by sieve and the extraction solution retained. The hair fibers were then placed in a solution of 4 L of 100 mM tris and shaken at 100 rpm for 2 hours at 37 ^{0}C. Hair was again recovered by sieve and placed in a freshly prepared 1 L solution of 0.5 M TGA adjusted to a pH of 10.5 and shaken at 100 rpm for 15 hours at 37 ^{0}C. The resulting extraction solution was retained and the hair was then placed in 2 L of 100 mM tris and shaken at 100 rpm for 2 hours at 37 ^{0}C. The hair was then recovered by sieve and discarded. The extraction solution was retained and pooled with extraction solutions obtained in previous steps to form a solution of crude keratin extract. The extract was clarified

of particulate matter by centrifugation through a solids separator running at 30,000 rpm, followed by filtration through a filter membrane with a 20-25 average pore size. Keratin nanomaterials were obtained from this clarified crude keratin extract by ultrafiltration using a 100 kDa NLMWCO polysulfone, tangential flow filtration (TFF) cartridge. TFF was conducted with 10 volume washes against a buffer consisting of 10 mM disodium phosphate and 100 mM sodium chloride at pH 9.1, followed by 5 volume washes against purified water. The purified keratin nanomaterial solution was concentrated, titrated to pH 8.5, frozen and freeze dried to produce a keratin nanomaterial powder.

6.3.4 Gel Electrophoresis and Western Blot

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate purified protein prior to Western Blot analysis. Samples were prepared in a 1:1 ratio of SDS buffer to protein and analyzed on a 10% acrylamide gel. Extracted kerateine (KTN) and keratose (KOS) were diluted at 10mg/ml in sodium phosphate at pH 7.4, and recombinant proteins were prepared at 5 mg/mL. Following SDS-PAGE, proteins were transferred onto nitrocellulose membranes at 0.35A for 2 hours. The membranes were blocked with 5% non-fat dry milk in Tris Buffer Saline with 0.25% Tween 20 (TBST) for 1 hour. Guinea pig anti-human keratin-31 (K31) and guinea pig anti-human keratin-81 (K81) antibodies (Progen Biotechnik, Heidelberg, Baden-Württemberg, Germany) were used as primary probes and were both diluted at 1:2000 in blocking buffer for 1 hour. The membranes were washed three times with TBST, submerged into a 1:3000 dilution of the rabbit anti-Guinea pig IgG-HRP (Life Technologies) secondary probe for 1 hour, then again washed three times with TBST. All incubation periods were conducted at room temperature. Pierce ECL Plus substrate (Thermo Fisher Scientific) mix was added to the membranes 3 minutes prior to been imaged in a Fujifilm LAS-3000 Imager (General Electric).

6.3.5 Dialysis

Following affinity purification and molecular weight verification by SDS-PAGE and MS analysis, K31 and K81 were individually dialyzed out of elution buffer pH 8 containing 300 mM NaCl, 50 mM Tris HCl, 300 mM imidazole, 10 mM βME, and 8 M urea. In the first step of dialysis the protein was dialyzed against buffer pH 8 with 10 mM Na_2HPO_4, 75 mM NaCl, 5 mM DTT, and 8 M urea. Four additional dialysis steps were completed with decreasing amounts of urea equal to 6, 4, 2, and 0 M. Each of the steps were completed at 3 hour intervals except for the last step, which was allowed to equilibrate overnight. Keratin proteins that were previously extracted and lyophilized were reconstituted and prepared following the same procedure.

6.3.6 Size exclusion chromatography

A Dionex chromatography system with an Ultimate 3000 UV/Vis detector was used for size exclusion chromatography (SEC). Proteins were detected at 280 nm and analyzed with Chromeleon v6.8 chromatography software. Samples were analyzed following each step of the dialysis process. Each samples was passed through a 0.22 μm filter after a 3 hour equilibration period in the appropriate dialysis buffer. The mobile phase used for each sample corresponds to the relevant dialysis buffer. Samples were analyzed with a flow rate of 0.5 mL/min.

6.3.7 DLS

Dynamic light scattering (DLS) was completed using a Malvern Zetasizer Nano-ZS to analyze the average particle size and size distribution of extracted and recombinant keratin biopolymers in solution. Prior to measurement samples were filtered using a 0.22 μm filter. Each sample corresponds to steps during the dialysis process, and thus samples contain the corresponding buffer and urea concentration as described in the dialysis section. The Malvern software converts the intensity percent size distribution to volume percent using Mie theory.

6.3.8 Transmission Electron Microscopy

Using a PhilipsEM420 microscope with an accelerating voltage of 120 kV transition electron microscopy (TEM) analysis was performed on extracted and recombinant keratin biopolymers. Samples were prepared using 300 mesh carbon-coated grids purchased from Electron Microscopy Science. Following deposition of the sample on the grid, a 1 minute drying period was allowed before excess sample was removed. All samples were stained using 2% uranyl acetate, with a 30 second drying period. Excess stain was then removed and samples were allowed to air-dry for 24 hours prior to analysis.

6.4 Results and Discussion

Recombinant DNA technology and protein engineering are greatly influencing the next-generation biomaterials landscape. Recombinant protein-based biomaterials provide the structural and mechanical properties of their natural counterparts while offering the potential for creating materials with tunable sequences, and thus tailored and improved characteristics. Cellular binding motifs, degradation sites, and protein fusions exemplify some of the benefits afforded from recombinantly expressed biopolymers.[58-62] Indeed, many protein-based biomaterials, including silk[59, 63], elastin[61, 64], collagen[65-66], and resilin[67] have benefited from recombinant DNA technology. In addition to providing a path to increased structural and functional complexity of biomaterials, recombinant biopolymers are indispensable in structure-property relationships studies.

6.4.1 SDS-PAGE and Western Blot

In addition to K31 and K81, hair fibers contain different types of keratin proteins, including gamma-, alpha-, and beta-keratins.[44] Beta-keratins provide protection to the hair fiber,

gamma-keratins serve as a crosslinking agent, and alpha-keratins function as the main structural component.[44] As such, the desired component for fabrication of biomaterials is the alpha-keratins due to its important structural properties. However, through the extraction and purification process it is often difficult to remove the other types of keratin proteins, as well as additional by-products, which results in a heterogeneous mixture following extraction and purification. On the other hand, recombinant protein are expressed and purified individually. The N-terminal histidine tag enables metal affinity purification allowing for efficient removal of all other proteins and by-products not containing the specific affinity tag. In Figure 6.1 A we first visually compare solutions of extracted and recombinant proteins. The purified recombinant protein solution is labeled "1" and the extracted protein solution is labeled "2". The recombinant protein solution is clear while the extracted solution, even after purification, appears brown. The observed color is from melanin not removed during purification, demonstrating the difficulty of removing small molecular weight by-products during extraction. In addition, SDS-PAGE analysis of the purified proteins further indicates the improved sample homogeneity of the recombinant keratins (Figure 6.1 B). In Figure 6.1 B, lanes 1 and 2 represent affinity purified rhK31 (lane 1) and rhK81 (lane 2). A single protein band is present in each lane, showing that both recombinant proteins have been successfully purified with no observable by-products or unwanted contaminants. However, the extracted KTN sample in lane 3 has many protein components present, indicating either the difficulty in removing residual hair fiber components or protein degradation, or both.

