THE FIRST 3D STRUCTURE OF A FULL LENGTH HOX PROTEIN AND EVIDENCE FOR A CONFORMATIONAL CHANGE

A Thesis

by

SYDNEY TIPPELT

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

MASTER OF SCIENCE

Chair of Committee,	Sarah Bondos
Committee Members,	Margaret Glasner
	Raquel Sitcheran
	Jae-Hyun Cho
Head of Department,	David Threadgill

December 2021

Major Subject: Genetics

ABSTRACT

Hox proteins are an important class of transcription factors that regulate development of region-specific features in bilaterian animals via binding and regulation of target genes. These proteins contain a conserved DNA binding homeodomain (HD), generally near their C-terminus. Despite the large degree of HD conservation between members of the Hox family, each of these proteins oversee unique and specific functions *in vivo.* The observed discrepancy between specificity of function and similarity in DNA binding has been termed the Hox paradox, and as of yet has no concrete explanation. This work investigates the possibility that proteins modulate DNA binding via intramolecular interactions between N-terminal intrinsically disordered regions (IDRs) and the HD.

Here we show via negative stain transmission electron microscopy evidence that the Hox protein Ultrabithorax (Ubx) exists in multiple conformations. In these structures, the arrangement of the HD relative to the N-terminal disordered domains varies, and thus conformation may modulate HD-DNA binding. Because the N-terminus of Hox proteins is less conserved than the HD, it's possible that these disorder-HD interactions are different in each member of the Hox family, providing a mechanism by which these proteins can uniquely diversify their DNA binding.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Sarah Bondos, and my committee members, Dr. Margy Glasner, Dr. Raquel Sitcheran, and Dr. Jae-Hyun Cho, for their guidance and support throughout the course of this work. I'd also like the thank my lab members, both current and former, Dr. Rebecca Booth, Dr. Gabriela Mendes, Liliana Abrego, and Amanda Jons, for their help and support both in and out of the lab.

I would also like to thank my collaborator, Dr. Anindito Sen, for his insight and contributions, without which this research would not have been possible. His patience and guidance during my training are very much appreciated. I'd like to thank the staff and faculty of the Genetics program; their commitment to their students helped make my graduate experience the best it could be.

Additional thanks goes to my family, whose unconditional love and support made the completion of this research possible. My parents, Robert and Shari Tippelt, were always willing to lend me an ear and encourage me, and I'm very grateful for their support. My siblings, Sam Tippelt, Scott Tippelt, and Maddie Flora also provided much needed emotional support. I'd also like to thank my dog, Nadia, for inspiring me; she always puts on her brave face, no matter how terrifying the ceiling fan.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Sarah Bondos, who also acted as advisor, and Raquel Sitcheran of the Texas A&M Health Science Center, and Professors Margaret Glasner and Jae-Hyun Cho of the Texas A&M Biochemistry and Biophysics Department.

Results in chapter one were obtained in collaboration with Dr. Anindito Sen of the Texas A&M Microscopy and Imaging Center. Computational analysis for the generation of the 2D and 3D structures were performed by Dr. Sen, while collection of raw images and particle selection were performed by the student.

All other work for this thesis was completed independently by the student.

Funding Sources

Funding for this work was provided by a fellowship awarded to the student from the Texas A&M Office of Graduate and Professional Students.

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGEMENTS iii

CONTRIBUTORS AND FUNDING SOURCES iv

Contributors iv
Funding Sources iv

TABLE OF CONTENTS v

LIST OF FIGURES vii

CHAPTER I INTRODUCTION 1

Hox Proteins 1
Hox Paradox 6
Potential Explanations for the Hox Paradox 7
Chromatin Remodeling 7
Protein-Protein Interactions 8
Intramolecular Regulation 9
Ultrabithorax 10
Disordered Regions in Ubx Impact DNA Binding 12
Conformational Change Hypothesis 13
Investigating Ubx Structure 15

CHAPTER II 3D STRUCTURE OF A FULL LENGTH HOX PROTEIN AND EVIDENCE FOR MULTIPLE CONFORMATIONS 18

Introduction 18
Materials and Methods 19
Growth and Purification 19
Imaging and Analysis 20
Results 21
Micrograph Images Show Conformational Heterogeneity 21
3D Reconstructions Show Open and Closed Conformations 25
Domains in the N-terminal Disorder may Correspond to Self-Associating Regions 31

Intermediate States Seen in 2D Class Averages 32
Conclusion 33
Sequence Collapse in the Ubx IDRs 34
Confirmation of the Open and Closed States 36

CHAPTER III CONCLUSIONS 37

Implications for the Hox Paradox 38
Implications for Posterior Prevalence 38
Implications for IDP and Transcription Factor Organization 41

REFERENCES 43

APPENDIX A FULL LENGTH UBX PURIFICATION 54

1 Column Purification 54
2 Column Purification 55

APPENDIX B DEVELOPMENT OF THE 2-COLUMN PURIFICATION PROCEDURE 57

LIST OF FIGURES

Page

Figure 1 Hox genes in *Drosophila*. 3

Figure 2 Homeotic transformations in *Drosophila*. 5

Figure 3 Homeodomain sequence alignment between *Drosophila* Hox transcription factors. 7

Figure 4 Characterized domains of UbxIb. 12

Figure 5 Functional model for full length Ultrabithorax. 15

Figure 6 2D class averages show Ubx heterogeneity. 22

Figure 7 Experimental flow and final particle counts. 24

Figure 8 3D reconstruction of small and medium particles gives the open conformation. 26

Figure 9 Independent 2D class averages verify the 3D model. 28

Figure 10 3D reconstruction of large particles gives the closed conformation. 30

Figure 11 Sequence motifs in UbxIa. 32

Figure 12 Intermediate states seen in Ubx. 33

Figure 13 Posterior prevalence may be a function of protein conformation. 40

Figure 14 Coomasie stain of samples from phosphocellulose purification. 58

Figure 15 Western blot of samples collected during phosphocellulose purification. 59

Figure 16 Western blot of samples collected from different lysis conditions. 60

CHAPTER I

INTRODUCTION

Protein transcription factors (TFs) play essential roles in the regulation and expression of genetic information. While general TFs are utilized in transcription of all genes, specific TFs must recognize and bind the regulatory elements of only their cognate target genes (Latchman 1997).

Most TFs can be sorted into evolutionarily conserved families which share one or more DNA binding domains (DBDs) flanked by regions of intrinsic disorder. These binding domains are largely identical between family members, and yet each TF regulates distinct biological responses (Hollenhorst 2011). The mechanisms by which these proteins recognize their unique genetic targets using a standard DBD is unknown. In this thesis, we explore this problem in the Hox family of transcription factors.

Hox Proteins

Hox transcription factors regulate the development of segment-specific morphologies in bilaterian animals. Hox proteins act as master regulators to coordinate the development of organs and appendages unique to their expression domains, as well as specifying the formation of features spanning the length of an organism such as the digestive, skin, musculature, and nervous system (Pearson 2005). Hox proteins act as both initiators and regulators of gene networks (Chen 1999, Breau 2013, Prin 2014). Hox proteins were discovered in *Drosophila melanogaster*, which contains a single Hox

complement consisting of eight genes. Hox genes are expressed colinearly along the anterior-posterior axis of the organism, with each protein regulating the development of specific morphology within its domain of action (Fig. 1) (Hughes and Kaufman 2002). The single Hox complement makes *Drosophila* an ideal study system; most higher order vertebrates have multiple Hox complements due to gene duplication events (Wagner 2003, Wellik 2007). Consequently, study of Hox proteins in these higher order systems is significantly complex due to the necessity of treating multiple gene copies.

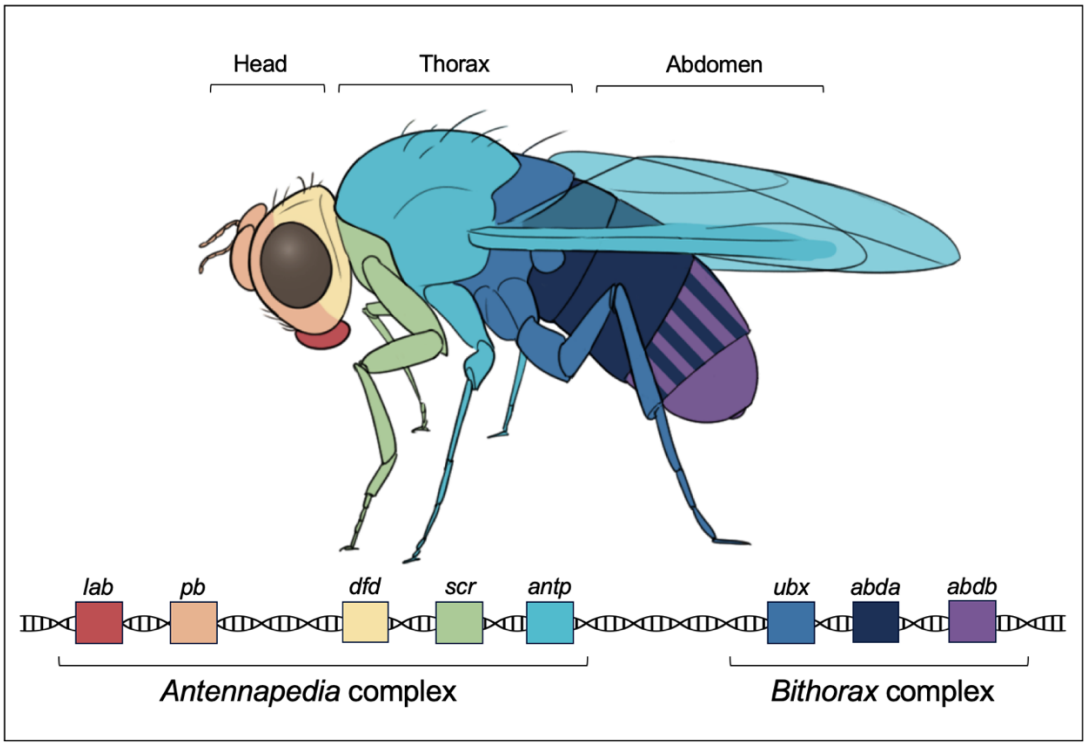

Figure 1 Hox genes in *Drosophila*.
The position of each Hox gene on the chromosome corresponds to the protein's domain of action along the anterior-posterior axis, and each protein regulates development of domain-specific morphology. Abbreviations: lab, labial; pb, proboscipedia; dfd, deformed; scr, sex-combs reduced; antp, antennapedia; ubx, ultrabithorax; abda, abdominal-A; abdb, abdominal-B.

Hox proteins are characterized by a conserved 60 amino acid motif near the C-terminus called the homeodomain (HD) (McGinnis 1984). This HD serves as the DNA binding region of the transcription factor. The HD is a helix-turn-helix motif consisting of three alpha helices and an unstructured N-terminal arm. The third helix interacts with the major groove and is most responsible for sequence recognition (Gehring 1994). Additional interactions between the minor groove and the HD N-terminal arm also

determine sequence recognition (Slattery 2011). A hexapeptide (HX) motif, which coordinates cofactor binding, is another common structural feature of this class of proteins, though a handful of Hox proteins such as AbdB lack this motif (Shanmugam 1997). Regions outside of the HD consist mainly of intrinsically disordered regions (IDRs) (Liu 2008). This has hindered further characterization of these proteins, as purified full-length Hox proteins are prone to proteolysis and aggregation.

Ectopic expression or deletion of Hox proteins *in vivo* results in disruption of morphology, and these mutations are termed homeotic transformations (Fig. 2). One example of such mutation occurs in *Drosophila* when a posterior Hox protein, such as Antp, is ectopically expressed in the head, transforming the antennae into legs (Denell 1981). Another classic homeotic transformation occurs when Ubx function is lost from the thorax. This loss of function causes duplication of the T2 features in the T3 segment, including the transformation of the halteres into a second set of wings (Lewis 1978). These mutations underscore both the importance of individual Hox protein contributions and the severe consequences of Hox protein disruption in an organism.

