PROTEIN INTERACTIONS IN REGULATION AND ASSEMBLY

A Dissertation

by

HAO-CHING HSIAO

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,	J. Martin Scholtz
	Rajesh Miranda
	Jun-yuan Ji
Intercollegiate Faculty Chair,	Craig J. Coates

May 2015

Major Subject: Genetics

ABSTRACT

The objectives of this work include: validation of yeast-based assays, investigation of protein-protein interactions in the regulatory role of an intrinsically disordered protein, Ultrabithorax (Ubx), and exploration of possible application of Ubx self-assemble protein materials for cell culture study.

Yeast based assays are useful for detecting DNA-protein and protein-protein interactions. Although yeast-based assays are sensitive, substantial amount of artifacts often hinder the correct interpretation of the results. The first part of the project was focused on the validation of yeast one-hybrid and two- hybrid assay and assessment for the possible source of artifacts. We found that media sources and preparation methods had significant influence on the results to yeast-one-hybrid and yeast-two hybrid assay.

Using condition optimized yeast-two hybrid assay to match *in vivo* experiments, we examined protein interactions formed by Ubx, a protein with intrinsically disordered regions. In the classical protein structure-function paradigm, protein function is determined by particular protein structure. However, some proteins do not have rigid structure, and these proteins extensively interact with proteins with diverse functions. The second part of this dissertation was focused on dissecting the topological features of binding partner of an intrinsically disordered protein, Ubx. We found that the binding partners of Ubx protein were enriched in particular folds and disordered regions on Ubx protein were essential for Ubx-partner protein interactions, which may explain why proteins with multiple binding partners often have disordered regions.

Ubx protein not only interacts with other proteins but also with itself. This self-association has been explored for the application for tissue engineering purposes. The cytocompatible and biocompatible characteristics of Ubx self-assemble protein materials could provide an unmet need for 3-dimensional (3D) scaffolds on which to culture cancer cells for carcinogenesis characterization. The third part of this dissertation was focused on determining the compatibility between isogenic human mammary gland cell lines with various degrees of tumorigenesis and characterization of cell behavior on Ubx materials. We found that Ubx materials were not toxic to human mammary gland cells with similar genetic background but different degree of tumorigenesis. In addition, Ubx materials enabled further characterization of cancer cell specific cell behaviors.

DEDICATION

To my friends, family, and people who have helped me reach this point in my life.

Grandma Chen, this is for you. Sorry, I missed the chance to say " Goodbye" to you.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Bondos, and my committee members, Dr. Scholtz, Dr. Ji, Dr.Miranda, and Dr. Coates, for their guidance and support throughout the course of this dissertation study.

Thank you to my friends, colleagues and the department faculties and staffs for making my time at Texas A&M University a great experience.

Thank you to Robbie Schultz, who is always there for me at the most difficult time.

Thank you to my family, particularly to my dad, my mom, and my sister, for their supports and sacrifice.

Finally, thank you to the God we trust.

NOMENCLATURE

2D	2-dimensional
3D	3-dimensional
Abd-A	Abdominal-A
AkUbx	*Akanthokara kaputensis* Ultrabithorax
Al	Aristaless
Aly	Always early
Antp	Antennapedia
Apt	Apontic
Arm	Armadillo
BRCA	Breast cancer 1
CBP	CREB-binding protein
CBP80	Cap-binding protein 1
c-Ha-ras	p21 protein coding gene
CM	Challenged mouse
CNS	Central nervous system
CREB	cAMP response element-binding protein
CycK	Cyclin K
DAPI	6-diamidino-2-phenylindole
Dfd	Deformed
DIC	Differential interference contrast

DIP1	Disconnected-interacting protein 1
Dll	Distal-less
DMEM/F12	Dulbecco's Modified Eagle Medium/ Nutrient Mixture F-12
DmUbx	*Drosophila melanogaster* Ultrabithorax
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's phosphate-buffered saline
Dsh	Dishevelled
DTT	Dithiothreitol
ECs	Endothelial cells
EF1γ	Elongation factor 1γ
Ef2b	Elongation factor 2b
EGFP	Enhanced Green Fluorescent Protein
eIF4E	Ribosome recruiting protein RNA
eIF4E-BP2	Eukaryotic translation initiation factor 4E-binding protein 2
ELISA	Enzyme-linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
En	Engrawolberiled
Exd	Extradenticle
FP3.38	Anti-homeodomain antibody
Fzo	Fuzzy onions
GPa	Giga pascal
HCV	Hepatitis-C Virus

HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hox	Homeobox
HRP	Horseradish peroxidase
HSC70-4	Heat shock protein cognate 4
Hth	Homothorax
IgG	Immunoglobulin G
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kd	Equilibrium dissociation constant
kDa	Kilo dalton
Keap1	Kelch-like ECH-associated protein 1
KP	Potassium phosphate
Mdm2	Mouse doubles minute 2 homolog/ E3 ubiquitin-protein ligase
Meis	Meis homeobox protein
MMP	Matrix metalloproteases
MMPs	Matrix metalloproteases
mRpL44	Mitochondrial ribosomal protein L44
Nmo	Nemo
NMR	Nuclear magnetic resonance
NO	Nitric oxide
Noc	No ocelli
Nrf2	Nuclear factor erythroid 2-related factor 2
Nrt	Neurotactin