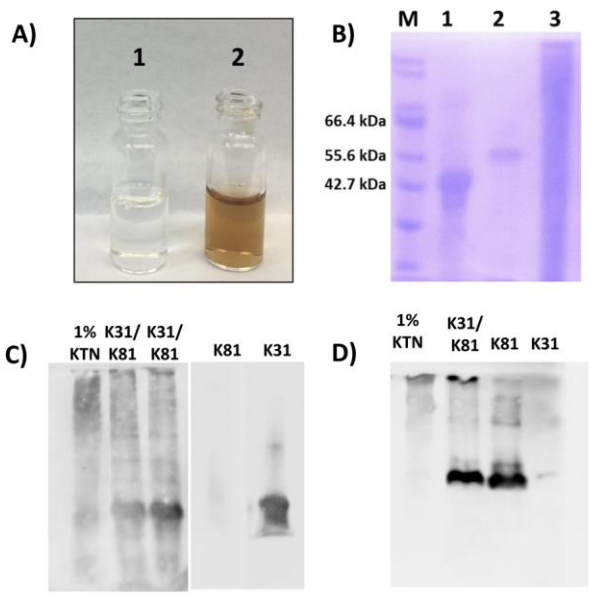


Figure 6.1. Comparison of purified recombinant and extracted keratins a) picture of purified recombinant (1) and extracted (2) solutions b) SDS-PAGE fractions: M-marker, 1-rhK31, 2-rhK81, 3-Extracted KTN c) Western blot with K31 antibody and d) Western blot with K81 antibody.

To verify the identity of the proteins observed in SDS-PAGE, we implemented a Western blot analysis. Antibodies specific for either K31 or K81 were used to analyze each sample (Figure 6.1 C and 6.1 D). Figure 6.1 C shows the results of the Western Blot with the K31 antibody, whereas the results in Figure 6.1 D were obtained from the K81 specific antibody. This analysis confirmed that the predominant proteins in the recombinant materials preparations were K31 and K81, as expected. Interestingly, however, in the extracted sample, the stained bands correspond to higher molecular weights even under denaturing and reducing gel conditions. This indicates the existence of only irreversible higher order oligomers in the extracted samples at the same solution conditions in which the recombinant sample is monomeric. The exception is the 1% KTN sample where a band corresponding to the molecular weight of monomeric K31 is observed (Figure 6.1 C). Furthermore, not all bands present in SDS-PAGE of the extracted

keratins also appear in the Western blot signifying the existence of additional proteins in the extracted sample. From the results obtained from SDS-PAGE and Western blot it appears that recombinant protein production and purification methods provide starting materials of improved homogeneity over that of extracted keratins. However, when rhK31 and rhK81 are combined, bands corresponding to higher molecular weights than monomers appear in Western blots, analogous to extracted KTN sample. Thus, the recombinant heterodimer is capable of heteropolymerization and formation of higher order structures. To further investigate heteropolymerization of rhK31 and rhK81 we utilized SEC and DLS.

6.4.2 SEC and DLS

To prepare samples for SEC analysis, rhK31 and rhK81 were mixed in the 8M urea buffer (Materials and Methods). KTN samples were resuspended in buffer of the same composition as the recombinant samples. SEC data obtained in buffer containing 8 M urea shows that both the extracted and recombinant keratin proteins contain structures that are larger than the K31/K81 heterodimer (Figure 6.2). Interestingly, the higher order oligomers are present in both samples even in the presence of a denaturant (urea) and reducing agent (DTT). It is important to note that all samples are passed through a 0.22 μm filter before SEC analysis. Therefore, all structures larger than the filter cut-off will be excluded from this method of analysis.

While five peaks are observed in the recombinant sample chromatogram the extracted heteropolymer chromatogram contains two broader peaks shifted toward shorter elution times (Figure 6.2). The estimated oligomeric states for each peak (labeled by the red numbers in the Figure 6.2) are shown in Table 6.1. Additionally, as seen in the SDS-PAGE of the KTN sample, SEC will show an overall signal of all proteins present in the extracted material. This may explain the broad peaks observed in the KTN chromatogram.

Table 6.1. SEC labels and estimated oligomeric states.

Peak #	Estimated oligomeric state-rhK31/K81	Estimated oligomeric state-KTN
1	Octamer	Not assigned (void volume)
2	Tetramer	Dimer
3	Dimer	----
4	Monomer	----

Interestingly, higher molecular weight peaks 1, 2, and 3 in the recombinant sample overlap with peaks 1 and 3 in the extracted sample suggesting that the K31/K81 octamer, tetramer, and dimer are present in both the extracted and recombinant samples. Peak 1 in the extracted sample, not observed in the recombinant proteins sample, elutes at the time corresponding to the column void volume. Thus sample components too large to be retained on the column are already present in the extracted sample. At the same time, there is no protein eluting at retention times corresponding to K31 and K81 monomers in the extracted sample. From the SEC analysis we conclude that in the recombinant sample under denaturing conditions the major fraction of the solution is monomeric rhK31 and rhK81 in equilibrium with higher order oligomers. On the contrary, the major fraction of the KTN solution corresponds to oligomers larger than octamers and the smallest observable component is a dimer. Dimer observation in the KTN sample is consistent with the formation of the obligate keratin dimer in nature, however it is interesting that this dimer is resistant to denaturing conditions.

The presence of larger heteropolymers in the extracted KTN sample, as observed in the SEC, SDS-PAGE, and Western blot analyses, are consistent with the extraction procedure that

relies on breaking down preformed, durable keratin-based structures. In order to efficiently extract the desired keratin biopolymers, the extensive network of intermolecular disulfide bonds must be reduced. Thus, the size of sample components acquired from extraction is dependent on the efficiency with which this network is disrupted, and the resulting higher order structures persist due to covalent interactions that are not affected by the solution conditions. Conversely, recombinant protein production facilitates assembly from each individual component. To further probe the solution behavior of the recombinant and extracted keratins we used DLS to monitor changes in oligomerization equilibrium as the concentration of denaturant in the sample was decreased.