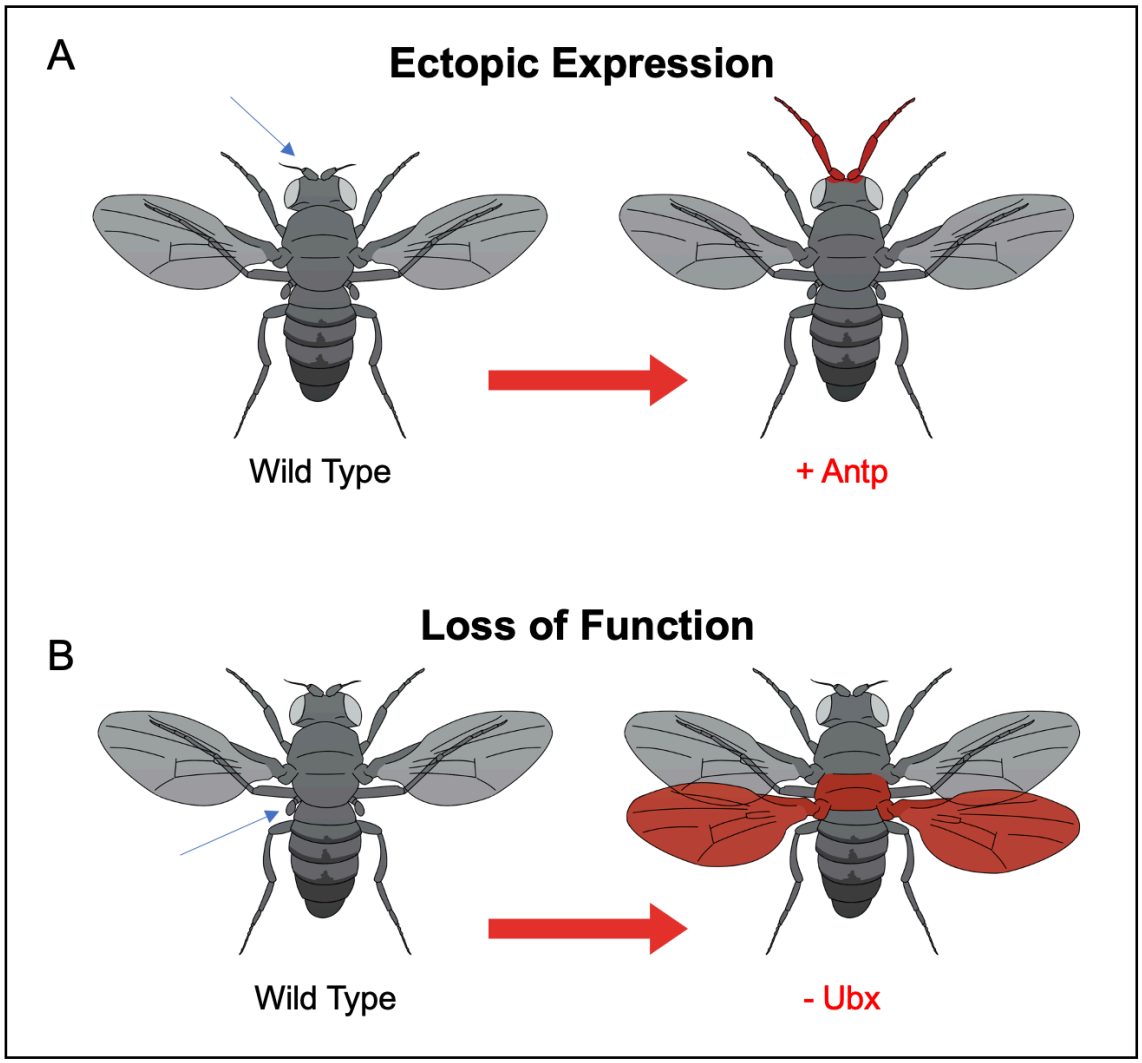

Figure 2 Homeotic transformations in *Drosophila*.
Hox misexpression results in transformations of segment morphology. The transformation on antennae into legs is caused by ectopic expression of Antp in the head (A). The bithorax mutation, in which a second set of wings develop in place of halteres, is caused by a loss of function of Ubx in the haltere imaginal disc during larval development, resulting in an expansion of the Antp domain and morphology into the third thoracic segment (B).

Hox Paradox

As previously noted, misexpression of individual Hox proteins has the capacity to drastically alter the body plan. As shown above in Figure 2B, complete loss of function of Ubx results in an expansion of the Antp domain into the posterior thorax; however, Antp is unable to recapitulate Ubx function in the T3 segment. The specificity of individual Hox contributions is also demonstrated in the development of the *Drosophila* heart. The posterior aorta and heart terminal chamber are specified by the Hox proteins Ubx and AbdA, respectively, and each of these Hox proteins regulate distinct target genes leading to the differentiation of these structures (Monier 2005). However, despite their specific functions *in vivo*, The DBD of Ubx and AbdA is over 90% identical, and all DNA contacting residues are conserved between the two HDs (Bondos 2001).

This degree of HD sequence conservation is not unique to Ubx and AbdA; all members of the Hox family show a large degree of conservation within their DNA binding region. Furthermore, residues that vary more don't contact DNA (Fig. 3). Consequently, HDs from different Hox proteins bind target DNA with similar affinity and specificity *in vitro* (Kalionis 1993). For example, Draganescu et al. (1995) demonstrated that Ubx and Dfd HDs share numerous conserved protein-DNA contacts and form analogous binding structures with target DNA.

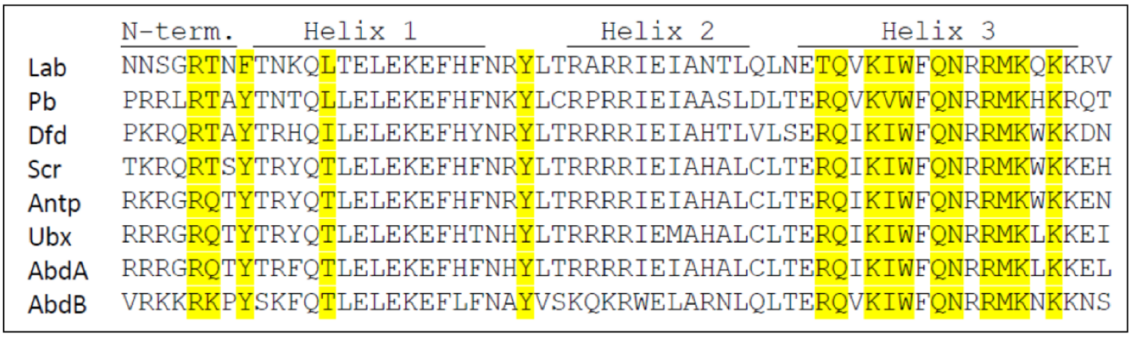

Figure 3 Homeodomain sequence alignment between *Drosophila* Hox transcription factors.
Hox HDs contain a large degree of amino acid sequence conservation, particularly in residues that contact DNA. Residues that directly contact DNA are highlighted in yellow. Alignment data from Passner et al., 1999.

This discrepancy between *in vivo* Hox specificity and similarity in HD binding behavior *in vitro* is known as the Hox paradox (Mann 1995). Currently, there is no known solution to the Hox paradox, although several potential explanations have been proposed, including chromatin remodeling, protein-protein interactions, and intramolecular interactions.

Potential Explanations for the Hox Paradox

Chromatin Remodeling

In most tissues, expression of each Hox protein is confined to specific regions along the anterior-posterior axis. Ectopic Hox expression in cell culture suggests that the chromatin structure within each region is distinctive, and thus local chromatin composition may dictate which target sequences are accessible for binding by Hox

proteins (Beh 2016). However, this explanation is inconsistent with observed homeotic transformants. In a Ubx deletion mutant, Antp expressed in the posterior thorax should be able to coordinate proper haltere development if target gene availability is the sole discriminator of Hox function. This is not seen, however; despite occupying Ubx's segments, Antp cannot reconstitute its function. Additionally, Hox proteins have been shown to bind chromatin remodeling proteins and can alter transcription regulation and chromatin composition (Lu 2003, Shen 2004).

Protein-Protein Interactions

Protein cofactors may increase Hox specificity through tandem binding to target DNA. The most prominent example of this is the well-studied Hox cofactor Extradenticle (Exd), which interacts via the conserved hexapeptide motif and modulates DNA recognition by interacting with the minor groove (Joshi 2007). Numerous target sites within the genome can only be bound by Hox proteins in tandem with a cofactor such as Exd, helping to refine Hox action (Mann 2009, Pearson 2005, Bondos 2006).

While protein cofactors can help contextualize individual Hox protein functions within their domain of action, it is insufficient to explain differences between Hox proteins. Cofactors such as Exd interact promiscuously with most Hox proteins; all but one *Drosophila* Hox proteins contain a HX motif. Exd interaction alone is unlikely to confer distinctive specificity to individual Hox proteins. Additionally, there are numerous Hox targets that lack sites for known Hox cofactors, indicating that Hox proteins are capable of DNA binding without the need for cofactors. Finally, Tan et al.

(2002) demonstrated that Hox proteins in cell culture were capable of regulating target genes without protein cofactors.

Intramolecular Regulation

The idea that ubiquitous Hox proteins are only contextualized in the presence of certain cofactors or specific chromatin landscapes is insufficient to fully explain Hox behaviors seen *in vivo*. It is likely that a major contributor to individual Hox specificity comes from intrinsic differences between the proteins themselves, likely in regions outside the HD. As these regions are less conserved than the HD, differences between Hox protein sequences could account for different functions *in vivo*. Intramolecular regulatory interactions have been documented in numerous transcription factors. For example, Li et al. (1996) demonstrated that the activity of the transcription factor ATF-2 could be repressed via direct binding between the DBD and the activation domain. Similarly, Dash et al. (1996) show that c-Myb activity is modulated by direct binding between the protein's conserved EVES motif and its N-terminal DBD.

In the case of Hox proteins, it is likely that motifs in the disordered regions modulate the DNA binding behavior of the HD. Because these regions are less conserved between members of the Hox family, these interactions may differ enough to confer the characteristic Hox behavior seen *in vivo*. However, full length Hox proteins are prone to proteolysis and aggregation and are consequently difficult to work with, and so the nature of these potential interactions is not yet known. In this thesis, we are primarily concerned with answering the question of whether intramolecular interactions occur and if differences in these interactions could explain the Hox paradox.

Ultrabithorax

In order to understand how these proposed intramolecular interactions lead to functional differences between Hox proteins, we must first understand how these interactions behave in a single protein.

Ultrabithorax (Ubx) is a well-studied Hox protein and has seen widespread use as a model for Hox protein function. As a result, Ubx is one of the most well understood Hox proteins. Ubx splicing has been characterized in detail; this protein has six known isoforms with established expression patterns (Reed 2010). Additionally, Ubx orthologues have been identified in numerous other species, allowing us to compare orthologue protein features (Grenier and Carroll 2000).

Additionally, Ubx has the largest number of characterized *in vivo* DNA binding sites (Pearson 2005). As our primary question involves Hox-DNA binding, knowledge of these sites is advantageous. Ubx's optimal DNA binding site is known (Beachy 1993) and has been used for *in vitro* DNA binding measurements; however, Liu et al. (2009) demonstrated that binding to this optimal sequence alone masks smaller contributions from non-HD regions, and thus knowledge of Ubx's natural DNA sites is necessary to fully understand this protein's DNA binding behavior.

Finally, numerous assays have been established for Ubx, including *in situ* and *in vivo* assays (Makhijani 2007, Tong 2014). These established assays provide numerous options for further investigating Ubx-DNA interactions in more complex environments.

In *Drosophila*, Ubx is responsible for specifying the identity of the posterior thorax and anterior abdomen. Its broad functions include promoting the development of

halteres in the thorax and repressing limb formation in the abdomen, as well as managing development of the ectoderm, musculature, and nervous system. Ubx also works on a smaller scale to "micromanage" segment identity, for example, in coordinating bristle formation in the legs (Hughes and Kaufman 2002).

Aside from the HD and hexapeptide motifs, Ubx has a handful of characterized regions (Fig. 4). Between the HD and hexapeptide sit three disordered microexons (b, mI, and mII) which are alternatively spliced to produce six Ubx isoforms. (Bomze 1993). The C-terminus contains two repression regions, a poly-A domain and a QA motif (Ronshaugen 2002, Galant 2002). Ubx also contains a transcription activating region N-terminal of the HD between residues 159-242. Interestingly, part of this region is computationally and experimentally predicted to form alpha-helical secondary structure (Tan 2002). However, of this region, only the structure of the hexapeptide has been conclusively solved and modelled via crystallography (Passner 1999).

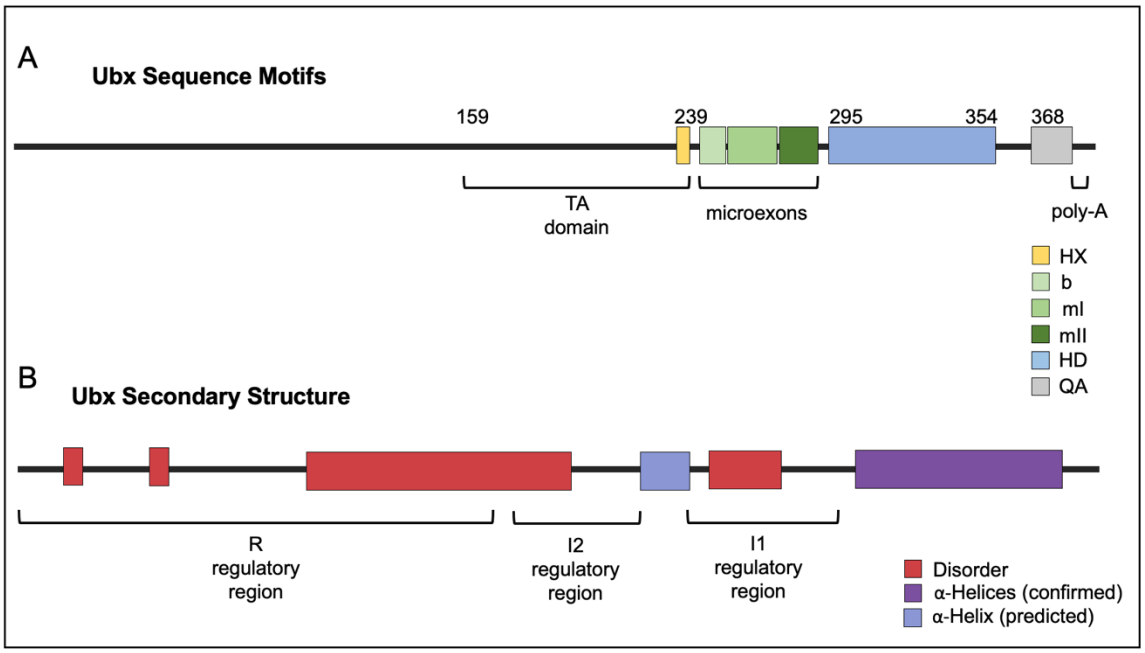

Figure 4 Characterized domains of UbxIb.
Ubx characterized motifs include the HD, HX, and alternatively spliced microexons, as well as a TA region and QA repression domain(A). The HD region contains three alpha helices, and the TA domain is predicted to contain alpha helical structure. Ubx regions of disorder and notable secondary structure are shown in B, along with three regions proposed by Liu et al. to be important in DNA binding regulation. Abbreviations: HX, hexapeptide; HD, homeodomain.

