HOX DNA BINDING: FROM MONOMERS TO FIBERS

A Dissertation

by

KELLY ANN CHURION

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
	James Erickson
	Jun-Yuan Ji
Interdisciplinary Faculty Chair,	Dorothy Shippen

May 2017

Major Subject: Genetics

Copyright 2017 Kelly Churion

ABSTRACT

Biological macromolecules can bind a wide range of ligands with high affinity and high specificity. In particular, proteins that bind DNA must recognize a particular DNA sequence within a vast array of non-target sites. In many cases, protein-DNA interactions must also be regulated by the cell and its environment. Hox proteins are transcription factors that interact with DNA via the homeodomain with extraordinarily high affinity (pM). Hox homeodomains are highly conserved and bind similar DNA sequences; therefore, it is unclear how different Hox proteins recognize different genes *in vivo*. Elucidating the molecular basis for high affinity binding and the mechanisms that enable regulation will allow us to understand how Hox proteins normally function in development and wound repair and how malfunction of Hox proteins leads to developmental malformations and cancer. In addition, this information can be exploited to develop novel therapies and biotechnologies.

This thesis explores the DNA binding interactions by the *Drosophila melanogaster* Hox transcription factor Ultrabithorax (Ubx). First, most transcription factors contain large intrinsically disordered domains, making these proteins prone to proteolysis and complicating protein purification. Furthermore in Hox proteins, proteolytic products out-compete full-length protein for binding to DNA. We developed an innovative protein purification technique to generate Ubx of sufficient quality and purity for sensitive DNA binding assays. Second, it was previously known that non-homeodomain regions influence DNA-binding. However the specific amino acids that

comprise this intraprotein interface have not been identified. We mapped specific amino acids involved in Ubx-DNA interactions that regulate high affinity binding, providing the first model of a full-length Hox protein structure. Third, materials composed of DNA are easily designed but lack the diverse structural and chemical properties thus limiting the range of structures and activities that can be achieved. Conversely, proteins can form materials with diverse structural and chemical properties, but are difficult to design. Incorporating both DNA and protein into composite materials can maximize the advantages of both types of molecules. We demonstrate that materials composed of Ultrabithorax (Ubx) retain the ability to bind DNA noncovalently in a sequence specific manner, which allows for the optimization and generation of novel composite biomaterials. Taken together, the knowledge reported in this thesis explains many aspects of Hox regulation and provides the basis for the development of novel composite materials.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Sarah Bondos, for being a great mentor and guiding me to be the scientist I am today. I would also like to thank my committee members, Dr.Kapler, Dr. Erickson and Dr. Ji, for their guidance and support throughout the course of this research. I especially thank former lab members Dr. Shang-Pu Tsai for teaching me essential molecular biology skills, Dr. Hao-Ching Hsiao for teaching me DNA binding techniques and Jan Patterson for all her help in expressing and purifying Ultrabithorax (Ubx) protein for all of my projects along with valuable life lessons. I would like to thank current lab members Dr. David Howell for his help with microscopy and fluorescent experiments and graduate student Rebecca Booth for her help in finishing the structure/function projects for Ubx. I would also like to thank Alexi Person for being a great friend and colleague.

Thanks also go to my friends, colleagues, the department faculty and staff for making my time at Texas A&M University an unforgettable experience. I will carry the aggie experience for the rest of my life. I would like to specially thank Dr. Kayla Bayless lab members for their numerous intellectual contributions and collaboration and also Dr. Hays Rye lab members for their help with fluorescence spectroscopy. Thanks to a great mentor Dr. Kathleen Matthews who helped me move along projects and special thanks to Dr. Payel Das and Dr. Ying Liu for their contribution to each project. I would like to thank a very special person in my life, Zinnia Galindo, her support, care and love lifted my spirit when I needed it the most. Thank you.

I would like to especially thank my brothers Raymond and Bryan and my wonderful parents Nelson Churion and Mariluz Churion, for being there for me when I needed it the most and never letting me give up. They have been my inspiration throughout the entire process and without them I would not be who I am today. I hope you are proud of me mommy and daddy. We did it!

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Sarah E. Bondos, Dr. Geoffrey Kapler, Dr. Jun-Yuan Ji of the Department of Molecular and Cellular Medicine, Texas A&M Health Science Center and Professor James Erickson of the Department of Biology, Texas A&M University.

All work for the dissertation was completed by the student, in collaboration with Dr. Payel Das from IBM Thomas J Watson Research Center, the Computational Biology Center, Dr. Ying Liu from the Department of Biochemistry and Cell Biology, Rice University, Dr. Sarah E. Bondos from the Department of Biochemistry and Cell Biology, Rice University and the Department of Molecular and Cellular Medicine, Texas A & M Health Science Center, Dr. Hays Rye from the Department of Biochemistry and Biophysics - Texas A&M University, Lauren Kustigian from the Department of Biochemistry and Biophysics - Texas A&M University. Dr. Kayla J. Bayless from the Department of Molecular and Cellular Medicine, Texas A&M Health Science Center, Robert E. Rogers from the Department of Molecular and Cellular Medicine, Texas A&M Health Science Center

Funding Sources

This work was made possible in part by the National Science Foundation, and the Robert A. Welch Foundation under Grant Number C-576 and 1151394 respectively.

Its contents are solely the responsibility of the author and do not necessarily represent the official views of these institutions.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. GENERAL INTRODUCTION AND LITERATURE REVIEW

1.1 Protein-DNA interactions

The association between protein and DNA molecules has been investigated by scientists for decades. Initial low-resolution X-ray structures of nucleic acid duplexes (Rosenberg et al. 1973) concluded that the major groove of the DNA helix offered a set of base-specific hydrogen bond donors and acceptors and non-polar groups that could be recognized by a complementary set of donors and acceptors present in amino acid side chains (Seeman et al. 1973). These observations suggested that short DNA sequences could serve as binding sites that were specifically read by a complementary sequence of amino acid side chains (Viswamitra et al. 1978) creating unique protein-DNA complexes crucial for life processes. Over time, improvements in the methods used to characterize these complexes helped elucidate the fundamental relationship between protein-DNA interactions and their biological function. This biological relationship contributes to animal development, disease progress, and potentially even biotechnological applications.

1.1.1 DNA binding by transcription factor proteins is essential for gene regulation

DNA-binding proteins include, but are not limited to, transcription factors which modulate the process of transcription, polymerases, nucleases which cleave DNA molecules, histones which are involved in chromosome packaging and transcription in the cell nucleus, and single stranded DNA and RNA binding proteins. These DNA

binding proteins are vital for the normal development of an organism, as well as for routine cellular functions and response to disease. A few of the most conserved DNA binding motifs are the zinc finger, leucine zipper, and helix-turn-helix domains. DNA binding by these proteins is a critical aspect of their function.

The transcriptional regulatory system plays a central part in coordinating many important biological processes, from cell cycle progression and physiological responses, to cell differentiation and development, to reliably express specific genes in a temporal and specific manner (Vaquerizas et al. 2009). Transcription factors (TFs) are the main players in transcriptional regulation. TFs regulate both chromatin structure and gene expression by binding to DNA. Thus, transcription factors have a central role in the transcriptional regulation of all cellular organisms, being present in all branches of the tree of life: bacteria, archaea, and eukaryotes (Mendoza et.al 2013). The proportion of proteins that encode transcription factors (TFs) increases with the complexity of the organism, ranging from approximately 5% of the protein-coding genes code in flies and nematodes to approximately 10% for mouse and humans (Martinez and Walhout 2009). They may bind directly to promoter regions of DNA, which lie upstream of the coding region in a gene, or directly to the RNA polymerase molecule (Cooper 2000). After binding occurs, alone or with other proteins in a complex, transcription factors may activate or repress the transcription of a gene, thus determining whether the associated is active.

Transcription factors are a very diverse family of proteins that range widely in structure and associated function. Basal, or general, transcription factors are necessary

for RNA polymerase to function at a site of transcription in eukaryotes and prokaryotes (Weinzierl 1999).They are considered the most basic set of proteins needed to activate gene transcription, and they include a number of proteins, such as TFIIA (transcription factor II A) and TFIIB (transcription factor II B), among others (Cooper et al. 2015). Significant progress has been made in defining the roles played by each of the proteins that compose the basal transcription factor complex.

Unlike general transcription factors, specific transcription factors control the transcription of specific target genes. For many genes, general transcription factors alone only activate transcription at very low (basal) levels. Instead, the right set of specific transcription factors must also be present to activate the gene. Consequently, the gene is expressed only under certain conditions, which are controlled by the specific transcription factors. A typical activating specific transcription factor is a DNA-binding protein that recognizes a target DNA sequence and recruits RNA polymerase or a general TF to the corresponding promoter. Transcription repressors typically block access to DNA binding sites; wither by binding them directly or by inducing chromatin remodeling. Regulation by specific transcription factors is often combinatorial, with a subset of transcription factors nucleating assembly of the complex. This allows information from different sources to be integrated into a single response to reach the appropriate level of gene transcription (Lodish et al. 2000, Griffiths et al. 1999, Griffiths et al. 2000).

1.2 Hox transcription factors general introduction

Hox proteins, the subject of this thesis, are a class of specific transcription factors that control the pattern of body formation. Clusters of Hox genes encode transcription factors that direct cells to form various parts of the body. Hox transcription factors were first identified in the fruit fly *Drosophila melanogaster* (Lewis 1978), although they function during normal development of all metazoans (Gehring and Hiromi 1986). A common feature of Hox proteins is the presence of a highly conserved, 60 amino acid helix-turn-helix motifs termed the homeodomain (HD) (McGinnis 1984) which mediates sequence-specific DNA binding at specific sites. Once bound, Hox proteins can either activate or repress transcription of their target genes to collectively guide development of organisms, as extensively reviewed (Pearson et al. 2005, Levine and Hoey 1988, Scott et al. 1989, McGinnis and Krumlauf 1992). Homeobox (*Hox)* genes were discovered following the observation of two striking mutations in the *Drosophila melanogaster* (Figure 1.1). In the *Antennapedia* mutation, the antennae develop as legs, whereas in the bithorax mutation, the haltere (a balancing organ on the third thoracic segment) is transformed into part of a wing.

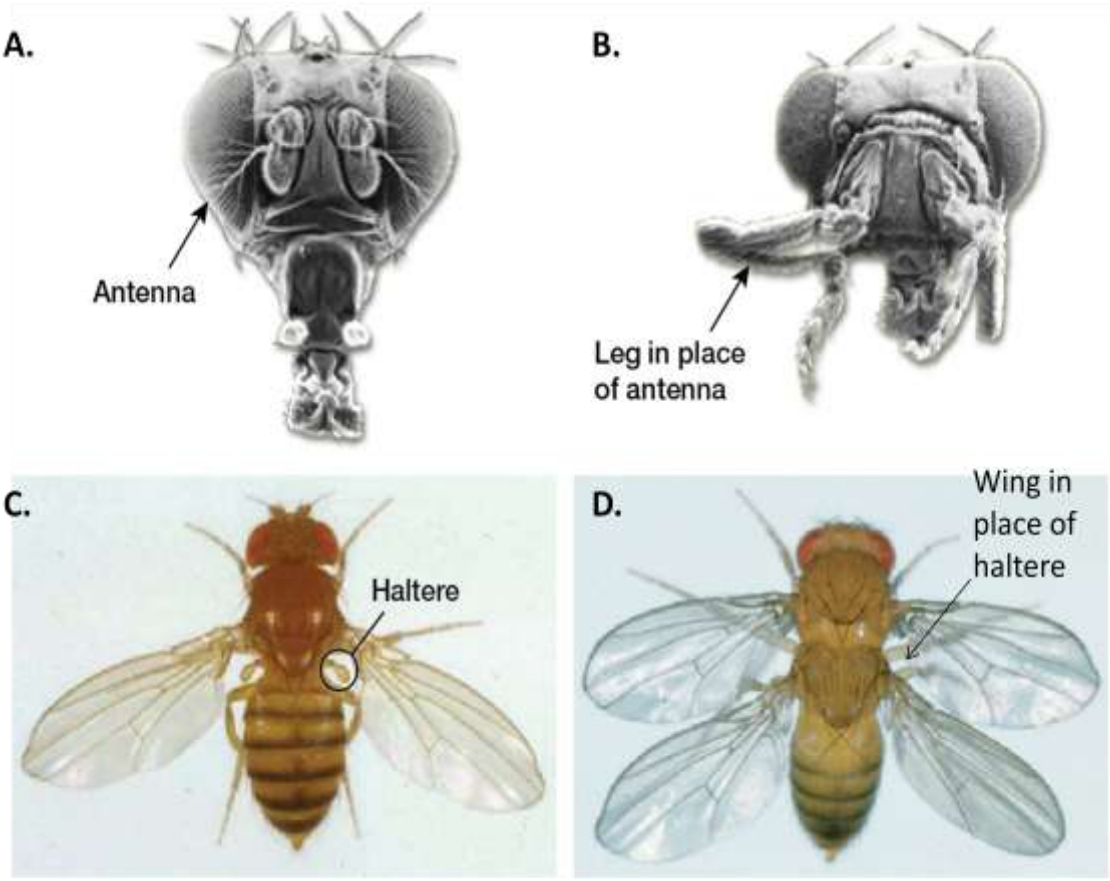

Figure 1.1 *Drosophila* homeotic transformations in the Antennapedia complex and the Bithorax complex. Top: **(A)** Normal *Drosophila* head; **(B)** *Drosophila* head with mutation in *Antennapedia* gene. **Bottom**: **(C)** Normal *Drosophila*; **(D)** *Drosophila* with a homeotic mutation producing two thoraxes. Bottom images courtesy of the Archives, California Institute of Technology. Modified from: Genetic Science Learning Center, University of Utah, http://learn.genetics.utah.edu.

These changes were described as homeotic transformations from the Greek word *homeosis*, signifying a change of a complete body structure into another. It was proposed that a master regulatory gene could control the development of different regions of the fly, which were subsequently localized to the *bithorax* complex with three Hox genes (*Ubx*, *Abd-A*, and *Abd-B*) and the *antennapedia* complex with five Hox genes (*Lab*, *Pb*,

Dfd, Scr and Antp). These genes, which have been highly conserved throughout evolution, are expressed during embryonic development in a highly coordinated manner (Lappin et al. 2006). Understanding how this positional information is encoded and deciphered is a major scientific challenge that will be addressed in this thesis, with a focus on DNA binding.

1.2.1 Developmental processes regulated by Hox proteins

Hox genes are key regulators of embryonic development and differentiation (Marx 1992) found in metazoans (Ulijaszek et al. 1998). Understanding the regulatory mechanism by which this is achieved, remains a major scientific challenge. In particular, how these highly conserved Hox proteins control a set of downstream regulatory genes that guide development and influence evolution of diverse body plans remains unclear. As Hox genes were first discovered in the fruit fly *Drosophila melanogaster*, much of our understanding of these important transcription factors originates from studies in this organism.

1.2.2 Hox gene organization, colinearity, and posterior prevalence

Hox genes have been studied in many species, but with more detail in *Drosophila melanogaster* (fruit fly) and *Mus musculus (*mouse*)*. The order of genes along the chromosome corresponds to the position of their expression domain along the anterior-posterior axis of *Drosophila*. For instance, the *labial* (*lab*) and *Deformed* (*Dfd*) genes specify the head segments, while *Sex combs reduced* (*Scr*) and *Antennapedia*

(*Antp*) are required for the identities of the first and second thoracic segments, respectively. *Ultrabithorax* (*Ubx*) is responsible for specifying third thoracic segment identity, and *Abdominal A* (*Abd-A*) and *Abdominal B* (*Abd-B*) determine abdominal segment fates. *Hox* genes confer axial positional information to emerging embryonic tissues in all three germ layers. In mammals, 39 *Hox* genes are organized in four different clusters (*HoxA, HoxB, HoxC*, and *HoxD*) found at four distinct chromosomal loci (Scott 1992, Taniguchi 2014) (Figure 1.2). These clusters are the result of two whole genome duplication events during the emergence of the vertebrates. Based on similarities between the *Hox* genes in vertebrates and *Drosophila*, these genes are classified into 13 homology groups, referred to as paralogs (Scott 1992, Meyer 1996, Greer et al. 2000).

As observed in *Drosophila*, in vertebrates the order of these paralogs on their respective chromosomes is colinear with expression pattern along the anterior-posterior axis of the developing animal. In *Drosophila melanogaster* all *Hox* genes are activated almost simultaneously during the cellular blastoderm stage (Negre et al. 2005, Wilkinson et al. 1989, Hunt et al. 1991, Iimura and Pourquie, 2006). In vertebrates, there is a temporal sequence of activation of Hox genes, such that 5′ genes are expressed later and more posteriorly along the A/P axis, whereas those located at the 3’ end of the complex are expressed earlier and more anteriorly- a phenomenon known as spatial colinearity (Iimura and Pourquié 2007, Soshnikova and Duboule 2008).

Gain-of-function and loss-of-function mutations in individual *Hox* genes alter the identity of tissues, resulting in a collection of bizarre phenotypes (Lewis 1978, Kaufman

et al. 1980). In the two most famous Hox mutations, the gain-of-function mutant *Antennapedia* (*Antp*) phenotype transforms the head antenna into an extra thoracic leg, and *Ultrabithorax (Ubx)* loss of function transforms the third thoracic segment (T3) into the second one (T2), including the development of a second pair of wings (Figure 1.1). This phenomenon, termed "homeosis", was originally described back in the late 19th century based on spontaneous variants (Bateson 1894). Once identified, the genes responsible for these mutations were named "HOM-C" and later "Hox" genes.

Homeotic transformations comparable to those in *Drosophila* occur in mice as well. Hox knockout mice show that the identity of the body along the anterior-posterior axis is primarily determined by the posterior-most *Hox* gene expressed in that segment (Iimura and Pourquié 2007). This came from the observation that disruption of all *Hox10* paralogs in mice results in the conversion of lumbar vertebrae into thoracic vertebra-like structures with rib projections. Similarly, when all *Hox11* paralogs in mice are deleted, sacral vertebrae are transformed into vertebrae with lumbar identity (Wellik and Capecchi 2003), indicating a similar mode of action. In both cases, the most posterior Hox protein determines the tissue identity. In general, any Hox protein is able to repress expression of any more anterior Hox gene, a phenomenon known as posterior prevalence. In Section 5, I discuss a possible mechanism explaining this phenomenon.

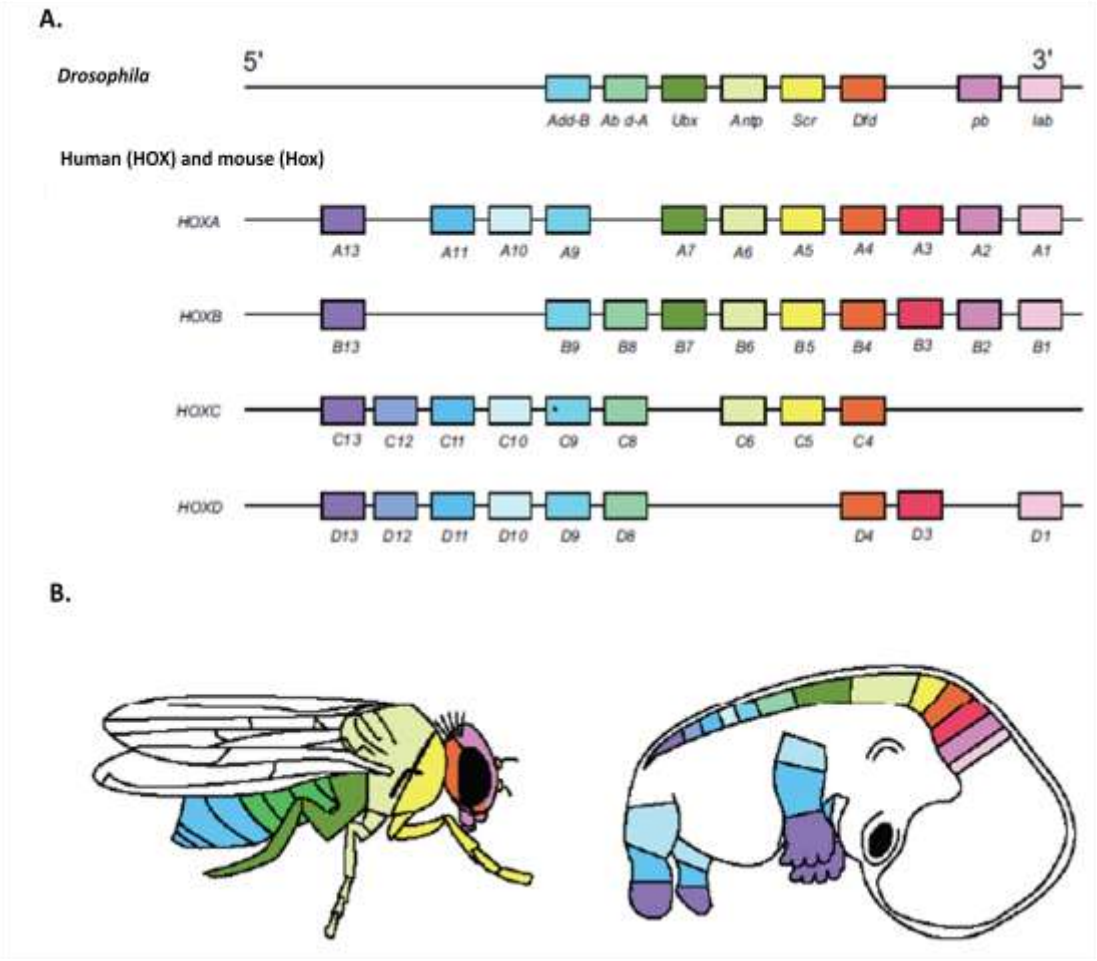

Figure 1.2 Arrangement of *Hox* genes in the *Drosophila* and mammalian genomes. A) In *Drosophila*, eight *Hox* genes clustered on a single chromosome, the homeotic complex (*HOM-C*), are divided into two groups: the Antennapedia complex (ANT-C) and Bithorax complex (BX-C) as indicated. In mammals, 39 *Hox* genes are divided into four separate clusters (*HoxA*, *HoxB*, *HoxC*, and *HoxD*) on four different chromosomes. In each cluster, *Hox* genes are tandem arranged in sequence from 3′ to 5′. *Hox* genes with the same number are referred to as paralogs. *Drosophila* and Human Hox protein with a common ancestor are depicted by boxes with the same color (Taniguchi 2014) **B)** The color coding in panels A and B shows the correspondence between the genomic order of Hox gene protein expression along the main body axis in Drosophila and human. This image was derived and modified with permission under the terms of the Creative Commons Attribution License https://creativecommons.org/licenses/by/2.5/legalcode. "This research was originally published in Journal of Current Genomics. Durston A, Wacker S, Bardine N, Jansen H. Time Space Translation: A Hox Mechanism for Vertebrate A-P Patterning. *Journal of Current Genomics*. 2012; 13(4):300-307. ©2012 Bentham Science Publishers."

1.2.3 Hox function in adults

Hox genes are also expressed in adults and may participate in various processes in the adult body. Expression in adult cells constitutes a form of positional memory, providing crucial information for a cell to map its location within a multicellular organism (Wang et al. 2009). For animals capable of regeneration such as the freshwater flatworm planarian, Hox proteins are involved in this process as well (Gurley et al. 2008, Petersen and Reddien 2008). Planarians regenerate when either a head or tail, at the appropriate location, if it is removed.

During wound healing upon wound repair, Hox genes are responsible for regionalization and specification of the newly formed structures. Missing tissues can be restored following injury in many animals (Sánchez 2000). These re-development programs are capable of generating many tissues and organs from a very different physical starting point than that of embryonic development.

Deficiency in Hox genes impairs embryonic development and wound healing. In adults dysregulation of Hox genes that maintain differentiated tissue can lead to pathological tissue remodeling as well as tumor formation (Mack et al. 2003, Carrio et al. 2005, Rauch et al. 2007).

Defining the mechanisms that regulate context-specific Hox function at the amino acid level is crucial to comprehend the complex biology of Hox function in normal and diseased tissue. Hox proteins regulate many growth factors, oncogenes, and tumor suppressor genes, including p53 (Chu et al. 2004, Raman et al. 2000, Zhang et al. 2003, Naora et al. 2001) and Hox mis-regulation impacts all stages of tumor progression

(Samuel and Naora 2005). Consequently, alterations in Hox activity have been implicated in the formation and progression of a wide variety of cancers including tumors in the brain, prostate, lung, breast, skin, kidney, colon, uterus, cervix, ovary, and hematopoetic cells (Cillo et al. 1996, Cillo et al. 2001, Abate-Shen 2002, Svingen and Tonissen 2003).

Although molecules that target Hox proteins show promise as cancer therapeutics (Morgan et al. 2007, Kelly et al. 2011), treatments are needed that abrogate tumor-promoting Hox activities without mis-regulating normal Hox function. The first step in developing such therapeutics is defining tissue-specific regulatory mechanisms in Hox proteins. In Section 3, we propose that long range interactions that regulate DNA binding are disrupted in Hox mutants associated with tumor progression.

1.2.4 DNA binding by the homeodomain in Hox proteins

The early work of Lewis (1978) was able to characterize key genetic loci controlling *Drosophila* development, including Hox genes (McGinnis et al. 1984, Scott and Weiner 1984). Years later, these labs discovered that these loci encode proteins containing a homeodomain, a structured, stable and conserved 60-amino acid section of protein. The first homeodomain structure was originally characterized by nuclear magnetic resonance (NMR) spectroscopy (Qian et al. 1989, Billeter et al. 1990) on the Antp homeodomain and revealed 3 helices and a disordered N-terminal arm (Muller et al. 1988) (Figure 1.3). DNA is done mainly by the third α-helix (αIII) and a flexible N-terminal arm. The N terminal arm establishes contacts through hydrogen bonds with

exposed bases in the minor groove and sugar phosphate backbone. The third helix contributes most to DNA recognition, and hence it is often called the "recognition helix". It binds to the major groove of DNA through hydrogen bonds and Van der Waals interactions with exposed bases. The second α-helix stabilizes the interaction between protein and DNA, but does not play a particularly strong role in its recognition (Matthews et al. 1982). The recognition helix and its preceding helix always have the same relative orientation (Wintjens and Rooman 1996, Brennan and Matthews 1989).

The recognition of DNA sequences by transcription factors, must be sequence specific to restrict binding to true target sites from the large excess of potential binding sites that are typically available in the genome. The overall affinity of a protein-DNA complex is normally ensured by opposite charges between DNA phosphate groups and DNA-binding proteins. Specificity, on the other hand, requires intermolecular interactions, direct or water-mediated, between the protein side-chains and the DNA bases (Gutmanas and Billeter 2004). Furthermore, indirect readout is defined as protein-DNA interactions that depend on base pairs that are not directly contacted by the protein (Rohs et al. 2010). Different high-throughput screening assays such as high resolution analysis of sequence preferences using protein binding microarrays (PBMs) (Mukherjee et al. 2004) and DNA footprinting analysis indicated a conserved seven or eight bases as the binding preference for a homeodomain monomer (Banerjee-Basu et al. 2003, Sandelin et al.2004). Numerous homeodomain-containing proteins tested all show a preference for the conserved consensus sequence 5’-TAATTA-3’ (Berger et al. 2008,

Noyes 2008, Gutmanas and Billeter 2004). However variations in this sequence, especially at the 3' end, elicit only small changes in affinity (Ekker et al. 1991).

Insights into the mechanisms of sequence-specific DNA binding by homeodomains have been elucidated by utilizing methods such as x-ray crystallography, NMR spectroscopy, molecular dynamics simulations, and atomic-resolution structural studies. The three-dimensional structures have been determined for individual protein-DNA complexes coupled with directed mutagenesis and biochemical analysis (Gehring et al. 1994, Ades and Sauer 1995, Joshi et al. 2007, Gutmanas and Billeter 2008, Garvie and Wolberger, 2001). Together, these high resolution views have classified homeodomains cognate DNA recognition (Garvie and Wolberger, 2001, Harrison and Aggarwal 1990, Pabo and Sauer 1992, Joshi et al. 2007).

Within this classification system, Homeodomain-DNA contacts are nearly identical for Hox proteins (Noyes et al.2008, Chu et al. 2012, Kissinger et al. 1990, Wolberger et al. 1991). This conserved binding geometry allows differences in amino acid sequence and DNA-binding specificity for various homeodomains to be interpreted within a common structural framework. Residues at positions 2, 3, and 5–8 on the N-terminal arm, as well as residues at positions 47, 50, 51, 54, and 55 on the recognition helix, all contribute to DNA-binding specificity (Figure 1.4) (Ades and Sauer 1995, Noyes et al. 2008, Wolberger et al. 1991, Passner et al. 1999, Piper et al. 1999). Most studies have focused on studying protein-DNA interactions with only the homeodomain (Beachy et al. 1993) since full length Hox proteins are prone to aggregation and proteolysis (Bondos and Bicknell 2003). To combat these problems, I have developed a

novel purification technique (Churion et al. 2016). The details of this method will be discussed in Section 2.

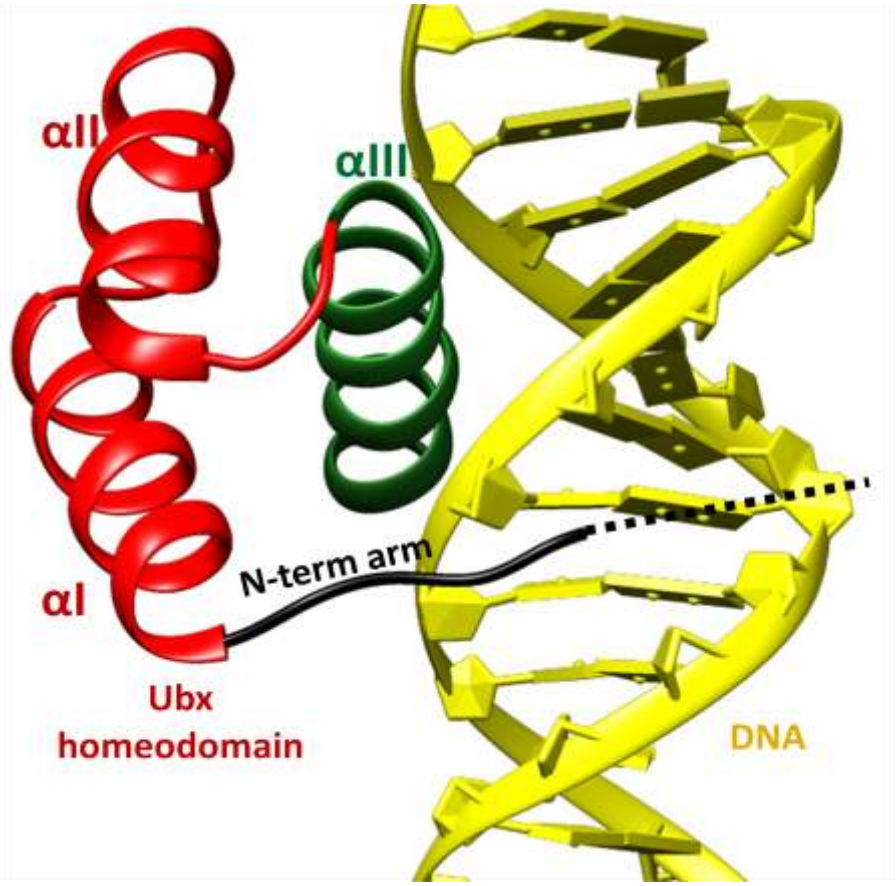

Figure 1.3 The Homeodomain Tertiary Structure. Homeodomain from the *Drosophila* Ultrabithorax protein (structure obtained from pdb 1B8I). The first two helices (in red marked αI, αII) ,the third recognition helix (in green marked αIII), and the disordered N-terminal arm (shown in black and dotted line to indicate disorder) making up the *Drosophila* Ultrabithorax homeodomain. This structure is prototypic for all homeobox proteins, serving as the core structure even for divergent homeobox families. DNA double helix is shown in yellow.

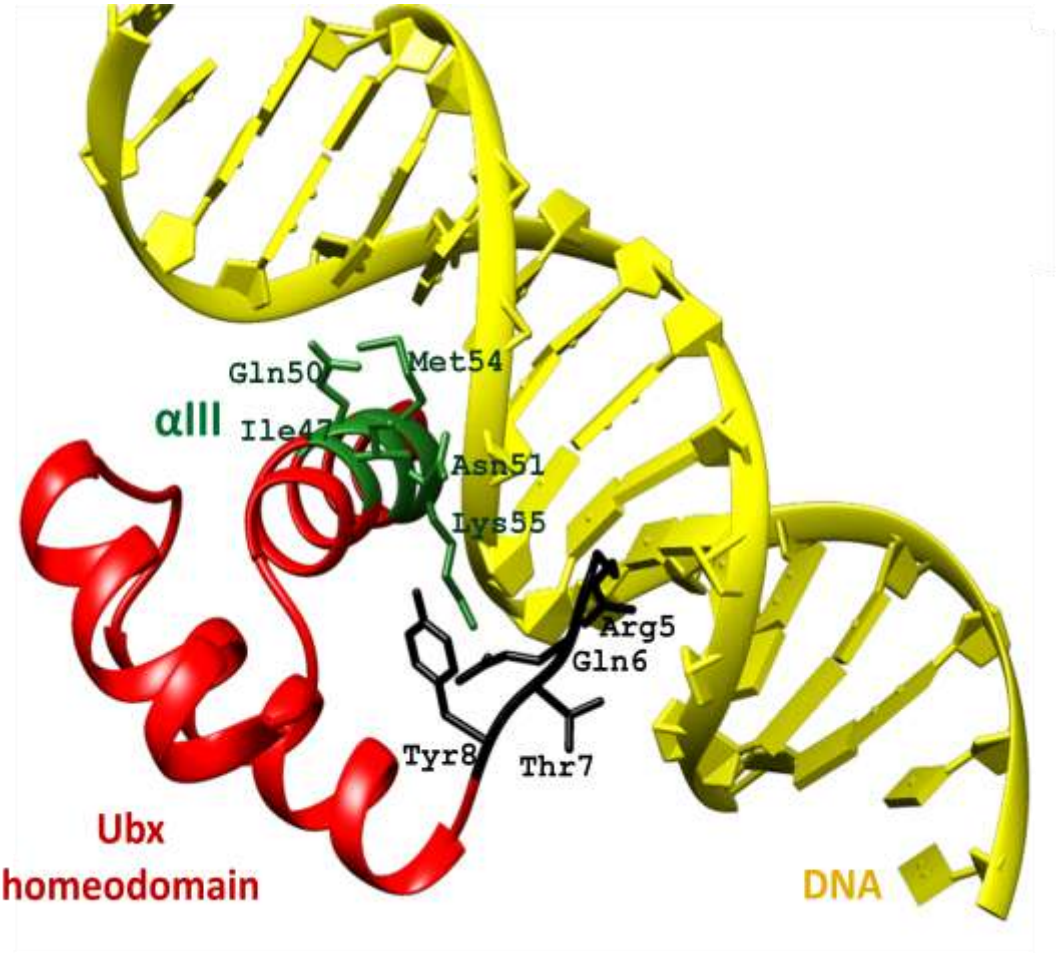

Figure 1.4 Protein-DNA interface for the *Drosophila* Ultrabithorax protein (structure obtained from pdb 1B8I). Amino acid backbone residues that contribute to Homeodomain-DNA Interactions Between Ultrabithorax N-terminal arm plus the recognition helix to the canonical core are shown. Contacting residues 2 and 3 (not shown, structure not available in pdb file due to intrinsic disorder) and residues 5-8 (shown in black) of the N terminal arm make contact with the minor groove of DNA. Residues of the recognition helix (αIII) 47, 50, 51, 54, and 55 (shown in green) make contacts to the DNA in the major groove. Amino acids docking in the major groove are further thought to make more stable contacts and thus ultimately impose a higher specificity. Nucleotides 5 and 6 are also thought to make specific contact with the homeodomain.

1.2.5 The Hox paradox

To regulate development, individual Hox family members control transcription of unique downstream genes, yet the homeodomain of each Hox protein mediates DNA binding with high affinity but extremely low specificity when binding to specific genes as monomer. Ultrabithorax and Antennapedia homeodomains are 90% identical, including all DNA-contacting residues (Table 1.1) (Hoey and Levine 1988, Kalionis and O'Farrell 1993, Bondos and Tan 2001). Consequently the DNA affinity and specificity of their homeodomains *in vitro* are also very similar (Beachy et al. 1993, Passner et al. 1999). However, *in vivo* Ultrabithorax and Antennapedia regulate different genes and drive development of unique body structures.

The disparity between the absolute requirement for distinct Hox activities *in vivo* and the similarity of homeodomain-DNA recognition *in vitro* has been termed the "Hox paradox" (Hueber and Lohmann 2008, Prince 2002, Mann and Chan 1996). Contributing to the Hox paradox is that fact that only a subset of Hox targets have been identified and characterized, however, given their hypothesized hierarchical functions, the number of such targets would be expected to be much larger. *In vitro* analysis suggests that all Hox proteins display poor DNA-binding specificity and recognize very similar consensus core with high affinity (Agrawal et al. 2011).

Further investigation of Hox genes *in vivo* suggested that the functional specificity seen for their encoded transcription factors would lay mainly within the highly conserved homeodomain (HD) and N-terminal arm plus a hexapeptide motif upstream of the N-terminal arm (Gehring et al. 1994, Ades and Sauer 1995, Joshi et al.

2007, Gutmanas and Billeter 2008, Garvie and Wolberger, 2001). These regions have largely been conserved across species, from flies to humans, whereas the remainder of the protein structure has not (Kuziora and McGinnis 1989, Mann and Hogness 1990; Lin and McGinnis 1992; Chan and Mann 1993; Zeng et al.1993). Obviously, homeodomain-containing Hox transcription factors elicited distinct and different effects on downstream target genes, but through what mechanisms? Clearly homeodomain-DNA interaction is not the only source for biological specificity.

This paradox is partially resolved through Hox interactions with other transcription factors, increasing specificity by requiring tandem Hox and partner DNA binding sites (Bondos et al. 2004, Bondos et al. 2006, Joshi et al.2007). Since the expression and activity of many Hox partners is limited to specific tissues, protein interactions potentially provide contextual information to Hox proteins as well as contribute to target site selection. However, a subset of Hox-regulated enhancers lack sites for known Hox partners. Furthermore, Hox proteins can regulate reporter genes in a DNA-specific manner in *Drosophila* cell culture in absence of partner proteins (Tan et.al 2002).