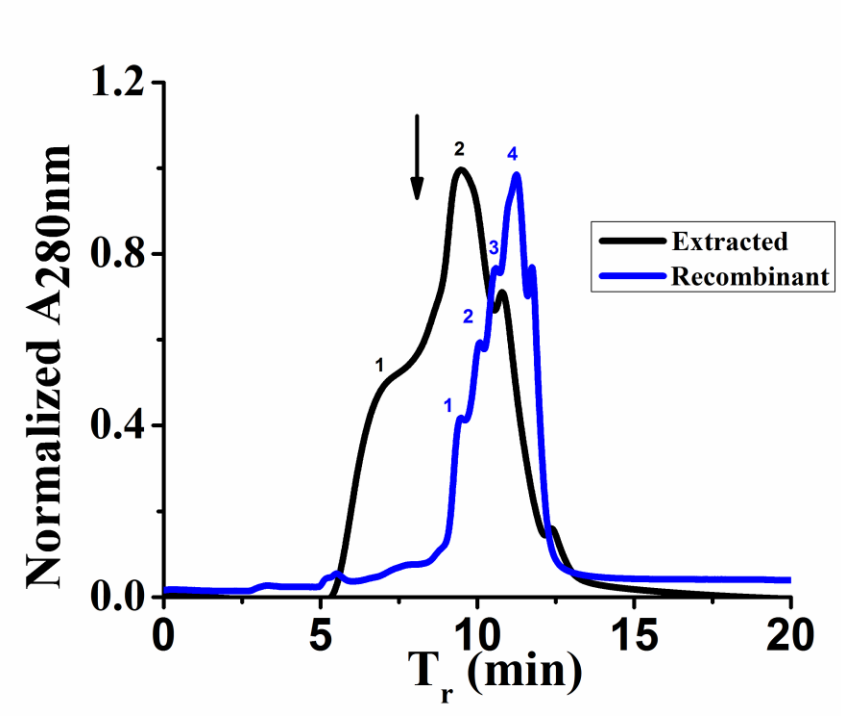


Figure 6.2. Chromatogram of recombinant (blue) and extracted (black) keratin in 8 M urea obtained from SEC. Red numbers correspond to peak labels listed in table 1. Absorbance is normalized to 1 for all samples.

We performed DLS on aliquots of samples obtained during each dialysis step. As the sample is dialyzed, the concentration of urea is removed in a step-wise manner in order to allow

the proteins to return to their native state, and in the case of keratin proteins, to allow for self-assembly to occur. Thus, as the urea concentration decreases we expect the oligomerization equilibrium to shift towards higher order oligomers and finally fibers. However, all samples are filtered prior to analysis using a 0.22 µm filter. Consequently, any sample components larger than 0.22 µm will not be observed, similarly to the SEC. Figure 6.3 shows recombinant (Figure 3A) and extracted (Figure 6.3B) heteropolymer samples at 8 M (blue), 4 M (red), and 0 M (black) urea concentrations. Consistent with SEC data, nanostructures present in the KTN sample are larger than in the recombinant sample at the same solution conditions. The recombinant heteropolymer is characterized by two size populations in 8 M urea and 0 M urea, and by the formation of a third, larger population when the urea concentration is decreased to 4 M. On the other hand, the extracted heteropolymer consists of two populations, in all urea concentrations, which steadily decrease in size. All peak labels and sizes are given in table 6.2.

Table 6.2. DLS peak labels and sizes.

Peak #	Sample	Population size-rhK31/K81	Population size- KTN
1	8 M	18	120
2	8 M	225	1100
1	4M	13	110
2	4 M	120	900
3	4 M	460	----
1	0 M	8	50
2	0 M	60	295

An overlay of the volume percent of recombinant and extracted keratins in 0 M urea is shown in Figure 6.3C. The volume percent distribution provides the relative proportion of the different sample components. The populations detected in the DLS at 0 M urea corresponds to

the keratin that has not been incorporated into IF after all denaturant has been removed from the system. The major scattering species in the recombinant keratin sample has an 8 nm hydrodynamic radius, consistent with K31 and/or K81 monomers.[68] The extracted keratin sample is mostly composed of 50 nm oligomers. This is consistent with measurements obtained at each urea concentration, as well as the SEC and Western blot analysis, and further indicates that upon extraction the starting material is not completely reduced to individual proteins.

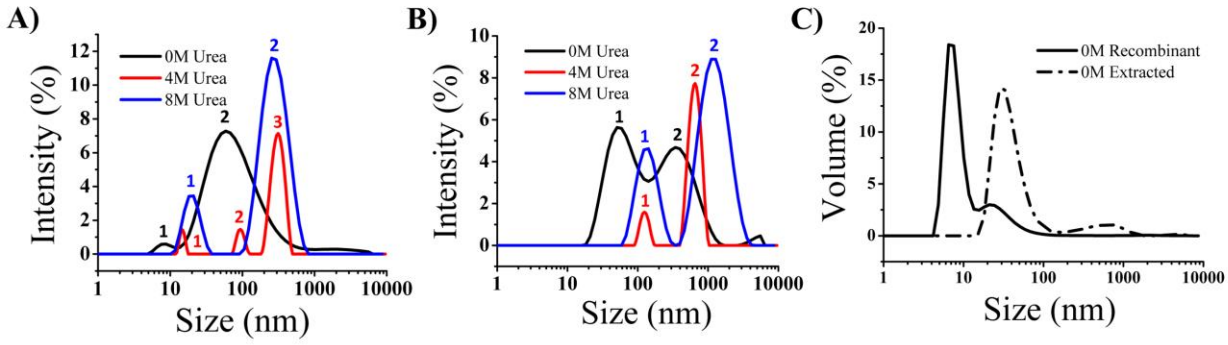


Figure 6.3. DLS from a) recombinant keratin urea series b) extracted keratin urea series and c) recombinant (solid line) and extracted (dotted line) keratin samples in 0 M urea.