Disordered Regions in Ubx Impact DNA Binding

Several studies have demonstrated that the IDRs of Ubx impact DNA binding of the full-length protein. Liu et al. (2008) demonstrated that full length Ubx binds its optimal target sequence 2.5-fold weaker than the Ubx HD and identified three regions in the N-terminal disordered regions that impact this binding affinity (Fig. 4). There is also evidence that three motifs in the I1 region shown in figure 4B interact complexly to

impact full length binding specificity. Liu et al. (2009) demonstrated that the HX, the six amino acids N-terminal to it, and the eight amino acids C-terminal to it each contribute both individually and synergistically to modulate DNA binding specificity. Ubx splicing isoforms, which differ based on the inclusion or exclusion of the microexons, have unique functions *in vivo*, further suggesting that these regions may influence DNA binding by the full-length protein (Reed 2010).

There are two mechanisms by which disorder may impact DNA binding. Firstly, disordered regions may indirectly affect the HD conformation through allostery. Due to the stability of the HD, this mechanism is unlikely. More likely, disordered regions physically prevent HD-DNA binding, either via dynamically blocking the HD, or through direct contacts between disordered and HD residues. In either of these cases, the conformation of Ubx must change to allow DNA binding to occur.

Conformational Change Hypothesis

As noted above, transcription factors can utilize direct intramolecular interactions to regulate their activity. Liu et al. (2008) determined that certain Ubx disordered regions which reduce binding affinity of the full-length protein were likely to do so by sterically impeding HD-DNA interactions. Additionally, *in vitro*, Ubx can self-assemble into materials in which monomers make specific bonds between tyrosines in the HD and the N-terminal disorder (Howell 2015). It is likely that these specific interactions, many of which occur between conserved residues, are also important in *in vivo* Ubx function.

These findings suggest that Ubx's disordered regions physically interact with the HD, and protein conformation must change so that HD-DNA interactions can occur.

We propose that Ubx populates an ensemble of conformations bounded by "open" to "closed" states. In the open state, disordered N-terminal domains are sequestered away from the HD, leaving it exposed and free to interact with DNA. In the closed state, however, disordered domains crowd around the HD, blocking it from target DNA. In this way, intramolecular interactions determine the protein's overall binding ability by directly affecting HD availability (Fig. 5). Data from a prior student demonstrated that Ubx makes specific tyrosine interactions between residues in the HD and conserved residues in the N-terminal disorder, and this hypothesis is additionally supported by the specific tyrosine bonds formed between many of these residues in materials formation (Churion 2017, Howell 2015).

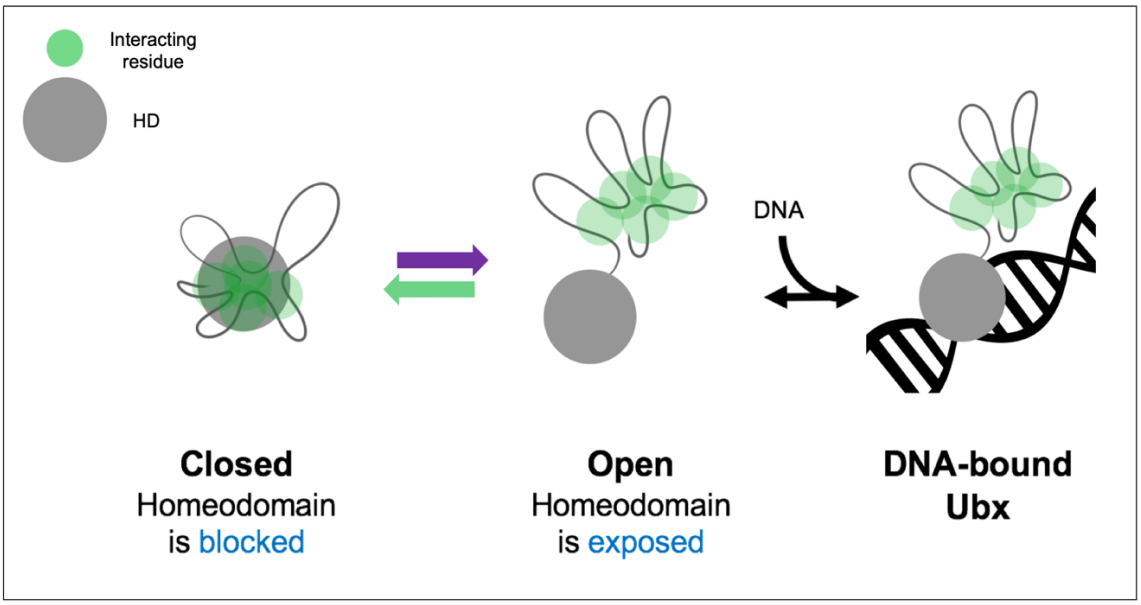

Figure 5 Functional model for full length Ultrabithorax.
Our model proposes that Ubx changes conformation from closed to open in order to tune DNA binding. The protein is closed when residues in the disordered N-terminus interact with residues on the HD (green arrow), and the protein is open when residues in the N-terminus interact with each other (purple arrow). The open conformation permits DNA binding via the exposed HD. Residues and regions shown are not to scale.

If this hypothesis is correct, then we should be able to observe evidence for multiple structures of full length Ubx, which we demonstrate in this thesis.

Investigating Ubx Structure

Due to challenges in working with full-length Hox proteins, no physical model has yet been shown for any Hox protein beyond the HD and HX motif.

Because regions outside the HD consist mainly of IDRs, x-ray crystallography cannot be used to resolve the full-length structure. NMR spectroscopy is a popular

technique for characterizing intrinsically disordered proteins, however, Ubx aggregates at concentrations above 80 μM, and mM concentrations are needed to perform this technique. Numerous aromatic residues in Ubx also render fluorescence approaches unsuitable, as there is a high probability that aromatic fluorophores would disrupt the protein's structure.

Microscopy is promising in that it can image particles at low concentrations without the need for probes. Negative stain electron microscopy (NSEM), the technique employed in this thesis, was developed in 1959 to image virus particles (Brenner and Horne 1959). This technique involves staining a sample with a heavy metal salt, such as uranyl acetate. Negative stains lightly coat particles but primarily interact with the background, providing a dark halo around the sample. This staining preserves the sample's structure, provides protection against radiation damage which may occur when the subject is exposed to an electron beam, and greatly increases image contrast (Bremer 1992). Consequently, NSEM can provide higher resolution images of much smaller particles. In contrast with cryo-EM, which requires particles to be greater than 200 kDa for efficient imaging, NSEM has been used to image particles as small as 28 kDa (Rames 2014). Because Ubx is about 40 kDa, negative staining is best suited for resolving its structure.

In this thesis, negatively stained samples were imaged using transmission electron microscopy (TEM). TEM involves passing an electron beam through a sample and using the resultant electron interactions to produce a micrograph image. TEM is able to yield images with higher resolution than both light and scanning electron microscopy,

with modern microscopes able to resolve particles at 1Å or less (Franken 2020). This high resolution makes this technique favorable for imaging small molecules.

CHAPTER II

3D STRUCTURE OF A FULL LENGTH HOX PROTEIN AND EVIDENCE FOR MULTIPLE CONFORMATIONS

Introduction

Hox proteins regulate numerous developmental processes in bilaterian organisms and are necessary for the proper formation of specific morphologies along the anterior-posterior axis (Hughes and Kaufman 2002). Hox protein misexpression results in disruptions of morphology, a prominent example being the bithorax mutant, which lacks Ubx and subsequently develops a second set of wings (Lewis 1978).

Each Hox protein regulates region-specific developmental outcomes; in the development of the *Drosophila* heart, for example, Ubx oversees processes that build the posterior aorta, while its neighbor AbdA regulates different genes to form the heart terminal chamber (Monier 2005). Despite these distinct and specific functions *in vivo*, the DNA binding HDs of Ubx and AbdA are over 90% identical (Bondos 2001). This contradiction between Hox proteins' observed specificity *in vivo* with their highly conserved HD has been termed the Hox paradox (Hoey and Levine 1988).

While interactions with cofactors and the cellular context in which Hox proteins are expressed partially resolve this paradox, it is likely that the structural characteristics of the proteins themselves also regulate DNA binding. Hox proteins are largely unstructured outside of the HD, which has made understanding potential intramolecular regulation challenging. However, Liu et al. (2008) demonstrated that that non-HD N-

terminal regions impact HD-DNA binding in the Hox protein Ubx. Based on the pH dependence of this interaction, some of these non-HD regions are likely to directly interact with the HD. These interactions between HD and non-HD residues may directly control DNA binding by blocking or exposing the HD via a conformational change.

In this thesis, we show evidence for this conformational change by resolving multiple 3D structures for the Hox protein Ubx using negative stain TEM. In one structure, which we refer to as the open state, the HD is far from the N-terminal disorder, and thus is capable of binding target DNA. In the second structure, which we refer to as the closed state, the HD is buried within the N-terminal disordered regions, which sterically impedes HD-DNA binding.

Materials and Methods

Growth and Purification

The UbxIa isoform was expressed in BL21 (DE3) pLysS *E. coli* grown in LB broth using the pET-19b plasmid in order to provide a His-tag on the protein's N-terminus. Following IPTG induction, cells were expressed for 2 hours before being spun down for 10 minutes at 17,568 x g. Cell pellets corresponding to either 1 or 2L of cell culture were collected and frozen at -20°C.

Protein was extracted and purified as described in Appendix A. All steps except lysis were performed at 4°C.

Imaging and Analysis

Purified protein was dialyzed overnight into a buffer containing 100 mM KCl, 20 mM Tris–HCl, and 10 mM β-mercaptoethanol (BME), and adjusted to pH 7.5. 2-4 µl of sample was deposited on a glow-discharged carbon 300-mesh grid for 1 minute before blotting off excess fluid using Whatman 1 filter paper. The grid was subsequently incubated with 2-4 µl filtered 2% uranyl acetate stain for 45 seconds before blotting dry.

Digital micrograph images were recorded on a 200 keV cryo Transmission Electron Microscope (cryo-TEM) ThermoFisher Scientific (TF20) using a Gatan Tridiem GIF-CCD camera at 67000X magnification, resulting in a final pixel size of 2.08 Å. Particles in these images were boxed using the program e2boxer.py of the EMAN2.0 software package (Tang 2007). Particles were corrected for contrast transfer function (CTF) using the bcft program of the software package BSOFT (Heymann and Belnap 2007).

Reference-free 2D class averages of the particles were generated using the program e2refine2d.py of the EMAN2.0 software. Class averages and the corresponding particles were then sorted by size. Initial reference maps for each of the selected groups were generated without imposing symmetry using the e2initialmodel.py program of EMAN2.0. Ten iterations of 3D reconstructions were performed in BSOFT for each group, after which 3D classification and a further 20 iterations of 3D reconstruction were performed to generate the final 3D models.

Results

Micrograph Images Show Conformational Heterogeneity

To observe protein conformation, purified Ubx was stained with uranyl acetate and imaged via TEM. Raw particle images were corrected for contrast transfer function (CTF), and these corrected particles were used to generate reference-free 2D class averages.

The first 2D class average dataset consisted of fifty 2D classes derived from a total of 5,273 particles. These classes varied by particle size and were thus grouped into three approximate categories: “small”, “medium”, and “large” (Fig. 6). 2D class averages in the initial “small” category were derived from 2,302 particles, while the initial “medium” 2D classes were derived from 2,325 particles. The initial “large” 2D classes were derived from 646 particles.

Based on preliminary EM images generated by a previous student in which Ubx monomers were imaged alongside GroEL, we were able to estimate the diameter of Ubx as approximately 5nm (Booth, unpublished). This measurement, when compared to that of other IDPs of similar sequence length, was determined to be plausible (Uversky 2012). The small and medium classes were grouped together for a 3D reconstruction, as it is likely that they reflect different viewing angles of identical particles. Because Ubx is not radially symmetric, identical particles viewed from different angles in 2D may appear to be different sizes; however, as the small and medium particles were both close to the 5nm diameter estimate for Ubx, we are confident in grouping these two categories. The 5nm diameter was, however, expected to vary somewhat for alternate

conformations. We hypothesized that the large 2D classes may represent an alternate conformation, and these particles were used to generate a separate, independent 3D reconstruction.