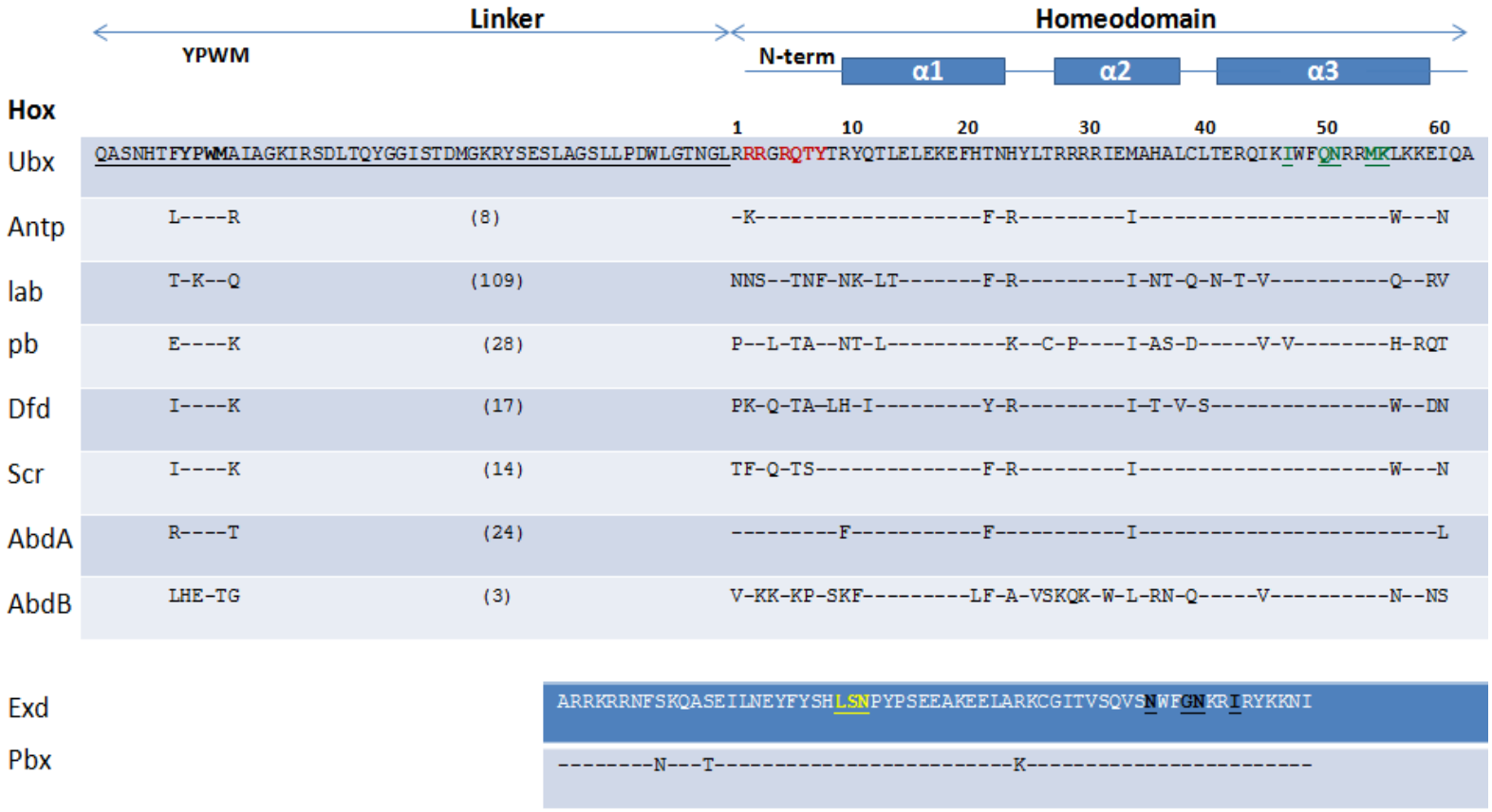

Table 1.1 *Drosophila* Hox and cofactor protein alignment. Top, The sequence of the Ubx is aligned with the seven other *Drosophila* Hox proteins. The homeodomain is numbered from 1 to 60 and residues N-terminal to the homeodomain that contain the YPWM (bold black underline) motif are underlined. The distance between the YPWM motif and the homeodomain are shown in parentheses for each Hox protein. There is no consensus YPWM motif in AbdB but a conserved tryptophan is required to form cooperative heterodimers with Exd. Shown above the homeodomain sequences are the relative **Table 1.1 continued.** locations of the α-helices. The residues that normally contact DNA bases in the major groove in homeodomains are in bold green and underlined and residues in the N-terminal arm that contact DNA in the minor groove are bold maroon (no underline). **Bottom**, the sequence of the Exd is aligned with the sequence of the human oncoprotein Pbx1. The three-amino-acid insert in the Exd homeodomains is underlined and bold yellow. The residues that normally contact DNA bases in the major groove in homeodomains are underlined bold black (**Passner et.al. 1999**).

1.2.6. Possible solutions to the Hox paradox

1.2.6.1 Regulation of binding in full-length Ubx by intrinsically disordered regions

The Ubx homeodomain is remarkably similar to previously characterized homeodomains. Superimposing the crystal structures of the Ubx homeodomain with Antennapedia (Antp) homeodomain revealed an average root mean-square deviation (r.m.s.d.) of only 0.41 Å (for Cα atoms in HD), suggesting only very small differences exist in the DNA-bound structures (Fraenkel and Pabo 1998; Passner 1999). The lack of variation in structure extends to the Ubx and Antp N-terminal arms in which the first four residues are disordered. Regions outside of the HD that don't directly interact with DNA also have effect on DNA binding affinity and specificity (Liu et al. 2008, Liu et al. 2009). Therefore our lab has been exploring the hypothesis that amino acid sequences outside the homeodomain contribute to differences in binding site selection by Hox proteins during animal development. These nonhomeodomain sequences vary significantly between Hox proteins and thus potentially permit distinct Hox functions *in vivo*. However, little is known about protein structure, function, and potential intramolecular interactions outside the homeodomain.

The presence of large intrinsically disordered regions is common among transcription factors compared to other types of proteins (Liu et al. 2006). Disorder is expected to play a regulatory role in target site selection for transcription factors with a large number of DNA binding targets (Singh et al. 2007). Since Hox proteins are known to regulate a vast array of target genes during development in the adult organism

(Pearson et al. 2005) the extensive disordered regions found in Ubx could potentially be involved in regulation of DNA binding. There is a difference in the DNA binding affinities of full-length Ubx and UbxHD (only Ubx HD) which confirms that sequences outside the Ubx homeodomain modulate DNA binding by this region (Liu et al. 2008, Liu et al. 2009). In order to determine which non-homeodomain regulatory regions were important for DNA binding, a series of Ubx deletion mutants were analyzed via DNA binding assays to determine regions important for DNA binding regulation (Liu et al. 2008). The sequential evaluation of the DNA binding affinity by these Ubx variants indicated that the majority of the UbxIa amino acid sequence either binds DNA or modulates DNA binding affinity. This finding indicated that non-homeodomain intrinsically disordered regions are crucial for DNA binding regulation.

Disorder outside of the homeodomain does not only occur in Ubx; this feature is conserved in all *Drosophila* Hox proteins (Figure 1.5) suggesting that intrinsically disordered regions must be crucial for function. The presence of disorder is also conserved across multiple species, further supporting this hypothesis (Liu et al. 2008, Galant and Carroll 2002). The observation that disordered regions in Ubx influence DNA binding affinity by the homeodomain opens new opportunities to further investigate potential mechanisms for regulating DNA binding.

In this thesis, we identified amino acids located outside the homeodomain that are able to interact with the surface residues on the homeodomain and form a regulatory network that regulates DNA binding (Section 3). With the available information we developed the first model for a full length Hox protein (Ubx).

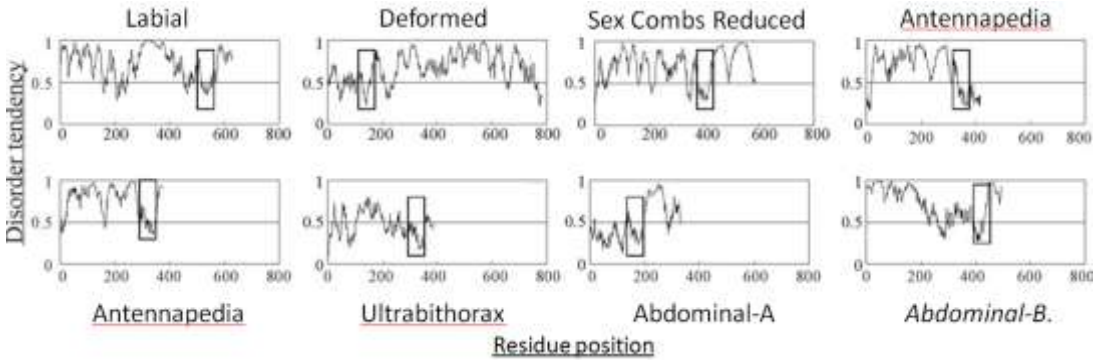

FIGURE 1.5 Prediction of disordered regions in all *Drosophila* melanogaster Hox proteins. Disorder tendency was predicted by the IUPred algorithm. Values above 0.5 (black solid line) indicate predicted disorder in the corresponding region. Homeodomains are enclosed in black boxes. Each graph depicts data for a different Drosophila Hox protein as labeled. In general, Hox proteins are highly disordered. Figure was obtained with the appropriate permission from Liu et al. 2008. "This research was originally published in Journal of Biological Chemistry. Liu Y, Matthews KS, Bondos SE. Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila hox protein ultrabithorax. *Journal of Biological Chemistry*. 2008; 30:20874-87. © the American Society for Biochemistry and Molecular Biology**."**

1.2.6.2 Protein interactions

Interactions with other proteins potentially provide Hox protein with specificity in transcription regulation (Mann and Chan 1996, Beachy et al. 1988). Hox complex formation with Exd can alter the interaction of residues within and beyond the N-terminal arm with DNA (Joshi et al. 2007). Furthermore, Exd is a well-studied Hox cofactor in *Drosophila*. The conserved Hox hexapeptide FYPWMA motif makes direct contacts with the TALE motif in Exd (Chan and Mann 1996, Joshi et al. 2007, Passner et al. 1999, Johnson et al. 1995). Thus, while the primary sequence determinants within the

N-terminal arm help define sequence preferences, intramolecular or intermolecular protein-protein interactions can also influence DNA recognition. (Noyes 2008).

To understand the structural basis of Hox-Extradenticle pairing, the crystal structure of an Ultrabithorax-Extradenticle-DNA complex at 2.4 Å resolution was determined, using the minimal polypeptides that form a cooperative heterodimer (Passner et al. 1999). The Extradenticle (Exd) homeodomain diverges from the canonical homeodomain, with two unusual DNA-recognition residues and three extra residues in the loop between helices α1 and α2 (Table 1.1 and Figure 1.6).

The Ultrabithorax and Extradenticle homeodomains bind opposite faces of the DNA, with their DNA-recognition helices almost touching each other. Most of the cooperative interactions arise from the hexapeptide motif of Ultrabithorax, located amino-terminal to its homeodomain, which forms a reverse turn and inserts into a hydrophobic pocket on the Extradenticle homeodomain surface. The conformation of DNA changes noticeably upon Hox-Exd interaction, significantly broadening the DNA minor groove occurs. Exd is orthologous to the protein Pbx protein family, which binds DNA cooperatively with the mammalian Hox proteins as Exd does with *Drosophila* Hox proteins by interacting with the same conserved motif (Passner et al.1999).

Hox and Exd/Pbx1 protein interactions independent of the YPWM motif have also been reported (Chan et al. 1994, Hudry et al. 2012) and mapped to residues predicted to interact with Exd located C-terminally to Hox homeodomains. Although protein-protein interactions might partially resolve the Hox paradox, a conundrum still remains. Seven out of the eight Hox proteins can bind Exd via the hexapeptide motif

(with the exception of AbdB which does not include a hexapeptide motif) to bind a composite Exd-Hox recognition sequence establishing similar contacts to DNA yet each Hox protein must bind different genes (Mann et al. 2009). Therefore this interaction is not unique to a specific Hox protein, and cannot explainhow different Hox proteins select different target genes.

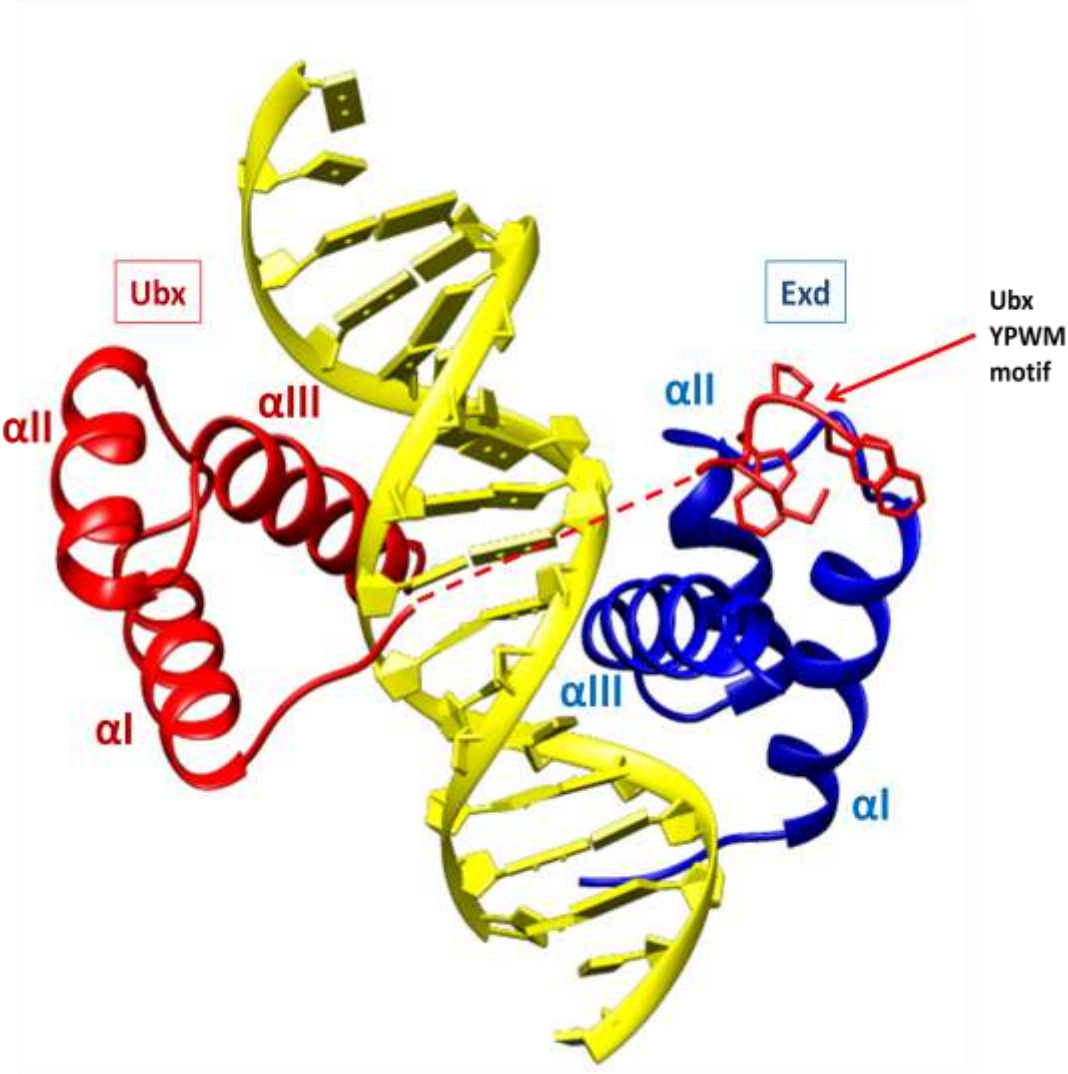

Figure 1.6 Ubx-Exd-DNA ternary complex. Ubx-Exd-DNA image from pdb 1B8I**.** The diagram shows the Ubx (red) and Exd (blue) homeodomains binding in a tandem manner on opposite faces of the DNA (yellow). The dashed red line represents the disordered linker between the Ubx homeodomain and its YPWM motif. The YPWM motif reaches into a hydrophobic pocket on the surface of the Exd homeodomain (Passner et al. 1999).

1.2.6.3 Other protein interactions

A handful of Hox cofactors have now been identified, including another TALE-class protein Meis, however their interfaces remain undefined. Ubx physically interacts with 39 known partner proteins with a wide variety of molecular functions (Hsiao et al. 2014), potentially improving binding specificity based on the partner protein present in specific tissues. Partner topology is a key aspect of protein interactions formed by the intrinsically disordered regions of the *Drosophila* Hox protein Ubx where it binds structurally similar, yet widely diverse proteins with very different chemical natures and molecular functions (Hsiao et al. 2014). Therefore, the disordered regions of Ubx, which interact with these partner proteins, may need to be positioned in a specific manner to maximize partner interactions and enable cooperative DNA binding *in vivo.*

1.2.6.4 Post-translational modifications and alternative splicing

Post-translational modifications alter the regulatory interfaces of proteins so as to induce gain, loose, or exchange binding partners, thereby affecting function at many levels (Van Roey et al. 2013). Alternative splicing and phosphorylation of Ubx are both tissue-specific, creating a mixture of Ubx functional states in any cell (O'Connor et al. 1988, Gavis and Hogness 1991, Lopez and Hogness, 1991). However, different forms of the Ubx protein, dominate in each tissue (Gavis and Hogness 1991, Liu et al. 2008, Kim et al. 2010, Reed et al. 2010, Fuxreiter et al. 2011, de Navas et al. 2011).

Alternative splicing is a mechanism that produces protein isoforms from the same precursor mRNA by either retaining or excluding different exons. This process

generates proteins with different primary sequences and hence, potentially alters protein functions. Alternative splicing occurs in all eukaryotic lineages (Black 2003) and becomes more prevalent as organism complexity, estimated by the number of different cell types, increases (Chen et al. 2014). A disproportionate number of transcription factors with intrinsically disordered protein (IDP) domains, are alternatively spliced, expanding their functional and regulatory diversity (Liu et al. 2006, Vuzman and Levy 2012).

The *Drosophila* Hox gene Ultrabithorax (Ubx) produces a family of protein isoforms through alternative splicing in a stage- and tissue-specific manner (Lopez et al. 1996). The isoforms are not interchangeable during development (Reed et al. 2010). The spatial distribution of Ubx isoforms has been conserved in species that diverged 60 million years ago, suggesting a functional requirement for the tissue-specific expression (Reed et al. 2010, Bennett et al. 1999).

The Ubx proteins are encoded in a large transcription unit of 77 kb that generates mRNAs of 3.2 kb and 4.3 kb, the Ubx unit. The differential use of the two mRNAs and the inclusion or exclusion of three microexons results in the translation of five protein isoforms (Figure 1.7) (Burnette et al. 1999). The five alternative *Ubx* mRNAs share large protein-coding 5' and 3' exons but differ in the pattern of incorporation of three elements: the B element is located between two alternative donor sites at the end of the first common exon, whereas mI and mII are internal cassette exons. These three elements are located in the protein sequence between the conserved hexapeptide motif and the homeodomain. Therefore splicing will change the length of a region known to

have an effect on DNA binding (Liu et al. 2009). The impact of splicing on DNA binding by Ubx is explored in the Appendix.

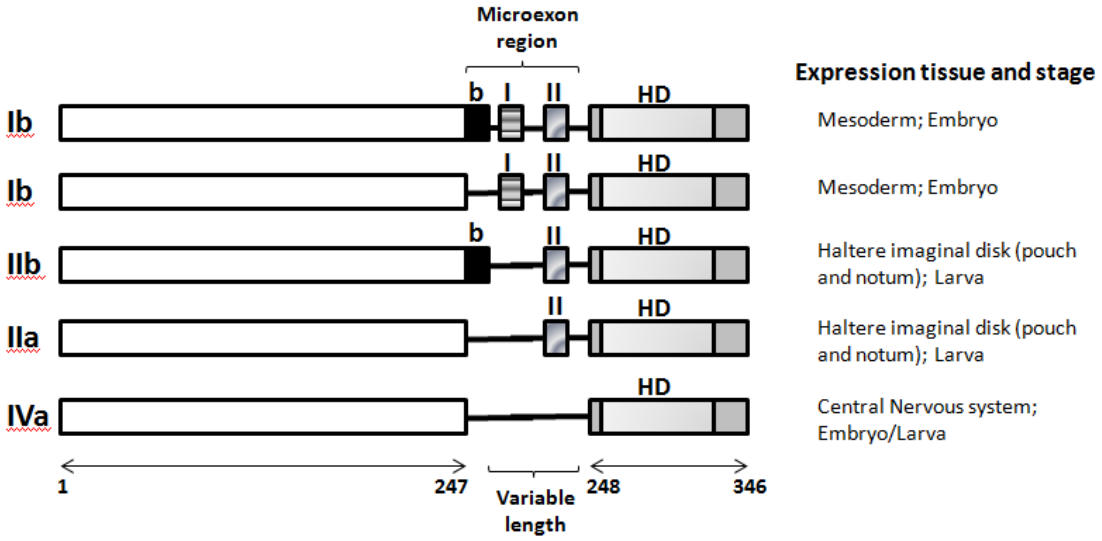

Figure 1.7 Schematic representations of natural Ubx isoforms and expression domain *in vivo*. Primary structure of the UBX protein isoforms. Regions encoded by separate exons are represented as rectangles connected by horizontal lines. Amino acid coordinates for the common regions are indicated beneath isoform IVa. HD identifies the homeodomain; b, I, and II identify the optional elements (Lopez et al. 1996).

1.2.6.5 Phosphorylation

The functions of IDPs are also modulated by phosphorylation (Iakoucheva et al. 2004, Romero et al. 2006, Singh and Dash 2007). For example, DNA binding by the transcription factor Ets-1 is allosterically coupled to a serine-rich region (Lee et al. 2008). The *Drosophila* Hox protein Ubx is multiply phosphorylated (Gavis and Hogness 1991), including at sites within its intrinsically disordered domain, which regulates DNA

binding, protein-protein interactions and transcription activation (Tan et al. 2002, Hsiao et al. 2011, Liu et al. 2008, Liu et al. 2009, Gavis and Hogness 1991). Given that phosphorylation has the potential to regulate as well as coordinate multiple transcription factor functions, Ubx has the potential to use this mechanism in order to increase specificity of interaction.

1.3 Ultrabithorax

Our model is Ultrabithorax (Ubx), a member of the *Drosophila* Hox gene family which functions specifies the posterior legs and halteres, as well as portions of the ectoderm, gut, heart, musculature, and central and peripheral nervous systems (Figure 1.2 B) (Hughes and Kaufman 2002). The *Drosophila melanogaster* Hox protein Ultrabithorax (Ubx) is the best model for exploring regions outside of the homeodomain that modulate DNA binding because it is one of the few full-length Hox proteins that can currently be purified in a soluble, active form (Liu et al. 2008, Liu et al. 2009, Bondos and Bicknell 2003). This information allows us understand the intramolecular mechanism of the Ubx Hox protein. In addition, more natural DNA binding sites have been identified for Ubx than for any other Hox protein in any species (Pearson et al. 2005) providing the substrates needed for DNA binding experiments. In Section 5 we discuss how intramolecular interactions might affect tissue-specific selection of DNA binding by a single Hox protein.

In Section 3 we show that evolutionarily conserved motifs located within intrinsically disordered regions interact with the homeodomain and/or each other to alter

the relative binding affinity for a single Ubx optimal site. Identifying Hox protein molecular mechanisms will develop a better understanding of Hox function in normal and diseased tissues provide new molecular tools for investigating these processes, and provide a guide for investigating other Hox proteins.

1.4 Ubx self assembly into protein biomaterials

Protein-based materials potentially combine outstanding mechanical properties, biocompatibility, and biodegradability with the potential to engineer these traits or append novel functions to create novel and useful hybrid biomaterials. Protein-based materials have been used in a variety of applications, including drug delivery, surgical sealants and tissue engineering (Deming 2007, Place et al. 2009). These applications require materials to be compatible with biological systems (Astbury and Woods 1934, Grevellec et al. 2001, Rodriguez-Cabello et al. 2007, Velema and Kaplan 2006) and to exhibit diverse morphologies, suitable mechanical properties, appropriate functional properties (Maskarinec and Tirrell 2005), and the ability to be further functionalized to form hybrid materials.

Incorporating DNA-binding proteins into materials is an important goal for biotechnology. DNA-binding protein interactions are very high affinity, functionally diverse and biocompatible. Therefore, exploiting these non-covalent interactions is of great interest. Due to the presence of intrinsically disordered regions, Ubx is prone to aggregation. Upon identifying conditions that inhibit amorphous aggregation (Bondos and Bicknell 2003) it was discovered Ubx spontaneously assembles into ordered

aggregates *in vitro* (Greer 2009). Ubx materials are transparent; diffract light, and X-ray diffraction and Thioflavin T binding studies of Ubx materials lack any indications of amyloid structure (Howell et al. 2015).

Further characterization of Ubx biomaterials determined that Ubx initially forms globular aggregates, which further rearrange to form nanoscale fibrils (Majithia et al. 2011). The fibrils further associate to generate macroscopic films, which are the foundation for a range of macroscale Ubx architectures such as fibers, sheets, and bundles (Greer et al. 2009, Huang et al. 2010). Ubx protein biomaterials are cytocompatible, biocompatible, and non-immunogenic (Patterson et al. 2014, Patterson et al. 2015).

An advantage of Ubx protein biomaterials is that they can be further functionalized (1) with full-length proteins via gene fusion, (Huang 2011**,** Tsai et al. 2015) (2) with nanoparticles by noncovalent surface interactions (Majithia et al. 2011) , and (3) with DNA by sequence-specific recognition to form potential hybrid materials for a variety of potential applications which will be discussed in detail in Section 4.

1.4.1 Dityrosine bonds in Ubx biomaterials

Fluorescent microscopy experiments demonstrated that Ubx materials autofluoresce when excited at 305 nm (Figure 1.8) which corresponds to the emission spectrum formed by oxidation of two tyrosine residues (Howell et al. 2015). The presence of dityrosine formation in Ubx materials was confirmed via immunohistochemistry experiments, in which antidityrosine antibodies were able to

specifically recognize Ubx fibers. Upon mutagenesis and subsequent fluorescence analysis of specific tyrosine residues, it was confirmed that only two dityrosine bonds are formed by Ubx in materials (Howell et al. 2015). The resulting model from this study concluded that the N-terminus of Ubx interacts with the homeodomain forming two mutually exclusive bonds between Y4 and Y296 or Y12 and Y296, while Y167 and Y240 form a second bond. The determined dityrosine content in this study directly correlates with the strength of the materials, suggesting these bonds are intermolecular. The presence of dityrosine intermolecular interactions in Ubx biomaterials might reflect inherent intramolecular interactions that occur in Ubx which also provided an important clue for developing the first model on a Hox protein structure in Section 3.

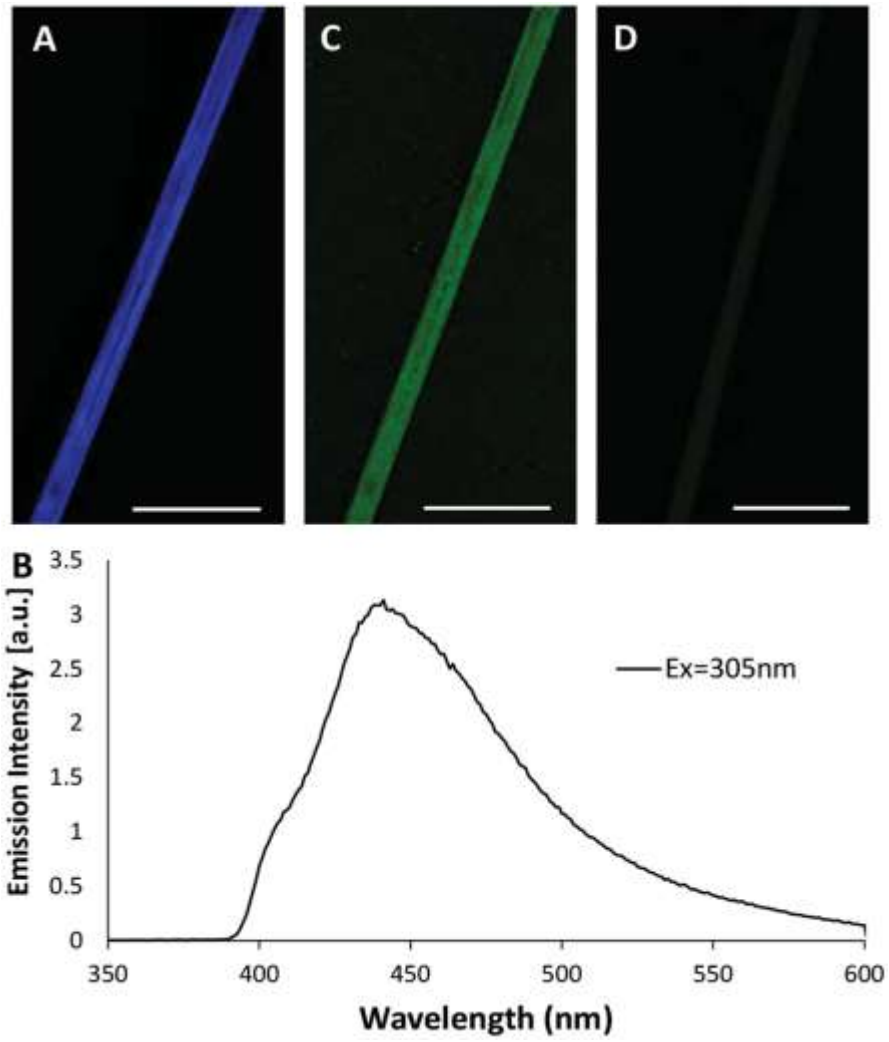

Figure 1.8 Ubx materials contain dityrosine. A) Fibers autofluoresce blue. **B)** The Ubx emission spectrum with a peak at 438 nm when excited at 305 nm is similar to other dityrosine containing proteins. **C**) Immunofl uorescence of Ubx fiber demonstrating antidityrosine primary antibodies recognize a Ubx fiber. **D)** A negative control experiment with the primary antibody omitted demonstrates that secondary antibodies do not adhere nonspecifically to Ubx fibers. Scale bar equals 30 μm in all panels. Reprinted with permission from Howell et al. 2015. License number: 3991190519734.

2. SEPARATING FULL-LENGTH PROTEIN FROM AGGREGATING PROTEOLYTIC PRODUCTS USING FILTER FLOWTHROUGH PURIFICATION[*]

2.1 Introduction

Proteolysis commonly occurs during the production and purification of recombinant proteins. Degradation or modification of proteins by proteases during expression and/or purification reduces yield and introduces contaminants similar to the full length protein. For example, proteolytic products that contain a tag or epitope may bind resin and/or full length protein during affinity purification. For a protease to cut, the protein target site must be embedded in a protein region that is at least 10 amino acids long and unfolded (Hubbard and Benyon 2001). Some proteins are more susceptible to proteolysis than others, including unstable proteins, proteins with long exposed loops, mutant proteins, molten globules, and intrinsically disordered proteins. A large proportion of gene sequences code for long stretches of amino acids that are likely to be unfolded in solution **(**Dyson and Wright 2005, Wright and Dyson 1999, Romero et al. 1998). The occurrence of unstructured regions greater than 40 residues in length is especially common in proteins that regulate critical cell responses **(**Dunker et al. 2002,

[*]Adapted with permission from "Separating Full-length Protein from Aggregating Proteolytic Products Using Filter Flow-through Purification" by Churion KA, Rogers RE, Bayless KJ, and Bondos SE.2016. Analytical Biochemistry. Copyright 2015 by Elsevier. DOI:10.1016/j.ab.2016.09.009

Uversky 2002, Dunker et al. 2015, Wright and Dyson 2015, Dunker et al. 1998, Garner et al. 1998). However, intrinsically disordered regions are more readily cut by proteases, and the presence of proteolytic contaminants can alter the apparent function of the purified protein. Because the presence of these contaminating proteolytic products can confound interpretation of downstream characterization assays, removal of these proteolytic products in a rapid and facile manner is crucial.

General approaches to solve the proteolysis problem include prevention of proteolysis before or during purification and removal of proteolytic products after the initial purification steps. Slowing protein expression can often decrease both aggregation and proteolysis before purification. This can be accomplished by lowering the incubation temperature after induction, shortening expression time, and/or changing to a less rich media (Lebendiker and Danieli 2014). A variety of protease inhibitors are frequently added to the extraction buffer to prevent and minimize proteolysis during protein purification. However, this technique is not completely effective for proteins extremely susceptible to proteolysis, such as intrinsically disordered proteins. Finally, if the proteolytic products bind the full-length protein, then they will co-purify with the full-length protein even if they lack affinity tags. An effective method to remove contamination by proteolytic products is to purify under denaturing conditions, yet this harsh method could render the protein of interest non-functional if it is not able to refold properly *in vitro*.

Existing methods to remove proteolytic products can be slow, providing time for the protein of interest to further degrade and possibly aggregate. In dialysis methods, the

full length protein is retained by the membrane, whereas proteolytic products smaller than the pores pass through. This process requires hours to days and thus is not ideal for a protein that degrades or aggregates on this timescale (Harper and Speicher 2010, Beloff and Anfinsen 1945). Gel filtration chromatography, which separates proteins by size, is also slow, generally taking hours to perform. Affinity chromatography using a tag on the protein or a specific antibody on the resin will capture full-length protein as well as any proteolytic products also containing the tag or epitope.

Alternatively, iterative affinity purification using different tags on each of the protein termini can ensure that only full-length protein, which contains both tags, is collected. Although this is an effective method, these fusion tags could potentially alter protein expression function and/or activity (Zhao et al. 2013, Tsai et al. 2015). A fast, reliable method is needed to retain protein activity as well as to safely and efficiently remove proteolytic products.

2.2 Materials and methods

Sec14L1 was amplified from Human Umbilical Vein Endothelial Cell cDNA using Accuprime Taq HIFI according to manufacturer's protocol (Invitrogen; Forward Primer: AGCGGATATCGATGGTGCAGAAATACCAG; Reverse Primer: AGCGCTCGAGAGAACCAGAACCCCTGGAGATCATGGAGCTGGAGTG). The amplicon was digested, purified, and ligated into the pQE-TriSystem vector using EcoRV and XhoI restriction enzymes. The resulting ligated product was transformed via

heat shock into BL21(DE3)pLysS cells (Promega), and positive clones were selected by colony PCR, restriction digest, and sequencing.

Sec14L1 was expressed from a single colony of transformed in BL21(DE3)pLysS cells that were grown overnight at 37 ^{o}C in a 5 mL LB starter culture with 25 µg/mL carbenicillin and 25 µg/mL chloramphenicol. The starter culture was diluted 1:100 into 200 mL LB with 25 µg/mL carbenicillin and 25 µg/mL chloramphenicol and grown at 37 ^{o}C until the OD_{600} measurement of the bacterial growth was 0.6. The bacterial culture was induced with 300 µM IPTG for 48 hours at 4 ^{o}C. To purify expressed Sec14L1 protein, 200 mL bacterial culture was centrifuged at 2,800 xg for 15 minutes at 4 ^{o}C to pellet the induced bacteria.

The pellet was resuspended in 10 mL lysis buffer (500 mM NaCl, 20 mM Na_2HPO_4, pH 8.0 with 1 mM phenylmethylsulfonyl fluoride (PMSF) and Roche protease inhibitor cocktail) and French-pressed at 1,250 psig before centrifuging at 20,000 xg for 15 minutes at 4 ^{o}C to pellet insoluble debris. The resulting supernatant was incubated with 2 mL Nickel charged His-binding resin (Novagen) at 4 ^{o}C for 2 hours. The resin was washed with 50 column volumes of lysis buffer, and loosely-bound contaminants were removed with 5 column volumes of lysis buffer with 20 mM imidazole, then 5 column volumes of lysis buffer with 40 mM imidazole, then 5 column volumes of lysis buffer with 80 mM imidazole. Protein was then eluted with 4 column volumes of lysis buffer with 200 mM imidazole. Fractions containing protein were determined by SDS PAGE and pooled. Gel electrophoresis showed Sec14L1 is on

average 57% pure, and western blots identified proteolytic product as the major contaminants (Figure 2.1B).

Ultrabithorax splicing isoform Ia (Ubx) was expressed with a his-tag in BL21 (DE3) pLysS and purified as described by Liu et al.2008 with the following modifications. Cells expressing UbxIa were harvested 150 min after 1 mM IPTG induction at mid-log phase as previously described (Liu et al. 2008, Liu et al. 2009). Cells were resuspended in 2 mL of collection buffer (100 mM NaCl, 20 mM NaH_2PO_4, 80 mM Na_2HPO_4, pH=7.5 with one tablet of Complete Proteinase Inhibitor Mixture EDTA-free (Roche Applied Science) and cell pellets corresponding to 2 L of fermentation were frozen at -20 °C.

To purify expressed Ubx from the pET19b plasmid (in order to provide a His10 tag at the UbxIa N terminus), a cell pellet corresponding to 2 L of fermented *E.coli* (final OD_{600nm} ~1.1) was thawed at room temperature and lysed in 20 mL of buffer containing 800 mM NaCl, 10 mM DTT, two tablets of Complete Proteinase Inhibitor Mixture EDTA free (Roche Applied Science), 5% glucose, 5 mM imidazole, and 50 mM NaH_2PO_4, pH 8.0, 1 mM phenylmethylsulfonyl fluoride (PMSF). Released DNA in the cell lysate was digested with DNase I (20 mg/mL, 40 µl) prior to centrifugation at 17,000 rpm for 30 min. The supernatant was mixed with polyethyleneimine (200 µl of 50% (w/v)) and recentrifuged at 17,000 rpm for 20 mins. The supernatant pH was then adjusted to column pH=8.0 and recentrifuged at 17,000 rpm for 20 mins. To prepare the affinity column, 4 mL of Ni^{+2} nitrilotriacetic acid-agarose resin (Thermo Scientific) was pre-equilibrated with wash buffer containing 500 mM NaCl, 10 mM-DTT, 5% glucose,

20 mM imidazole, and 50 mM NaH_2PO_4, pH 8.0. The clear lysate was incubated with gentle agitation at 4 °C for 20 min before being poured into the column. Packed resin was washed with 50 mL of wash buffer containing 0 mM imidazole followed by 100 mL wash of buffer containing 10 mM imidazole, 50 mL wash buffer containing 20 mM imidazole, 50 mL wash containing 40 mM imidazole and finally 25 mL wash buffer containing 80 mM imidazole. Protein was eluted with 14 mL wash buffer containing 200 mM imidazole. All steps in the protein purification were performed at 4 °C.

2.3 Results

In this section, we demonstrate that two intrinsically disordered proteins generate aggregating proteolytic products that can be removed in minutes by filter flow-through purification. In this rapid technique, full-length protein passes through the filter and aggregated proteolytic products create particles larger than the pores and are thus retained by the filter. Our approach has been demonstrated with two distinct proteins: the putative human lipid-binding protein Sec14-like 1 (Sec14L1), and the *Drosophila* Hox transcription factor, Ultrabithorax (Ubx). Both of these proteins contain intrinsically disordered regions.

The first protein tested was the 80 kDa human protein, Sec14L1, in which approximately 32% of the sequence is intrinsically disordered (Figure 2.1A). To examine if the filter flow-through purification assay would improve purity, we filtered 500 µL of purified Sec14L1 protein through an Amicon concentrator with 100 nm pores (EMD-Millipore). The supernatant was added to the filter column prior to centrifuging at

10,000 x g for 10 minutes at 4 °C to remove aggregated material. Soluble full length Sec14L1 and a soluble variant (either a long proteolysis product or phosphorylated protein) passed through the filter. However, aggregated proteolytic products were retained by the filter. Band density for SDS PAGE gels and western blot were quantitated using Image Studio Lite version 5.2 software for Sec14L1 and subsequently for Ubx. Quantification of each fraction by western blot demonstrated that soluble Sec14L1 reached approximately 98% purity with no visible contamination from aggregating proteolytic products (Figure 2.1B).

To test whether this assay was applicable to other types of proteins or specific to Sec14L1, we examined the 40 kDa *Drosophila* Hox transcription factor, Ultrabithorax (Ubx). Ubx is 53% intrinsically disordered, and contains a predominately structured DNA-binding homeodomain (HD) that binds to the sequence 5'-TAAT-3' (Figure 2.1A) (Liu et al. 2008, Liu et al. 2009). Most of the Ubx sequence is intrinsically disordered, and these regions are extremely susceptible to proteolysis (Liu et al. 2008). Since the homeodomain binds the target DNA with higher affinity than the full length Ubx protein, proteolytic contaminants containing this structured domain can out-compete full-length Ubx in DNA binding assays, preventing accurate affinity measurements (Liu et al. 2008). Gel electrophoresis and Western blot analysis showed Ubx to be less than 34% pure, with the major contaminants being proteolytic products of Ubx (Figure 2.1C,D). Because Ubx is prone to aggregation (Bondos and Bicknell 2003**)** and will adsorb non-specifically to many chromatography resins, removing aggregates by further chromatography was not feasible.