We performed DLS on fractions of each sample obtained during each dialysis step. As the sample is dialyzed, the concentration of urea is removed in a step-wise manner in order to allow the proteins to return to their native state, and in the case of keratin proteins, to allow for self-assembly to occur. Figure 6.3 shows recombinant (figure 6.3A) and extracted (figure 6.3B) heteropolymer samples at 8 M (blue), 4 M (red), and 0 M (black) urea concentrations. The recombinant heteropolymer appears to shift from two populations of 60 nm and 8 nm in 8 M urea, to three populations of 13 nm, 120 nm, and 460 nm in 4 M urea, and then back to two populations of 18 nm and 255 nm in 0 M urea. Overall, the DLS shows an initial shift to larger

species followed by a subsequent change back to smaller oligomeric states. Conversely, the extracted heteropolymer sample demonstrated a consistent increase in two size populations as it changes from initial sizes of 50 nm and 295 nm in 8 M urea, and then increases to 110 and 900 nm in 4 M urea, with a final two populations of 120 nm and 1100 nm in 0 M urea.

6.4.3 TEM

One of the essential features of keratin protein materials is their intrinsic capacity for self-assembly. Retaining this important biological feature provides the ability for generation of materials such as films[40], hydrogels[38], and sponges[45]. Self-assembly of keratin IFs is a well-studied process.[28, 69] During IF assembly a dimer composed of one type I (acidic) and one type II (basic) keratin is formed. Following dimer formation, tetramers form through antiparallel alignment of two heterodimers to create a staggered conformation. Subsequent parallel head to tail stacking of tetramers results in protofilaments, which further assemble to form 10 nm IFs.[28, 70]

In order for recombinant keratin proteins to be viable for use in biomaterials, they too must possess the ability to self-assemble into fibers after expression and purification. Results from solution characterization point to the formation of higher order structures in both the recombinant and extracted samples.

Extracted and recombinant keratins were prepared using the dialysis protocol described in materials and methods. Samples, in buffer without urea and with a mild reducing agent, 5 mM DTT, were prepared for TEM analysis at a concentration of 5 mg/mL. Figures 6.4 A and 6.4 B are representative TEM images of rhK31/K81 nanostructures. The recombinant proteins do self-assemble into standard 10 nm keratin IFs, and further form large bundles through additional IF interactions. These structures are several μm long and an average of 150 nm wide. Figures 6.4 C and 6.4 D are representative TEM images of the extracted keratin nanostructures. At a 5 mg/mL

concentration, the extracted materials readily formed films on the TEM grid (Figure S2). These films appeared featureless and individual fibers or fiber bundles were not observed. Therefore, the extracted samples were diluted to 0.5 mg/mL. This sample concentration resulted in fibers observable by TEM, but the overall number of fibers, in comparison to recombinant sample, was low. Moreover, these fibers do not have the typical IF morphology. The width of fibers was estimated to be between 70 and 100 nm, with lengths of a few μm (Figure 6.4 C and 6.4 D). However, similar to the recombinant sample, some bundling of fibers is apparent (Figure 6.4 C). In conclusion, the recombinant heteropolymer retains the ability to self-assemble into characteristic IF. Additionally, under the described solution conditions extracted keratins have the propensity to self-assemble although the fibers observed differ in morphology from standard keratin IF.

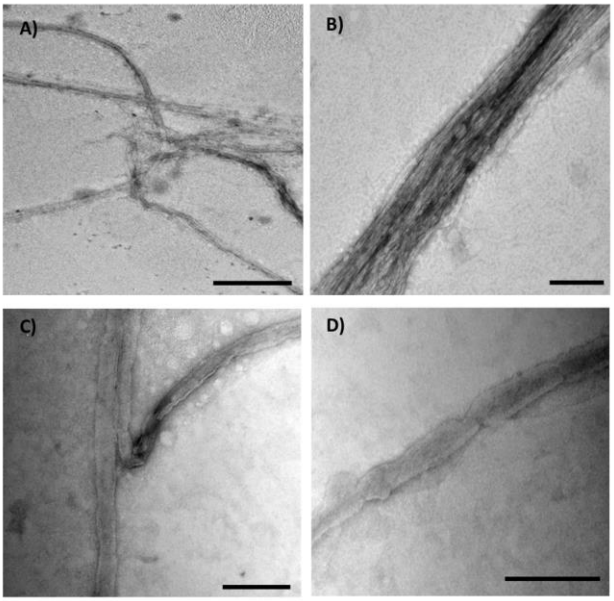


Figure 6.4. TEM images of recombinant (a-b) and extracted (c-d) keratin proteins. Scale bars are (a) 500 nm and (b-d) 300 nm. All images are stained with 2% uranyl acetate.

6.5 Conclusions

Keratin-based biomaterials have been successfully used as scaffolds in regenerative medicine and tissue engineering.[41, 53, 71-72] However, structure-property studies were impeded by the lack of homogenuous and consistent samples. We have shown here that trychocytic keratins, K31 and K81, can be produced recombinantly, resulting in samples of increased homogeneity. Not surprisingly, the composition of extracted versus recombinant materials is shifted towards higher molecular weight oligomers. The distribution of oligomers and their sizes are dependent on the efficiency with which the extensive network of intermolecular disulfide bonds is disrupted in hair fibers, further contributing to the sample heterogeneity.

The recombinant heteropolymer forms standard IF that further associate to create large bundled structures, while extracted keratins fiber did not appear to have the typical IF morphology. Nonetheless, extracted keratins have a remarkable propensity for self-assembly in films that was not observed for recombinant samples under the same conditions. This further reinforces the notion that recombinant and extracted samples have different composition. There are two different scenarios that could justify this observation. Perhaps oligomerization equilibrium in the extracted samples is shifted towards fiber formation that further leads to more efficient entanglement and film formation. Alternatively, it is possible that there is an additional biomolecule in the extracted sample that is facilitating fiber interaction and film formation. The latter observation would be consistent with the small molecular weight species observed in the SDS-PAGE shown in Figure 1B. With recombinant keratin technology in hand, we are now in the position to continue investigations of trichocityc keratin self-assembly.

Design flexibility afforded by recombinant technology will result in multifunctional, dynamic keratin materials. Tailoring of the function, composition, and nanostructure can now be

achieved at the amino acid sequence level. Both extracted and recombinant keratin biopolymers provide a promising tool for engineering novel biomaterials with controlled chemical and physical properties, and future work into the fabrication and characterization of these materials will allow for new advances in regenerative medicine and tissue engineering.

6.6 Supplemental Information

Wool fiber schematic; Extracted keratin film from TEM

6.7 Acknowledgements

The authors would like to thank Virginia Tech, the VT Chemistry Department, and the Department of Biomedical Engineering and Mechanics for funding. The authors are also thankful to the Grove and Van Dyke labs for discussion and careful reading of the manuscript.