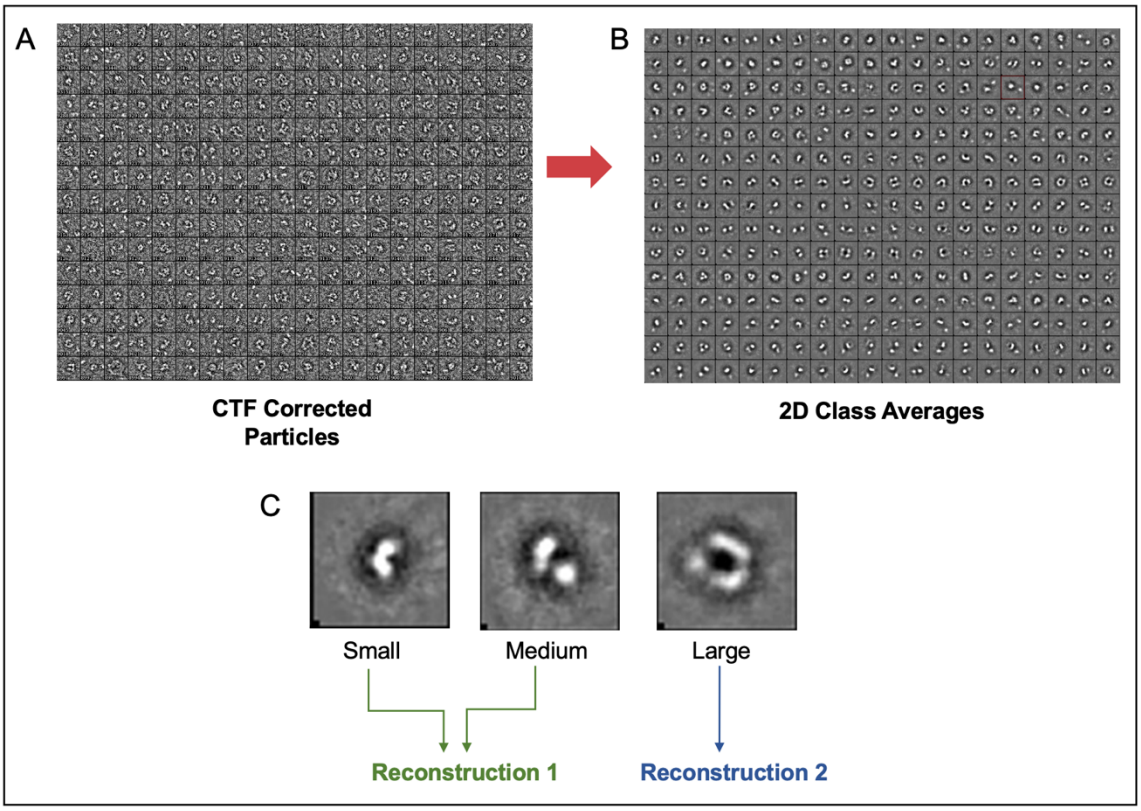

Figure 6 2D class averages show Ubx heterogeneity.
CTF corrected particles (A) were used to generate 2D class averages (B). 2D class averages were sorted into three groups based on particle size. Small and medium were grouped together for reconstruction, while large particles were used for an independent reconstruction (C). Note that due to the size of the dataset, only a subset of the total particles is shown in A.

As, noted above, the “large” 2D classes were derived from 646 particles and were the least represented in the dataset. This limited the resolution we could achieve in a 3D model. To determine if this subset of particles represented an alternate conformation as we suspected, it was necessary to increase the number of particles in this state. Because Ubx is a transcription factor, and our functional model suggests that DNA binding is tied to a conformational change, it is possible that DNA contaminants could alter the conformational equilibrium of our protein. Our initial dataset was obtained using Ubx monomers purified on a Ni-NTA resin, which binds protein via a Histidine tag on the N-terminus. We hypothesized that DNA contaminants present after lysis could remain bound to the Ubx HD and persist in the final sample. These contaminants would trap Ubx in the DNA-bound open state, reducing the concentration of particles present in the closed state. To overcome this, we switched to a 2-column purification procedure. This protocol uses a phosphocellulose resin, which mimics DNA and binds Ubx via the HD, outcompeting any DNA contaminants that may be in complex with the protein. We used this purification protocol for the remainder of our sample generation in order to reduce DNA contaminants and their potential effects.

To produce sufficiently refined 3D models, we aimed to produce a minimum of 6,000 particles per group. Our final dataset was derived from particles collected from multiple sample preparations and imaged over multiple sessions. In total, over 30,000 particles collected from 700 micrographs were used to generate our 3D models. The total number of particles contributing to each final reconstruction, along with the general experimental methodology followed, is shown in Figure 7.

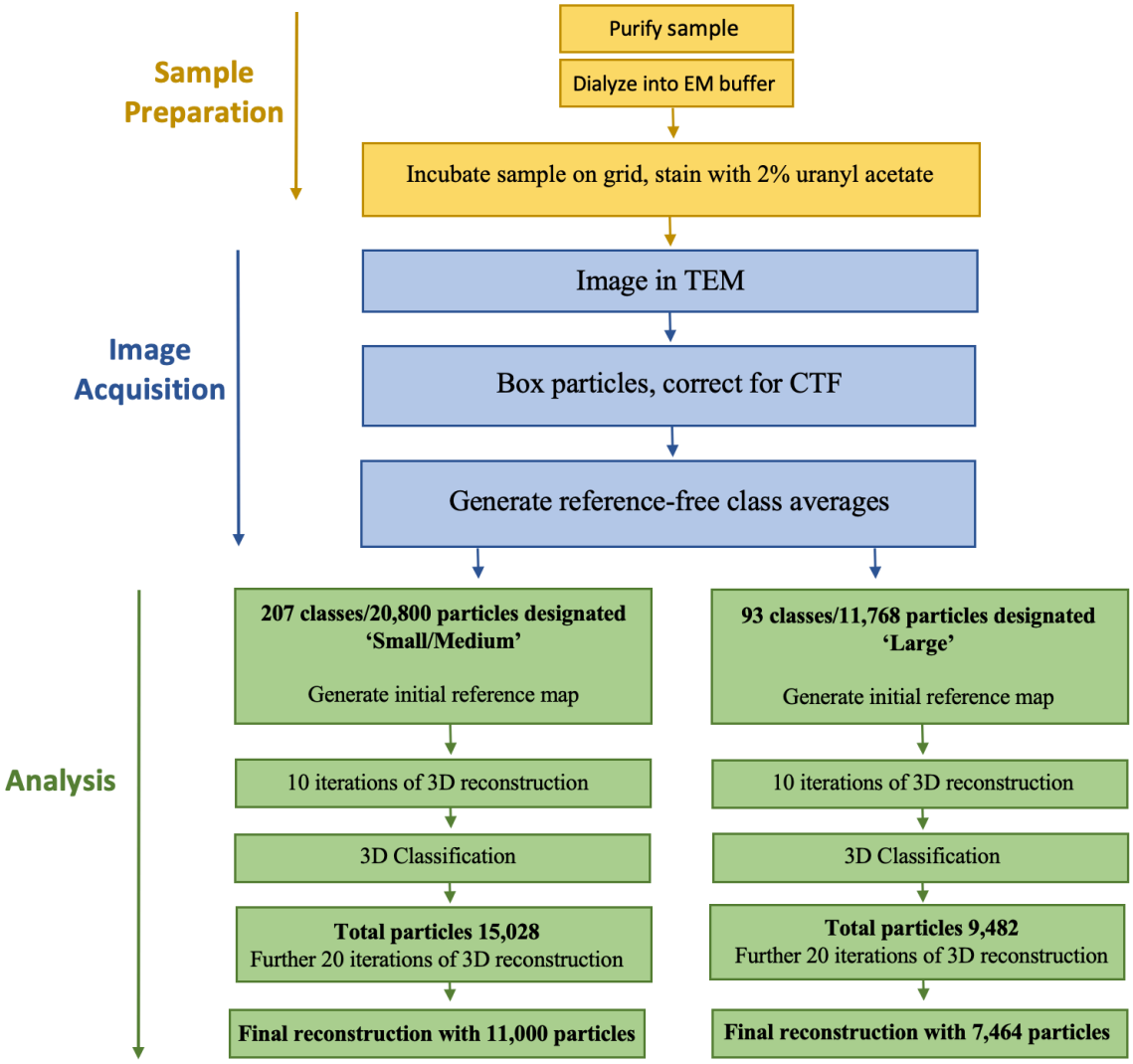

Figure 7 Experimental flow and final particle counts.
The experimental methodology was followed as shown here.

3D Reconstructions Show Open and Closed Conformations

Because small and medium particles most closely matched our expectation for the size of Ubx monomers, these particles were first grouped and used to produce a 3D reconstruction. A total of 207 classes consisting of 20,800 particles were designated as either small or medium. Thirty iterations of 3D reconstructions were used to further refine the dataset, and the final 3D reconstruction was generated from a total of 11,000 particles. (Fig. 8). Previous EM work done by Rebecca Booth again helped us confirm the shape of our monomers; the general orientation of particles seen in the old 2D class averages suggested a large, roughly spherical hollow ball attached to a smaller domain. This layout was also seen in both our current 2D class averages and final 3D reconstruction.

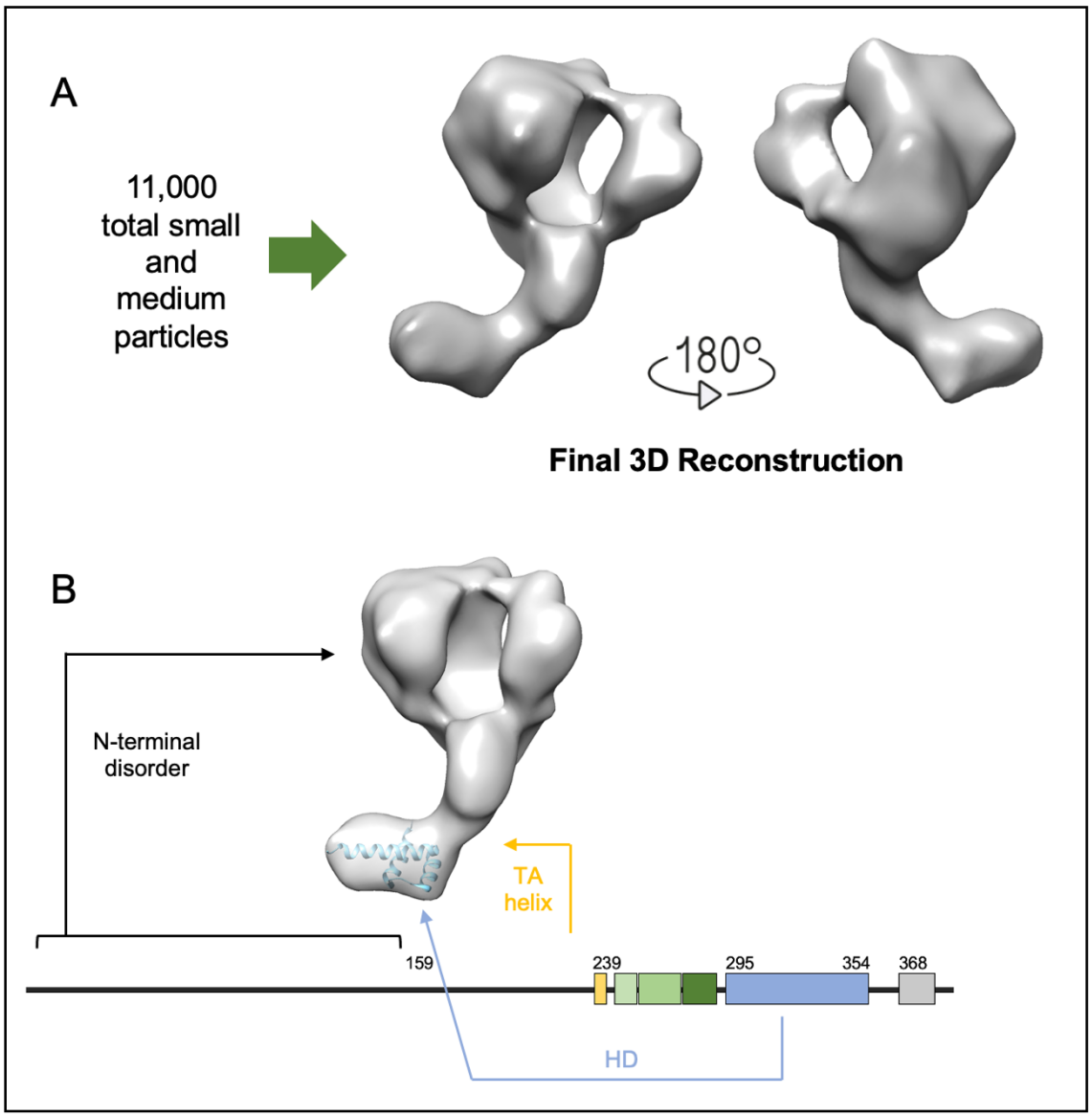

Figure 8 3D reconstruction of small and medium particles gives the open conformation.
A total of 11,000 small and medium particles contributed to the open state model (A). Several domains can be identified on this structure, including the HD, N-terminal IDRs, and TA helix (B).

We propose that the small domain at the bottom of our model corresponds to the HD, while the larger regions at the top correspond to the N-terminal disorder. The neck region of our structure, which connects the two domains, is narrow enough that it can't be consistently seen in our 2D class averages, and thus it is unlikely that this region contains more than a single alpha helix or beta sheet. Tan et al. (2002) have proposed that a portion of the Ubx TA domain contains alpha helical characteristic, and so we propose that the neck region of our structure contains a single alpha helix and represents the C-terminal portion of the TA region. As this region remains narrow in both our density map and 3D reconstructions, it is unlikely for the protein to loop back though this domain. As a result, we are confident that the top and bottom of our structure represents the N and C-terminus of the protein, respectively. As the HD has well characterized size, structure, and location, we assign this motif to the smaller domain in our 3D model, which represents the C-terminus of the protein. The large domains at the top of our model likely correspond to the remainder of the protein's N-terminus, which consists predominantly of intrinsically disordered regions (IDRs). Based on the HD's orientation relative to the N-terminal IDRs, we classify this structure as the open state.

The open state 3D model was verified before generating the final reconstruction via comparison between a preliminary 3D reconstruction and an independently generated set of 2D class averages derived from a separate sample preparation (Fig. 9).

