PROTEIN-BASED MATERIALS IN NEXT GENERATION MEDICINE

A Dissertation

by

REBECCA MICHELLE BOOTH

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Sarah E. Bondos
Committee Members,	Geoffrey Kapler
	Michael McShane
	Jun-Yuan Ji
Intercollegiate Faculty Chair,	David Threadgill

December 2020

Major Subject: Genetics

ABSTRACT

Protein-based materials can combine impressive mechanical properties and biocompatibility with a variety of functions through protein fusions, which make the materials useful in a diverse range of applications. *Drosophila* Ultrabithorax (Ubx) is an intrinsically disordered transcription factor which self-assembles into a variety of materials *in vitro*. The materials are biocompatible and capable of functionalization with proteins of interest, which retain their function while fused to materials. These properties make Ubx materials ideal for use in many fields.

Here I show that Ubx materials can reversibly, noncovalently bind and protect DNAs harboring the Ubx target DNA sequence. Ubx fibers bind DNA sequence specifically, and DNA with multiple binding sites has a slower rate of release from fibers. Ubx fibers have sustained release of DNA capable of transforming cells for up to 60 days.

Ubx materials can also be functionalized with other proteins, such as a biosensing protein. One proof-of-concept sensor system that has been developed is a set of truncated circular permutants of GFP, collectively known as leave-one-out GFP (LOO*n*-GFP), in which the n^{th} β-strand has been 'left out' of the protein. Without this β-strand the GFP is unable to fold properly and is not fluorescent. Rebinding of peptide leads to recovery of fluorescence, making this construct a self-reporting biosensor for the missing piece of its sequence. However, LOO-GFPs have a tendency to oligomerize,

which hampers ligand binding and leads to fluorescence in the unbound state. Ubx offers a transparent support system for LOO-GFPs and has the potential to maintain LOO-GFP stability and binding affinity while preventing LOO-GFP aggregation. Indeed, I found that immobilization on Ubx materials stabilizes LOO-GFP without negatively impacting the chromophore and improves the maximum relative fluorescent intensity of the protein.

This data shows that Ubx materials have promise to be used someday as a vehicle for gene therapy. Ubx materials are particularly suited for use as a DNA delivery scaffold that releases nucleic acids over a long period of time. LOO-GFP-Ubx materials are also promising as biosensors for point of care medical testing or to monitor zoonotic mosquito-borne viruses like dengue fever virus and zika virus.

ACKNOWLEDGEMENTS

I would like to thank my committee chair and mentor, Dr. Sarah Bondos, and my committee members, Dr. Kapler, Dr. McShane, and Dr. Ji, for their guidance and support throughout the course of this research. Thanks go to my former and current lab members, Dr. Kelly Churion, Dr. David Howell, Dr. Shang-Pu Tsai, Gabriela Mendes, Sydney Tippelt, and Amanda Jons, for their training and assistance on my research projects.

Thanks also go to my friends, colleagues, the departmental faculty and staff for making my time at Texas A&M University a great experience. I would like to especially thank the Genetics Program for all of the support and enrichment experiences that have helped me grow as a scientist. Thanks to Dr. Kayla Bayless and her lab members for all their help and support at our joint lab meetings and also Dr. Hays Rye and his lab members for their help with fluorescence spectroscopy experiments. I would like to thank my friends Adelaide Bradicich and Maureen Hayden for all the time they spent studying with me, encouraging me to write my papers and dissertation. My dog Buddy also kept me company during many writing sessions, and helped me get out of the house and stay active.

I would like to thank my family for all the love and encouragement they have given me throughout my time in graduate school. My parents, Ernie and Julie Booth, have always

believed in me and encouraged me to pursue my goals and dreams, and let me vent my frustrations any time I need. Grant and Emily Wilde were my best friends and always supported by spending time with me and feeding me delicious meals. I will be forever grateful that they lived in the same town as me for the majority of my time in grad school. Brian and Kimberly Booth have also been a great help to me; my weekend getaways to visit them always leave me feeling happy and refreshed when I get back. Finally, I would like to thank my nephews and niece Tommy, Zachary, and Holly. Getting to see their adorable pictures and videos never failed to put a smile on my face, regardless of how busy or stressed out I was.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Sarah Bondos, Geoffrey Kapler, and Jun-yuang Ji of the Department of Molecular and Cellular Medicine and Professor Mike McShane of the Department of Biomedical Engineering.

Chapter 2 was co-authored with former graduate student Dr. Kelly Churion. Data concerning sequence specific binding of DNA by Ubx materials was collected by Dr. Churion, while data related to extended release of DNA was collected by me. My collaborator, Dr. Chris Bystroff of Rensselaer Polytechnic Institute, designed the construct and provided the DNA to start the research in Chapter 3, and regularly offered feedback on experiments and results. The data in Chapter 3 was collected by myself, with assistance in the final months of data collection by Amanda Jons, the new graduate student I trained to continue the project. I obtained assistance with equipment for fluorescence spectroscopy data in Chapter 3 from Dr. Lauren Kustigan and Andrew Roth from Dr. Hays Rye's laboratory.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by the Weizmann—Texas A&M University Collaborative Program and a summer fellowship from the Interdisciplinary Program in Genetics. This work was also made possible by the National Institute of Health and the National Science Foundation, under Grant Numbers GM099827-06 and 1151394 respectively. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of these institutions.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

GENERAL INTRODUCTION AND LITERATURE REVIEW*

1.1 Introduction

Protein-based biomaterials are a next generation technology with uses in diverse applications, such as drug delivery, biosensing, tissue engineering, photonics, and defense (Howell et al. 2016; Jo, Gao, & Tabata 2019). The diverse range of applications possible is due to the wide variety of structural, mechanical, and functional properties possible with biomaterials. The focus of this dissertation is on Ultrabithorax (Ubx) biomaterials. Ubx materials are non-immunogenic, biodegradable, biocompatible, and have stabilizing effects on incorporated proteins (Patterson et al. 2014; Patterson et al. 2015; Hsiao et al. 2016; Huang et al. 2011; Tsai et al. 2015). These unique properties are needed for many applications. In this dissertation, I explore using Ubx materials in the fields of DNA delivery and biosensing, with focus on the capability of Ubx materials to bind and deliver DNA and the ability of the materials to stabilize a split protein biosensor.

This literature review will focus on protein-based materials and their uses in next-generation medicines and biosensing. The first section of this chapter will examine the dissertation, its molecular function of nucleic acid binding, and its formation of materials

*Part of Section 1.3 is reprinted with permission from "Generating novel materials using the intrinsically disordered protein Ubx" by Mendes GG, Booth RM, Pattison DL, Alvarez AJ, and Bondos SE, 2018. *Methods in Enzymology*, 611, 583-605, Copyright 2018 by Elsevier Inc.

through intermolecular protein-protein interactions. The second section of the literature review will discuss the advantages of protein-based materials and some examples of these materials and their uses. Finally, we will investigate the use of other protein materials in DNA delivery and biosensing.

1.2 Hox Proteins

Hox transcription factors are fundamental to the processes of development, wound repair, and carcinogenesis. Although Hox proteins were first discovered in *Drosophila melanogaster* (Lewis 1978), homologous Hox genes are found in all bilateral animals (Prince 2002). An ancestral invertebrate Hox gene was likely duplicated to create paralogs within organisms, that have developed different functions over time (Prince 2002). Hox proteins contain the homeodomain (HD) sequence, a 60 amino acid region which comprises the DNA binding HD motif (Scott, Tamkun, & Hartzell 1989). This motif is conserved between organisms, although the sequence outside the HD can vary widely (Hoey & Levine 1988; Kissinger et al. 1990). The HD binds DNA with high affinity, but low specificity. The disordered characteristic of some regions outside the HD is conserved, but the sequence within them is variable (Liu et al. 2008, Galant & Carroll 2002).

Individual Hox protein expression is responsible for development of organs and appendages within an organism (Figure 1-1) (Hughes & Kaughman 2002). Hox proteins are expressed in stripes along the anterior-posterior of an organism and regulate

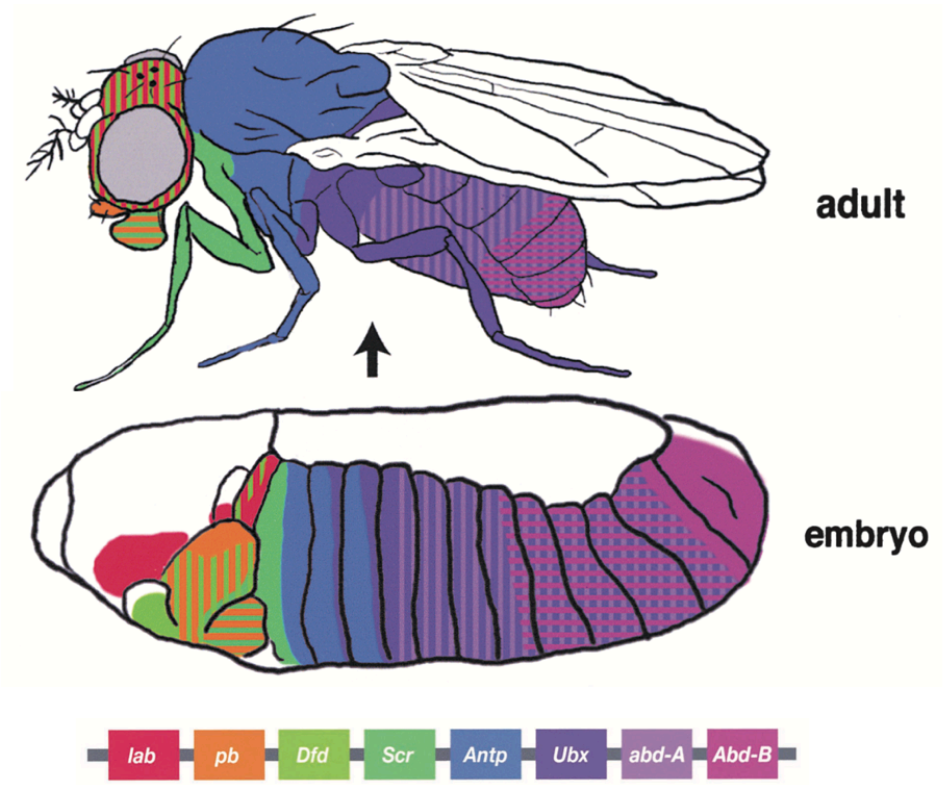

Figure 1-1 Hox genes in Drosophila melanogaster development.
Area of expression in embryo and adult life stages are shown. Image reprinted with permission from Hughes CL and Kaufman TC. Hox genes and the evolution of the arthropod body plan. Evolution & Development (6): 459-99, Copyright 2002 by Wiley Periodicals, Inc.

differentiation and patterning in a region-specific manner (Mace et al. 2005). Hox proteins are responsible for the development of organs and appendages. Misexpression and mutation of Hox proteins during embryo development is linked to birth defects in

humans such as synpolydactyly (Brison, Tylzanowski, & Debeer 2012) and hand-foot-genital syndrome (Mortlock & Innis 1997). Mutation and deregulation of Hox expression in adulthood is associated with increased cancer risk, lower survivability, and tumorigenesis (Li, Huang, & Wei 2019; Bondos, Mendes, & Jons 2020).

1.2.1 Intrinsic Disorder

Hox proteins are disordered, which has made them difficult to study experimentally. Disordered proteins present unique problems when analyzed by many common structural studies, such as missing electron density in X-ray crystallography (Dunker et al. 2001). Structured proteins are dynamic only in the motions of side chains and flexible loops. In lieu of a defined three dimensional structure, some proteins are so flexible that they adopt an ensemble of conformations (Habchi et al. 2014). These are referred to as "intrinsically disordered" proteins (IDPs) or regions (IDRs), and the variability in positioning of these regions leads to the missing electron density mentioned above.

Disordered proteins are present in all organisms, with a general increase in disorder as complexity increases. Eukaryotes have the highest predicted percentage of their proteome disordered (35-51%), relative to archaea (9-37%) or bacteria (6-33%) (Dunker 2002; Dunker et al. 2008; Habchi et al. 2014). Like most transcription factors, Hox proteins contain large intrinsically disordered regions. Proteins that function in signaling and regulation, such as differentiation, transcription, cell cycle, and protein transport are more likely to contain IDRs, while proteins with functions in metabolism and

biosynthesis, such as enzymes, are less likely to be disordered (Habchi et al. 2014; Dunker et al. 2008).

The amino acids composition of disordered proteins typically contains fewer order-promoting amino acids (C, F, I, L, N, V, W, and Y) and more disorder-promoting amino acids (A, E, G, K, P, R, S, and Q) (Dunker et al. 2001). Many IDPs are polyampholytes, meaning they include positively and negatively charged amino acids (Das & Pappu 2013). Strong polyampholytes can form coil-like structures. If the opposite charges are segregated, conformations caused by long-range electrostatic attractions, like hairpins, form. Post-translational modification can alter the regional charge and thus the structure of the region (Das & Pappu 2013). Other disordered proteins, including elastomeric proteins that form materials, are composed of fewer charged residues than structured proteins.

1.2.2 Ultrabithorax Features and Interactions

We use Ultrabithorax (Ubx) as a model protein. This is one of the most studied Hox proteins and is expressed in thorax segments 2 and 3, and abdominal segments 1-8 of *Drosophila melanogaster* (Hughes & Kaufmann 2002) (Beachy et al.1988) (Figure 1-1). Ubx is responsible for formation of posterior legs and halteres and portions of the gut, heart, ectoderm, musculature, and central and peripheral nervous systems (Hughes & Kaufmann 2002).

Six isoforms of Ubx are made *in vivo.* Isoforms are termed “a” or “b” if they lack (a) or include (b) a nine amino acid segment at the end of the first microexon known as the “B” element (Figure 1-2). Numbers (I-IV) refer whether microexons mI and mII are included. Each Ubx isoform is expressed in a stage and tissue specific manner and are not interchangeable during the process of development (Lopez et al. 1996; Reed et al. 2010). The alternatively spliced region is located between the conserved YPWM motif and the homeodomain. The YPWM motif alters affinity and sequence selectivity of DNA binding. Alternative splicing in this region also alters affinity and sequence selectivity (Liu et al. 2009). The alternatively spliced region also plays a role in interactions with partner proteins, causing partner binding to vary between Ubx isoforms (Hsiao et al. 2014).

Ubx has the structured DNA binding homeodomain (HD) common to all Hox proteins (Figure 1-3), as discussed above, which is composed of three alpha helices. Helices 1 and 2 pack in an antiparallel arrangement, and lie perpendicular to helix 3, the recognition helix, which lies within the major groove of DNA (Figure 1-3) (Kissinger et al. 1990), forming contacts with the bases and phosphate backbone. Other structured regions of Ubx include a coiled coil outside the HD (Liu et al. 2008) and the YPWM motif forms a reverse turn and interacts with Extradenticle (Passner et al. 1999) by inserting in a hydrophobic pocket of the Exd HD. Residues Il332, Gln335, Asn336, and Met339, from the third alpha helix, lay in major groove and form contacts along the DNA (Passner et al. 1999). The N-terminal arm of the HD inserts into the minor groove

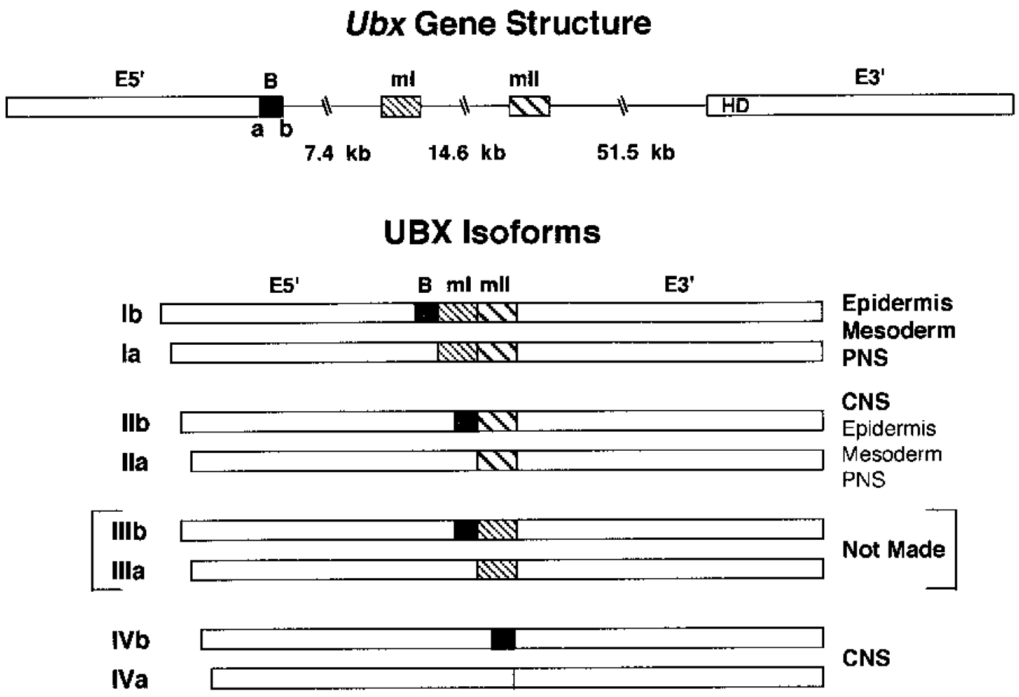

Figure 1-2 Ubx isoforms.
All possible isoforms are shown, along with the tissue they are expressed in. Not all are produced in *Drosophila*. Figure reprinted with permission from Hatton AR, Subramaniam V, and Lopez AJ. Generation of Alternative Ultrabithorax Isoforms and Stepwise Removal of a Large Intron by Resplicing at Exon-Exon Junctions. Molecular Cell 2: 787-796. Copyright 1998 by Elsevier Inc.

for specificity (Kissinger et al. 1990; Joshi et al. 2007). Many of the areas outside the HD are disordered (Figure 1-4A), and in some regions the sequence is evolutionarily conserved and predicted to be involved in protein interactions (Howell et al. 2015; Liu et al. 2008). There are also regions where the disordered nature but not the sequence are conserved.

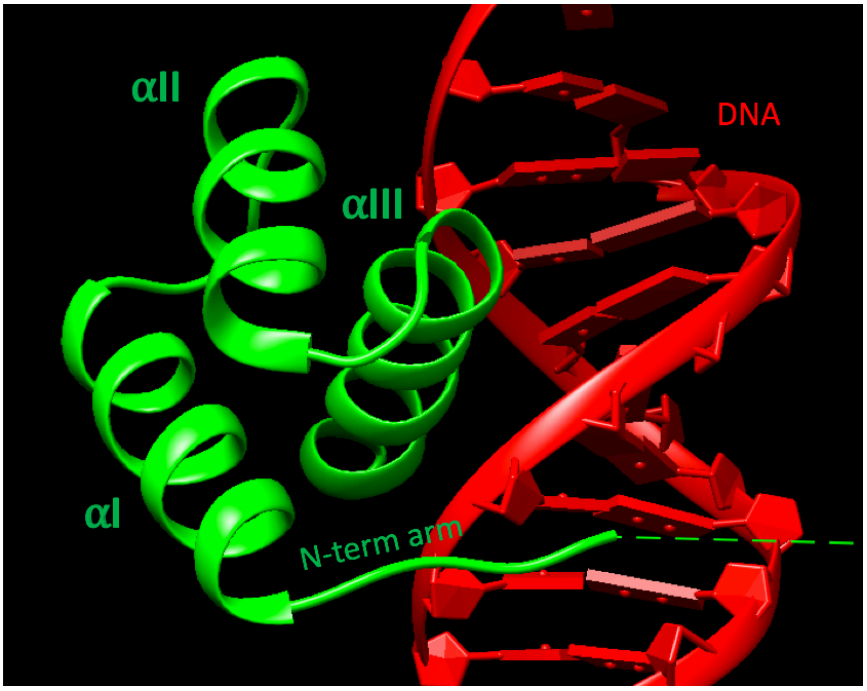

Figure 1-3 The Ubx homeodomain.
Structure obtained from PDB 1B8I. The homeodomain (green) contacts DNA (red). The third alpha helix lies in the major groove of DNA and forms contacts with the bases and phosphate backbone, while the N-terminal arm inserts in the minor groove of DNA.

The function of Ubx is controlled by the regions within the protein. The region aa 159-242 is essential for transcription activation (Figure 1-4 B). This activation domain is predicted to have a β-sheet and an α-helix. The α-helix is required for transcriptional activation, although it alone is not sufficient (Tan et al. 2002). A transcriptional repression domain, known as the QA domain, is located in the carboxy-terminal region of Ubx, which contains 24 amino acids, including a QAQAQK peptide motif and a poly-

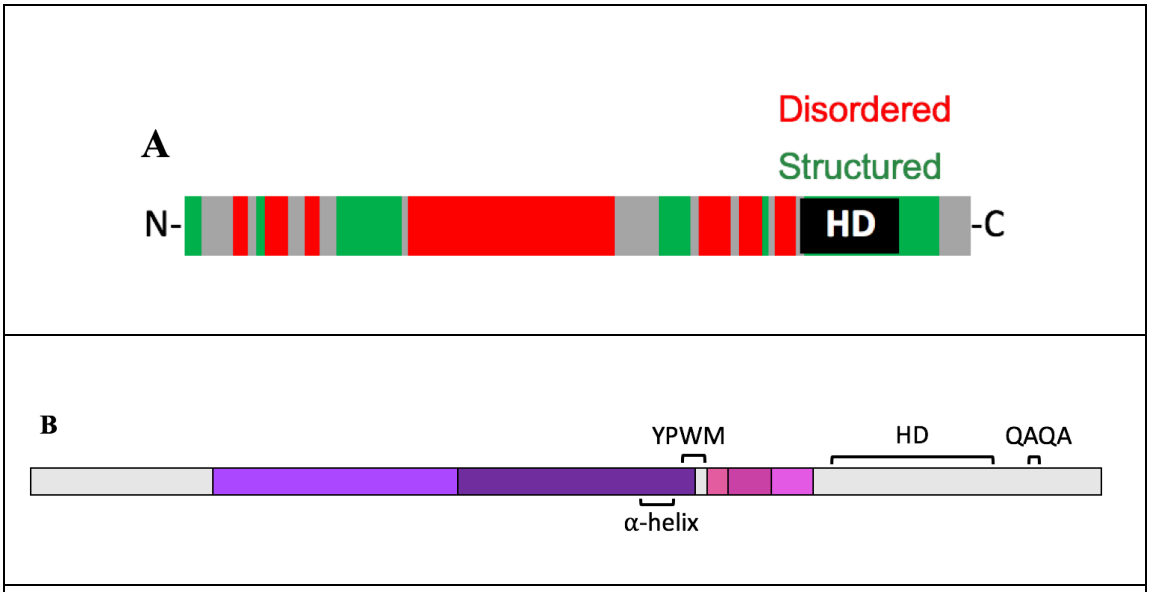

Figure 1-4 Ubx schematic.
(A) Regions of Ubx outside the homeodomain (black) that are predicted to be disordered and structured are marked in red and green, respectively. (B) Ubx sequence schematic, showing the homeodomain (HD). The Ubx mRNA is alternatively spliced, which alters the length of the amino acid sequence between the YPWM motif and the HD. Amino acid sequences included or excluded by alternative splicing are named b, mI, and mII, and indicated by boxes colored pink, dark pink, and light pink, respectively. The transcription activation core (purple box) and the transcription activation enhancing region (light purple box) are indicated. The YPWM motif is used in some contexts to bind the Hox cofactor Extradenticle (Johnson et al, 1995) and inhibits DNA binding (Liu et al, 2008). Finally, the QAQA motif confers the ability to repression transcription (Galant and Carroll 2002; Ronshaugen et al, 2002) Figure 1-4 B reprinted with permission from Liu et al. 2018.

alanine stretch (Galant & Carroll 2002). The region between the YPWM motif and the HD, which includes the microexons, inhibits DNA binding (Liu et al. 2008). The intrinsically disordered N-terminal region restores binding affinity of the HD. (Liu et al. 2008). Unlike HD-DNA binding, transcription regulation of Hox proteins has evolved.

Both the sequence and location of activation and repression domains in Ubx orthologues has changed over time (Liu et al. 2018).

Ubx isoforms are phosphorylated at serine and threonine residues but not tyrosine residues (Gavis & Hogness 1991). The phosphorylated residues are threonine 170, serine 174, and serine 177 (Zhai et al. 2008). A region or regions near the C-terminus might interact with the phosphorylated N-constant region to make a phosphorylation dependent conformational change. The Ubx ortholog class of proteins have evolved an increased ability to repress *Distal-less* through loss of serine and threonine residues in casein kinase 2 (CK2) sites. The presence of CK2 sites C-terminal to the homeodomain inhibits repression functions of Ubx proteins (Taghli-Lamallem et al. 2008).

In addition to its *in vivo* function as a transcription factor, Ubx also self assembles into biomaterials through intermolecular dityrosine bonds (Howell et al. 2015). Our lab has characterized the materials and explored potential uses in a variety of areas. My dissertation focuses on applications of Ubx materials. The next section of this review will introduce protein-based materials.

1.3 Protein-Based Materials

A major goal in materials engineering is the development of functionalized materials for use in diverse applications, including tissue engineering, biosensing, drug delivery, and photonics (Costa et al. 2018; Elsharkawy et al. 2018; Howell et al. 2016; Wei et al.

2017; Wolinsky, Colson, & Grinstaff 2012). Realizing the potential of this exciting field requires the development of novel materials with a broad range of structural, mechanical, and functional properties optimized for each application. The inspiration for such materials is often drawn from biology, including the concept of hierarchical self-assembly of biopolymers to generate the materials' structure. Furthermore, embedding active biomolecules—often proteins—that are capable of functioning with unparalleled affinity and specificity has the potential to create bioactive materials (Huang et al. 2011; Nileback et al. 2017; Tsai et al. 2015; Zhou et al. 2015). Proteins offer many advantages in forming the scaffold, or basic structure, in these materials. First, recombinant protein technology allows facile production of monomers with uniform length and composition. Unlike organic polymers, the amino acid sequence of proteins can be easily engineered using standard molecular biology techniques to tune the properties of the resulting materials (Howell et al. 2015; Kim et al. 2014). Proteins are naturally biodegradable, and many protein-based materials are also biocompatible. Protein-based materials can exhibit a wide variety of useful mechanical properties: the tensile strength of spider dragline silk can rival that of structural steel, materials composed of elastin are extremely extensible, and collagen can form stiff three-dimensional gels (Koebley, Vollrath, & Schniepp 2017; Li & Daggett 2002; Wang, Lai, & Yang 2016).

A key advantage of protein-based materials is the capacity to incorporate a specific protein, and thus its function, into materials with simple molecular biology approaches (Girotti et al. 2015; Huang et al. 2011; Jansson et al. 2014). In gene fusion, the DNA

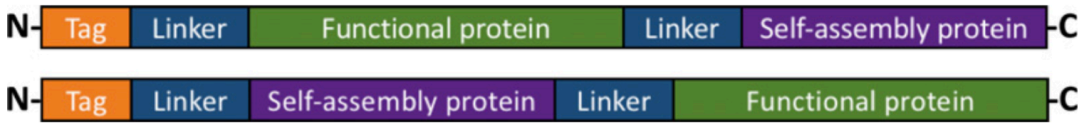

Figure 1-5 Fusion protein schematic.
Primary sequence schematic of fusion proteins used to create functionalized materials, depicting a purification tag (orange), flexible linkers (blue), the functional protein sequence (green), and the self-assembling protein (purple). The functional protein can potentially be attached N-terminal or C-terminal of the self-assembling protein. Figure reprinted with permission from Mendes GG, Booth RM, Pattison DL, Alvarez AJ, and Bondos SE. Generating novel materials using the intrinsically disordered protein Ubx 2018. Methods in Enzymology, 611, 583-605, Copyright 2018 by Elsevier Inc.

sequence encoding the self-assembling protein is attached to the gene encoding the functional protein without intervening stop codons. When this DNA sequence is transformed into a cell, the cell will produce a single polypeptide containing the amino acid sequences of both proteins (Figure 1-5). Ideally, the materials-forming protein will retain the ability to self-assemble, and the activity of the functional protein will be preserved in the resulting materials. This approach has numerous advantages for functionalizing materials. First, the functional protein does not have to be separately produced, reducing cost. Second, the material is uniformly and completely functionalized: every self-assembling protein monomer is attached to a functional protein. Third, only a single attachment site—a peptide bond—is formed between the

functional protein and the self-assembling protein, thus uniformly orienting the functional protein. Fourth, the assembly and functionalization of the materials occur in a single step, eliminating the extra time and cost required for a multicomponent, multistep process. Finally, functional proteins, once embedded in protein materials, are often substantially stabilized, preserving their activity in otherwise hazardous conditions, such as heat, acid, and other denaturing conditions (Boothby et al. 2017; Tsai et al. 2015).

Several factors have prevented the widespread application of functionalization by gene fusion. The size, charge, or quaternary structure of the functional protein may misposition the self-assembling protein, slow assembly kinetics, or even prevent materials assembly (Jansson et al. 2015; Tsai et al. 2015). Conversely, the function of an appended protein can be compromised either by the materials structure, or by harsh environmental conditions (e.g., heat, low pH, organic solvents) often encountered during materials assembly (Falconnet et al. 2006; Hahn, Miller, & West 2005; Huang et al. 2011; Tsai et al. 2015). This challenge has limited many self-assembling proteins to functionalization via peptides, which can still be a powerful tool. However, full length proteins offer many more avenues for functionalization of materials than peptides. A subset of proteins has successfully incorporated single-domain monomeric proteins via gene fusion, such as Enhanced Green Fluorescent Protein-Ultrabithorax (EGFP-Ubx) (Huang et al. 2011) and Z domain-4RepCT (Jansson et al. 2014). Multidomain, multimeric proteins have been incorporated into materials composed of the Hox transcription factor Ubx (Tsai et al. 2015).

1.3.1 Role of Intrinsic Disorder in Protein Materials

The structural characteristics of materials-forming proteins are important, and there are some characteristics that are often found in these proteins, such as disorder. Intrinsically disordered protein sequences can play a major role in the design, formation, properties, and functionalization of protein-based materials. In many proteins that form materials, regions of intrinsic disorder alternate with structured protein regions, mimicking the arrangement of alternating hydrophobic–hydrophilic regions in block copolymers (Rabotyagova, Cebe, & Kaplan 2011; Wesley et al. 2005). Intrinsically disordered proteins with low sequence complexities and multiple short-linear motifs may drive self-association and ultimately formation of phase-separated assemblies (Chavali, Gunnarsson, & Babu 2017; Simon et al. 2017). Sequence enrichment in the structure-breaking amino acids proline and glycine is an integral feature of intrinsically disordered proteins that form materials, while the hydropathy and charge of the polypeptide determine the dependence of self-association on temperature (Quiroz & Chilkoti 2015).

Intrinsically disordered protein regions contribute to the mechanical properties of many protein-based materials, including silk, resilin, and Ubx (Balu et al. 2015; Giesa et al. 2017; Huang et al. 2010). Typically, materials composed of regularly arranged structured proteins are stiff (Young's modulus ~GPa) and less than 20% extensible (Guthold et al. 2007). In contrast, inclusion of intrinsically disordered regions in a protein sequence generates soft, extensible protein-based fibers (Young's modulus ~MPa, ε_{max} >100%) (Guthold et al. 2007). Indeed, the elastomeric properties of many

proteins, including elastin, resilin, Ubx, and mussel adhesive protein, have been attributed to the presence of highly disordered sequences in these proteins (Huang et al. 2010; Qin et al. 2012; Tamburro, Bochicchio, & Pepe 2005). In addition, the disordered regions of silk are required for supercontraction, the shrinkage of silk fibers along the main axis at high humidity (Giesa et al. 2017). Mutation of residues in the disordered regions can prevent or even reverse this behavior. Finally, the correlation of intrinsic disorder with extensibility suggests that these regions are relatively compact when the materials are not under tension. For the self-assembling protein Ubx, we previously hypothesized that these disordered regions can extend to accommodate the presence of a functional protein without sacrificing the interprotein contacts that stabilize the materials (Tsai et al. 2015). In support of this hypothesis, fibers composed of GFP-Ubx fusion proteins were less extensible than fibers composed only of the Ubx protein (Tsai et al. 2015).