6.8 References

1. Goldberg, M.; Langer, R.; Jia, X. Q., Nanostructured materials for applications in drug delivery and tissue engineering. *Journal of Biomaterials Science-Polymer Edition* **2007,** *18* (3), 241-268.
2. Kuratomi, Y.; Nomizu, M.; Tanaka, K.; Ponce, M. L.; Komiyama, S.; Kleinman, H. K.; Yamada, Y., Laminin gamma 1 chain peptide, C-16 (KAFDITYVRLKF), promotes migration, MMP-9 secretion, and pulmonary metastasis of B16-F10 mouse melanoma cells. *Br. J. Cancer* **2002,** *86* (7), 1169-1173.
3. Nectow, A. R.; Marra, K. G.; Kaplan, D. L., Biomaterials for the Development of Peripheral Nerve Guidance Conduits. *Tissue Engineering Part B-Reviews* **2012,** *18* (1), 40-50.
4. Pakulska, M. M.; Ballios, B. G.; Shoichet, M. S., Injectable hydrogels for central nervous system therapy. *Biomedical Materials* **2012,** *7* (2).
5. Park, S. H.; Park, S. R.; Chung, S. I.; Pai, K. S.; Min, B. H., Tissue-engineered cartilage using fibrin/hyaluronan composite gel and its in vivo implantation. *Artif. Organs* **2005,** *29* (10), 838-845.
6. Rouse, J. G.; Van Dyke, M. E., A Review of Keratin-Based Biomaterials for Biomedical Applications. *Materials* **2010,** *3* (2), 999-1014.
7. Schleicher, I.; Parker, A.; Leavesley, D.; Crawford, R.; Upton, Z.; Xiao, Y., Surface modification by complexes of vitronectin and growth factors for serum-free culture of human osteoblasts. *Tissue Eng.* **2005,** *11* (11-12), 1688-1698.
8. Silva, S. S.; Mano, J. F.; Reis, R. L., Potential applications of natural origin polymer-based systems in soft tissue regeneration. *Crit. Rev. Biotechnol.* **2010,** *30* (3), 200-221.
9. Stoppel, W. L.; Ghezzi, C. E.; McNamara, S. L.; Black, L. D.; Kaplan, D. L., Clinical Applications of Naturally Derived Biopolymer-Based Scaffolds for Regenerative Medicine. *Ann. Biomed. Eng.* **2015,** *43* (3), 657-680.

10. Vepari, C.; Kaplan, D. L., Silk as a biomaterial. *Prog. Polym. Sci.* **2007,** *32* (8-9), 991-1007.
11. Wittmer, C. R.; Phelps, J. A.; Saltzman, W. M.; Van Tassel, P. R., Fibronectin terminated multilayer films: Protein adsorption and cell attachment studies. *Biomaterials* **2007,** *28* (5), 851-860.
12. von der Mark, K.; Park, J.; Bauer, S.; Schmuki, P., Nanoscale engineering of biomimetic surfaces: cues from the extracellular matrix. *Cell Tissue Res.* **2010,** *339* (1), 131-153.
13. Anderson, J. M.; Rodriguez, A.; Chang, D. T., Foreign body reaction to biomaterials. *Semin. Immunol.* **2008,** *20* (2), 86-100.
14. Brinckmann, J., Collagens at a glance. In *Collagen: Primer in Structure, Processing and Assembly*, Brinckmann, J.; Notbohm, H.; Muller, P. K., Eds. 2005; Vol. 247, pp 1-6.
15. McKeownLongo, P. J.; Panetti, T. S., Structure and function of vitronectin. *Trends in Glycoscience and Glycotechnology* **1996,** *8* (43), 327-340.
16. Pankov, R.; Yamada, K. M., Fibronectin at a glance. *J. Cell Sci.* **2002,** *115* (20), 3861-3863.
17. Parenteau-Bareil, R.; Gauvin, R.; Berthod, F., Collagen-Based Biomaterials for Tissue Engineering Applications. *Materials* **2010,** *3* (3), 1863-1887.
18. Vrhovski, B.; Weiss, A. S., Biochemistry of tropoelastin. *Eur. J. Biochem.* **1998,** *258* (1), 1-18.
19. Callahan, L. A. S.; Xie, S. B.; Barker, I. A.; Zheng, J. K.; Reneker, D. H.; Dove, A. P.; Becker, M. L., Directed differentiation and neurite extension of mouse embryonic stem cell on aligned poly(lactide) nanofibers functionalized with YIGSR peptide. *Biomaterials* **2013,** *34* (36), 9089-9095.
20. de Luca, A. C.; Faroni, A.; Downes, S.; Terenghi, G., Differentiated adipose-derived stem cells act synergistically with RGD-modified surfaces to improve neurite outgrowth in a co-culture model. *J. Tissue Eng. Regen. Med.* **2016,** *10* (8), 647-655.
21. Lutolf, M. P.; Hubbell, J. A., Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. *Nat. Biotechnol.* **2005,** *23* (1), 47-55.
22. Stevens, M. M.; George, J. H., Exploring and engineering the cell surface interface. *Science* **2005,** *310* (5751), 1135-1138.
23. Williams, D. F., On the nature of biomaterials. *Biomaterials* **2009,** *30* (30), 5897-5909.
24. Gomes, S.; Leonor, I. B.; Mano, J. F.; Reis, R. L.; Kaplan, D. L., Natural and genetically engineered proteins for tissue engineering. *Prog. Polym. Sci.* **2012,** *37* (1), 1-17.
25. Engler, A. J.; Sen, S.; Sweeney, H. L.; Discher, D. E., Matrix elasticity directs stem cell lineage specification. *Cell* **2006,** *126* (4), 677-689.
26. Kohn, J.; Welsh, W. J.; Knight, D., A new approach to the rationale discovery of polymeric biomaterials. *Biomaterials* **2007,** *28* (29), 4171-4177.
27. Almine, J. F.; Bax, D. V.; Mithieux, S. M.; Nivison-Smith, L.; Rnjak, J.; Waterhouse, A.; Wise, S. G.; Weiss, A. S., Elastin-based materials. *Chem. Soc. Rev.* **2010,** *39* (9), 3371-3379.
28. Coulombe, P. A.; Fuchs, E., ELUCIDATING THE EARLY STAGES OF KERATIN FILAMENT ASSEMBLY. *J. Cell Biol.* **1990,** *111* (1), 153-169.
29. Altman, G. H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R. L.; Chen, J. S.; Lu, H.; Richmond, J.; Kaplan, D. L., Silk-based biomaterials. *Biomaterials* **2003,** *24* (3), 401-416.
30. Tachibana, A.; Furuta, Y.; Takeshima, H.; Tanabe, T.; Yamauchi, K., Fabrication of wool keratin sponge scaffolds for long-term cell cultivation. *J. Biotechnol.* **2002,** *93* (2), 165-170.