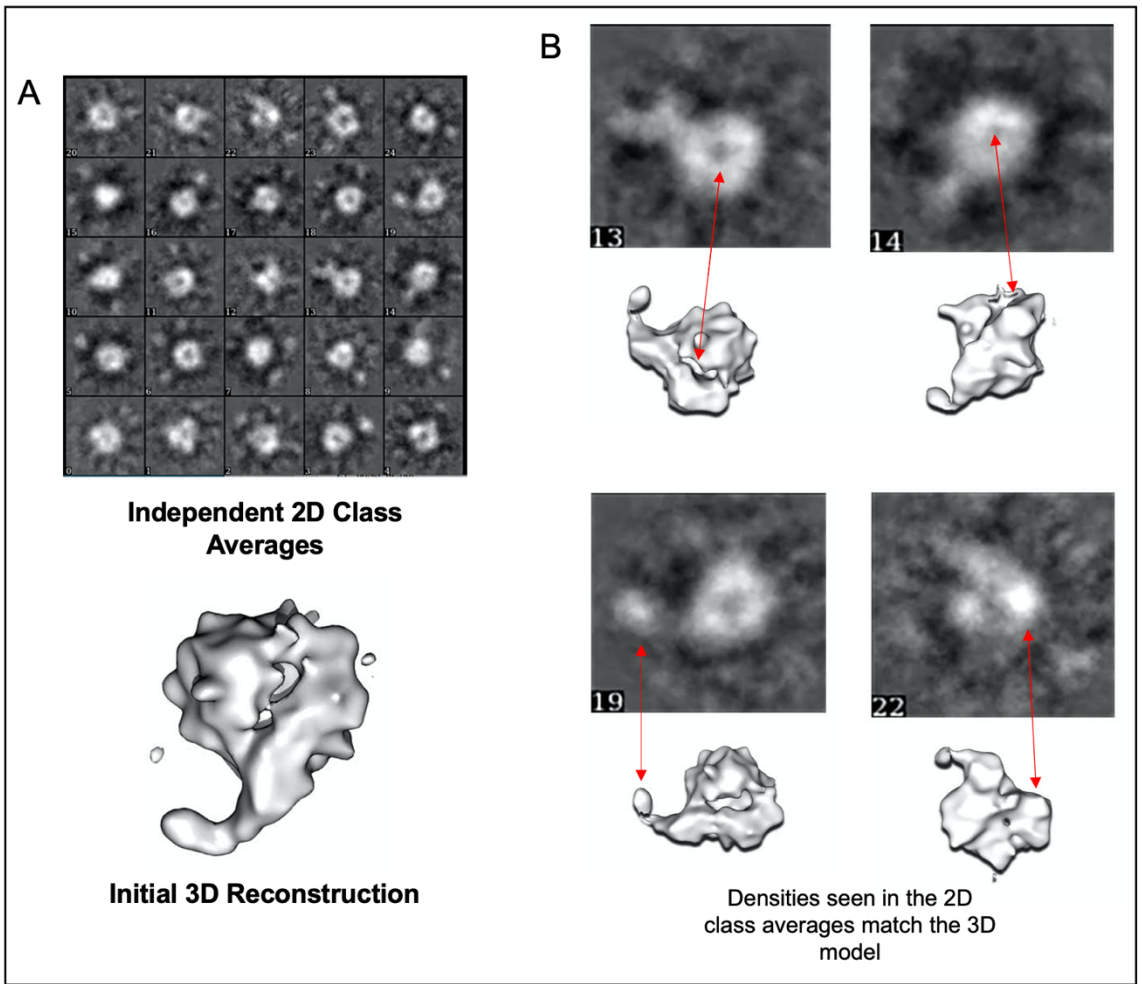

Figure 9 Independent 2D class averages verify the 3D model.
The preliminary 3D model was compared with an independent subset of 2D class averages (A). The 2D class averages show similar structural features to the preliminary 3D reconstruction (B).

Several domains in the independent class averages closely match the structural features seen in the preliminary 3D model. In the 2D images, bright areas correspond to regions of the particle that are closer to the viewer, while duller, more transparent regions are farther away. Many of the bright areas in the 2D class averages correspond to raised

protrusions on the 3D model, while many dull areas correspond to cavities. Additionally, the silhouette of the protein appears to be conserved between 3D model and the 2D class averages. These similarities indicate that the 3D model is most likely derived from the same protein from which the independent 2D class averages are.

A total of 93 classes with 11,768 particles were classified as large. Thirty iterations of 3D reconstructions further refined this dataset, and the final 3D reconstruction was generated from 7,464 particles. As this model was derived from fewer particles, it could not be refined to the same degree as the open state model. Consequently, structural features are not as defined. However, it is still possible to propose the location of both the HD and the IDRs. Notably, in this model, and HD is in much closer proximity to the N-terminal IDRs, and so we propose that this model corresponds to the closed state of Ubx (Fig. 10).

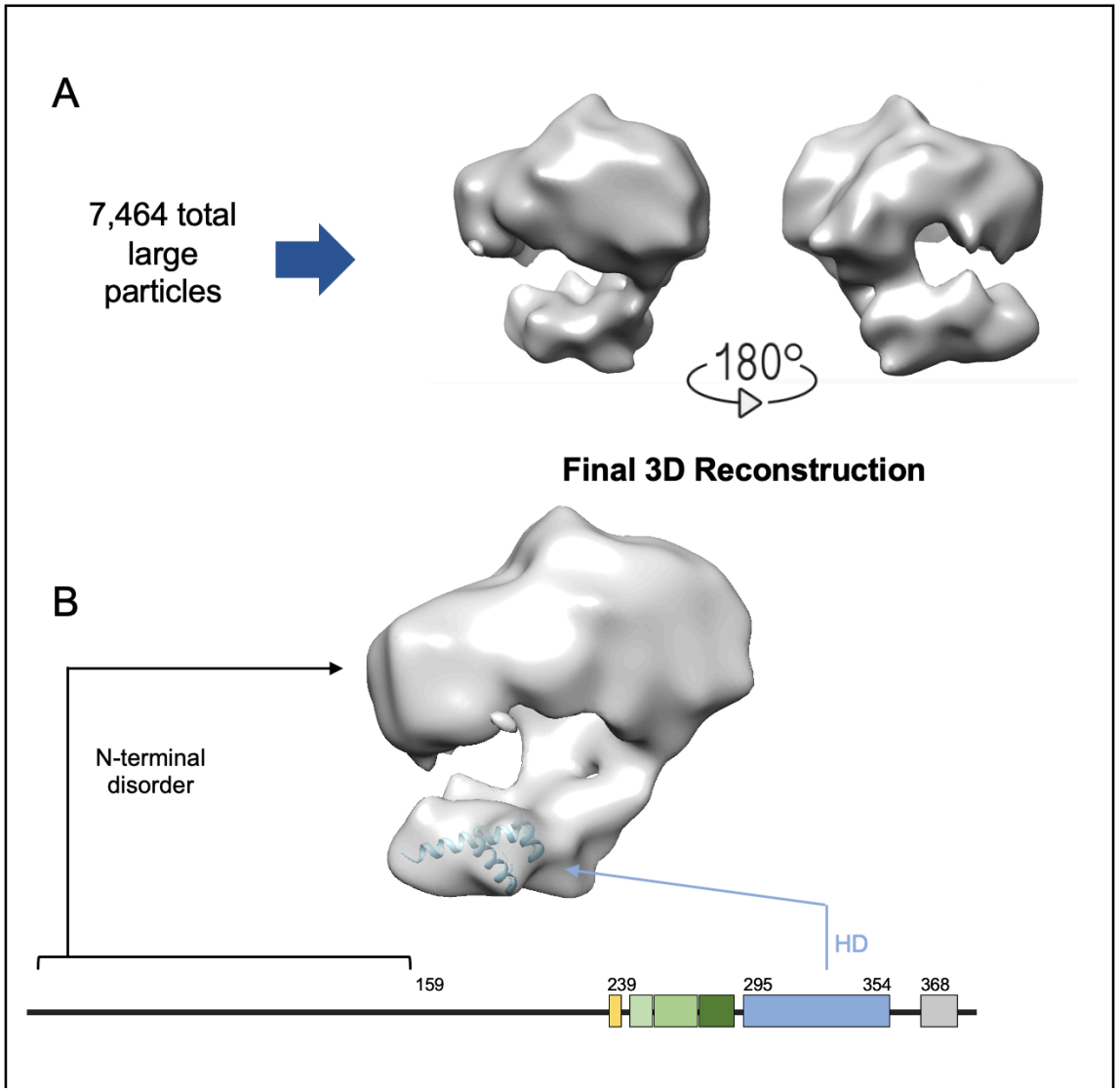

Figure 10 3D reconstruction of large particles gives the closed conformation. 7,464 large particles contributed to the final 3D model of the closed state (A). Proposed assignments for the N-terminal disorder and HD are indicated (B).

As proposed by Churion (2017), it is likely that the IDR motifs interact with the HD via specific tyrosine residues. Specifically, tyrosine residues 4, 12, 167, and 240 in the non-

HD motifs and residues 293 and 296 in the HD are likely to be important. These residues form specific bonds during *in vitro* self-assembly and fall in conserved regions of the protein (Howell 2015). Because of this, we propose that these residues are also likely to be important in *in vivo* protein function. While the HD is exposed in the open state model and therefore free to interact with DNA, it is likely that the proximity of the HD and disordered regions impedes DNA binding by the HD in the closed state. The existence of both the open and closed state supports our model for HD-IDR interactions, in which motifs in the HD form contacts with motifs in the N-terminal disordered regions. These interactions are likely to modulate the DNA binding activity of the full-length protein by controlling access to the HD.

Domains in the N-terminal Disorder may Correspond to Self-Associating Regions

Our 3D models show distinct globular densities in the N-terminal disorder. There is negligible charge or structure in the Ubx N-terminus, so it is unlikely that these properties account for the lobes seen in our 3D structures. Based on the sequence of Ubx, we hypothesize that these densities may correspond to self-associating motifs. As shown in Figure 11, Ubx contains several tracts of similar amino acids, the most notable of which are poly-glycine, poly-alanine, and poly-arginine. Because molecules with similar hydrophobicity tend to dissolve each other, we propose that these relatively uniform regions may preferentially self-associate rather than mixing with the remainder of the N-terminal domains.

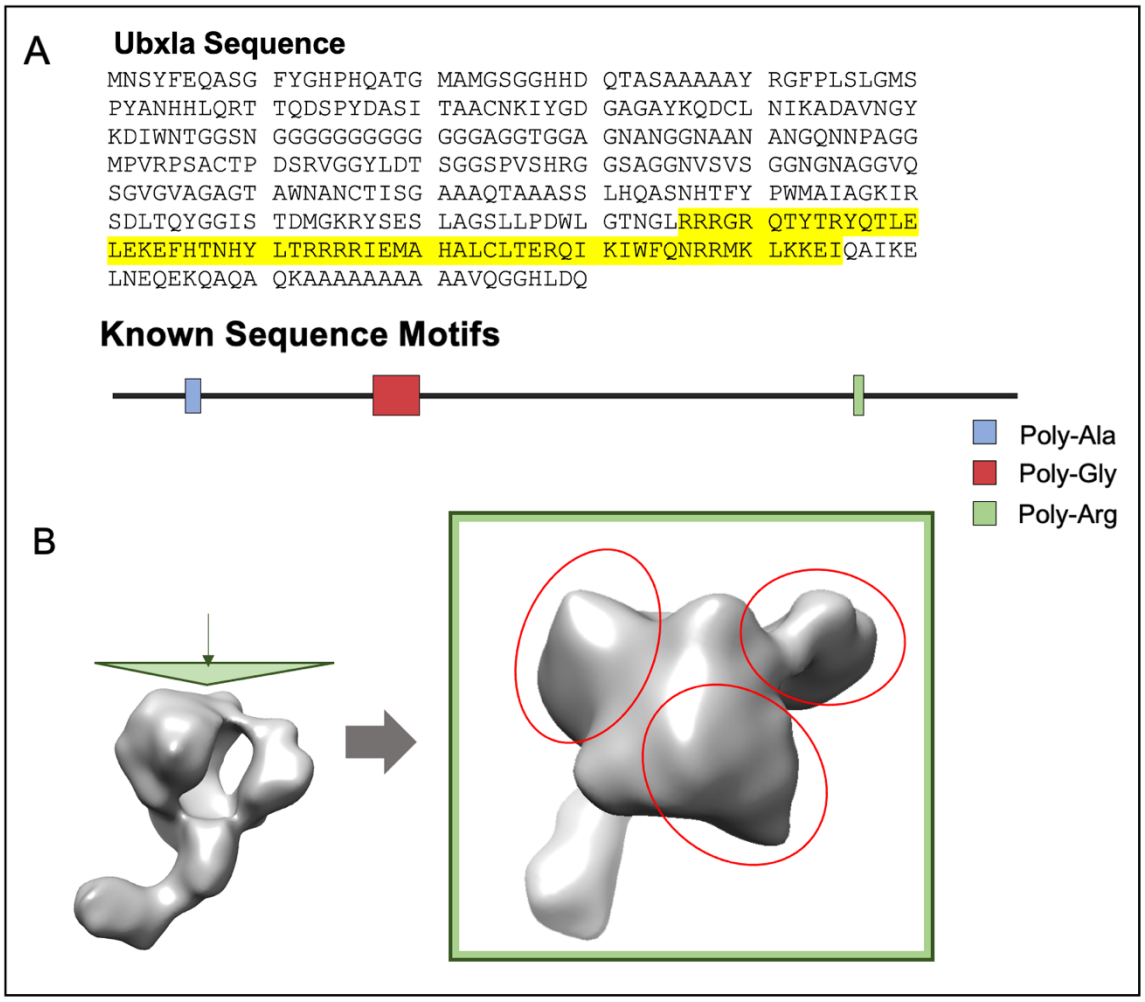

Figure 11 Sequence motifs in UbxIa.
Different amino acids are enriched in certain regions of the Ubx protein sequence, particularly in the N-terminal IDRs (A). These regions may self-associate in the Ubx 3D structures we've produced (B). The highlighted sequence corresponds the HD.