A subset of proteins naturally self-assembles into useful materials. Common examples include silks, elastin, mussel adhesive proteins, bacterial S-layer proteins, and virus particles (Bellas et al. 2015; Bidwell et al. 2007; Dinjaski et al. 2017; Kim et al. 2011; Lee et al. 2009; Pum et al. 2013). This molecular toolbox is constantly being expanded by the enhancement of natural protein sequences or the creation of novel proteins that form nanoscale to macro- scale scaffolds (Yang, Holmberg, & Olsen 2017).

1.3.2 Collagen Materials

Collagen is the most prevalent protein in the extracellular matrix, which is a composite material of proteins and carbohydrates that occurs naturally in the body. Collagen provides strength and support to the extracellular matrix, and makes up 30% of the dry weight of the body (Deshmukh et al. 2016). Collagen has a triple helical structure, with three repeating chains containing the $(Gly\text{-}X\text{-}Y)_n$ motif, in which X and Y are most often proline and hydroxyproline (Miranda-Nieves & Chaikof 2017). Formation of hydroxyproline improves stability of materials through formation of hydrogen bonds. There are 29 different types of collagen, each with unique characteristics and functions in different tissue types (Parenteau-Bareil, Gauvin, & Berthod 2010). Fibril forming collagens (types I, II, III, V, and XI) can assemble into fibers, with diameters ranging from 25 nm to 400 nm. Hydrolysis of collagen by heat or chemicals creates gelatin, another common biomaterial (Hu et al. 2012).

Collagen materials are commonly used in biomedical applications, such as gene delivery (Park, Jeong, & Kim 2006; Yamamoto & Tabata 2006; Pannier & Shea 2004; Urello, Kiick, & Sullivan 2014), tissue engineering (Gomes et al. 2012), bone regeneration, stents, and wound healing (Deshmukh et al. 2016), because they are biodegradable and mostly nontoxic and nonimmunogenic to cells. Collagen for these applications is often harvested from bovine skin and tendons, rat tail, porcine skin, fish, jellyfish, and sponges, and recombinant collagen has also been used (Parenteau-Bareil et al. 2010; Miranda-Nieves & Chaikof 2017).

Collagen materials for biomedical applications need to be sterile, but this is not easily achieved. Collagen materials are fragile and temperature-sensitive, can't be autoclaved, and gamma irradiation can alter their structure (Parenteau-Bareil et al. 2010). Animal-derived collagen is the most frequently used; however, it has problems with variability between batches and contamination with pathogens that make it too unreliable for human treatments. Recombinant collagen for polymer synthesis has its own challenges, mainly in achieving the post-translational modification of proline hydroxylation that contributes to the integrity of collagen fibrils (Miranda-Nieves & Chaikof 2017). Collagen is not mechanically stable, which limits its applications (Dinjaski et al. 2017).

1.3.3 Silk Materials

Natural silk comes from spiders and silkworms. It is difficult to harvest spider silk due to difficulties breeding spiders in lab. Natural spider silks are also composed of several different silk proteins (Deptuch & Dams-Kozlowska 2017). Silkworm cocoons are made of two fibroin proteins coated in sericins. Sericins are controversial because there is conflicting data regarding the biosafety of sericin, with some claiming sericins must be removed from materials due to their immunogenicity, while others suggest there is only a mild response (Deptuch & Dams-Kozlowska 2017; Jiao et al. 2018; Siavashani et al. 2020). Silk proteins have repeating domains with high glycine and alanine content. The alanine rich motifs form β-sheet structures, which give the materials their strength, while the glycine rich motifs form amorphous regions in the fiber that provide flexibility and elasticity (Deptuch & Dams-Kozlowska 2017).

Silks have high tensile strength and extensibility, and are biocompatible and biodegradable (Gomes et al. 2012). They can adopt multiple forms, such as films, fibers, mats, hydrogels, and capsules, and can be functionalized by fusing additional peptides or proteins (Saric & Scheibel 2019). These characteristics make them ideal for use in many applications, such as drug and nucleic acid delivery, diagnostics, tissue and bone regeneration (Dinjaski et al. 2017; Koh 2018; Bellas et al. 2015), sensors (Xu, Jiang, Pradhan, & Yadavalli 2019), coatings (Deputch & Dams-Kozlowska 2017), and wearable technology (Blamires, Spicer, & Flanagan 2020).

Bioengineered silks have tunable characteristics that lead to better performance for their defined application than natural silks. However, recombinant silks are difficult to produce. Modifications such as adding functional proteins and tags to silk materials can affect self-assembly into materials and mechanical properties within the final materials (Saric & Scheibel 2019). Silk proteins typically have a high molecular weight and are highly repetitive, which leads to difficulties in recombinant silk production. Silks can be engineered to have the minimal required motifs for materials formation to improve expression, yet scale-up to large production remains challenging (Saric & Scheibel 2019). Many applications of silk proteins are biomedical and require very pure materials, and contamination of the silk with endotoxin or lipopolysaccharide can be a problem (Deptuch & Dams-Kozlowska 2017).

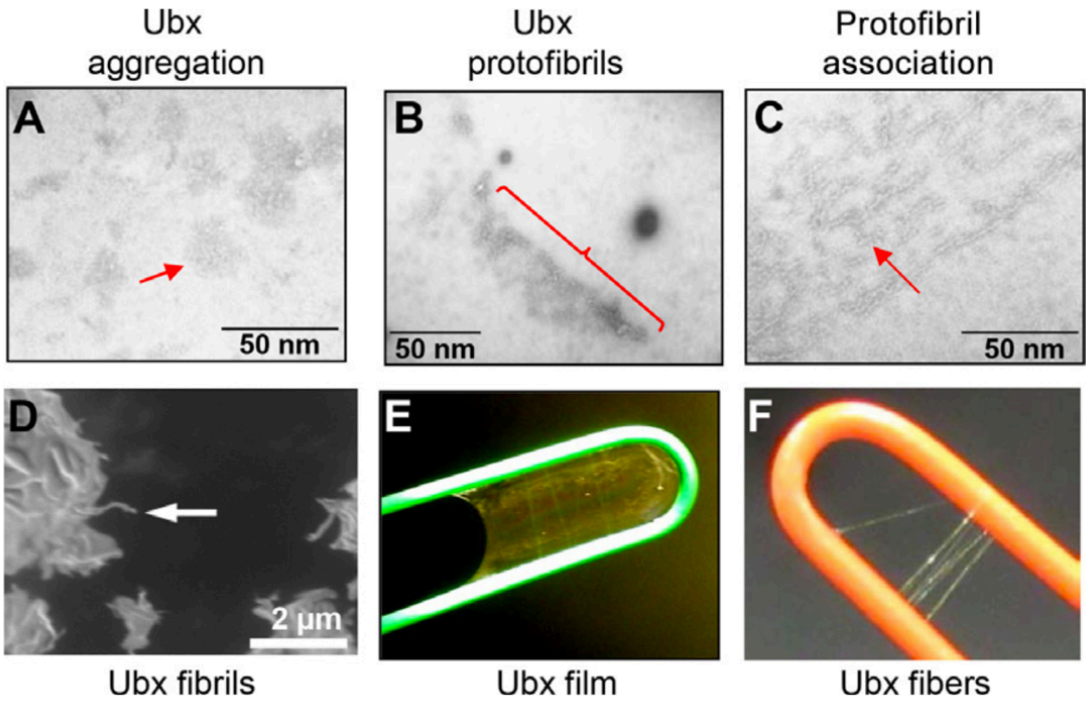

Figure 1-6 Ubx monomers form materials *in vitro*.
(A) About 15 min after dilution into a tray, Ubx begins to form irregular aggregates (red arrow). (B) Within 1h, these aggregates begin to spontaneously reshape into lines of Ubx proteins, termed protofibrils. (C) After 1–2 h, the protofibrils align in parallel and begin to form side-to-side contacts. (D) After 2h multiple protofibrils condense to form fibrils about 50nm in diameter (white arrow). (E) Fibrils also interact laterally to form films which can be lifted from the surface. (F) Alternately, films on the air–water interface can be drawn into fibers composed of aligned fibrils. Figure reprinted with permission from Mendes GG, Booth RM, Pattison DL, Alvarez AJ, and Bondos SE. Generating novel materials using the intrinsically disordered protein Ubx 2018. Methods in Enzymology, 611, 583-605, Copyright 2018 by Elsevier Inc.

1.3.4 Ultrabithorax Materials

Although capable of cooperative DNA binding (Beachy et al. 1993), Ubx does not form materials as part of its natural function. Despite the lack of evolutionary selection for self-assembly, Ubx monomers hierarchically form ordered materials on a variety of length scales in vitro (Figure 1-6). In buffers near neutral pH, at room temperature, and under gentle conditions, Ubx coacervates at the air–water interface, generating disorganized clusters (aggregates) of molecules, with no need for high temperatures of organic chemicals (Greer et al. 2009; Majithia et al. 2011). The disorganized clusters reshape themselves into lines, or protofibrils. Lateral interactions between protofibrils condense them into fibrils, which are similar in diameter to amyloid fibrils but lack amyloid structure. Fibrils interact laterally to form film, and film can be drawn into fibers that are meters long, with diameters ranging from 2 to 50 μm, depending on the thickness of the film. Ubx self-assembles into materials in as little as two hours. Ubx materials are resistant to heat, with Ubx ropes capable of withstanding hours at high temperatures (Greer et al. 2009). Fresh Ubx materials are extremely flexible, with better extensibility than silks and collagen. Upon drying, Ubx materials become brittle, but are remarkably durable, capable of storage at room temperature for years. Newly synthesized materials can adhere to other Ubx materials, as well as surfaces such as plastic, glass, Teflon, and metal (Greer et al. 2009).

The mechanical properties of Ubx fibers are tunable; the extensibility, breaking stress, and breaking strain are linked to fiber diameter. Narrow fibers stretch elastically, and the

breaking strength and Young's modulus decrease as fiber diameter increase (Huang et al. 2010). This control over the properties of the materials make them well suited for multiple applications. The strength of Ubx materials is derived from the rapid and spontaneous formation of dityrosine bonds, which autofluoresce blue within the materials (Howell et al. 2015). Two dityrosine bonds are known to form, one between the N-terminus and the homeodomain (Y4/Y296 or Y12/Y293) and one between Y167 and Y240. Materials can be strengthened or weakened through mutagenesis, increasing or decreasing the number of tyrosines present in Ubx (Howell et al. 2015).

The large intrinsically disordered regions within the Ubx sequence (Figure 1-4) enable the successful incorporation into materials of a wide variety of proteins, including fusions nearly 3 times the size of Ubx, charged proteins, and dimers. Fused proteins retain their function within the material, and are substantially stabilized (Tsai et al. 2015; Huang et al. 2011). mCherry-Ubx fibers retain 80% of fluorescence after a 30 minute ethanol incubation, and retain 20% of fluorescence after sterilization in an autoclave for 20 minutes (Tsai et al. 2015). Production of the fusion protein is most successful when the appended protein is soluble, while materials formation is not significantly impacted by the identity of the appended protein. However, the mechanical properties of the materials can be impacted by the appended protein (Tsai et al. 2015). Ubx materials can be patterned with different fusion proteins, by arranging the placement of monomers as materials assemble, or after they have formed due to their self-adhesive properties, allowing for construction of patterned 3D structures (Huang et al. 2011).

Ubx materials are cytocompatible, biocompatible, and non-immunogenic. There is no toxicity observed when fibers are cultured with umbilical vein aortic smooth muscle cells, endothelial cells, or brain vascular pericytes (Patterson et al. 2014). Cells also readily attach to Ubx fibers. Attachment alters cell morphology, with cells aligning along the long axis of the fiber (Patterson et al. 2014). Mice fail to produce antibodies that recognize Ubx materials, and the materials do not stimulate an inflammatory or immune response (Patterson et al. 2015; Hsiao et al. 2016).

Fusion of vascular endothelial growth factor (VEGF) to Ubx creates bioactive materials capable of influencing cell behavior and directing vascular growth (Howell et al. 2016). Although VEGF is an unstable dimer with poor solubility, it remains active when immobilized on Ubx materials. VEGF-Ubx fibers can enhance migration and prolong cell survival in human endothelial cells, and is capable of inducing blood vessel formation in a chicken embryo chorioallantoic membrane (CAM) assay.

1.4 Applications of Ubx Materials

The unique properties of Ubx materials discussed above make them suitable for use in many exciting, cutting edge technologies. Although there are many possible avenues of function for Ubx materials, here we will focus on DNA delivery and biosensing, two applications of Ubx materials explored in Chapters 2 and 3 of this dissertation, respectively.

1.4.1 Nucleic Acid Therapies

The first technology I will discuss is nucleic acid therapy, which is a promising treatment with some issues that must be overcome for utilization in humans. Nucleic acid therapy is the transfer of genetic material or introduction of exogeneous DNA to a patient in order to treat a disease (Anguela & High 2019). The goal is to have long-term expression of the delivered gene at therapeutic levels. The therapy could be the wild type version of a mutated gene causing a disease, or suppression of expression of a gene using RNA interference, which is useful for cancer (Bondos et al. 2020). The therapy could be delivered *in vivo*, or cells may be removed from the body, treated with the therapy, and given back to the person. In the future, it may be possible to correct mutated human genes using genome editing techniques, though there are many ethical considerations and advances in technology that need to be made.

The idea for gene therapy has been around for almost 5 decades, and the first clinical trials began in the 1990s (Dunbar et al. 2018). However, gene therapy suffered several major setbacks in the late 1990s and early 2000s. Jesse Gelsinger passed away from a negative reaction to the adenovirus used in his therapy for ornithine transcarbamylase deficiency (Zallen 2000; Couzin & Kaiser 2005). Several children in Europe being treated for severe combined immune deficiency developed leukemia due to integration of the viral vector into an oncogene in the genome(Couzin & Kaiser 2005; Zechiedrich & Fogg 2019). These incidents triggered public attention and pushback against gene therapy. The reasons behind these setbacks, a negative reaction to the viral vector and

integration into an unintended area of the genome, provided guidance on what is needed in nucleic acid therapies. Since then, the scientific community has worked to overcome these setbacks and completed more than 2,500 clinical studies, with 6 therapies approved for use since 2012 (Anguela & High 2019). Glybera®, Strimvelis® were approved by the European Medicines Agency (EMA), while IMLYGIC®, KYMRIAH®, YESCARTA®, and LUXTURNA® were approved by the US Food and Drug Administration (FDA).

1.4.1.1 Types of Gene Therapy

Viral vectors are commonly used for gene therapy, because it's an easy mechanism to enter cells and deliver DNA. However, viral vectors have some drawbacks. Lentiviral and retroviral vectors insert the nucleic acid therapy into the genome, and have a high risk of gene disruption (Hardee et al. 2017). Adenoviral vectors are not inserted into the genome, but can cause immunogenicity and toxicity, as the human body is designed to fight viral infections. For all viral vectors, patients may develop immunity if the same serotype is used to deliver treatment to a patient multiple times (Sinn, Burnight, & McCray Jr 2009; Hardee et al. 2017). For viral vectors that do not insert into the genome, treatment length is not long, but the treatments must be done repeatedly.

Non-viral vectors have advantages over viral vectors, such as lower immunogenicity. RNA vectors, such as microRNAs and synthetic messenger RNAs, are easily translated in the cell, but they are much less stable than DNA. Plasmid DNA vectors are more stable and easier to produce than RNA vectors, do not often integrate into the genome,

and can be delivered to patients repeatedly without development of immunity to the treatment. However, plasmids have difficulty transfecting many cell types compared to viral vectors (Gaspar et al. 2014). Minivectors are small constructs of DNA (100-200 bp without an added gene) that can carry a microRNA or a small gene, and transfect cells more readily than larger plasmid DNA (Zechiedrich & Fogg 2019; Hardee et al. 2017).

1.4.1.2 Strategies to Improve Gene Therapy

Controlled release of DNA can improve gene therapy by increasing both the level of expression of the gene and the length of time that the gene is expressed (Pannier & Shea 2004). A slower release rate lengthens the amount of time the therapy is effective, which increases the time before another treatment is needed. Release rate of the therapeutic is regulated by binding affinity of the materials for the DNA vector. A higher binding affinity translates to a slower dissociation rate (K_{off}), and thus prolongs DNA release.

Non-viral vectors have a lower transfection efficiency for many cell types, and delivery vehicles can help overcome this problem (Hardee et al. 2017). DNA vectors are usually bound to a delivery vehicle polymer by nonspecific interactions, such as electrostatic, van der Waals, and hydrophobic interactions (Pannier & Shea 2004). Polymers both protect the nucleic acid therapy and can help to deliver it to its site of action. Protein-based materials have advantages over synthetic polymers because they are biocompatible, sustainable, and biodegradable (Jao et al. 2017). Materials such as

collagen, PLG polymers, ethylene vinyl-co-acetate (EVAc) polymers, and others can release DNA from hours up to 60 days (Pannier & Shea 2004).

Cell targeting is also possible with a delivery vehicle. For example, the RGD sequence (Arginyl-glycyl-aspartic acid) selectively binds to integrins, which are present on the surface of many cell types, such as endothelial cells, macrophages, melanomas, osteoclasrs, and platelets (Saric et al. 2019; Numata et al. 2009). RGD containing peptides are frequently used to control biodistribution of drugs by targeting cells with cell surface integrins (Espin et al. 2012). Addition of this sequence to the delivery vehicle of therapies can improve control over where the therapy is delivered; if the delivery vehicle is bound to integrins on a cell, the delivery vehicle is in closer proximity to the cell of interest when it releases the therapy.

1.4.1.3 Protein-Based Materials and DNA Therapeutics

Collagen materials are frequently used in biomedical applications, as discussed earlier. The materials are good for delivering protein and DNA therapeutics because of their structural and biochemical properties. Multiple collagen scaffolds have been tested as gene-activated scaffolds to deliver DNA for tissue engineering purposes (Mao et al. 2009; Raftery et al. 2016) . Collagen scaffolds are typically combined with gene delivery vectors, such as cationic lipids, to control the release and increase cellular uptake of the DNA. Early tests in mice proved collagen scaffolds loaded with a VEGF gene/N,N,N-trimethyl chitosan chloride (TMC) complex are able to increase VEGF expression and

angiogenesis within the scaffold (Mao et al. 2009). VEGF and other genes for growth factors, such as BMP-2 and PGDF have been investigated for use in bone repair, with some success bridging defects 4 to 8 weeks after treatment with gene-activated collagen scaffolds in canine and rodent models (Raftery et al. 2016; Huang et al. 2015; Endo et al. 2006) . However, most applications for gene-activated collagen scaffolds involve implantation of the scaffold at the site of injury, with treatment only needed for a short period of time. Long-term treatments using other methods of delivery have not been investigated.

Silk is also frequently used for biomedical applications and has been tested as a delivery vehicle for DNA. Silk-based block copolymers with added poly(L-lysine) domains were tested as a vehicle for gene delivery to human embryonic kidney (HEK) cells (Numata et al. 2009). Poly(amino acids)s, like Poly(L-lysine), and cationic polymers can bind DNA with electrostatic interactions and form polyelectrolyte complexes, however these interactions are not specific and release somewhat quickly. Poly(L-lysine)-silk transformed human embryonic kidney cells (HEK), but the transfection efficiency was too low for gene therapy use (Numata et al. 2009). For improved transfection efficiency, they repeated transfection experiments in HEK and HeLa cells with an RGD sequence genetically fused to the silk protein, which helps target it to cells (Numata et al. 2010). Addition of 11 RGD sequences improved the transfection efficiency in cells, but addition of one or two motifs was not enough to reliably improve transfection efficiency. The maximum transfection efficiency compared to transfection with Lipofectamine

(LPS), which is frequently used to facilitate transfection in laboratory experiments, was about 15% in HeLa cells and 10% for HEK cells (Numata et al. 2010). Further improvement in transfection efficiency is needed for Poly(L-lysine)-silk to be useful for gene therapy treatments in humans. In summary, for DNA delivery purposes, silk has a relatively fast rate of release, which is not ideal for extended function of materials in delivering therapies.

Ubx materials may be well suited for use delivering nucleic acid therapies. As a transcription factor, Ubx has an extremely high affinity for DNA (Liu et al. 2008). A tunable, biocompatible biomaterial with high affinity DNA binding has many ideal characteristics for such an application. The high affinity of Ubx for DNA would translate to slow K_{off} for the materials, making Ubx materials suitable for extended release of DNA therapeutics. In addition, protein fusion makes it possible to functionalize the materials further, such as with cell targeting peptides like the RGD motif. In chapter 2, Ubx materials are tested for DNA binding and release.

1.4.2 Protein-Based Sensors

Sensors offer information on the chemical environment by measuring an analyte of interest. The analyte could be physical, chemical, electrical, or biological in nature (Figure 1-7) (Adhikari & Majumada 2004; Bhalla et al. 2016). Analyte is sensed by a recognition element, which produces a signal, such as pH change, light, heat, or voltage, when the analyte is bound. The signal is then converted into a measurable form by a

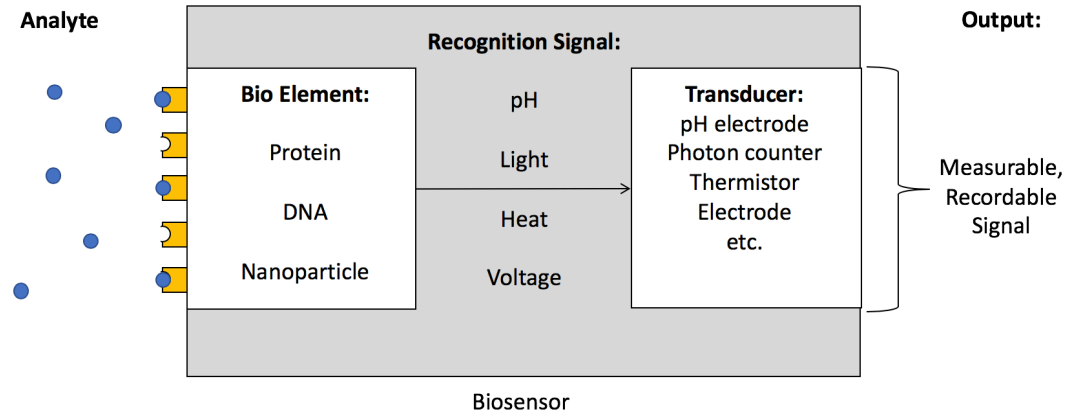

Figure 1-7 Biosensor schematic.
Analyte is bound by a biorecognition element, producing a recognition signal. The signal is turned into a measurable output by a transducer.

transducer. Sensors are versatile and widely used in many fields, such as environmental monitoring, drug development, food safety, agriculture, disease detection, and defense (Adhikari & Majumada 2004; Bhalla et al. 2016). Biomaterials can be used to immobilize sensors, as the recognition elements, or as the transducer.

Protein-based sensors are usually biosensors, with the protein featured as the biorecognition element. Proteins make good biorecognition elements in sensors, because many proteins already bind to specific ligands with high affinity, and proteins can be engineered to bind to novel ligands of interest (Lakowicz 2006). The biorecognition

element can also be based on distance or proximity to another protein of interest, or be sensitive to environmental conditions such as pH and ion concentrations. Sensors are frequently used in areas such as disease and infection monitoring, drug discovery, environmental monitoring for contaminants, and biomedical research.

Many protein-based sensors use fluorescent proteins, which give a convenient readout. Green fluorescent protein (GFP) is a widely used reporter in cell biology and is also the most commonly used split protein. GFP is a β barrel made of 11 β strands. Its chromophore, which gives the protein its fluorescence, lies at the center of the barrel. The fluorescence of the chromophore comes from the post-translational modification of the motif Ser65-Tyr66-Gly67 (Crone et al. 2013). Once folded, GFP is kinetically stable, but it will unfold rapidly at low pH. The chromophore retains its structure after unfolding, whether it be chemical, thermal, or pH-induced. However, fluorescence is quenched when GFP is unfolded due to hula-twist motions in the excited state of the chromophore (Huang et al. 2015).

Several heavy metal protein biosensors have been produced. The first used enhanced green fluorescent protein (EGFP). Metal binding loops (MBL) were added to a loop region of EGFP, preventing fluorescence except when MBLs bind to a metal ion, induce a conformational change, and restore fluorescence (Kim et al. 2019). Another metal detection biosensor was developed that senses mercury ions (Hg^{2+}). The mercury-sensing transcription factor MerR was fused to enhanced yellow fluorescent protein

(EYFP). Binding of Hg^{2+} induced a conformational change that lowered fluorescent signal. Loss of fluorescence was proportional to increasing concentration of Hg^{2+} (Ozyurt et al. 2019). These sensors could be used to test for heavy metals like cadmium and mercury in soil and water samples.

Non-metal ions are also a common target of protein biosensors. For example, a genetically encoded calcium ion (Ca^{2+}) sensor that contains mutated calmodulin and its binding peptide has been made. The sensor uses Förster resonance energy transfer (FRET) between enhanced cyan fluorescent protein (ECFP) and a yellow fluorescent protein variant (YPet) to monitor Ca^{2+} signaling within the endoplasmic reticulum (Kim, Kim, & Jung 2017).

Voltage sensors to detect neural activity, known as genetically encoded voltage indicators (GEVIs), are an emerging technology for recording voltage changes within the cell. One example was developed by fusing mScarlet, a red fluorescent protein, to Ace2N, a voltage-sensitive inhibitory rhodopsin. It uses FRET to detect changes in membrane voltage due to neural activity (Beck & Gong 2019).

Many pH sensors have been made from proteins. Green fluorescent proteins (GFPs) have a phenol group on the chromophore, so most variants are sensitive to pH. One type of GFP has been designed to measure pH within different cellular compartments

(Bizzarri et al. 2009). A sensor called pHRed was developed for intracellular imaging of pH, based on mKeima, a red fluorescent protein (Tantama, Hung, & Yellen 2012).

1.4.2.1 Split Proteins

Split protein assays are very useful and have been developed for many different purposes. Split proteins have been separated into two pieces, and lack the function or activity of the whole protein when separated. When the two pieces of the protein come together, function is restored. Proteins of interest can be fused to the split proteins, making it possible to measure interactions or distance between the proteins of interest by the response from the split proteins coming together. Many of these protein-fragment complementation assays have been designed and used to investigate protein-protein interactions in a wide variety of fields (Romei & Boxer 2019; Nguyen & Silberg 2010). Readouts can be anything from fluorescence or bioluminescence to cell survival and gene transcription (Romei & Boxer 2019.)

Split GFPs separate the barrel into two non-fluorescent fragments, which bind to each other and restore fluorescence. Split GFPs are most commonly split after β strand 7 or 8, which results in fragments with a size ratio of 2:1 (Romei & Boxer 2019)**.** Although they are widely used, split GFP systems commonly encounter two problems. First, the split fragments tend to self-associate (Figure 1-8), which leads to high background and false positives (Kodama & Hu 2012; Lakowicz 2006). Split GFPs also take time to reconstitute fluorescence, as proper folding and chromophore maturation must occur

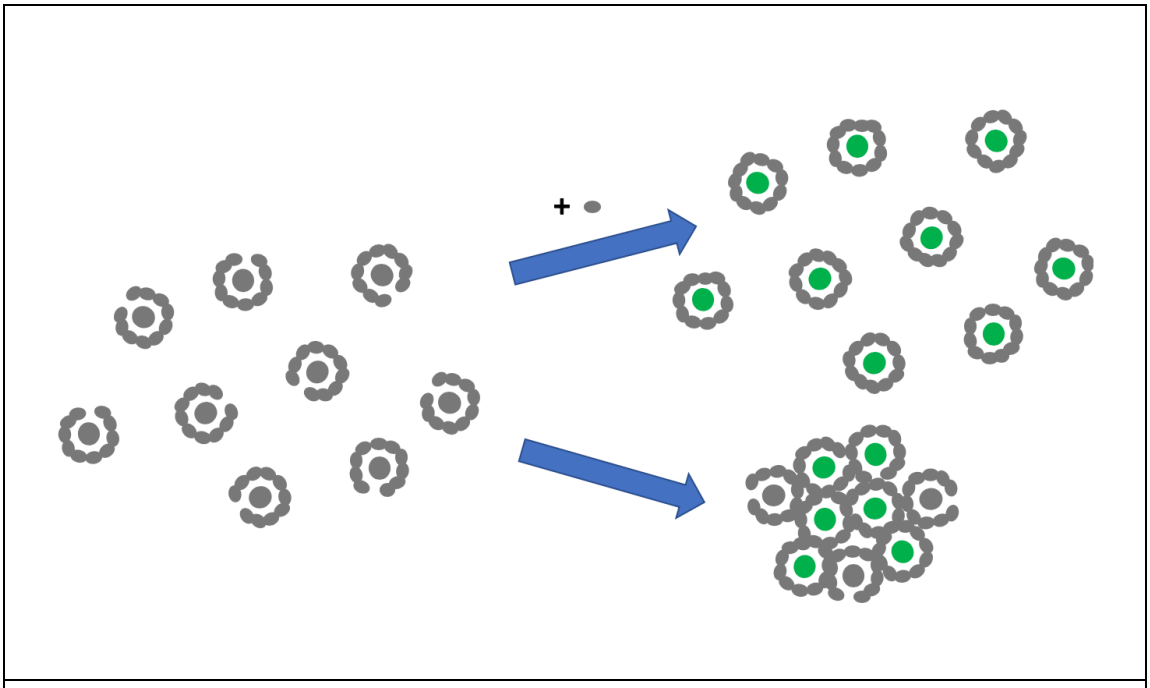

Figure 1-8 Schematic of split-GFP aggregation.
Incomplete split GFPs should bind their left out piece and recover fluorescence (top right). However, incomplete split GFPs can aggregate, causing them to fluoresce even in the unbound state.

after the two fragments bind. Faster folding variants of GFP have been made to overcome this problem (Romei & Boxer 2019).

1.4.2.2 Leave-One-Out-GFP

In leave-one-out GFP (LOO*n*-GFP), a split-GFP system based on a version of GFP known as superfolder GFP OPT that is optimized for higher solubility, the protein sequence was circularly permuted and truncated so one β strand is left out, with *n*

referring to the particular strand (Huang & Bystroff 2009; Huang et al. 2011). Removal of one β strand creates a GFP sensor that fluoresces upon binding the left-out-peptide. Removal of the peptide prevents proper folding, but addition of the left-out strand reconstitutes the structure (Huang & Bystroff 2009). The chromophore maintains its structure during unfolding of the rest of the protein, but loses its fluorescence due to quenching from hula-twist motions (Huang et al. 2015). LOO-GFP can be computationally redesigned to fluoresce upon binding a novel target peptide, such as Hemagglutinin (HA) protein from H5N1 influenza virus (Huang et al. 2015). There is the possibility for development of biosensors for multiple ligands all based on same LOO-GFP design.

LOO-GFP suffers from two of the problems with split-GFPs discussed above, such as self-association and high background signal. As part of my dissertation I will immobilize LOO-GFP biosensors on Ubx materials to overcome these split GFP problems (see Chapter 3).