To determine whether filter flow-through purification can remove these proteolytic products, we filtered 500 μL of Ubx protein through an Amicon concentrator with a 100 kDa cutoff filter (EMD-Millipore). The filter was centrifuged at 10,000 x g for 10 minutes at 4 °C, and samples were collected for subsequent analysis. The SDS-PAGE results were quantified as mentioned previously to determine purity (Figure 2.1C,D). The full-length protein increased from < 35% to > 90% pure. The filter-purified protein had little detectable contamination when visualized by coomasie stain or the more sensitive western blot analysis (Figure 2.1C,D). However, any trace proteolytic products can be detrimental in DNA binding assays.

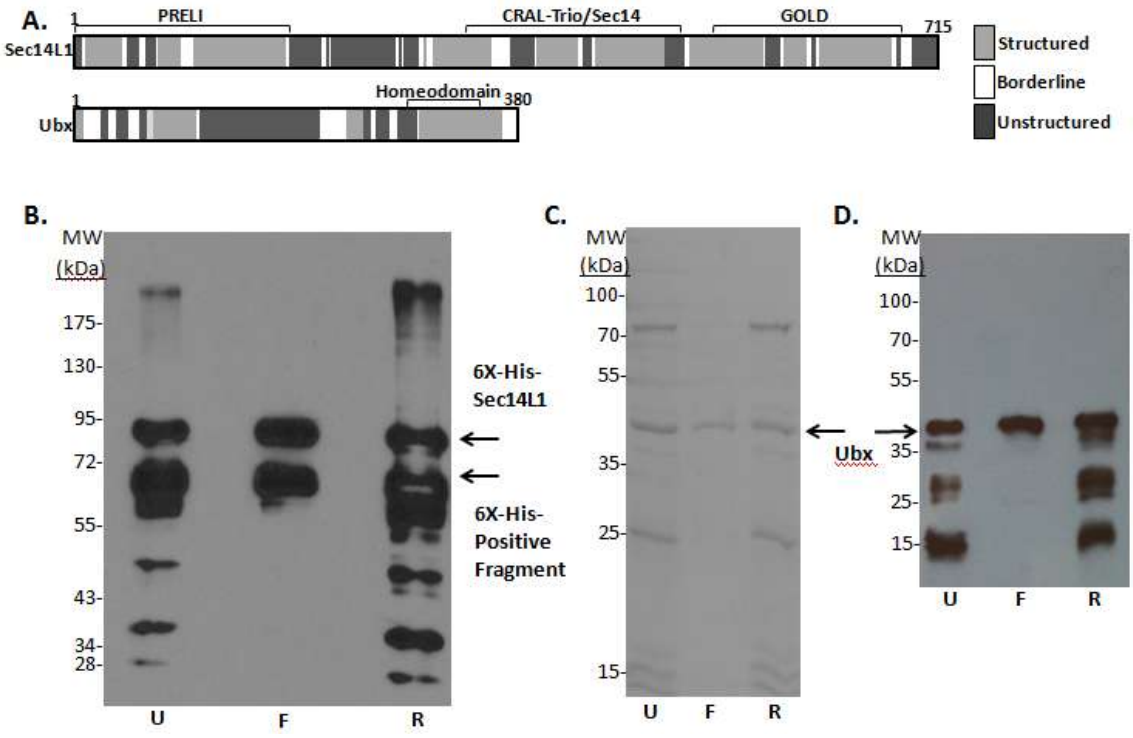

Figure 2.1 Filter flow-through purification of the Sec14L1 and Ubx proteins. A) Protein schematics depicting the structured (medium grey, score ≤ 0.4) and disordered (dark grey, score ≥ 0.6) regions of Sec14L1 (top) and Ubx (bottom). White boxes indicate ambiguous regions, in which the score lies between 0.4 and 0.6. Ordered and disordered regions for Sec14L1 were determined from the disorder prediction algorithm PONDR, and scores for Ubx were the average of the IuPred, DisEMBL and PONDR programs as well as native state proteolysis experiments **[11]**. B) Western blot analysis with Mouse anti-6X His (GeneTex) of a filter flow-through purification of Sec14L1, comparing unfiltered protein (U), the filtrate (F), and retentate (R). Band quantification revealed that the dual Sec14L1 bands were 98% pure, and that 97% of the total Sec14L1 protein was recovered in the filtrate. C,D) SDS-PAGE analysis of filter flow-through purification of Ubx, visualized by coomasie stain (C) and western blot using the anti-Ubx primary antibody FP3.38 (D). Based on band quantification, Ubx purified by filtrate was >90% pure, and 96% of the total full-length Ubx protein was recovered in the filtrate. In both cases, the filter captures aggregated protein fragments, and filtration substantially increases the purity of the protein. *Reprinted with permission from Analytical Biochemistry (Churion et al. 2016).

We used DNA binding electrophoretic mobility shift assays, performed as previously described (Liu et al. 2008, Liu et al. 2009) as a more sensitive method to detect Ubx proteolytic products. When assessing filter purified Ubx, only 2 major bands were detected in the gel, corresponding to free DNA and DNA bound to full length protein Ubx (Figure 2.2B). In contrast, the gel with unfiltered Ubx displayed additional bands, located in between these extremes, which correspond to the proteolytic products of Ubx bound to DNA (Figure 2.2A). The dominant band in this region corresponds to the Ubx HD binding to DNA. Although the HD has a higher affinity for DNA than Ubx, it is present at lower concentrations, and thus the HD-DNA band first appears in the same sample as the Ubx-DNA band. We conclude that filter purification yields full length Ubx protein with no detectable proteolytic products.

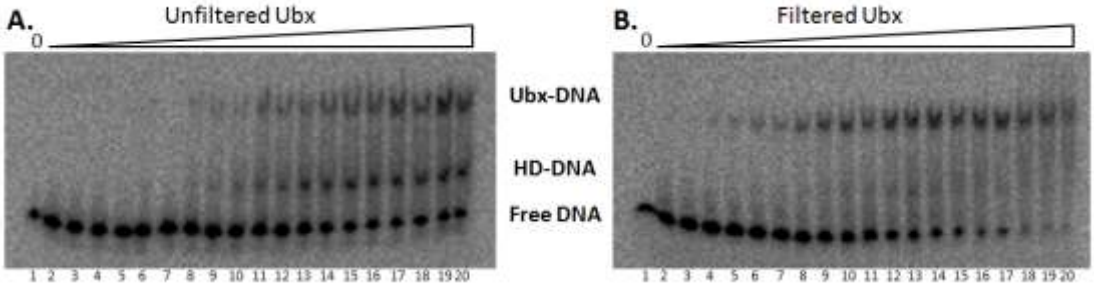

Figure 2.2 Comparison of DNA binding by unfiltered and filtered UbxIa. Both panels show equilibrium DNA gel retardation (gel shift) experiments. All lanes contain radiolabeled 2 pM 40AB DNA **[11]**, and lanes 2–20 also include Ubx, increasing from 2.5 pM to 10 nM. A) Unfiltered full-length UbxIa binding the 40AB DNA. B) Filtered Ubx binding 40AB. Filtration largely removes binding by contaminating proteolytic products, most notably the Ubx homeodomain (HD). *Reprinted with permission from Analytical Biochemistry (Churion et al. 2016).

2.4 Discussion

Purification is required to quantitatively assess the function of specific proteins. However, many proteins, especially those harboring intrinsically disordered regions, are prone to proteolysis during expression, purification, and storage. In this paper we demonstrate that filter flow-through purification can be a fast, efficient and inexpensive method for separating proteolytic product contaminants from full length purified protein. For filter purification to work, the proteolytic products must aggregate, causing them to be retained in a filter with pores sufficiently large for the full length protein to pass to the filtrate. We tested this approach on two proteins with significant intrinsically disordered regions, and found both proteins generated aggregating proteolytic products. Although Sec14L1 and Ubx have distinct properties and features, filter flow-through purification successfully removed proteolytic product contaminants from both preparations. Importantly, different procedures were used to express and initially purify these proteins and thus the success of filter purification is not dependent on the protocols used to produce either protein. This fast and efficient method can purify full-length proteins to near homogeneity, without compromising protein activity.

3. FIRST FULL-LENGTH HOX PROTEIN MODEL BY UNDERSTANDING UBX INTRAMOLECULAR INTERACTIONS BY PROTEIN-DNA BINDING ANALYSIS

3.1 Introduction

The development of all bilaterally symmetric animals requires Hox proteins to reliably define regional identities along the anterior-posterior (A-P) axis. This interaction determines the fate of a variety of tissues, including appendages and organs, as well as regionalizes systems that span the length of the developing organism, such as the musculature, ectoderm, nervous system, and digestive tract (Hughes and Kaufman 2002, Pearson et al. 2005, Hueber and Lohmann 2008). Individual Hox proteins control transcription of unique batteries of downstream genes. Each Hox protein must accurately recognize and regulate its unique battery of DNA binding sites and target genes (McGinnis and Krumlauf 1992, Zandvakili and Gebelein 2016). The sequence, structure, function, DNA binding targets, and the DNA binding affinities of Hox homeodomains appear indistinguishable, a sharp contrast to the necessity for differential Hox functions *in vivo* (Hoey and Levine 1988).

For example, *in vivo* Ultrabithorax and Antennapedia regulate different genes and drive development of unique body structures; *Drosophila* midthoracic legs and wings are formed within the Antennapedia expression domain, while development of halteres and the posterior-most pair of thoracic legs from analogous tissues requires Ultrabithorax. In order to enable these different Hox protein functions, the regulation of

different genes in a specific manner is required. However, the Hox proteins Ultrabithorax and Antennapedia homeodomains are 90% identical, which include all DNA-contacting residues (as shown in Section 1). Consequently, the affinity and specificity for DNA are also very similar (Ekker et al. 1994, Beachy et al. 1993, Passner et al. 1999). This disparity between the absolute requirement for distinct Hox activities *in vivo* and the similarity of homeodomain-DNA recognition *in vitro* has been termed the "Hox paradox" (Mann and Chan 1996, Mann 1995).

Regions outside the homeodomain influence DNA binding and have a large impact on DNA site selection, either by recruiting other transcription factors and binding DNA as a complex (LaRonde-LeBlanc and Wolberger 2003, Mann and Carroll 2002, Bondos et al. 2004, Bondos et al. 2006), or by directly impacting the DNA binding affinity and specificity of the homeodomain (Liu 2008, Liu 2009). The *Drosophila melanogaster* Hox protein Ultrabithorax (Ubx) is purified as a soluble, active full-length protein (Liu 2008, Liu 2009, Bondos and Bicknell 2003). This information allows us investigate the intramolecular mechanism of the Ubx Hox protein including the effect of nonhomeodomain regions. In addition, more natural DNA binding sites have been identified for Ubx than for any other Hox protein in any species (Pearson et.al. 2005). Therefore information from the studies presented in this section on binding a single DNA site can one day be extended to determine whether these mechanisms also alter DNA sequence selection by Hox proteins. These studies used UbxIa, which has the highest expression level among the *ubx* isoforms found *in vivo* (Busturia et al. 1990, O'Connor et al. 1988; Kornfeld et al. 1989).

The studies by Liu et al. demonstrate that most of the protein is able to regulate homeodomain-DNA interactions (Liu et al. 2008, Liu et al. 2009). These non-homeodomain sequences vary significantly between Hox proteins and thus may diversify DNA binding by Hox proteins. However, no full-length Hox protein structures are available, and no long-range intramolecular interactions have been identified in Hox proteins that could suggest a mechanism for regulating DNA binding. Directly probing the thermodynamics of intramolecular protein interactions in Hox proteins has been hindered by the high Ubx-DNA affinity, preventing analysis via isothermal titration calorimetry (Liu et al. 2008, Jelesarov and Bosshard 1999) and the prevalence of intrinsic disorder outside the homeodomain (Liu et al. 2008) which precludes structural studies. Furthermore, Ubx has few recognizable structural features outside of its DNA binding domain that could guide site specific mutant design to direct point mutagenesis. In addition, the presence of unstructured regions is refractory to mutagenesis which renders site-specific methods difficult (Romero et al. 2001).

To identify long-range interactions at amino acid resolution in the *Drosophila* Hox protein Ultrabithorax, conservation of the disordered regions in Ubx was analyzed via sequence alignment (Liu et al. 2008). Since disorder is important for Ubx function, this feature, although not the amino acid sequence, is conserved across multiple species. A small disordered region predicted near the N terminus is conserved in insects and crustaceans and a very large disordered region near the middle of the protein is additionally predicted to be conserved in arthropods (Figure 3.1). Consistent with this hypothesis, disordered regions are present in the other seven *Drosophila* Hox proteins

(Figure 1.5). Mutations in one of these conserved motifs (the hexapeptide motif FYPWMA) alters DNA binding affinity and specificity indicating this region is important for regulation upon DNA binding by the homeodomain (Liu et al. 2009).

Based on the locations of these conserved motifs and the identification of intermolecular Ubx interactions that stabilize Ubx materials, we have generated a structural model of full-length Ubx, in which non-homeodomain tyrosine residues interact to form aromatic interactions with the homeodomain tyrosine residues creating a large aromatic cluster. In this model, the resulting tyrosine cluster blocks access to the DNA binding helix. DNA binding assays under oxidizing conditions, fluorescence spectroscopy, osmotic stress, native state proteolysis, and heat capacity measurements all support this model, in which these long-range interactions must be disrupted to enable DNA binding. Mutagenesis of homeodomain tyrosines removes these regulatory interactions, and causes the full-length, mutant Ubx to bind DNA with properties similar to DNA binding by the isolated Ubx homeodomain (UbxHD). These long-range interactions have the potential to be disrupted in a tissue-specific manner by protein interactions, alternative splicing, and/or phosphorylation which will be further discussed in the Appendix. Thus these interactions have the potential play an important role in differentiation and regulation of Hox proteins *in vivo*.

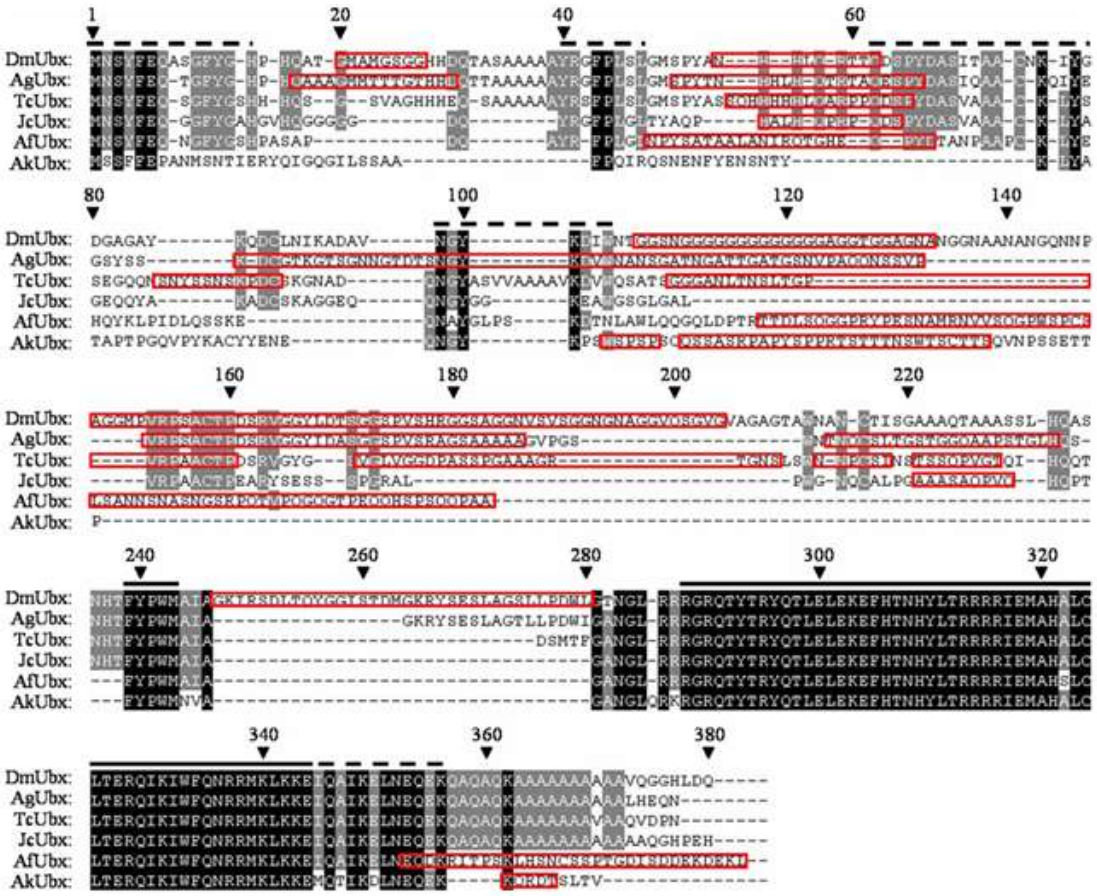

Figure 3.1. Conservation of the disordered regions in Ubx. An amino acid sequence alignment is shown for Ubx derived from the fruit fly *D. melanogaster*. (DmUbx), the mosquito *Anopheles gambiae* (AgUbx), the beetle *Tribolium castaneum* (TcUbx), the butterfly *Juonia coenia* (JcUbx), the shrimp *Artemia franciscana* (AfUbx), and the Onychophoran velvet worm *Akanthokara kaputensis* (AkUbx). The consensus results from the disorder prediction algorithms PONDR, IUPred, and DisEMBL are marked for each Ubx orthologue by red boxes. Identical and conserved residues are marked by black and gray shading, respectively, and previously identified conserved motifs are identified by dashed lines. The YPWM motif and homeodomain are marked by solid lines. Reprinted with permission from Liu et al. 2008. "This research was originally published in Journal of Biological Chemistry. Liu Y, Matthews KS, Bondos SE. Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the *Drosophila* hox protein ultrabithorax. *Journal of Biological Chemistry.* 2008; 30:20874-87. © the American Society for Biochemistry and Molecular Biology."

3.2 Materials and methods

3.2.1 Ubx protein expression

His-tagged Ultrabithorax spliceoform Ia was expressed from the pET-19b in BL21 (DE3) pLysS and purified as described by Liu et al.2008 (Liu 2008) with the following modifications. Cells expressing UbxIa were harvested 150 min after induction. Cell pellets corresponding to 2 L of fermented *E. coli* (final OD_{600nm} ~1.1) were resuspended in 2 mL of collection buffer (100 mM NaCl, 20 mM NaH_2PO_4, 80 mM Na_2HPO_4, pH=7.5 with one tablet of Complete Proteinase Inhibitor Mixture EDTA-free (Roche Applied Science) and frozen at -20 °C.

3.2.2 Ubx protein purification

A cell pellet corresponding to 2 L of fermented *E. coli* was thawed at room temperature and lysed in 20 mL of buffer containing 800 mM NaCl, 10 mM DTT, two additional tablets of Complete Proteinase Inhibitor Mixture EDTA free (Roche Applied Science), 5% glucose, 5 mM imidazole, and 50 mM NaH_2PO_4 at pH 8.0, 1 mM phenylmethylsulfonyl fluoride (PMSF, diluted from a 0.5 M PMSF stock dissolved in ethanol). Released DNA in the cell lysate was digested with DNase I (20 mg/mL, 40 µL) prior to centrifugation at 17,000 rpm for 30 min. The supernatant was mixed with polyethyleneimine (200 µL of 50% (w/v)) and recentrifuged at 17,000 rpm for 20 mins. The supernatant pH was then adjusted to pH=8.0 and recentrifuged at 17,000 rpm for 20 mins. To prepare the affinity column, 4 mL of Ni+2 nitrilotriacetic acid-agarose resin

(Thermo Scientific) was pre-equilibrated with wash buffer (500 mM NaCl, 10 mM-DTT, 5% glucose, 20 mM imidazole, and 50 mM NaH_2PO_4 at pH 8.0). The clear lysate was incubated with gentle agitation at 4 °C for 20 min before being poured into the column. Packed resin was washed with 50 mL of wash buffer containing 0 mM imidazole followed by 100 mL wash of buffer containing 10 mM imidazole, 50 mL wash buffer containing 20 mM imidazole, 50 mL wash containing 40 mM imidazole and finally 25 mL wash buffer containing 80 mM imidazole. Protein was eluted with 14 mL wash buffer containing 200 mM imidazole.

To remove aggregated proteolytic products from the final protein product, 500 μL of Ubx protein was filtered through an Amicon concentrator with a 100 kDa cutoff filter (EMD-Millipore). The filter was centrifuged at 10,000 x g for 10 minutes at 4 °C and samples were collected for subsequent analysis. The SDS-PAGE results were quantified as described previously in Section 2. The full-length protein increased from less than 35% to greater than 90% pure, and was used in DNA binding assays for subsequent analysis.

3.2.3 Homeodomain purification

The UbxHD, plus the 12 amino acids C-terminal to the homeodomain was expressed in *E. coli* using the pET3c expression vector and purified as described previously (Liu et al. 2008).

3.2.4 DNA labeling

Annealed double stranded DNAs were radiolabeled with [γ- ^{32}P]-ATP (Perkin Elmer- Product description: ATP,[g-32P]-Catalog # BLU002A500UC, concentration 10 mCi/mL) using polynucleotide kinase (New England Biolabs). The DNAs were radiolabeled for 90 minutes in a reaction containing 1 μL 0.1 M DTT, 1 μL annealed DNA (40AB 3 x 10-7 M), 1.5 μl 10X PNK buffer , 10.5 μl γ32P ATP, 2 μl T4-PNK enzyme for a total volume of 16 μL at 37 °C. The reactions were terminated by adding 1 μl of 0.5 M EDTA at pH 8.0. Unincorporated [γ- ^{32}P]-ATP was removed with a Nick column (pre-packed with Sephadex G-50). The Nick column was pre-equilibrated in TE (10mM Tris, 1mM Ethylenediaminetetraacetic acid or EDTA) at pH 7.5. The column was washed with 400 μL TE buffer and eluted with an additional 400 μL TE buffer. The radioactivity of DNAs was measured with a scintillation counter, in which 1-3 μl of ^{32}P-labeled DNA was added to 5 mL of scintillation cocktail (Research Products International Corp.). Typical results were around 15,000 cpm/μL. Aliquots of labeled DNA were stored at -20 °C.

3.2.5 Electrophoretic mobility shift analysis (gel shift) assays

Stoichiometric and equilibrium DNA binding assays were performed as previously described (Liu et al. 2008) with the following modifications in the binding buffer: DNA binding buffer contained 1 mM DTT, 100 μg/mL bovine serum albumin (BSA), and 10% glycerol, and 20 mM Tris-HCl at pH 7.5, and 0.1M KCl. Retardation gels contained 10%, 19:1 polyacrylamide for UbxHD, or 4%, 37.5:1 polyacrylamide for

UbxIa wild type full length and UbxIa full length mutants, 0.5 M TBE (0.045 M Tris borate, pH 8.0, 0.001 M EDTA), and 3% glycerol.

Gel shift assays under osmotic conditions were performed as previously described (Li and Matthews 1997) with modifications mentioned above. For heat capacity measurements, the binding reactions containing 20 μL of buffer/protein and 10 μL of DNA were incubated for 20 to 25 min at 4°C, 22 ˚C, and 37°C prior to running the gel at room temperature. Binding affinity experiments run entirely at 4°C (in a refrigerated cold box) yielded the same results as experiments in which only the binding reaction is incubated at 4°C, confirming that the reaction does not re-equilibrate during gel loading.

Activity gel retardation experiments measured the fraction of protein capable of binding DNA. Activity assays were performed under stoichiometric conditions ([DNA] >> K_d). In typical activity experiments, the concentrations of freshly thawed proteins were at least one order of magnitude above the K_d. The unlabeled oligonucleotide concentration exceeded K_d by at least 10-fold. A small fraction of radio-labeled DNA was added to the reaction to yield radiation signals. Binding reactions containing 20 μL of buffer/protein and 10 μL of DNA (mixture of "cold" DNA and radio-labeled DNA) were incubated for 20 to 25 min at 4°C, 22 ˚C, and 37°C.

Retardation gels contained were pre-electrophoresed at 110 V with buffer recirculation for 0.5 hours. A 15 µl aliquot of each sample was loaded onto the gels while running at 300 V. After approximately 5-8 minutes, the voltage was reduced to

120 V for a further 1.5 hours. The gels were blotted on filter paper, dried on a vacuum slab gel dryer, and exposed to a FUJI phosphorimaging plate for 12 to 16 hours.

To analyze the DNA binding data, image plates were scanned on a Fuji Imaging Analyzer BAS1000 (Fuji Photo Film Co., Ltd, Japan). The relative amount of radioactivity within the "bound" and "free" species on the gel for each condition was determined by digitizing the gel images using MacBAS 2.0, ImageGauge 4.0, and MultiGauge 2.3 software. Analysis was done as described in Churion 2014 (Churion et al. 2014).

Gel shift assays under oxidizing conditions were performed as previously mentioned with the following modifications: the DNA binding buffer for oxidizing conditions contained 5 mM oxidized glutathione (GSSG), 100 μg/mL bovine serum albumin (BSA), 10% glycerol, and 20 mM Tris-HCl at pH 7.5, and 0.1M KCl.

3.2.6 Fluorescence spectroscopy

Fluorescence measurements were conducted with a PTI fluorometer, with temperature regulation through a jacketed cuvette holder (Neslab). The protein was diluted to 250 nM in dialysis buffer containing 5% glycerol, 10 mM β-mercaptoethanol, 150 mM NaCl, 50 mM Tris and pH = 8.0. The peptides were excited at 283 nm and emission spectra were taken from 295-450 nm. A buffer background was subtracted from all spectra.

3.2.7 Proteolysis experiments

Native state proteolysis experiments were performed as previously described in Liu et al. 2008.

3.2.8 Molecular dynamics simulations

The molecular dynamics simulation details are as follows: the initial coordinates of the missing loop region (pdb id: 1B8I) is obtained from homology-based modeling and then further refined by satisfying spatial restraints. Then, the system containing Ubx/Exd/DNA complex in explicit TIP3P water was simulated with atomistic molecular dynamics (MD) coupled with a simulated annealing protocol to further refine the structure of the missing region. First, the system was heated to 800° K and the temperature was lowered gradually to 300° K. During this simulated annealing-MD, the backbone coordinates of Ubx-HD and Exd cofactor, as well as the DNA chains, were constrained. Finally, the whole system was allowed to relax at 300° K by removing all constraints. The structure of Ubx protein at the end of relaxation was immersed in a cubic box containing aqueous NaCl solution for further study. All molecular dynamics simulations were performed using namd2 software in IBM BlueGene supercomputer with charmm32 forcefield. The production MD was performed at 300° K temperature and 1 atm (NPT ensemble). Electrostatics were calculated with Particle Mesh Ewald (PME) method. All production MD simulations were >50 ns long. Simulations were performed for both fragments of wild-type ubx protein (as in Passner et al. 1999) and its GPGG mutant in 0 mM, 100 mM, 300 mM, and 500 mM aqueous NaCl. At each salt

concentration, three different MD simulations were run starting from different initial conformation.

3.3 Results

3.3.1 Regions outside the homeodomain impact DNA binding affinity

Previous DNA binding affinity studies demonstrated that large regions outside the UbxHD can modulate the binding affinity of Ubx (Liu et al. 2008) yet did not identify the specific amino acids are involved. To verify DNA binding measurements work properly, we measured UbxHD and full-length UbxIa DNA binding affinity to the optimal UbxHD-DNA binding sequence TAAT within a 40-bp oligonucleotide termed 40AB (Table 3.1). Consistent with previous measurements by our laboratory and others (Li et al. 1996, Li and Matthews 1997, Liu et al. 2008; Liu et al. 2009), the isolated UbxHD binds to 40AB with an affinity of 60 ± 24 pM; in contrast, full-length UbxIa has a binding affinity of 160 ± 33 pM, approximately 2.5 fold weaker than that of UbxHD (Table 3.2 and Figure 3.5). This difference in the DNA binding affinities of full-length UbxIa and UbxHD confirms that sequences outside the Ubx homeodomain modulate DNA binding by this region.

DNA Name	Structure	# Binding Sites	Length	Sequence 5'- 3'
40AB	linear	1	40 bp	CCGGGCTGCACATGGTTAATGGCCAGTCCA CGCGTAGATC

Table 3.1 Characteristics of DNA used in Ubx DNA binding experiments

3.3.2 Conserved motifs overlap with anchor positive regions

In Ubx orthologues, the presence and location of intrinsically disordered regions is conserved. However the amino acid sequence of these regions is not conserved. Within these intrinsically disordered regions, there are short motifs that can mediate protein-protein interactions. The amino acid sequence of these short motifs is more hydrophobic than the rest of the intrinsically disordered regions and thus these motifs can be located by algorithms such as ANCHOR. In Ubx, the amino acid sequence and the placement of these motifs are both conserved (Figure 3.2). This general arrangement (homeodomain, intrinsically disordered regions interspersed with conserved motifs) is common to all Hox proteins. Mutations in one of these conserved motifs for example the highly conserved hexapeptide motif (FYPWMA) alters DNA binding affinity and specificity.

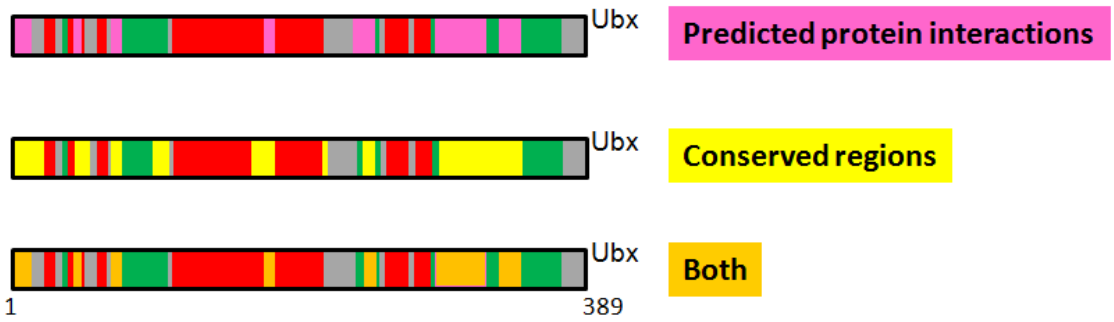

FIGURE 3.2 Conserved motifs overlap with ANCHOR positive regions. ANCHOR algorithm predicts the location of molecular recognition features, motifs capable of engaging in protein interactions. For Ubx ANCHOR algorithm identifies the N-terminus as a region likely to engage in protein interactions. In Ubx, the amino acid sequence and the placement of these motifs are both conserved. The Ubx amino acid sequence evolutionarily conserved sequences in yellow boxes, sequences predicted to be involved in protein interactions via the ANCHOR algorithm are in pink boxes, and sequences both conserved and interactive in orange boxes. In green boxes the structured regions are indicated, in red are intrinsically disordered regions.

3.3.3 A Digression: Ubx self-assembles into materials *in vitro*

We hypothesized that these conserved, ANCHOR positive motifs bind the homeodomain, accounting for the long-range regulatory interactions observed by Liu et al. (Liu et al. 2008, Liu et al. 2009). Therefore mutating these motifs should modulate DNA binding affinity. However, some of these motifs are reasonably long. Which amino acid should be mutated to yield the maximal effect? A critical clue was provided by materials formed by the Ubx protein.

The mechanical strength and blue fluorescence of Ubx materials is due to the oxidation of tyrosines to create dityrosine bonds between Ubx monomers (Howell et al. 2015). Bonds form between specific amino acids, including between the N-terminus and

the homeodomain (where 2 mutually exclusive bonds form Y4-Y293 and Y12-Y296), and between Y167 and Y240. These tyrosine residues are all located in conserved motifs in Ubx and predicted to be involved in protein-protein interactions. Mutations that abolish one or both dityrosine bonds reduce the fiber strength; therefore, molecular these specific interactions are responsible for the strength of Ubx materials.

3.3.4 Ubx interactions in materials are similar to interactions in Ubx monomers

We hypothesized that *inter*molecular interaction that occur between Ubx monomers to stabilize Ubx materials might reflect inherent *intra*molecular interactions that occur in Ubx monomers to regulate DNA binding function. Evidence to support this hypothesis came from multiple studies ranging from studies utilizing different N-terminal truncation mutants with Ubx monomers to Ubx protein biomaterials. Each terminal truncation mutant was previously examined for the ability to bind DNA and for the ability to form Ubx fibers. Ubx does not form materials as part of its natural function. Therefore, Ubx intermolecular interactions may be non-native aggregates or amyloid. However, in contrast to amorphous protein aggregates, which often appear as white flocculates, Ubx materials are transparent and can diffract light (Majithia et al. 2011) suggesting a regular structure. Furthermore, X-ray diffraction and Thioflavin T binding studies of Ubx materials lack any indications of amyloid structure (Howell et al. 2015).

Indeed, the materials retain the ability to bind DNA in a sequence specific manner, (see Section 4, Howell et al. 2015) suggesting that the homeodomain structure

is intact as well. When bound to clusters of DNA binding sites, Ubx can oligomerize in multiple orientations: side-to-side cooperative interactions when binding to linear DNA, and back-to-back interactions between clusters of cooperatively bound Ubx proteins to form the stem of a DNA loop (Beachy et al. 1993). These interactions may mimic the side-to-side interactions that form nanoscale fibrils and the back-to-back interactions that allow fibrils to associate into films and fibers. Finally, an evaluation of Ubx truncation mutants revealed that regions required to form Ubx biomaterials (as measured by fiber length) (Greer et al. 2009) also improve DNA binding (Liu et al. 2008). The simplest explanation is that the intramolecular interactions that favor DNA binding occur in an intermolecular fashion in materials (Figure 3.3).

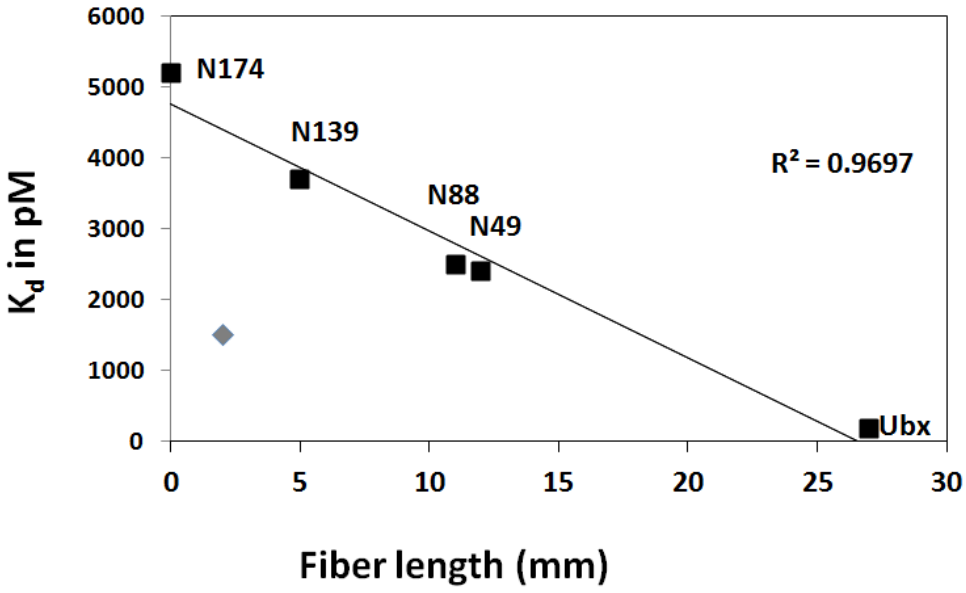

FIGURE 3.3 Deletion of conserved regions that modulate binding also affect biomaterial formation. In truncation mutants the ability to self-assemble is proportional to DNA binding affinity. As nonhomeodomain regions that modulate DNA binding are deleted, the regulatory intramolecular interactions that modulate DNA binding are affected. Since the inherent intramolecular interactions are reflected via intermolecular interactions to form Ubx fibers we see an effect in Ubx material formation.

Interestingly, the tyrosines that were specifically mapped to interact in protein biomaterials are also i) located in conserved regions, ii) in regions that modulate protein-DNA binding affinity, and iii) are also in motifs predicted to engage in protein-DNA interactions (Figure 3.4). Therefore this section will determine if the tyrosines are involved intramolecular interactions to regulate DNA binding.

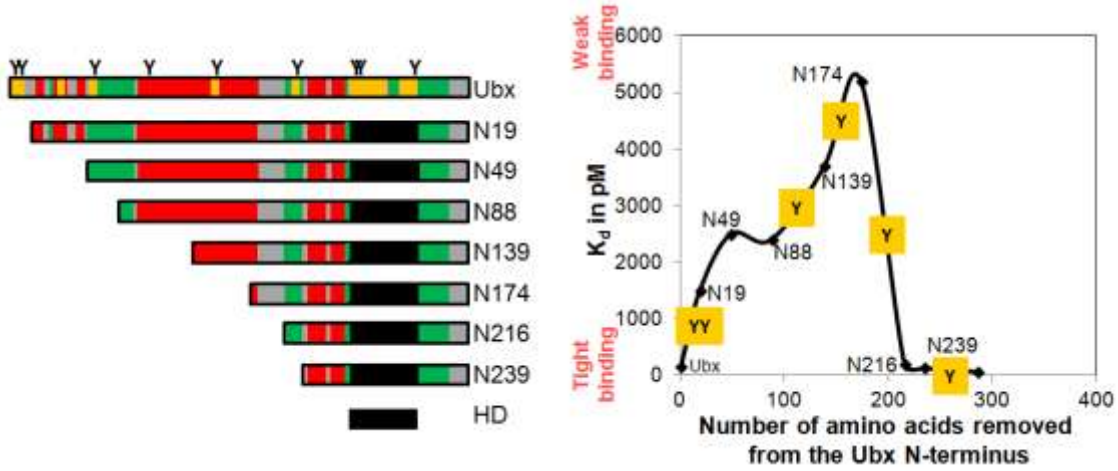

FIGURE 3.4 Tyrosines that form dityrosine bonds are embedded in evolutionarily conserved motifs predicted to be involved in protein interactions and modulate DNA binding affinity. The equilibrium dissociation constant (black diamonds), measured for Ubx, Ubx truncation mutants and Ubx homeodomain binding to the optimal DNA sequence (40AB) is only significantly altered when a conserved motif is removed. Analysis of DNA binding affinity assays indicate that the tyrosines that were specifically mapped to interact in protein biomaterials are also located in conserved regions that modulate protein-DNA binding affinity, and in motifs predicted to engage in protein-DNA interactions. Key tyrosines are in large, black bold text and tyrosine residues that modulate DNA binding affinity are in black bold text in yellow boxes, intrinsically disordered regions are indicated in red boxes, structured regions in green boxes, and the homeodomain in black box (Liu et al. 2008).

3.3.5 Tyrosine mutants significantly alter DNA binding

The same tyrosine residues predicted to be important for biomaterial formation were targeted for EMSA analysis to determine the effect on DNA binding affinity and regulation. Tyrosines were changed to residues retaining as much of the chemical nature of tyrosine as possible to prevent mutagenesis-induced structural rearrangements. Tyrosines in intrinsically disordered regions outside the homeodomain were mutated to serine, because the transfer coefficient of serine best mimics that of tyrosine as a free amino acid, leading to their similar values on the Kyte-Doolittle hydropathy scale. (Kyte and Doolittle 1982). Tyrosines within the homeodomain were mutated to leucine, because leucine most closely resembles the hydropathy of tyrosine on the surface of a structured protein (Pace et al. 2014). These mutations do not dramatically alter the function of the homeodomain, because fibers composed of homeodomain mutants can successfully bind DNA in a sequence-specific manner (Howell et al. 2015). The tyrosine point mutants tested in full length UbxIa for this analysis are listed in Table 3.2. To facilitate the purification of these mutants, a his-tag was engineered at the N-terminus of this protein. Prior studies demonstrated that the presence of the his-tag does not alter DNA binding affinity (Liu et al. 2008). EMSA analysis revealed these tyrosine mutations elicited a vast range of effects on DNA binding affinity under reducing conditions (Figure 3.5).