Intermediate States Seen in 2D Class Averages

In addition to the open and closed state, several 2D class averages show what appeared to be intermediate conformations. In these classes, the HD appears to be situated at variable distance from the N-terminal regions (Fig 12).

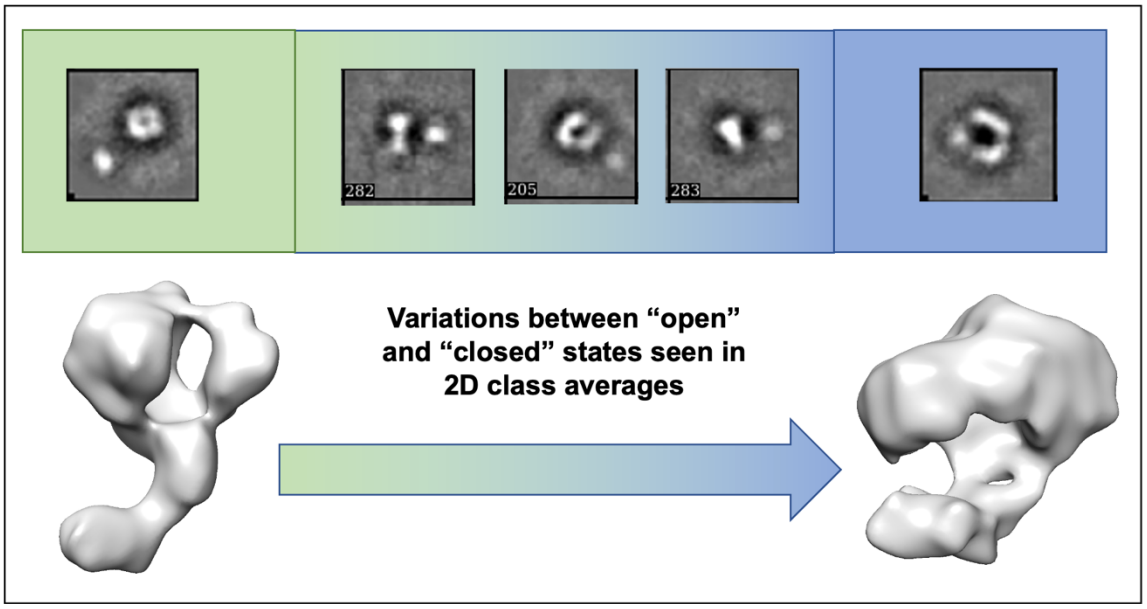

Figure 12 Intermediate states seen in Ubx.
Several 2D class averages appear to show intermediate conformations of Ubx. Three examples of such classes are shown here.

These 2D class averages suggest that Ubx populates a continuum of states that are bounded by the open and closed conformation.

Conclusion

Despite vast similarities between HD sequence, individual Hox proteins bind DNA with distinct affinities and specificities. This observation, known as the Hox paradox, has of yet remained largely unexplained, but it is likely that disordered regions form intramolecular interactions with the HD to regulate HD-DNA binding. However,

lack of a full-length structure has hindered more in-depth characterization of Hox protein function, particularly in regard to DNA binding regulation. Here we have shown the first structural model of a full-length Hox protein using Ubx. Several prominent sequence features, such as the HD and TA helix, can be distinguished on our 3D model. Additionally, we have demonstrated that monomers of the Hox protein Ubx have multiple conformations, and that these conformations differ in degree of HD availability for DNA binding.

These structures support a model of Hox protein function in which motifs in the Ubx IDRs modify protein-DNA binding by directly controlling HD availability. When residues in non-HD motifs interact with residues in the HD, the N-terminus forms a blockade around the HD, and the protein is "closed." When non-HD motifs interact with each other, the HD is available for DNA binding and the protein is "open." Lui et al. (2008) determined that truncations in the N-terminus can drastically change Ubx-DNA binding behavior, and Churion (2017) demonstrated that point mutations of specific tyrosine residues similarly altered DNA binding and in some cases resulted in proteins that strongly favored either the open or closed conformation. Based on these previous data and our own, we propose that this interaction occurs via direct interactions between residues in the N-terminal motifs and residues on the HD.

Sequence Collapse in the Ubx IDRs

The majority of the Ubx sequence, spanning from the N-terminus through the start of the HD, is intrinsically disordered. As very little of this region is predicted to form stable secondary structures, we initially expected that this region would resolve as

a dense spherical globule with few defining structural characteristics. However, our 3D model resolved the Ubx IDRs as distinct domains surrounding a hollow cavity. We hypothesize that these domains are the result of local sequence collapse and can be understood through the properties of liquid-liquid demixing. Holehouse and Pappu (2018) propose three primary factors in determining protein collapse: backbone interactions, side chain interactions, and solvent-polypeptide interactions. These interactions govern protein collapse in both folding and disordered proteins and can thus be used to understand the organization of Ubx's IDRs. For example, one of the most prominent sequence features of the Ubx IDR is the poly-glycine region spanning amino acids 107-131 (Bondos 2004). Favorable amide-amide backbone interactions and water release would dictate the local collapse of this region (Drake and Pettitt 2015). The collapse of other IDR sequences may be governed by more complex interplay between side chain interactions as well as solvent and backbone interactions, but, as with the poly-G region, these interactions would be encoded in the amino acid sequence (Holehouse and Pappu 2018). Thus, we propose that the distinct densities that we see in our 3D model are driven by local amino acid composition.

Because our dataset consists of a heterogenous population of Ubx conformations, it is difficult to faithfully determine the structure of these densities. Furthermore, because most of these regions are intrinsically disordered, they are highly dynamic, and thus we are unlikely to be able to resolve the backbone. We propose that truncation mutants may be used to test this local self-association hypothesis; if these regions correspond to specific densities in the N-terminus, then we would expect to see a

reduction of individual protrusions in mutants with the corresponding amino acid tracts deleted. The structural characteristics of these mutants would again be determined using electron microscopy.

Confirmation of the Open and Closed States

We have shown both an open and closed state for Ubx. However, as shown in Fig 12, several of the 2D classes generated by our final dataset showed what appeared to be intermediate states, with the HD density at varying orientations relative to the IDRs. These intermediates limit the number of particles we can obtain for the fully open and fully closed states and may interfere with our final models if any of these intermediates were included in either category.

Previously, our lab has generated tyrosine point mutations that disrupt Ubx's intramolecular interactions (Churion 2017). Notably, two of these point mutants show significantly altered DNA binding affinity. The first, Y4S/Y12S, binds DNA approximately 7-fold weaker than wild type Ubx. The second, Y293S/Y296S, bind DNA approximately 3-fold stronger than the full-length protein. We hypothesize that this change in binding affinity is a direct result of protein conformation; the 4/12 mutant binds DNA weakly because it favors the closed conformation, while the 293/296 mutant favors the open state. 3D models generated from these mutants may therefore be used to confirm or fine-tune the structures we've proposed, as we expect each protein to more strongly favor either the fully closed or fully open state.

CHAPTER III

CONCLUSIONS

In this thesis, we have used negative stain TEM to resolve multiple structures for the Hox protein Ubx. These structures, which we term "open" and "closed," differ on the basis of HD availability. In the closed state, the HD is in close contact with disordered regions, which sterically impede DNA binding. In the open conformation, the protein's disordered regions self-associate and are sequestered away from the HD, which remains exposed and free to bind target DNA.

Hox proteins are vital in proper morphological development of bilaterian animals (Hughes and Kaufman 2002). However, limited knowledge of the full-length structure of Hox proteins has hindered our understanding as to how these important proteins carry out their functions *in vivo*. Our work here represents the first structural model for a full-length Hox protein and provides new and valuable insight into how these proteins may regulate their DNA binding. The existence of the multiple conformations we've shown here supports the idea that intramolecular protein interactions can control HD-DNA binding in Hox proteins. This conclusion supports previous findings in the lab. Churion (2017) demonstrated that inhibitory interactions were likely to form between N-terminal motifs and the HD, which would necessitate a conformational change to allow DNA binding. The role of non-HD regions in influencing DNA binding ability can also explain the observed differences between full-length Hox-DNA binding affinity and HD-DNA binding affinity (Liu 2008).

Implications for the Hox Paradox

The conformational change model provides new insight into the long-standing Hox paradox. Individual Hox proteins achieve *in vivo* functional specificity despite the large degree of conservation in their DNA binding HDs. Our model suggests that non-HD disordered regions play a key role in determining HD availability and subsequent DNA binding. Because many non-HD regions are less conserved between Hox proteins, the interactions between these motifs and the HD may differ substantially between members of the Hox family. The equilibrium between the open and closed conformations is what ultimately determines the unique binding behaviors of each Hox protein. These interactions could explain observed differences between individual Hox functions.

Electrophoretic mobility shift assays (EMSA) have been used previously in our lab to determine the DNA binding affinity of full length Ubx (Churion 2013). Using this technique to measure the DNA binding affinities of different Hox proteins for the same consensus DNA target could help to validate our hypothesis. If intramolecular regulation accounts for differences between Hox proteins, then we expect each Hox protein to display unique binding affinities for this consensus sequence, even *in vitro.*

Implications for Posterior Prevalence

Aside from the Hox paradox, our model also has the potential to answer some long-standing questions in the field of Hox research.

Posterior prevalence describes the observation that, when two Hox proteins are co-expressed, the Hox protein that normally occupies the more posterior expression domain is functionally dominant (Gibson and Gehring 1988). This can be seen in the homeotic mutants; ectopic expression of Antp in the Lab expression domain transforms antennae into legs, the more posterior feature. However, ectopic expression of the anterior protein Lab in the Antp expression domain cannot transform legs into antennae (Lewis 2004). Only loss of function of a posterior protein can result in the phenotype of an anterior Hox protein moving back along the organism, as is the case in the Ubx loss of function mutant (Lewis 1978). Posterior prevalence is violated in a few select cases, but this observation largely holds true across all bilaterian organisms (Capovilla 1998). While this phenomenon has been known for several years, no mechanistic explanation has yet been proposed (Lewis 2004).

Because Hox proteins are transcription factors, DNA binding may play a key role in the determination of functional dominance. Examining this problem in the context of our model, we hypothesize that, in general, posterior Hox proteins spend more time in the open conformation and are thus more readily able to bind their target DNA. More anterior Hox proteins, in contrast, likely spend more time in the closed conformation and don't bind DNA as readily. This conformational preference should be reflected in the number of contacts that form between the HD and non-HD motifs, with posterior proteins forming fewer contacts than anterior proteins (Fig. 13).

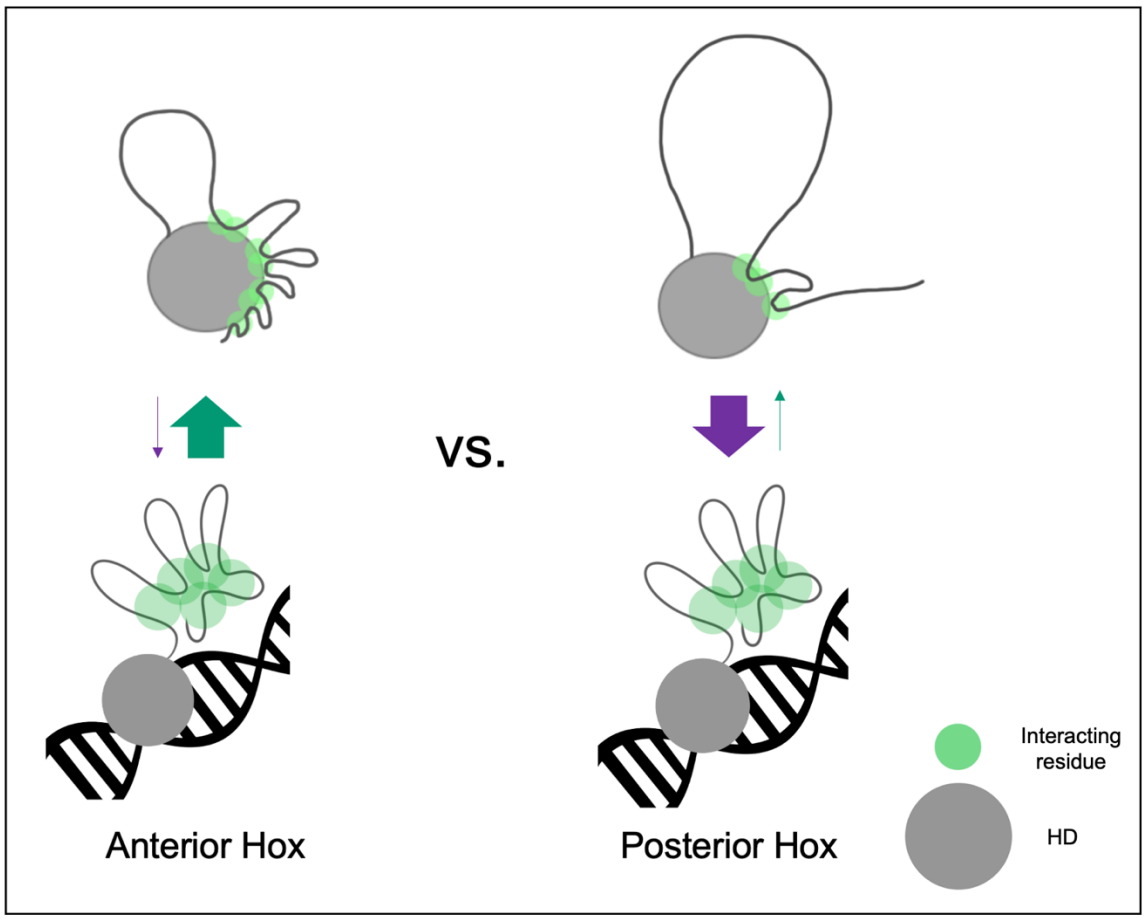

Figure 13 Posterior prevalence may be a function of protein conformation. We hypothesize that protein conformation and subsequent DNA binding affinity may account for posterior prevalence. Anterior Hox proteins may form more contacts between HD and non-HD regions, causing these proteins to favor the closed conformation and bind DNA weakly. Posterior proteins may form fewer HD-to-non-HD contacts, and thus favor the open conformation and bind DNA tightly.