CHAPTER II

SEQUENCE-SPECIFIC DNA BINDING, PROTECTION, AND EXTENDED RELEASE BY PROTEIN MATERIALS

2.1 Introduction

Nucleic acids have been developed for diverse functions in recent years, with applications in vaccines, gene and immunotherapy, nanodevices, and biosensors (Zechidrich and Fogg 2019; Villalonga 2020; Kong 2020; Dobrovolskia 2019). With many applications designed to be deployed *in vivo,* a mechanism is needed to display DNAs that is biocompatible, non-toxic, and capable of reversible but long-lived binding. Protein materials are capable of reversible binding and display of nucleic acids, but until now have lacked the high affinity needed for extended binding and release. Here we demonstrate that materials composed of the *Drosophila* transcription factor Ultrabithorax (Ubx) reversibly bind and protect DNA. Ubx materials can sustain release of small plasmids to cells for up to 2 months, in which the release rate can be tuned by the number and position of Ubx binding sites in the DNA sequence. These results introduce protein-DNA composite materials that allow construction of stable and tunable complexes with orientation specific binding.

2.2 Results and Discussion

In vivo, the Ubx protein binds the double stranded DNA sequence 5'-TAAT-3' to fulfill its role as a transcription factor (Liu, Matthews, & Bondos 2008; Liu, Matthews, and

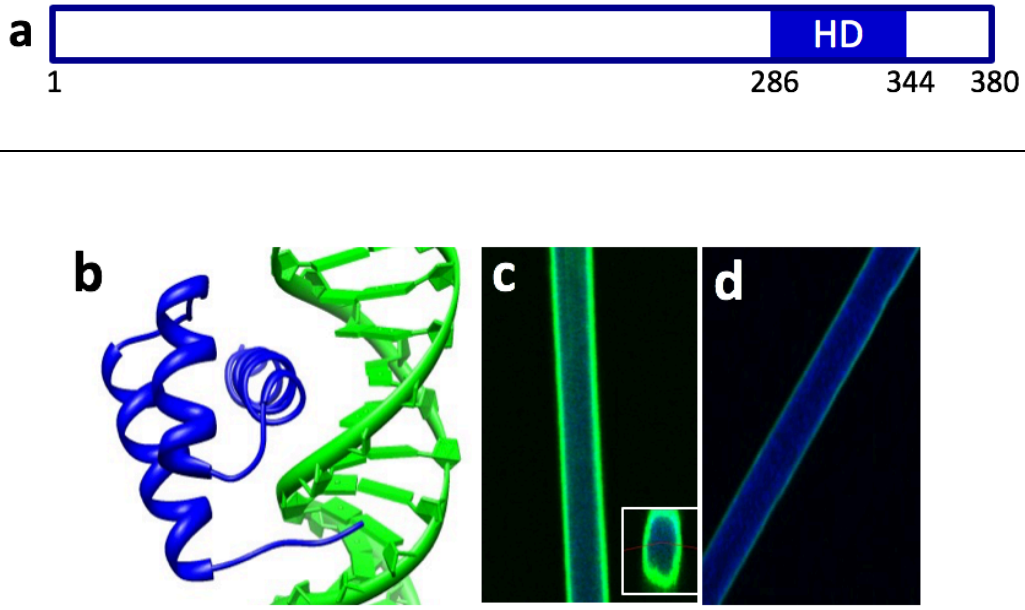

Figure 2-1 Homeodomain is capable of DNA binding in Ubx fibers.
(A) Schematic of the UbxIa amino acid sequence, showing the placement and relative size of the DNA-binding homeodomain. (B) DNA sequences for 40Sp, 40Nsp, 100Sp and 100Nsp. Ubx binding sites (underlined) are present in the specific (Sp) DNAs, but not the nonspecific (Nsp) DNAs. (C) Structure of the Ubx homeodomain (blue) bound to DNA (green), derived from PDB 1B8I (Passner 1999). (D) The FP3.38 antibody recognizes the homeodomain on the fiber surface. This pattern of binding to the surface of the Ubx fiber has been observed with other antibodies for dityrosine (Howell et al. 2015).

Bondos 2009). *In vitro*, Ubx self assembles to form nanoscale fibrils, films, and fibers with diameters ranging from 2-100 μm and meters in length (Greer et al. 2009; Majithia et al. 2011; Mendes et al. 2018). These materials are produced in a single-component, one-pot reaction at remarkably low protein concentrations (Greer et al. 2009; Majithia et al. 2011). Ubx materials are biocompatible, non-immunogenic, and biodegradable (Patterson et al. 2014; Patterson et al. 2015; Huang et al. 2016), and can be functionalized with proteins via gene fusion (Huang et al. 2011; Tsai et al. 2015). The Ubx protein contains a homeodomain (HD) that mediates DNA binding (Passner et al. 1999). If the HD is folded and accessible in Ubx materials, then they could also bind DNA in a sequence-specific manner. Given the high Ubx-DNA affinity (K_D = 160 pM) (Liu et al. 2008), the materials may allow orientation and slow DNA release.

FP3.38, an antibody that recognizes the Ubx homeodomain (White & Wilcox 1984), binds to Ubx fibers, which suggests that the homeodomain is accessible in the materials (Figure 2-1). If the Ubx homeodomain is indeed folded and the DNA binding interface is accessible, then Ubx materials should bind 5'-TAAT-3' with higher affinity than non-specific DNA (Liu et al. 2008; Kalodimos et al. 2004). Binding of Ubx fibers to Alexa 488-tagged 40 bp dsDNA containing a single 5'-TAAT-3' site (40Sp) and to a similar sequence lacking a Ubx binding site (40NSp) (Table 2-1, Table 2-2) was detected by confocal microscopy. Ubx fibers readily bind 40Sp DNA, but not 40NSp DNA (Figure 2-2 a,b). To ensure the fluorescent tag was not enabling binding, we verified these results using PCR to detect binding of a second set of longer dsDNAs, 100Sp and

Table 2-1 Characteristics of DNA sequences used in Ubx fiber binding experiments.

DNA	Type of Binding	Structure	# Binding Sites	Length
40Sp	Specific	linear	1	40 bp
40NSp	Non-specific	linear	0	40 bp
100Sp	Specific	linear	4	100 bp
100NSp	Non-specific	linear	0	100 bp
SM	Specific	minicircle, supercoiled	4	336 bp
LM	Specific	minicircle, linear	4	336 bp
LV	Specific	linear	1	504 bp
LV+6	Specific	linear	7	504 bp
SP	Specific	circular, supercoiled	42	7360 bp
LP	Specific	linear	42	7360 bp
SP+6	Specific	circular, supercoiled	48	7360 bp
LP+6	Specific	linear	48	7360 bp
pUC57	Specific	circular, supercoiled	22	2710 bp

Table 2-2 DNA sequences used in Ubx fiber binding experiments.
Ubx binding sites are in bold and underlined text. The *vegf* gene, used for PCR detection in Figure 2-2, is in red text. The DNA sequence amplified for PCR detection in Figure 2-4 is in blue text. DpnI restriction enzyme cutting sites within this second PCR region are highlighted in yellow. Finally, novel Ubx binding sites created by mutation within the *vegf* gene sequence are highlighted in light grey. (* The 3' end of the 40Sp/40NSp sense oligonucleotide strand was labeled with Alexa Fluor® 488 (NHS Ester) by Integrated DNA Technologies.)

DNA Name	Sequence 5'-3'
40Sp*	CCGGGCTGCACATGGTTAATGGCCAGTCCACG CGTAGATC
40NSp*	GATCGTGTCTACATGTCAGACAGTCAGCTGCATGACG AGTC
100Sp	CTTATTAGCTCACGCAGTCGATCACGTATTATCAGGAT GTAGAATTATCATAGATGCATCGCTGCTCATTATGCTA GTCACGCAGTCAGTCAGTCGTATGCTTGTCACTTCAGT CAGCAGATACCGTAGCATCAGTATGTAGAGTTCTCAC AGATGTATCACTGATCAGGATGCTAGTCAGTACTTCA

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
100NSp (Continued)	GTCAGTCGTATG
mc336 supercoiled or linearized with EcoRV	ATCACCGAAACGCGCGAGGCAGCTGTATGGCATGAAA GAGTTCTTCCCGGAAAACGCGGTGGAATATTTCGTTTC CTACTACGACTACTATCAGCCGGAAGCCTATGTACCG AGTTCCGACACTTTATTGAGAAAGATGCCTCAGCTCTG TTACAGGTCACTAATACCATCTAAGTAGTTGATTCATA GTGACTGCATATGTTGTGTTTTACAGTATTATGTAGTC TGTTTTTTATGCAAAATCTAATTTAATATTTGATATTTA TATCATTTTACGTTTCTCGTTCAGCTTTTTTATACTAAC TTGAGCGAAACGGGAAGGGTTTTCACCGAT
LV	CATATGAGTGCACCCATGGCAGAAGGAGGAGGGCAG AATCTCACGAAGTGGTGAAGTTCATGGATGTCTATCA GCGCAGCTACTGCCATCCAATCGAGACCCTGGTGGAC ATCTTCCAGGAGTACCTGATGAGATCGAGTACATCTTC AAGCCATCCTGTGTGCCCCTGATGCGATGCGGGGGCT GCTGCAATGACGAGGGCCTGGAGTGTGTGCCCACTGA GGAGTCCAACATCACCATGCAGATTATGCGGATCAAA CCTCACCAAGGCCAGCACATAGGAGAGATGAGTTCCT ACAGCACAACAAATGTGAATGCAGACCAAAGAAAGA TAGAGCAAGACAAGAAAATCCCTGTGGGCCTTGCTCA GAGCGGAGAAAGCATTTGTTTGTACAAGATCCGCAGA CGTGTAAATGTTCCTGCAAAAACACAGACTCGCGTTG CAAGGCGAGGCAGCTTGAGTTAAACGAACGTACTTGC AGATGTGACAAGCCAAGACGAAATCATATG
LV+6	CATATGAGTGCACCCATGGCAGAAGGAGGAGGGCATT AATCATCACGAAGTGGTGAAGTTCATGGATGTCTATC AGCGCAGCTACTGCCATCTAATCGAGACCCTGGTGGA CATCTTCCAGGAGTACCCTGATGAGATCGAGTACATC TTCAAGCCATCCTGTGTGCCCCTGATGCGATGCGGGG GCTGCTGTAATGACGAGGGCCTGGAGTGTGTGCCCAC TGAGGAGTCCAACATCACCATGCAGATTATGCGGATC AAACCTCACCAAGGCCAGCACATAGGAGAGATGAGCT TCCTACAGCACAACAAATGTTAATGCAGACCAAAGAA AGATAGAGCAAGACAAGAAAATCCCTGTGGGCCTTGC

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
LV+6 (continued)	TCAGAGCGGAGAAAGCATTAGTTTGTACAAGATCCGC AGACGTGTAAATGTTCCTGCAAAAACACAGACTCGCG TTGCAAGGCGAGGCAGCTTGAGTTAATCGAACGTACT TGCAGATGTGACAAGCCAAGACGAAATCATATG
SP supercoiled or linearized with HindIII	AGCTTATCGATGATAAGCTGTCAAACATGAGAATTCT TGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTAT AGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCA GGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTA TTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGC TCATGAGACAATAACCCTGATAAATGCTTCAATAATA TTGAAAAAGGAAGAGTATGAGTACAACATTTCCGTGT CGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGT TTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGAT GCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCG AACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTT TCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTA AAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGAC GCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATT CTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGA AAAGCATCTTACGGATGGCATGACAGTAAGAGAATTA TGCAGTGCTGCCATAACCATGAGTGATAACACTGCGG CCAACTTACTTCTGACAACGATCGGAGGACCGAAGGA GCTAACCGCTTTTTTGCACAACATGGGGGATCATGTA ACTCGCCTTGATCGTTGGGAACCGGAGCTGAATAAGC CATACCAAACGACGAGCGTGACACCACGATGCCTGCA GCAATGGCAACAACGTTGCGCAAACTATTAACTGGCG AACTACTTACTCTAGCTTCCCGGCAACAATTAATAGAC TGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGC GCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAA TCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGT AGTTATCTACACGACGGGGAGTCAGGCAACTATGGAT GAACGAATAGACAGATCGCTGAGATAGGTGCCTCACT GATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCAT ATATACTTTAGATTGATTTAAAACTTCATTTTTAATTT

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
SP supercoiled or linearized with HindIII (continued)	AAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCAT GACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAA CAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCC GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTC TAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTC TGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGT TACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTT ACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACA GCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGA TACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTC CCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCG GCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC CAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGAT GCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCA GCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGG CCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGAT TCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGA TACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAG TCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGG TATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCG CATATATGGTGCACTCTCAGTACAATCTGCTCTGATGC CGCATAGTTAAGCCAGTATACACTCCGCTATCGCTAC GTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAAC ACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCG GCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGA GCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAA ACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGG TCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGC GTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCT GGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTT

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
SP supercoiled or linearized with HindIII (continued)	TTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGAT TTCTGTTCATGGGGGTAATGATACCGATGAAACGAGA GAGGATGCTCACGATACGGGTTACTGATGATGAACAT GCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGG CGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCA GGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGT GTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGA TCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGT TTCCAGACTTTACGAAACACGGAAACCGAAGACCATT CATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCA GTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCT GCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGT CCTCAACGACAGGAGCACGATCATGCGCACCCGTGGC CAGGACCCAACGCTGCCCGAGATGCGCCGCGTGCGGC TGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCA AGGGTTGGTTTGCGCATTCACAGTTCTCCGCAAGAATT GATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCG AGGTGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCG GCTCCATGCACCGCGACGCAACGCGGGGAGGCAGAC AAGGTATAGGGCGGCGCCTACAATCCATGCCAACCCG TTCCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGA CGATCAGCGGTCCAGTGATCGAAGTTAGGCTGGTAAG AGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGTCG TCATCTACCTGCCTGGACAGCATGGCCTGCAACGCGG GCATCCCGATGCCGCCGGAAGCGAGAAGAATCATAAT GGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGC AAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCGA TAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGG ACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATT CCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGC TCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAG CGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAG AAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCC GCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCT CAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTG

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
SP supercoiled or linearized with HindIII (continued)	AGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGC TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAT GAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTAT TGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACG GGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAG AGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAG CAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGG ATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCAC TACCGAGATATCCGCACCAACGCGCAGCCCGGACTCG GTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGT TGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATT CAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCA CTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTG ATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGC AGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACA GCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTC CACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATA ATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAA ATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGC AATGGCATCCTGGTCATCCAGCGGATAGTTAATGATC AGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCG CCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATC GACACCACCACGCTGGCACCCAGTTGATCGGCGCGAG ATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAG GGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGA CTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAA TGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCC CGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCA CGCGGGAAACGGTCTGATAAGAGACACCGGCATACTC TGCGACATCGTATAACGTTACTGGTTTCACATTCACCA CCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATA CCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGA TCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAA GCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCC GCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCC

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
SP supercoiled or linearized with HindIII (continued)	AACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCA CGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGC CCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCG CCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCA CGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCC GCGAAATTAATACGACTCACTATAGGGGAATTGTGAG CGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAAC TTTAAGAAGGAGATATACCATGGGCCATCATCATCAT CATCATCATCATCATCACAGCAGCGGCCATATCGACG ACGACGACAAGCATATGAGTGCACCCATGGCAAAGG GAGGGCAGAATCATCACGAAGTGGTGAAGTTCATGGA TGTCTATCAGCGCAGCTACTGCCATCCAATCGAGACC CTGGTGGACATCTTCCAGGAGTACCCTGATGAGATCG AGTACATCTTCAAGCCATCCTGTGTGCCCCTGATGCGA TGCGGGGGCTGCTGCAATGACGAGGGCCTGGAGTGTG TGCCCACTGAGGAGTCCAACATCACCATGCAGATTAT GCGGATCAAACCTCACCAAGGCCAGCACATAGGAGA GATGAGCTTCCTACAGCACAACAAATGTGAATGCAGA CCAAAGAAAGATAGAGCAAGACAAGAAAATCCCTGT GGGCCTTGCTCAGAGCGGAGAAAGCATTTGTTTGTAC AAGATCCGCAGACGTGTAAATGTTCCTGCAAAAACAC AGACTCGCGTTGCAAGGCGAGGCAGCTTGAGTTAAAC GAACGTACTTGCAGATGTGACAAGCCAAGACGAAATC ATATGAACTCGTACTTTGAACAGGCCTCCGGCTTTTAT GGCCATCCGCACCAGGCCACCGGAATGGCGATGGGCA GCGGTGGCCACCACGACCAGACGGCCAGTGCAGCGG CGGCCGCGTACAGGGGATTCCCTCTCTCGCTGGGCAT GAGTCCCTATGCCAACCACCATCTGCAGCGCACCACC CAGGACTCGCCCTACGATGCCAGCATCACGGCCGCCT GCAATAAGATATACGGCGATGGAGCCGGAGCCTACAA ACAGGACTGCCTGAACATCAAGGCGGATGCGGTGAAT GGCTACAAAGACATTTGGAACACGGGCGGCTCGAATG GCGGCGGGGGTGGCGGCGGAGGCGGTGGTGGCGGCG GAGCGGGCGGAACAGGTGGAGCCGGCAATGCCAATG GCGGTAATGCGGCCAATGCAAACGGACAGAACAATCC

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
SP supercoiled or linearized with HindIII (continued)	GGCGGGCGGTATGCCCGTTAGACCCTCCGCCTGCACC CCAGATTCCCGAGTGGGCGGCTACTTGGACACGTCGG GCGGCAGTCCCGTTAGCCATCGCGGCGGCAGTGCCGG CGGTAATGTGAGTGTCAGCGGCGGCAACGGCAACGCC GGAGGCGTACAGAGCGGCGTGGGCGTGGCCGGAGCG GGCACTGCCTGGAATGCCAATTGCACCATCTCGGGCG CCGCTGCCCAAACGGCGGCCGCCAGCAGTTTACACCA GGCCAGCAATCACACATTCTACCCCTGGATGGCTATC GCAGGTAAGATAAGATCTGATTTAACACAATACGGCG GCATATCAACAGACATGGGTAAGAGATACTCAGAATC TCTTGCGGGCTCACTTCTACCAGACTGGCTAGGTACAA ATGGTCTGCGAAGACGCGGCCGACAGACATACACCCG CTACCAGACGCTCGAGCTGGAGAAGGAGTTCCACACG AATCATTATCTGACCCGCAGACGGAGAATCGAGATGG CGCACGCGCTATGCCTGACGGAGCGGCAGATCAAGAT CTGGTTCCAGAACCGGCGAATGAAGCTGAAGAAGGA GATCCAGGCGATCAAGGAGCTGAACGAACAGGAGAA GCAGGCGCAGGCCCAGAAGGCGGCGGCGGCAGCGGC TGCGGCGGCGGCGGTCCAAGGTGGACACTTAGATCAG TAATAGGGTTAGGCTGCTAACAAAGCCCGAAAGGAAG CTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGC ATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGT TTTTTGCTGAAAGGAGGAACTATATCCGGATATCCCG CAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATG CCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGA TGAGCGCATTGTTAGATTTCATACACGGTGCCTGACT GCGTTAGCAATTTAACTGTGATAAACTACCGCATTAA
SP+6 supercoiled or linearized with HindIII	AGCTTATCGATGATAAGCTGTCAAACATGAGAATTCT TGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTAT AGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCA GGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTA TTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGC TCATGAGACAATAACCCTGATAAATGCTTCAATAATA TTGAAAAAGGAAGAGTATGAGTACAACATTTCCGTGT CGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGT

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
SP+6 supercoiled or linearized with HindIII (continued)	TTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGAT GCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCG AACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTT TCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTA AAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGAC GCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATT CTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGA AAAGCATCTTACGGATGGCATGACAGTAAGAGAATTA TGCAGTGCTGCCATAACCATGAGTGATAACACTGCGG CCAACTTACTTCTGACAACGATCGGAGGACCGAAGGA GCTAACCGCTTTTTTGCACAACATGGGGGATCATGTA ACTCGCCTTGATCGTTGGGAACCGGAGCTGAATAAGC CATACCAAACGACGAGCGTGACACCACGATGCCTGCA GCAATGGCAACAACGTTGCGCAAACTATTAACTGGCG AACTACTTACTCTAGCTTCCCGGCAACAATTAATAGAC TGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGC GCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAA TCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGT AGTTATCTACACGACGGGGAGTCAGGCAACTATGGAT GAACGAATAGACAGATCGCTGAGATAGGTGCCTCACT GATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCAT ATATACTTTAGATTGATTTAAAACTTCATTTTTAATTT AAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCAT GACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAA CAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCC GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTC TAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTC TGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGT TACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTT ACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACA

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
SP+6 supercoiled or linearized with HindIII (continued)	GCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGA TACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTC CCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCG GCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC CAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGAT GCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCA GCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGG CCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGAT TCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGA TACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAG TCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGG TATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCG CATATATGGTGCACTCTCAGTACAATCTGCTCTGATGC CGCATAGTTAAGCCAGTATACACTCCGCTATCGCTAC GTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAAC ACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCG GCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGA GCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAA ACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGG TCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGC GTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCT GGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTT TTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGAT TTCTGTTCATGGGGGTAATGATACCGATGAAACGAGA GAGGATGCTCACGATACGGGTTACTGATGATGAACAT GCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGG CGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCA GGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGT GTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGA TCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGT TTCCAGACTTTACGAAACACGGAAACCGAAGACCATT CATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCA GTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCT GCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGT

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
SP+6 supercoiled or linearized with HindIII (continued)	CCTCAACGACAGGAGCACGATCATGCGCACCCGTGGC CAGGACCCAACGCTGCCCGAGATGCGCCGCGTGCGGC TGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCA AGGGTTGGTTTGCGCATTCACAGTTCTCCGCAAGAATT GATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCG AGGTGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCG GCTCCATGCACCGCGACGCAACGCGGGGAGGCAGAC AAGGTATAGGGCGGCGCCTACAATCCATGCCAACCCG TTCCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGA CGATCAGCGGTCCAGTGATCGAAGTTAGGCTGGTAAG AGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGTCG TCATCTACCTGCCTGGACAGCATGGCCTGCAACGCGG GCATCCCGATGCCGCCGGAAGCGAGAAGAATCATAAT GGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGC AAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCGA TAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGG ACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATT CCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGC TCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAG CGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAG AAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCC GCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCT CAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTG AGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGC TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAT GAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTAT TGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACG GGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAG AGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAG CAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGG ATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCAC TACCGAGATATCCGCACCAACGCGCAGCCCGGACTCG GTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGT TGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATT CAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCA

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
SP+6 supercoiled or linearized with HindIII (continued)	CTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTG ATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGC AGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACA GCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTC CACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATA ATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAA ATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGC AATGGCATCCTGGTCATCCAGCGGATAGTTAATGATC AGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCG CCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATC GACACCACCACGCTGGCACCCAGTTGATCGGCGCGAG ATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAG GGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGA CTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAA TGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCC CGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCA CGCGGGAAACGGTCTGATAAGAGACACCGGCATACTC TGCGACATCGTATAACGTTACTGGTTTCACATTCACCA CCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATA CCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGA TCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAA GCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCC GCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCC AACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCA CGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGC CCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCG CCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCA CGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCC GCGAAATTAATACGACTCACTATAGGGGAATTGTGAG CGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAAC TTTAAGAAGGAGATATACCATGGGCCATCATCATCAT CATCATCATCATCATCACAGCAGCGGCCATATCGACG ACGACGACAAGCATATGAGTGCACCCATGGCAAAGG AGGAGGGCATAATCATCACGAAGTGGTGAAGTTCATG GATGTCTATCAGCGCAGCTACTGCCATCTAATCGAGA

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
SP+6 supercoiled or linearized with HindIII (continued)	CCCTGGTGGACATCTTCCAGGAGTACCCTGATGAGAT CGAGTACATCTTCAAGCCATCCTGTGTGCCCCTGATGC GATGCGGGGGCTGCTGTAATGACGAGGGCCTGGAGTG TGTGCCCACTGAGGAGTCCAACATCACCATGCAGATT ATGCGGATCAAACCTCACCAAGGCCAGCACATAGGAG AGATGAGCTTCCTACAGCACAACTAATGTGAATGCAG ACCAAAGAAAGATAGAGCAAGACAAGAAAATCCCTG TGGGCCTTGCTCAGAGCGGAGAAAGCATTAGTTTGTA CAAGATCCGCAGACGTGTAAATGTTCCTGCAAAAACA CAGACTCGCGTTGCAAGGCGAGGCAGCTTGAGTTAAT CGAACGTACTTGCAGATGTGACAAGCCAAGACGAAAT CATATGAACTCGTACTTTGAACAGGCCTCCGGCTTTTA TGGCCATCCGCACCAGGCCACCGGAATGGCGATGGGC AGCGGTGGCCACCACGACCAGACGGCCAGTGCAGCG GCGGCCGCGTACAGGGGATTCCCTCTCTCGCTGGGCA TGAGTCCCTATGCCAACCACCATCTGCAGCGCACCAC CCAGGACTCGCCCTACGATGCCAGCATCACGGCCGCC TGCAATAAGATATACGGCGATGGAGCCGGAGCCTACA AACAGGACTGCCTGAACATCAAGGCGGATGCGGTGAA TGGCTACAAAGACATTTGGAACACGGGCGGCTCGAAT GGCGGCGGGGGTGGCGGCGGAGGCGGTGGTGGCGGC GGAGCGGGCGGAACAGGTGGAGCCGGCAATGCCAAT GGCGGTAATGCGGCCAATGCAAACGGACAGAACAAT CCGGCGGGCGGTATGCCCGTTAGACCCTCCGCCTGCA CCCCAGATTCCCGAGTGGGCGGCTACTTGGACACGTC GGGCGGCAGTCCCGTTAGCCATCGCGGCGGCAGTGCC GGCGGTAATGTGAGTGTCAGCGGCGGCAACGGCAACG CCGGAGGCGTACAGAGCGGCGTGGGCGTGGCCGGAG CGGGCACTGCCTGGAATGCCAATTGCACCATCTCGGG CGCCGCTGCCCAAACGGCGGCCGCCAGCAGTTTACAC CAGGCCAGCAATCACACATTCTACCCCTGGATGGCTA TCGCAGGTAAGATAAGATCTGATTTAACACAATACGG CGGCATATCAACAGACATGGGTAAGAGATACTCAGAA TCTCTTGCGGGCTCACTTCTACCAGACTGGCTAGGTAC AAATGGTCTGCGAAGACGCGGCCGACAGACATACACC

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
	CGCTACCAGACGCTCGAGCTGGAGAAGGAGTTCCACA CGAATCATTATCTGACCCGCAGACGGAGAATCGAGAT GGCGCACGCGCTATGCCTGACGGAGCGGCAGATCAAG ATCTGGTTCCAGAACCGGCGAATGAAGCTGAAGAAGG AGATCCAGGCGATCAAGGAGCTGAACGAACAGGAGA AGCAGGCGCAGGCCCAGAAGGCGGCGGCGGCAGCGG CTGCGGCGGCGGCGGTCCAAGGTGGACACTTAGATCA GTAATAGGGTTAGGCTGCTAACAAAGCCCGAAAGGAA GCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAG CATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGG TTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCCG CAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATG CCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGA TGAGCGCATTGTTAGATTTCATACACGGTGCCTGACTG CGTTAGCAATTTAACTGTGATAAACTACCGCATTAA
pUC57	TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACA CATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAA GCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCG TCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACT ATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCA TATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGA AAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGC GCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTC GCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCA AGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGT CACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCT CGGTACCTCGCGAATGCATCTAGATATCGGATCCCGG GCCCGTCGACTGCAGAGGCCTGCATGCAAGCTTGGCG TAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTA TCCGCTCACAATTCCACACAACATACGAGCCGGAAGC ATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTC CAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGG GCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCT

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
pUC57 (continued)	CGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTC AAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAA AAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGT TTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAA AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTC CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCG GATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCG CTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGT GTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAA CCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAA CTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTAT CGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAG AGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG TGGTGGCCTAACTACGGCTACACTAGAAGAACAGTAT TTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCA CCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAG ATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTT TGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGA AAACTCACGTTAAGGGATTTTGGTCATGAGATTATCA AAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAAT GAAGTTTTAAATCAATCTAAAGTATATATGAGTAAAC TTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCA CCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGT TGCCTGACTCCCCGTCGTGTAGATAACTACGATACGG GAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATAC CGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGC AATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAG TGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTA ATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGT TAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCA TCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTC AGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGAT

Table 2-2 Continued	
DNA Name	**Sequence 5'-3'**
pUC57 (continued)	CCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGT CCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTT ATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTA CTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGT GAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC GGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGA TAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTC ATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAA GGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCC ACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTT CACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAA AATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAA TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGA AGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT ATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTT CCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCT AAGAAACCATTATTATCATGACATTAACCTATAAAAA TAGGCGTATCACGAGGCCCTTTCGTC

100NSp (Table 2-1, Table 2-2). The 100Sp sequence contains four 5'-TAAT-3' binding sites, whereas the 100NSp sequence lacks binding sites. Again, binding was only observed when the 5'-TAAT-3'sequence was present (Figure 2-2 c). This specific binding to DNA is also seen in competition experiments (Figure 2-3). These data demonstrate that i) Ubx fibers are capable of binding DNA, ii) DNA binding by the fibers is sequence specific, and iii) the homeodomain must be folded and accessible in Ubx fibers.

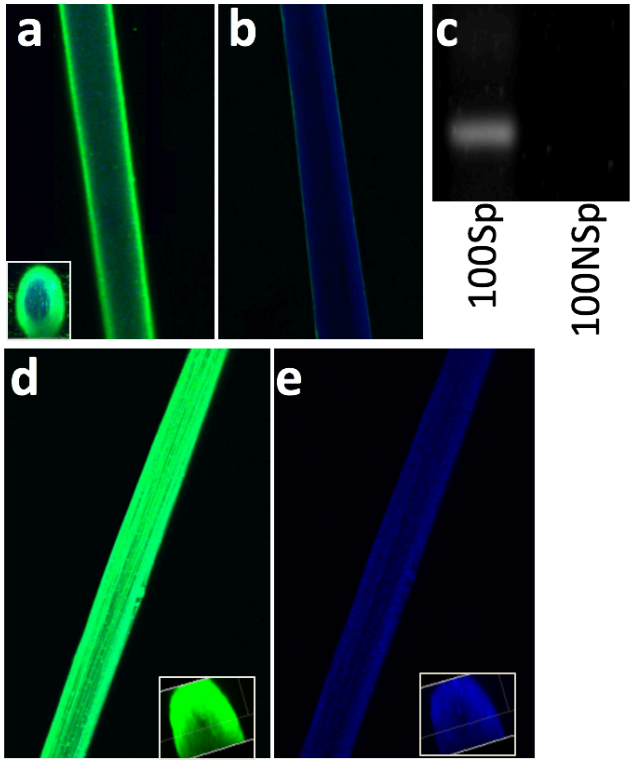

Figure 2-2 In Ubx fibers, the homeodomain is on the surface and capable of binding DNA in sequence-specific manner.
Ubx fibers self-assemble in the presence of 40Sp DNA, which contains 1 binding site but not 40NSp, which lacks binding sites. (A) 40Sp DNA with a fluorescent tag (green) bind Ubx fibers but not 40NSp DNA (B) a similar sequence lacking Ubx binding sites. (C) PCR of Ubx fibers exposed to DNA detects the presence of 100Sp DNA, but not 100NSp DNA. The characteristics of the sequences of DNAs used in this work are available in Supplementary Tables 2-1 and -2. (D) green fluorescence from the Alexa488 tag on 40Sp DNA in a fiber assembled from Ubx-40Sp dimers. Inset, a cross-section of the fiber demonstrating DNA is present throughout the fiber. (E) Blue auto-fluorescence from the same fiber, with a cross-section in the inset.