	Kd (pM) Reducing conditions	Kd (pM) Oxidizing conditions	Increased/ Decreased DNA binding	Possible Regulatory role
Ubx FL [2]WT	160 ± 33	0	WT binding	WT
UbxHD	63 ± 24	60 ± 24	Increased	N/A
Ubx FL Y4S+Y12S	1380 ± 70	2460 ± 300	Decreased	Enhancing
Ubx FL Y52S	648 ± 20	544 ± 19	Decreased	Enhancing
Ubx FL Y85S	172	290 ± 8	No effect	No effect
Ubx FL Y100S*	100*	100*	Increased*	Inhibitory*
Ubx FL Y167S	52 ± 9	56 ± 11	Increased	Inhibitory
Ubx FL Y240S	52 ± 9	52	Increased	Inhibitory
Ubx FL Y293L	49	N/A	Increased	Inhibitory
Ubx FL Y296L	42	N/A	Increased	Inhibitory
Ubx FL Y293L+Y296L	28 ± 17	35 ± 14	Increased	Inhibitory
Ubx FL Y310L	62 ± 1	56 ± 2	Increased	Inhibitory

FL = full length [2]WT = Wild-Type

***Ubx FL Y100S had aggregation issues during purification. Out of one DNA binding experiment , the K_d was similar to Ubx FL WT. The solubility problems are informative because it suggests that this amino acid center of cluster. Due to solubility and aggregation issues it suggests that it destabilizes the protein cluster dramatically and protein structure and hence the effects.**

Table 3.2 Ubx, UbxHD, and Ubx tyrosine mutants equilibrium dissociation constant (K_d) by measured via DNA binding affinity assays under reducing and oxidizing conditions. Tyrosine mutants significantly alter DNA binding and their significant involvement in forming the tyrosine cluster to regulate DNA binding.

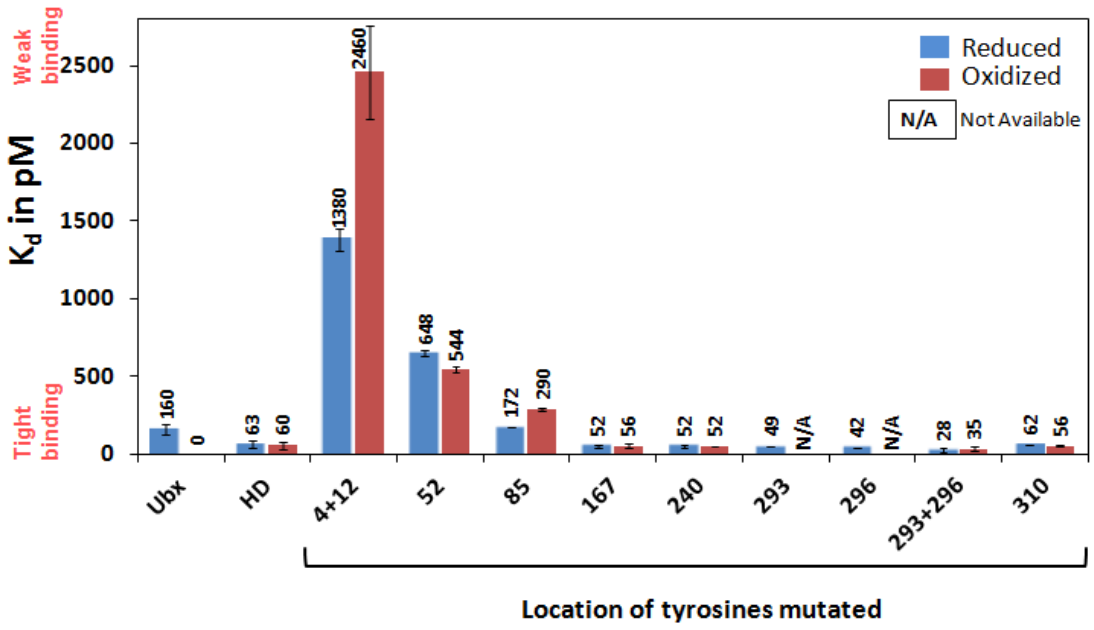

FIGURE 3.5 Tyrosine mutations significantly alter DNA binding. The equilibrium dissociation constant was determined via DNA binding affinity assays for Ubx, UbxHD, and Ubx tyrosine mutants under reducing (red bars) and oxidizing conditions (blue bars). Tyrosine mutations on the homeodomain (Y293L/Y296L or Y310L) or in the intrinsically disordered regions of the protein (all other tyrosine mutants) modulate DNA binding affinity. Oxidation prevents DNA binding only in wild-type Ubx, where the tyrosine cluster is locked in a closed conformation (see text).

All tyrosine residues predicted to be involved in biomaterial formation also modify DNA binding in Ubx monomers. As a control, tyrosine residues that do not fall within the requirements mentioned above are not predicted to be involved in DNA binding regulation and, therefore, no effect on DNA binding is expected. The most N-terminal mutations Y4S+Y12S, and Y52S decreased the affinity (increased K_d) of Ubx to optimal DNA sequence 40AB (Table 3.2 and Figure 3.5). These tyrosines enhance DNA binding affinity. Tyrosine mutations of conserved residues closer to the

homeodomain or within the homeodomain Y100S, Y167S, Y240S, Y293L, Y296L, Y293L+Y296L, and Y310L increase the affinity of Ubx to optimal DNA sequence 40AB. Therefore, in the wild-type protein these tyrosines inhibit DNA binding regulatory interactions. Two single mutants Y293L and Y296L and double tyrosine mutant Y293L+Y296L bind DNA 8 fold tighter than the wild-type, full length Ubx and 3 fold tighter than the UbxHD. This very high affinity is investigated and discussed in Section 5. In contrast, the tyrosine residue Y85, which is not in a conserved region and not involved in biomaterial formation, did not have an effect in DNA binding affinity. These data raised two questions: 1) how can 5 tyrosines located in non-homeodomain regions simultaneously interact with 3 tyrosines on the homeodomain surface to regulate DNA binding? and 2) why do similar mutations produce such a wide variety of effects?

3.3.6 Aromatic clusters in other proteins provide an important clue for interpreting Ubx-DNA binding data

Aromatic compounds are unsaturated cyclic and planar molecules that contain an aromatic ring. Above and below the aromatic ring, cyclic delocalization and resonance of these π electrons form a π electron cloud which harbors a partial negative charge (Anjana et al. 2012, Balaban et al. 2005). Out of the 20 amino acids found in protein structures only phenylalanine, tyrosine, tryptophan and histidine are aromatic (Brocchieri and Karlin 1991). The interactions that take place between the side-chains of the aromatic amino acid residues are referred to as aromatic-aromatic interactions and are relatively non-polar in nature. These interactions play an important role in

maintaining the overall structure of the protein molecules such as the sweet potato purple acid phosphatase, in which an extended aromatic network spans the length of the protein (Anjana et al. 2012, Lanzarotti et al. 2011). In addition, thermophiles have a higher propensity to form aromatic clusters, which enhance protein stability (Kannan and Vishveshwara 2000). These studies suggest the tyrosines in Ubx might form a cluster to simultaneously regulate intramolecular interactions and DNA binding. Disruption of any tyrosine from this network would affect these activities.

3.3.7 Where on the homeodomain do clusters bind?

In the homeodomain, Y293 minimally contacts the DNA phosphate backbone via a hydrogen bond. The other two tyrosine residues (Y296 and Y310) do not interact with DNA (Passner et al. 1999) (Figure 3.6). Multiple mechanisms exist by which a tyrosine cluster can perturb DNA-binding. The first mechanism by which the tyrosine cluster can allosterically impact DNA binding is by directly interacting with the DNA-binding helix (helix III) to sterically preclude homeodomain-DNA interactions. The second possible mechanism by which the tyrosine cluster might bind other regions of the homeodomain (which does not include directly interacting with helix III) providing allosteric mechanisms that impact DNA binding by the homeodomain (back side cluster formation) (Figure 3.7). To test for the difference between these two hypothesis molecular dynamics simulations and fluorescence spectroscopy experiments were performed and data obtained provided the basis for the structural model of Ubx.

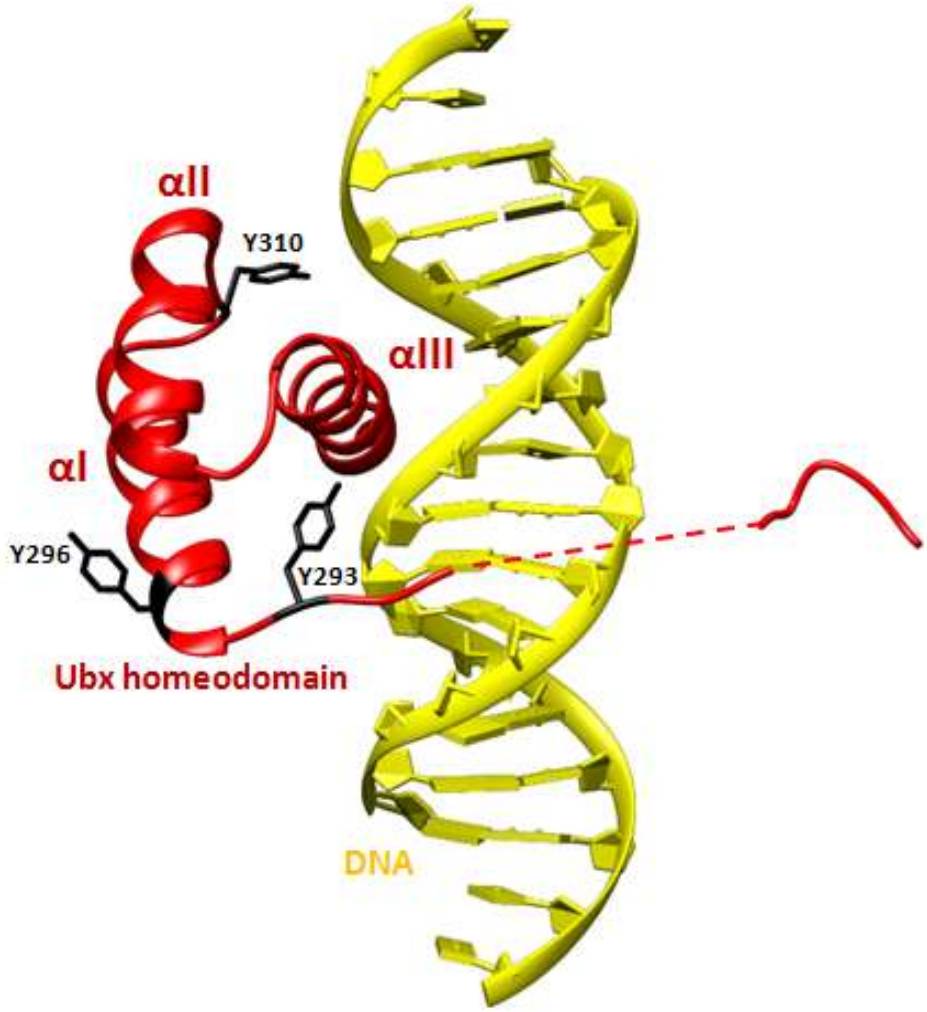

Figure 3.6. The *Drosophila* Ultrabithorax Homeodomain Tertiary Structure and conserved tyrosines residues. The three helices (in red helices marked αI, αII, αIII) making up the *Drosophila* Ultrabithorax homeodomain are prototypic for all homeobox proteins, serving as the core structure. In the homeodomain, Y293 minimally contacts the DNA phosphate backbone via a hydrogen bond. The other two tyrosine residues (Y296 and Y310) do not interact with DNA (tyrosine residue backbone is shown in black along with tyrosine number in bold black). Ubx structure and DNA obtained from pdb 1B8I.

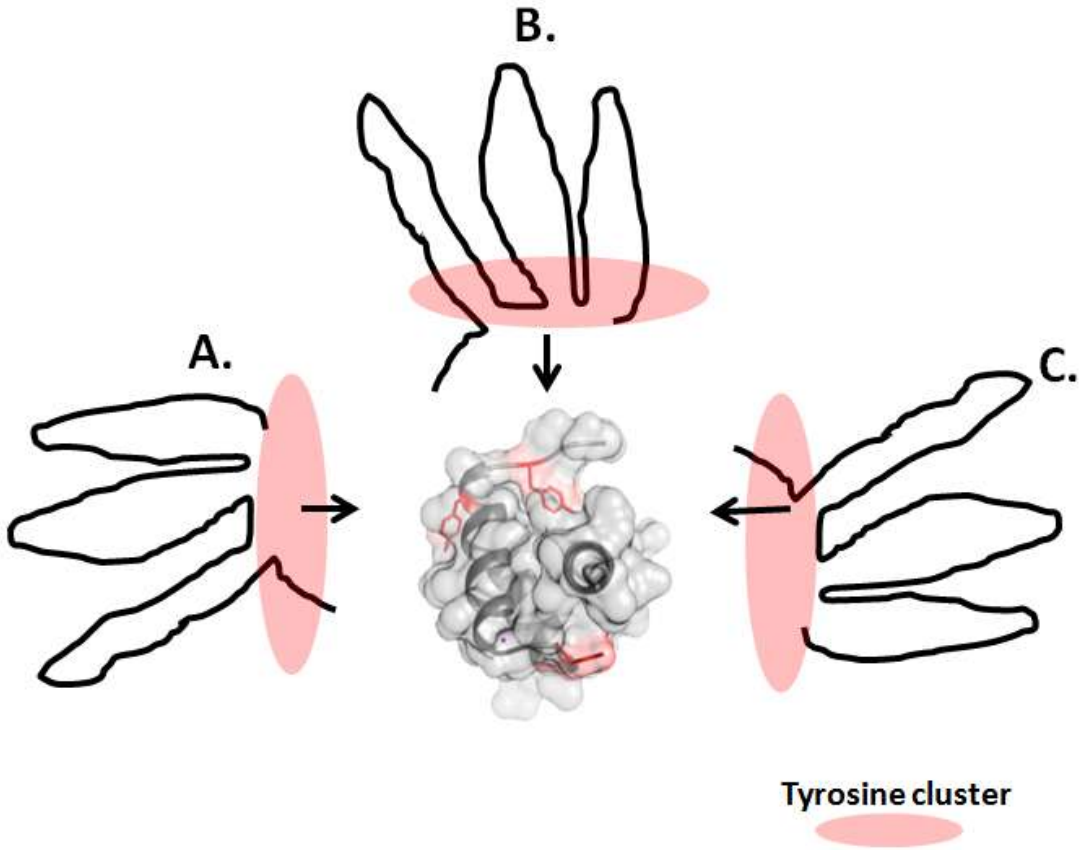

FIGURE 3.7 The tyrosines outside of the homeodomain interact with the conserved tyrosines on the homeodomain surface where a tyrosine cluster forms. All involved tyrosines must be present to simultaneously impact DNA binding. Possibilities where the tyrosine cluster could bind on the homeodomain include: **A and B)**The tyrosine cluster may bind other portions of the homeodomain to allosterically impact binding or **C)** The tyrosine cluster may bind the DNA-binding interface (outer surface of helix 3) to directly impact binding.

3.3.8. Molecular dynamics simulations suggest aromatics interact with DNA binding helix in homeodomain*

Molecular dynamics simulations can provide detail concerning individual atom motion as a function of time. The understanding at the atomic level of detail is important for a complicated reaction like protein folding. The simulations are done in explicit

solvent and starting from nearly extended configurations. Molecular dynamics is a detailed method to model large scale motions. Due to computational limitations, our molecular dynamics simulations use a truncated variant of Ubx which includes only the hexapeptide motif, a shortened linker mimicking the UbxIVa splice variant, and the homeodomain (Figure 3.8 A, B). At physiological conditions, the hexapeptide motif, which consists of amino acids FYPWMA, always binds to hydrophobic residues of the C-terminal end of helix 3 by first establishing hydrophobic interactions. This effect is independent of salt concentration, confirming the hydrophobic nature of the interaction. The initial hydrophobic interactions are followed long-lived cation-π interactions (Figure 3.8 C). The hexapeptide residues thus interact with the third helix of the homeodomain, essentially blocking the DNA binding site to prevent binding, suggesting other conserved motifs which also contain aromatic residues, may do the same.

*Molecular Dynamics simulations were done in collaboration with Dr. Payel Das from IBM Thomas J Watson Research Center, Computational Biology Center

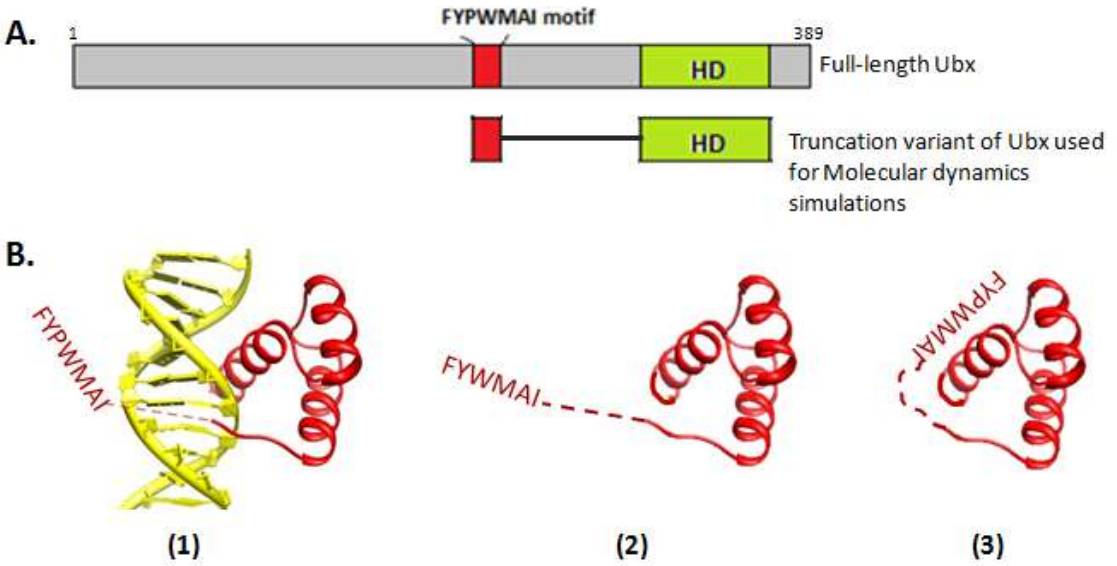

Figure 3.8 Molecular dynamics simulations suggest aromatics interact with DNA binding helix in homeodomain. A) a truncated variant of Ubx which includes only the hexapeptide motif, a shortened linker mimicking the UbxIVa splice variant, and the homeodomain. **B)** Molecular dynamics simulations were performed using namd2 software in IBM BlueGene supercomputer with charmm32 forcefield, briefly: (1) started with UbxHD-DNA crystal structure in water (Passner et al. 1999) (2) remove DNA in silico and (3) Heat and cool in silico. Detailed explanation of this experiment is described in materials and methods 3.2.8.

C.

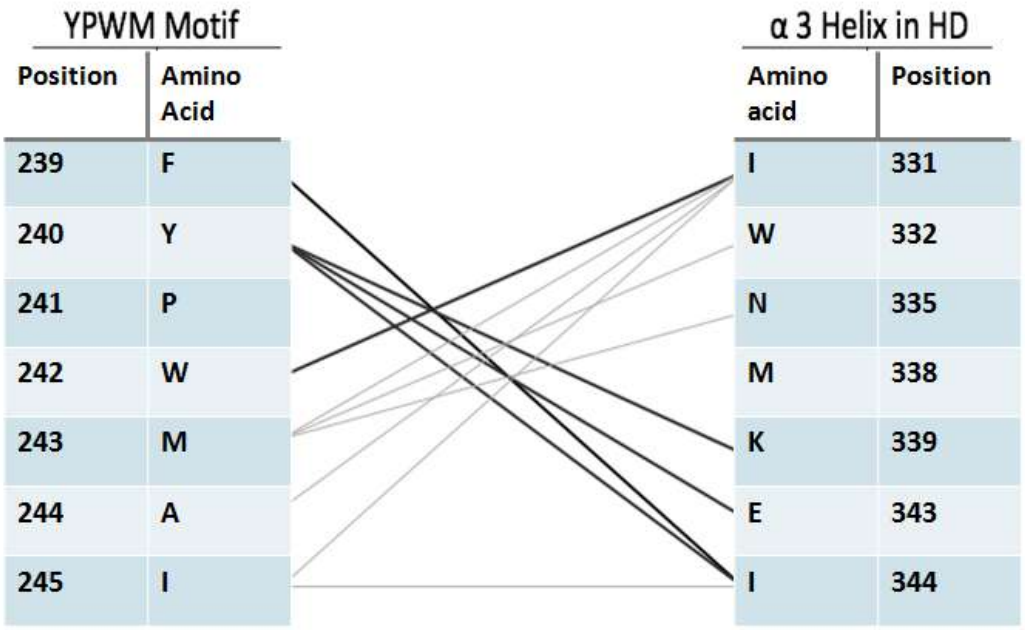

Figure 3.8 continued C) Contact map between FYPWMAI motif and C-terminal helix of the homeodomain indicate that the initial hydrophobic interactions are followed long-lived cation-π interactions suggesting that this motif binds helix 3 of the homeodomain, blocking its ability to bind the major groove of DNA (short lived contacts are indicated in gray lines and long lived contacts are indicated in black bold lines)

3.3.9 The first structural model of a Hox protein

Given the information and data obtained, we propose a model in which non-HD tyrosine residues interact with each other and tyrosines on the homeodomain surface to form an aromatic cluster that "blockades" the DNA binding helix (helix 3), preventing the homeodomain from accessing the DNA (the closed conformation). This structure is in equilibrium with a second structure, in which a different tyrosine cluster forms without the homeodomain tyrosines. In this structure, the "blockade" is removed, allowing the homeodomain to bind DNA (the open conformation) (Figure 3.9). These

interactions are specific, since different tyrosines differentially impact Ubx structure and thus DNA binding. Tyrosines not in conserved motifs nor involved in biomaterial formation produced no effect on DNA binding affinity, indicating that not all tyrosines contribute to cluster formation.

This model provides an explanation for the large range of observed effects generated by tyrosine mutations. Mutants that destabilize the open state more than the closed state shift the equilibrium towards the closed state thus decrease DNA binding. Conversely tyrosine mutants that destabilize the closed conformation shift the equilibrium towards the open state, thus favoring HD-DNA interaction (Figure 3.9).

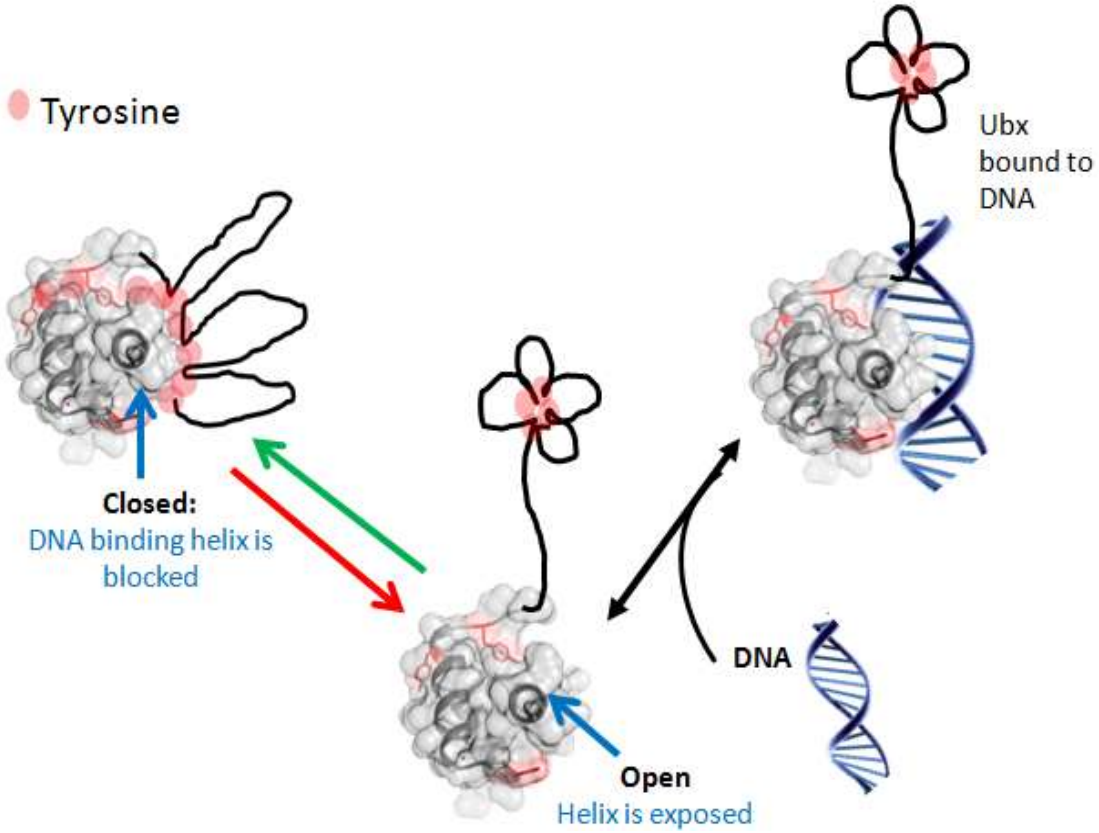

Figure 3.9 Model of Ubx structure. Free Ubx is in equilibrium between two conformations. In the "closed" state, a cluster of aromatic residues (red) connected by disordered loops (shortened for clarity) bind to the DNA binding helix of the homeodomain (pdb 1B8I) In the "open" state, the aromatic residues cluster separate from the homeodomain, freeing the DNA binding helix. Mutations that destabilize the open state relative to the closed state inhibit DNA binding (green arrow), whereas mutations that destabilize the closed state relative to the open state facilitate DNA binding (red arrow).

3.3.10 Tyrosine mutations in the HD abolished long-range interactions with the homeodomain*

The biochemical applications of fluorescence often utilize intrinsic protein fluorescence. There are three amino acids with intrinsic fluorescence properties,

phenylalanine, tyrosine and tryptophan but only tyrosine and tryptophan are used experimentally because their quantum yields (emitted photons/exited photons) is high enough to give a reliable fluorescence signal. Those residues can be used to monitor changes in protein structure (either by folding/unfolding events) because their fluorescence properties are sensitive to the environment. In the native folded state, tryptophan and tyrosine are generally located within the core of the protein, whereas in a partially folded or unfolded state, they become exposed to solvent. In a hydrophobic environment (buried within the core of the protein), tyrosine and tryptophan have a high quantum yield and therefore high fluorescence intensity (red shifted). In contrast in a hydrophilic environment (exposed to solvent) their quantum yield decreases leading to low fluorescence intensity (blue shifted).

The Ubx homeodomain has a predicted net charge of +11, and most of these charged residues are located on helix 3 (Howell et al. 2015). Therefore, if the aromatic cluster composed of tyrosine residues binds helix 3, as predicted by the molecular dynamics simulations, then the fluorescence emission spectrum of this cluster should be red-shifted (Vivian and Callis 2001). To test our model, the fluorescence emission spectrum of wild-type, full length Ubx was compared to that of the Ubx Y293L+Y296L mutant, which disrupts the formation of the aromatic structure in the closed (HD bound) state and thus favors the open configuration.

*Fluorescence spectroscopy experiments were done in collaboration with Dr. Hays Rye from the Department of Biochemistry and Biophysics - Texas A&M University, Lauren Kustigian from the Department of Biochemistry and Biophysics

If the cluster is indeed positioned away from the positively charged homeodomain, the emission spectrum is expected to blue-shift to a wavelength typical for an aromatic cluster (lower wavelengths). Indeed, compared to the wild-type full length Ubx λ_{max} =344 nm the fluorescent emission maximum of the full lenght Y293L+Y296L Ubx mutant is λ_{max} =339 nm which indicates a blue shift (Figure 3.10). Thus the aromatic cluster is not near the positively charged residues on the homeodomain in the conformation of Ubx best able to bind DNA.

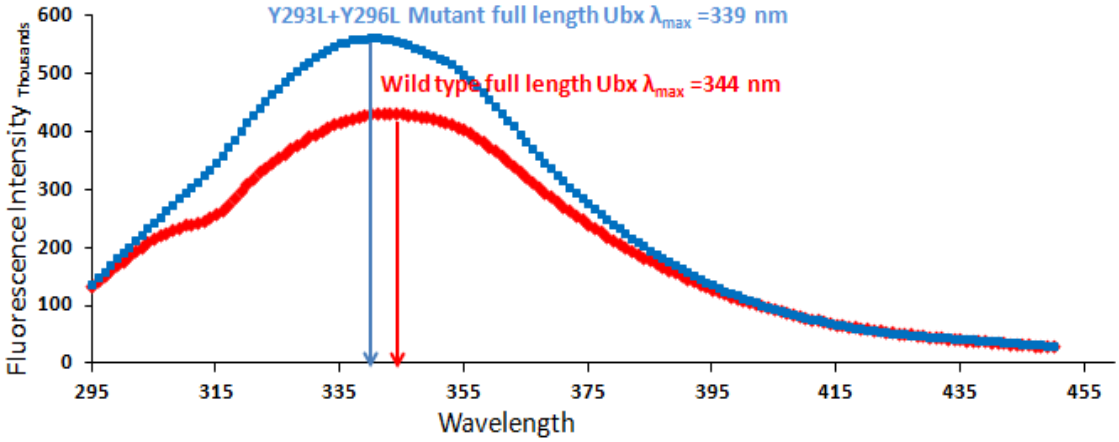

Figure 3.10 Fluorescence emission of wild-type and Y293L/Y296L Ubx monomers. Tyrosine mutations in the HD altered Ubx structure and abolished HD-nonHD region regulatory interactions. The wild-type full length Ubx fluorescence emission spectra is red shifted compared to the fluorescent emission spectra of the full-length Y293L+Y296L Ubx mutant which is blue shifted. These results indicate that the aromatic cluster is not near the positively charged residues on the homeodomain in the conformation of Ubx best able to bind DNA.

3.3.11 Oxidation of Ubx freezes the conformation in the closed state

Ubx forms di-tyrosine bonds upon oxidation of Ubx (Howell et al. 2015). If the intermolecular interactions that form Ubx materials are similar to the intramolecular interactions that regulate DNA binding, then oxidation of Ubx monomers has the potential to "lock" Ubx in the closed conformation. To test this hypothesis, DNA binding was measured in the presence of oxidized glutathione, a relatively gentle oxidant.

UbxHD binds under oxidizing and reducing conditions with the same affinity (same K_d) (Table 3.2 and Figure 3.5). Therefore, oxidation does not perturb homeodomain-DNA interaction. In contrast, full length Ubx loses the ability to bind

DNA under oxidizing conditions, in agreement with the model and our hypothesis. However, the loss of DNA binding could be due to dityrosine bond formation or disulfide bond formation.

In order to test which amino acids are contributing to the effect observed, all 6 cysteines in Ubx were mutated to alanine (6CA Ubx). The 6CA Ubx mutant still binds under oxidizing conditions; therefore, cysteines are not responsible for the loss of DNA binding by full-length Ubx under oxidizing conditions (preliminary data not shown). In contrast, every tyrosine mutant tested enables binding by full-length protein under oxidizing conditions. Furthermore, for each mutant the affinity is similar in oxidizing and reducing conditions.

We conclude from this analysis that oxidation is able to cross-link tyrosines in the cluster, thus blocking DNA binding. The fact that mutation of any tyrosine prevents the "redox effect" suggests that every tyrosine tested contributes to form the aromatic cluster. The stability of this cluster is dependent on the contribution of each tyrosine, hence the difference in DNA binding affinity for each tyrosine mutant.

3.3.12 A conformational change is required to bind DNA

For Ubx to bind DNA, Ubx must first undergo a conformational change to remove the aromatic cluster blockade. We have searched for this conformational change by three methods: 1) DNA-binding under osmotic stress, 2) Native state proteolysis, and 3) measurement of heat capacity.

3.3.13 Osmotic Stress experiments: evidence of conformational change in full length Ubx

We measured DNA binding as a function of osmotic pressure to compare changes in surface area accompanying DNA binding by UbxHD and full length Ubx. In solution, macromolecules and ligands are hydrated by water molecules. The thermodynamics of these water molecules are altered by the presence of the macromolecule (Timasheff 2002). For binding to occur between two molecules, a portion of these "bound" water molecules must be released from the surface to the bulk solution (Figure 3.11). Osmolytes that interact more favorably with bulk water than bound water are preferentially excluded from the macromolecule's surface. This exclusion is entropically favorable. As a result, the addition of osmolytes disrupts the reaction equilibrium in favor of the components with the least amount of surface area, usually forcing proteins to adopt the ligand-bound conformation (Li and Matthews 1997). The extent to which an osmolyte perturbs a reaction is dependent on both the concentration of osmolyte and the change in surface area. This effect is monitored through DNA binding affinity as a function of increasing osmolyte concentration. This osmotic stress technique is sufficiently sensitive to distinguish binding events by related proteins (reviewed in Li and Matthews 1997).

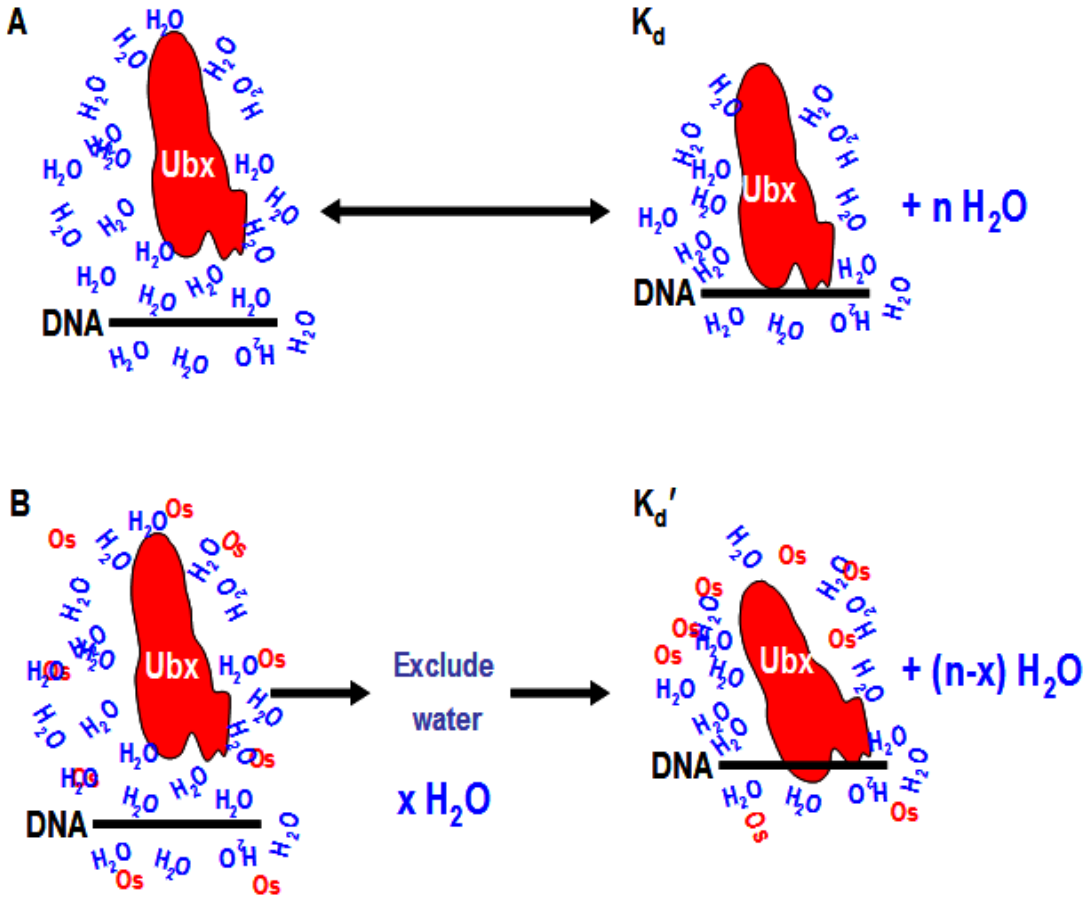

Figure 3.11 Osmotic Stress experiments. A) For binding between protein-DNA to occur, a portion of the water bound to each of their surfaces must be released to the bulk solution. **B)** Osmolytes that react more favorably with bulk water shift the reaction equilibrium to the components with the least amount of surface area.

Our model predicts a conformational change occurs in full-length Ubx upon binding DNA that exposes a significant amount of surface area for hydration. If our model is correct, then the impact of osmolytes on DNA binding by UbxHD and full length Ubx should be very different. DNA binding affinity was determined as a function of increasing osmolyte concentration using the DNA sequence 40AB, which contains a

single optimal DNA binding site (Table 3.1). We initially used glycerol, a neutral solute. Experimental results indicate that with increasing glycerol concentration, DNA binding by the full-length UbxIa was minimally affected, whereas binding by the UbxHD was substantially enhanced (Figure 3.12). Use of betaine and TEG as the osmolyte produced similar results, demonstrating that these osmolytes do not interact with the Ubx surface. These results suggest a conformational change occurs in full-length Ubx upon binding that exposes a significant amount of surface area for hydration, comparable to the surface area buried in the HD-DNA interface.

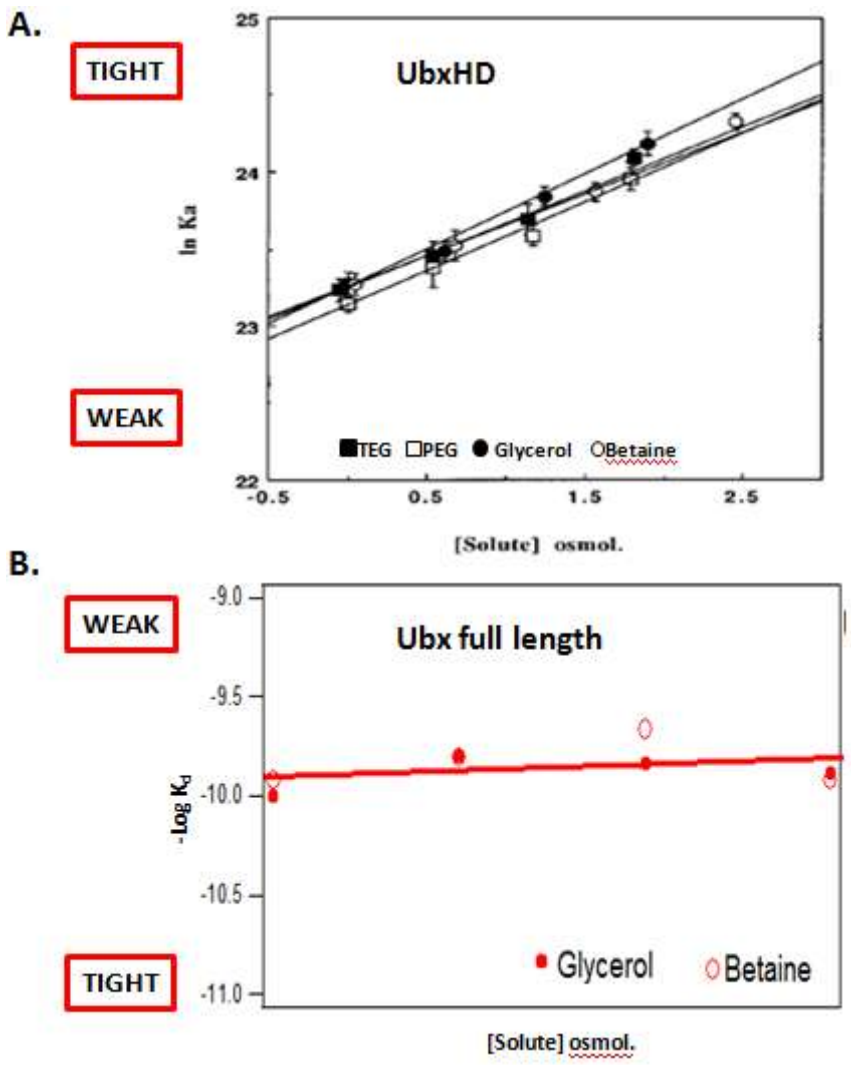

FIGURE 3.12 Monitoring DNA binding affinity as a function of osmolyte concentration: Ubx undergoes a large conformational change upon DNA binding. **A)** The equilibrium association binding constant for the Ubx homeodomain is strongly dependent on osmotic strength, but independent of the osmolyte used. (Adapted from Li and Matthews 1997). **B)** In contrast, binding by full-length Ubx appears independent of both the osmotic strength and the osmolyte used indicating that a Conformational change occurs in full-length Ubx upon binding that exposes a significant amount of surface area for hydration.