Preliminary *in silico* data suggest that this may be the case: there is a negative correlation between the position of a protein on the Hox cluster and the percentage of its sequence predicted to form intramolecular interactions (Churion 2017). EMSA experiments could once again shed further light on this hypothesis. If our model is

correct, then we would expect to see a lower DNA binding affinity (tighter binding) for full length posterior Hox proteins, and a higher affinity (weaker binding) for anterior Hox proteins.

Implications for IDP and Transcription Factor Organization

IDPs make up approximately one third of all human proteins and are vital for important functions such as molecular assembly and cell signaling (Dunker 2002). These unstructured proteins must somehow avoid aggregation and proteolysis; however, it is unknown how this is accomplished without recognizable secondary structure. Simister et al. (2011) propose that IDRs may spontaneously form organized domains via association with a structured domain on the protein's N-terminus, allowing disordered regions to adopt a more stable, condensed state. This N-terminal structured domain (NTSD) spontaneously folds, providing a nucleation site on which the protein's largely disordered C-terminal domains can associate via intramolecular interactions. In this model, numerous motifs within the C-terminal IDRs interact with the NTSD, leading to the compaction of the entire polypeptide chain. Our Ubx structural model appears to follow a similar level of IDP organization, except that the Ubx structured domain is on the C-terminus rather than the N-terminus.

The Hox transcription factor family is like other TF families in that its members share a highly conserved DBD embedded in larger, less conserved intrinsically disordered regions (Pabo 1992). For this reason, our findings in Hox proteins may be applicable to transcription factor organization as a whole.

Previous literature supports the idea that IDRs are important in conveying characteristic binding behavior to TFs. In their 2017 review, Staby et al. give an overview of the numerous mechanisms by which TF disordered regions can tune protein behavior. Disorder flexibility allows TFs to bind numerous partners, potentially diversifying their DNA targets, while short disordered motifs can tune both protein-protein and protein-DNA interactions. Notably, these authors observe that identical sequence motifs can have drastically different context-dependent functions. This finding may be tied to our Hox model; different contexts such as splicing, protein cofactor interactions, or post-translational modifications may drive different motif interactions, resulting in the protein favoring either the open or closed conformation.

Brodsky and Jana et al. (2020) report a system wherein numerous specificity elements scattered throughout IDRs mediate DNA binding by the TFs Msn2 and Yap1. In these proteins, numerous weak interactions between IDRs and target DNA inform promoter selection of the full-length protein's DBD. Rather than direct DNA-IDR interactions, our model suggests that non-covalent interactions between IDRs and the HD dictate overall promoter binding in Hox proteins.

REFERENCES

Beachy, P. A., Varkey, J., Young, K. E., von Kessler, D. P., Sun, B. I., & Ekker, S. C. (1993). Cooperative binding of an Ultrabithorax homeodomain protein to nearby and distant DNA sites. *Molecular and Cellular Biology, 13(11),* 6941-6956.

Beh, C. Y., El-Sharnouby, S., Chatzipli, A., Russell, S., Choo, S. W., & White, R. (2016). Roles of cofactors and chromatin accessibility in Hox protein target specificity. *Epigenetics and Chromatin. 9(1).*

Bondos, S. E., & Tan, X. X. (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),* 145-171.

Bondos, S. E., Catanese, D. J., Tan, X. X., Bicknell, A., Li, L., & Matthews, K. S. (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),* 26433–26444.

Bondos, S. E., Tan, X. X., & Matthews, K. S. (2006). Physical and genetic interactions link Hox function with diverse transcription factors and cell signaling proteins. *Molecular and Cellular Proteomics, 5(5),* 824-834.

Bomze, H. M., & Lopez, A. J. (1994). Evolutionary conservation of the structure and expression of alternatively spliced Ultrabithorax isoforms from Drosophila. *Genetics, 136(3),* 965-977.

Breau, M. A., Wilkinson, D. G., & Xu, Q. (2013). A Hox gene controls lateral line cell migration by regulating chemokine receptor expression downstream of Wnt signaling. *PNAS, 110 (42),* 16892-16897.

Bremer, A., Henn, C., Engel, A., Baumeister, W., & Aebi, U. (1992). Has negative staining still a place in biomacromolecular electron microscopy? *Ultramicroscopy, 46 (1992),* 85-111.

Brenner, S., & Horne, R. W. (1959). A negative staining method for high resolution electron microscopy of viruses. *BBA - Biochimica et Biophysica Acta, 34,* 103-10.

Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D. K., Carmi, M., & Barkai, N. (2020). Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Molecular Cell, 79,* 459–471.

Capovilla, M., & Botas, J. (1998). Functional dominance among Hox genes: Repression dominates activation in the regulation of dpp. *Development, 125,* 4949-4957.

Chen, F., & Capecchi, M. R. (1999). Paralogous mouse Hox genes, Hoxa9, Hoxb9, and Hoxd9, function together to control development of the mammary gland in response to pregnancy. *Proceedings of the National Academy of Sciences of the United States of America, 96 (2),* 541-546.

Churion, K. A. (2017) Hox DNA binding: from monomers to fibers. Texas A&M University, College Station, Texas.

Churion, K., Liu, Y., Hsiao, H. C., Matthews, K. S., & Bondos, S. E. (2014). Measuring Hox-DNA binding by electrophoretic mobility shift analysis. *Methods in Molecular Biology,* 1196, 211-230.

Dash, A. B., Orrico, F. C., & Ness, S. A. (1996). The EVES motif mediates both intermolecular and intramolecular regulation of c-Myb. *Genes and Development, 10,* 1858-1869.

Denell, R. E., Hummels, K. R., Wakimoto, B. T., & Kaufman, T. C. (1981). Developmental studies of lethality associated with the Antennapedia gene complex in Drosophila melanogaster. *Developmental Biology, 81,* 43-50.

Drake, J. A., & Pettitt, B. M. (2015). Force field-dependent solution properties of glycine oligomers. *Journal of Computational Chemistry, 36(17),*1275-85.

Draganescu, A., Levin, J. R., & Tullius, T. D. (1995). Homeodomain proteins: What governs their ability to recognize specific DNA sequences? *Journal of Molecular Biology, 250,* 595–608.

Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., & Obradović, Z. (2002). Intrinsic disorder and protein function. *Biochemistry, 41(21),* 6573–6582.

Franken, L. E., Grünewald, K., Boekema, E. J., & Stuart, M. C. A. (2020). A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. *Small, 16 (14),* 1906198.

Galant, R., & Carroll, S. B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. *Nature, 415,* 910-913.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Wüthrich, K. (1994). Homeodomain-DNA recognition. *Cell, 78(2),* 211-223.

Gibson, G., & Gehring, W. J. (1988). Head and thoracic transformations caused by ectopic expression of Antennapedia during Drosophila development. *Development, 102,* 657-675.

Grenier, J. K., & Carroll, S. B. (2000). Functional evolution of the ultrabithorax protein. *Proceedings of the National Academy of Sciences of the United States of America, 97(2),* 704–709.

Heymann JB and Belnap DM (2007) Bsoft: Image processing and molecular modeling for electron microscopy. *Journal of Structural Biology, 157(1),* 3-18.

Hoey, T., & Levine, M. (1988). Divergent homeo box proteins recognize similar DNA sequences in Drosophila. *Nature, 332(6167),* 858-861.

Holehouse, A. S., & Pappu, R. V. (2018). Collapse Transitions of Proteins and the Interplay among Backbone, Sidechain, and Solvent Interactions. *Annual Review of Biophysics, 47,* 19-39.

Hollenhorst, P. C., McIntosh, L. P., & Graves, B. J. (2011). Genomic and biochemical insights into the specificity of ETS transcription factors. *Annual Review of Biochemistry. 80,* 437-471.

Howell, D. W., Tsai, S. P., Churion, K., Patterson, J., Abbey, C., Atkinson, J. T., Porterpan, D., You, Y. H., Meissner, K. E., Bayless K. J., & Bondos, S. E. (2015). Identification of multiple dityrosine bonds in materials composed of the Drosophila protein Ultrabithorax. *Advanced Functional Materials, 25,* 5988–5998.

Hughes, C. L., & Kaufman, T. C. (2002). Hox genes and the evolution of the arthropod body plan. *Evolution and Development, 4(6),* 459–499.

Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A., Jacob, V., Aggarwal, A. K., Honig, B., & Mann, R. S. (2007). Functional Specificity of a Hox Protein Mediated by the Recognition of Minor Groove Structure. *Cell, 131(3),* 530–543.

Kalionis, B., & O'Farrell, P. H. (1993). A universal target sequence is bound in vitro by diverse homeodomains. *Mechanisms of Development, 4773(93),* 90023-Q.

Latchman, D. S. (1997). Transcription factors: an overview. *The International Journal of Biochemistry & Cell Biology, 29(12),* 1305-1312.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature, 276(5688),* 565-70.

Lewis, E. B. (2004). The bithorax complex: The first fifty years. *Genes, Development, and Cancer: The Life and Work of Edward B. Lewis,* 503-526.

Li, X. Y., & Green, M. R. (1996). Intramolecular inhibition of activating transcription factor-2 function by its DNA-binding domain. *Genes and Development, 10,* 517-527.

Liu, Y., Matthews, K. S., & Bondos, S. E. (2008). Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila Hox protein Ultrabithorax. *Journal of Biological Chemistry, 283(30),* 20874-20887.

Liu, Y., Matthews, K. S., & Bondos, S. E. (2009). Internal Regulatory Interactions Determine DNA Binding Specificity by a Hox Transcription Factor. *Journal of Molecular Biology, 390(4),* 760-774.

Lu, Y. F., Goldenberg, I., Bei, L., Andrejic, J., & Eklund, E. A. (2003). HoxA1O Represses Gene Transcription in Undifferentiated Myeloid Cells by Interaction with Histone Deacetylase 2. *Journal of Biological Chemistry, 278 (48),* 47792–47802.

Makhijani, K., Kalyani, C., Srividya, T., & Shashidhara, L. S. (2007). Modulation of Decapentaplegic gradient during haltere specification in Drosophila. *Developmental Biology, 302 (1),* 243-255.

Mann, R. S. (1995). The specificity of homeotic gene function. *BioEssays, 17 (10),* 855-863.

Mann, R. S., Lelli, K. M., & Joshi, R. (2009). Chapter 3 Hox Specificity. Unique Roles for Cofactors and Collaborators. *Current Topics in Developmental Biology, 88,* 63-101.

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

Monier, B., Astier, M., Sémériva, M., & Perrin, L. (2005). Steroid-dependent modification of Hox function drives myocyte reprogramming in the Drosophila heart. *Development, 132,* 5283-5293.

Pabo, C. (1992). Transcription Factors: Structural Families and Principles of DNA Recognition. *Annual Review of Biochemistry, 61,* 1053-1095.

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., & Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature, 397,* 714–719.

Pearson, J. C., Lemons, D., & McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nature Reviews Genetics, 6(12),* 893-904.

Prin, F., Serpente, P., Itasaki, N., & Gould, A. P. (2014). Hox proteins drive cell segregation and non-autonomous apical remodelling during hindbrain segmentation. *Development (Cambridge), 141,* 1492-1502.

Rames, M., Yu, Y., & Ren, G. (2014). Optimized negative staining: A high-throughput protocol for examining small and asymmetric protein structure by electron microscopy. *Journal of Visualized Experiments, (90),* 51087.

Reed, H. C., Hoare, T., Thomsen, S., Weaver, T. A., White, R. A. H., Akam, M., & Alonso, C. R. (2010). Alternative splicing modulates Ubx protein function in Drosophila melanogaster. *Genetics, 184(3),* 745-U192.

Ronshaugen, M., McGinnis, N., & McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. *Nature, 415,* 914-917.

Shanmugam, K., Featherstone, M. S., & Saragovi, H. U. (1997). Residues flanking the HOX YPWM motif contribute to cooperative interactions with PBX. *Journal of Biological Chemistry, 272 (30),* 19081-19087.

Shen, W., Chrobak, D., Krishnan, K., Lawrence, H. J., & Largman, C. (2004). HOXB6 protein is bound to CREB-binding protein and represses globin expression in a DNA binding-dependent, PBX interaction-independent process. *Journal of Biological Chemistry, 279 (38),* 39895–39904.

Simister, P. C., Schaper, F., O'Reilly, N., McGowan, S., & Feller, S. M. (2011). Self-organization and regulation of intrinsically disordered proteins with folded N-termini. *PLoS Biology, 9(2).*

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H. J., Mann, R. S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between hox proteins. *Cell, 147,* 1270-1282.

Staby, L., O'Shea, C., Willemoës, M., Theisen, F., Kragelund, B. B., & Skriver, K. (2017). Eukaryotic transcription factors: Paradigms of protein intrinsic disorder. *Biochemical Journal, 474,* 2509–2532.