31. Lu, Q.; Hu, X.; Wang, X. Q.; Kluge, J. A.; Lu, S. Z.; Cebe, P.; Kaplan, D. L., Water-insoluble silk films with silk I structure. *Acta Biomater.* **2010,** *6* (4), 1380-1387.
32. An, B.; Lin, Y. S.; Brodsky, B., Collagen interactions: Drug design and delivery. *Adv. Drug Del. Rev.* **2016,** *97*, 69-84.
33. Fearing, B. V.; Van Dyke, M. E., In vitro response of macrophage polarization to a keratin biomaterial. *Acta Biomater.* **2014,** *10* (7), 3136-3144.
34. Vasconcelos, A.; Freddi, G.; Cavaco-Paulo, A., Biodegradable materials based on silk fibroin and keratin. *Biomacromolecules* **2008,** *9* (4), 1299-1305.
35. Belcarz, A.; Ginalska, G.; Zalewska, J.; Rzeski, W.; Slosarczyk, A.; Kowalczulk, D.; Godlewski, P.; Niedzwiadek, J., Covalent Coating of Hydroxyapatite by Keratin Stabilizes Gentamicin Release. *Journal of Biomedical Materials Research Part B-Applied Biomaterials* **2009,** *89B* (1), 102-113.
36. Dias, G. J.; Mahoney, P.; Swain, M.; Kelly, R. J.; Smith, R. A.; Ali, M. A., Keratin-hydroxyapatite composites: Biocompatibility, osseointegration, and physical properties in an ovine model. *Journal of Biomedical Materials Research Part A* **2010,** *95A* (4), 1084-1095.
37. Dias, G. J.; Peplow, P. V.; McLaughlin, A.; Teixeira, F.; Kelly, R. J., Biocompatibility and osseointegration of reconstituted keratin in an ovine model. *Journal of Biomedical Materials Research Part A* **2010,** *92A* (2), 513-520.
38. Hill, P. S.; Apel, P. J.; Barnwell, J.; Smith, T.; Koman, L. A.; Atala, A.; Van Dyke, M., Repair of Peripheral Nerve Defects in Rabbits Using Keratin Hydrogel Scaffolds. *Tissue Engineering Part A* **2011,** *17* (11-12), 1499-1505.
39. Lee, K. Y.; Kong, S. J.; Park, W. H.; Ha, W. S.; Kwon, I. C., Effect of surface properties on the antithrombogenicity of silk fibroin/S-carboxymethyl kerateine blend films. *Journal of Biomaterials Science-Polymer Edition* **1998,** *9* (9), 905-914.
40. Reichl, S., Films based on human hair keratin as substrates for cell culture and tissue engineering. *Biomaterials* **2009,** *30* (36), 6854-6866.
41. Reichl, S.; Borrelli, M.; Geerling, G., Keratin films for ocular surface reconstruction. *Biomaterials* **2011,** *32* (13), 3375-3386.
42. Saul, J. M.; Ellenburg, M. D.; de Guzman, R. C.; Van Dyke, M., Keratin hydrogels support the sustained release of bioactive ciprofloxacin. *Journal of Biomedical Materials Research Part A* **2011,** *98A* (4), 544-553.
43. Shen, D. L.; Wang, X. F.; Zhang, L.; Zhao, X. Y.; Li, J. Y.; Cheng, K.; Zhang, J. Y., The amelioration of cardiac dysfunction after myocardial infarction by the injection of keratin biomaterials derived from human hair. *Biomaterials* **2011,** *32* (35), 9290-9299.
44. Sierpinski, P.; Garrett, J.; Ma, J.; Apel, P.; Klorig, D.; Smith, T.; Koman, L. A.; Atala, A.; Van Dyke, M., The use of keratin biomaterials derived from human hair for the promotion of rapid regeneration of peripheral nerves. *Biomaterials* **2008,** *29* (1), 118-128.
45. Tachibana, A.; Kaneko, S.; Tanabe, T.; Yamauchi, K., Rapid fabrication of keratin-hydroxyapatite hybrid sponges toward osteoblast cultivation and differentiation. *Biomaterials* **2005,** *26* (3), 297-302.
46. Thilagar, S.; Jothi, N. A.; Omar, A. R. S.; Kamaruddin, M. Y.; Ganabadi, S., Effect of Keratin-Gelatin and bFGF-Gelatin Composite Film as a Sandwich Layer for Full-Thickness Skin Mesh Graft in Experimental Dogs. *Journal of Biomedical Materials Research Part B-Applied Biomaterials* **2009,** *88B* (1), 12-16.
47. Yamauchi, K.; Maniwa, M.; Mori, T., Cultivation of fibroblast cells on keratin-coated substrata. *Journal of Biomaterials Science-Polymer Edition* **1998,** *9* (3), 259-270.