3.3.14 Further evidence of regulatory conformational change in full length Ubx by native state proteolysis

Native state proteolysis experiments are used to probe conformational features of proteins. Studies using the Lac repressor protein, have demonstrated that the sites of limited proteolysis along the polypeptide chain of a protein are characterized by enhanced backbone flexibility, implying that proteolytic probes can pinpoint the sites of local unfolding in a protein chain (Bondos 2012). If Ubx must undergo a conformational change to allow DNA binding that exposes additional surface area, then Ubx should become more prone to proteolysis in the DNA bound state. In order to test this hypothesis, Ubx was pre-incubated with buffer (as the control) or DNA initially. Different proteases were added to the pre-incubated reaction containing either Ubx-buffer or Ubx-DNA. Native state proteolysis for each reaction was monitored as a function of time.

This effect should be independent of the protease used. Preliminary Ubx native state proteolysis experiments demonstrate that DNA protects and exposes different regions of the protein (Figure 3.13) and that the conformation of Ubx in the presence and absence of DNA is different. In particular, the trypsin cut sites adjacent to the homeodomain are much more protease sensitive. This reveals that Ubx solvent accessible cut sites altered by DNA binding and conformational change. Together, these studies suggest that DNA acts as an allosteric effector of Ubx, exposing different surfaces (and potentially protein binding sites) upon binding each DNA sequence.

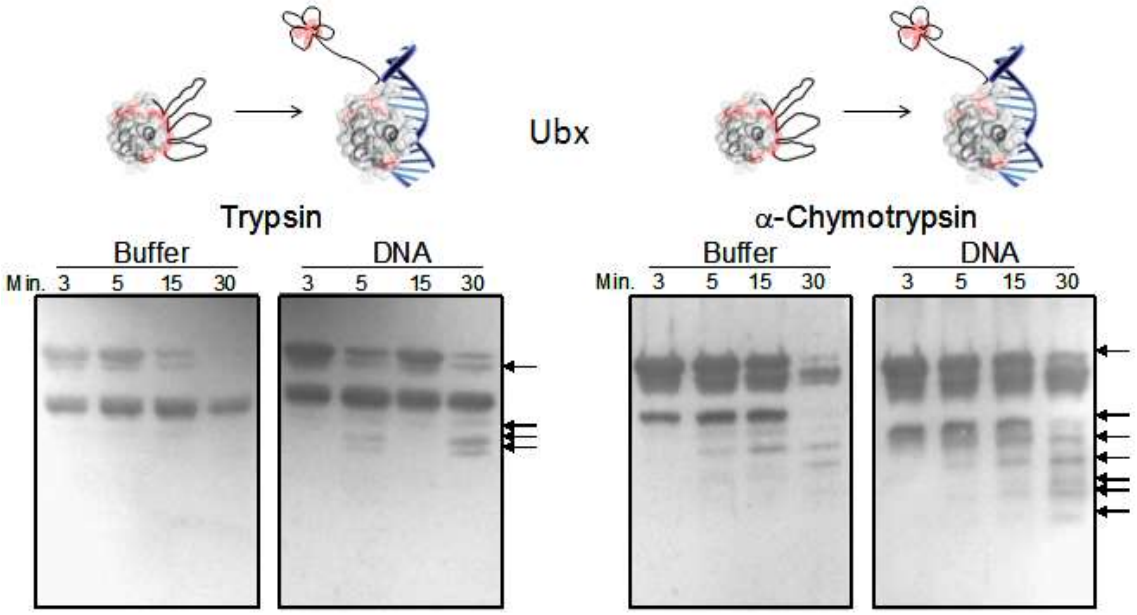

Figure 3.13 Ubx undergoes a conformational change upon DNA binding: protein becomes more exposed to solvent. DNA binding increases proteolysis of Ubx monomers (arrows) with bound to DNA. In contrast, when Ubx is not bound to Ubx there is less proteolysis when monitored over time indicating a conformational change is occurring upon DNA binding that exposes a significant amount of surface area that proteases are able to target. Furthermore, these Ubx solvent accessible cut sites are altered by DNA binding and a conformational change occurs that displaces regions of the Ubx protein which makes it more prone to proteolysis confirmed by native state proteolysis.

3.3.15 Examine DNA binding as a function of temperature to determine *ΔCp* to experimentally show conformational change in Ubx

If the protein shape changes as a result of the interaction, then that conformational change will also have an associated change in heat capacity. A nonzero change in heat capacity (*ΔCp*) is indicative of protein folding or protein unfolding. Negative Δ*Cp* of DNA binding is a thermodynamic property of the majority of

sequence-specific DNA-protein interactions. A negative heat capacity change in a biomolecular process is caused by the burial of nonpolar surface area. Over half of the systems examined display negative ΔCp values (Liu et al. 2008 (B)). In contrast, ligand binding-induced unfolding events expose buried surface area and give rise to a positive change in heat capacity.

We tested the DNA binding affinity of Ubx full length wild-type, tyrosine mutant Y293L+Y296L, and the UbxHD under reducing conditions at 4 °C, 25 °C, and 37 °C. The results in Table3.3 and the corresponding vant Hoff plot in Figure 3.14. The upwards curvature of this plot indicates a large positive change in heat capacity derived from increased solvent accessible surface area and coupled protein unfolding. In contrast, the full length Ubx mutant Y293L+Y296L, which removes the aromatic cluster from the homeodomain surface, shows little change in DNA binding affinity as a function of temperature, indicating structural changes are not required to bind DNA. The UbxHD also had similar results, which corroborate previous results.

Temperature (°C)	Full length Ubx K_d (pM)	Full length Y293L+Y296L K_d (pM)
4	26	39
22	160	28
37	44	21

Single runs only. These were one time measurements.

Table 3.3 The dissociation constant (K_d) for Ubx and Ubx tyrosine mutant Y293L+Y296L was measured via DNA binding affinity assays at different temperatures. The corresponding vant Hoff plot is in Figure 3.

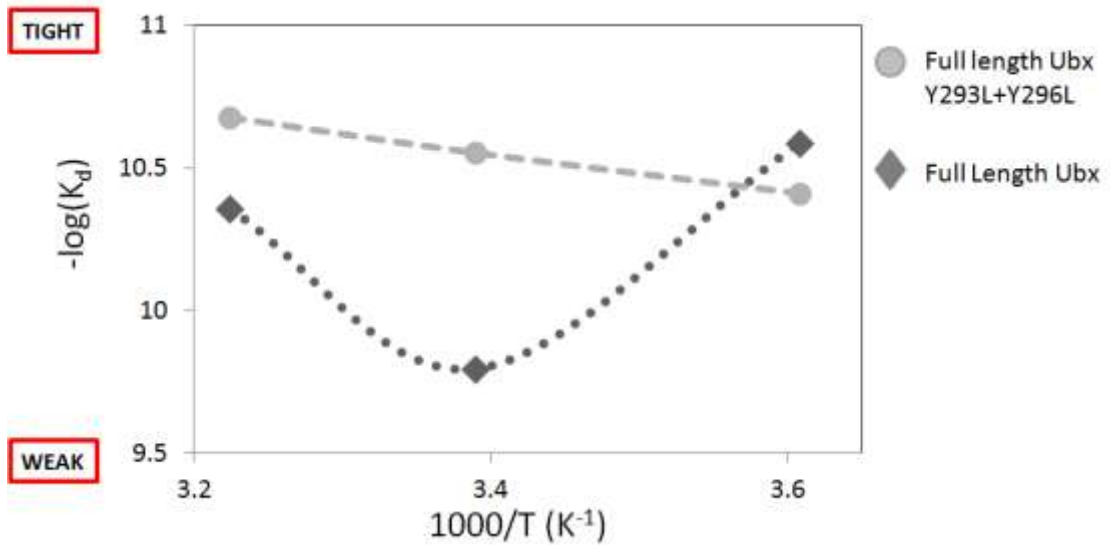

Figure 3.14 Change in heat capacity (ΔCp) confirms conformational change in Ubx. Measurement of DNA binding affinity as a function of temperature (in Kelvin) demonstrates that wild-type Ubx unfolds upon DNA binding (upward curve), whereas the Y293L/Y296L mutant, which destroys the tyrosine cluster, does not undergo a conformational change to bind DNA.

3.4 Conclusion

The sequence, structure, DNA binding affinities, and DNA sequence preferences of Hox homeodomains are extremely similar, a sharp contrast to the necessity for differential Hox functions *in vivo*. Regions outside the homeodomain may therefore be required to regulate HD•DNA interactions. Collectively, our experimental results indicate that conserved tyrosine residues in conserved motifs may form a network of interactions on the surface of the homeodomain to regulate DNA binding.

Based on our data, we have generated the first structural model of a full-length Hox protein. In this model, a tyrosine cluster blocks access to the DNA binding helix in the homeodomain. Osmotic stress, native state proteolysis, and heat capacity measurements all demonstrate that long range interactions must be disrupted to enable DNA binding. Mutagenesis of homeodomain tyrosines removes these regulatory interactions, and causes the full-length, mutant Ubx to bind DNA with properties similar to DNA binding by the UbxHD.

Such a conformational change provides an opportunity for DNA binding by the homeodomain to regulate other Ubx functions, including transcription activation and repression, protein interactions, or post-translational modifications. Conversely, formation of long-range interactions has the potential to be disrupted in a tissue-specific manner by alternative splicing, protein interactions, and/or phosphorylation. These implications will be further discussed in Section 5 and the Appendix. Thus these interactions have the potential play an important role in differentiation and regulation of Hox proteins *in vivo*. Interestingly the density of these aromatic interactions is predicted

to vary within the Hox family. Therefore, this mechanism has the potential to diversify DNA binding by Hox proteins, providing a partial solution to the Hox paradox.

4. SEQUENCE-SPECIFIC INCORPORATION OF DNA INTO PROTEIN-BASED BIOMATERIALS

4.1 Introduction

Biomolecules, such as proteins and DNA, can self-assemble to form ordered biomaterials. These biomaterials are exquisitely engineerable: (i) the composition of the constituent molecules can be precisely controlled using standard molecular biology techniques, and (ii) well-established rules of intra-and intermolecular interactions predictably and reliably drive self-assembly to defined structures (King and Lai 2013, Busseron et al. 2013). These features allow the design of novel structures and incorporation of useful functions, such as ligand binding and catalysis (Romano et al. 2011, Looger et al. 2003, Goh et al. 2000, Liberles et al. 2011). Because these molecules are natural components of living organisms, they are generally non-toxic, biodegradable, and bioresorbable, properties that are necessary for many *in vivo* applications (Sengupta and Heilshorn 2010). Furthermore, the sensitivity of these structures to environmental conditions enables their use in smart materials (Bawa et al. 2009, Michelotti et al. 2012). These properties make materials composed of biomolecules highly desirable for diverse technological and biomedical applications.

DNA is a promising building block for self-assembled materials. Its simple composition of 4 bases, combined with the well-established rules of Watson-Crick base pairing, allows the facile design of oligonucleotides that fold to form novel and complex 3D structures with unique recognition capabilities, physicochemical stability and

mechanical rigidity (Sacca and Niemeyer 2011). However, the limited number of components preclude design of very complicated structures, and design of high affinity, high specificity ligand binding sites is challenging. In contrast, proteins, composed of 20 chemically diverse amino acids, are the cellular components designed to recognize biomolecules and transform ligand binding into a cellular response. While the greater variety of chemical interactions underlying protein structure enables creation complex structures and molecular functions, this variety also complicates the design of protein-based materials (Sengupta and Heilshorn 2010, Sacca and Niemeyer 2011).

Hybrid materials, composed of both protein and DNA, potentially combine the designability of DNA structures with the functional variety of protein materials to create and pattern complex multi-scale functional devices. Hybrid DNA/protein biomaterials have led to the development of new approaches to build sensing devices, nanostructured objects, and machines (Busseron et al. 2013, de Vries et al. 2013, Michelotti et al. 2012). One difficulty in creating these materials is designing the protein-DNA interface. Current methods often rely on semisynthetic approaches involving chemical modification of the DNA, such that it can either directly react with protein or with a tag affixed to the protein (Erkelenz et al. 2011, Saghatelian et al. 2003, Niemeyer 2010). Although these approaches have many advantages, the additional steps increase materials cost. A less expensive approach is to rely on non-specific charge-charge interactions between protein and DNA, although this method lacks the exquisite spatial control characteristic of the more complicated chemical approaches (Pannier and Shea 2004). Furthermore, the rapid off-rate for non-specific protein-DNA interactions means

the DNA will not be retained by the materials for the longer timescales required for many applications.

We have circumvented these issues by constructing biomaterials using a *Drosophila* protein, Ultrabithorax (Ubx). Ubx is a transcription factor which specifically binds the DNA sequence 5'-TAAT-3'. DNA binding is mediated by a portion of the Ubx protein sequence, termed the homeodomain (Liu et al. 2008, Liu et al. 2009). In general, proteins designed for specific DNA binding are estimated to bind non-specific DNA with a K_d in the 10^{-6} to 10^{-7} range (Kalodimos et al. 2004, Von Hippel et al. 1986), and thus the high affinity of the Ubx homeodomain for its specific DNA sequence (K_d = 60 pM) represents an enhancement of 4 to 5 orders of magnitude. Materials composed of Ubx have many desirable properties: they rapidly self-assemble under very mild conditions (Greer et al. 2009), they have tunable mechanical properties that can mimic extracellular matrix proteins (Huang et al. 2010), and they are biocompatible and non-immunogenic (Patterson et al. 2014, Patterson et al. 2015). Ubx materials can be readily functionalized with proteins via gene fusion or with nanoparticles via adsorption or co-assembly (Huang et al. 2011, Majithia et al. 2011, Tsai et al. 2014). In this study, we demonstrate that pre-formed Ubx fibers retain the ability to bind DNA. Ubx fibers specifically recognize 5'-TAAT-3' binding sites in linear, supercoiled, and minicircle (strained) DNA. By adjusting the number of Ubx binding sites in the DNA sequence, DNA retention can be tuned from minutes to more than one week. Finally, DNA binding is compatible with materials assembly. This bottom-up approach allows single-pot self-assembly of protein-DNA hybrid materials without the use of tags or chemical

modifications. The DNA-sequence specificity of Ubx materials should allow precise orientation of DNA within composite materials.

4.2 Materials and Methods

4.2.1 Expression of Ubx in *E. coli*

The UbxIa splicing isoform was expressed without a histidine tag from the pET-3c-UbxIa plasmid in BL21 (DE3) pLysS *E. coli* as previously reported (Liu et al. 2008, Liu et al. 2009). Although this version of Ubx lacks an N-terminal histidine tag, it still forms materials. To purify Ubx, a cell pellet from a 2 L fermentation was lysed in 20 mL of 50 mM Tris–HCl pH=8.0, 10 mM β-mecaptoethanol, 10 mM Ethylenediaminetetraacetic acid (EDTA), 800 mM NaCl, and one tablet of Complete Proteinase Inhibitor Mixture (Roche). Following cell lysis, 40 µL of 20 mg/mL DNase I stock and 600 µL of 1 M $MgCl_2$ were added, and the mixture was incubated for 2 minutes. The lysate was centrifuged at 40,000 x *g* for 30 minutes to remove cell debris. The lysis supernatant was treated with polyethyleneimine (50% w/v, 200 µL) and centrifuged at 40,000 x *g* for 20 minutes to remove DNA fragments. The pH of the supernatant was adjusted to 6.8 with NaH_2PO_4 and centrifuged at 40,000 x *g* for 20 minutes 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. After washing with 500 mL of buffer Z, Ubx was eluted with

a 0.15–1.2 M NaCl gradient over 100 mL in buffer Z. Fractions containing Ubx were collected and dialyzed in 4 L of dialysis buffer (5% glycerol, 10 mM β-mercaptoethanol, 150 mM NaCl, and 50 mM Tris, pH = 8.0) for 25-30 minutes. Protein samples were then incubated for 15 minutes at 4 °C with 4 mL Ni-NTA agarose resin (Qiagen) that was previously equilibrated with 100 mL of dialysis buffer. The resulting slurry was poured into the column and, once settled, washed with 40 mL dialysis buffer then three successive washes: 1) 50 mL wash buffer (50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM β-mercaptoethanol and 5% glucose), 50 mL wash buffer plus 10 mM imidazole, and 3) 5 mL wash buffer plus 20 mM imidazole. Protein was eluted with 14 mL of elution buffer (wash buffer with 200 mM imidazole), analyzed by SDS-PAGE, and quantified using the standard BioRad assay.

4.2.2 Assembly of Ubx fibers using the drop method

A 250 µL volume of a solution containing 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl,10 mM β-mercaptoethanol, 5% glucose, 200 mM imidazole, and eluted Ubx, at concentrations ranging from 3 µM–6 µM, was carefully transferred using a pipette on the surface of a siliconized glass slide. The drop was covered with a 50 mL centrifuge tube cap and the tray was covered to prevent condensation. After 6 to 16 hours of incubation at room temperature and humidity, a film formed at the air-water interface that was drawn into fibers using a sterile inoculating loop and wrapping on average four times per loop. Fibers drawn from solution were immediately washed three times times in PBS buffer (20 mM NaH_2PO_4, 80 mM Na_2HPO_4, 100 mM NaCl) and stored for 1-2

hours at room temperature on sterile disposable inoculation loops in a container with wet paper towels to provide between 40%-50% humidity.

4.2.3 Immunofluorescence of Ubx fibers

Loops supporting fibers were placed in a well of a sterile 4-well cell culture plate. Blocking solution (250 µL PBS containing 0.1% Triton X-100, 1% BSA, 0.2% sodium azide, and 5% goat serum) was added to the center of the loop and allowed to incubate for 1 hour at room temperature. Fibers were then washed twice for 10 minutes in 250 µL PBS containing 0.1% Triton X-100. Primary antibody recognizing the Ubx homeodomain (FP3.38) (White and Wilcox 1984) was diluted 1:1000 in blocking solution and incubated in the wells with Ubx for an hour. After two washes, loops were incubated with goat anti-rabbit 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 immediately using confocal microscopy on a Nikon Eclipse Ti equipped with NIS Elements AR 4.10.01 software for fluorescent intensity analysis.

4.2.4 Preparation of DNAs for binding experiments

DNA sequences can be found in Table 4.3. Oligonucleotides from Table 4.1 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 the 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. The concentration of each dsDNA was calculated based on the absorbance of DNA at 260 nm (A_{260}) using the following equation:

$$[DNA]\ ln\frac{\mu g}{\mu L} = \frac{\left((A_{260})(dilution\ factor)(50)\right)}{1000}$$

Plasmid pET19b-VEGF-UbxIa plasmid (Tsai et al. 2014) 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 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.

4.2.5 Binding DNA to the fiber surface

Double stranded fluorophore-labeled DNA was diluted to 1.2 µM in PBS. Loops supporting fibers were placed in a well of a 24-well culture plate and 200 µM of the corresponding diluted fluorescent DNA was added to the center of each loop. The plate was wrapped in aluminum foil to prevent photobleaching and incubated at room temperature for six hours. Loops with fibers were lifted from the well, incubated in PBS buffer for 3-5 minutes, and washed three times in PBS buffer to remove excess DNA before viewing 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.

4.2.6 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). Primer sequences can be found in Table 4.1 below. 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. PCR products were analyzed by agarose gel electrophoresis. Gels of 2% agarose were used for product size smaller than 500 bp. Gels of 1% agarose were used for product size larger than 500bp.

Primer name	DNA sequence
vegf F' primer	CCTTAATCATATGAGTGCACCCATGG
vegf R' primer	GAACATATGATTTCGTCTTGGCTTGT
100 bp F' primer	CTTGTCACTTCAGTCAGCAG
100 bp R' primer	CATACGACTGACTGAAGTAATG
mc336 F' primer	ATCACCGAAACGCGCGAGGC
mc336 R' primer	ATCGGTGAAAACCCTTCCCG
DpnI assay F' primer	CAGATCGCTGAGATAGGTGCCTC
DpnI assay R' primer	CTTGATCCGGCAAACAAACCACC

Table 4.1 Primer sequences used in Ubx fiber binding experiments.

4.2.7 DNA binding competition experiments

100 bp DNA stocks were diluted to a final concentration of 10 µg/mL. Loops supporting fibers were then placed in a well of a 24-well culture plate. Ubx fibers were incubated for 16 hours at room temperature in the initial (1^{st}) DNA, incubated in PBS buffer for 3-5 minutes, and washed three times in PBS buffer to remove excess DNA, then incubated for 16 hours at room temperature in the competitor DNA sequence (2^{nd}), and washed three times before DNA detection by PCR.

4.2.8 DNA binding protocol for DNAs spanning 4 orders of magnitude with different structures

Loops supporting fibers were placed in a well of a 24-well culture plate. Ubx fibers were incubated separately for 16 hours at room temperature in SM, LM, SPwt, LPwt, LVwt, and LVmut DNA. The DNA stocks were diluted to a final concentration of 10 µg/mL for SM and LM. The DNA stocks for SPwt, LPwt, LVwt and LVmut DNA were diluted to four separate final concentrations: 4000 pM, 400 pM, 40 pM and 4 pM for each one. Loops supporting fibers were placed in separate wells and incubated with DNA at the specified concentrations. Each fiber was then incubated in PBS buffer for 3-5 minutes washed three times in PBS buffer to remove excess DNA, and analyzed by PCR to detect and quantify DNA binding.

4.2.9 DNA release protocol

Loops supporting fibers were placed in separate wells of a 24-well culture plate and incubated for 16 hours at room temperature in the appropriate DNA (Table 4.3). All fibers were then washed in PBS buffer, subsequently transferred to new wells of a 24-well culture plate and allowed to incubate for 0, 1, 3, 5, 7, or 9 days in PBS buffer to monitor DNA release. Each fiber was then analyzed by PCR to detect the *vegF* site within the plasmid.

4.2.10 DNA protection assay

Loops supporting fibers were placed in a well of a 24-well culture plate and incubated for 16 hours at room temperature in plasmid LPwt diluted to a final concentration of 0.02 µg/mL. All fibers were then washed in PBS buffer, subsequently transferred to new wells of a 24-well culture plate and incubate for 0, 15, and 30 minutes in a reaction mixture containing a final concentration of 1X Cutsmart buffer, 5 mg/mL BSA, and 2 units of DpnI enzyme and subsequently washed with PBS buffer. Ubx fibers were removed from the inoculating, using micro-scissors and micro-tweezers, and transferred from the inoculating loop to a 50 µL PCR reaction containing PBS buffer to heat inactivate the enzyme for 20 minutes at 80 °C. The PBS buffer was removed with a pipette and PCR was performed.

4.2.11 Co-assembly of Ubx protein and DNA fibers

Fibers co-assembled with fluorescent DNA (Table 4.3) were produced as described above with the following variations. For samples with DNA, the DNA stock was diluted to 6 μM in PBS buffer and 100 μL of the diluted DNA (or buffer for negative controls) was directly added to the Ubx drop to yield a final volume of 350 μL per drop. Fibers drawn from solution containing Ubx with and without DNA were washed immediately three times in PBS buffer and analyzed by fluorescent microscopy as described above.

4.3 Results and Discussion

For Ubx materials to bind DNA, the Ubx homeodomain (Figure 4.1a) must be both folded and accessible on the surface of the materials (Figure 4.1b). We used immunohistochemistry to determine whether FP3.38, an antibody that recognizes the Ubx homeodomain, can bind Ubx fibers. Although the antibody does not penetrate the fiber, it does recognize the homeodomain on the fiber surface (Figure 4.1c). This pattern of binding has been observed with other antibodies (Howell et al. 2015). Furthermore, the secondary antibody is incapable of recognizing the fiber in the absence of the primary antibody, verifying that the signal originates from specific recognition of the homeodomain by the FP3.38 primary antibody (Figure 4.1d).

Because the epitope for FP3.38 binding within the homeodomain is unknown, antibody binding does not guarantee the homeodomain structure is intact. This question is particularly important because Ubx self-assembles into materials at the air-water

interface (Huang et al. 2009), which induces many proteins to form beta-rich amyloid structure (Jean et al. 2012). However, we did not observe binding by Thioflavin T, which binds amyloid, either to pre-formed fibers or during materials assembly (data not shown). Conversely, if the homeodomain is folded and oriented correctly, then it should bind 5'-TAAT-3' with higher affinity than non-specific DNA (Liu et al. 2008, Kalodimos et al. 2004).

We initially used two 40 bp linear DNA constructs to determine whether Ubx fibers can bind specific (40Sp) and non-specific (40NSp) DNA sequences (Tables 4.2 and 4.3). Binding was detected using Alexa 488 fluorescent tags on the 3' end of one of the oligonucleotides in each double-stranded DNA (dsDNA) construct. Fibers were first incubated in buffer containing the dsDNA, and then rinsed three times in buffer to remove unbound DNA before detection by microscopy. The 40Sp DNA, which contains only a single 5'-TAAT-3' site, readily binds to the surface of the fibers. However, we did not detect binding by the 40NSp DNA, which lacks Ubx binding sites (Figure 4.1e,f). We verified these results using PCR to detect binding by a second set of longer dsDNAs, 100Sp and 100NSp (Tables 4.2 and 4.3). The 100Sp sequence contains four 5'-TAAT-3' binding sites, whereas the 100NSp sequence lacked any specific binding sites. Again, binding was only observed when the 5'-TAAT-3'sequence was present (Figure 4.1g).

We also performed competition experiments, in which fibers were first allowed to bind to one DNA (either a specific or non-specific sequence) and subsequently exposed to the other DNA sequence. The DNA sequence with the highest affinity should

out-compete the other sequence for binding, irrespective of the order in which the DNA is applied. In these experiments, 100Sp, but not 100NSp, was detected bound to Ubx fibers (Figure 4.2). 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.

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	linear	4	336 bp
LVwt	Specific	linear	1	504 bp
LVmut	Specific	linear	7	504 bp
SPwt	Specific	circular, supercoiled	42	7360 bp
LPwt	Specific	linear	42	7360 bp
SPmut	Specific	circular, supercoiled	48	7360 bp
LPmut	Specific	linear	48	7360 bp

Table 4.2 Characteristics of DNAs used in Ubx fiber binding experiments.

DNA Name	Sequence 5'- 3'
40Sp*	CCGGGCTGCACATGGTTAATGGCCAGTCCACG CGTAGATC
40NSp*	GATCGTGTCTACATGTCAGACAGTCAGCTGCA TGACGAGTC
100Sp	CTTATTAGCTCACGCAGTCGATCACGTATTATC AGGATGTAGAATTATCATAGATGCATCGCTGC TCATTATGCTAGTCACGCAGTCAGTCAGTCGTA TG
100NSp	CTTGTCACTTCAGTCAGCAGATACCGTAGCATC AGTATGTAGAGTTCTCACAGATGTATCACTGAT CAGGATGCTAGTCAGTACTTCAGTCAGTCGTAT G
mc336 supercoiled or linear linearized with EcoRV	ATCACCGAAACGCGCGAGGCAGCTGTATGGCA TGAAAGAGTTCTTCCCGGAAAACGCGGTGGAA TATTTCGTTTCCTACTACGACTACTATCAGCCG GAAGCCTATGTACCGAGTTCCGACACTTTCATT GAGAAAGATGCCTCAGCTCTGTTACAGGTCAC TAATACCATCTAAGTAGTTGATTCATAGTGACT GCATATGTTGTGTTTTACAGTATTATGTAGTCT GTTTTTTATGCAAAATCTAATTTAATATATTGA TATTTATATCATTTTACGTTTCTCGTTCAGCTTT TTTATACTAACTTGAGCGAAACGGGAAGGGTT TTCACCGAT

Table 4.3 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, is in red text. The DNA sequence amplified for PCR detection in Figure 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.

LVwt	CATATGAGTGCACCCATGGCAGAAGGAGGAGG GCAGAATCATCACGAAGTGGTGAAGTTCATGG ATGTCTATCAGCGCAGCTACTGCCATCCAATCG AGACCCTGGTGGACATCTTCCAGGAGTACCCT GATGAGATCGAGTACATCTTCAAGCCATCCTG TGTGCCCCTGATGCGATGCGGGGGCTGCTGCA ATGACGAGGGCCTGGAGTGTGTGCCCACTGAG GAGTCCAACATCACCATGCAGATTATGCGGAT CAAACCTCACCAAGGCCAGCACATAGGAGAGA TGAGCTTCCTACAGCACAACAAATGTGAATGC AGACCAAAGAAAGATAGAGCAAGACAAGAAA ATCCCTGTGGGCCTTGCTCAGAGCGGAGAAAG CATTTGTTTGTACAAGATCCGCAGACGTGTAA ATGTTCCTGCAAAAACACAGACTCGCGTTGCA AGGCGAGGCAGCTTGAGTTAAACGAACGTACT TGCAGATGTGACAAGCCAAGACGAAATCATAT G
SPwt supercoiled or linear	AGCTTATCGATGATAAGCTGTCAAACATGAGA ATTCTTGAAGACGAAAGGGCCTCGTGATACGC
linearized with HindIII	CTATTTTTATAGGTTAATGTCATGATAATAATG GTTTCTTAGACGTCAGGTGGCACTTTTCGGGGA AATGTGCGCGGAACCCCTATTTGTTTATTTTTC TAAATACATTCAAATATGTATCCGCTCATGAG ACAATAACCCTGATAAATGCTTCAATAATATT GAAAAAGGAAGAGTATGAGTATTCAACATTTC CGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTT GCCTTCCTGTTTTTGCTCACCCAGAAACGCTGG TGAAAGTAAAAGATGCTGAAGATCAGTTGGGT GCACGAGTGGGTTACATCGAACTGGATCTCAA CAGCGGTAAGATCCTTGAGAGTTTTCGCCCCG AAGAACGTTTTCCAATGATGAGCACTTTTAAA GTTCTGCTATGTGGCGCGGTATTATCCCGTGTT GACGCCGGGCAAGAGCAACTCGGTCGCCGCAT ACACTATTCTCAGAATGACTTGGTTGAGTACTC ACCAGTCACAGAAAAGCATCTTACGGATGGCA TGACAGTAAGAGAATTATGCAGTGCTGCCATA ACCATGAGTGATAACACTGCGGCCAACTTACT TCTGACAACGATCGGAGGACCGAAGGAGCTAA CCGCTTTTTTGCACAACATGGGGGATCATGTAA CTCGCCTTGATCGTTGGGAACCGGAGCTGAAT GAAGCCATACCAAACGACGAGCGTGACACCAC GATGCCTGCAGCAATGGCAACAACGTTGCGCA AACTATTAACTGGCGAACTACTTACTCTAGCTT CCCGGCAACAATTAATAGACTGGATGGAGGCG GATAAAGTTGCAGGACCACTTCTGCGCTCGGC

Table 4.3 Continued

SPwt supercoiled or linear (continued)	CCTTCCGGCTGGCTGGTTTATTGCTGATAAATC TGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA TTGCAGCACTGGGGCCAGATGGTAAGCCCTCC CGTATCGTAGTTATCTACACGACGGGGAGTCA GGCAACTATGGATGAACGAAATAGACAGATCG CTGAGATAGGTGCCTCACTGATTAAGCATTGG TAACTGTCAGACCAAGTTTACTCATATATACTT TAGATTGATTTAAAACTTCATTTTTAATTTAAA AGGATCTAGGTGAAGATCCTTTTTGATAATCTC ATGACCAAAATCCCTTAACGTGAGTTTTCGTTC CACTGAGCGTCAGACCCCGTAGAAAAGATCAA AGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT AATCTGCTGCTTGCAAACAAAAAAACCACCGC TACCAGCGGTGGTTTGTTTGCCGGATCAAGAG CTACCAACTCTTTTTCCGAAGGTAACTGGCTTC AGCAGAGCGCAGATACCAAATACTGTCCTTCT AGTGTAGCCGTAGTTAGGCCACCACTTCAAGA ACTCTGTAGCACCGCCTACATACCTCGCTCTGC TAATCCTGTTACCAGTGGCTGCTGCCAGTGGCG ATAAGTCGTGTCTTACCGGGTTGGACTCAAGA CGATAGTTACCGGATAAGGCGCAGCGGTCGGG CTGAACGGGGGGTTCGTGCACACAGCCCAGCT TGGAGCGAACGACCTACACCGAACTGAGATAC CTACAGCGTGAGCTATGAGAAAGCGCCACGCT TCCCGAAGGGAGAAAGGCGGACAGGTATCCG GTAAGCGGCAGGGTCGGAACAGGAGAGCGCA CGAGGGAGCTTCCAGGGGGAAACGCCTGGTAT CTTTATAGTCCTGTCGGGTTTCGCCACCTCTGA CTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG GGGCGGAGCCTATGGAAAAACGCCAGCAACG CGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGC CTTTTGCTCACATGTTCTTTCCTGCGTTATCCCC TGATTCTGTGGATAACCGTATTACCGCCTTTGA GTGAGCTGATACCGCTCGCCGCAGCCGAACGA CCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC GGAAGAGCGCCTGATGCGGTATTTTCTCCTTAC GCATCTGTGCGGTATTTCACACCGCATATATGG TGCACTCTCAGTACAATCTGCTCTGATGCCGCA TAGTTAAGCCAGTATACACTCCGCTATCGCTAC GTGACTGGGTCATGGCTGCGCCCCGACACCCG CCAACACCCGCTGACGCGCCCTGACGGGCTTG TCTGCTCCCGGCATCCGCTTACAGACAAGCTGT GACCGTCTCCGGGAGCTGCATGTGTCAGAGGT TTTCACCGTCATCACCGAAACGCGCGAGGCAG CTGCGGTAAAGCTCATCAGCGTGGTCGTGAAG CGATTCACAGATGTCTGCCTGTTCATCCGCGTC

Table 4.3 Continued

SPwt supercoiled or linear (cont.)	CAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGT CTGGCTTCTGATAAAGCGGGCCATGTTAAGGG CGGTTTTTTCCTGTTTGGTCACTGATGCCTCCG TGTAAGGGGGATTTCTGTTCATGGGGGTAATG ATACCGATGAAACGAGAGAGGATGCTCACGAT ACGGGTTACTGATGATGAACATGCCCGGTTAC TGGAACGTTGTGAGGGTAAACAACTGGCGGTA TGGATGCGGCGGGACCAGAGAAAAATCACTCA GGGTCAATGCCAGCGCTTCGTTAATACAGATG TAGGTGTTCCACAGGGTAGCCAGCAGCATCCT GCGATGCAGATCCGGAACATAATGGTGCAGGG CGCTGACTTCCGCGTTTCCAGACTTTACGAAAC ACGGAAACCGAAGACCATTCATGTTGTTGCTC AGGTCGCAGACGTTTTGCAGCAGCAGTCGCTT CACGTTCGCTCGCGTATCGGTGATTCATTCTGC TAACCAGTAAGGCAACCCCGCCAGCCTAGCCG GGTCCTCAACGACAGGAGCACGATCATGCGCA CCCGTGGCCAGGACCCAACGCTGCCCGAGATG CGCCGCGTGCGGCTGCTGGAGATGGCGGACGC GATGGATATGTTCTGCCAAGGGTTGGTTTGCGC ATTCACAGTTCTCCGCAAGAATTGATTGGCTCC AATTCTTGGAGTGGTGAATCCGTTAGCGAGGT GCCGCCGGCTTCCATTCAGGTCGAGGTGGCCC GGCTCCATGCACCGCGACGCAACGCGGGGAGG CAGACAAGGTATAGGGCGGCGCCTACAATCCA TGCCAACCCGTTCCATGTGCTCGCCGAGGCGG CATAAATCGCCGTGACGATCAGCGGTCCAGTG ATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGA TCCTTGAAGCTGTCCCTGATGGTCGTCATCTAC CTGCCTGGACAGCATGGCCTGCAACGCGGGCA TCCCGATGCCGCCGGAAGCGAGAAGAATCATA ATGGGGAAGGCCATCCAGCCTCGCGTCGCGAA CGCCAGCAAGACGTAGCCCAGCGCGTCGGCCG CCATGCCGGCGATAATGGCCTGCTTCTCGCCG AAACGTTTGGTGGCGGGACCAGTGACGAAGGC TTGAGCGAGGGCGTGCAAGATTCCGAATACCG CAAGCGACAGGCCGATCATCGTCGCGCTCCAG CGAAAGCGGTCCTCGCCGAAAATGACCCAGAG CGCTGCCGGCACCTGTCCTACGAGTTGCATGAT AAAGAAGACAGTCATAAGTGCGGCGACGATA GTCATGCCCCGCGCCCACCGGAAGGAGCTGAC TGGGTTGAAGGCTCTCAAGGGCATCGGTCGAG ATCCCGGTGCCTAATGAGTGAGCTAACTTACA TTAATTGCGTTGCGCTCACTGCCCGCTTTCCAG TCGGGAAACCTGTCGTGCCAGCTGCATTAATG AATCGGCCAACGCGCGGGGAGAGGCGGTTTGC

Table 4.3 continued

SPwt supercoiled or linear (cont.)	GTATTGGGCGCCAGGGTGGTTTTTCTTTTCACC AGTGAGACGGGCAACAGCTGATTGCCCTTCAC CGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGT CCACGCTGGTTTGCCCCAGCAGGCGAAAATCC TGTTTGATGGTGGTTAACGGCGGGATATAACA TGAGCTGTCTTCGGTATCGTCGTATCCCACTAC CGAGATATCCGCACCAACGCGCAGCCCGGACT CGGTAATGGCGCGCATTGCGCCCAGCGCCATC TGATCGTTGGCAACCAGCATCGCAGTGGGAAC GATGCCCTCATTCAGCATTTGCATGGTTTGTTG AAAACCGGACATGGCACTCCAGTCGCCTTCCC GTTCCGCTATCGGCTGAATTTGATTGCGAGTGA GATATTTATGCCAGCCAGCCAGACGCAGACGC GCCGAGACAGAACTTAATGGGCCCGCTAACAG CGCGATTTGCTGGTGACCCAATGCGACCAGAT GCTCCACGCCCAGTCGCGTACCGTCTTCATGGG AGAAAATAATACTGTTGATGGGTGTCTGGTCA GAGACATCAAGAAATAACGCCGGAACATTAGT GCAGGCAGCTTCCACAGCAATGGCATCCTGGT CATCCAGCGGATAGTTAATGATCAGCCCACTG ACGCGTTGCGCGAGAAGATTGTGCACCGCCGC TTTACAGGCTTCGACGCCGCTTCGTTCTACCAT CGACACCACCACGCTGGCACCCAGTTGATCGG CGCGAGATTTAATCGCCGCGACAATTTGCGAC GGCGCGTGCAGGGCCAGACTGGAGGTGGCAAC GCCAATCAGCAACGACTGTTTGCCCGCCAGTT GTTGTGCCACGCGGTTGGGAATGTAATTCAGC TCCGCCATCGCCGCTTCCACTTTTTCCCGCGTT TTCGCAGAAACGTGGCTGGCCTGGTTCACCAC GCGGGAAACGGTCTGATAAGAGACACCGGCAT ACTCTGCGACATCGTATAACGTTACTGGTTTCA CATTCACCACCCTGAATTGACTCTCTTCCGGGC GCTATCATGCCATACCGCGAAAGGTTTTGCGC CATTCGATGGTGTCCGGGATCTCGACGCTCTCC CTTATGCGACTCCTGCATTAGGAAGCAGCCCA GTAGTAGGTTGAGGCCGTTGAGCACCGCCGCC GCAAGGAATGGTGCATGCAAGGAGATGGCGCC CAACAGTCCCCCGGCCACGGGGCCTGCCACCA TACCCACGCCGAAACAAGCGCTCATGAGCCCG AAGTGGCGAGCCCGATCTTCCCCATCGGTGAT GTCGGCGATATAGGCGCCAGCAACCGCACCTG TGGCGCCGGTGATGCCGGCCACGATGCGTCCG GCGTAGAGGATCGAGATCTCGATCCCGCGAAA TTAATACGACTCACTATAGGGGAATTGTGAGC GGATAACAATTCCCCTCTAGAAATAATTTTGTT TAACTTTAAGAAGGAGATATACCATGGGCCAT