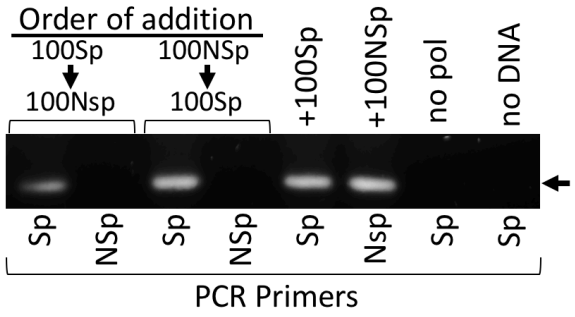

Figure 2-3 Competition experiments demonstrate DNA-binding specificity by Ubx fibers.
Ubx fibers, incubated first with 100Sp DNA, are not able to subsequently bind 100NSp non-specific DNA. In contrast, Ubx fibers incubated first in 100NSp DNA are able to bind 100Sp DNA. The order of DNA addition is listed above each reaction, and the primers used to detect the presence of DNA by PCR are listed below each lane. Sp primers detected the 100Sp DNA; NSp primers detected the 100NpP DNA. Positive control reactions (labeled +100Sp and +100NSp) contained 10 mg/mL of 100Sp or 100NSp template DNA. Negative control experiments for the Sp PCR reactions lacked polymerase (labeled no pol) or template DNA (labeled no DNA).

The Ubx monomer, even when fused to proteins over nine-times its size, still self-assembles (Greer et al. 2009), suggesting that Ubx monomers, pre-bound to DNA, may also self-assemble into materials. To test this possibility, we mixed Ubx monomers with fluorescently labeled 40Sp or 40NSp DNA at protein and DNA concentrations well above K_D, ensuring binding. This mixture was then incubated on siliconized glass slides for 4-5 hours to allow assembly. Ubx bound to 40Sp DNA self-assembled,

incorporating DNA throughout the fiber (Figure 2-2 d and e), whereas the non-specific DNA did not accumulate in the fiber. Therefore, the homeodomain is folded and functional throughout Ubx materials and not just on the fiber surface.

Many potential applications of protein-DNA composite materials require non-linear DNAs, such as supercoiled plasmids or DNA origami (Michelotti et al. 2012; Niemeyer 2010; Numata et al. 2010). To determine whether non-linear DNA structures bind to Ubx fibers, we tested the ability of Ubx fibers to bind to mv336, a 336 bp minicircle. Minicircles are small circular supercoiled dsDNAs being developed for gene therapy (Mayrhofer, Schleef, & Jechlinger 2009; Kobelt et al. 2013; Zhao et al. 2011). We compared Ubx fiber binding of supercoiled mv336 (SM) and linear mv336 (LM), each of which contains four 5'-TAAT-3' sites (Fogg et al. 2006). Ubx fibers bound both the supercoiled and linear DNA (Figure 2-4 a). To test the effects of DNA size on binding by fibers, we also compared binding of a supercoiled plasmid DNA (SP), which contains 42 5'-TAAT-3' sites, with binding by the same plasmid, linearized with HindIII (LP) (Table 2-1, Table 2-2). Ubx fibers bound both the supercoiled and linear DNA comparably at DNA concentrations spanning four orders of magnitude (Figure 2-4 b). This data demonstrates that similar binding of supercoiled and linearized DNAs is not an artifact of either DNA sequence or DNA length. We conclude that Ubx fibers are capable of binding non-linear DNA structures.

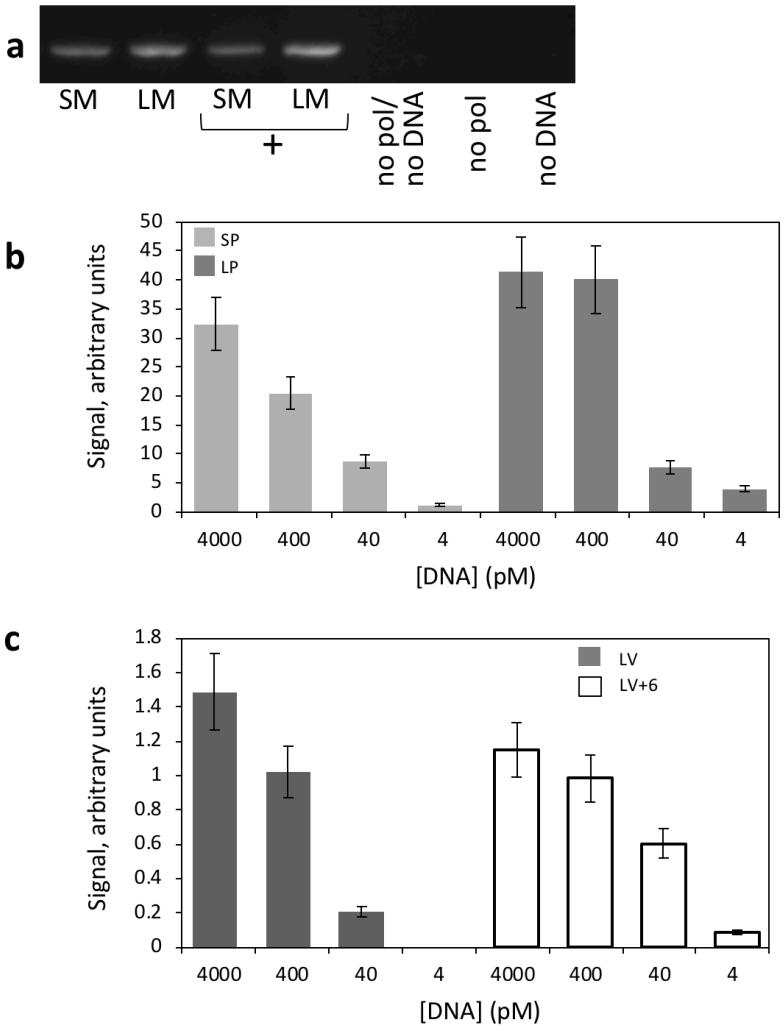

Figure 2-4 Ubx fibers bind both supercoiled DNA and linear DNA.
(A) Supercoiled, circular minivector (SM) and linearized minivector DNA (LM) bind Ubx fibers. Included are PCR reactions of positive controls using template DNA in solution (+) and negative controls lacking polymerase (no pol) and/or DNA (no DNA). (B-C) Quantitation of PCR reactions, varying DNA concentration reveals qualitative differences in affinity for the four DNA samples. (B) Both the SP and LP bound Ubx fibers to similar extents at DNA concentrations spanning four orders of magnitude. (C) Significant differences in binding were not observed at high (saturating) concentrations of DNA for all DNAs, but LV+6 DNA binds better than the LV DNA sequence at lower DNA concentrations.

Because Ubx recognizes specific DNA sequences, we should be able to tune the apparent affinity of Ubx materials for DNA by altering the number of Ubx binding sites in the DNA sequence. We compared binding of the linear wild type *vegf* DNA sequence (LV) which has a single binding site, and a mutant version (LV+6) in which an additional six binding sites have been added. At low (pM) DNA concentrations, significantly more LV+6 than LV DNA binds and is retained by Ubx fibers. As expected, these differences are not observed at saturating DNA concentrations (nM). Thus, the differences observed in binding are due to changes in affinity based on DNA sequence, and not an unanticipated change in DNA structure or protein-DNA interactions (Figure 2-4 c). Therefore, the extent of DNA binding by Ubx fibers can be tuned by altering the number of binding sites.

Interactions between Ubx fibers and DNA must be long-lived to generate useful composite materials. Although the Ubx homeodomain has a remarkably high affinity for DNA, the half-life ($t_{1/2}$) for interaction of a Ubx homeodomain monomer with a single DNA site is only 27 min (Beachy et al. 1993), too short to be useful for many applications. However, $t_{1/2}$ theoretically increases to 260 min for DNA with four binding sites and to 747 min for DNA with 12 binding sites (Beachy et al. 1993). To measure DNA retention times on Ubx fibers, we incubated fibers in the presence of supercoiled plasmid containing the *vegf* gene (SP). After repeated rinsing, the fibers were incubated in a large volume of buffer to approximate infinite dilution, to prevent reassociation of released DNA. After a second series of washes, DNA still bound to the

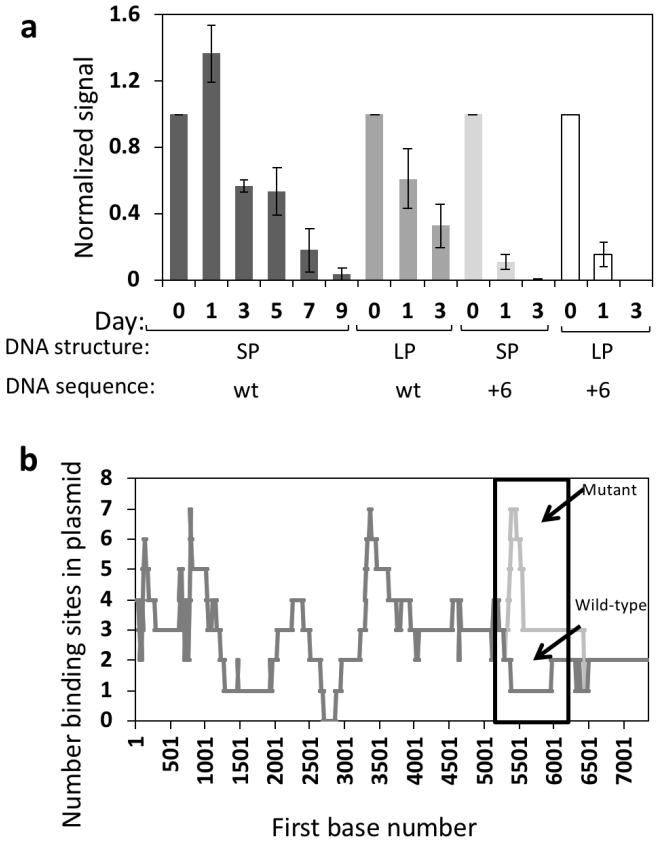

Figure 2-5 DNA is retained by Ubx fibers for several days, and DNA reorients on the fibers to maximize binding.

(A) Quantitation of multiple experiments with error bars indicating standard deviation. Ubx fibers were incubated in the presence of SP, LP, SP+6, or LP+6 which contains 42 (wild type) or 48 (mutant) Ubx binding sites, respectively. PCR was used to detect the presence of DNA bound to Ubx fibers. The differences in number of binding sites present and the retention time differs for each sequence indicating re-orientation of DNA on the fiber surface and the ability to detect bound DNA. The SP reorients between the 0- and 1-day measurements, but no reorientation was observed for LP, SP+6 and LP+6 indicating that binding site density is able to orient DNA on materials to maximize the number of binding sites in contact with Ubx. (B) Graphic representation of Ubx binding sites (TAAT or ATTA) density in SP/LP wt and SP/LP mut, using a sliding window of 500 bp at one base pair intervals, the approximate length of the *vegf* gene detection site. The *vegf* detection site contains 1 binding site within the detected region for wild type sequence, whereas 7 sites are in the detected region for mutant sequences.

Ubx fiber was detected by PCR. DNA remained bound for up to one week (Figure 2-5 a), representing an extremely long $t_{1/2}$. We attribute the long $t_{1/2}$ to the high affinity of the Ubx homeodomain for DNA (Liu et al. 2008), the presence of many binding sites on the plasmid, and the fact that the Ubx homeodomains are tethered together by the structure of the materials.

Both SP and LP signal decayed over time. However, more SP signal was consistently observed after 1-day wash than immediately after DNA binding and washing. Because DNA is not added to the fibers during the extended wash period, the only remaining possibility is that supercoiled DNA is more easily detected on Day 1 than on Day 0 (Figure 2-5 a). Indeed, we used PCR to detect the *vegf* gene, and 5'-TAAT-3' sites are less abundant in this region than in the rest of the gene (Figure 2-5 b). If initial plasmid binding to the Ubx fiber is random, then in some molecules the *vegf* gene would face the fiber, potentially hindering the PCR reaction. The extended wash period provides an opportunity for the binding sites in the *vegf* gene on the plasmid to release, and then compete for re-binding with regions of the plasmid with more densely populated sites. Over time, the system should reach equilibrium with the *vegf* gene exposed and detectable by PCR, thus improving the ability to detect DNA, but not changing the number of DNA molecules actually present. To test this hypothesis, we added six Ubx binding sites to the *vegf* region of the plasmid, creating supercoiled SP+6 and linearized LP+6. Unlike the wild-type plasmid, less of the supercoiled mutant plasmid was detected (Figure 2-5 a), demonstrating that the effect is indeed caused by binding site

density, and that DNA orients on materials to maximize the number of binding sites in contact with Ubx. This binding reorientation behavior is similar to "self-healing", which has been observed for other interactions involving multivalent, non-covalent bonds (Ahn et al. 2014).

Although Ubx fibers are remarkably resistant to proteolysis (Patterson et al. 2015), nuclease digestion is a major concern with DNA-containing materials (de Vries, Zhang, & Hermann 2013). Because fiber binding alters the efficiency of a PCR reaction, we speculated that binding to materials might also protect DNA from digestion by restriction enzymes. After binding SP+6 to Ubx fibers and washing away excess DNA, we added the restriction enzyme DpnI. DpnI digests DNA, which prevents DNA amplification by PCR. We amplified a 379 bp region that contains both DpnI restriction sites (five) and Ubx binding sites (four) (Table 2-1, Table 2-2). Free DNA was used in a positive control for DpnI digestion reaction. Although fiber binding did not prevent DNA digestion, DNA bound to fiber was digested more slowly than DNA in the positive control reaction (Figure 2-6).

In order to be a useful vehicle for gene therapy, Ubx fibers must be capable of delivering DNA to cells. A slow, controlled release of DNA is ideal for this type of application, as it increases both the level and length of gene expression (Pannier & Shea 2004). To measure the duration of DNA release, we used Ubx fibers in transformation experiments with *E. coli* cells. Fibers were soaked in pUC57, a small plasmid with only 22 Ubx

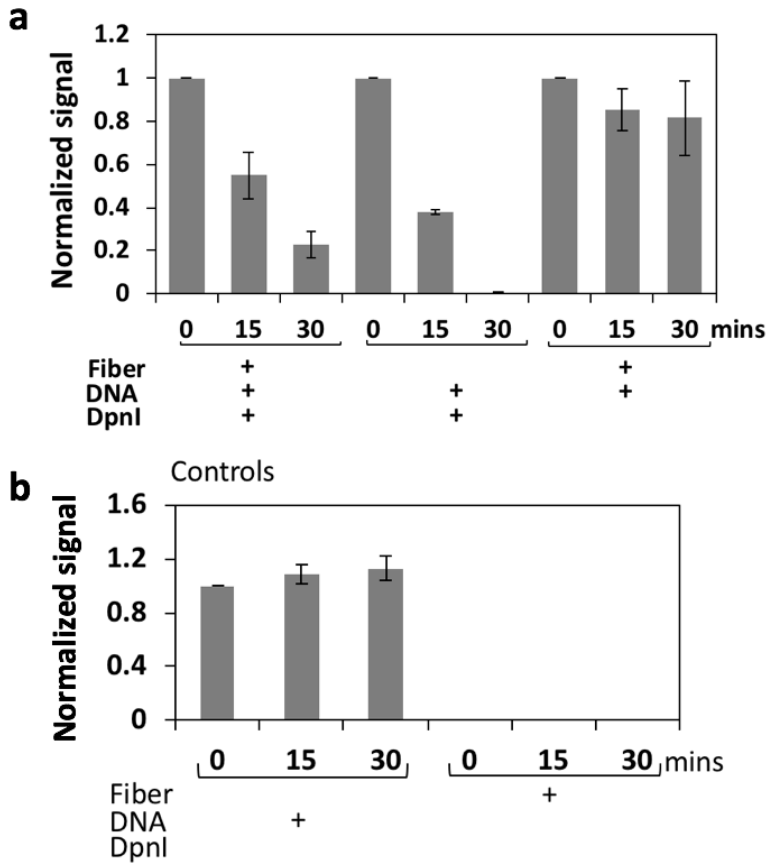

Figure 2-6 DNA is protected from degradation once bound to Ubx fibers.
(A) Ubx fibers were incubated in LP DNA. DNA, bound to Ubx fiber, was exposed to the restriction enzyme DpnI for varying times. The amount of DNA remaining intact was analyzed by PCR. Fiber binding protected DNA from the restriction enzyme to some extent. (B) Free LP DNA in the absence of fiber and DpnI restriction enzyme was not degraded over time as expected. Ubx fibers alone were not able to elicit any signal after PCR analysis, thus confirming our results that the Ubx fiber is able to protect DNA from degradation somewhat.

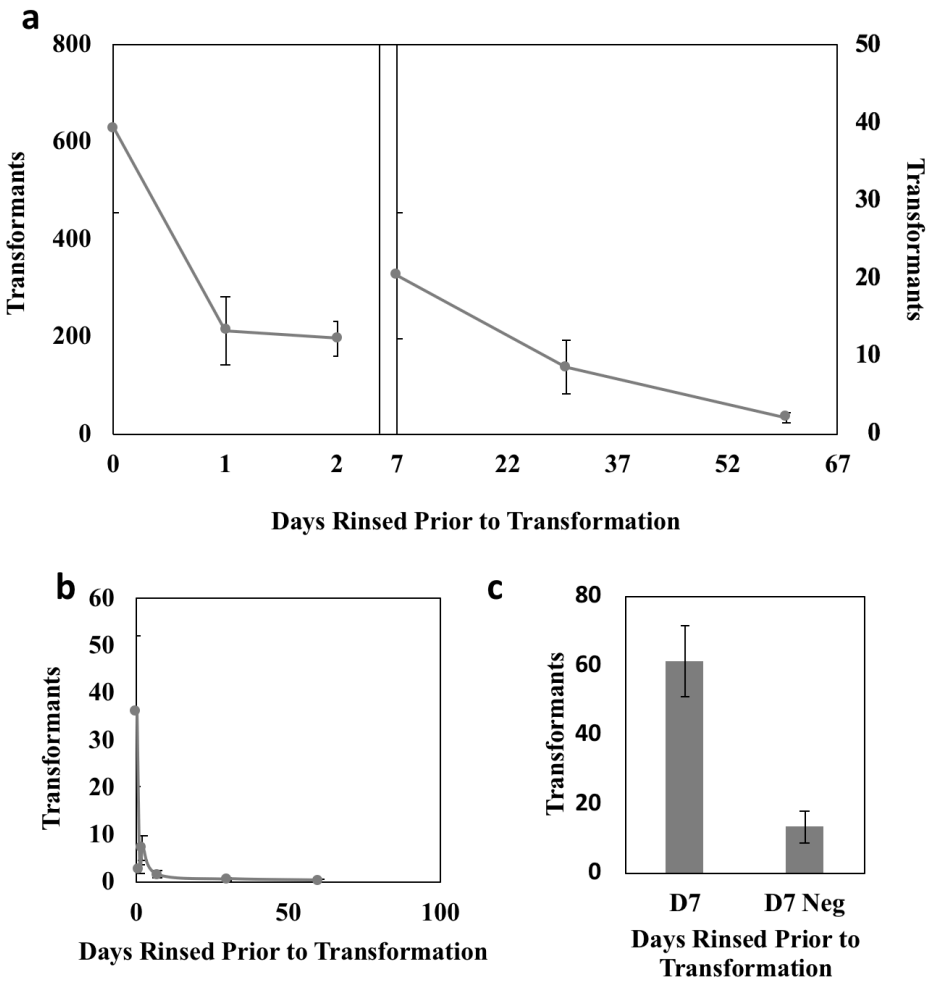

Figure 2-7 Extended release of DNA by Ubx fibers for 60 days.
(A) Ubx fibers were incubated in pUC57 DNA for 30 minutes, rinsed in PBS for a varying amount of time, and used to deliver DNA to *E. coli* cells in a transformation. Cells were transformed after up to 2 months of rinsing. Standard error bars are shown. (B) Controls with no fibers, which underwent the same treatments as (A). (C) A 10x increase in rinse volume does not remove more DNA. The ratio of transformants from positive to negative treatments decreased slightly, but there was still a significant difference between positive (loop with fiber) and negative (loop with no fiber). The DNA is not rebinding the fiber due to rinse volume.

binding sites (Table 2-1, Table 2-2), then had a series of rinse steps at infinite dilution that varied from minutes to 2 months. We found that Ubx fibers with bound DNA can be used to deliver DNA and transform *E. coli* cells after up to 2 months of rinsing (Figure 2-7a, Figure 2-7b). Increasing the rinse volume by 1000% does not remove the DNA from the fibers (Figure 2-7 c). The combination of DNA binding, materials forming Ubx with plasmid DNA allows for creation of materials that are able to release DNA over an extended period of time and transform prokaryotic cells *in vitro*.

Combining protein and DNA in materials both increases the possibilities for structural design and extends the range of potential functions that can be incorporated. We have devised methods for creating protein-DNA composite biomaterials that can be accomplished in either a single step (co-assembly) or two steps (surface binding) without harsh chemicals. DNA can be added only to the Ubx fiber surface or throughout the materials as needed. The combination of the high affinity of the Ubx homeodomain for its cognate DNA sequence and multi-site binding allows construction of complexes that are stable, a characteristic normally associated with covalent cross-linking, able to release DNA, and capable of self-healing to orient the DNA, properties of non-covalent complexes. The ability to control DNA affinity by varying the number of binding sites allows the lifetime of the protein-DNA complex to be tuned. This level of control enables engineering protein-DNA composite materials for different applications requiring brief associations, such as DNA delivery, and more stable complexes, as in construction of customizable protein-DNA structures.

2.3 Methods

Production of Ubx materials: For all experiments except delivery of DNA to cells, Ubx was expressed and purified as previously reported (Liu et al. 2008). In brief, the *ubxIa* gene was cloned into pET- 3c vector and transformed into BL21 (DE3) pLysS *E. coli* cells (Novagen). Single colonies were used to inoculate overnight liquid cultures. Protein expression was induced at mid-log phase with 0.001 M isopropyl-β-D-1-thiogalactopyranoside (IPTG) for 3 h and cells were harvested by centrifugation and stored at -20°C. Frozen cell pellets were lysed in 50 mM Tris–HCl pH=8.0, 10 mM β-mecaptoethanol, 10 mM Ethylenediaminetetraacetic acid (EDTA), 800 mM NaCl, Complete Proteinase Inhibitor Mixture (Roche), and 0.04 mg/mL DNase I. The lysate was centrifuged to remove cell debris. The lysis supernatant was treated with polyethyleneimine (50% w/v, 200 μL) and centrifuged to remove DNA fragments. The pH of the supernatant was adjusted to 6.8 with NaH_2PO_4 and centrifuged to remove precipitates. The final supernatant was diluted to 100 mL with buffer Z (5% glycerol, 10 mM β-mercaptoethanol, 0.1 mM EDTA, 25 mM NaH_2PO_4, and 150 mM NaCl, pH = 6.8) and loaded onto a P11 phosphocellulose column (Whatman) equilibrated with buffer Z. Ubx protein was eluted with a 0.15–1.2 M NaCl gradient in buffer Z. Fractions containing Ubx were collected and dialyzed into dialysis buffer (5% glycerol, 10 mM β-mercaptoethanol, 150 mM NaCl, and 50 mM Tris, pH = 8.0). Protein samples were then incubated with Ni-NTA agarose resin (Qiagen) and clarified with Ni-NTA chromatography. Protein was eluted with pH 8 elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM β-mercaptoethanol, 5% glucose, and 200mM imidazole).

Fibers were pulled using the drop method, as previously described (Mendes et al. 2019). In brief, drops of protein elution were transferred to a siliconized glass slide, covered and incubated for 6 to 16 hours at room temperature. Sterile inoculating loops were used to pull fibers from the drops, with an average of four fiber wraps per loop. Fibers were immediately washed in PBS buffer (20 mM NaH2PO4, 80 mM Na2HPO4, 100 mM NaCl) and stored for 1-2 hours at room temperature on sterile disposable inoculating loops in a container with wet paper towels to provide between 40%-50% humidity.

For delivery of DNA to cell experiments, Ubx was expressed in a pET-19b vector as described above. Cell pellets were lysed in buffer containing 50 mM NaH_2PO_4, 500 mM NaCl, 5% glucose, 20 mM imidazole, 1.5 mM PMSF, 5 mM DTT, 0.2 mg/mL lysozyme, Complete Proteinase Inhibitor Mixture (Roche), and 0.04mg/mL DNase I. The lysis was centrifuged to remove cell debris and purified using Ni-NTA chromatography. Protein was eluted with pH 7 buffer containing 50 mM NaH_2PO_4, 500 mM NaCl, 300 mM imidazole, and 5% glucose.

Fibers were pulled from a tray as previously described (Mendes et al. 2018). In brief, protein elutions were added to a tray with buffer (50mM NaH_2PO_4, 500 mM NaCl, and 5% glucose). The tray was covered and left at room temperature overnight. The film which had formed overnight was pulled to one side of the tray, and fibers were pulled from the surface of the film using disposable inoculating loops.

Immunofluorescence of Ubx fibers: Loops supporting Ubx fibers were placed in the wells of a sterile 24-well cell culture plate and incubated for 1 hour in blocking solution (250 µL PBS with 0.1% Triton X-100, 1% BSA, 0.2% sodium azide, and 5% goat serum) then washed twice for 10 minutes in PBS containing 0.1% Triton X-100. Primary antibody recognizing the Ubx homeodomain (FP3.38) (White & Wilcox 1984) was diluted 1:1000 in blocking solution and incubated in the wells with the loops for an hour. After two washes, loops were incubated with goat anti-mouse secondary antibodies conjugated to Alexa 488 (Molecular Probes), diluted 1:300 in blocking solution, for an hour. Loops were washed twice, placed on a 22 mm x 55 mm coverslip, and imaged using confocal microscopy on a Nikon Eclipse Ti equipped with NIS Elements AR 4.10.01 software for fluorescence intensity analysis.

Preparation of DNAs for binding experiments: DNA sequences can be found in Table 2-2. Oligonucleotides were purchased from Integrated DNA Technologies. The 3' end of the 40Sp/40NSp sense oligonucleotide strand was labeled with Alexa Fluor® 488 (NHS Ester) by Integrated DNA Technologies. Complementary oligonucleotides were annealed as follows**:** Each oligonucleotide was diluted in sterile water to a concentration of 100 µM. The reaction for annealing DNA included 1X PCR buffer (Invitrogen), 20 µM of each oligonucleotide, and 1.5 µM $MgCl_2$. The annealing reaction was then incubated in water bath at 95 °C for 10 minutes. The temperature of the water bath was allowed to cool to less than 40 °C. The annealed DNA was stored at -20 °C until needed.

pET19b-VEGF-UbxIa plasmid (Tsai et al. 2015) was purified from *E. coli* using the Qiagen MidiPrep kit. This plasmid was linearized by restriction digestion with HindIII (New England Biolabs), followed by purification with Zymoclean Gel DNA recovery kit. The *vegf* DNA was amplified from this plasmid using PCR and purified using Zymoclean Gel DNA recovery kit.

Supercoiled 336 bp minicircle was obtained from Twister Biotech, Inc. (Houston, TX). Supercoiled minicircle was digested with EcoRV according to the manufacturer's protocol (New England Biolabs). The resulting linear minicircle was purified using the Qiagen QIAquick PCR purification kit and eluted in 10 mM Tris-HCl, pH 8.0, and 0.1 mM disodium EDTA.

pUC57 plasmid DNA was obtained from GeneScript and produced and purified from *E. coli* using the Qiagen MiniPrep kit.

Binding DNA to the fiber surface: Loops supporting Ubx fibers were placed in the wells of a sterile 24-well culture plate and incubated in 200 µL of fluorophore-labeled DNA (Table 2-1) diluted to 1.2 µM. The plate was wrapped in aluminum foil to prevent photobleaching and incubated at room temperature for 6 hours. Fibers were then incubated in PBS buffer for 3-5 minutes and washed three times in PBS buffer to remove excess DNA before viewing the fibers. For experiments detected by PCR, non-fluorescent DNA was diluted in PBS to a final concentration of 10 µg/mL unless

otherwise specified. Loops supporting fibers were then placed in a well of a 24-well culture plate; subsequently, a volume of 200 μL of the appropriate DNA was pipetted in each well, allowed to incubate at room temperature, and wrapped in parafilm overnight. Following DNA incubation, fibers were incubated in PBS buffer for 3-5 minutes, washed three times in PBS buffer to remove excess DNA, and analyzed by PCR as described below.

Detection of bound DNA by PCR: Ubx fibers bound to DNA were removed from the inoculating loop, using micro-scissors and micro-tweezers, and transferred to a 50 μL PCR reaction containing 1X Pfx AccuPrime Reaction mix buffer (Invitrogen), 50 μM of each dATP, dCTP, dGTP, dTTP, 0.5 μM of each primer, and 10 units of AccuPrime Pfx DNA polymerase (Invitrogen). PCR parameters were: 95 °C for 2 min, then 20 cycles of 95 °C for 2 min, anneal at 59 °C for 30 sec, extension at 68 °C for 45 sec per kb amplified. Reactions were stored at 4 °C prior to analysis by agarose gel electrophoresis, using 1% or 2% agarose gels as appropriate, based on the size of the PCR product.

DNA binding competition experiments: Ubx fibers were placed in the wells of a sterile 24-well culture plate and incubated in 10 ug/mL of the appropriate DNA (100Sp and 100NSp DNA) for 16 hours at room temperature. Fibers were incubated in PBS buffer for 3-5 minutes and washed three times in PBS buffer to remove excess DNA. Fibers were then incubated in the second competitor DNA sequence for 16 hours at room temperature and washed three times before DNA detection by PCR as specified above.

DNA release assay: pET-19b-Vegf-Ubx plasmid DNA (SP) was used as template to insert an additional six 5'-TAAT-3' in the *vegf* gene, creating SP+6 (sequence information in Table 2-2). Ubx fibers were bound to DNA as specified above, washed three times, and incubated in a large volume of buffer to approximating infinite dilution for the designated time, ranging from zero to nine days to monitor DNA release. Each fiber was then analyzed by PCR to detect the *vegf* gene within the plasmid.

DNA protection assay: Ubx fibers were placed in the wells of a 24-well culture plate then incubated for 16 hours at room temperature in LP DNA diluted to a final concentration of 0.02 μg/mL. Fibers incubated in DNA were subsequently incubated in PBS buffer for 3-5 minutes, washed three times in PBS buffer, and then placed in a new 24-well culture plate. Ubx fibers were incubated in a reaction mixture containing DpnI and allowed to incubate for 0, 15, or 30 minutes prior to heat inactivation and analysis by PCR.

Co-assembly of protein and DNA into materials: Fibers co-assembled with fluorescent 40Sp and 40NSp DNA were produced as described above with the following variations: For samples with DNA, the DNA stock was diluted to two-fold higher than the protein concentration in PBS buffer and directly pipetted into the Ubx drop. For samples serving as negative controls in Ubx-DNA co-assembly experiments, an equivalent volume of PBS buffer, in lieu of DNA, was added to the Ubx drop. Fibers drawn from

solution containing both Ubx with and without DNA were washed immediately three times in PBS buffer and analyzed by fluorescent microscopy as described above.

Extended release transformation assay: Ubx fibers were wrapped around a 1 uL disposable inoculating loop, which was siliconized with Sigmacote (Sigma-Aldrich). The loops, with and without fibers, were soaked in 6 ng/μL pUC57 DNA for 30 minutes, rinsed in 1 mL of PBS twice, then placed in 200 μL PBS for a rinse time of 30 minutes to 60 days. Buffer was replaced every 15 days. After rinsing, loops were used to transform DH5α *E. coli* cells (Zymo Research). Transformed colonies were counted. Statistical outliers (1.5IQR (Interquartile Range) away from the average). To ensure the PBS rinse step was a large enough volume, a 7 day rinse experiment following the protocol above was performed, using 2 mL of PBS rather than 200 μL.