48. Richter, J. R.; de Guzman, R. C.; Van Dyke, M. E., Mechanisms of hepatocyte attachment to keratin biomaterials. *Biomaterials* **2011,** *32* (30), 7555-7561.
49. Kellner, J. C.; Coulombe, P. A., Keratins and protein synthesis: the plot thickens. *J. Cell Biol.* **2009,** *187* (2), 157-159.
50. Kowalczewski, C. J.; Tombyln, S.; Wasnick, D. C.; Hughes, M. R.; Ellenburg, M. D.; Callahan, M. F.; Smith, T. L.; Van Dyke, M. E.; Burnett, L. R.; Saul, J. M., Reduction of ectopic bone growth in critically-sized rat mandible defects by delivery of rhBMP-2 from kerateine biomaterials. *Biomaterials* **2014,** *35* (10), 3220-3228.
51. Burnett, L. R.; Richter, J. G.; Rahmany, M. B.; Soler, R.; Steen, J. A.; Orlando, G.; Abouswareb, T.; Van Dyke, M. E., Novel keratin (KeraStat (TM)) and polyurethane (Nanosan (R)-Sorb) biomaterials are hemostatic in a porcine lethal extremity hemorrhage model. *J. Biomater. Appl.* **2014,** *28* (6), 869-879.
52. Apel, P. J.; Garrett, J. P.; Sierpinski, P.; Ma, J. J.; Atala, A.; Smith, T. L.; Koman, A.; Van Dyke, M. E., Peripheral Nerve Regeneration Using a Keratin-Based Scaffold: Long-Term Functional and Histological Outcomes in a Mouse Model. *Journal of Hand Surgery-American Volume* **2008,** *33A* (9), 1541-1547.
53. Ham, T. R.; Lee, R. T.; Han, S.; Haque, S.; Vodovotz, Y.; Gu, J.; Burnett, L. R.; Tomblyn, S.; Saul, J. M., Tunable Keratin Hydrogels for Controlled Erosion and Growth Factor Delivery. *Biomacromolecules* **2016,** *17* (1), 225-236.
54. de Guzman, R. C.; Merrill, M. R.; Richter, J. R.; Hamzi, R. I.; Greengauz-Roberts, O. K.; Van Dyke, M. E., Mechanical and biological properties of keratose biomaterials. *Biomaterials* **2011,** *32* (32), 8205-8217.
55. Yu, J. L.; Yu, D. W.; Checkla, D. M.; Freedberg, I. M.; Bertolino, A. P., HUMAN HAIR KERATINS. *J. Invest. Dermatol.* **1993,** *101* (1), S56-S59.
56. Strnad, P.; Usachov, V.; Debes, C.; Grater, F.; Parry, D. A. D.; Omary, M. B., Unique amino acid signatures that are evolutionarily conserved distinguish simple-type, epidermal and hair keratins. *J. Cell Sci.* **2011,** *124* (24), 4221-4232.
57. Mark Van Dyke, M. B. R. Keratin nanomaterials and methods of production. 2015.
58. Dinjaski, N.; Kaplan, D. L., Recombinant protein blends: silk beyond natural design. *Curr. Opin. Biotechnol.* **2016,** *39*, 1-7.
59. Harris, T. I.; Gaztambide, D. A.; Day, B. A.; Brock, C. L.; Ruben, A. L.; Jones, J. A.; Lewis, R. V., Sticky Situation: An Investigation of Robust Aqueous-Based Recombinant Spider Silk Protein Coatings and Adhesives. *Biomacromolecules* **2016,** *17* (11), 3761-3772.
60. Jang, Y.; Champion, J. A., Self-Assembled Materials Made from Functional Recombinant Proteins. *Acc. Chem. Res.* **2016,** *49* (10), 2188-2198.
61. Rodriguez-Cabello, J. C.; Arias, F. J.; Rodrigo, M. A.; Girotti, A., Elastin-like polypeptides in drug delivery. *Adv. Drug Del. Rev.* **2016,** *97*, 85-100.
62. Zhou, M. L.; Qian, Z. G.; Chen, L.; Kaplan, D. L.; Xia, X. X., Rationally Designed Redox-Sensitive Protein Hydrogels with Tunable Mechanical Properties. *Biomacromolecules* **2016,** *17* (11), 3508-3515.
63. Spiess, K.; Lammel, A.; Scheibel, T., Recombinant Spider Silk Proteins for Applications in Biomaterials. *Macromol. Biosci.* **2010,** *10* (9), 998-1007.
64. Teng, W. B.; Cappello, J.; Wu, X. Y., Recombinant Silk-Elastinlike Protein Polymer Displays Elasticity Comparable to Elastin. *Biomacromolecules* **2009,** *10* (11), 3028-3036.
65. An, B.; Kaplan, D. L.; Brodsky, B., Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering. *Frontiers in Chemistry* **2014,** *2*.

66. An, B.; Abbonante, V.; Xu, H. F.; Gavriilidou, D.; Yoshizumi, A.; Bihan, D.; Farndale, R. W.; Kaplan, D. L.; Balduini, A.; Leitinger, B.; Brodsky, B., Recombinant Collagen Engineered to Bind to Discoidin Domain Receptor Functions as a Receptor Inhibitor. *J. Biol. Chem.* **2016,** *291* (9), 4343-4355.
67. Elvin, C. M.; Carr, A. G.; Huson, M. G.; Maxwell, J. M.; Pearson, R. D.; Vuocolo, T.; Liyou, N. E.; Wong, D. C. C.; Merritt, D. J.; Dixon, N. E., Synthesis and properties of crosslinked recombinant pro-resilin. *Nature* **2005,** *437* (7061), 999-1002.
68. Steinert, P. M.; Marekov, L. N.; Fraser, R. D. B.; Parry, D. A. D., KERATIN INTERMEDIATE FILAMENT STRUCTURE - CROSS-LINKING STUDIES YIELD QUANTITATIVE INFORMATION ON MOLECULAR DIMENSIONS AND MECHANISM OF ASSEMBLY. *J. Mol. Biol.* **1993,** *230* (2), 436-452.
69. Steinert, P. M.; Steven, A. C.; Roop, D. R., STRUCTURAL FEATURES OF EPIDERMAL KERATIN FILAMENTS REASSEMBLED INVITRO. *J. Invest. Dermatol.* **1983,** *81* (1), S86-S90.
70. Hatzfeld, M.; Weber, K., The coiled coil of in vitro assembled keratin filaments is a heterodimer of type I and II keratins: use of site-specific mutagenesis and recombinant protein expression. *The Journal of Cell Biology* **1990,** *110* (4), 1199-1210.
71. Tomblyn, S.; Kneller, E. L. P.; Walker, S. J.; Ellenburg, M. D.; Kowalczewski, C. J.; Van Dyke, M.; Burnett, L.; Saul, J. M., Keratin hydrogel carrier system for simultaneous delivery of exogenous growth factors and muscle progenitor cells. *Journal of Biomedical Materials Research Part B-Applied Biomaterials* **2016,** *104* (5), 864-879.
72. Pace, L. A.; Plate, J. F.; Mannava, S.; Barnwell, J. C.; Koman, L. A.; Li, Z. Y.; Smith, T. L.; Van Dyke, M., A Human Hair Keratin Hydrogel Scaffold Enhances Median Nerve Regeneration in Nonhuman Primates: An Electrophysiological and Histological Study. *Tissue Engineering Part A* **2014,** *20* (3-4), 507-517.