Tan, X. X., Bondos, S., Li, L. K., & Matthews, K. S. (2002). Transcription activation by ultrabithorax Ib protein requires a predicted alpha-helical region. *Biochemistry, 41(8),* 2774-2785.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. & Ludtke, S. J., 2007, EMAN2: an extensible image processing suite for electron microscopy. *Journal of Structural Biology, 157(1),* 38-46

Tong, X., Hrycaj, S., Podlaha, O., Popadic, A., & Monteiro, A. (2014). Over-expression of Ultrabithorax alters embryonic body plan and wing patterns in the butterfly *Bicyclus anynana. Developmental Biology, 394 (2),* 357-366.

Uversky, V. N. (2012). Size-exclusion chromatography in structural analysis of intrinsically disordered proteins. *Methods in Molecular Biology, 896,* 179-194.

Wagner, G. P., Amemiya, C., & Ruddle, F. (2003). Hox cluster duplications and the opportunity for evolutionary novelties. *Proceedings of the National Academy of Sciences of the United States of America, 100(25),* 14603–14606.

Wellik, D. M. (2007). Hox patterning of the vertebrate axial skeleton. *Developmental Dynamics, 236,* 2454-2463.

APPENDIX A

FULL LENGTH UBX PURIFICATION

1 Column Purification

5 mL of Ni^2-NTA Qiagen resin was preequilibrated with 80 mL wash buffer containing 500 mM NaCl, 10 mM BME, 5% glucose, 5 mM imidazole, and 50 mM $NaH_2\ PO_4$, and the pH was adjusted to 8.0. 1 liter of Ubx cell pellet was thawed at room temperature and lysed for 25 min in 40 mL of buffer containing 800 mM NaCl, 5% glucose, 5 mM imidazole, 50 mM $NaH_2\ PO_4$, 10 mM BME, 180 µL DNase I, 300 µL 200mM PMSF, a small amount of lysozyme, and 2 tablets of Complete Proteinase Inhibitor Mixture (Roche), at pH 8.0.

The lysate was centrifuged at 14,000 rpm and 4°C for 8 min. The supernatant was treated with polyethyleneimine and re-centrifuge an additional 8 minutes. The resulting supernatant's pH was adjusted to 8.0 using $NaH_2\ PO_4$ and was centrifuged for another 8 minutes.

The supernatant was incubated with the equilibrated resin at 4°C for one minute before being poured into the column and allowed to flow through slowly. The packed resin was washed with 80 mL of 5 mM imidazole wash buffer, followed by 80 mL of 20 mM imidazole wash buffer, and 20 mL of 80 mM imidazole wash buffer. Protein was eluted with 200 mM imidazole wash buffer.

2 Column Purification

2L of Ubx cell pellet were lysed in 40 mL of buffer containing 800 mM NaCl, 50 mM Tris–HCl pH 8.0, 100 mM arginine, 10 mM BME, 2 tablets of Complete Proteinase Inhibitor Mixture (Roche), 300 µL 200mM PMSF, a small amount of lysozyme, and 320 µL 20 mg/mL DNase I stock. After incubating for 25 minutes at room temperature, the resulting lysate was centrifuged at 4°C and 14,000 rpm for 10 minutes. The lysis supernatant was treated with PEI and re-centrifuged at 4°C and 14,000 rpm for 8 minutes. The pH of the supernatant was adjusted to 6.8 with NaH_2PO_4 and centrifuged at 4°C and 14,000 rpm for 8 minutes.

The final supernatant was diluted to 250 mL with buffer Z containing 150 mM NaCl, 25 mM NaH_2PO_4, 5% glucose, 100 mM arginine, 10 mM BME, and 0.1 mM EDTA at a pH of 6.8. The diluted supernatant was loaded onto a P11 phosphocellulose column (Whatman) pre-equilibrated with 200 mL of buffer Z. The resin was washed with 300 mL buffer Z, and Ubx was eluted with 50mL of buffer Z plus 1M NaCl.

Fractions containing Ubx were collected and dialyzed in 4 L of dialysis buffer containing 5% glucose, 10 mM BME, 150 mM NaCl, and 50 mM Tris, pH = 8.0 for 25-30 minutes. The sample was then poured on 4 mL Ni-NTA agarose resin (Qiagen) pre-equilibrated with 100 mL dialysis buffer and, once all liquid had run through, washed with:

1) 50 mL wash buffer containing 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM BME, 5% glucose, and 10 mM imidazole

2) 8 mL of 20 mM imidazole wash buffer

Protein was eluted with 14 mL of 200 mM imidazole wash buffer.

APPENDIX B

DEVELOPMENT OF THE 2-COLUMN PURIFICATION PROCEDURE

The 2-column purification used in this thesis was adapted from Churion, 2017, with the following initial adjustments: Lysis buffer was doubled to 40mL, EDTA was removed, DNase was added immediately, and PMSF, lysozyme, and 10 mM BME were added. Buffer Z was adjusted to 100 mM NaCl, and glucose was used instead of glycerol. The 100 mL dialysis buffer wash was removed from the second purification, and the protein-resin incubation was reduced to about one minute on the column.

Under these conditions, we found that Ubx did not elute well over the NaCl gradient; concentrations in the elutions were poor, and an SDS-page gel revealed a large amount of protein matching Ubx's molecular weight remaining on the resin (Fig. 14).

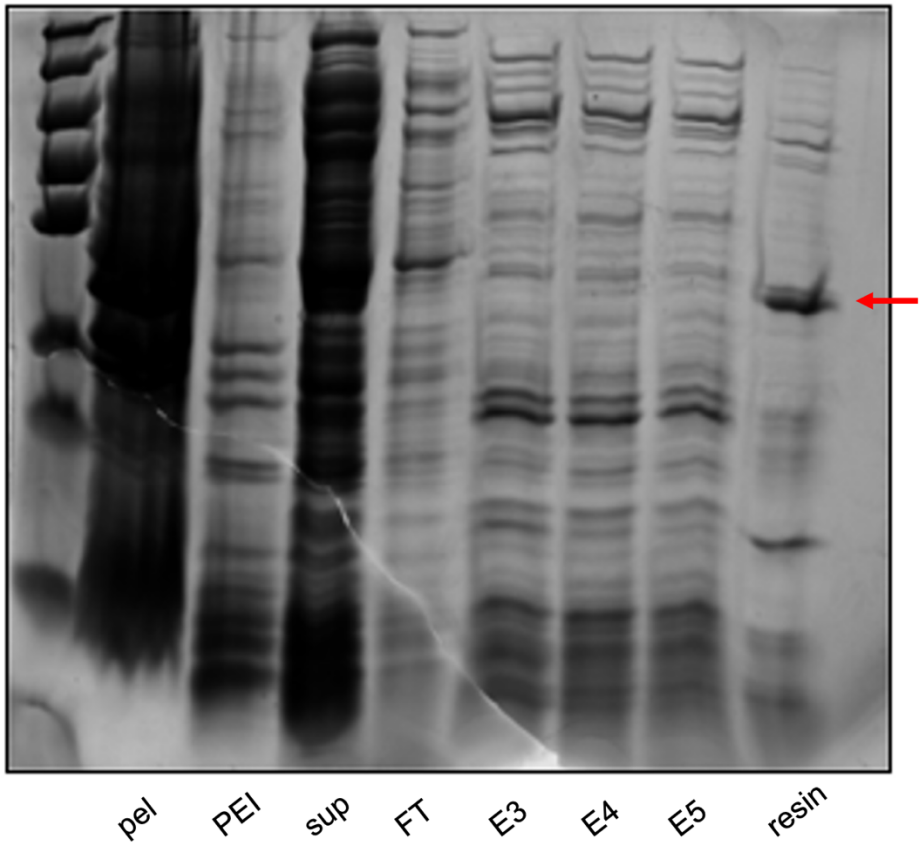

Figure 14 Coomasie stain of samples from phosphocellulose purification.
Lanes correspond to samples from the pellet, the PEI pellet, the supernatant, column flow through, elutions 3-5, and the resin. The red arrow indicates the level at which Ubx appears on the gel.

This problem was overcome by foregoing the gradient and switching to an elution buffer containing buffer Z with 1M NaCl. When eluting with this condition, Ubx tended to elute faster, appearing in elutions 2 through 5, and was more concentrated.

During this and subsequent purifications, we also noted that a large amount of Ubx was retained in the pellet, thus limiting the amount of soluble protein we could recover during the elution step (Fig. 15).

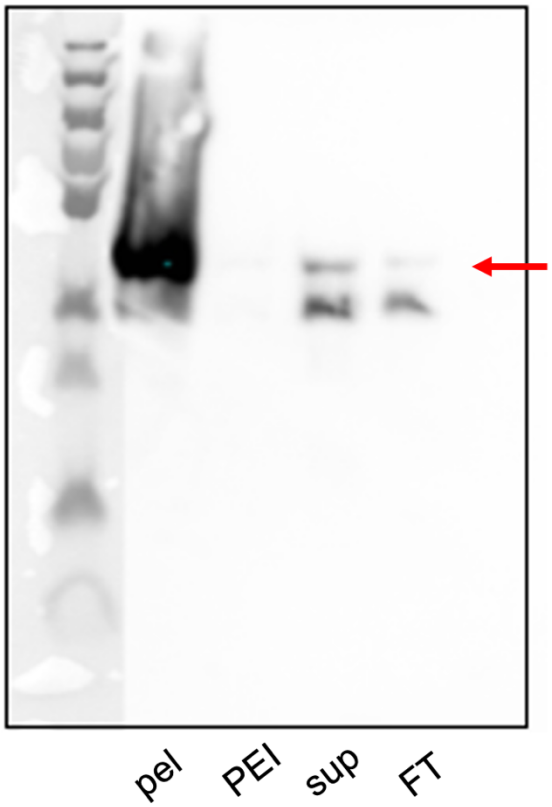

Figure 15 Western blot of samples collected during phosphocellulose purification. Protein was detected using a primary antibody that binds the Ubx HD. Lanes correspond to samples taken from the primary pellet, the PEI pellet, the lysis supernatant, and the column flow through. The red arrow indicates the level of Ubx.

We tested three buffer conditions to improve Ubx solubility; lysis buffer plus 200 mM urea, lysis plus 50 mM arginine, and lysis plus 200 mM arginine. 0.5 L of *E. coli* cell pellet were tested with each condition, and samples from each condition were analyzed via Western blot (Fig. 16).

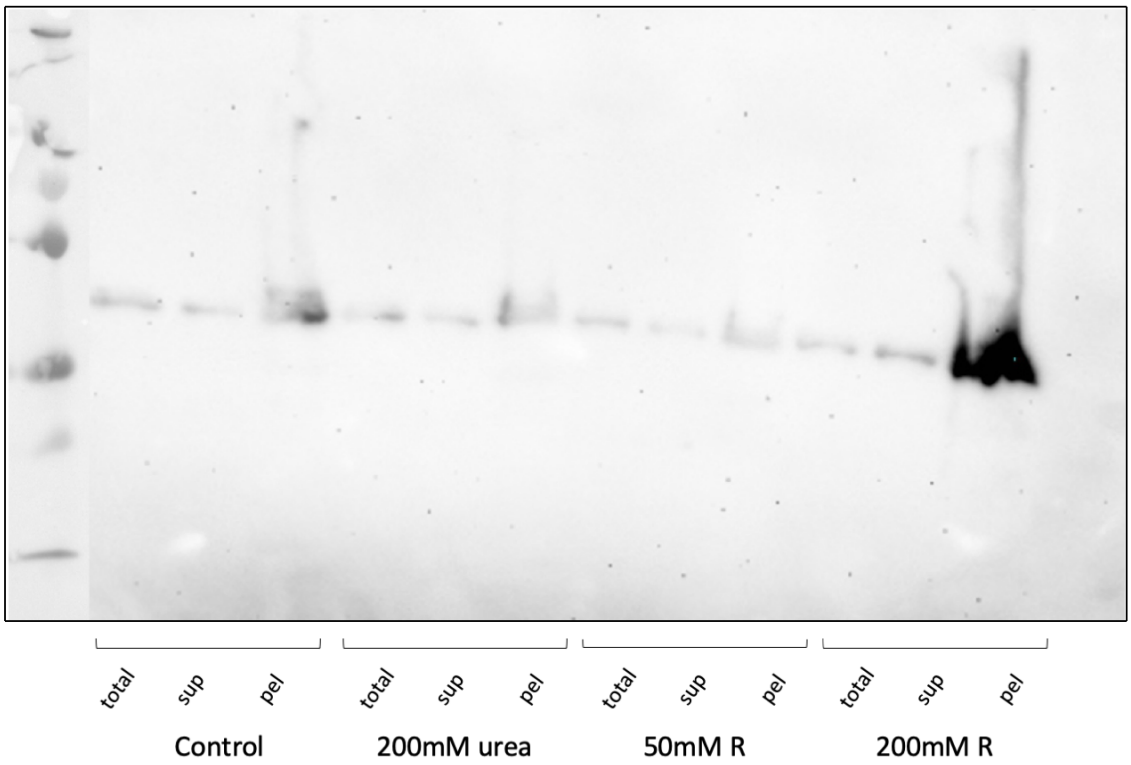

Figure 16 Western blot of samples collected from different lysis conditions. Protein was detected using a primary antibody that binds the Ubx HD.

The 200mM arginine buffer had the greatest amount of Ubx present in the supernatant. However, this sample also had the loosest pellet, which impeded efficient collection of the supernatant (it is possible that some of the Ubx seen in the pellet lane corresponds to uncollected supernatant). Additionally, arginine is a positively charged amino acid, so large amounts may interfere with Ubx-phosphocellulose binding and

cause protein loss in the flow through. For these reasons, 100 mM arginine was used in subsequent purifications.