Table 4.3 continued

SPwt supercoiled or linear (cont.)	CATCATCATCATCATCATCATCACAGCAGC GGCCATATCGACGACGACGACAAGCATATGAG TGCACCCATGGCAAAGGAGGAGGGCAGAATC ATCACGAAGTGGTGAAGTTCATGGATGTCTAT CAGCGCAGCTACTGCCATCCAATCGAGACCCT GGTGGACATCTTCCAGGAGTACCCTGATGAGA TCGAGTACATCTTCAAGCCATCCTGTGTGCCCC TGATGCGATGCGGGGGCTGCTGCAATGACGAG GGCCTGGAGTGTGTGCCCACTGAGGAGTCCAA CATCACCATGCAGATTATGCGGATCAAACCTC ACCAAGGCCAGCACATAGGAGAGATGAGCTTC CTACAGCACAACAAATGTGAATGCAGACCAAA GAAAGATAGAGCAAGACAAGAAAATCCCTGT GGGCCTTGCTCAGAGCGGAGAAAGCATTTGTT TGTACAAGATCCGCAGACGTGTAAATGTTCCT GCAAAAACACAGACTCGCGTTGCAAGGCGAGG CAGCTTGAGTTAAACGAACGTACTTGCAGATG TGACAAGCCAAGACGAAATCATATGAACTCGT ACTTTGAACAGGCCTCCGGCTTTTATGGCCATC CGCACCAGGCCACCGGAATGGCGATGGGCAGC GGTGGCCACCACGACCAGACGGCCAGTGCAGC GGCGGCCGCGTACAGGGGATTCCCTCTCTCGC TGGGCATGAGTCCCTATGCCAACCACCATCTG CAGCGCACCACCCAGGACTCGCCCTACGATGC CAGCATCACGGCCGCCTGCAATAAGATATACG GCGATGGAGCCGGAGCCTACAAACAGGACTGC CTGAACATCAAGGCGGATGCGGTGAATGGCTA CAAAGACATTTGGAACACGGGCGGCTCGAATG GCGGCGGGGGTGGCGGCGGAGGCGGTGGTGG CGGCGGAGCGGGCGGAACAGGTGGAGCCGGC AATGCCAATGGCGGTAATGCGGCCAATGCAAA CGGACAGAACAATCCGGCGGGCGGTATGCCCG TTAGACCCTCCGCCTGCACCCCAGATTCCCGAG TGGGCGGCTACTTGGACACGTCGGGCGGCAGT CCCGTTAGCCATCGCGGCGGCAGTGCCGGCGG TAATGTGAGTGTCAGCGGCGGCAACGGCAACG CCGGAGGCGTACAGAGCGGCGTGGGCGTGGCC GGAGCGGGCACTGCCTGGAATGCCAATTGCAC CATCTCGGGCGCCGCTGCCCAAACGGCGGCCG CCAGCAGTTTACACCAGGCCAGCAATCACACA TTCTACCCCTGGATGGCTATCGCAGGTAAGAT AAGATCTGATTTAACACAATACGGCGGCATAT CAACAGACATGGGTAAGAGATACTCAGAATCT CTTGCGGGCTCACTTCTACCAGACTGGCTAGGT ACAAATGGTCTGCGAAGACGCGGCCGACAGAC ATACACCCGCTACCAGACGCTCGAGCTGGAGA

Table 4.3 continued

	AGGAGTTCCACACGAATCATTATCTGACCCGC AGACGGAGAATCGAGATGGCGCACGCGCTATG CCTGACGGAGCGGCAGATCAAGATCTGGTTCC AGAACCGGCGAATGAAGCTGAAGAAGGAGAT CCAGGCGATCAAGGAGCTGAACGAACAGGAG AAGCAGGCGCAGGCCCAGAAGGCGGCGGCGG CAGCGGCTGCGGCGGCGGCGGTCCAAGGTGGA CACTTAGATCAGTAATAGGGTTAGGCTGCTAA CAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTG CCACCGCTGAGCAATAACTAGCATAACCCCTT GGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG CTGAAAGGAGGAACTATATCCGGATATCCCGC AAGAGGCCCGGCAGTACCGGCATAACCAAGCC TATGCCTACAGCATCCAGGGTGACGGTGCCGA GGATGACGATGAGCGCATTGTTAGATTTCATA CACGGTGCCTGACTGCGTTAGCAATTTAACTGT GATAAACTACCGCATTAA
SPmut supercoiled or linear	AGCTTATCGATGATAAGCTGTCAAACATGAGA ATTCTTGAAGACGAAAGGGCCTCGTGATACGC
linearized with HindIII	CTATTTTTATAGGTTAATGTCATGATAATAATG GTTTCTTAGACGTCAGGTGGCACTTTTCGGGGA AATGTGCGCGGAACCCCTATTTGTTTATTTTTC TAAATACATTCAAATATGTATCCGCTCATGAG ACAATAACCCTGATAAATGCTTCAATAATATT GAAAAAGGAAGAGTATGAGTATTCAACATTTC CGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTT GCCTTCCTGTTTTTGCTCACCCAGAAACGCTGG TGAAAGTAAAAGATGCTGAAGATCAGTTGGGT GCACGAGTGGGTTACATCGAACTGGATCTCAA CAGCGGTAAGATCCTTGAGAGTTTTCGCCCCG AAGAACGTTTTCCAATGATGAGCACTTTTAAA GTTCTGCTATGTGGCGCGGTATTATCCCGTGTT GACGCCGGGCAAGAGCAACTCGGTCGCCGCAT ACACTATTCTCAGAATGACTTGGTTGAGTACTC ACCAGTCACAGAAAAGCATCTTACGGATGGCA TGACAGTAAGAGAATTATGCAGTGCTGCCATA ACCATGAGTGATAACACTGCGGCCAACTTACT TCTGACAACGATCGGAGGACCGAAGGAGCTAA CCGCTTTTTTGCACAACATGGGGGATCATGTAA CTCGCCTTGATCGTTGGGAACCGGAGCTGAAT GAAGCCATACCAAACGACGAGCGTGACACCAC GATGCCTGCAGCAATGGCAACAACGTTGCGCA AACTATTAACTGGCGAACTACTTACTCTAGCTT CCCGGCAACAATTAATAGACTGGATGGAGGCG GATAAAGTTGCAGGACCACTTCTGCGCTCGGC

Table 4.3 continued

SPmut supercoiled or linear (cont.)	CCTTCCGGCTGGCTGGTTTATTGCTGATAAATC TGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA TTGCAGCACTGGGGCCAGATGGTAAGCCCTCC CGTATCGTAGTTATCTACACGACGGGGAGTCA GGCAACTATGGATGAACGAAATAGACAGATCG CTGAGATAGGTGCCTCACTGATTAAGCATTGG TAACTGTCAGACCAAGTTTACTCATATATACTT TAGATTGATTTAAAACTTCATTTTTAATTTAAA AGGATCTAGGTGAAGATCCTTTTTGATAATCTC ATGACCAAAATCCCTTAACGTGAGTTTTCGTTC CACTGAGCGTCAGACCCCGTAGAAAAGATCAA AGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT AATCTGCTGCTTGCAAACAAAAAAACCACCGC TACCAGCGGTGGTTTGTTTGCCGGATCAAGAG CTACCAACTCTTTTTCCGAAGGTAACTGGCTTC AGCAGAGCGCAGATACCAAATACTGTCCTTCT AGTGTAGCCGTAGTTAGGCCACCACTTCAAGA ACTCTGTAGCACCGCCTACATACCTCGCTCTGC TAATCCTGTTACCAGTGGCTGCTGCCAGTGGCG ATAAGTCGTGTCTTACCGGGTTGGACTCAAGA CGATAGTTACCGGATAAGGCGCAGCGGTCGGG CTGAACGGGGGGTTCGTGCACACAGCCCAGCT TGGAGCGAACGACCTACACCGAACTGAGATAC CTACAGCGTGAGCTATGAGAAAGCGCCACGCT TCCCGAAGGGAGAAAGGCGGACAGGTATCCG GTAAGCGGCAGGGTCGGAACAGGAGAGCGCA CGAGGGAGCTTCCAGGGGGAAACGCCTGGTAT CTTTATAGTCCTGTCGGGTTTCGCCACCTCTGA CTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG GGGCGGAGCCTATGGAAAAACGCCAGCAACG CGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGC CTTTTGCTCACATGTTCTTTCCTGCGTTATCCCC TGATTCTGTGGATAACCGTATTACCGCCTTTGA GTGAGCTGATACCGCTCGCCGCAGCCGAACGA CCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC GGAAGAGCGCCTGATGCGGTATTTTCTCCTTAC GCATCTGTGCGGTATTTCACACCGCATATATGG TGCACTCTCAGTACAATCTGCTCTGATGCCGCA TAGTTAAGCCAGTATACACTCCGCTATCGCTAC GTGACTGGGTCATGGCTGCGCCCCGACACCCG CCAACACCCGCTGACGCGCCCTGACGGGCTTG TCTGCTCCCGGCATCCGCTTACAGACAAGCTGT GACCGTCTCCGGGAGCTGCATGTGTCAGAGGT TTTCACCGTCATCACCGAAACGCGCGAGGCAG CTGCGGTAAAGCTCATCAGCGTGGTCGTGAAG CGATTCACAGATGTCTGCCTGTTCATCCGCGTC

Table 4.3 continued

SPmut supercoiled or linear (cont.)	CAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGT CTGGCTTCTGATAAAGCGGGCCATGTTAAGGG CGGTTTTTTCCTGTTTGGTCACTGATGCCTCCG TGTAAGGGGGATTTCTGTTCATGGGGGTAATG ATACCGATGAAACGAGAGAGGATGCTCACGAT ACGGGTTACTGATGATGAACATGCCCGGTTAC TGGAACGTTGTGAGGGTAAACAACTGGCGGTA TGGATGCGGCGGGACCAGAGAAAAATCACTCA GGGTCAATGCCAGCGCTTCGTTAATACAGATG TAGGTGTTCCACAGGGTAGCCAGCAGCATCCT GCGATGCAGATCCGGAACATAATGGTGCAGGG CGCTGACTTCCGCGTTTCCAGACTTTACGAAAC ACGGAAACCGAAGACCATTCATGTTGTTGCTC AGGTCGCAGACGTTTTGCAGCAGCAGTCGCTT CACGTTCGCTCGCGTATCGGTGATTCATTCTGC TAACCAGTAAGGCAACCCCGCCAGCCTAGCCG GGTCCTCAACGACAGGAGCACGATCATGCGCA CCCGTGGCCAGGACCCAACGCTGCCCGAGATG CGCCGCGTGCGGCTGCTGGAGATGGCGGACGC GATGGATATGTTCTGCCAAGGGTTGGTTTGCGC ATTCACAGTTCTCCGCAAGAATTGATTGGCTCC AATTCTTGGAGTGGTGAATCCGTTAGCGAGGT GCCGCCGGCTTCCATTCAGGTCGAGGTGGCCC GGCTCCATGCACCGCGACGCAACGCGGGGAGG CAGACAAGGTATAGGGCGGCGCCTACAATCCA TGCCAACCCGTTCCATGTGCTCGCCGAGGCGG CATAAATCGCCGTGACGATCAGCGGTCCAGTG ATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGA TCCTTGAAGCTGTCCCTGATGGTCGTCATCTAC CTGCCTGGACAGCATGGCCTGCAACGCGGGCA TCCCGATGCCGCCGGAAGCGAGAAGAATCATA ATGGGGAAGGCCATCCAGCCTCGCGTCGCGAA CGCCAGCAAGACGTAGCCCAGCGCGTCGGCCG CCATGCCGGCGATAATGGCCTGCTTCTCGCCG AAACGTTTGGTGGCGGGACCAGTGACGAAGGC TTGAGCGAGGGCGTGCAAGATTCCGAATACCG CAAGCGACAGGCCGATCATCGTCGCGCTCCAG CGAAAGCGGTCCTCGCCGAAAATGACCCAGAG CGCTGCCGGCACCTGTCCTACGAGTTGCATGAT AAAGAAGACAGTCATAAGTGCGGCGACGATA GTCATGCCCCGCGCCCACCGGAAGGAGCTGAC TGGGTTGAAGGCTCTCAAGGGCATCGGTCGAG ATCCCGGTGCCTAATGAGTGAGCTAACTTACA TTAATTGCGTTGCGCTCACTGCCCGCTTTCCAG TCGGGAAACCTGTCGTGCCAGCTGCATTAATG AATCGGCCAACGCGCGGGGAGAGGCGGTTTGC

Table 4.3 continued

SPmut supercoiled or linear (cont.)	GTATTGGGCGCCAGGGTGGTTTTTCTTTTCACC AGTGAGACGGGCAACAGCTGATTGCCCTTCAC CGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGT CCACGCTGGTTTGCCCCAGCAGGCGAAAATCC TGTTTGATGGTGGTTAACGGCGGGATATAACA TGAGCTGTCTTCGGTATCGTCGTATCCCACTAC CGAGATATCCGCACCAACGCGCAGCCCGGACT CGGTAATGGCGCGCATTGCGCCCAGCGCCATC TGATCGTTGGCAACCAGCATCGCAGTGGGAAC GATGCCCTCATTCAGCATTTGCATGGTTTGTTG AAAACCGGACATGGCACTCCAGTCGCCTTCCC GTTCCGCTATCGGCTGAATTTGATTGCGAGTGA GATATTTATGCCAGCCAGCCAGACGCAGACGC GCCGAGACAGAACTTAATGGGCCCGCTAACAG CGCGATTTGCTGGTGACCCAATGCGACCAGAT GCTCCACGCCCAGTCGCGTACCGTCTTCATGGG AGAAAATAATACTGTTGATGGGTGTCTGGTCA GAGACATCAAGAAATAACGCCGGAACATTAGT GCAGGCAGCTTCCACAGCAATGGCATCCTGGT CATCCAGCGGATAGTTAATGATCAGCCCACTG ACGCGTTGCGCGAGAAGATTGTGCACCGCCGC TTTACAGGCTTCGACGCCGCTTCGTTCTACCAT CGACACCACCACGCTGGCACCCAGTTGATCGG CGCGAGATTTAATCGCCGCGACAATTTGCGAC GGCGCGTGCAGGGCCAGACTGGAGGTGGCAAC GCCAATCAGCAACGACTGTTTGCCCGCCAGTT GTTGTGCCACGCGGTTGGGAATGTAATTCAGC TCCGCCATCGCCGCTTCCACTTTTTCCCGCGTT TTCGCAGAAACGTGGCTGGCCTGGTTCACCAC GCGGGAAACGGTCTGATAAGAGACACCGGCAT ACTCTGCGACATCGTATAACGTTACTGGTTTCA CATTCACCACCCTGAATTGACTCTCTTCCGGGC GCTATCATGCCATACCGCGAAAGGTTTTGCGC CATTCGATGGTGTCCGGGATCTCGACGCTCTCC CTTATGCGACTCCTGCATTAGGAAGCAGCCCA GTAGTAGGTTGAGGCCGTTGAGCACCGCCGCC GCAAGGAATGGTGCATGCAAGGAGATGGCGCC CAACAGTCCCCCGGCCACGGGGCCTGCCACCA TACCCACGCCGAAACAAGCGCTCATGAGCCCG AAGTGGCGAGCCCGATCTTCCCCATCGGTGAT GTCGGCGATATAGGCGCCAGCAACCGCACCTG TGGCGCCGGTGATGCCGGCCACGATGCGTCCG GCGTAGAGGATCGAGATCTCGATCCCGCGAAA TTAATACGACTCACTATAGGGGAATTGTGAGC GGATAACAATTCCCCTCTAGAAATAATTTTGTT TAACTTTAAGAAGGAGATATACCATGGGCCAT

Table 4.3 continued

SPmut (cont.)	supercoiled or linear
	CATCATCATCATCATCATCATCACAGCAGC
	GGCCATATCGACGACGACGACAAGCATATGAG
	TGCACCCATGGCAGAAGGAGGAGGGCATAATC
	ATCACGAAGTGGTGAAGTTCATGGATGTCTAT
	CAGCGCAGCTACTGCCATCTAATCGAGACCCT
	GGTGGACATCTTCCAGGAGTACCCTGATGAGA
	TCGAGTACATCTTCAAGCCATCCTGTGTGCCCC
	TGATGCGATGCGGGGGCTGCTGTAATGACGAG
	GGCCTGGAGTGTGTGCCCACTGAGGAGTCCAA
	CATCACCATGCAGATTATGCGGATCAAACCTC
	ACCAAGGCCAGCACATAGGAGAGATGAGCTTC
	CTACAGCACAACTAATGTGAATGCAGACCAAA
	GAAAGATAGAGCAAGACAAGAAAATCCCTGT
	GGGCCTTGCTCAGAGCGGAGAAAGCATTAGTT
	TGTACAAGATCCGCAGACGTGTAAATGTTCCT
	GCAAAAACACAGACTCGCGTTGCAAGGCGAGG
	CAGCTTGAGTTAATCGAACGTACTTGCAGATG
	TGACAAGCCAAGACGAAATCATATGAACTCGT
	ACTTTGAACAGGCCTCCGGCTTTTATGGCCATC
	CGCACCAGGCCACCGGAATGGCGATGGGCAGC
	GGTGGCCACCACGACCAGACGGCCAGTGCAGC
	GGCGGCCGCGTACAGGGGATTCCCTCTCTCGC
	TGGGCATGAGTCCCTATGCCAACCACCATCTG
	CAGCGCACCACCCAGGACTCGCCCTACGATGC
	CAGCATCACGGCCGCCTGCAATAAGATATACG
	GCGATGGAGCCGGAGCCTACAAACAGGACTGC
	CTGAACATCAAGGCGGATGCGGTGAATGGCTA
	CAAAGACATTTGGAACACGGGCGGCTCGAATG
	GCGGCGGGGGTGGCGGCGGAGGCGGTGGTGG
	CGGCGGAGCGGGCGGAACAGGTGGAGCCGGC
	AATGCCAATGGCGGTAATGCGGCCAATGCAAA
	CGGACAGAACAATCCGGCGGGCGGTATGCCCG
	TTAGACCCTCCGCCTGCACCCCAGATTCCCGAG
	TGGGCGGCTACTTGGACACGTCGGGCGGCAGT
	CCCGTTAGCCATCGCGGCGGCAGTGCCGGCGG
	TAATGTGAGTGTCAGCGGCGGCAACGGCAACG
	CCGGAGGCGTACAGAGCGGCGTGGGCGTGGCC
	GGAGCGGGCACTGCCTGGAATGCCAATTGCAC
	CATCTCGGGCGCCGCTGCCCAAACGGCGGCCG
	CCAGCAGTTTACACCAGGCCAGCAATCACACA
	TTCTACCCCTGGATGGCTATCGCAGGTAAGAT
	AAGATCTGATTTAACACAATACGGCGGCATAT
	CAACAGACATGGGTAAGAGATACTCAGAATCT
	CTTGCGGGCTCACTTCTACCAGACTGGCTAGGT
	ACAAATGGTCTGCGAAGACGCGGCCGACAGAC
	ATACACCCGCTACCAGACGCTCGAGCTGGAGA

Table 4.3 continued

SPmut supercoiled or linear (cont.)	AGGAGTTCCACACGAATCATTATCTGACCCGC AGACGGAGAATCGAGATGGCGCACGCGCTATG CCTGACGGAGCGGCAGATCAAGATCTGGTTCC AGAACCGGCGAATGAAGCTGAAGAAGGAGAT CCAGGCGATCAAGGAGCTGAACGAACAGGAG AAGCAGGCGCAGGCCCAGAAGGCGGCGGCGG CAGCGGCTGCGGCGGCGGCGGTCCAAGGTGGA CACTTAGATCAGTAATAGGGTTAGGCTGCTAA CAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTG CCACCGCTGAGCAATAACTAGCATAACCCCTT GGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG CTGAAAGGAGGAACTATATCCGGATATCCCGC AAGAGGCCCGGCAGTACCGGCATAACCAAGCC TATGCCTACAGCATCCAGGGTGACGGTGCCGA GGATGACGATGAGCGCATTGTTAGATTTCATA CACGGTGCCTGACTGCGTTAGCAATTTAACTGT GATAAACTACCGCATTAA

Table 4.3 continued

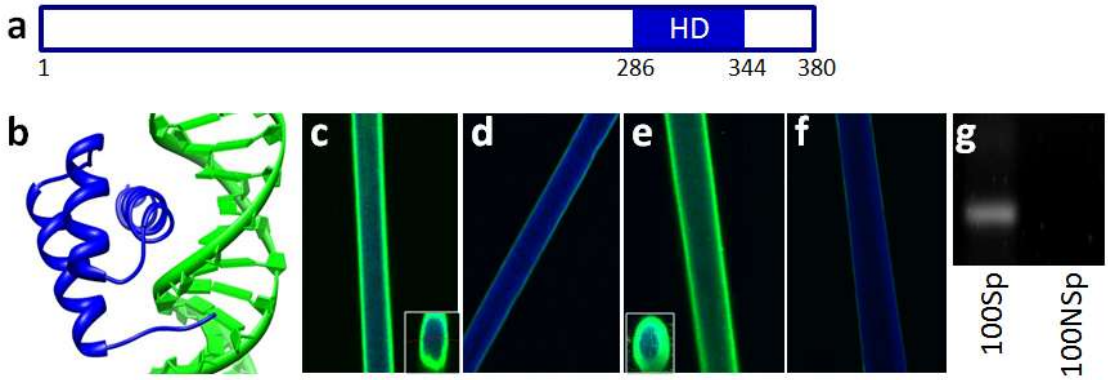

Figure 4.1 In Ubx fibers, the homeodomain is on the surface and capable of binding DNA in sequence-specific manner. a, Schematic of the UbxIa amino acid sequence, showing the placement and relative size of the DNA-binding homeodomain **b,** Structure of the Ubx homeodomain (blue) bound to DNA (green), derived from PDB 1B8I (Passner et al. 1999). **c,** Immunofluorescence assays (green) demonstrate the anti-homeodomain antibody FP3.38 (White and Wilcox 1984), binds Ubx fibers, which autofluoresce blue due to dityrosine bonds (Howell et al. 2015). **d,** Ubx fibers are not **Figure 4.1 continued,** recognized by secondary antibody in the absence of FP3.38. **e,** 40Sp DNA with a fluorescent tag (green) bind Ubx fibers but not 40NSp DNA **f**, a similar sequence lacking Ubx binding sites. **g,** PCR of Ubx fibers exposed to DNA detects the presence of 100Sp DNA, but not 100NSp DNA. The characteristics of sequences of DNAs used in this work are available in Tables 1 and 2.

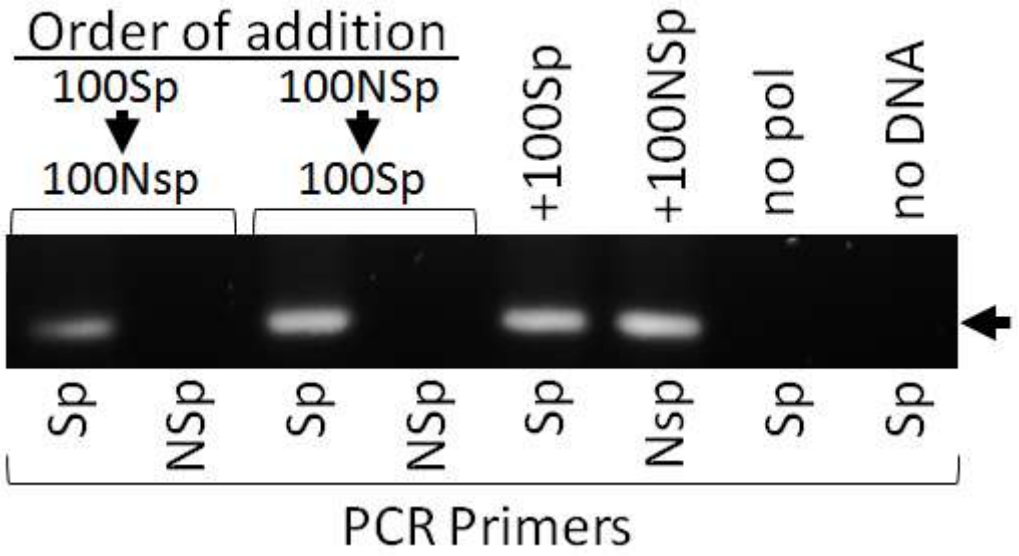

Figure 4.2 Competition experiments demonstrate DNA-binding specificity by Ubx fibers. Ubx fibers, incubated first with 100Sp DNA, are not able to subsequently bind 100NSp non-specific DNA. In contrast, Ubx fibers incubated first in 100NSp DNA are able to subsequently bind 100Sp DNA without retaining 100NSp 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 100NSp 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).

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 impact binding to Ubx fibers, we tested the ability of Ubx fibers to bind to mv336, a 336 bp minicircle containing four 5'-TAAT-3' sites. Minicircles are small circular supercoiled dsDNAs being developed for gene therapy (Mayrhofer et al. 2009, Kobelt et al. 2013, Zhao et al 2011). The strain induced by bending DNA into smaller circles is

partially relieved by supercoiling. We combared Ubx fiber binding by supercoiled mv336 (SM) and minicircle linearized by digestion with EcoRV (LM) (Fogg et al. 2006). Both SM and LM DNA bind Ubx fibers (Figure 4.3a). Therefore neither DNA supercoiling nor strain prevents fiber binding. We also compared binding of a supercoiled plasmid DNA (SPwt) with binding by the same plasmid, linearized with HindIII (LPwt) (Tables 4.2 and 4.3, Figure 4.4a). Both the supercoiled and linear plasmids, each of which contain 42 5'-TAAT-3' sites, bound Ubx fibers to similar extents at DNA concentrations spanning four orders of magnitude (Figure 4.3b). Comparison of the minicircle and plasmid DNA binding experiments demonstrates that observation of similar binding for supercoiled and linearized DNAs is not an artifact of DNA sequence or DNA length. We conclude that non-linear DNA structures do not preclude binding to Ubx fibers.

To test whether the number of DNA binding sites present modulates binding to Ubx fibers, we compared binding of the wild-type *vegf* DNA sequence (LVwt) which has a single binding site, and a mutant version (LVmut) in which an additional six binding sites have been added. Although significant differences in binding were not observed at high (saturating) concentrations of DNA, a larger amount of the mutant DNA was bound at lower DNA concentrations (Figure 4.2c, Figure 4.4b). Therefore, the extent of DNA binding can be tuned by altering the number of binding sites.

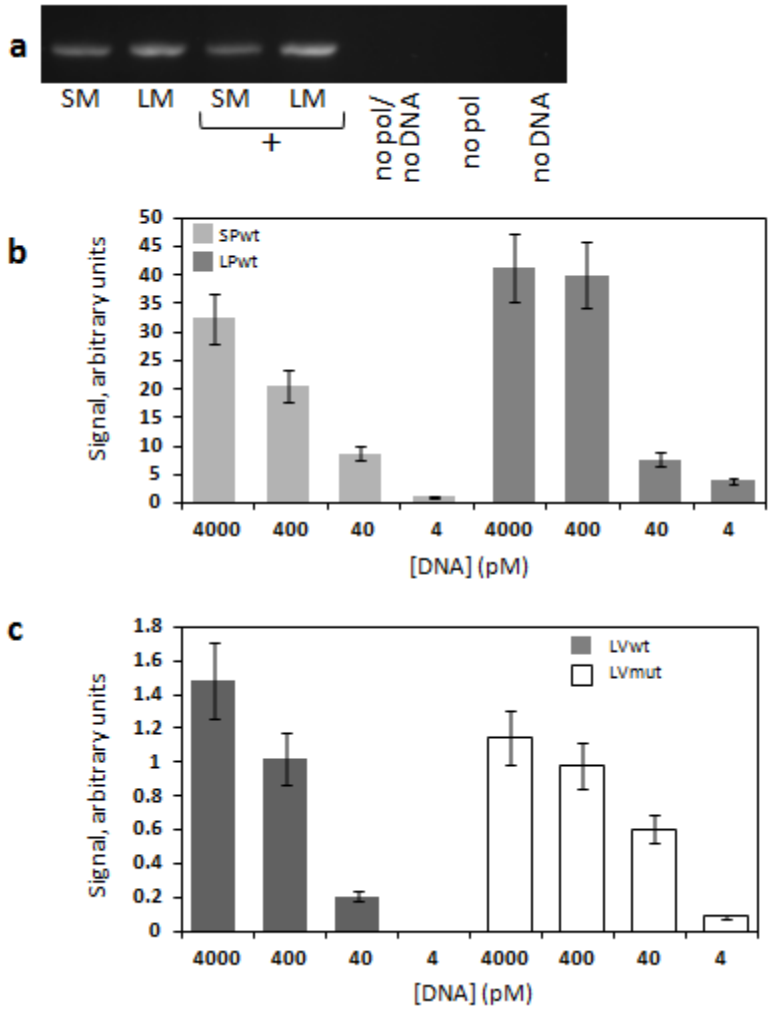

Figure 4.3 Ubx fibers bind supercoiled DNA and linear DNA. a, Supercoiled, circular minivector (SM) and linearized minivector DNA (LM) both bind Ubx fibers. Included are PCR reactions of positive controls using template DNA in solution (+) and negative controls lacking polymerase (no pol) and/or DNA (no DNA). **b-c,** Quantitation of data in supplementary figure 2, varying DNA concentration reveals qualitative differences in affinity for the four DNA samples. The characteristics and sequences of DNAs used are reported in Supplementary Information Tables 1 and 2. **b,** Both the SPwt and LPwt bound Ubx fibers to similar extents at DNA concentrations spanning four **Figure 4.3 continued,** orders of magnitude. **c,** Significant differences in binding were not observed at high (saturating) concentrations of DNA for all DNAs, and the mutant sequence binds better at lower DNA concentrations for Lvmut.

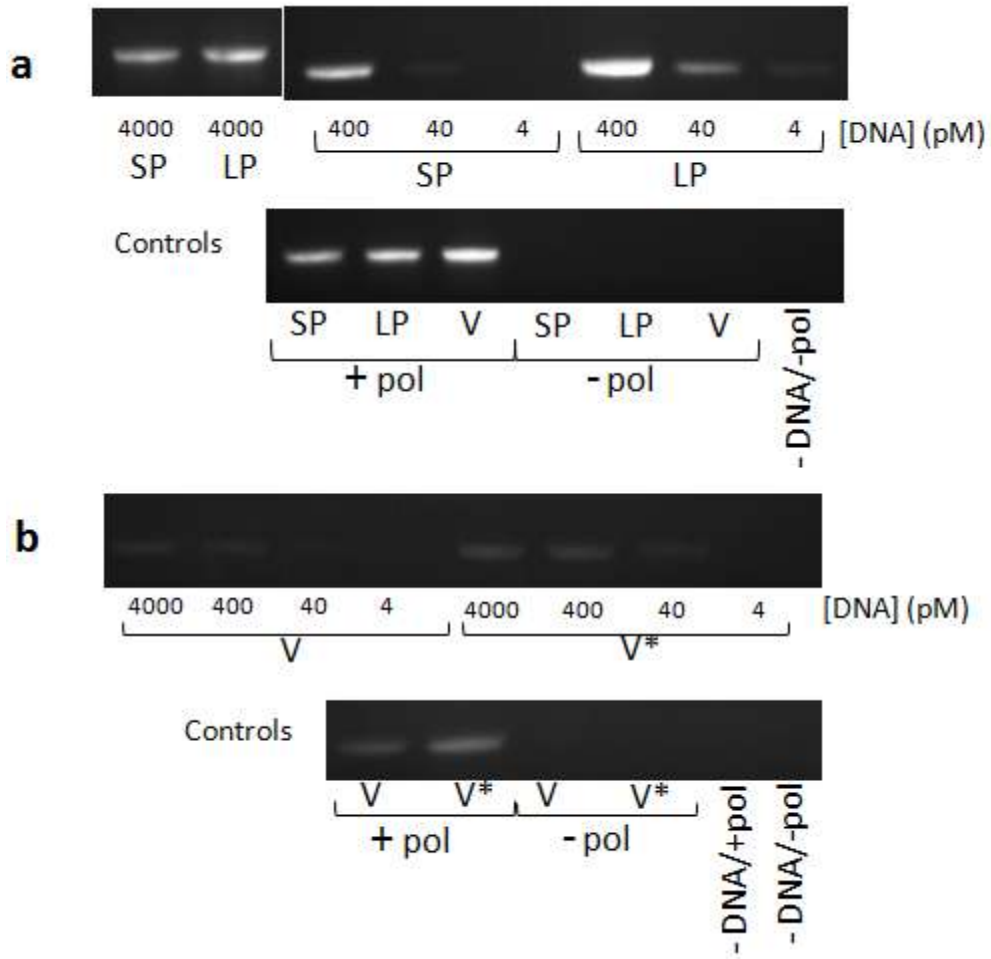

Figure 4.4 The number of binding sites matters for binding Ubx fibers spanning 4 orders of magnitude – non-linear structures do not preclude binding (raw gels) a, SPwt and a linearized version, LPwt were detected by PCR of the *vegf* gene present in all samples. The concentration of DNA used in each binding reaction is listed below the gel image. Both the supercoiled and linear DNA sequences bound Ubx fibers to similar extents at DNA concentrations spanning four orders of magnitude. Included are PCR reactions of positive controls using template DNA in solution (+) and negative controls lacking polymerase (no pol), DNA (no DNA) or both. **b,** significant differences in **Figure 4.4 continued,** binding were not observed at high (saturating) concentrations of DNA, LVmut DNA binds better than the LVwt DNA sequence at lower DNA concentrations. Positive and negative controls are labeled as above. All four DNA sequences bind Ubx fibers.

Interactions between DNA and Ubx materials 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, $t_{1/2}$ 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 (SPwt). After repeated rinsing, the fibers were incubated in a large volume of buffer for up to nine days to approximate infinite dilution, preventing reassociation of released DNA. Following this incubation, we used PCR to detect any DNA remained bound to the Ubx fiber. Although the amount of bound DNA declined, some DNA remained bound to fibers for up to 1 week (Figure 4.5a, Figure 4.6), representing an extremely long half-life for a non-covalent, reversible binding reaction. 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.

Surprisingly, these experiments consistently yielded a more intense signal on Day 1 relative to Day 0 for supercoiled plasmid, an effect not observed for linearized plasmid. Since DNA isn't added to the fibers during the incubation period, the only remaining possibility is that supercoiled DNA is more easily detected on Day 1 than on Day 0 (Figure 4.5a , Figure 4.6). One explanation could be that we used PCR to detect the *vegf* gene, and that Hox sites are less abundant in this region than in the rest of the

gene (Figure 4.5b). 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 incubation 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 (Fig 4.5b), creating supercoiled SPmut and linearized LPmut. Unlike the wild-type plasmid, less of the supercoiled mutant plasmid was detected (Figure 4.5a, Figure 4.6), 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 phenomenon is similar to self-healing, a trait observed for other interactions involving multivalent, non-covalent bonds (Ahn et al. 2014).

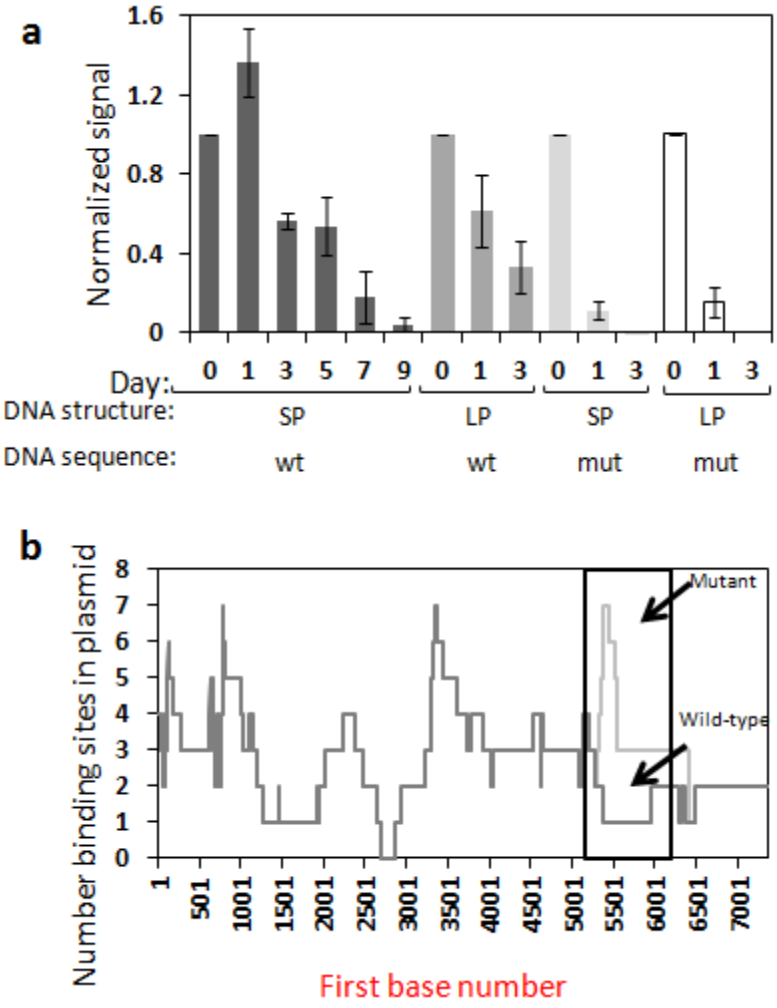

Figure 4.5 DNA is retained by Ubx fibers for several days, and DNA reorients on the fibers to maximize binding. a, Quantitation of multiple experiments with error bars indicating standard deviation. Ubx fibers, we incubated fibers in the presence of SPwt, LPwt, SPmut, or LPmut which contains 42 (wild 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 SPwt reorients between the 0 and 1 day measurements, but no reorientation was observed for LPwt, Spmut and LPmut 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.