CHAPTER III

IMMOBILIZING SPLIT-GFP BIOSENSORS IN PROTEIN BASED MATERIALS

3.1 Introduction

Biosensors are devices that rely on living organisms and/or biological molecules to detect specific chemicals, often disease agents or environmental contaminants (Jarczewska and Malinowska 2020). Biosensors must have specific molecular recognition of the analyte with a mechanism to detect binding. Because biosensors are often either deployed in or designed to detect living organisms, they must also be biocompatible. Proteins offer a unique opportunity to achieve all three goals: proteins can bind other molecules with unparalleled specificity, they can create chemical, electric or optical signals, and they are composed of amino acids, a natural component of all life.

Protein-fragment complementation assays (PCAs) provide a mechanism to link analyte binding with signal generations by a protein. In this approach, a protein into two chains which individually lack function, but reconstitute their native structure and function when combined. Many PCAs have been designed to investigate protein-protein interactions in a wide variety of fields (Romei and Boxer 2019; Nguyen and Silberg 2010). These assays use a variety of readouts, including fluorescence, bioluminescence, cell survival, and gene transcription (Romei and Boxer 2019). In particular, split green fluorescent protein (GFP) systems offer the advantage of a self-reporting fluorescent readout that requires no reagents, which makes on-site detection and analysis possible

using only a blue light source and an orange high-pass filter. The GFP chromophore, generated by post-translational modification of three residues (Crone et al. 2013), lies at the center of an 11 β-strand barrel (Figure 3-1A). When the GFP protein sequence is split, oxygen enters the core and quenches fluorescence by the chromophore. Binding by the excluded portion of GFP sequesters the core and restores fluorescence. This approach unites the sensor and method of detection into a single biocompatible molecule, which can be performed in a single step requiring minimal equipment.

To link fluorescence of the split GFP to analyte binding, the hole created by the missing GFP fragment is redesigned to recognize the analyte. Generally, the two GFP fragments are synthesized or pre-incubated together promote proper folding and chromophore maturation, which can require several hours (Romei and Boxer 2019; Cabantous et al. 2005). Consequently, the small fragment must be removed to regenerate the empty biosensor binding site. The Leave-one-out GFP (LOO-GFP) split GFP system was engineered from the superfolder GFP-OPT protein, a variant of GFP that has been optimized for higher solubility in the leave-one-out state (Cabantous et al. 2005). In LOO-GFP, the protein sequence is circularly permuted and truncated, allowing any β strands that does not contribute an amino acid to the chromophore can to be omitted (Figure 3-1B) (Huang and Bystroff 2009). Variations of LOO*n*-GFP with each of the 11 β strands have been tested, with '*n*' denoting the number of the left-out strand in the original GFP-OPT protein. Without the missing peptide, LOO-GFP monomers exists in a partially unfolded state that fluoresces weakly or not at all. Addition of the left out β-

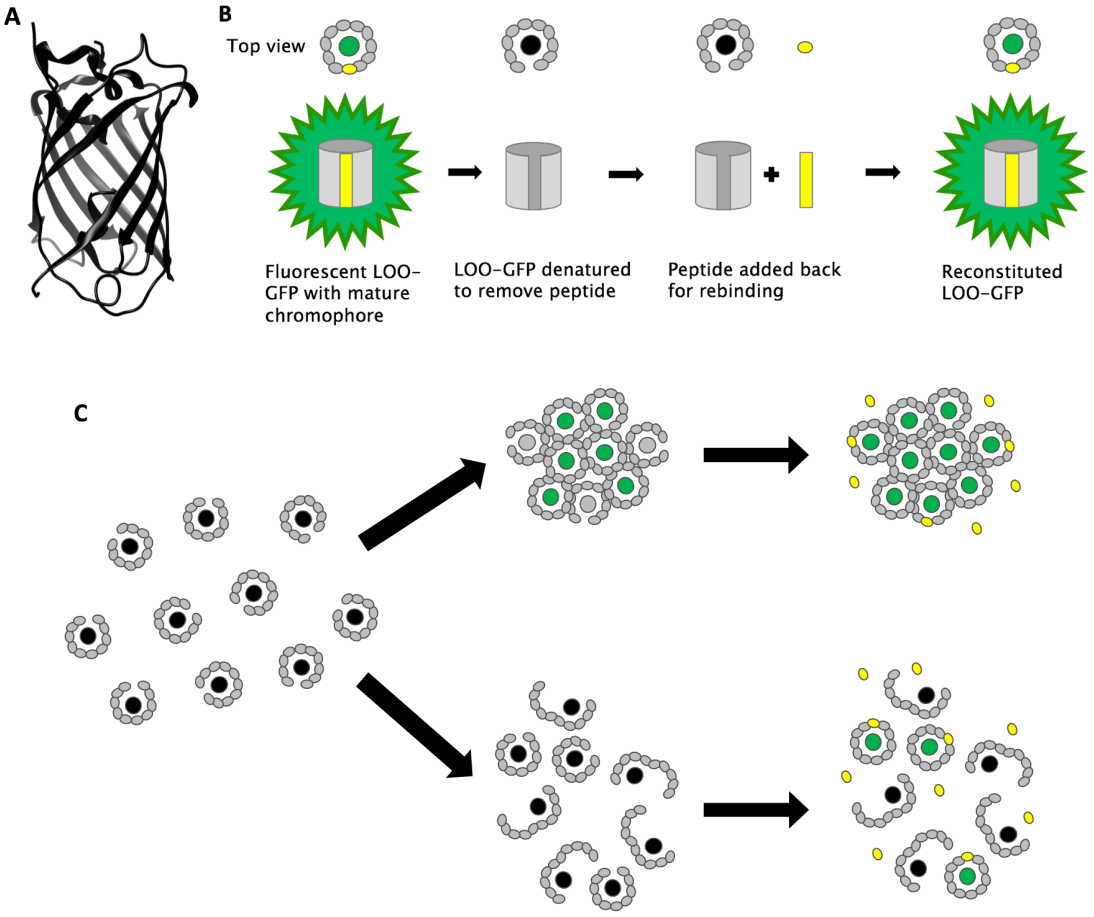

Figure 3-1 LOO-GFP design and basic problems.
(A) Ribbon diagram of LOO8-GFP, from GFP-OPT with strand 8 removed (PDB 2b3p). (B) LOO-GFP overview cartoon. LOO-GFP is co-expressed with its left out peptide for proper folding and chromophore maturation. Denaturing conditions remove the peptide and fluorescence. Peptide rebinding allows for reconstitution of fluorescence. (C) Schematics of basic problems – LOO-GFP monomers aggregate, causing fluorescence even in the incomplete state. Improper folding of incomplete monomers can also prevent full peptide rebinding.

strand enables proper folding and chromophore fluorescence (Figure 3-1B) (Huang and Bystroff 2009), making LOO-GFP a sensor for its left-out peptide. Because the small left out fragment is only one β strand, rather than the 4 or 5 which is typical in split-GFPs, LOO-GFP can be more readily computationally redesigned to fluoresce upon binding a novel target peptide.

Although widely used, split GFP systems commonly encounter two problems (Figure 3-1C). First, the large fragments tend to bind to each other in the absence of the small fragments, sequestering the chromophore and leading to high background (Kodama and Hu 2012; Huang et al. 2015). This problem is aggravated by the preference for high concentrations of the split GFP to increase the signal detected. Second, the split GFP may have difficulty refolding, even in the presence of the missing peptide, lowering the maximum obtainable signal. In this work, will address both of these problems by immobilizing the split GFP biosensor within protein materials.

In this proof-of-concept paper, we test whether immobilization of LOO8-GFP in materials composed of Ultrabithorax (Ubx), a *Drosophila melanogaster* protein, can address these issues. In mild buffers, Ubx self assembles into biocompatible, non-immunogenic materials at the air-water interface (Greer et al. 2009; Patterson et al. 2014; Patterson et al. 2015). The resulting materials are stabilized by spontaneously-forming dityrosine bonds (Howell et al. 2015). Functional proteins can be incorporated into materials through gene fusion, by adding the gene encoding of the functional protein

adjacent to the gene encoding Ubx, with no intervening stop codons (Huang et al. 2011; Tsai et al. 2015). The resulting fusion gene produces a single polypeptide harboring both the LOO8-GFP and the Ubx amino acids sequences. Thus far, all Ubx fusion proteins tested are able to both self-assemble into materials and retain the function of the appended protein (Huang et al. 2011; Tsai et al .2015, Howell et al. 2015). We find that the LOO8-GFP-Ubx fusion protein also self-assembles into materials, and examined the fluorescent characteristics of Ubx fibers with LOO-GFPs incorporated. The fluorescent intensity of LOO8-GFP fibers is improved by immobilization in Ubx fibers. Conditions for peptide removal and rebinding from LOO8-GFP fibers are examined and compared to free monomers. Immobilization in Ubx materials does indeed reduce the levels of background fluorescence. Although LOO8-GFP interactions with Ubx appear to hamper peptide rebinding, these interactions can be reduced by increasing the concentration of salt, suggesting the interactions are ionic in nature.

3.2 Results and Discussion

Protein complementation assays based on GFP often encounter two difficulties: an increase in background signal due to oligomerization of the apo-protein, and mis-folding of the apo-protein, reducing the maximum signal. To address these issues for LOO8-GFP, we fused LOO8-GFP to Ubx (Figure 3-2A), and co-expressed it with the left-out-peptide (strand 8, or s8) to improve protein folding and chromophore maturation. However, attaching LOO8-GFP to self-assembling protein biomaterials has the potential to alter the properties of one or both proteins, so it is necessary to assess both materials

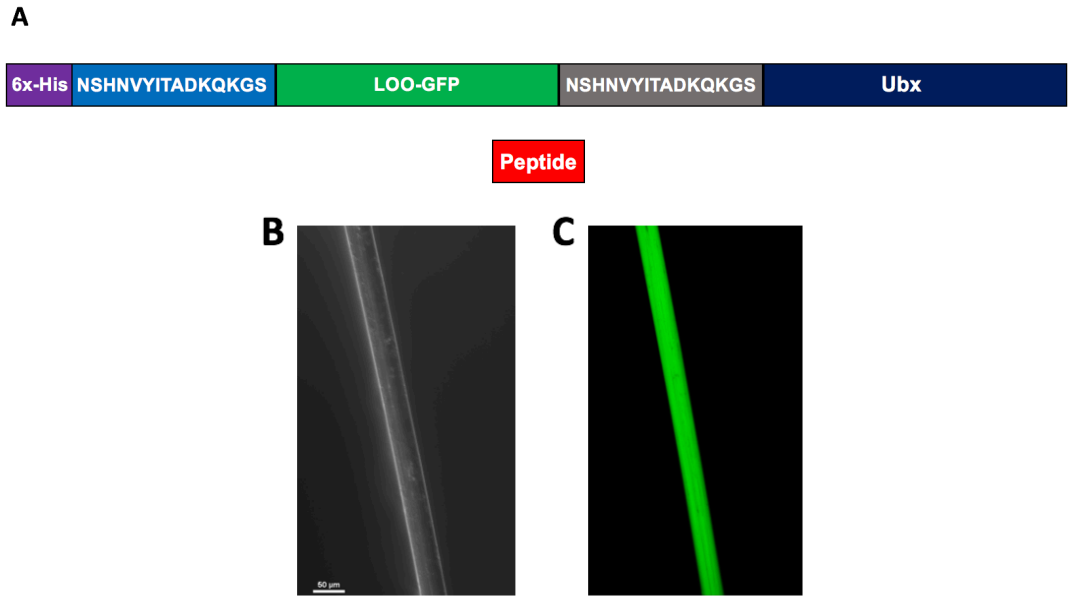

Figure 3-2 LOO8-GFP-Ubx fusion forms normal fibers.
(A) LOO-GFP-Ubx fusion schematic, with linkers between proteins included. (B) Light confocal microscopy image of LOO8-GFP-Ubx fibers shows normal fiber morphology. (C) Fluorescein isothiocyanate (FITC) channel of confocal image of the same LOO8-GFP-Ubx fiber from (B) is green fluorescent. (D) Two fiber pictures- light and green fluorescence of same fiber after strong wash, show we can remove all fluorescence, get rid of background fluorescence.

formation and LOO8-GFP function. Because the strand 8 peptide bound to LOO8-GFP (s8:LOO8-GFP) is less stable than EGFP, a significant concern is whether either fusion or materials assembly will significantly damage this protein. To test whether s8:LOO8-GFP is damaged by fusion with Ubx, we compared the fluorescence of s8:LOO-GFP-Ubx monomers to that of EGFP-Ubx monomers. EGFP is very stable, and is known to retain activity in Ubx monomers and fibers (Huang et al. 2011). We found that purified

s8:LOO-GFP-Ubx monomer fluoresce green, and the absorption and emission spectra of LOO8-Ubx monomers matched that of EGFP-Ubx monomers (Figure 3-3 A,B) and that of unfused EGFP monomers (Figure 3-4). These data suggest that formation of the fusion protein did not harm LOO8-GFP activity.

The fusion protein also self-assembled into green fluorescent fibers (Figure 3-2 B,C). To test whether fusion to P8:LOO-GFP impacted the ability of Ubx to self-assemble into materials, we examined the morphology of the fibers. Conditions that damage Ubx function also harm the ability of Ubx to self-assemble into materials and thus impact fiber structure. For instance, Ubx truncations mutants generate weak and/or malformed fibers (Greer et al. 2009) and Ubx films drawn through organic solvents create mis-shapen fibers (Majithia et al. 2011). The morphology of the s8:LOO-GFP-Ubx fibers was typical for Ubx materials. Thus, the normal appearance of LOO8-GFP fibers suggests that fusion to LOO8-GFP did not impact Ubx assembly. This result is consistent with data from materials composed of other Ubx fusion proteins (Tsai et al. 2015).

Finally, the absorption and emission spectra of fibers composed of LOO8-GFP-Ubx matched those of fibers composed of EGFP-Ubx (Figure 3-3 A,B). Thus, immobilization in Ubx materials did not appear to alter the function of immobilized LOO8-GFP relative to its more stable counterpart, immobilized EGFP. Interestingly, the fluorescence emission spectra of both LOO8-GFP and EGFP immobilized in fibers is

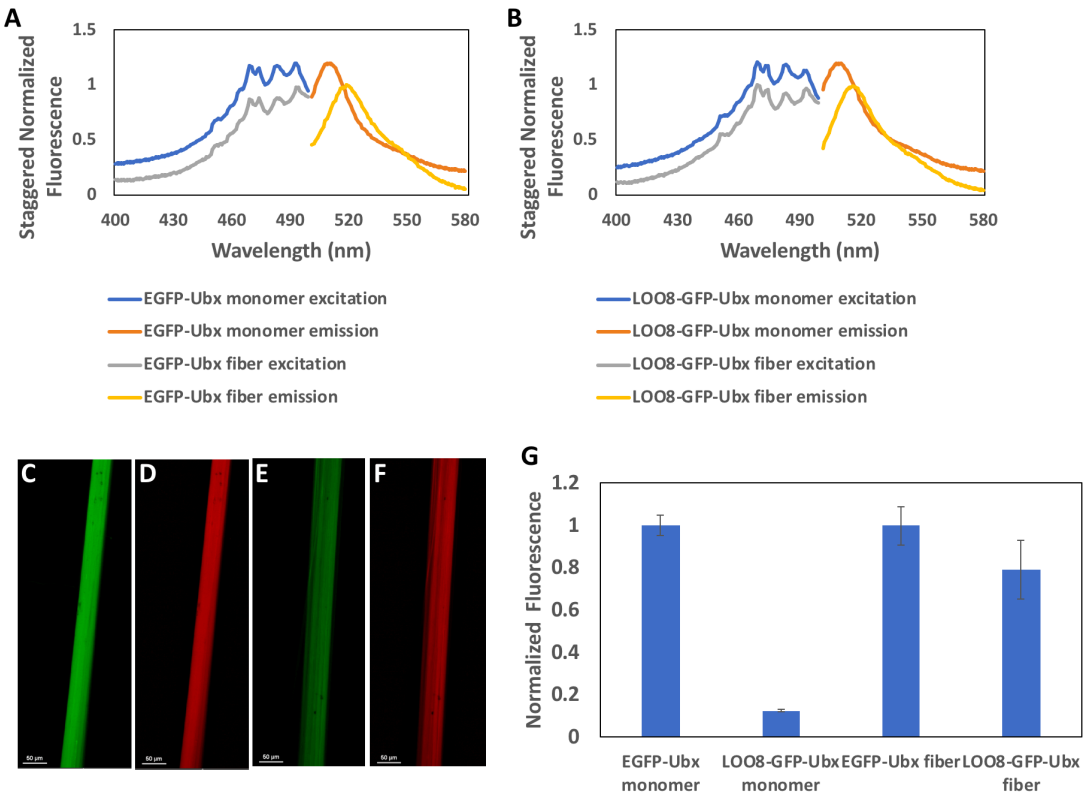

Figure 3-3 Immobilization on fibers increases the fluorescent intensity of LOO8-GFP.

A-B. Absorption and emission of monomer and fiber EGFP-Ubx (A) and monomer and fiber LOO8-GFP-Ubx (B). C-D. Confocal microscopy images of 50%/50% mCherry-Ubx and EGFP-Ubx on FITC channel (C) and TRITC channel (D). E-F. Confocal microscopy images of 50%/50% mCherry-Ubx and LOO8-GFP-Ubx on FITC channel (E) and TRITC channel (F). (G). Quantitative relative intensity data compiled from A-F.

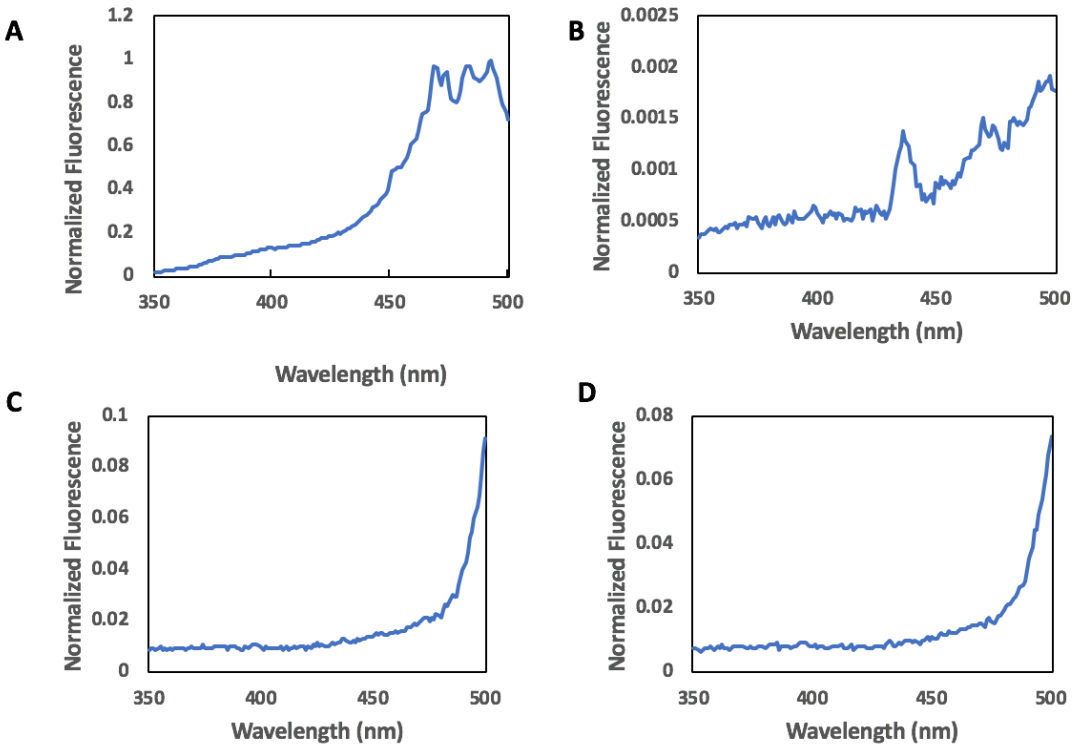

Figure 3-4 Excitation spectrum controls show Ubx fusion proteins are similar to EGFP excitation.

There are multiple extra peaks in the excitation spectrum for LOO-GFP-Ubx and EGFP-Ubx, in both the monomers and the fibers, compared to previously published EGFP monomers (Figure 3-3 a). However, EGFP monomer excitation (A) measured on the same instrument used for the above experiments has similar extra peaks. Ubx monomers (B) and Ubx fibers (C), however, do not show any significant peaks with the same excitation and emission settings, nor does the paperclip used as a scaffold to hold the fibers in the light path (D). The small peak just before the cutoff for the excitation is likely just scatter of the beam.

red-shifted relative to the emission spectra of both monomers. A likely explanation is that materials formation places charged amino acids near the fluorescent proteins, thus shifting the spectrum (Vivian and Callis 2001). The Ubx protein includes a homeodomain, in which 11 of the 60 amino acids are lysine or arginine. This phenomenon has previously been observed for dityrosine bonds in Ubx, in which the blue fluorescence emission peak is shifted from the normal range of 410-430 nm to 438 nm (Howell et al. 2015).

Split fluorescent proteins typically have a lower quantum yield than the original protein due to a less stable β-barrel structure (Koker, Fernandez, and Pinaud 2018). One approach to assessing the stability of the LOO-GFP protein is to measure fluorescent intensity. Because the removed peptide is non-covalently attached to the protein, the LOO-GFP exists in an equilibrium between the bound and free states. When the peptide is unbound, the protein is unable to fluoresce, and so its quantum yield is 0. The less stable the protein is, the less time that protein spends folded and thus able to bind the peptide. Consequently, unstable LOO-GFP variants spend most of their time in the unbound state, reducing the fluorescent intensity of that sample. Indeed, the quantum yield for s7:LOO7-GFP, in which strand 7 is replaced by a peptide, was 40% that of the parent protein (Huang and Bystroff 2009). Comparing the fluorescent intensities of EGFP-Ubx and s8:LOO8-GFP-Ubx, normalized for concentration, revealed that s8:LOO8-GFP-Ubx monomers were only 12.5% as fluorescent as EGFP-Ubx (Figure 3-3 G).

We hypothesized that incorporating a LOO-GFP into Ubx materials could stabilize the LOO-GFP, and thus increase the protein's fluorescence, which increases the maximum possible signal from the sensor. Ubx materials had previously been shown to stabilize proteins remarkably well; mCherry fibers retain fluorescence after incubation in ethanol for 30 minutes or after being autoclaved for 20 minutes (Tsai et al 2015). We compared the fluorescent intensity of two fibers by confocal microscopy. To normalize the signal for the amount of fiber in the image, we divided the green fluorescent intensity generated by s8:LOO-GFP-Ubx by the red fluorescent intensity generated by an equal concentration of mCherry-Ubx present in the Ubx materials. Using this approach, we determined that the s8:LOO-GFP-Ubx fibers are 79% as fluorescent as EGFP-Ubx fibers (Figure 3-3 C-G). This represents a >6-fold increase in fluorescence relative to protein monomers, a substantial increase in both protein stability and the maximum signal that the protein can provide.

The left-out peptide s8 is co-expressed with LOO8-GFP-Ubx to improve protein folding and chromophore maturation. Thus this peptide must be removed from the materials before the fibers can be used as a biosensor. Previous LOO-GFP studies were removed peptide by exposing the monomers to either 6 M guanidinium hydrochloride (GdnHCl) or reducing the pH to 2 (Huang et al. 2009; Huang et al. 2015). Both methods were also able to remove peptide from fibers (Figure 3-5), although we were unable to recover fluorescence with the addition of peptide when the fibers had been incubated in the guanidinium hydrochloride. Likewise, a 40 minute denaturation step at pH 2 removes

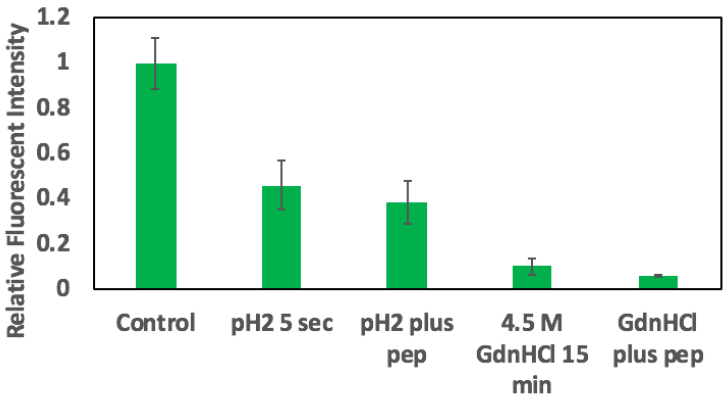

Figure 3-5 No fluorescence is recovered after low pH or GdnHCl wash.
Fluorescent intensity of fibers is shown after no treatment (control), a 5 second pH 2 wash with and without peptide added, and a 15 minute GdnHCl wash with and without peptide added added after. No recovery of fluorescence is observed in either condition.

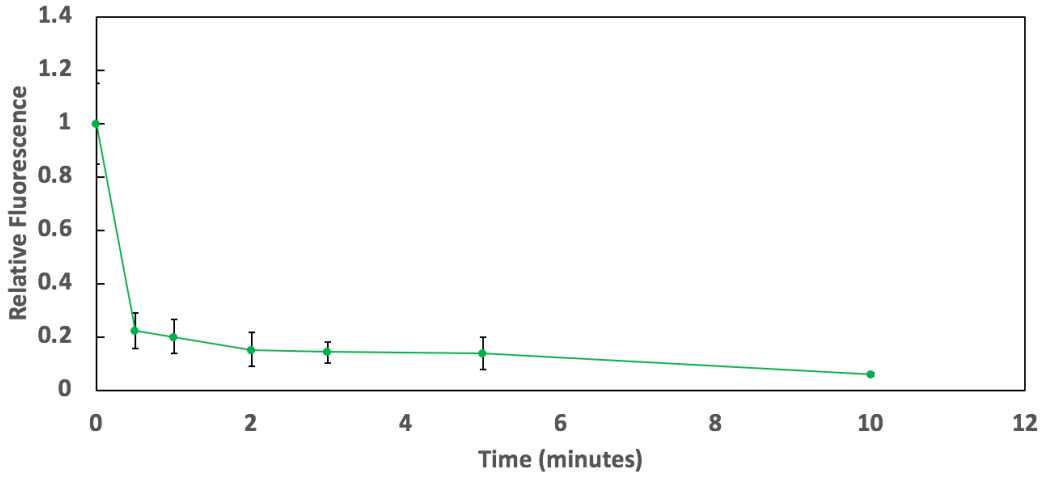

Figure 3-6 Most fluorescence is removed quickly in a pH 2.2 wash.
Fluorescent intensity after a pH2.2 denaturation wash time for 30 seconds, 1, 2, 3, 5, and 10 minutes. Most fluorescence is removed within the first minute.

all peptide, but we observed no peptide rebinding. To reduce this problem, we examined the loss of fiber fluorescence versus incubation time at pH 2.2. A shorter incubation in denaturing conditions is expected to facilitate LOO-GFP refolding by reducing conformational drift, which has been proposed to be an issue with LOO-GFPs (Huang and Bystroff 2009). Indeed, 5 seconds in the denaturation buffer allowed most of the LOO8-GFP to fluoresce upon addition of peptide (Figure 3-5). However, these very short denaturation times may not completely remove the peptide, especially since the peptide must diffuse out of the fiber. Consequently, we examined the amount of background fluorescence remaining as a function of the incubation time in the denaturing buffer (Figure 3-6). Since most of the protein is removed in the first minute, we adopted a standard 1-minute denaturation time for all subsequent experiments.

The final challenge is to identify conditions in which the peptide reliably rebinds the LOO8-GFP-Ubx fiber, creating the fluorescent signal. Our initial studies using PBS (20 mM NaH2PO4, 80 mM Na2HPO4, 100 mM NaCl) as the rebinding buffer did not reveal any peptide rebinding above background levels. An important clue was provided by a prior study, in which ionic interactions between Ubx (predicted net charge +9) with EGFP (predicted net charge -2) were proposed to the be source of the increased strength of EGFP-Ubx fibers, relative to fibers composed of only Ubx (Huang et al. 2015). Many of the negatively charged residues in LOO8-GFP surround the gap created when the P8 peptide was removed (Figure 3-7). We hypothesized that, in the absence of peptide, this

negatively charged face of LOO8-GFP binds Ubx, which then physically blocks peptide re-binding. If true, increasing the salt concentration should alleviate this effect.

Consequently, we tested the impact of salt concentration on peptide rebinding. We repeated peptide binding experiments with different concentrations of salt added to all buffers (denaturation, wash, and peptide rebinding) (Figure 3-8). For each salt concentration, we compared samples in which peptide was added in the final step with sample in which the peptide was omitted to assess the levels of background fluorescence. While both the background fluorescence and the signal increased with increasing salt concentration, the increase was larger for the samples in which peptide had been added. As a result, a statistically significant difference between samples with and without peptide was observed for several elevated salt concentrations. These results confirm that the LOO8-GFP biosensors work when immobilized within protein fibers.

3.3 Conclusions

We conclude that protein biosensors can work when immobilized throughout protein-based materials. The sensor protein did not interfere with materials assembly. However the materials did alter sensor function. On the positive side, immobilization in materials stabilized the sensor, allowing >6-fold fluorescent signal when compared with monomers. However, interactions between APO-LOO8-GFP and Ubx also hampered

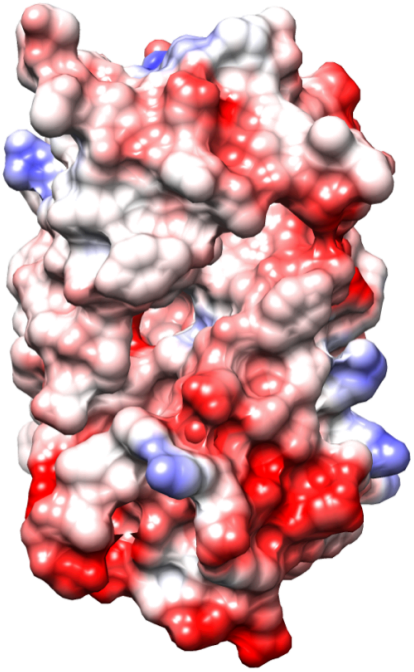

Figure 3-7 Surface charge around the binding pocket for the left out peptide. There are multiple areas of negative charge around the binding pocket in LOO8-GFP, which may interact with the positively charged Ubx. These interactions would block peptide rebinding.

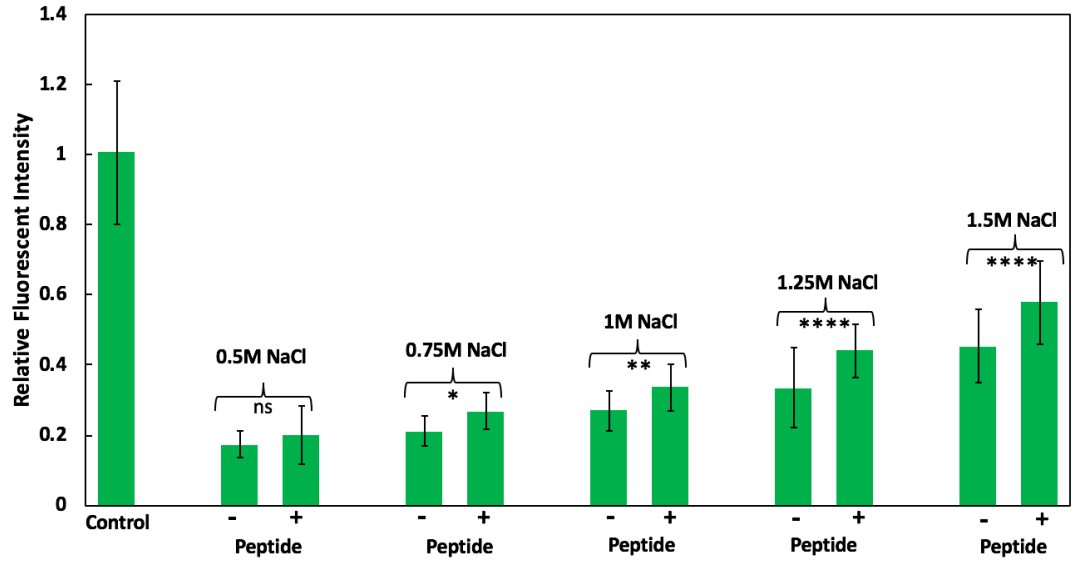

Figure 3-8 NaCl concentration is critical for peptide rebinding.
Fluorescent intensity is shown before and after peptide rebinding in denaturing, wash, and rebinding conditions of varying NaCl concentrations. Higher NaCl concentrations have statistically significant amounts of peptide rebinding. (ns is not significant; * is $P \leq 0.05$; ** is $P \leq 0.01$; **** is $P \leq 0.0001$)

peptide rebinding. While this problem could be partially alleviated by increasing the salt concentration, an ideal sensor would function in more physiological conditions.