6.9 Supplemental Figures

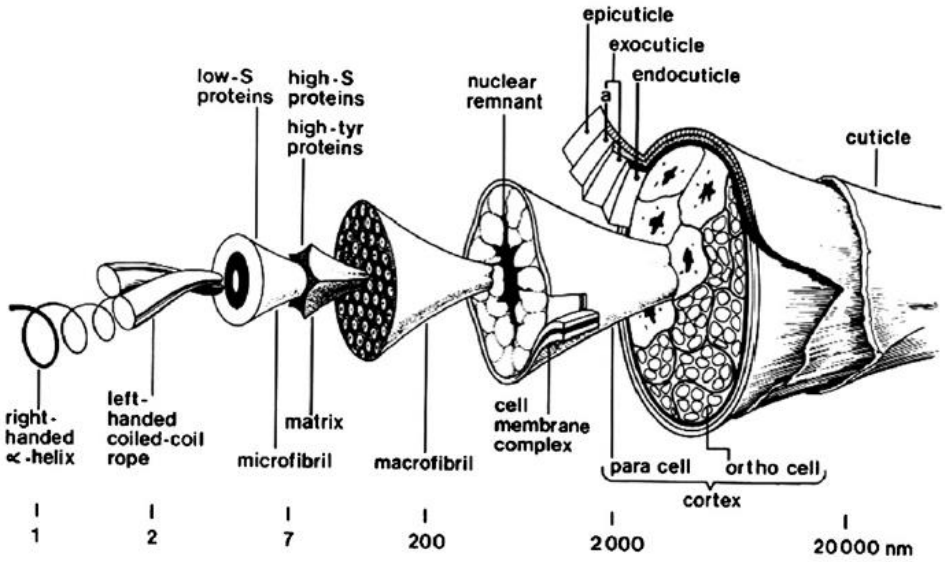


Figure S1. Wool fiber schematic. Copyright CSIRO Australia 1996. Reproduced with permission from *The Lennox Legacy* (Rivett DE, Ward SW, Belkin LM, Ramshaw JAM and Wilshire JFK). Published by CSIRO PUBLISHING, Melbourne Australia

Figure S2. Extracted KTN film (left) formed on TEM grid (right).

Chapter 7. Conclusions and Future Work

7.1 Overall Project Conclusions

Protein engineering offers a powerful tool for designing specific and functional biomedical materials. The two projects described in this dissertation exemplify the utility of recombinant DNA technology combined with various protein design strategies for development of molecular recognition motifs and biomaterials. Our work has demonstrated how natural proteins can be used as the starting point for designing functional materials through further enhancement and tuning of chemical and physical properties to create specific and tailored scaffolds through the use of protein engineering.

7.1.1 Leucine-rich Repeat Proteins for Molecular Recognition

In the first presented project, a consensus based design approach was implemented to engineer leucine-rich repeat (LRR) proteins based on the proposed molecular recognition domain of innate immunity receptors. The resultant protein, CLLR2, exhibited characteristics necessary for successful use as an alternative protein scaffold. First, CLLR2 demonstrated good stability and, as observed from the CD analysis and homology modeling, displayed secondary structure comparable to that of natural LRRs. These results point to the effectiveness of consensus design for creating stable and structurally precise proteins that retain the physical properties of proteins from which they are designed. Furthermore, CLRR2 showed binding affinity to muramyl dipeptide, a known ligand of NOD2. This finding was especially interesting as no affinity maturation techniques were implemented to enhance binding properties of CLRR2, suggesting that biochemical information as well as biophysical properties may have been retained through the consensus design method. In addition to CLRR2 displaying binding affinity for MDP, it also

exhibited selective binding and enantiomeric specificity, as it specifically recognized the biologically active form of MDP. Overall, the results of the design and characterization of LRRs provide a promising scaffold for development of highly selective molecular recognition modules.

7.1.2 Recombinant Human Hair Keratins

Recombinant hair keratins present a promising alternative for biomaterials engineering that may overcome many of the challenges associated with naturally extracted keratin materials. In this work we use protein engineering techniques to produce recombinant human hair keratin 31 and keratin 81. Characterization of their solution phase behavior and self-assembly mechanism, as well as comparisons to extracted keratin proteins, offered favorable results as to the advantages of recombinant keratins. The work described in chapters five and six provide a foundation for the development of recombinant keratin-based scaffolds that will enable specific design and tailoring of chemical and mechanical materials properties.

Through the use of recombinant DNA technology we produced synthetic keratin proteins that were isolated with excellent purity and with no unwanted sequence alterations. Each full length keratin, 31 and 81, can be produced individually, opening up many possibilities for independent modulation of their sequences. As previously noted, extracted materials can only be obtained in their heterooligomeric form, and thus only the properties of this structure can be modified. Additionally, the recombinant keratins retained their inherent self-assembly properties following heterologous expression, purification, and reassembly through dialysis. Interestingly, we determined the ability of K31 to homopolymerize, presenting the possibility for constructing novel materials from only one keratin polymer. Through the characterization and assessment of recombinant and extracted keratin proteins, we conclude that recombinant keratin proteins preserve structural and functional properties necessary for generation of biomaterials.

7.2 Importance of Presented Research and Proposed Future Work

We have described the first functional design of an LRR protein, as well as one of the first reports of LRR binding to a biologically relevant target. Thus, this work has many exciting implications for future research. Previous efforts from other research groups have failed to develop a functional alternative scaffold based on the LRR protein. Additionally, there is a lack of information regarding the mechanisms governing interaction of the natural Nod-like receptors with their targets. As LRRs are the putative binding domain of these proteins, our developed CLLR2 can potentially serve as a model system to further study the role of LRRs in pathogen recognition. Determination of the ligand binding mode of CLRR2 to MDP, as well as other relevant biological targets would offer insight into how LRRs recognize their ligands in nature. This study would also provide useful information for future rational and tailored design of recombinant LRRs for recognition of new targets.

Additionally, there is a great need for molecular recognition agents that can specifically and selectively recognize whole cell pathogens for use in point of care diagnostics, field tests, and clinical applications requiring rapid results. LRRs present an excellent candidate for biosensors due to their inherent function, and we have demonstrated that recombinant LRRs have the potential to be specific and selective molecular recognition agents. Our work lays the foundation for future engineering of CLLR2 towards medically relevant ligands allowing for design of biosensors and diagnostic tools for whole cell pathogen sensing.

In the second research project presented, we described the homopolymerization of K31. This phenomena has not been previously reported, and presents an interesting finding from a fundamental and materials design point of view. These results open up many exciting

possibilities for mechanistic studies of keratin homopolymer self-assembly, as well as expands the breadth of possible scaffolds for biomaterials design through homopolymer-based materials. Additionally, keratin biomaterials hold great promise for clinical applications, and the work described demonstrates the potential for recombinant keratins to overcome the current challenges of keratin materials and improve their properties for desired clinical applications. Consequently, future work will enable the tailored design of materials using recombinant keratin proteins. Engineering of the sequences to incorporate functional cell binding motifs will aid in cellular attachment and addition of degradation sites will facilitate controlled degradation of the material. Each of the research projects described provide the ground work for future fundamental studies and the use of each respective material in biomedical applications.