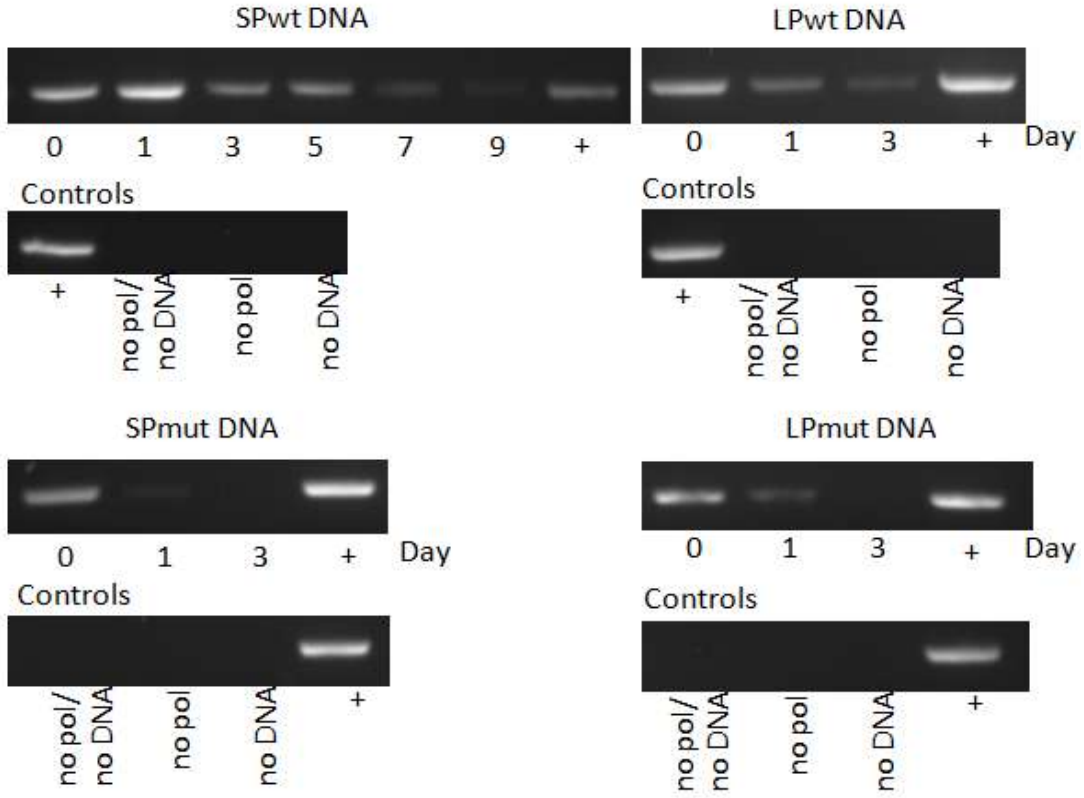

Figure 4.6 DNA is retained by Ubx fibers for several days, and DNA reorients on the fibers to maximize binding. Differences in binding are observed for SPwt and LPwt as well as SPmut and LPmut. Ubx fibers, we incubated fibers in the presence of SPwt, LPwt, SPmut or LPmut which contains 42 (wild type) or 48 (mutant) Ubx binding sites, respectively. PCR was used to detect the presence of DNA bound to Ubx fibers. DNA release was monitored over a period of 9 days for the SPwt sequence. LPwt DNA was monitored over a period of 3 days as well as the SPmut and LPmut DNA which contains the heavy cluster of binding sites inserted in the *vegf* gene detection site.

Although Ubx fibers are remarkably resistant to proteolysis (Patterson et al. 2015), nuclease digestion is a major concern with DNA-containing materials (de Vries et al. 2013). Since 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 LPwt to Ubx fibers and washing away excess DNA, we added the restriction enzyme DpnI. Digestion by DpnI will prevent DNA amplification by PCR. We amplified a 379 bp region that contains both DpnI restriction sites (five) and Ubx binding sites (four) (Tables 4.2 and 4.3). Free LPwt DNA was used in a positive control DpnI digestion reaction, which was designed to yield a similar PCR signal at 0 min in the absence of fibers. Although fiber binding did not prevent DNA digestion, DNA bound to fiber was digested slower than DNA in the positive control reaction (Figure 4.7, Figure 4.8).

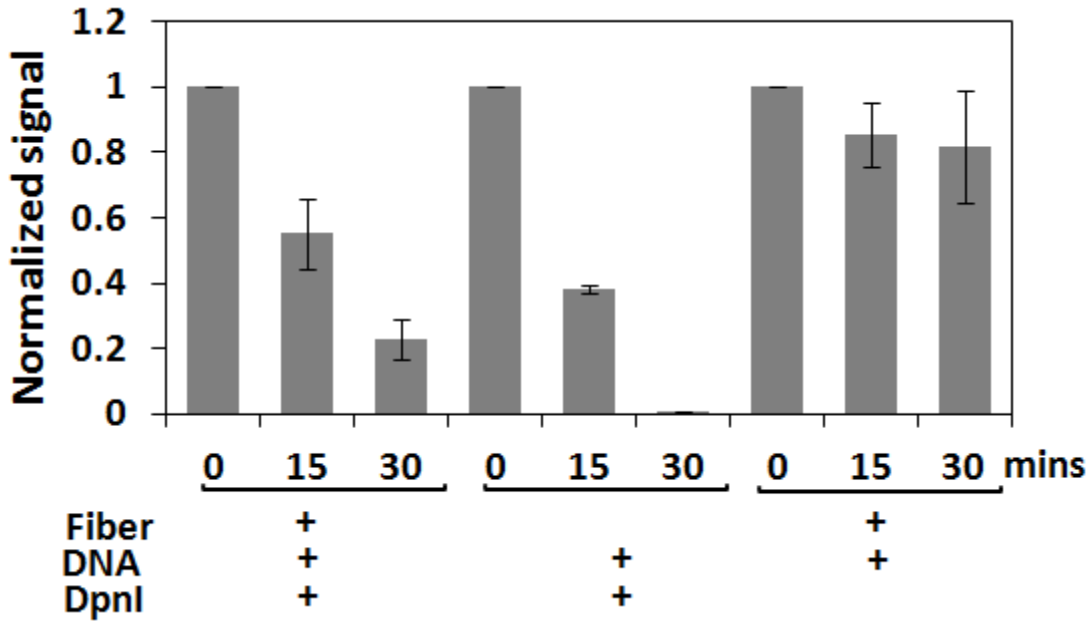

Figure 4.7 DNA is protected from degradation once bound to Ubx fibers. Ubx fibers were incubated in LPwt DNA. DNA, bound to Ubx fiber, was exposed to the restriction enzyme DpnI for varying times. The amount of DNA remaining intact was analyzed by PCR. Fiber binding protected DNA from the restriction enzyme to some extent.

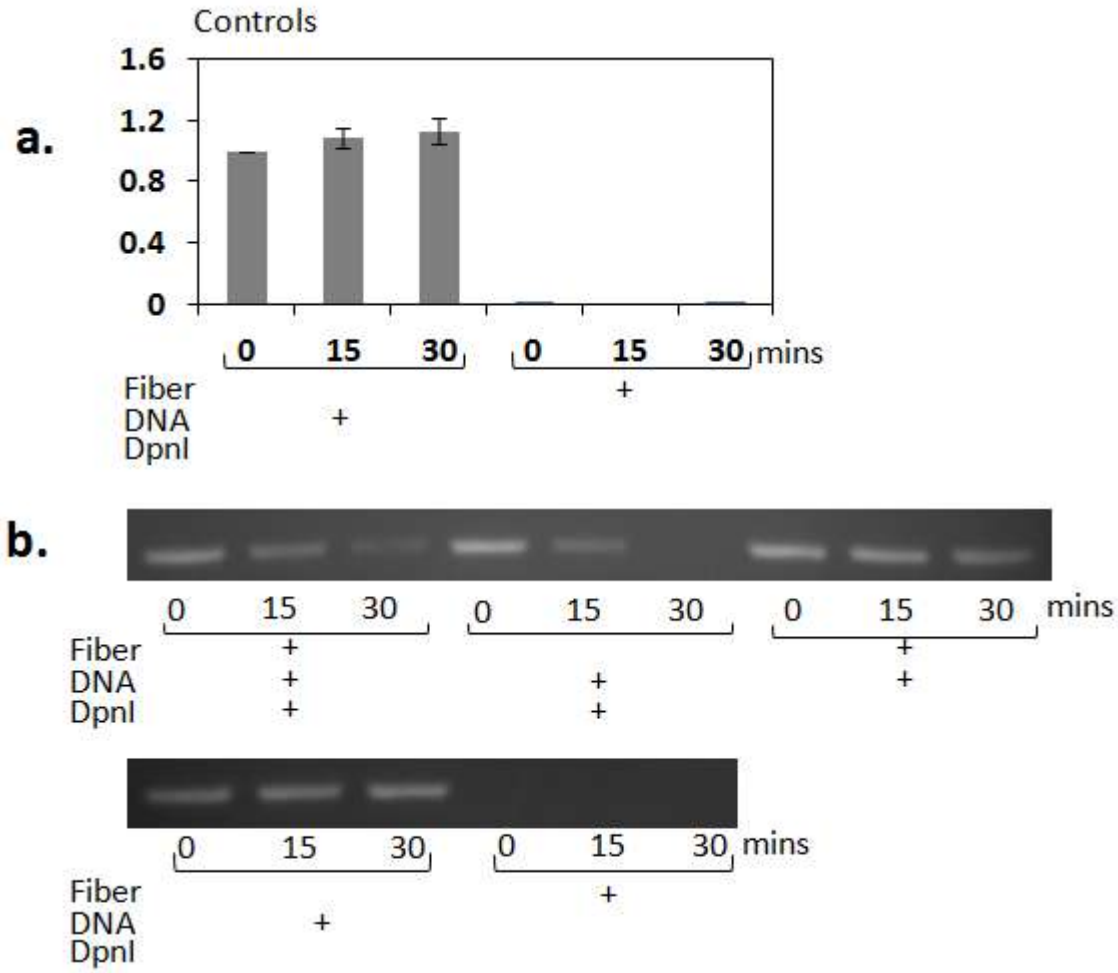

Figure 4.8 controls, a. Free LPwt DNA in the absence of fiber and DpnI restriction enzyme was not degraded over time as expected. Ubx fibers alone were not able to elicit any signal after PCR analysis thus confirming our results that the Ubx fiber is able to protect DNA from degradation somewhat. **b,** Ubx fibers incubated in LPwt, washed and subsequently placed in a reaction mixture containing DpnI restriction enzyme to monitor DNA degradation over time in the absence and presence of the Ubx fiber analyzed by PCR.

In all of the above experiments, DNA binding to the surface of pre-formed fibers was detected. Ubx, when genetically fused to proteins up to three-times its size, still self-assembles normally (Tsai et al. 2015), suggesting that Ubx monomers, pre-bound to DNA, might 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 Ubx assembly in the presence of DNA. Ubx bound to 40Sp DNA did self-assemble, incorporating DNA throughout the fiber (Figure 4.9). Therefore, the homeodomain is folded and functional throughout Ubx materials and not just on the fiber surface. Surprisingly, Ubx monomers were unable to self-assemble in the presence of 40NSp DNA (Figure 4.9). There are significant differences in the structure and dynamics of the homeodomain-DNA interface for specific *versus* non-specific DNA (Iwahara and Clore 2006, Vuzman et al. 2010, Frazee et al. 2002). We surmise that structural features, unique to the non-specific complex, preclude the Ubx-Ubx interactions that are vital to materials assembly. This hypothesis is supported by the fact that the homeodomain engages in covalent dityrosine bonds that substantially strengthen Ubx materials (Howell et al. 2015).

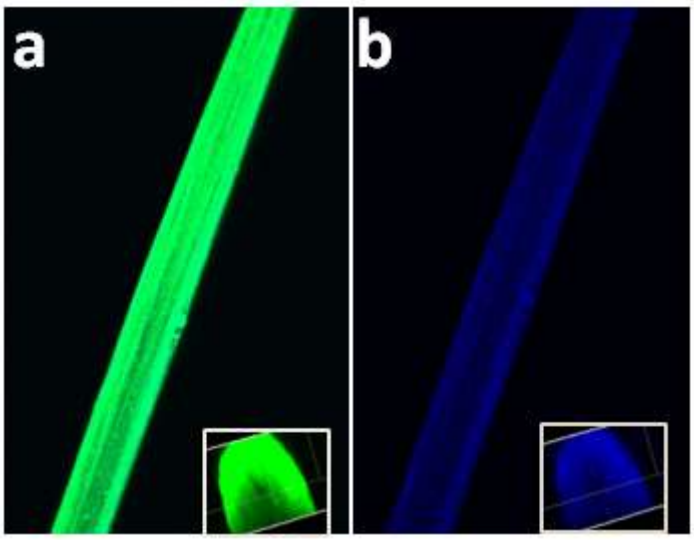

Figure 4.9 Ubx fibers self-assemble in the presence of 40Sp DNA, which contains 1 binding site but not 40NSp, which lacks binding sites. a, green fluorescence from the Alexa488 tag on 40Sp DNA in a fiber assembled from Ubx-40Sp dimers. Inset, a cross-section of the fiber demonstrating DNA is present throughout the fiber. **b,** Blue auto-fluorescence from the same fiber, with a cross-section in the inset.

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. These approaches allow DNA to be applied only to the 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.

5. CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

The Hox paradox is derived from two very contrasting points. *In vivo*, very specific and reliable function is required by Hox proteins. Misregulation of this process can lead to abnormalities or death in developing animals, or promote carcinogenesis and impede wound repair in adults. In contrast, *in vitro*, all Hox homeodomains recognize highly similar DNA sequences containing an 5'-TAAT-3' core, due to a very similar mode of DNA recognition by the homeodomain (Gehring et al. 1994). In addition, all Hox homeodomains display poor *in vitro* DNA-binding specificity. Despite more than 30 years of previous research on this important protein family, no molecular mechanisms have been developed that define how a Hox protein is able to regulate DNA binding at the molecular level.

A potential solution stems from regions of the Ubx sequence outside the homeodomain which establish long range regulatory interactions that influence DNA binding by the homeodomain. Previous studies on truncation mutants demonstrated that most of the Ubx sequence regulates DNA binding (Liu et al. 2008, Liu et al. 2009). Since these regions often play regulatory roles, *in vitro* characterization of the impact of these intrinsically disordered sequences is essential to understand protein function. These differences potentially provide a means to vary DNA recognition between different Hox members and thereby diversify Hox function. Regulation of interactions by cellular events (e.g. protein interactions, alternative splicing, phosphorylation) could allow a

single hox protein to function differently in different tissues. Since these nonhomeodomain regions differ between family members, it could potentially diversify of Hox function. Nonhomeodomain regions also diverge in evolution, potentially enabling orthologue-specific functions. The studies in this thesis create a structural model that can be used to understand the role of regions of Hox proteins outside the homeodomain: (i) what is the mechanism by how these long range interactions are occurring and (ii) which amino acids are responsible to establish this long range regulatory interaction.

Investigating how non-homeodomain regions interact with the rest of the protein has been hampered by the fact that the rest of the protein sequence is intrinsically disordered. Very little was known about protein structure or function outside the homeodomain and the prevalence of intrinsic disorder outside the homeodomain precludes structural studies. In addition, intrinsically disordered regions are prone to proteolysis and aggregation, making protein purification a challenging task (Churion et.al. 2016). Furthermore, mutagenesis in intrinsically disordered regions typically has little effect (Romero et al. 2001).

To overcome these issues, we developed a protocol to remove aggregating proteolytic products in minutes by filter flow-through purification. In this rapid technique, full-length protein passes through the filter and aggregated proteolytic products create particles larger than the pores and are thus retained by the filter (Churion et.al. 2016). We were also able to bypass targeted mutation issues by combining two observations: (i) dityrosine bonds were observed in materials and (ii) protein algorithm

programs revealed that the tyrosines observed in these interactions in materials were found in conserved motifs important for protein-protein interactions (Howell et.al. 2015).

We have generated a structural model of full-length Hox protein Ubx, in which conserved non-homeodomain tyrosines interact to bridge surface homeodomain tyrosines. In this model, the resulting line of tyrosines crosses the DNA binding helix. These long range interactions must be disrupted to enable DNA binding. Mutagenesis of homeodomain tyrosines removes these regulatory interactions, and causes the full-length, mutant Ubx to bind DNA with properties similar to DNA binding by the isolated homeodomain. Each tyrosine mutation has a different effect on DNA binding affinity which means that each conserved motif is important for the correct function of the cluster. In addition, we can hypothesize that each tyrosine contributes to the cluster by positioning each tyrosine in a specific manner.

Understanding the mechanism regulating binding by a full length Hox protein would help elucidate answers to the Hox paradox. Unlike the homeodomain, the nonhomeodomain regions that alter DNA binding have diverse amino acid sequences across this family. These differences provide a means to vary DNA recognition between different Hox members and thereby drive tissue-specific, Hox protein-specific, or orthologue-specific functions.

5.1.1 Tissue-specfic functions

The Ubx homeodomain binds to all DNA sites with similar, high affinities. A prior study from our lab establishes that monomer wild-type full length Ubx does not bind Hox-Exd composite sites unless cofactor Exd is present, yet Ubx is able to bind well to monomeric sites (Liu et al. 2009). Mutating the hexapeptide motif to GPGG allows Ubx to also bind the composite Ubx-Exd site as a monomer. Because Exd binds Ubx via the conserved hexapeptide motif, which contains essential tyrosine residue Y240, Liu and colleagues hypothesized that the hexapeptide motif inhibits binding to Hox-Exd composite sites and that Exd interaction removes this inhibition. However, the mechanism by which the hexapeptide motif inhibits binding and how Exd relieves this inhibition remained unknown. Our model provides a potential explanation.

In our model the mutation Y240L destabilizes the closed conformation and thus shifts the equilibrium towards the open protein conformation, accounting for its ability to bind Hox-Exd sites with high affinity, like homeodomain. On the other hand Y240L does not have a dramatic effect on the open conformation and thus it does not have a key role in forming the cluster in the open conformation. We can then hypothesize that Y240 is then forming less (or no) crucial contacts in the open conformation and is potentially found in the periphery of the cluster which makes it available to bind to Exd. Therefore Exd binding to Ubx favors the open state, and also enables Ubx to bind Hox-Exd composite sequences.

5.1.2 Ubx Hox orthologues in Hox evolution

A key question regarding animal evolution is how do orthologues of a single Hox protein vary their function to specify distinct body plans? The complex regulatory interactions observed for Ubx create the required intramolecular contacts stemming from conserved motifs, located in disordered regions, to the homeodomain. Although the corresponding amino acid sequences of Hox orthologues are surprisingly divergent, conservation in Ubx orthologues occurs at three levels: (1) the highly homologous homeodomain, (2) the less strictly conserved motifs, and (3) the preservation of disordered character but not amino acid sequence (Figure 3.1).

Tyrosine residues in conserved motifs are shifting relative to the original position in *Drosophila* or are lost throughout evolution, allowing for changes in intramolecular interactions which might diversify the body plan (Figure 3.1). This mechanism may preserve the necessary intramolecular contacts to establish a network of regulation and patterning in crucial core structures, such as the central nervous system, muscle, and gut, while still permitting evolution of new morphologies and body plans. Furthermore, the DNA binding regulatory regions overlap domains associated with other Hox functions (Figure 5.1). Thus, these regulatory mechanisms potentially coordinate DNA binding with protein interactions or transcription regulation. This creates further opportunities to diversify DNA binding by orthologues of Ubx Hox proteins.

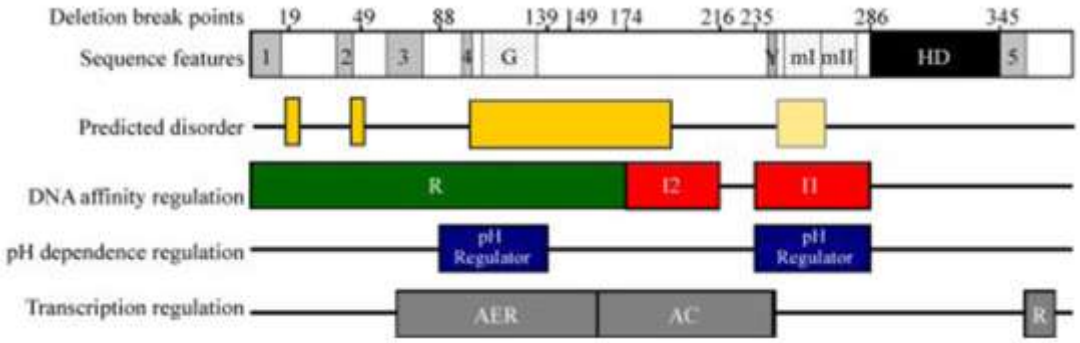

Figure 5.1 Overlap between intrinsically disordered regions and functional domains within UbxIa. The conserved motifs (1–5), polyglycine region (G),YPWM motif (Y), and homeodomain (HD) are indicated. Regions of UbxIa predicted to be disordered are indicated by dark yellow boxes. The light yellow box is also disordered in UbxIb and predicted to also be disordered in UbxIa. All three DNA binding regulatory regions contain disordered sequences. High affinity DNA binding by the homeodomain can be directly inhibited by amino acids 235–286 (I1; red), which includes the conserved YPWM motif and Ubx microexons. The I2 inhibitory region (amino acids 174–216; red) inhibits DNA binding via conformational fluctuations in the absence of the R region (amino acids 1–174; green), which restores binding in a length dependent manner. The I1 region and amino acids 88–139 (blue) regulate the pH dependence of Ubx protein-DNA binding, demonstrating that they can directly impact the pKa of homeodomain residues important for binding and influence access to this region in a pH-dependent manner. Most of the activation domain (activation enhancement region (AER) and activation core (AC)) is encompassed within the R and I2 regions, suggesting that DNA binding and transcription regulation may be functionally linked via overlapping domains. Reprinted with permission from Liu et al. 2008. "This research was originally published in Journal of Biological Chemistry. Liu Y, Matthews KS, Bondos SE. Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila hox protein ultrabithorax. Journal of Biological Chemistry. 2008; 30:20874-87. © the American Society for Biochemistry and Molecular Biology."

5.1.3 Hox protein homologues

Our model is based on regulatory interactions between the Ubx structured homeodomain and conserved motifs embedded in intrinsically disordered regions in the remainder of the protein. This domain organization is present in all Hox proteins in

Drosophila melanogaster. Even though all features are conserved (as mentioned in orthologues above), some key differences between homologues are present.

One such difference is that anterior Hox proteins devote a bigger percentage of their sequence to forming motifs (Figure 5.2). Therefore, a higher number of motifs are able to interact and form an inhibitory cluster on the homeodomain. Conversely, the intervening loops must be shorter, but more numerous, potentially allowing more protein-protein interactions to form. We hypothesize that the anterior Hox proteins bind DNA with lower affinity, and are more reliant on protein interactions to relieve inhibition of DNA binding by their larger tyrosine clusters.

In contrast, posterior Hox proteins devote a lower percentage of their sequence to forming motifs (Figure 5.2); smaller regulatory cluster is able to form on the homeodomain surface, and longer intervening loops are established. This provides an opportunity for stronger DNA binding and less protein-protein interactions to form (relative to anterior proteins). In both cases, different regulatory interactions potentially determine the extent of protein-protein interactions and Hox-DNA interactions to diversify Hox function. This mechanism provides an opportunity to explore different regulatory mechanisms while preserving essential contacts in order to reach different specificity for DNA sequences in the organism. Our model potentially provides an explanation for diversification of Hox function within the protein family and accounts for differences in DNA binding by the isolated homeodomain versus the full-length protein. This mechanism also provides a partial solution to the Hox paradox which

highlights the contrast between the similarity of DNA binding by the Hox homeodomains versus the diverse and specific Hox functions *in vivo*.

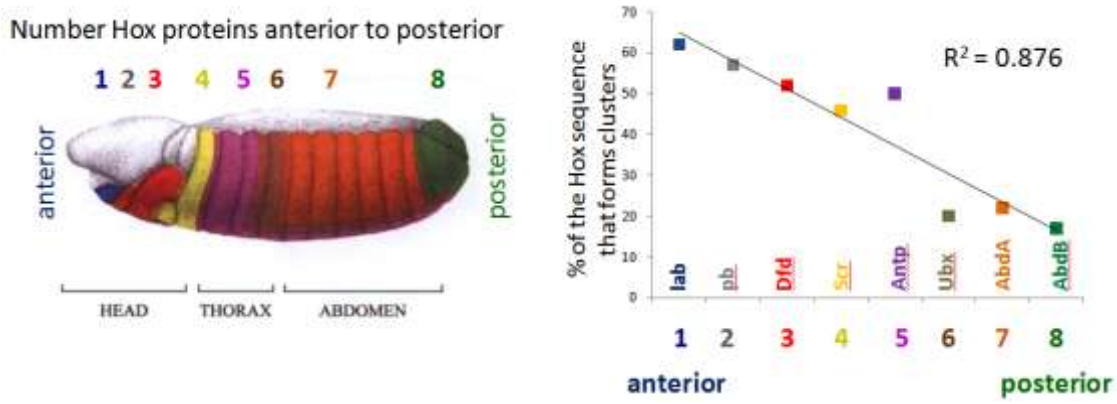

Figure 5.2 Plot of the percent of *Drosophila* Hox protein sequences composed of ANCHOR-identified motifs versus the position of Hox expression in the embryo (left) from anterior to posterior (colored circles). In *Drosophila* this single gene cluster is divided into two clusters (Antennapedia cluster and Bithorax cluster), with the breakpoint between the antp and ubx genes. Plotting the percentage of non-homeodomain sequence predicted by ANCHOR to contain interactive motifs with aromatic residues versus the order in which Hox proteins are expressed in vivo reveals a surprisingly straight line and strong correlation.

5.1.4 Hox role in cancer

Hox mis-regulation contributes to all stages of tumor progression, including stem cell proliferation, oncogenic transformation, stimulation of angiogenesis, and metastasis. Homeodomain mutations potentially disrupt DNA binding and, as a consequence, disrupt normal biological effects. Previously there was no hypothesis of what mutations outside of the homeodomain were “doing” to disrupt the normal biological effect.

I hypothesize that mutations in cancer disrupt long range regulatory interactions. Alignment of mutants isolated from tumors revealed that tyrosine mutations associated with cancer are located in conserved motifs predicted to be involved in protein-protein or even intramolecular interactions (Figure 5.3). If protein monomer intramolecular interactions are affected, then DNA binding misregulation occurs which then leads to disrupted interactions and downstream biological effects.

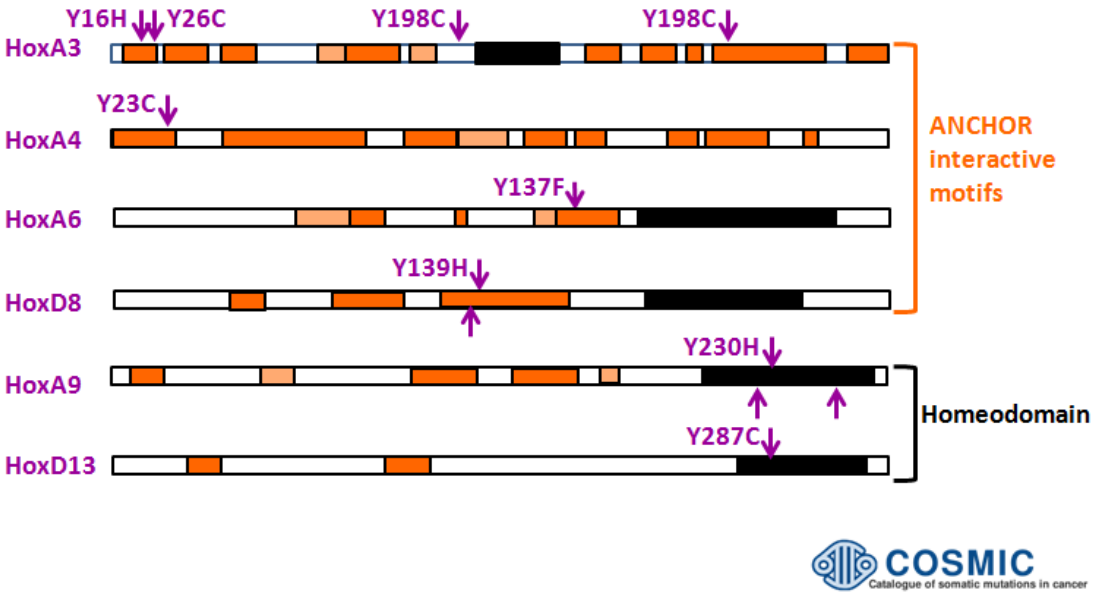

Figure 5.3 Alignment of tyrosine mutants reveals that mutations are present in ANCHOR positive regions of Hox proteins in cancer. Sequences were obtained from the COSMIC database (online database of somatically acquired mutations found in human cancer) and aligned. Alignment revealed that for some cancers, tyrosine residues were mutated outside the homeodomain (ANCHOR interactive motifs- orange text top four sequence alignments) and in the homeodomain (black text-bottom two sequence alignments), potentially disrupting regulatory interactions. Purple arrows indicate location of tyrosine mutations and purple text above each sequence schematic reveals the identity of the tyrosine mutation. The Hox protein affected is listed in purple text to the left.

5.1.5 Ubx materials

Ubx materials form via intermolecular interactions similar to the intramolecular interactions. Understanding monomer intramolecular interactions can then: allow us to develop better intermolecular bonds to engineer stronger/more flexible materials or even improve the stability of Ubx materials by potentially being able to add
These interacting conserved sequences to loops or unstructured regions of Ubx and thus add specific dityrosine bonds to increase the strength Ubx materials (Howell et al. 2015). A deep understanding on how the cluster forms and regulates Ubx at the monomeric level would then allow to build materials with specific properties and diverse features.

5.2 Future directions

Intrinsic conformational dynamics of proteins have been suggested to play crucial roles in ligand binding and dissociation (Vuzman and Levy 2012). The isolated UbxHD is by far the tightest monomeric unit in the literature binding at K_d = 60 pM (Bloom and Zamble 2004, Jana et al. 1998, Nalefski et al. 1996, Moll et al. 2002, Moore et al. 2001, Moorefield et al. 2004, Swinger and Rice 2007, Wilson et al. 2007). However the Y293L+Y296L mutant bind the same DNA sequence with a K_d = 28 pM (Table 3.2, Figure 3.5). Likewise the K_d for UbxIVa binding to DNA is 30pM (Table 1 in Appendix). There are no alternative DNA binding domains or even clusters of positively charged residues in Ubx outside of the homeodomain that could be used to enhance binding, therefore the difference in DNA binding affinity observed is still unclear.

There are two possible options why we observe a difference in DNA binding affinity by the full length Ubx Y293L+Y296L mutant: (1) the homeodomain (HD)-DNA effect, where: (1a) tyrosine mutations are directly changing the HD-DNA interface, (1b) tyrosine mutations do not directly affect the HD-DNA interface but indirectly improve HD binding by increasing favorable contacts to DNA, or (1c) tyrosine mutations are changing HD-DNA energetics. The second option is the: (2) Homeoodomain(HD)-Intrinsically disordered regions (IDR) effect, where: (2a) tyrosine mutations are directly changing the intraprotein (HD-IDR) interface /structure of the protein, which means tyrosine mutation are directly having an effect on IDR regulatory interactions to the HD, or (2b) tyrosine mutations are changing intraprotein (HD-IDR) dynamics.

In order to investigate these effects I propose we begin by determining the dissociation constant (K_d) for the isolated UbxHD wildtype and the isolated UbxHD Y293L+Y296L on the optimal DNA sequence 40AB (Table 3.1). If the result is that the isolated UbxHD Y293L+Y296L mutant has a greater DNA binding affinity (lower K_d) than isolated UbxHD wild type, and similar DNA binding affinity to full length Ubx Y293L+Y296L mutant, I would conclude that tyrosine mutations in the homeodomain changed HD-DNA interaction, confirming option 1. In contrast, if the result is that the isolated UbxHD Y293L+Y296L mutant has the same relative DNA binding affinity (close to the same Kd) than the isolated UbxHD wild type, I would conclude that the tyrosine mutations in the homeodomain changed the regulatory interactions between HD and the IDR in the full-length Ubx Y293L+Y296L mutant , confirming option 2.

There is evidence to expect option 2 over option 1. From the crystal structure in Passner 1999 (Passner et al. 1999), tyrosine residue 296 does not contact DNA, and tyrosine residue 293 contacts DNA backbone by establishing a hydrogen bond to the DNA phosphate backbone. A mutation in the contacting tyrosine residue 293 should then decrease the DNA binding affinity instead of increasing it. Furthermore stability studies of the HD (Shukla et al. 2012, Subramaniam et al. 2001, Torrado et al. 2009) demonstrate that the HD is a very stable unit. In addition, due to this stability, it would be very difficult for tyrosine surface mutations to destabilize the core and perturb the DNA binding energetics and high DNA binding affinity (Liu et al. 2008, Liu et al. 2009).

To test the possibility of option 2, I propose to first examine DNA binding affinity with tyrosine→serine mutants in Ubx full length protein and isolated UbxHD to corroborate the results are not an effect of the mutation to leucine. If the result is different then we can attribute this effect to the amino acid being substituted. If the results are the same as the leucine mutations then I would conclude that the tyrosine mutations in the homeodomain are altering HD-IDR interactions in the full-length Ubx Y293L+Y296L mutant, further confirmation of option 2. Another possibility for option 2 is that the tyrosine mutations in the homeodomain (HD) created/destroyed secondary structure elsewhere in the protein. To test this possibility I propose circular dichroism experiments for full length Ubx mutants and isolated UbxHD mutants as well as their respective wild type counterparts. Previously published data (Liu et al. 2008) revealed no structure outside the HD therefore it is expected that mutations do not create/destroy

secondary structure elsewhere in the protein and the effect in affinity is due to other factors. Furthermore, a change in structure for these mutants could also be experimentally proven by fluorescence emission spectra and native state proteolysis (as done in Section 3). Although these experiments could assess why there was a difference in DNA binding affinity, what remains to be investigated is: why is binding of the full-length Y293L+Y296L mutant better than that of the isolated UbxHD?

HOX proteins have a single DNA binding domain, the homeodomain, which consists of a flexible N-terminal arm and three α-helices. When bound to DNA the third helix, makes a number of DNA contacts in the major groove. The intrinsically disordered N-terminal arm extends from the packed helical structure and lies along the minor groove. In a free homeodomain, the N-terminal arm is highly dynamic (Qian et al. 1992, Qian et al. 1994). The high entropic cost of keeping the N-terminal arm still while it is bound to DNA affects protein association energetic and thus the dissociation constant of the protein.

In order to investigate this question further, I will focus on the N-terminal arm dynamics of the homeodomain. One possible answer is that the N-terminal arm dynamics in the isolated UbxHD is artificially high because it is the N-terminus of the protein. Consequently, the requirement to hold the N-terminal arm still to establish the appropriate contacts to DNA has a high entropic/energetic cost, therefore, an effect on the association energetic and dissociation constant of the protein is observed. In contrast, the full-length Ubx Y293L+Y296L mutant has the same N-terminal arm but is tethered to the rest of the protein. When bound to DNA, the N-terminal arm establishes contacts

to the DNA minor groove but in the case of the full-length Ubx mutant, the N-terminal arm is less dynamic because it must move the rest of protein through solvent and thus it can establish more stable contacts with a lower entropic/energetic cost. To test this possibility I propose to create two UbxHD fusions that consist of:

a. UbxHD-RNase (structured fused protein) (Figure 5.4)

b. UbxHD-p53 (Intrinsically disoredered fused protein) (Figure 5.4)

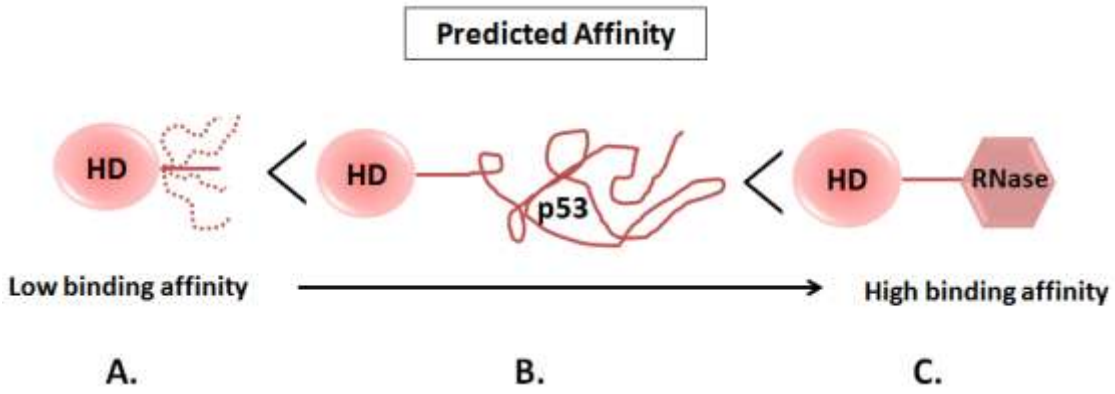

Figure 5.4 UbxHD fusions to assess the role of the N-terminal arm in DNA binding energetics. A) Isolated UbxHD:expect lowest affinity because N-term arm is untethered and highly dynamic (indicated with dashed lines). **B)** p53-HD: expect middle affinity because the N terminal arm is less dynamic because it must move the rest of protein through the solvent but the rest of the protein is intrinsically disordered. **C)** RNase-HD: expect highest affinity because N-terminal arm is less dynamic due to structured protein fused and the high energetics it takes to move it through solvent.

The first Ubx fusion would consist of UbxHD-RNase where highest affinity is expected because N-terminal arm is less dynamic due to structured protein fused and the high energetics it takes to move it through solvent. The second Ubx fusion UbxHD-p53

a DNA binding affinity that falls between the UbxHD-RNase fusion and the isolated Ubx homeodomain (UbxHD), because the N terminal arm is less dynamic because it must move the rest of the protein through solvent but rest of protein is intrinsically disordered. The isolated Ubx homeodomain UbxHD is expected to have lowest affinity because the N-terminal arm is untethered and highly dynamic. Therefore, when comparing these fusions with the isolated UbxHD we can assess the role of the N-terminal arm in DNA binding. These experiments have the potential to define a second novel Hox regulatory mechanism. Furthermore, this mechanism could be exploited by the cell to regulate DNA binding affinity via alternative splicing

REFERENCES

A

Abate-Shen C (2002) Deregulated homeobox gene expression in cancer: cause or consequence? Nature Reviews Cancer 2, 777-785.

Ades SE, Sauer RT (1995) Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. Biochemistry 34(44):14601-8.

Agrawal P, Habib F, Yelagandula R, Shashidhara LS (2011) Genome-level identification of targets of Hox protein Ultrabithorax in *Drosophila*: Novel mechanisms for target selection. Scientific Reports 1, Article number: 205.

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

Anjana R, Vaishnavi MK, Sherlin D, et al. (2012) Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: A study based on orientation and distance. Bioinformation 8(24):1220-1224.

Astbury WT, and Woods HJ (1934) X-Ray Studies of the structure of hair, wool, and related fibres. II. The molecular structure and elastic properties of hair keratin. Philosophical Transactions of the Royal Society A 232 333-394.

B

Balaban AT, Schleyer Pv, Rzepa HS (2005) Crocker, not Armit and Robinson, begat the six aromatic electrons. Chemical Reviews 105(10):3436-47.

Banerjee-Basu S, Moreland T, Hsu BJ, Trout KL, Baxevanis AD (2003) The Homeodomain Resource: 2003 update. Nucleic Acids Research 31(1):304-6.

Barash Y, Calarco JA, Gao W, et al. (2010) Deciphering the splicing code. Nature 465(7294):53-9.

Bateson W (1894) Materials for the study of variation. London, Macmillan 598.

Bawa P, Pillay V, Choonara YE, du Toit LC (2009) Stimuli-responsive polymers and their applications in drug delivery. Biomedical Materials 4(2):022001.

Beachy PA, Varkey J, Young KE, et al. (1993) Cooperative binding of an Ultrabithorax homeodomain protein to nearby and distant DNA sites. Molecular and Cellular Biology 13(11):6941-56.

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

Beloff A and Anfinsen CB (1948) The products of proteolysis of some purified proteins. Journal of Biological Chemistry 176(2):863-72.

Berndt KD, Güntert P, Orbons LP, Wüthrich K (1992) Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. Journal of Molecular Biology 227(3):757-75.

Bennett RL, Brown SJ, Denell RE (1999) Molecular and genetic analysis of the Tribolium Ultrabithorax ortholog, Ultrathorax. Development Genes and Evolution 209(10):608-19.

Berger MF, Badis G, Gehrke AR, et al. (2008) Variation in homeodomain DNA-binding revealed by high-resolution analysis of sequence preferences. Cell. 2008;133(7):1266-1276.

Berget SM, Moore C, Sharp PA (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proceedings of the National Academy of Sciences of the United States of America 74(8):3171-5.

Billeter M, Qian Y, Otting G, Müller M, Gehring WJ, Wüthrich K. (1990) Determination of the three-dimensional structure of the *Antennapedia* homeodomain from *Drosophila* in solution by nuclear magnetic resonance spectroscopy. Journal of Molecular Biology 214(1): 183-197.

Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. Annual Review of Biochemistry 72:291-336.

Bloom SL and Zamble DB (2004) Metal-selective DNA-binding response of Escherichia coli NikR. Biochemistry 43:10029-10038.