3.4 Materials and Methods

Construction of plasmids: To create the LOO-GFP-Ubx construct, pCDFDuet-1—LOOn-GFP was used as the parent vector (Huang 2011). Ubx splicing isoform Ia (GenBank AAN13718.1) was inserted between BamHI and EcoRI sites in the pCDFDuet-1 plasmid, forming LOOn-GFP-Ubx in MSC1. MSC2 contains the left out

peptide “n” fused to the carrier protein Ssp-DnaB mini-intein, as previously described (Huang 2011).

pET19b—UbxIa and pET19-b—EGFP-UbxIa plasmid were used. As previously described (Tsai 2015), this construct contains the Ubx mRNA splicing isoform Ia (GenBank AAN13718.1) inserted between the NdeI and BamHI sites of pET19b vector (Novagen), with EGFP inserted in the NdeI site between the N-terminal His-tag and Ubx.

Protein expression, purification, and materials production: Protocols were used as established in the Bondos lab for expression and purification of protein fusions into materials. [Greer 2009, Tsai 2015]. In brief, the construct (pCDFDuet-1—LOO-GFP-Ubx, pET19b—Ubx, or pET19b—EGFP-Ubx) was transformed into Rosetta (DE3) pLysS cells (Novagen). Single colonies were used to inoculate overnight liquid cultures. Protein expression was induced at mid-log phase with 1×10^{-3} M isopropyl-β- D -1-thiogalactopyranoside (IPTG) for 5 h, then the cells were harvested by centrifugation and stored at −20 °C. Frozen cell pellets were lysed according to established protocols [Patterson 2014, Tsai 2015], in a lysis buffer containing 50 mM NaH_2PO_4, 500 mM NaCl, 5% glucose, 20 mM imidazole, 1.5 mM PMSF, 5 mM DTT, 0.2 mg/mL lysozyme, Complete Proteinase Inhibitor Mixture (Roche), and 0.04mg/mL DNase I, pH 8.75. Cell debris was removed by centrifugation for 10 min at 4°C, 9,000 rpm. Protein was purified from the clarified cell lysate by HisPur™ Ni-NTA Superflow Agarose

(ThermoFisher Scientific), which was pre-equilibrated with wash buffer containing 50 mM NaH_2PO_4, 500 mM NaCl, 5% glucose, pH 8.0. Bound resin wash washed with wash buffer and wash buffer containing 20 mM imidazole. Protein was eluted using a buffer containing 50 mM NaH_2PO_4, 500 mM NaCl, 5% glucose, and 300 mM imidazole, with a pH of 7.0. The concentration of purified protein samples were determined using the BioRad protein assay (BioRad). Fibers were pulled as previously described from films produced in a "buffer reservoir" (Mendes et al. 2018) with 1-2 mg of protein per reservoir, using the wash buffer described above. All fibers used for peptide removal and rebinding experiments were 75% UbxIa and 25% LOO-GFP-Ubx, unless otherwise noted. After overnight incubation at room temperature (approximately 25 °C) , fibers were wrapped around a 5 mm sterile plastic inoculation loop and stored in a sterile tissue culture dish until use.

Spectroscopic measurements: Fluorescence spectroscopy was performed using a JASCO-815 fluorometer and a 10-mm pathlength quartz cuvette. Proteins were in wash buffer, described above, at a concentration of 250 nM, Excitation spectra at 488 nm and emission spectra read at 488 nm were monitored at 25 °C using a bandwidth of 1 nm for samples in solution and 2 nm for fibers, at a scan rate of 100 nm min−1. For LOO-GFP-Ubx chimeric fibers, fluorescence excitation and emission spectra were obtained by placing fibers in the excitation path, and the maximum peak intensity was defined as 1. Fluorescent intensity of monomers was measured using a Cary Eclipse Fluorescence Spectrophotometer and a 10-mm pathlength quartz cuvette. Proteins were in wash buffer

at 500 nM and 750 nM, for EGFP-Ubx and LOO8-GFP-Ubx, respectively. Emission spectra were monitored at 488 nm at 25°C using a bandwidth of 5 nm and a scan rate of 600 nm min−1. Intensity was corrected for the difference in concentration. Fluorescent intensity of fibers was measured using confocal microscopy on a Nikon Eclipse Ti equipped with NIS Elements AR software, as described below. To compare the intensity of EGFP fibers to LOO8-GFP fibers, chimera fibers made of 50% mCherry-Ubx and 50% GFP construct were made. mCherry fluorescence was used as a control fluorescence level between fibers.

Denaturing wash time assay: Loops with 25% LOO-GFP fibers were placed in a 24-well plate and denatured for 0.5, 1, 2, 3, 5, or 10 minutes to remove bound peptide with 250 μL of 0.05M glycine-HCl buffer, pH 2.2, containing 0.5 M NaCl. Fibers were rinsed by placing the loops in two different rinse wells containing high salt PBS (20 mM NaH2PO4, 80 mM Na2HPO4, 1.25 M NaCl), pH 8.0 Fibers were then placed either in high salt PBS or 0.12 mM synthetic left out peptide (GenScript) diluted in high salt PBS and left overnight at 25°C to bind peptide before imaging as described below.

High NaCl assay: Loops with 25% LOO-GFP fibers were placed in a 24-well plate and denatured for 1 minute to remove bound peptide with 250 μL of 0.05 M glycine-HCl buffer, pH 2.2, containing x M NaCl, with x corresponding to the stated NaCl concentration (0.5, 0.75, 1, 1.25, or 1.5 M NaCl). Fibers were rinsed by placing the loops in two different rinse wells containing phosphate buffered saline (PBS) (20 mM

NaH2PO4, 80 mM Na2HPO4, xM NaCl), pH 8.0, with x corresponding to the same NaCl concentrations used in the denaturing wash. Fibers were then placed either in xM high salt PBS (again corresponding to the NaCl concentration used in the denaturing wash) or 0.12 mM synthetic left out peptide (GenScript) diluted in high salt PBS and left overnight at 25 °C to bind peptide before imaging as described below. Synthetic LOO8-GFP peptide was dissolved in a small volume of dimethylformamide and diluted with PBS to a stock concentration, then stored at -80 °C until use.

Confocal images and fluorescent intensity measurements: Fibers were rinsed in the 24 well plate by removal and replacement of buffer with PBS. For imaging, loops with fibers were transferred from the 24 well plate to a glass slide and imaged using confocal microscopy on a Nikon Eclipse Ti A1R equipped with NIS Elements AR software to normalize fiber width and for fluorescent intensity analysis. Z-stack images were captured using a 20x objective with a step size of 1.025 μM to obtain measurements throughout the fiber. Fluorescent intensity of LOO-GFP or EGFP was normalized using fiber width and dityrosine fluorescence, measured in the 4′,6-diamidino-2-phenylindole (DAPI) channel, which accounts for the amount of materials present.

CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

4.1 Conclusions and Future Direction: DNA Delivery

Ubx materials are non-immunogenic, biodegradable, biocompatible, and have stabilizing effects on incorporated proteins (Patterson et al. 2014; Patterson et al. 2015; Hsiao et al. 2016; Huang et al. 2011; Tsai et al. 2015). These unique properties and the ability to incorporate functional proteins make the materials ideal for use in a variety of fields. In this dissertation, I explored using Ubx materials in the fields of DNA delivery and biosensing.

I found that Ubx materials can reversibly, noncovalently bind and protect DNAs harboring the Ubx target DNA sequence. Ubx fibers bind DNA sequence specifically, and DNA with multiple binding sites has a slower rate of release from fibers. Ubx fibers have sustained release of DNA for up to 60 days, depending on the DNA sequence used. These results prove the Ubx materials can form protein-DNA composite materials that are tunable and stable, with a high affinity for DNA that leads to extended release of DNA. The materials have the potential for use in nucleic acid technologies such as therapies.

Since Ubx materials have been proven to bind DNA and release it over a long period of time, the next step is to deliver them to eukaryotic cells. I propose to test the ability of

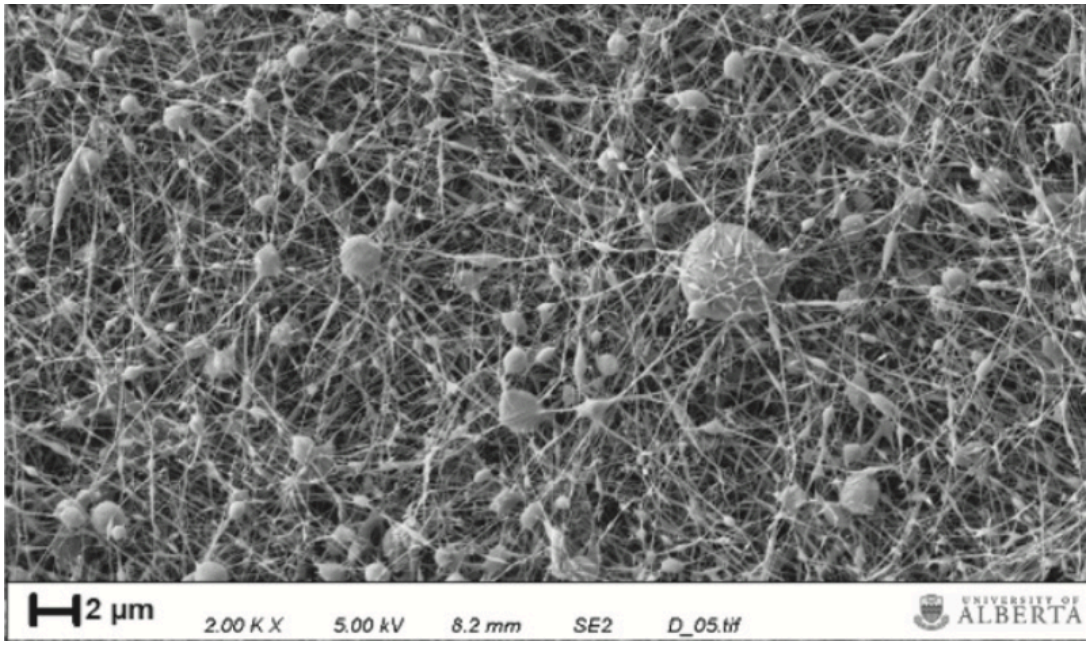

Figure 4-1 An electrospun mat containing PEI/DNA complexes.
An electrospun gelatin-PEG-50 fiber mat containing pDNA/PEI complexes capable of transfecting cells. Image reprinted with permission from Pankongadisak P, Tsekoura E, Suwantong O, and Uludag H. Electrospun gelatin matrices with bioactive pDNA polyplexes. International Journal of Biological Macromolecules 149: 296-308. Copyright 2018 by Elsevier Inc.

Ubx materials to transfect eukaryotic cell lines using bound DNA, but careful consideration is needed to design a successful experiment. Transforming E. coli cells is much easier than transfecting eukaryotic cells, which do not take up extracellular DNA as readily as bacteria. Plasmids have difficulties transfecting eukaryotic cells and often require precise "help" to make it into the cell from delivery vehicles such as cationic lipids or polymers (Hardee et al. 2017). Some of this difficulty lies in the large size

typical of bacterial plasmids; a large vector size leads to lower transfection efficiency in mammalian cells (Gaspar et al. 2014). One way to improve transfection efficiency with Ubx materials would be to include a cationic lipid, such as polyethylenimine (PEI), in the materials. This may be possible by electrospinning Ubx materials with PEI, which is commonly used in electrospinning protocols, and has recently been used to spin bioactive materials containing plasmid DNA (pDNA) (Figure 4-1) (Pankongadisak et al. 2020).

Another way to improve the transfection efficiency of Ubx materials would be to use smaller DNA constructs. Minicircles or minivectors (Figure 4-2), which are small DNA constructs carrying one gene or microRNA, have a much better transfection rate than larger plasmid DNA (Zechiedrich and Fogg 2019; Hardee et al. 2017). Their smaller size improves transfection efficiency, and removal of the bacterial origin of replication and antibiotic resistance gene make them safer for use in mammalian cells (Gaspar et al. 2014). It is also possible to include sequences in the vector to increase gene transfer in cells, such as by adding tissue- or organelle-specific targeting signals or promoters, or by including scaffold attachment regions (Gaspar et al. 2014).These characteristics of minicircles make them an ideal form of nucleic acid to use in future transfection experiments with Ubx materials.

A longer term goal for this project would be to use Ubx materials as the delivery vehicle for DNA in nucleic acid therapies. The type of Ubx materials (fibers, films, hydrogels,

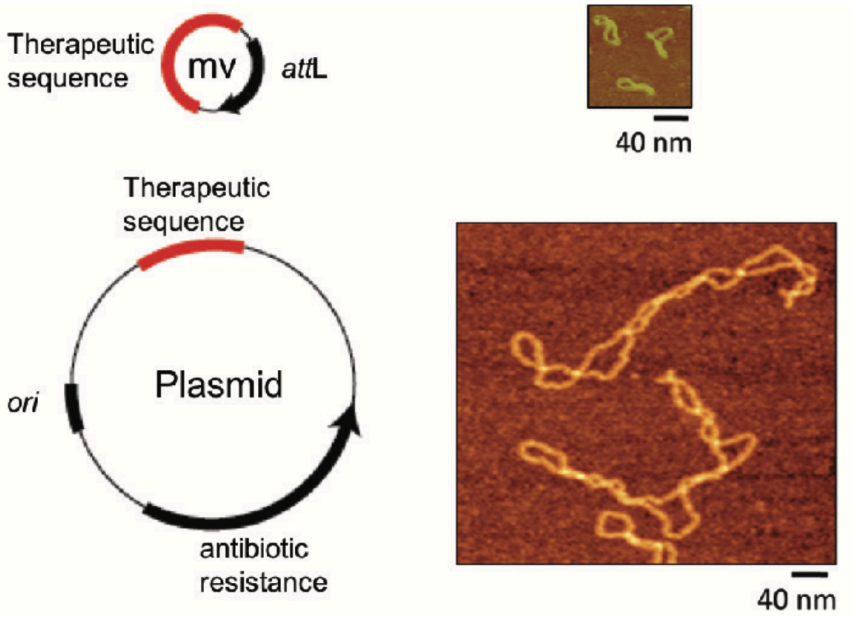

Figure 4-2 Minivector size compared to plasmid size.
Minivectors are up to 50 times smaller than plasmids. Figure reprinted with permission from Zechiedrich L and Fogg JM. Biophysics meets gene therapy: How exploring supercoiling-dependent structural changes in DNA led to the development of minivector DNA. Technology and Innovation 20: 427-440. Copyright 2019 by Cognizant LLC.

particles) used would determine how this technology could be applied, but the extended release of DNA is ideal to lengthen time between therapy treatments/application. Once Ubx materials have been developed for delivering DNA to cells, the materials could be developed as hybrid treatments for tissue engineering. Currently, proteins such as growth factors are fused to the materials to elicit a response from cells in this application.

Growth factor materials could also carry DNA for the tissue engineering application, which may prolong the pro-angiogenic effects of the fiber.

Other applications could also benefit from materials that bind DNA. The controlled orientation of DNA is important for applications such as sensors and nanodevices (Villalonga, Perez-Calabuig, & Villalonga 2020; Kong et al. 2020; Chakraborty et al. 2016). Nucleic acid aptamers are used as biorecognition elements in some biosensors, where the orientation of the displayed DNA is critical for proper recognition function (Villalonga, Perez-Calabuig, & Villalonga 2020). The biocompatible and biodegradable aspects of Ubx materials may be suitable for nucleic acid vaccines and other therapies, where compatibility with the human body is necessary (Dobrovolskaia 2019).

4.2 Conclusions and Future Directions: Biosensors

Split GFPs are useful as sensors, but have problems. I tested immobilization of LOO-GFP sensors on Ubx materials as a potential method to overcome the limitations associated with split proteins. I found that immobilization on Ubx materials stabilizes LOO-GFP without negatively impacting the chromophore and improves the maximum relative fluorescent intensity. Rate of peptide removal is sensitive to NaCl concentrations. Higher NaCl concentrations during peptide rebinding lead to a higher recovery of fluorescence, likely by screening electrostatic interactions between LOO-GFP and Ubx. Ubx offers a transparent support system for LOO-GFPs and has the

potential to maintain LOO-GFP stability and binding affinity while preventing LOO-GFP aggregation.

Although the immobilized LOO-GFPs used earlier in this dissertation offer an improvement in the relative fluorescent intensity of the construct, there are issues with refolding the LOO-GFP on the materials that must be overcome to make them useful. Further testing of denaturing and refolding conditions may identify a more successful experimental setup, and lead to a better understanding of what is hindering recovery of fluorescence.

Although the versions of LOO-GFP used in this dissertation are simple proof of the flexibility of this concept, it is possible to computationally redesign LOO-GFPs to bind other proteins of interest (Figure 4-3). Our collaborator, Dr. Chris Bystroff (Rensselaer Polytechnic Institute), and his laboratory have worked on designs that bind proteins from influenza virus (Huang 2015), dengue fever virus, and recently, SARS-CoV-2. Functioning LOO-GFP-Ubx materials could be used as biosensors for point of care medical testing. The final version of these biosensors would combine fibers that sense different targets derived from a single virus, with a positive result only coming from signal from every experimental (vs. control) fiber. This would increase the specificity of the sensor.

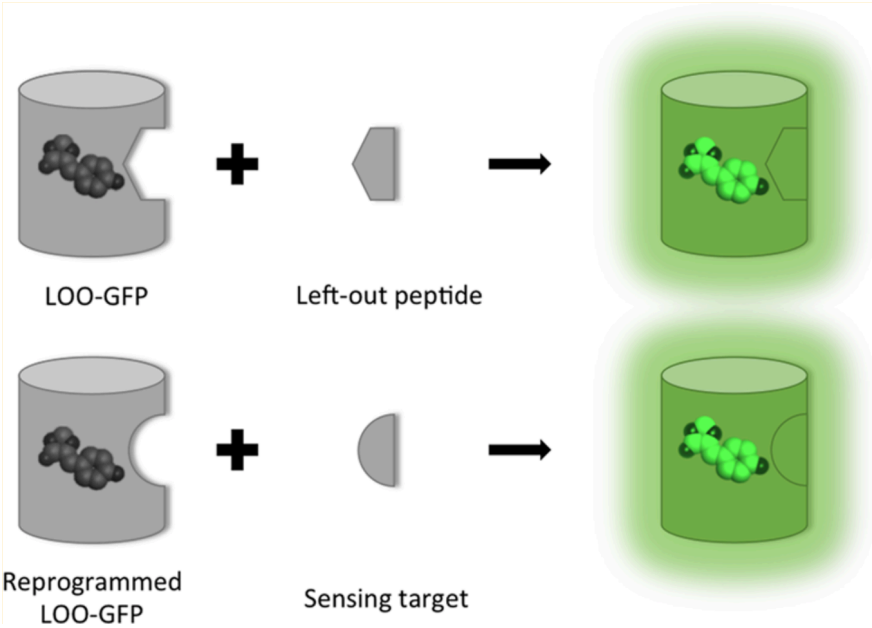

Figure 4-3 LOO-GFP biosensor design.
Removal of a β-strand eliminates the signal. Rebinding of the strand recovers fluorescence. LOO-GFP can be redesigned to sense other target proteins. Reprinted with permission from Huang YM, Banerjee S, Crone DE, Schenkelberg CD, Pitman DJ, Buck PM, and Bystroff C (2015) Toward Computationally Designed Self-Reporting Biosensors Using Leave-One-Out Green Fluorescent Protein. Biochemistry 54: 6263-6273. Copyright 2015 by American Chemical Society.

Writing my dissertation during the SARS-CoV-2 public health emergency has made it clear that there is an urgent need for quick, accurate testing for presence of virus, and in the future we need quick ways to develop these tests for new pandemics. The tests based on PCR are time consuming and require samples to be sent to distant laboratories for completion, which leads to a very slow turnaround time for results. This slow turnaround time is a significant problem with identifying and tracing cases of a novel virus like

SARS-CoV-2 (Babiker et al. 2020). There were also shortages of supplies needed to perform the tests, such as reagents and collection swabs. A viral test that can be completed on-site with minimal chemicals and equipment required would speed up the process and increase the reliability of testing. LOO-GFP can be quickly redesigned to bind viral peptides, and once there are established methods for testing new designs in the materials, it may be a useful tool against new viral outbreaks.

In addition to testing medical samples for presence of virus, LOO-GFP materials may also be used to monitor zoonotic mosquito-borne viruses like dengue fever virus and zika virus. Currently, mosquitoes are trapped in the field in a variety of traps, and must be sent to a laboratory for testing. Most labs use RT-PCR or qPCR to detect if mosquitoes carry dengue or zika (Figueiredo et al. 2013; Leandro et al. 2019; Gourinat et al. 2015). Sentinel chickens can also be screened using antibody tests to determine if they have been infected (Unlu et al. 2009; Burkett-Cadena et al. 2016). An inexpensive, reliable, single-step field test for these viruses would save time and money. LOO-GFP materials could be placed in a mini-device to test homogenized mosquito samples at the site of collection, or other desired locations, without the need for specialized equipment or a time delay to send samples off.

4.3 Other Ideas and Applications for Ubx Materials

4.3.1 Benefits of Immobilization

As discussed earlier, immobilization of LOO-GFPs on Ubx materials improved the

characteristics of the sensor and prevented aggregation of multiple incomplete LOO-GFPs. Immobilization of proteins is frequently used in many applications due to the benefits it provides to the proteins. For example, immobilization controls the orientation a protein is held in while protecting the protein from degradation. Controlled orientation of proteins is key to the performance of biosensors and biochips. Proteins in these applications must be held with the binding site available (Liu & Yiu 2015). Immobilization stabilizes and improves the functional properties of enzymes, and is frequently used to enhance industrial biocatalytic processes (Cowan & Fernandez-Lafuente 2011). The use of Ubx fusion protein materials for immobilization of enzymes and other proteins would allow for controlled orientation through a stable peptide bond between the materials and the protein of interest.

4.3.2 Ubx for 3D Printing

The use of proteins to make bioactive printed structures had increased in recent years. 3D-bioprinting has been used to develop scaffolds for tissue engineering, implants, and drug delivery systems (Chia & Wu 2015). Ubx can be investigated as a bioprinting ink. If Ubx materials are compatible with this fabrication method, they would be particularly useful as printed scaffolds for tissue engineering and as drug delivery systems. Using Ubx growth factor fusion proteins, it would be possible to print a 3D scaffold to regrow tissues with stem cells for personalized medicine. The location of each type of growth factor-Ubx materials could guide growth and development of the cells around the scaffold.

4.3.3 Fusions for Other Applications

There are millions of proteins that exist in nature, many with very unique characteristics and functions. The ability to fuse other proteins to Ubx materials makes the possible applications of this material endless. Fusion of proteins with exclusive, distinctive functions will add those functions to Ubx materials, opening the door for new research directions. For example, one protein that would be interesting to fuse to Ubx materials is the cephalopod protein reflectin.

Reflectin proteins are responsible for many of the remarkable optical components of iridocyte and leucophore cells, which are responsible for the iridescence and color-changing camouflage for which cephalopods are famous for (Figure 4-4A)(Chatterjee et al. 2018). Reflectins have a conserved (M/F-D-X_5)(M-D-X_5)$_n$(M-D-$X_{3/4}$) motif separated by disordered linkers high in charged and aromatic amino acids. Reflectins form *in vivo* structures within reflective cells, which are responsible for many of the optical characteristics of the cells.

Reflectin is similar to Ubx in many ways. Reflectins tend to aggregate, are sensitive to changes in salt concentration, lack secondary structure, and have unique amino acid sequence (Naughton et al. 2016). Similar to Ubx, reflectins self-assemble to form structures in vitro (Figure 4-4 B and C), though unlike Ubx, micro- and nano-structures also form in vivo. Reflectin monomers aggregate into nanoparticles, which then form larger ensembles such as films, which can be drawn into fibers (Chatterjee et al. 2018).

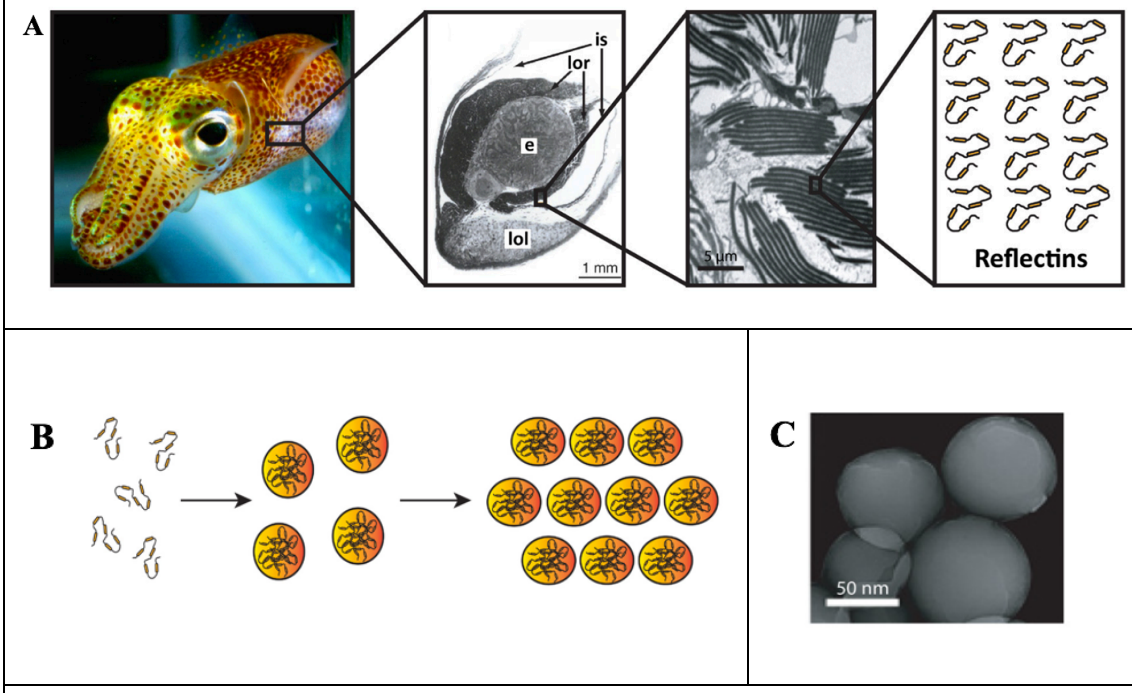

Figure 4-4 Reflectins *in vivo* and *in vitro*.
(A) Far left: a picture of the Hawaiian bobtail squid *E. scolopes.* Middle left: a light micrograph of an *E. scolopes* light organ cross section, where the ink sack diverticula (is) surrounds the reflector (lor), which in turn surrounds the central epithelium(e). The lens tissue (lol) on the ventral surface of the light organ is also shown. Middle right: a cross-sectional transmission electron microscopy image of reflectin platelet stacks within the light organ of *E. scolopes*. Far right: an illustration of the reflectin proteins contained within the platelets. (B) An illustration of the assembly of reflectins into nanoparticles, followed by aggregation of the nanoparticles into larger multimeric ensembles. (D) A transmission electron microscopy image of aggregated nanoparticles from *E. scolopes* reflectin. Figure reprinted with permission from Chatterjee et al. 2018. Chatterjee A, Norton-Baker B, Bagge LE, Patel P et al. (2018) An introduction to color-changing systems from the cephalopod protein reflectin. Bioinspiration and Biomimetics 12: 045001. Copyright 2018 by IOP Publishing Ltd.

Reflectin materials support proliferation and differentiation of cells—neuronal stem cells, are responsive to pH, ionic strength, humidity, and charge, and films can be used for stimulus-responsive infrared camouflage, color changing coatings (Naughton et al. 2016; Chatterjee et al. 2018). Reflectin materials have good electrical properties that are similar to those of artificial proton conductors. Fusion of reflectin to Ubx could lend Ubx materials the ability to change color in response to stimulus, which could be used to create bio-inspired color changing materials such as fabrics, or improve the electrical properties of Ubx materials for use in other applications. This is just one example of a protein fusion that would offer Ubx materials new functions. There is a vast world of opportunity open to exploring through protein fusions with Ubx.

REFERENCES

Adhikari B and Majumada S (2004) Polymers in sensor applications. Progress in Polymer Science 29: 699-766.

Ahn BK., Lee DW, Israelachvili JN, and Waite JH (2014) Surface-initiated self-healing of polymers in aqueous media. Nature Materials 13: 867-887.

Anguela XM and High KA (2019) Entering the modern era of gene therapy. Annual Review of Medicine 70: 273-288.

Babiker A, Myers CW, Hill CE, and Guarner J (2020) SARS-CoV-2 testing: Trials and tribulations. American Journal of Clinical Pathology 153: 706-708.

Bah A and Forman-Kay JD (2016) Modulation of intrinsically disordered protein function by post-translational modifications. The Journal of Biological Chemistry 291(13): 6696-6705.

Balu R, Knott R, Cowieson NP, Elvin CM, Hill AJ et al. (2015) Structural ensembles reveal intrinsic disorder for the multi-stimuli responsive bio-mimetic protein Rec1-resilin. Scientific Reports 5: 10896.

Beachy PA, Varkey J, Young KE, von Kessler DP, Sun BI et al. (1993) Cooperative binding of an Ultrabithorax homeodomain protein to nearby and distant DNA sites. Molecular and Cellular Biology 13: 6941-6956.

Beachy PA, Krasnow MA, Gavis ER, and Hogness DS (1988) An Ultrabithorax protein binds sequences near its own and the Antennapedia P1 promoters. Cell 55: 1069-1081.

Beck C and Gong Y (2019) A high-speed, bright, red fluorescent voltage sensor to detect neural activity. Scientific Reports 9: 15878.

Beh CY, El-Sharnouby A, Chatzipli A, Choo SW, and White R (2016) Roles of cofactors and chromatin accessibility in Hox protein target specificity. Epigenetics & Chromatin 9:1.

Bellas E, Lo TJ, Fournier EP, Brown JE, Abbott RD et al. (2015) Injectable silk foams for soft tissue regeneration. Advanced Healthcare Materials 4(3): 452-459.

Berry M and Gehring W (2000) Phosphorylation status of the SCR homeodomain determines it's functional activity: Essential role for protein phosphatase 2A,B. The EMBO Journal 19(12): 1946-2957.

Bhalla N, Jolly P, Formisano N, and Estrela P (2016) Introduction to biosensors. Essays in Biochemistry 60: 1-8.

Bidwell GL 3rd, Fokt I, Priebe W, and Raucher D (2007) Development of elastin-like polypeptide for thermally targeted delivery of doxorubicin. Biochemical Pharmacology 73(5): 620-631.

Bizzarri R, Serresi M, Luin S, and Beltram F (2009) Green fluorescent protein based pH indicators for in vivo use: A review. Analytical and Bioanalytical Chemistry 393: 1107-1122.