Bondos SE and Tan XX (2001) Combinatorial transcriptional regulation: the interaction of transcription factors and cell signaling molecules with homeodomain proteins in *Drosophila* development. Critical Reviews in Eukaryotic Gene Expression 11(1-3):145-71

Bondos SE and Bicknell A (2003) Detection and prevention of protein aggregation before, during, and after purification. Analytical Biochemistry 316(2):223-31.

Bondos SE, Catanese DJ Jr, Tan XX, et al. (2004) Hox transcription factor Ultrabithorax Ib physically and genetically interacts with disconnected interacting protein 1, a double-stranded RNA-binding protein. Journal of Biological Chemistry 279(25):26433-44.

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

Bondos SE and Hsiao HC (2012) Roles for intrinsic disorder and fuzziness in generating context-specific function in Ultrabithorax, a Hox transcription factor. Advances in Experimental Medicine and Biology 725:86-105.

Brennan RG and Matthews BW (1989) The helix-turn-helix DNA binding motif. Journal of Biological Chemistry 264(4):1903-6.

Briza P, Winkler G, Kalchhauser H, Breitenbach M (1986) Dityrosine is a prominent component of the yeast ascospore wall. A proof of its structure. Journal of Biological Chemistry. 261(9):4288-94.

Brocchieri L and Karlin S (1994) Geometry of interplanar residue contacts in protein structures. Proceedings of the National Academy of Sciences of the United States of 91 (20), 9297-9301.

Burnette JM, Hatton AR, Lopez AJ (1999) Trans-acting factors required for inclusion of regulated exons in the Ultrabithorax mRNAs of *Drosophila melanogaster*. Genetics 151(4):1517-1529.

Busseron E, Ruff Y, Moulin E, and Giuseppone N (2013) Supramolecular self-assemblies as functional nanomaterials. Nanoscale 5:7098-7140.

Busturia A, Vernos I, Macias A, Casanova J, Morata G (1990) Different forms of Ultrabithorax proteins generated by alternative splicing are functionally equivalent. EMBO Journal 9(11):3551-5.

C

Carrio M, Arderiu G, Myers C, Boudreau NJ (2005) Homeobox D10 induces phenotypic reversion of breast tumor cells in a three-dimensional culture model. Cancer Research 65:7177–85.

Chan SK and Mann RS (1993) The segment identity functions of UltraBithorax are contained within its homeodomain and carboxy-terminal sequences. Genes & Development 7(5): 796-811.

Chan SK, Jaffe L, Capovilla M, Botas J, Mann RS (1994) The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. Cell 78(4):603-15.

Chan SK, Mann RS (1996) A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. Proceedings of the National Academy of Sciences of the United States of America 93: 5223-5228.

Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO (2014) Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. Molecular Biology and Evolution 31(6):1402-13.

Chow LT, Gelinas RE, Broker TR, Roberts RJ (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell 12(1):1-8.

Chu MC, Selam FB, Taylor HS (2004) HOXA10 regulates p53 expression and matrigel invasion in human breast cancer cells. Cancer biology & therapy 3(6):568-72.

Chu SW, Noyes MB, Christensen RG, Pierce BG, Zhu LJ, Weng Z, Stormo GD, Wolfe SA (2012) Exploring the DNA-recognition potential of homeodomains. Genome Research (10):1889-98.

Churion K, Liu Y, Hsiao HC, Matthews KS, Bondos SE (2014) Measuring Hox-DNA binding by electrophoretic mobility shift analysis. Methods in Molecular Biology 1196:211-30.

Churion KA, Rogers RE, Bayless KJ, Bondos SE (2016) Separating full-length protein from aggregating proteolytic products using filter flow-through purification. Analytical Biochemistry 514:8-11.

Cillo C, Cantile M, Mortarini R, Barba P, Parmiani G, Anichini A (1996) Differential patterns of HOX gene expression are associated with specific Integrin and ICAM profiles in clonal populations isolated from a single human melanoma metastasis. International Journal of Cancer 66, 692-697.

Cillo C, Cantile M, Faiella A, Boncinelli E (2001) Homeobox genes in normal and malignant cells. Journal of Cellular Physiology 188, 161-169.

Cooper GM (2000) The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; regulation of transcription in eukaryotes. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9904/.

Cooper CDO, Newman JA, Aitkenhead H, Allerston CK, Gileadi O (2015) Structures of the Ets Protein DNA-binding Domains of Transcription Factors Etv1, Etv4, Etv5, and Fev: Determinants of DNA binding and redox regulation by disulfide bond formation. The Journal of Biological Chemistry 290(22), 13692–13709.

D

de Navas LF, Reed H, Akam M, 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.

Deming TJ (2007) Synthetic polypeptides for biomedical applications. Progress in Polymer Science 32: 858-875.

de Vries JW, Zhang F, and Herrmann A (2013) Drug delivery systems based on nucleic acid nanostructures. Journal of Controlled Release 172:467–483.

Duboule D, Dolle P (1989) The structural and functional organization of the murine HOX gene family resembles that of *Drosophila* homeotic genes. The EMBO Journal 8(5):1497–1505.

Dunker AK, Garner E, Guilliot S, et al. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pacific Symposium on Biocomputing 3:473-484.

Dunker AK1, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z (2002) Intrinsic disorder and protein function. Biochemistry 41(21):6573-82.

Dunker AK, Bondos SE, Huang F, Oldfield CJ (2015) Intrinsically disordered proteins and multicellular organisms. Seminars in Cell & Developmental Biology 37:44-55.

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

Dyson HJ and Wright PE (2005) Intrinsically unstructured proteins and their functions. Nature Reviews Molecular Cell Biology 6(3):197-208.

E

Ekker SC, Young KE, Kessler DPV and Beachy PA (1991) Optimal DNA sequence recognition by the Ultrabithorax homeodomain of *Drosophila*. EMBO Journal 10, 1179–1186.

Ekker SC, Jackson DG, von Kessler DP, et al. (1994) The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. The EMBO Journal 13(15):3551-3560.

Erkelenz M, Kuo CH, Niemeyer CM (2011) DNA-mediated assembly of cytochrome P450 BM3 subdomains. Journal of the American Chemical Society 133(40):16111-8.

F

Fogg JM, Kolmakova N, Rees I, et al. (2006) Exploring writhe in supercoiled minicircle DNA. Journal of physics Condensed matter : an Institute of Physics journal 18(14):S145-S159.

Fraenkel E and Pabo CO (1998) Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex. Nature Structural Biology 5(8):692-7.

Frazee RW, Taylor JA, and Tullius TD (2002) Interchange of DNA-binding modes in the Deformed and Ultrabithorax homeodomains: A structural role for the N-terminal arm. Journal of Molecular Biology 323(4):665-83.

Fuxreiter M, Simon I, Bondos S (2011) Dynamic protein-DNA recognition: beyond what can be seen. Trends in Biochemical Science 36(8):415-23.

G

Garner E, Cannon P, Romero P, Obradovic Z, Dunker AK (1998) Predicting disordered regions from amino acid sequence: common themes despite differing structural characterization. Genome Inform Ser Workshop Genome Inform. 9:201-213.

Garvie CW and Wolberger C (2001) Recognition of specific DNA sequences. Molecular Cell 8:937–946.

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

Gehring WJ, Hiromi Y (1986) Homeotic genes and the homeobox. Annual Review of Genetics 1986 20:147-73.

Gehring WJ, Qian YQ, Billeter M, et al. (1994) Homeodomain-DNA recognition. Cell 78(2):211-23.

Goh CS, Bogan AA, Joachimiak M, Walther D, Cohen FE (2000) Co-evolution of proteins with their interaction partners. Journal of Molecular Biology 299(2):283-93.

Graba Y, Aragnol D, Pradel J (1997) *Drosophila* Hox complex downstream targets and the function of homeotic genes. Bioessays 19(5):379–388.

Greer JM, Puetz J, Thomas KR, Capecchi MR (2000) Maintenance of functional equivalence during paralogous Hox gene evolution. Nature 403:661-664.

Greer AM, Huang Z, Oriakhi A, et al. (2009) The *Drosophila* transcription factor Ultrabithorax self-assembles into protein-based biomaterials with multiple morphologies. Biomacromolecules 10(4):829-37.

Grevellec J, Marquié C, Ferry L, Crespy A, Vialettes V (2001) Processability of cottonseed proteins into biodegradable materials. Biomacromolecules 2(4):1104-9.

Griffiths AJF, Gelbart WM, Miller JH, Lewontin RC (1999) Transcription: Gene regulation in eukaryotes-an overview. Modern genetic analysis. New York, NY: W. H. Freeman. Retrieved from www.ncbi.nlm.nih.gov/books/NBK21346/.

Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM (2000) Transcription: An overview of gene regulation in eukaryotes. In An introduction to genetic analysis (7th ed.) New York, NY: W. H. Freeman. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK21780/.

Gurley KA, Rink JC, Alvarado AS (2008) β-Catenin defines head versus tail identity during planarian regeneration and homeostasis. Science (New York, NY) 319(5861):323-327.

Gutmanas A, Billeter M (2004) Specific DNA recognition by the Antp homeodomain: MD simulations of specific and nonspecific complexes. Proteins 57(4):772-82.

H

Harper S and Speicher DW (2010) Purification of proteins fused to glutathione S-transferase. Methods in Molecular Biology 681:259-80

Harrison SC and Aggarwal AK (1990) DNA recognition by proteins with the helix-turn-helix motif. Annual Review of Biochemistry 59:933–969.

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

Howell DW, Tsai SP, Churion K, Patterson J, Abbey C, Atkinson JT, Porterpan D, You YH, Meissner KE, Bayless KJ, Bondos SE (2015) Identification of multiple dityrosine bonds in materials composed of the *Drosophila* protein Ultrabithorax. Advanced Functional Materials 25:5988-5998.

Hsiao HC, Gonzalez KL, Catanese DJ Jr, 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.

Huang Z, Lu Y, Majithia R, et al. (2010) Size dictates mechanical properties for protein fibers self-assembled by the *Drosophila* Hox transcription factor Ultrabithorax. Biomacromolecules 11(12):3644-51.

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

Hubbard S and Benyon R J (2001) in Proteolytic enzymes, 2nd Ed. (Benyon, R., and Bond, J. S., eds) 248-249, Oxford University Press, Oxford, UK.

Hudry B, Remacle S, Delfini MC, et al. (2012) Hox proteins display a common and ancestral ability to diversify their interaction mode with the PBC class cofactors. PLoS Biology 10(6):e1001351.

Hueber SD and Lohmann I (2008) Shaping segments: Hox gene function in the genomic age. BioEssays 30(10):965-79.

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

Hunt P, Gulisano M, Cook M, et al. (1991) A distinct Hox code for the branchial region of the vertebrate head. Nature 353(6347):861–864.

I

Iakoucheva LM, Radivojac P, Brown CJ, et al. (2004) The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Research 32(3):1037-49.

Iimura T, Pourquié O (2006) Collinear activation of Hoxb genes during gastrulation is linked to mesoderm cell ingression. Nature 442(7102):568-71.

Iimura T, Pourquié O (2007) Hox genes in time and space during vertebrate body formation. Development Growth and Differentiation 49(4):265–275.

Iwahara J and Clore GM (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. Nature 440, 1227-1230.

J

Jana R, Hazbun TR, Fields JD, Mossing MC (1998) Single-chain Lambda Cro repressors confirm high intrinsic dimer-DNA affinity. Biochemistry 37:6446-6455.

Jean L, Lee CF, Vaux DJ (2012) Enrichment of amyloidogenesis at an air-water interface. Biophysical Journal 102(5):1154-1162.

Jelesarov I and Bosshard HR (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. Journal of Molecular Recognition 12(1):3-18.

Johnson FB, Parker E, Krasnow MA (1995) Extradenticle protein is a selective cofactor for the *Drosophila* homeotics- role of the homeodomain and ypwm amino-acid motif in the interaction. Proceedings of the National Academy of Sciences of the United States of America 92: 739-743.

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

K

Kalionis B and O'Farrell PH (1993) A universal target sequence is bound in vitro by diverse homeodomains. Mechanisms of Development 43(1):57-70.

Kalodimos CG, Biris N, Bonvin AM, et al. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 305(5682):386-9.

Kannan N and Vishveshwara S (2000) Aromatic clusters: a determinant of thermal stability of thermophilic proteins. Protein Engineering, Design & Selection 13 (11): 753-761.

Kaufman TC, Lewis R, Wakimoto B (1980) Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: The homeotic gene complex in polytene chromosome interval 84a-B. Genetics 94(1): 115-133.

Kelly ZL, Michael A, Butler-Manuel S, Pandha HS, Morgan RGL (2011) Hox genes in ovarian cancer. Journal of Ovarian Research 4:16.

Kim Y, Coppey M, Grossman R, et al. (2010). MAPK substrate competition integrates patterning signals in the *Drosophila* embryo. Current Biology 20(5):446-51.

King NP and Lai Y (2013) Practical approaches to designing novel protein assemblies. Current Opinion in Structural Biology. 23:632–638.

Kissinger CR, Liu BS, Martin-Blanco E, Kornberg TB, 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.

Knoepfler PS and Kamps MP (1995) The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1.Molecular and Cellular Biology 15(10):5811-5819.

Kobelt D, Schleef M, Schmeer M, et al. (2013) Performance of high quality minicircle DNA for in vitro and in vivo gene transfer. Molecular Biotechnology 53(1):80-9.

Kornfeld K, Saint RB, Beachy PA, et al. (1989) Structure and expression of a family of Ultrabithorax mRNAs generated by alternative splicing and polyadenylation in *Drosophila*. Genes & Development 3(2):243-58.

Krumlauf R (1994) Hox genes in vertebrate development. Cell 78:191–201.

Kuziora MA and McGinnis W (1989) A homeodomain substitution changes the regulatory specificity of the Deformed protein in *Drosophila* embryos. Cell 59: 563–571.

Kyte J and Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 157(1):105-32.

L

Lanzarotti E, Biekofsky RR, Estrin DA, Marti MA, Turjanski AG (2011) Aromatic-aromatic interactions in proteins: beyond the dimer. Journal of Chemical Information and Modeling 51(7):1623-33.

Lappin TR, Grier DG, Thompson A, Halliday HL (2006) Hox genes: seductive science, mysterious mechanisms. The Ulster Medical Journal *75*(1), 23–31.

LaRonde-LeBlanc NA1, Wolberger C (2003) Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes & Development 17(16):2060-72.

Latchman, DS (1995) Eukaryotic transcription factors, 2nd Ed, Academic Press, London.

Lebendiker M and Danieli T (2014) Production of prone-to-aggregate proteins. FEBS Letters 588(2):236-46.

Lee GM, Pufall MA, Meeker CA, et al. (2008) The affinity of Ets-1 for DNA is modulated by phosphorylation through transient interactions of an unstructured region. Journal of Molecular Biology 382(4):1014-30.

Leff SE, Rosenfeld MG, Evans RM (1986) Complex transcriptional units: diversity in gene expression by alternative RNA processing. Annual Review of Biochemistry 55:1091-117.

Levine M and Hoey T (1988) Homeobox proteins as sequence-specific transcription factors. Cell 55: 537-540.

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

Liberles DA, Tisdell MDM, Grahnen JA (2011) Binding constraints on the evolution of enzymes and signalling proteins: the important role of negative pleiotropy. Proceedings of the Royal Society B: Biological Sciences 278(1714):1930-1935.

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

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

Lin L and McGinnis W (1992) Mapping functional specificity in the Dfd and Ubx homeodomains. Genes & Development 6(6):1071-1081.

Liu J, Perumal NB, Oldfield CJ, et al. (2006) Intrinsic disorder in transcription factors. Biochemistry 45(22):6873-88.

Liu Y, Matthews KS, Bondos SE (2008) Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the *Drosophila* Hox protein Ultrabithorax. Journal of Biological Chemistry 283(30):20874-87.

(B) Liu C-C, Richard AJ, Datta K, LiCata VJ (2008) Prevalence of temperature-dependent heat capacity changes in protein-DNA Interactions. Biophysical Journal 94(8):3258-3265.

Liu Y, Matthews KS, Bondos SE (2009) Internal regulatory interactions determine DNA binding specificity by a Hox transcription factor. Journal of Biological Chemistry 390(4):760-74.

Lodish H, Berk A, Zipursky SL, et al. (2000) Molecular mechanisms of eukaryotic transcriptional control. In Molecular cell biology *(*4th ed., section 10.7). New York, NY: W. H. Freeman. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK21677/.

Looger LL, Dwyer MA, Smith JJ, Hellinga HW (2003) Computational design of receptor and sensor proteins with novel functions. Nature 423(6936):185-90.

Lopez AJ, Hogness DS (1991) Immunochemical dissection of the Ultrabithorax homeoprotein family in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America 88(22):9924-8.

Lopez AJ, Artero RD, 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.

Lu Q, Knoepfler PS, Scheele J, Wright DD, Kamps MP (1995) Both Pbx1 and E2A-Pbx1 bind the DNA motif ATCAATCAA cooperatively with the products of multiple murine Hox genes, some of which are themselves oncogenes. Molecular and Cellular Biology 15(7):3786-95.

M

Mack JA, Abramson SR, Ben YX, et al. (2003) Hoxb13 knockout adult skin exhibits high levels of hyaluron and enhanced wound healing. FASEB Journal 17, 1352-1354.

Majithia R, Patterson J, Bondos SE, Meissner KE (2011) On the design of composite protein-quantum dot biomaterials via self-assembly. Biomacromolecules 12(10):3629-37.

Mann RS and Hogness DS (1990) Functional dissection of Ultrabithorax proteins in *D. melanogaster*. Cell 60: 597–610.

Mann RS (1995) The specificity of homeotic gene function. Bioessays 17(10):855-63.

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 and Carroll SB (2002) Molecular mechanisms of selector gene function and evolution. Current Opinion in Genetics & Development 12:592-600.

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

Martinez NJ, and Walhout AJM (2009) The interplay between transcription factors and microRNAs in genome-scale regulatory networks. BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology 31(4) 435-44.

Marx J (1992) Homeobox genes go evolutionary. Science 255: 399-401.
Maskarinec SA and Tirrell DA (2005) Protein engineering approaches to biomaterials design. Current Opinion in Biotechnology 16(4):422-6.

Matthews BW, Ohlendorf DH, Anderson WF, Takeda Y (1982) Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor. Proceedings of the National Academy of Sciences of the United States of America 79 (5): 1428-32.

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

McGinnis W, Garber RL, Wirz J, Kuroiwa A, Gehring WJ (1984) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. Cell 37(2): 403-408.

McGinnis W, Krumlauf R (1992) Homeobox genes and axial patterning. Cell 68:283-302.

Mendoza AD, Sebé-Pedrós A, Šestak MS, et al. (2013) Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proceedings of the National Academy of Sciences of the United States of America 110 (50) E4858-E4866.

Meyer A (1996) The evolution of body plans: Hom/Hox cluster evolution, model systems, and the importance of phylogeny. New uses for new phylogenies. P. H. Harvey, Brown, A.J.L., Smith, J.M. and Nee, S. Oxford, Oxford University Press, pp. 322-340.

Moens CB and Selleri L (2006) Hox cofactors in vertebrate development. Developmental Biology 291:193-206.

Moll JR, Acharya A, Gal J, Mir AA, Vinson C (2002) Magnesium is required for specific DNA binding of hte CREB B-ZIP domain. Nucleic Acids Research 30:1240-1246.

Moore M, Choo Y, Klug A (2001) Design of polyzinc finger peptides with structured linkers. PNAS 98:1432-1436.

Moorefield KS, Fry SJ, Horowitz JM (2004) Sp2 DNA binding activity and trans-activation are negatively regulated in mammalian cells. Journal of Biological Chemistry 279:13911-13924.

Morgan R, Pirard PM, Shears L, et al. (2007) Antagonism of Hox/Pbx dimer formation blocks the in vivo proliferation of melanoma. Cancer Research 67(12):5806-13.

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

Mukherjee S, Berger MF, Jona G, et al. (2004) Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nature Genetics 36(12):1331-9.

Müller M, Affolter M, Leupin W, et al. (1988) Isolation and sequence-specific DNA binding of the *Antennapedia* homeodomain. EMBO Journal **7**(13):4299-4304.

N

Nalefski EA, Nebelitsky E, Lloyd JA, Gullans SR (2006) Sing-molecule detection of transcription factor binding to DNA in real time: Specificity, equilibrium, and kinetic parameters. Biochemistry 45:13794-13806.

Naora H, Yang YQ, Montz FJ, et al. (2001) A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. Proceedings of the National Academy of Sciences of the United States of America 98(7):4060-5.

Neuteboom ST, Peltenburg LT, van Dijk MA, Murre C (1995) The hexapeptide LFPWMR in Hoxb-8 is required for cooperative DNA binding with Pbx1 and Pbx2 proteins. Proceedings of the National Academy of Sciences of the United States of America 92(20):9166-70.

Negre B, Casillas S, Suzanne M, et al. (2005) Conservation of regulatory sequences and gene expression patterns in the disintegrating *Drosophila* Hox gene complex. Genome Research 15(5):692-700.

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

Niklas KJ, Bondos SE, Dunker AK, Newman SA (2015) Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. Frontiers in Cell and Developmental Biology 3:8.

Noyes MB, Christensen RG, Wakabayashi A, et al. (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell 133(7):1277-89.

Numata K, Hamasaki J, Subramanian B, Kaplan DL (2010) Gene delivery mediated by recombinant silk proteins containing cationic and cell binding motifs. Journal of Controlled Release 146(1):136-43.

O

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

P

Pabo CO and Sauer RT (1992) Transcription factors: structural families and principles of DNA recognition. Annual Review of Biochemistry 61:1053-1095.

Pace CN, Scholtz JM, Grimsley GR (2014) Forces Stabilizing Proteins. FEBS letters 588(14):2177-2184.

Pallavi SK., Kannan R. Shashidhara LS (2006) Negative regulation of Egfr/Ras pathway by Ultrabithorax during haltere development in *Drosophila*. Developmental Biology. 296, 340–352.

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

Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK (1999) Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature 397(6721):714-9.

Patterson JL, Abbey CA, Bayless KJ, Bondos SE (2014) Materials composed of the *Drosophila melanogaster* protein Ultrabithorax are cytocompatible. Journal of Biomedical Materials Research Part A 102(1):97-104.

Patterson JL, Arenas-Gamboa AM, Wang TY, et al. (2015) Materials composed of the *Drosophila* Hox protein Ultrabithorax are biocompatible and nonimmunogenic. Journal of Biomedical Materials Research Part A 103(4):1546-53.

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

Petersen CP, Reddien PW (2008) Smed-βcatenin-1 is required for anteroposterior blastema polarity in planarian regeneration. Science 319:327-330.

Phelan ML, Rambaldi I, Featherstone MS (1995) Cooperative interactions between Hox and Pbx proteins mediated by a conserved peptidemotif. Molecular and Cellular Biology 15(8):3989-3997.

Piper DE, Batchelor AH, Chang CP, Cleary ML, Wolberger C (1999) Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. Cell 96(4):587-97.

Place ES, Evans ND, Stevens MM (2009) Complexity in biomaterials for tissue engineering. Nature Materials 8, 457-470.

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

Q

Qian YQ, Billeter M, Otting G, Müller M, Gehring WJ, Wüthrich K (1989) The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell 59:573-580.

Qian YQ, Otting G, Furukubo-Tokunaga K, Affolter M, Gehring WJ, Wüthrich K (1992) NMR structure determination reveals that the homeodomain is connected through a flexible linker to the main body in the *Drosophila* Antennapedia protein. Proceedings of the National Academy of Sciences of the United States of America 89(22):10738-10742.

Qian YQ, Resendez-Perez D, Gehring WJ, Wüthrich K (1994) The des(1-6)antennapedia homeodomain: comparison of the NMR solution structure and the DNA-binding affinity with the intact Antennapedia homeodomain. Proceedings of the National Academy of Sciences of the United States of America 91(9):4091-4095.

R

Raman V, Martensen SA, Reisman D, et al. (2000) Compromised HoxA5 function can limit p53 expression in human breast tumours. Nature 22;405(6789):974-8.

Rauch T, Wang Z, Zhang X, et al (2007) Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. Proceedings of the National Academy of Sciences of the United States of America 104:5527–32.

Reed HC, Hoare T, Thomsen S, et al. (2010) Alternative splicing modulates Ubx protein function in *Drosophila melanogaster*. Genetics 184(3):745-58.

Rodríguez-Cabello JC, Prieto S, Reguera J, Arias FJ, Ribeiro A (2007) Biofunctional design of elastin-like polymers for advanced applications in nanobiotechnology. Journal of Biomaterials Science, Polymer Edition 18(3):269-86.

Rohs R, Jin X, West SM, et al. (2010) Origins of specificity in protein-DNA recognition. Annual Review of Biochemistry 79:233-269.

Romano N, Sengupta D, Chung C, Heilshorn SC (2011) Protein-engineered biomaterials: Nanoscale mimics of the Extracellular Matrix. Biochimica et biophysica acta 1810(3):339-349.

Romero P, Obradovic Z, Kissinger, et al. (1998) Thousands of proteins likely to have long disordered regions. Pacific Symposium on Biocomputing 3:437-448.

Romero P, Obradovic Z, Li X, et al. (2001) Sequence complexity of disordered protein. Proteins: Structure, Function and Genetics 42(1):38-48.

Romero P, Obradovic Z, Dunker AK (2004) Natively disordered proteins: functions and predictions. Applied Bioinformatics 3(2-3):105-13.

Rosenberg JM, Seeman NC, Kim JJ, et al. (1973) Double helix at atomic resolution. Nature 243:150-154.

S

Sacca B and Niemeyer CM (2011) Functionalization of DNA nanostructures with proteins. Chemical Society Reviews 40, 5910-5921.

Saghatelian A, Guckian KM, Thayer DA, Ghadiri MR (2003) DNA detection and signal amplification via an engineered allosteric enzyme. Journal of the American Chemical Society 125(2):344-345.

Samuel S and Naora H (2005) Homeobox gene expression in cancer: insights from developmental regulation and deregulation. European Journal of Cancer (16):2428-37.

Sánchez AA (2000) Regeneration in the metazoans: why does it happen? BioEssays. 22:578–590.

Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Research 32(Database issue):D91-4.

Scott MP and Weiner AJ (1984) Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi

tarazu loci of *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America 81(13):4115-4119.

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

Scott MP (1992) Vertebrate homeobox gene nomenclature. Cell 71(4):551–553.

Seeman NC, Rosenberg JM, Rich A (1976) Sequence-specific recognition of double helical nucleic acids by proteins. Proceedings of the National Academy of Sciences of the United States of America 73:804–808.

Sengupta D and Heilshorn SC (2010) Protein-engineered biomaterials: highly tunable tissue engineering scaffolds. Tissue Engineering Part B: Reviews 16, 285-93.

Shaoqiang C, Yue Z, Yang L, et al. (2013) Expression of HoxD3 correlates with shorter survival in patients with invasive breast cancer. Clinical & Experimental Metastasis (30):155-163.

Shukla A, Burton NM, Jayaraman P-S, Gaston K (2012) The proline rich homeodomain protein PRH/Hhex forms stable oligomers that are highly resistant to denaturation. Defossez P-A, ed. PLoS ONE 7(4):e35984.

Singh GP and Dash D (2007) Intrinsic disorder in yeast transcriptional regulatory network. Proteins 68, 602-605.

Soshnikova N and Duboule D (2008) Epigenetic regulation of Hox gene activation: the waltz of methyls. Bioessays 30(3):199-202.

Subramaniam V, Jovin TM, Rivera-Pomar RV (2001) Aromatic amino acids are critical for stability of the Bicoid homeodomain. Journal of Biological Chemistry 276(24):21506-11.

Svingen T, Tonissen KF (2003) Altered Hox gene expression in human skin and breast cancer cells. Cancer Biology and Therapy 2, 518-523.

Swinger KK and Rice PA (2007) Structure-based analysis of HU-DNA binding. Journal of Molecular Biology 365:1005-1016.

T

Tan XX, Bondos S, Li L, Matthews KS (2002) Transcription activation by Ultrabithorax Ib protein requires a predicted alpha-helical region. Biochemistry 41(8):2774-85.

Taniguchi Y (2014) Hox Transcription Factors: modulators of cell-cell and cell-extracellular matrix adhesion. BioMed Research International 2014:591374.

Timasheff SN (2002) Protein hydration, thermodynamic binding, and preferential hydration. Biochemistry 41(46):13473-82.

Torrado M, Revuelta J, Gonzalez C, et al. (2009) Role of conserved salt bridges in homeodomain stability and DNA binding. The Journal of Biological Chemistry 284(35):23765-23779.

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

U

Ulijaszek SJ, Johnston FE, Preece MA (1998) The cambridge encyclopedia of human growth and development. Cambridge University Press: Cambridge.

Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Science: A Publication of the Protein Society 11(4):739-756.

Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. Journal of Molecular Recognition 18(5):343-84.

V

Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, Davey NE (2013) The switches. ELM resource: a compendium of conditional regulatory interaction interfaces. Science Signaling 6(269):rs7.

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. Nature Reviews Genetics 10, 252-263.

Velema J and Kaplan D (2006) Biopolymer-based biomaterials as scaffolds for tissue engineering. Journal of Biomaterials Science, Polymer Edition 102:187-238.

Venables JP, Tazi J, Juge F (2012) Regulated functional alternative splicing in *Drosophila*. Nucleic Acids Research 40(1):1-10.

Viswamitra MA, Kennard O, Jones PG, et al. (1978) DNA double helical fragment at atomic resolution. Nature 273:687-688.

Vivian JT and Callis PR (2001) Mechanisms of tryptophan fluorescence shifts in proteins. Biophysical Journal 80(5):2093-109.

Von Hippel PH and Berg OG (1986) On the specificity of DNA-protein interactions. Proceedings of the National Academy of Sciences of the United States of America 83(6):1608-1612.

Vuzman D, Azia A, Levy Y (2010) Searching DNA via a "monkey bar" mechanism: the significance of disordered tails. Journal of Molecular Biology 396(3):674-84.

Vuzman D and Levy Y (2012) Intrinsically disordered regions as affinity tuners in protein-DNA interactions. Molecular BioSystems 8(1):47-57.

W

Wang KC, Helms JA, Chang HY (2009) Regeneration, repair and remembering identity: the three Rs of Hox gene expression. Trends in Cell Biology, 19(6), 268-275.

Weinzierl ROJ (1999) Mechanisms of Gene Expression: structure, function and evolution of the basal transcriptional machinery. World Scientific Publishing Company. *ISBN 1-86094-126-5*.

Wellik DM and Capecchi MR (2003) Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. Science *3*01(5631):363–367.

White RAH and Wilcox M (1984) Protein products of the bithorax complex in *Drosophila*. Cell 39(1):163-71.

Wilkinson DG, Bhatt S, Cook M, Boncinelli D, Krumlauf R (1989) Segmental expression of Hox-2 homoeobox-containing genes in the developing mouse hindbrain. Nature 341(6241):405-409.

Wilson DS and Desplan C (1999) Structural basis of Hox specificity. Nature Structural & Molecular Biology 6(4):297-300.

Wilson CJ, Zhang H, Swint-Kruse L, Matthews KS (2007) Ligand interactions with lactose repressor protein and the repressor-operator complex: The effects of ionization and oligomerization on binding. Biophysical Chemistry 126:94-105.

Wintjens R, Rooman M (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. Journal of Molecular Biology 262 (2): 294-313.

Wolberger C, Vershon AK, Liu B, Johnson AD, Pabo CO (1991) Crystal structure of a MATalpha2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell 67(3):517-28.

Wright PE and Dyson, HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. Journal of Molecular Biology 293(2):321-31.

Wright PE and Dyson HJ (2015) Intrinsically disordered proteins in cellular signaling and regulation. Nature reviews Molecular cell biology 16(1):18-29.

Z

Zandvakili A, Gebelein B (2016) Mechanisms of specificity for Hox factor activity. Journal of developmental biology 4(2):16.

Zeng W, Andrew DJ, Mathies LD, Horner MA, Scott MP (1993). Ectopic expression and function of the Antp and Scr homeotic genes: the N terminus of the homeodomain is critical to functional specificity. Development 118(2): 339-352.

Zhang X, Zhu T, Chen Y, Mertani HC, Lee KO, Lobie PE (2003) Human growth hormone-regulated HoxA1 is a human mammary epithelial oncogene. Journal of Biological Chemistry 278(9):7580-90.

Zhao X, Li G, Liang S (2013) Several affinity tags commonly used in chromatographic purification. Journal of Analytical Methods in Chemistry 1-8.

APPENDIX

ALTERNATIVE SPLICING

1. Ubx is alternatively spliced

Ultrabithorax is a protein 389 amino acids long ranging in size due to alternative splicing that generates six different Ubx mRNAs (O'Connor et al. 1988; Kornfeld et al. 1989). The alternative splicing patterns of Ubx are evolutionarily conserved, suggesting these functional differences are important (Bennett et al. 1999). The six corresponding protein isoforms of Ubx (Lopez et al. 1996) (Figure 1.7) share common amino and carboxy terminal regions, but differ in internal sequence length according to the pattern of incorporation of three elements: the "B element" a sequence of 9 amino acids encoded between alternative donor splice sites at the end of the common 5' exon, and two small exons (mI and mII), each 17 amino acids long (O'Connor et al. 1988; Kornfeld et al. 1989). The Ubx homeodomain is located in the carboxy-terminal region and is identical for all isoforms, and it is separated from the differential elements by only four amino acid residues.

2. Alternative splicing potentially differentiates the sequence and function of Ubx proteins

As a consequence of stage and tissue specific alternative splicing, the *ubx* gene encodes a family of transcription factor isoforms that control segmental identity in the epidermis, mesoderm and nervous system of *Drosophila* (Burnette et al. 1999). The *ubx*

mRNA is alternatively spliced in a stage- and tissue-specific manner (Lopez et al. 1996). Ubx splicing isoforms exhibit functional differences *in vivo.* For example: within the central nervous system (CNS), different neurons express distinct ratios of *Ubx* isoforms (Burnette et al. 1999). Ubx isoforms differentially regulate dpp expression and muscle development (Venables et al. 2012). Alternative splicing impacts the length of a region known to alter DNA binding affinity and specificity (Liu et al. 2009). Therefore, alternative splicing has the potential to regulate DNA binding by Ubx in a tissue-specific manner (Mann and Hogness 1990).

3. Splicing changes the linker length between the hexapeptide motif and the homeodomain

The tyrosine residue Y240 found in the hexapeptide motif is essential in forming the closed conformation of Ubx. Altering the linker length between the hexapeptide and the homeodomain has the potential to alter DNA binding by changing the probability of hexapeptide:HD interaction. This biophysical feature may directly affect the ability of Ubx to bind target DNA by affecting the dynamics and energetics of the homeodomain N-terminal arm and/or by altering the positioning of Y240 in the cluster. We began investigating the role of alternative splicing in altering Ubx affinity for the 40AB optimal binding sequence. DNA binding affinity was measured for a subset of Ubx isoforms including: UbxIb, UbxIa, and UbxIVa (Figure 1.7). They range in amino acid length from longest to shortest, accordingly (Table 1 and Table 2). An interesting observation from our data indicates that the longer the linker length between the

hexapeptide motif and the homeodomain the lower the DNA binding affinity and the shortest isoform UbxIVa had the tightest affinity.

Isoform	**Hexapeptide-HD linker length (# amino acid)**	**DNA binding affinity to 40AB**
UbxIb	49	300 pM
UbxIa	40	160 pM
UbxIVa	6	30 pM

Table 1. Amino acid length from hexapeptide motif to the homeodomain for different Ubx isoforms and dissociation constant (K_d) for each.

Isoform	**Ubx isoform amino acid sequence**
UbxIb	MNSYFEQASGFYGHPHQATGMAMGSGGHHDQTASAAAAAYR GFPLSLGMSPYANHHLQRTTQDSPYDASITAACNKIYGDGAGAY KQDCLNIKADAVNGYKDIWNTGGSNGGGGGGGGGGGGAGG TGGAGNANGGNAANANGQNNPAGGMPVRPSACTPDSRVGGYL DTSGGSPVSHRGSAGGNVSVSGGNGNAGGVQSGVGVAGAGTA WNANCTISGAAAQTAAASSLHQASNHTFYPWMAIAGECPEDPT KSKIRSDLTQYGGISTDMGKRYSESLAGSLLPDWLGTNGLRRRG RQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWF QNRRMKLKKEIQAIKELNEQEKQAQAQKAAAAAAAAAAAVQGG HLDQ
UbxIa	MNSYFEQASGFYGHPHQATGMAMGSGGHHDQTASAAAAAYR GFPLSLGMSPYANHHLQRTTQDSPYDASITAACNKIYGDGAGAY KQDCLNIKADAVNGYKDIWNTGGSNGGGGGGGGGGGGAGG TGGAGNANGGNAANANGQNNPAGGMPVRPSACTPDSRVGGYL DTSGGSPVSHRG GSAGGNVSVSGGNGNAGGVQSGVGVAGAGTAWNANCTISGAA AQTAAASSLHQASNHTFYPWMAIAGKIRSDLTQYGGISTDMGK RYSESLAGSLLPDWLGTNGLRRRGRQTYTRYQTLELEKEFHTNH

	YLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQ EKQAQAQKAAAAAAAAAAVQGGHLDQ
UbxIVa	MNSYFEQASGFYGHPHQATGMAMGSGGHHDQTASAAAAAYR GFPLSLGMSPYANHHLQRTTQDSPYDASITAACNKIYGDGAGAY KQDCLNIKADAVNGYKDIWNTGGSNGGGGGGGGGGGGAGG TGGAGNANGGNAANANGQNNPAGGMPVRPSACTPDSRVGGYL DTSGGSPVSHRG GSAGGNVSVSGGNGNAGGVQSGVGVAGAGTAWNANCTISGAA AQTAAASSLHQASNHTFYPWMAIAGTNGLRRRGRQTYTRYQTL ELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKK EIQAIKELNEQEKQAQAQKAAAAAAAAAAVQGGHLDQ

Table 2. Amino acid sequence for Ubx isoforms tested is indicated. The amino acid sequence in between the hexapeptide motif to the homeodomain (linker region) for different Ubx isoforms is highlighted in yellow, and the microexon region is underlined (b element, microexon I, and microexon II).

3.1 The hexapeptide-HD linker length influences DNA binding affinity in different Ubx splicing isoforms

In Section 3, UbxIa that contained mutation Y293S+Y296S destabilized the closed conformation of the Ubx protein by affecting the positioning of crucial tyrosine residues and thus affecting the core formation of the tyrosine cluster. This mutation destabilized the closed conformation and shifted the equilibrium towards the open conformation protein. Furthermore, it caused Ubx to bind DNA at a very high affinity (20pM), tighter than that of the isolated homeodomain (UbxHD). We hypothesize that differences in the N-terminal arm dynamics generate this difference in DNA binding

affinity. The UbxHD has an untethered N-terminal arm, which is therefore free to be highly mobile. In contrast, the full length mutant UbxIa (Y293S+Y296S) is locked in the open conformation, so the homeodomain is free to bind DNA and the N-terminal arm is tethered to the rest of the protein, making it less dynamic. Consequently, less entropy is lost upon DNA binding, thus enhancing binding affinity.

Alternative splicing has previously been proposed to change the dynamics of both the hexapeptide-HD linker region and the N-terminal arm of the homeodomain (Hsiao et al. 2014). Therefore alternative splicing could alter DNA binding affinity by modulating the dynamics of the N-terminal arm. This model fits our data thus far: UbxIb, which has the longest and most dynamic linker, also binds with the poorest affinity (highest K_d). The linker in UbxIa is 9 amino acids shorter, and lacks the highly dynamic sequence encoded by the b element, and thus binds better than UbxIb. Finally, the linker in UbxIVa is substantially shorter and less dynamic therefore UbxIVa binds with a correspondingly high affinity. Our results suggest that splicing alters Ubx function, in part, by altering DNA binding affinity or specificity.