Blamires SJ, Spicer PT, and Flanagan PJ (2020) Spider silk biomimetics to inform the development of new wearable technologies. Frontiers in Materials 7: 29.

Bondos SE, Mendes GG, & Jons A (2020) Context-dependent HOX transcription factor function in health and disease. Progress in molecular biology and translational science. In press. DOI: 10.1016/bs.pmbts.2020.05.003

Bondos SE, Tan XX, and Matthews KS (2006) Physical and genetic interactions link Hox function with diverse transcription factors. Molecular and Cellular Proteomics 5: 824-834.

Boothby TC, Tapia H, Brozena AH, Piszkiewicz S, Smith AE et al. (2017) Tardigrades use intrinsically disordered proteins to survive desiccation. Molecular Cell 65(6): 975-984.

Brison N, Tylzanowski P, and Debeer P (2012) Limb skeletal malformations—What the HOX is going on? European Journal of Medical Genetics 55: 1-7.

Burkett-Cadena ND, Gibson J, Lauth M, Stenn T, Acevedo C et al. (2016) Evaluation of the honey-card technique for detection of transmission of arboviruses in Florida and comparison with sentinel chicken seroconversion. Journal of Medical Entomology 53(6): 1449-1457.

Cabantous S, Terwilliger TC, and Waldo GS (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nature Biotechnology 23(1):102-107.

Chakraborty K, Veetil AT, Jaffrey SR, and Krishnan Y (2016) Nucleic acid-based nanodevices in biological imaging. The Annual Review of Biochemistry 85: 349-373.

Chatterjee A, Norton-Baker B, Bagge LE, Patel P, and Gorodetsky AA (2018) An introduction to color-changing systems from the cephalopod protein reflectin. Bioinspiration and Biomimetics 12: 045001.

Chavali S, Gunnarsson A, and Babu MM (2017) Intrinsically disordered proteins adaptively reorganize cellular matter during stress. Trends in Biochemical Sciences 42(6): 410-412.

Chia HN and Wu BM (2015) Recent advances in 3D printing of biomaterials. Journal of Biological Engineering 9: 4.

Costa SA, Simon JR, Amiram M, Tang L, Zauscher S et al. (2018) Photo-crosslinkable unnatural amino acids enable facile synthesis of thermoresponsive nano- to microgels of intrinsically disordered polypeptides. Advanced Materials 30(5): 1704878.

Couzin J and Kaiser J (2005) As Gelsinger case ends, gene therapy suffers another blow. Science 307: 1028.

Cowan DA and Ternandez-Lafuente R (2011) Enhancing the functional properties of thermophilic enzymes by chemical modification and immobilization. Enzyme and Microbial Technology 49: 326-346.

Crone DE, Huang YM, Pitman DJ, Schenkelberg C, Fraser K et al. (2013) GFP-based biosensors. State of the Art in Biosensors-General Aspects, ed. T Rinken London, United Kingdom. InTech. 3-36.

Das RK and Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. PNAS 110(33): 13392-13397.

de Navas LF, Reed H, Akam M, Barrio R, Alonso CR et al. (2011) Integration of RNA processing and expression level control modulates the function of the *Drosophila* Hox gene Ultrabithorax during adult development. Development 138(1):107-16.

Deptuch T and Dams-Kozlowska H (2017) Silk materials functionalized via genetic engineering for biomedical applications. Materials 10: 1417.

Deshmukh SN, Dive AM, Moharil R, and Munde P (2016) Enigmatic insight into collagen. Journal of Oral and Maxillofacial Pathology 20(2): 276-283.

Dinjaski N, Plowright R, Zhou S, Belton DJ, Perry CC et al. (2017) Osteoinductive recombinant silk fusion proteins for bone regeneration. Acta Biomaterialia: 49: 127-139.

Dobrovolskaia MA (2019) Nucleic acid nanoparticles at a crossroads of vaccines and immunotherapies. Molecules 24: 4620.

Domingo-Espín J, Petegnief V, de Vera N, Conchillo-Sole O, Saccardo P et al. (2012) RGD-based cell ligands for cell-targeted drug delivery act as potent trophic factors. Nanomedicine: Nanotechnology, Biology, and Medicine 8: 1263-1266.

Duboule D (1991) Patterning in the vertebrate limb. Current Opinion in Genetics and Development 1: 211-216.

Dunbar CE, High KA, Joung JK, Kohn DB, Ozawa K et al. (2018) Gene therapy comes of age. Science 359(6372): 175.

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P et al. (2001) Intrinsically disordered protein. Journal of Molecular Graphics and Modelling 19: 26-59.

Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JK et al. (2008) The unfoldomics decade: An update on intrinsically disordered proteins. BMC Genomics. 2008; 9 Suppl 2(Suppl 2):S1.

Durston A, Wacker S, Bardine N, and Jansen H (2012) Time space translation: A Hox mechanism for vertebrate A-P patterning. Current Genomics 13(4):300-307.

Elsharkawy S, Al-Jawad M, Pantano MF, Tejeda-Montes E, Mehta K et al. (2018) Protein disorder-order interplay to guide the growth of hierarchical mineralized structures. Nature Communications 9: 2145.

Endo M, Kuroda S, Kondo H, Maruoka Y, Ohya K et al. (2006) Bone regeneration by modified gene-activated matrix: Effectiveness in segmental tibial defects in rats. Tissue Engineering 12(3): 489-497.

Falconnet D, Csucs G, Grandin HM, and Textor M (2006) Surface engineering approaches to micropattern surfaces for cell-based assays. Biomaterials 27(16): 3044-3063.

Figueiredo RMP, Mourao MPH, Abi-Abib YEC, Oliveira CM, Roque R et al. (2013) Identification of dengue viruses in naturally infected *Aedes aegypti* females captured with BioGents (BG)- Sentinel traps in Manaus, Amazonas, Brazil. Revista da Sociedade Brasileira de Medicina Tropical 46(2): 221-222.

Fogg JM, Kolmakova N, Rees I, Magonov S, Hansma H et al. (2006) Exploring writhe in supercoiled minicircle DNA. Journal of Physics: Condensed Matter 18: S145-S159.

Galant R and Carroll SB (2002) Evolution of a transcriptional repression domain in an insect HOX protein. Nature 415: 910-913.

Gaspar V, de Melo-Diogo D, Costa E, Moreira A, Queiroz J et al. (2014) Minicircle DNA vectors for gene therapy: Advances and applications. Expert Opinion on Biological Therapy 15(3): 353-379.

Gavis ER and Hogness DS (1991) Phosphorylation, expression and function of the Ultrabithorax protein family in *Drosophila melanogaster*. Development 112(4): 1077-1093.

Geyer A, Koltsaki I, Hessinger C, Renner S, and Rogulja-Ortmann A(2015) Impact of *Ultrabithorax* alternative splicing on *Drosophila* embryonic nervous system development. Mechanism of Development 138: 177-189.

Giesa T, Schuetz R, Fratzl P, Buehler MJ, and Masic A (2017) Unraveling the molecular requirements for macroscopic silk supercontraction. ACS Nano 11(10): 9750-9758.

Girotti A, Orbanic D, Ibanez-Fonseca A, Gonzalez-Obeso C, and Rodriguez- Cabello JC (2015) Recombinant technology in the development of materials and systems for soft-tissue repair. Advanced Healthcare Materials 4(16): 2423-2455.

Goh YY, Ho B, and Ding JK (2002) A novel fluorescent protein-based biosensor for gram- negative bacteria. Applied and Environmental Microbiology 68(12): 6343-52.

Gomes S, Leonor IB, Mano JF, Reis RL, and Kaplan DL (2012) Natural and genetically engineered proteins for tissue engineering. Progress in Polymer Science 37(1): 1-17.

Gonzalez-Reyes A, Urquia N, Gehring WJ, Struhl G, and Morata G (1990) Are cross-regulatory interactions between homeotic genes functionally significant? Nature 344:78.

Gourinat AC, O'Connor O, Calvez E, Goarant C, and Duport-Rouzeyrol M (2015) Detection of zika virus in urine. Emerging Infectious Diseases 21 (1): 84-86.

Greer AM, Huang Z, Oriakhi A, Lu Y, Lou J et al. (2009) The *Drosophila* transcription factor Ultrabithorax self-assembles into protein-based biomaterials with multiple morphologies. Biomacromolecules 10: 829-837.

Guthold M, Liu W, Sparks EA, Jawerth LM, Peng L, Falvo M et al. (2007) A comparison of the mechanical and structural properties of fibrin fibers with other protein fibers. Cell Biochemistry and Biophysics 49(3): 165-181.

Habchi J, Tompa P, Longhi S, and Uversky VN (2014) Introducing protein intrinsic disorder. Chemical Reviews 114: 6561-6588.

Hahn MS, Miller JS, and West JL (2005) Laser scanning lithography for surface micro-patterning on hydrogels. Advanced Materials 17(24): 2939-2942.

Hatton AR, Subramaniam V, and Lopez AJ (2018) Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. Molecular Cell 2: 787-796.

Hardee CL, Arevalo-Soliz LM, Hornstein BD, and Zechiedrich L (2017) Advances in non-viral DNA vectors for gene therapy. Genes 8; 65.

Hoey T and Levine M (1988) Divergent homeo box proteins recognize similar DNA sequences in *Drosophila*. Nature 332(6167): 858-61.

Hollenhorst PC, McIntosh LP, and Graves BJ (2011) Genomic and biochemical insights into the specificity of ETS transcription factors. Annual Reviews of Biochemistry 80: 437-471.

Howell DW, Duran CL, Tsai SP, Bondos SE, and Bayless KJ (2016) Functionalization of Ultrabithorax materials with vascular endothelial growth factor enhances angiogenic activity. Biomacromolecules 17(11): 3558-3569.

Howell DW, Tsai SP, Churion K, Patterson J, Abbey C et al. (2015) Identification of multiple dityrosine bonds in materials composed of the *Drosophila* protein Ultrabithorax. Advanced Functional Materials 25: 5988-98.

Hsiao HC, Gonzalez KL, Catanese DJ Jr, Jordy KE, Matthews KS et al. (2014) The intrinsically disordered regions of the *Drosophila melanogaster* Hox protein Ultrabithorax select interacting proteins based on partner topology. PLOS One 9(10): e108217.

Hsiao HC, Santos A, Howell DW, Patterson JL, Fuchs-Young SL et al. (2016) Culture of tumorigenic cells on protein fibers reveals metastatic cell behaviors. Biomacromolecules 17: 3790-3799.

Hu X, Cebe P, Weiss AS, Omenetto F, and Kaplan DL (2012) Protein-based composite materials. Materials Today 15: 208-215.

Huang CL, Leblond AL, Turner EC, Kumar AH, Martin K et al. (2015) Synthetic chemically modified mRNA-based delivery of cytoprotective factor promotes early cardiomyocyte survival post-acute myocardial infarction. Molecular Pharmaceutics 12: 991-996.

Huang Y and Bystroff C (2009) Complementation and reconstitution of fluorescence from circularly permuted and truncated Green Fluorescent Protein. Biochemistry 48: 929-940.

Huang YM, Banerjee S, Crone DE, Schenkelberg CD, Pitman DJ et al. (2015) Toward computationally designed self-reporting biosensors using Leave-One-Out Green Fluorescent Protein. Biochemistry 54: 6263-6273.

Huang YM, Nayak S, and Bystroff C (2011) Quantitative *in vivo* solubility and reconstitution of truncated circular permutants of green fluorescent protein. Protein Science 20: 1775-1780.

Huang Z, Lu Y, Majithia R, Shah J, Meissner K et al. (2010) Size dictates mechanical properties for protein fibers self-assembled by the Drosophila Hox transcription factor Ultrabithorax. Biomacromolecules 11(12): 3644-3651.

Huang Z, Salim T, Brawley A, Patterson J, Matthews KS et al. (2011) Functionalization and patterning of protein-based materials using active Ultrabithorax chimeras. Advanced Functional Materials 21: 2633-2640.

Hughes CL and Kaufman TC (2002) Hox genes and the evolution of the arthropod body plan. Evolution & Development (6): 459-99.

Jakob U, Kriwacki R, and Uversky VN (2014) Conditionally and transiently disordered proteins: Awakening cryptic disorder to regulate protein function. Chemical Reviews 114: 6778-6805.

Jansson R, Courtin CM, Sandgren M, and Hedhammar M (2015) Rational design of spider silk materials genetically fused with an enzyme. Advanced Functional Materials 25(33): 5343-5352.

Jansson R, Thatikonda N, Lindberg D, Rising A, Johansson J et al. (2014) Recombinant spider silk genetically functionalized with affinity domains. Biomacromolecules 15(5): 1696-1706.

Jarczewska M and Malinowska E (2020) The application of antibody–aptamer hybrid biosensors in clinical diagnostics and environmental analysis. Analytical Methods 12(25): 3183-3199.

Jao D, Xue Y, Medina J, and Hu X (2017) Protein-based drug-delivery materials. Materials 10(5): 517.

Jiao Z, Song Y, Jin Y, Zhang C, Peng D et al. (2018) *In vivo* characterizations of the immune properties of sericin: An ancient material with emerging value in biomedical applications. Macromolecular Bioscience 17: 1700229.

Jo J, Gao JQ, and Tabata Y (2019) Biomaterial-based delivery systems of nucleic acid for regenerative research and regenerative therapy. Regenerative Therapy 11: 123-130.

Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A et al. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. Cell 131(3): 530-43.

Kalodimos CG, Biris N, Bonvin AMJJ, Levandoski MM et al. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 305: 386-389.

Kim CS, Choi YS, Ko W, Seo JH, Lee J et al. (2011) A mussel adhesive protein fused with the BC domain of protein a is a functional linker material that efficiently immobilizes antibodies onto diverse surfaces. Advanced Functional Materials 21(21): 4101-4108.

Kim H, Lee W, and Yoon Y (2019) Heavy metal(loid) biosensor based on split-enhanced green fluorescent protein: Development and characterization. Applied Microbiology and Biotechnology 103: 6345-6452.

Kim SY, Wong AH, Abou Neel EA, Chrzanowski W, and Chan HK (2014) The future perspectives of natural materials for pulmonary drug delivery and lung tissue engineering. Expert Opinion on Drug Delivery 12(6): 869-887.

Kim TJ, Kim KA, and Jung SH (2017) Development of an endoplasmic reticulum calcium sensor based on fluorescence resonance energy transfer. Sensors and Actuators B 247: 520-525.

Kissinger CR, Liu BS, Martin-Blanco E, Kornberg TB, and Pabo CO (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: A framework for understanding homeodomain-DNA interactions. Cell 63(3):579-90.

Kobelt D, Schleef M, Schmeer M, Aumann J, Schlag PM et al. (2013) Performance of high quality minicircle DNA for *in vitro* and *in vivo* gene transfer. Molecular Biotechnology 53: 80-89.

Kodama Y and Hu CD (2012) Bimolecular fluorescence complementation (BiFC): A 5 year update and future perspectives. BioTechniques 53: 285-298.

Koebley SR, Vollrath F, and Schniepp HC (2017) Toughness-enhancing metastructure in the recluse spider's looped ribbon silk. Materials Horizons 4(3): 377-382.

Koh LD, Yeo J, Lee YY, Ong Q, Han M et al. (2018) Advancing the frontiers of silk fibroin protein-based materials for futuristic electronics and clinical wound-healing. Materials Science & Engineering C 86: 151-172.

Koker T, Fernandez A, and Pinaud F (2018) Characterization of split fluorescent protein variants and quantitative analyses of their self-assembly process. Scientific Reports 8: 53-44.

Kong G, Zhang M, Xiong M, Fu X, Ke G et al. (2020) DNA nanostructure-based fluorescent probes for cellular sensing. Analytical Methods 12: 1415-1429.

Kozlowska AK, Florczak A, Smialek M, Dondajewska E, Mackiewicz A et al. (2017) Functionalized bioengineered spider silk spheres improve nuclease resistance and activity of oligonucleotide therapeutics providing a strategy for cancer treatment. Acta Biomaterialia 59: 221-233.

Kribelbauer JF, Rastogi C, Bussemaker HJ, and Mann RS (2019) Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. Annual Review of Cell and Developmental Biology 35: 357-379.

Lakowicz, J. (2006). Principles of fluorescence spectroscopy. Third Edition University of Maryland School of Medicine. Baltimore, Maryland, USA. Springer.

Leandro AS, Britto AS, Rios JA, Galvao SR, Kafka R et al. (2020) Molecular detection of dengue virus in mosquitoes as an early indicator to aid in the prevention of human infection in endemic areas. Vector-Borne and Zoonotic Diseases 20(1): 54-59.

Lee YJ, Yi H, Kim WJ, Kang K, Yun DS, Strano MS, et al. (2009) Fabricating genetically engineered high-power lithium-ion batteries using multiple virus genes. Science 324(5930); 1051-1055.

Lewis EB (1978) A gene complex controlling segmentation in *Drosophila*. Nature 276: 565-570.

Li B, Huang Q, and Wei G-GH (2019). The role of HOX transcription factors in cancer predisposition and progression. Cancers 11(4); 528.

Li B and Daggett V (2002). Molecular basis for the extensibility of elastin. Journal of Muscle Research and Cell Motility, 23(5–6), 561–573.

Li L and Matthews KS (1997) Differences in water release with DNA binding by Ultrabithorax and Deformed homeodomains. Biochemistry 36(23): 7003-7011.

Li L, von Kessler D, Beachy PA, and Matthews KS (1996) pH-dependent enhancement of DNA binding by the Ultrabithorax homeodomain. Biochemistry 35(30): 9832-9839.

Liu Y, Huang A, Booth RM, Mendes GG, Merchant Z et al. (2018) Evolution of the activation domain in a Hox transcription factor. The International Journal of Developmental Biology 62: 745-753.

Liu Y, Matthews KS and Bondos SE (2008) Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila Hox protein Ultrabithorax. Journal of Biological Chemistry 283: 20874-20887.

Liu Y, Matthews S and Bondos SE (2009) Internal regulatory interactions determine DNA binding specificity by a Hox transcription factor. Journal of Molecular Biology 390**;** 760-774.

Liu Y and Yu J (2016) Oriented immobilization of proteins on solid supports for use in biosensors and biochips: A review. Microchimica Acta 183: 1-19.

Lopez AJ, Artero RD, and Perez-Alonso M (1996) Stage, tissue, and cell specific distribution of alternative Ultrabithorax mRNAs and protein isoforms in the *Drosophila* embryo. Roux's Archives of Developmental Biology 205: 450-459.

Lufkin T, Mark M, Hart CP, Dolle P, LeMeur M et al. (1992) Homeotic transformation of the occipital bones of the skull by ectopic expression of a homeobox gene. Nature 359: 835-841.

Mace KA, Hansen SL, Myers C, Young DM, and Boudreau N (2005) HOXA3 induces cell migration in endothelial and epithelial cells promoting angiogenesis and wound repair. Journal of Cell Science 118: 2567-77.

Majithia R, Patterson J, Bondos SE and Meissner KE (2011) On the design of composite protein-quantum dot biomaterials via self-assembly. Biomacromolecules 12: 3629-3637.

Mann RS and Chan SK (1996) Extra specificity from extradenticle: The partnership between Hox and Pbx/Exd homeodomain proteins. Trends in Genetics 12(7): 258-62.

Mann RS, Lelli KM, and Joshi R (2009) Hox specificity unique roles for cofactors and collaborators. Current Topics in Developmental Biology 88:63-101.

Mao Z, Shi H, Guo R, Ma L, Gao C et al. (2009) Enhanced angiogenesis of porous collagen scaffolds by incorporation of TMC/DNA complexes encoding vascular endothelial growth factor. Acta Biomaterialia 5: 2983-2994.

Mayrhofer P, Schleef M and Jechlinger W (2009) Use of minicircle plasmids for gene therapy. Methods in Molecular Biology 542: 87-104.

Mendes GG, Booth RM, Pattison DL, Alvarez AJ, and Bondos SE (2018) Generating novel materials using the intrinsically disordered protein Ubx. Methods in Enzymology 611: 583-605.

Michelotti N, Johnson-Buck A, Manzo AJ and Walter NG (2012) Beyond DNA origami: A look on the bright future of nucleic acid nanotechnology. WIREs Nanomedicine and Nanobiotechnology 4: 139-152.

Miranda-Nieves D and Chaikof EL (2017) Collagen and elastin biomaterials for the fabrication of engineered living tissues. ACS Biomaterials Science & Engineering 3: 694-711.

Mortlock DP and Innis JW (1997) Mutation of HOXA13 in hand-foot-genital syndrome. Nature Genetics15(2):179-180.

Muller DA, Frentiu FD, Rojas A, Moreira LA, O'Neill SL, and Young PR (2012) A portable approach for the surveillance of dengue virus-infected mosquitoes. Journal of Virological Methods 183: 90-93.

Naughton KL, Phan L, Leung EM, Kautz R et al. (2016) Self-assembly of the cephalopod protein reflectin. Advanced Materials 28: 8405-8412.

Nguyen PQ and Silberg JJ (2010) A selection that reports on protein-protein interactions within a thermophilic bacterium. Protein Engineering, Design & Selection 23(7):529-536.

Niemeyer, C. M. (2010) Semisynthetic DNA-protein conjugates for biosensing and nanofabrication. Angewandte Chemie International Edition 49: 1200-1216.

Nileback L, Hedin J, Widhe M, Floderus LS, Krona A et al. (2017) Self-assembly of recombinant silk as a strategy for chemical-free formation of bioactive coatings: A real-time study. Biomacromolecules 18(3): 846-854.

Numata K, Hamasaki J, Subramanian B, and Kaplan DL (2010) Gene delivery mediated by recombinant silk proteins containing cationic and cell binding motifs. Journal of Controlled Release 146: 136-143.

Numata K, Subramanian B, Currie HA, and Kaplan DL (2009) Bioengineered silk protein-based gene delivery systems. Biomaterials 30: 5775-5784.

O'Connor MB, Binari R, Perkins LA, and Bender W (1988) Alternative RNA products from the Ultrabithorax domain of the bithorax complex. EMBO Journal 7(2): 435-45.

Ozyurt C, Ustukarci H, Evran S, and Telefoncu A (2019) MerR-fluorescent protein chimera biosensor for fast and sensitive detection of Hg^{2}+ in drinking water. Biotechnology and Applied Biochemistry 66(5): 731-737.

Pankongadisak P, Tsekoura E, Suwantong O, and Uludag H (2020) Electrospun gelatin matrices with bioactive pDNA polyplexes. International Journal of Biological Macromolecules 149: 296-308.

Pannier AK, Shea LD (2004) Controlled release systems for DNA delivery. Molecular Therapy 10(1):19-26.

Parenteau-Bareil R, Gauvin R, and Berthod F (2010) Collagen-based biomaterials for tissue engineering applications. Materials 3: 1863-1887.

Park TG, Jeong JH, and Kim SW (2006) Current status of polymeric gene delivery systems. Advanced Drug Delivery Reviews 58: 467-486.

Passner JM, Ryoo HD, Shen L, Mann RS and Aggarwal AK (1999) Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature 397: 714-719.

Patterson JL, Abbey CA, Bayless KJ, and Bondos SE (2014) Materials composed of the *Drosophila melanogaster* protein Ultrabithorax are cytocompatible. Journal of Biomedical Materials Research Part A 102A: 97-104.

Patterson JL, Arenas-Gamboa AM, Wang TY, Hsiao HC et al. (2015) Materials composed of the *Drosophila* Hox protein Ultrabithorax are biocompatible and nonimmunogenic. Journal of Biomedical Materials Research Part A. 103A: 1546-1553.

Pearson JC, Lemons D, and McGinnis W (2005) Modulating Hox gene functions during animal body patterning. Nature Reviews Genetics 6: 893-904.

Pitman DJ, Schenkelberg CD, Huang YM, Teets FD et al. (2014) Improving computational efficiency and tractability of protein design using a piecemeal approach. A strategy for parallel and distributed protein design. Bioinformatics 30(8): 1138-1145.

Prince V (2002) The Hox paradox: More complex(es) than imagined. Developmental Biology 249(1):1-15.

Pum D, Toca-Herrera JL, and Sleytr UB (2013). S-layer protein self-assembly. International Journal of Molecular Sciences 14(2): 2484-2501.

Qin G, Hu X, Cebe P, and Kaplan DL (2012) Mechanism of resilin elasticity. Nature Communications 3: 1003.

Quiroz FG, and Chilkoti A (2015) Sequence heuristics to encode phase behavior in intrinsically disordered protein polymers. Nature Materials 14: 1164-1171.

Rabotyagova OS, Cebe P, and Kaplan DL (2011) Protein-based block copolymers. Biomacromolecules 12(2): 269-289.

Raftery RM, Walsh DP, Castano IM, Heise A, Duffy GP et al. (2016) Delivering nucleic-acid based nanomedicines on biomaterial scaffolds for orthopedic tissue repair: Challenges, progress and future perspectives. Advanced Materials 28: 5447-5469.

Romei MG and Boxer SG (2019) Split Green Fluorescent Proteins: scope, limitations, and outlook. Annual Reviews of Biophysics 48:19-44.

Saric M and Scheibel T (2019) Engineering of silk proteins for materials applications. Current Opinion in Biotechnology 60: 213-220.

Schneider CP, Shukla D, and Trout BL (2011) Arginine and the Hofmeister series: The role of ion-ion interactions in protein aggregation suppression. The Journal of Physical Chemistry B 115: 7447-7458.

Scott MP, Tamkun JW, Hartzell GW (1989) The structure and function of the homeodomain. Biochimica et Biophysica Acta 989(1): 25-48.

Shi R, Pan Q, Guan Y, Hua Z, Huang Y et al. (2007) Imidazole as a catalyst for in vitro refolding of enhanced green fluorescent protein. Archives of Biochemistry and Biophysics 459: 122-128.

Siavashani AZ, Mohammadi J, Rottmar M, Senturk B, Nourmohammadi J et al. (2020) Silk fibroin/sericin 3D sponges: The effect of sericin on structural and biological properties of fibroin. International Journal of Biological Macromolecules 153: 317-326.

Simon JR, Carroll NJ, Rubinstein M, Chilkoti, A., & Lopez GP (2017). Programming molecular self-assembly of intrinsically disordered proteins containing sequences of low complexity. Nature Chemistry 9: 509-515.

Sinn PL, Burnight ER, and McCray Jr PB (2009) Progress and Prospects: prospects of repeated pulmonary administration of viral vectors. Gene Therapy 16: 1059-1065.

Taghli-Lamallem O and Hsia C (2008) Context-dependent regulation of Hox protein functions by CK2 phosphorylation sites. Developmental Genes and Evolution 218: 321-332.

Tamburro AM, Bochicchio B, and Pepe A (2005). The dissection of human tropoelastin: From the molecular structure to the self-assembly to the elasticity mechanism. Pathologie Biologie, 53(7): 383-389.

Tan XX, Bondos S, Li L, and Matthews KS (2002). Transcription activation by Ultra-bithorax Ib protein requires a predicted α-helical region. Biochemistry 41(8): 2774-2785.

Tantama M, Hung YP, and Yellen G (2012) Imaging intracellular pH in live cells with a genetically encoded red fluorescent protein sensor. Journal of the American Chemical Society 133: 10034-10037.

Tsai SP, Howell DW, Huang Z, Hsiao HC, Lu Y et al. (2015) The effect of protein fusions on the production and mechanical properties of protein-based materials. Advanced Functional Materials 25:1442-1450.

Unlu I, Roy AF, Garrett D, Bell H et al. (2009) Evaluation of surveillance methods for detection of West Nile Virus activity in East Baton Rouge Parish, Louisiana, 2004–2006. Journal of the American Mosquito Control Association 25(2): 126-133.

Urello MA, Kiick KL, and Sullivan MO (2014) A CMP-based method for tunable, cell-mediated gene delivery from collagen scaffolds. Journal of Materials Chemistry B 2: 8174-8185.

Villalonga A, Perez-Calabuig AM, and Villalonga R (2020) Electrochemical biosensors based on nucleic acid aptamers. Analytical and Bioanalytical Chemistry 412: 55-72.

Vivian JT and Callis PR. (2001) Mechanisms of tryptophan fluorescence shifts in proteins. Biophysical Journal 80(5): 2093-2109.

Wang T, Lai JH, and Yang F (2016) Effects of hydrogel stiffness and extracellular compositions on modulating cartilage regeneration by mixed populations of stem cells and chondrocytes in vivo. Tissue Engineering Part A 22(23–24): 1348-1356.

Wei G, Su Z, Reynolds NP, Arosio P, Hamley IW et al. (2017) Self-assembling peptide and protein amyloids: From structure to tailored function in nanotechnology. Chemical Society Reviews 46(15): 4661-4708.

Wesley RD, Dreiss CA, Cosgrove T, Armes P, Thompson L et al. (2005). Structure of a hydrophilic-hydrophobic block copolymer and its interactions with salt and an anionic surfactant. Langmuir 21(11): 4856–4861.

White RAH and Wilcox M (1984) Protein products of the bithorax complex in *Drosophila*. Cell 39: 163-171.

Wolinsky JB, Colson YL, and Grinstaff MW (2012). Local drug delivery strategies for cancer treatment: Gels, nanoparticles, polymeric films, rods, and wafers. Journal of Controlled Release, 159(1): 14–26.

Xu M, Jiang Y, Prandhan S, and Yadavalli VK (2019) Use of silk proteins to form organic, flexible, degradable biosensors for metabolite monitoring. Frontiers in Materials 6: 331.

Yamamoto M and Tabata Y (2006) Tissue engineering by modulated gene delivery. Advanced Drug Delivery Reviews 58: 535-554.

Yang YJ, Holmberg AL, and Olsen BD (2017). Artificially engineered protein polymers. Annual Review of Chemical and Biomolecular Engineering 8: 549-575.

Zallen DT (2000) US gene therapy in crisis. Trends in Genetics 16(6): 272-275.

Zechiedrich L and Fogg JM (2019) Biophysics meets gene therapy: How exploring supercoiling-dependent structural changes in DNA led to the development of minivector DNA. Technology and Innovation 20: 427-440.

Zhai B, Villen J, Beausoleil SA, Mintseris J, and Gygi SP (2008) Phosphoproteome analysis of *Drosophila melanogaster* embryos. Journal of Proteome Research 7: 1675-1682.

Zhao N, Fogg JM, Zechiedrich L and Zu Y (2011) Transfection of shRNA-encoding minivector DNA of a few hundred base pairs to regulate gene expression in lymphoma cells. Gene Therapy 18; 220-224.

Zhou S, Huang W, Belton DJ, Simmons LO, Perry CC et al. (2015) Control of silicification by genetically engineered fusion proteins: Silk-silica binding peptides. Acta Biomaterialia, 15: 173-180.

APPENDIX A

A NEW CONSTRUCT TO REDUCE LOO-GFP AND UBX INTERACTIONS: SGMU

A.1 Introduction and Results

Immobilization of LOO8-GFP by Ubx materials increases the relative fluorescence of LOO8-GFP while lowering the background fluorescence of the sensor (Chapter 3). By removing the majority of background fluorescence and using a combination of high NaCl concentration, imidazole, and arginine to rebind peptide, we were able to increase of the fluorescent signal to 164% that of background fluorescence. The intensity when recovered was only 23% of the maximum intensity for LOO8-GFP fibers, which leaves much room for improvement. Future improvements will focus on peptide rebinding by the materials, both in increasing the amount of peptide rebound, and increasing the maximum fluorescence recovered.

One of the issues with LOO8-GFP-Ubx fibers was incomplete peptide rebinding, which we attribute to ionic interaction between the negatively charged LOO8-GFP and positively charged Ubx (Figure A-1A). Thus, high NaCl concentrations screen the ionic interactions between the proteins, allowing peptide binding to LOO8-GFP. The best peptide rebinding occurred between 1-1.5M NaCl, which is a much higher concentration than ideal for use with many proteins the sensor may be redesigned to bind. The other problem is that not all LOO-8-GFP-Ubx proteins are purified after expression in the peptide bound state. Although the left-out strand and LOO-8-GFP-Ubx are co-expressed

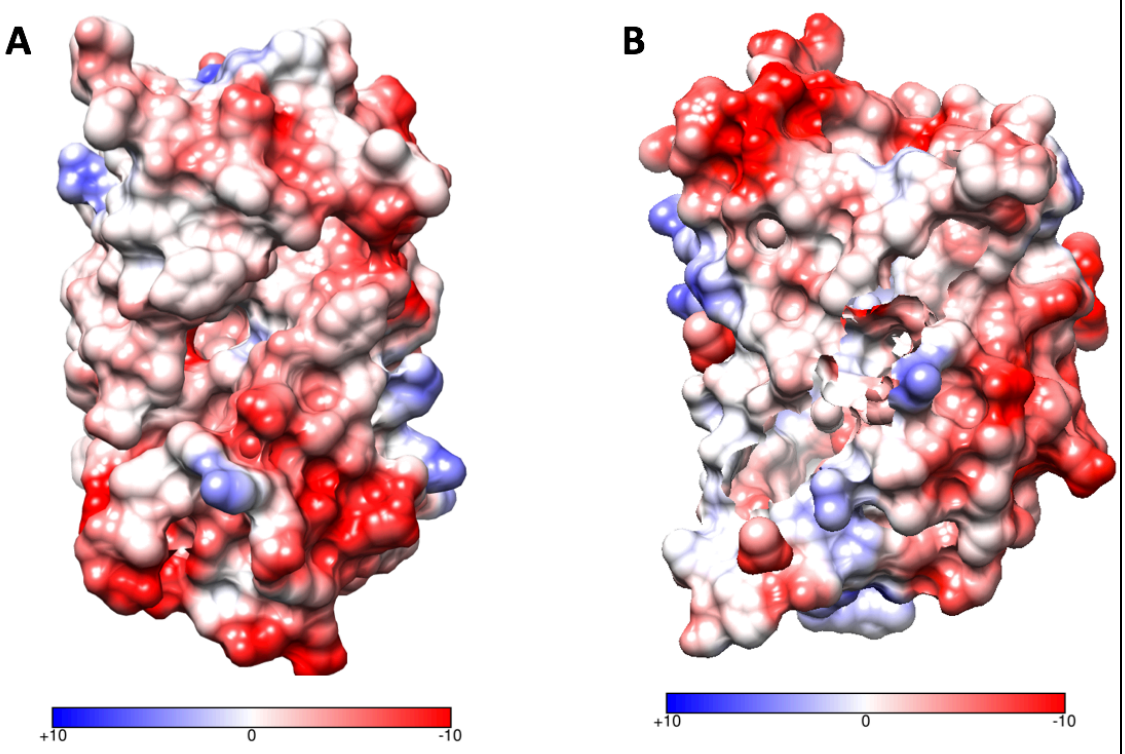

Figure A-1 GFP-OPT surface charge at pH 8.0.
Superfolder GFP-OPT is the version of GFP used to design LOO-GFPs (PDB 2b3p). (A) The surface of the barrel distant to the strand 8 binding pocket in LOO8-GFP. There are multiple negatively charged residues (red) around the left-out strand 8. (B) The section of the barrel near the strand 11 binding pocket of LOO11-GFP. There appear to be fewer negatively charged residues (red) around the left-out strand 11 than strand 8. This may prevent interactions between Ubx and the binding pocket of LOO11-GFP.

from the same plasmid, it does not guarantee that every LOO-GFP protein will be bound to peptide. This can lead to aggregation and proteolysis of the LOO-GFP, lowering the purification yield.

A.2 Results

To assist in overcoming these problems, we designed a new LOO-GFP construct known as SGMU (s11—**S**ortase A—LOO11-**G**FP—Maltose Binding Protein—**U**ltrabithorax), which is expressed as a single fusion protein (Figure A-2). S11, the left out peptide, is expressed as part of the fusion protein, tethering the peptide to LOO11-GFP to increase the probability of binding. Sortase A (SrtA) is a self-cleaving enzyme activated by calcium. It is able to cleave itself and the s11 from the rest of the fusion protein, making peptide removal possible. MBP increases solubility of the fusion protein and provides steric hindrance between the open LOO-GFP and Ubx. We theorize that the steric hinderance may block interactions between the proteins and allow for more peptide binding, and thus a higher increase in fluorescence. The SGMU construct uses LOO11-GFP, rather than LOO8-GFP. There are fewer negatively charged residues around strand 11 than strand 8 (Figure A-1 A and B), which we hypothesize will also decrease interactions between LOO-GFP and Ubx.

Due to incorporation of SrtA, a step must be added to the fiber treatment protocol, incorporating a sortase buffer step that provides calcium for cleavage of the left out peptide from the rest of the protein (Figure A-3). However, it appears that peptide removal occurs even without treating fibers with sortase buffer (Figure A-4 A-C). To inhibit SrtA activity during materials formation, we added EDTA to the materials forming buffer. Removal of the peptide despite this extra step (Figure A-4 D-F) implies

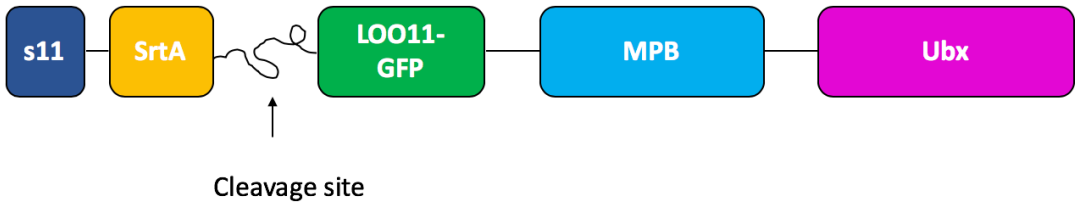

Figure A-2 SGMU Fusion Protein.
The fusion protein contains, from N-terminus to C-terminus, the left out strand 11 (dark blue); Sortase A (SrtA, yellow), a self-cleaving enzyme activated by calcium; LOO11-GFP (green); Maltose Binding Protein (MBP, light blue) provides steric hinderance between the open LOO11-GFP and Ubx and increases the overall solubility of the construct; Ubx (pink), forms materials to immobilize LOO11-GFP.

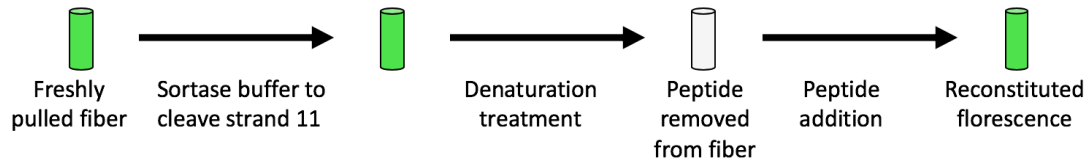

Figure A-3 Experimental procedure
SGMU fibers are produced with strand 11 at the end of the fusion protein. Treatment with sortase buffer containing calcium allows SrtA to cleave itself and s11. These fibers can then go through the denaturation treatment to remove the left out peptide, which is no longer fused to the protein.

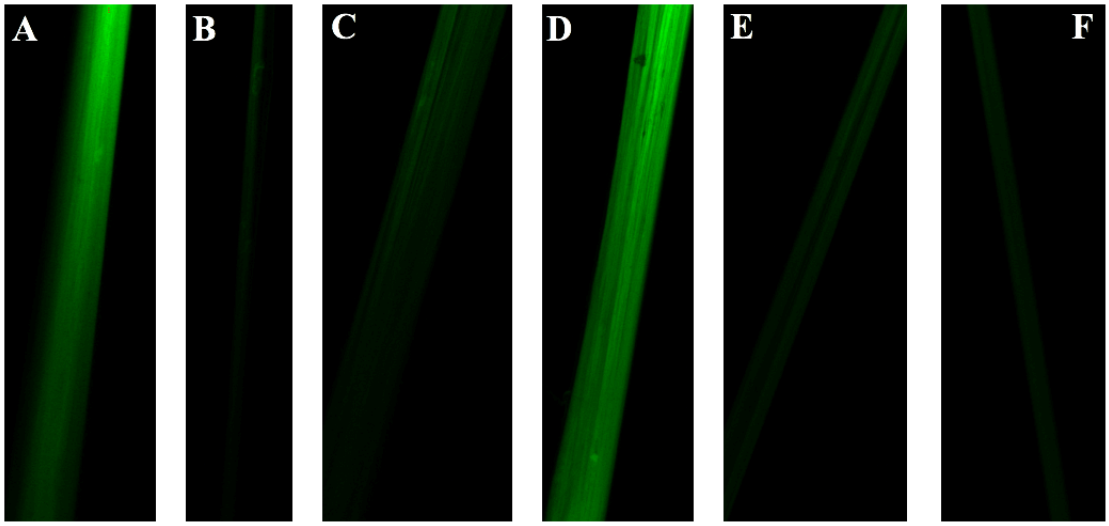

Figure A-4 SGMU is cleaved during expression or unable to refold.
(A) Control SGMU fiber that underwent no treatments. (B) and (C) SGMU fibers washed for 1 minute at pH 2.2, with 0.5M NaCl (B) or 1M NaCl (C) with no sortase buffer treatment. (D-F) SGMU fibers produced with EDTA to prevent cleavage of the peptide by Sortase A during materials formation. (D) Control SGMU fiber (E) and (F) SGMU fibers washed for 1 minute at pH 2.2, with 0.5M NaCl (E) or 1M NaCl (F) with no sortase buffer treatment. Fluorescence decreases without treatment to cleave the peptide, and with precautions taken during materials formation to prevent cleavage.

that SrtA cleavage occurs before materials formation. SrtA does have a 10% activity without the presence of calcium, so it is possible that SrtA may cleave the peptide during protein expression in *E. coli*. It is also possible that the 1 minute low pH wash is irreparably unfolding SGMU, preventing fluorescence. Further testing of conditions on EGFP-Ubx is needed to determine which possibility is occurring.

We repeated the combinations of low pH and NaCl concentrations performed in Chapter 3 with the new SGMU construct and found that SGMU yields much more signal upon peptide rebinding, and at low NaCl concentrations also dramatically reduces the background signal. (Figure A-5 A-C). This lower reliance on salt concentrations to rebind peptide is likely due to a combination of MBP blocking interactions between LOO1-GFP and Ubx, and fewer charged residues around the binding site of LOO11-GFP.

A.3 Materials and Methods

To test for cleavage by SrtA without the presence of calcium, loops with 25% LOO11-GFP/75% UbxIa fibers were placed in a 24-well plate and denatured for 1 minute to remove bound peptide with 250 μL of 0.05 M glycine-HCl buffer, pH 2.2, containing NaCl concentration of 0.5 or 1 M as specified. Fibers were rinsed by placing the loops in two different rinse wells containing PBS (20 mM NaH_2PO_4, 80 mM Na2HPO4, NaCl as specified), pH 8.0. Fibers were then imaged on a Nikon Eclipse Ti A1R equipped with NIS Elements AR software.

To test the salt dependence of peptide rebinding by LOO11-GFP, loops with 25% LOO11-GFP/75% UbxIa fibers were placed in a 24-well plate and treated with 250 μL of sortase buffer (50 mM Tris-HCl, 150 mM NaCl, 10 mM $CaCl_2$, pH 7.5) for 30

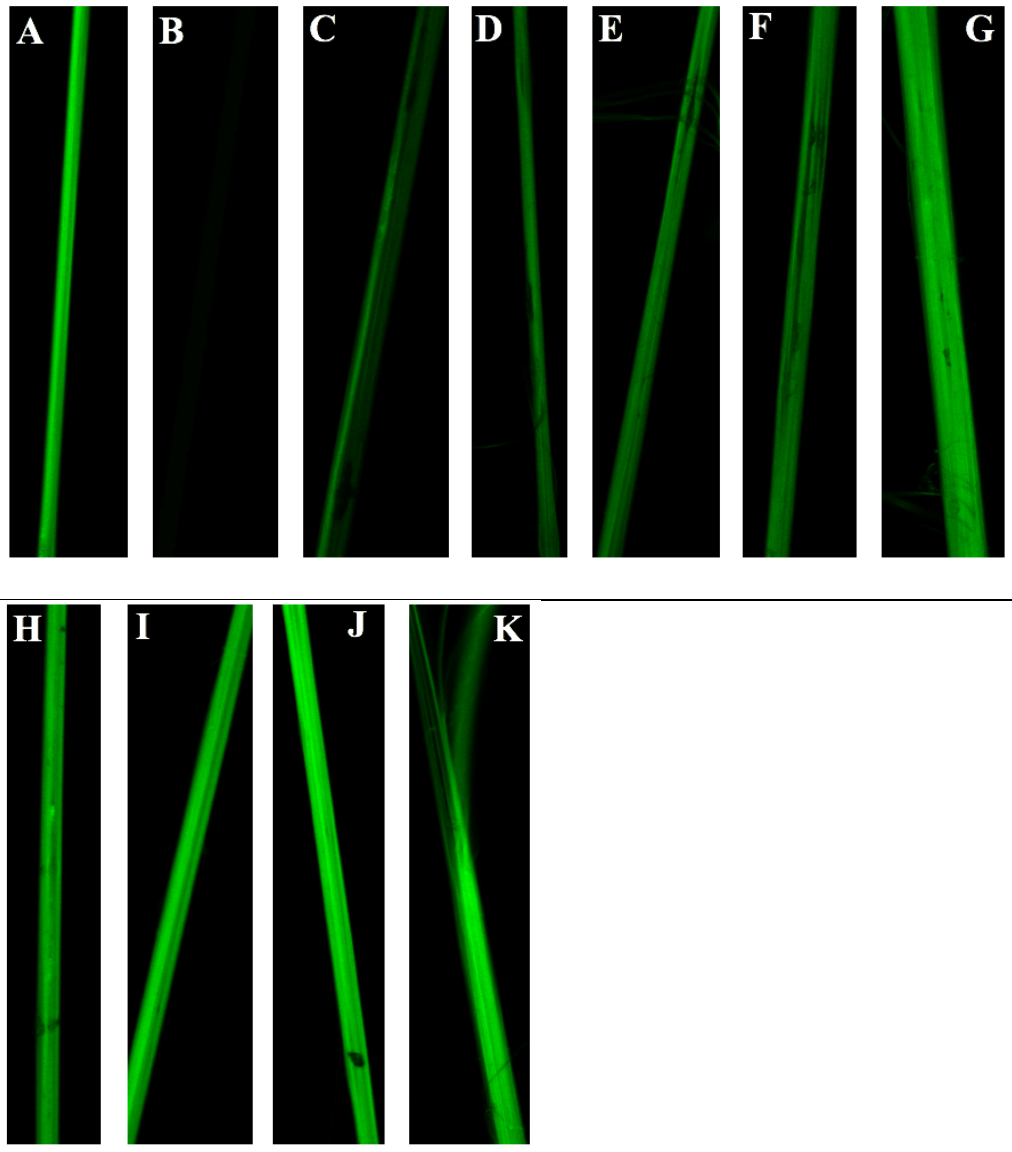

Figure A-5. Removal and rebinding of peptide is sensitive to NaCl concentrations. (A) Control SGMU fiber that underwent no treatments. (B) and (C) SGMU fibers washed for 1 minute at pH 2.2, with 0.5M NaCl. Peptide is added in 0.5M NaCl in (C). (D) and (E) SGMU fibers washed for 1 minute at pH 2.2, with 0.75M NaCl. Peptide is added in 0.75M NaCl in (E). (F) and (G) SGMU fibers washed for 1 minute at pH 2.2, with 1M NaCl. Peptide is added in 1M NaCl in (G). (H) and (I) SGMU fibers washed for 1 minute at pH 2.2, with 1.25M NaCl. Peptide is added in 1.25M NaCl in (I). (J) and (K) SGMU fibers washed for 1 minute at pH 2.2, with 1.5M NaCl. Peptide is added in 1.5M NaCl in (K).

minutes. After a quick rinse in phosphate buffered saline (PBS, 20 mM NaH_2PO_4, 80 mM Na_2HPO_4, 100 mM NaCl), fibers were denatured for 1 minute to remove bound peptide with 250 μL of 0.05 M glycine-HCl buffer, pH 2.2, with NaCl ranging from 0.5 to 1.5 M NaCl as specified. Fibers were rinsed by placing the loops in two different rinse wells containing high salt PBS, which is no-salt PBS (20 mM NaH_2PO_4, 80 mM Na2HPO4), pH 8.0 with NaCl added as indicated in the figures. Fibers were then placed either in high salt PBS or synthetic left out peptide (GenScript), diluted in high salt PBS to 0.12 mM, and left overnight at 25 ^{O}C to bind peptide before imaging as described below. Synthetic LOO11-GFP peptide was dissolved in a small volume of dimethylformamide and diluted with PBS to a stock concentration, then stored at -80 ^{O}C until use. Fibers were then imaged on a Nikon Eclipse Ti A1R equipped with NIS Elements AR software.

APPENDIX B

A FULL LENGTH MODEL AND FIRST IMAGES OF A HOX PROTEIN

B.1 Hox Structure

Hox transcription factors are fundamental to the processes of development, wound repair, and carcinogenesis. Although Hox proteins were first discovered in *Drosophila melanogaster* (Lewis 1978), homologous Hox genes are found in all bilateral animals (Prince 2002). Hox proteins all contain the same homeodomain (HD) sequence, a 60 amino acid region which contains the DNA binding HD motif (Scott, Tamkun, and Hartzell 1989). This motif is conserved between organisms, although the sequence outside the HD can vary widely (Hoey and Levine 1998, Kissinger et al. 1990). The HD binds DNA with high affinity, but low specificity. The disordered characteristic of some regions outside the HD is conserved, but the sequence within them is variable (Liu et al. 2008, Galant and Carroll 2002).

No complete structure of a full-length Hox protein is available for any Hox protein from any organism. Although the structure of the homeodomain, which is heavily conserved among Hox proteins, is known to be three helices with a helix-turn-helix motif (Qian et al. 1989), little structural information is available for regions outside the homeodomain. For Ultrabithorax (Ubx), a *Drosophila* Hox protein, the only verified structural information outside of the HD is a reverse turn formed by the YPWM motif (Passner et al. 1999); all other information is based upon predictions. For a few well studied Hox

proteins, functional roles have been assigned to the non-HD regions. In Ubx, the transcription activation domain is located N-terminal to the HD and covers a large section of the protein (~200 amino acids) (Tan et al. 2002). Regions of repression have also been mapped to areas of Hox proteins outside of the HD (Ronshaugen et al. 2002; Tour et al. 2005). These functional regions have never been linked to a structure and the mechanism by which they work is still unclear.

Ubx, one of the best studied Hox proteins, has the structured DNA binding homeodomain (HD) common to all Hox proteins, as discussed above, which is composed of three alpha helices. Helices 1 and 2 pack in an antiparallel arrangement, while helix 3, the recognition helix, lies within the major groove of DNA (Kissinger et al. 1990), forming contacts with the bases and phosphate backbone. The N-terminal arm of the HD inserts into the minor groove for specificity (Kissinger et al. 1990)(Joshi et al. 2007). There is a region predicted to be a coiled coil outside the HD (Liu et al. 2008) and the YPWM motif, which is N-terminal to the HD, forms a reverse turn and interacts with Extradenticle (Passner et al. 1999) by inserting in a hydrophobic pocket of the Exd HD. Many of the areas outside the HD are disordered and some regions are evolutionarily conserved and predicted to be involved in protein interactions (Figure B-1) (Howell et al. 2015, Liu et al. 2008).

A

MNS**Y**FEQASGF**Y**GHPHQATGMAMGSGGHHDQTASAAAAAYRGFPLSLGMS
PYANHHLQRTTQDSPYDASITAACNKIYGDGAGAYKQDCLNIKADAVNG**Y**K
DIWNTGGSNGGGGGGGGGGGGAGGTGGAGNANGGNAANANGQNNPAGG
MPVRPSACTPDSRVGG**Y**LDTSGGSPVSHRGGSAGGNVSVSGGNGNAGGVQS
GVGVAGAGTAWNANCTISGAAAQTAAASSLHQASNHTF**Y**PWMAIAGKIRSD
LTQYGGISTDMGKRYSESLAGSLLPDWLGTNGLRRRGRQT**Y**TR**Y**QTLELEKE
FHTNH**Y**LTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQEKQ
AQAQKAAAAAAAAAAVQGGHLDQ

B

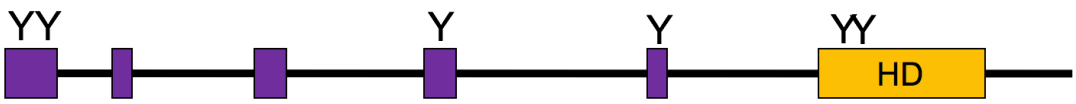

Figure B-1. Ubx has conserved regions predicted to be involved in protein interactions. (A) The UbxIa amino acid sequence. The DNA-binding homeodomain is in orange underlined text and sequences predicted to be conserved sequences predicted to be involved in protein interactions via the ANCHOR algorithm are in purple underlined text. Tyrosines involved in dityrosine bond formation in Ubx materials are in large, bold text. (B) Ubx sequence schematic, with conserved regions containing aromatics that are predicted to be involved in protein interactions via the ANCHOR algorithm marked in purple, and the homeodomain (HD) marked in orange. There are 5 disordered loops in the areas between these regions and the HD. The tyrosines involved in materials formation are marked above.

B.2 Regions Outside the Homeodomain Influence DNA Binding

For a few well studied Hox proteins, functional roles have been assigned to the non-HD regions. In Ubx, the transcription activation domain is located N-terminal to the HD and covers a large section of the protein (~200 amino acids). This activation domain is predicted to have a β-sheet and an α-helix. The α-helix is necessary, but not sufficient, for transcriptional activation (Tan et al. 2002). Regions of repression have also been mapped to areas of Hox proteins outside of the HD (Ronshaugen et al. 2002); Tour et al. 2005). There is a transcriptional repression domain in the carboxy-terminal region of Ubx known as the QA domain (Galant and Carroll 2002). The region between the YPWM motif and the HD, which includes the microexons and is known as I1, inhibits DNA binding, as does inhibition region I2 (aa 174-216) (Liu et al. 2008). The intrinsically disordered N-terminal region (R region) restores binding affinity of the HD. (Liu et al. 2008). Evolutionarily conserved, intrinsically disordered regions outside the HD affect DNA binding specificity and transcription activation, even though they do not directly interact with DNA (Liu et al. 2008; Liu et al. 2009). Full length UbxIa compared to HD binding affinity to a variety of DNA sequences can vary more than an order of magnitude (Liu et al. 2009). Binding of the full length protein is more specific than binding of HD alone, based on DNA binding to the optimal binding site for the HD 40AB (TAAT), and 5 genomic binding targets. These functional regions have never been linked to a structure and the mechanism by which they work still unclear.

B.3 A Proposed Model

Although much of Ubx is intrinsically disordered, there are conserved motifs that may adopt local structure, are enriched in aromatic amino acids, and which are predicted to be involved in protein interactions (Liu et al. 2008) (Figure B-1). Truncations mutants show that the presence of these motifs impact DNA binding affinity (Liu et al. 2008). Ubx forms intermolecular dityrosine bonds to self-assemble into materials *in vitro*, and many of the tyrosines involved fall within conserved motifs (Howell et al. 2015). We theorize that these intermolecular interactions may also occur intramolecularly to regulate DNA binding. We propose a model in which an aromatic cluster, composed of the conserved motifs mentioned above, interacts with the DNA binding interface of Ubx (closed state) and a conformational change occurs to free the interface to bind DNA (open state) (Figure B-2).

B.4 Electron Microscopy as a Method to Test the Model

Ubx is a small (~40 kD) protein in which long intrinsically disordered regions complicate structural analysis. In Negative Stain Electron Microscopy (NS-EM), protein is coated in a thin layer of charged heavy metal salts to increase contrast (Booth et al., 2011), allowing small biological molecules to be examined by EM. Optimized Negative Staining has been used to determine the mechanics of a 53kDa protein (CETP) and can be used to study dynamic proteins with equilibrium-fluctuating structures (Zhang et al., 2012). We expect that with the resolution obtainable with NS-EM, we will be able to see

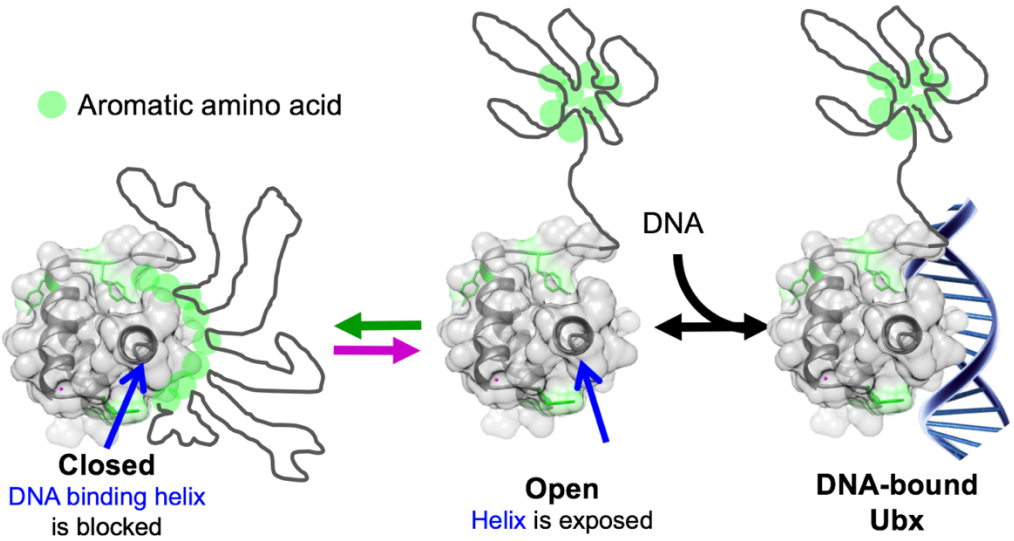

Figure B-2. Model of Conformations of Ubx. Ubx exists in equilibrium between two conformations. In the closed state, a cluster of aromatic residues (green circles) connected by disordered loops that bind to the DNA binding helix of the homeodomain (pdb IB81). In the open state, the aromatic cluster separates from the homeodomain, freeing the DNA binding helix. This model is not drawn to scale.

the difference between the open and closed state in our model (Figure B-2).

B.5 Results

We collaborated with Dr. Junji Zhang (Texas A&M University) to examine UbxIa by negative stain electron microscopy (NS-EM). Using Ubx purified via the one column method (Greer et al. 2009; Tsai et al. 2015), we obtained an initial 2-dimensonal characterization to look at the morphology and structure of Ubx (Figure B-3). Our model

predicts that interactions between aromatic amino acids and the homeodomain will create 5 intrinsically disordered loops (Figure B-1). Our preliminary results from NS-EM of Ubx reveal that a compact structure can form with 5 lobes (Figure B-3 B and C), likely corresponding to the closed state depicted in Figure B-2. In Figure B-3 C and D, this 5-lobed structure is paired with an additional blob. We tentatively assigned these images as the open state in Figure B-2.

B.6 Conclusions

Our proposed model, the first of any Hox protein, suggests that regions outside the HD form aromatic clusters which modulate DNA binding. The first EM images of a Hox protein show the multiple conformations of Ubx, which may explain the effect of regions outside the HD on DNA binding. The model provides insight into how Hox proteins are regulated *in vivo* and defines how Hox proteins recognize DNA, and will lead the way in development of additional models and more detailed structural characterization of Ubx.

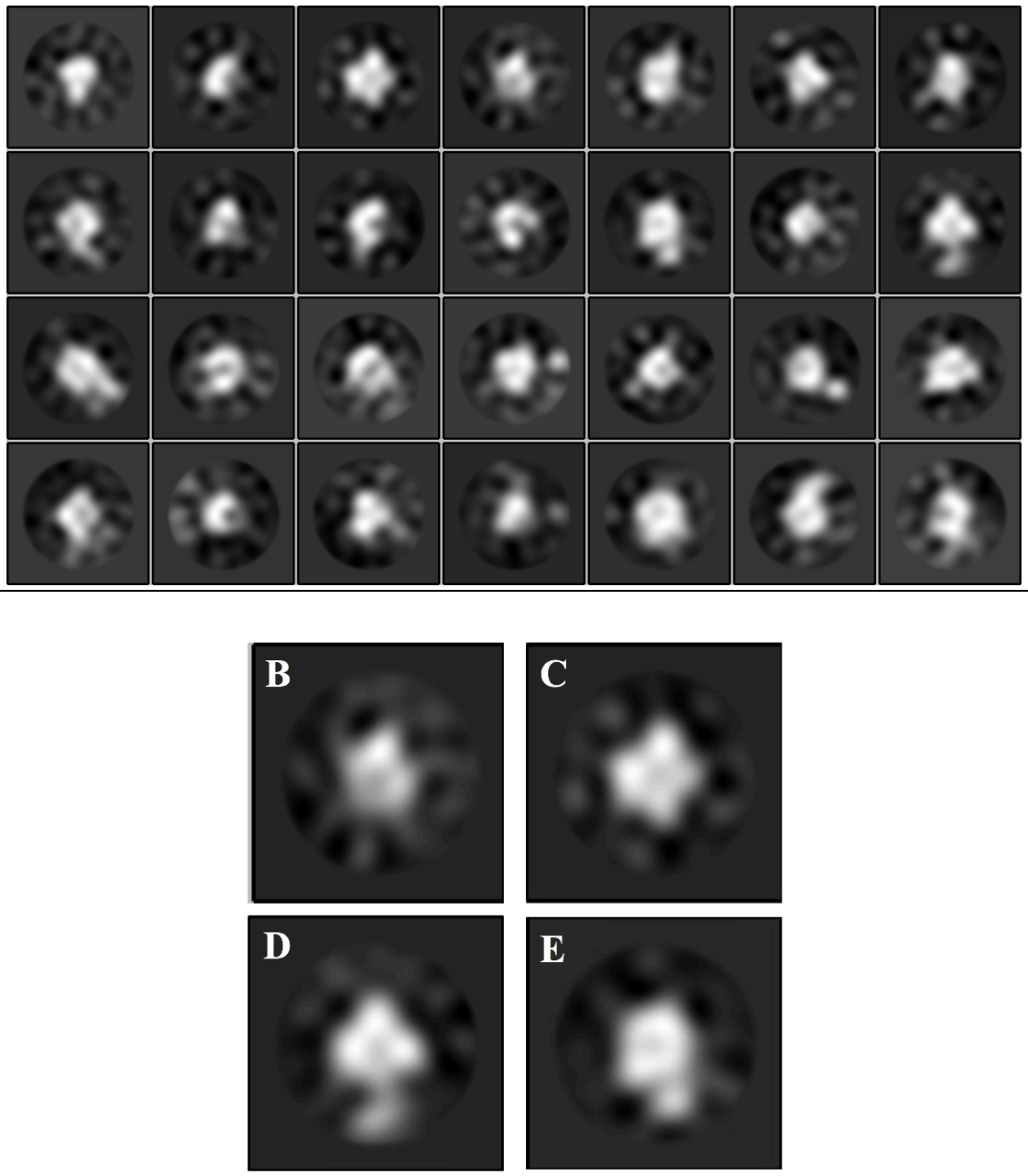

Figure B-3. Two-dimensional classification of Ubx. (A) Negative stain electron microscopy of wild type UbxIa. (B-E) Examples of open and closed conformation of UbxIa. Panel B and C may show Ubx in the closed state, unable to bind DNA. Panel D and E may show Ubx in the open state, with the smaller removed segment likely to be the homeodomain.