NS5A	Non-structural protein 5A
MoRF	Molecular recognition feature
ONP	o-Nitrophenol
ONPG	ortho-nitrophenyl-β-galactoside p21
Otu	Ovarian tumor
p120ctn	Adherens junction protein p120
PBC	Pre B cell proteins
Pbx	Pre-B cell leukemia transcription factor
PCs	Pericytes
PK17E	Protein kinase-like 17E
PONDR	Predictor of Naturally Disordered Regions
PRM	Proline-rich motif
ProTα	Prothymosin α
PS6	Parasegment 6
RpL22	Ribosomal protein L22
Rpn6	Protease p44.5 subunit
Rps13	Ribosomal protein S13
Sal	Spalt
SCOP	Structural Classification of Proteins
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH3	SRC homology 3 domain
SMCs	Smooth muscle cells

Smox	Smad on X
SRC	Proto-oncogene tyrosine-protein kinase
TALE	Three amino acid loop extension
Term	Terminus
TFIIEβ	Transcription factor IIEβ
TGFβ-1	Transforming growth factor β-1
TNF-α	Tumor necrosis factor-α
Trn	Transportin
Ubx	Ultrabithorax
UM	Unchallenged mouse
VAMP	Vesicle-associated membrane protein
VAP	Vesicle-associated membrane protein (VAMP)-associated protein
VAPB	Vesicle-associated membrane protein-associated protein B
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
YNB	Yeast nitrogen base
Zf30C	Zinc finger protein 30
Zn72D	Zinc-finger protein at 72D

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Introduction

Proteins control and coordinate biological functions in living organisms. Protein-protein interactions often regulate diverse *in vivo* functions including, but not limited to: transcription regulation, cell signaling, chemical analysis, and cell structure. A single protein may execute different functions: for example, many transcription factors repress and activate transcription in a context-specific manner [1-4]. How can a single protein execute multiple functions? One possibility is that protein function can be regulated by protein-protein interactions [5-7]. However, many proteins without rigid structure, called intrinsically disordered proteins, have multiple regulatory and signal transduction functions [8-11].

The first part of this literature review will focus on discussing different types of protein-protein interactions, the role of intrinsically disordered regions in facilitating context-specific function of Homeotic (Hox) protein, and molecular functions of a Hox protein, Ultrabithorax (Ubx). In addition to the endogenous biological functions executed through protein-protein, protein-protein interaction is essential to material formation and different materials formations can be observed among different protein building blocks. The second part of this literature review will focus on the general review on a variety of protein based materials and Ubx self-assemble protein materials.

Moiety of Protein- Protein Interactions

The structure function paradigms states that the unique amino acid sequence provide the unique structure of a protein, which determines the function of this protein. When a structured protein interacts with another structured protein requires the complimentary shape and proper surface chemistry, a concept is known as “lock and key ” model, a concept extended from enzymatic activity to macromolecule interactions [7,12]. The interface is composed of complementary shapes, and stabilizing hydrophobic hydrogen bonds, and ionic interactions. This type of protein interactions is commonly involved in critical biological functions such as enzymatic catalysis, immunological recognition, and receptor/ligand interactions [13-15]. However, transcription factors and cell signaling proteins often have disordered regions [11]. The following section will focus on reviewing literatures describing the role of disorder in protein-protein interaction.

Structured-disordered or disordered-disordered protein interaction

Proteins without rigid structure also carry out important biological functions such as cell signaling and transcription regulation. Statistical data has shown 66% of cell signaling proteins and >80 % of transcription factors are intrinsically disordered or have large disordered regions [11,16]. In addition, intrinsically disordered proteins are often involved in human diseases. Mis-regulation in intrinsically disordered proteins often causes catastrophically consequences [17]. These relationships seen in: p53, p21, and BRCA in cancer, prion protein in prion diseases, α-synuclein in synucleinopathies, and Tau protein in Alzheimer's disease etc. [17-19]. It is reasonable to wonder what

characteristics of intrinsically disordered proteins caused so many diseases. The inherited structural flexibility of intrinsically disordered protein provides the significant advantage in protein-protein interaction, which is the binding promiscuity [20,21]. Intrinsic disorder provides the conformational plasticity for biological functions such as transcription regulation and cell signal transduction [11,22]. Furthermore, intrinsically disordered regions are modified by posttranslational modifications and alternative splicing [23]. The flexibility of the intrinsically disordered proteins presents the higher probability for the required biological modification on the peptides or amino acids which ultimately imprint the functions of the protein. Ultimately, these modifications have the potential to regulate partner binding by the intrinsically disordered protein.

A disordered protein can form extensive interactions with numerous proteins, known as the "one to many" concept [20,24]. The p53 protein is a good example because it has been extensively studied with detailed structural data available on multiple partner interactions [25-27]. The malfunction of p53 is linked to carcinogenesis including but not limited to colon, lung, esophagus, breast, liver, brain, and hemopoietic tissues [25,28-33]. Large disordered regions can be found at the N- terminal domain and C-terminal domain of p53 protein (**Figure 1.1**). A comparison of order/disorder tendencies in p53 as revealed by PONDR disorder prediction algorithm together large scale bioinformatics study reveal that over 70% of p53 partners interact with disordered regions [20]. Individual disordered regions in p53 can change their conformations when they interact with different partners [18]. For example, the disordered C-terminal of p53 interacts with multiple proteins by conformational changes at individual disordered

regions, which has gone through a structural transition from disordered to secondary structures [34-36]. The type of secondary structure formed by this single disordered region include: a helix when associating with S100ββ, a β sheet with sirtuin, and distinct irregular structures with CBP and cyclin A2 [20] (**Figure 1.2**).

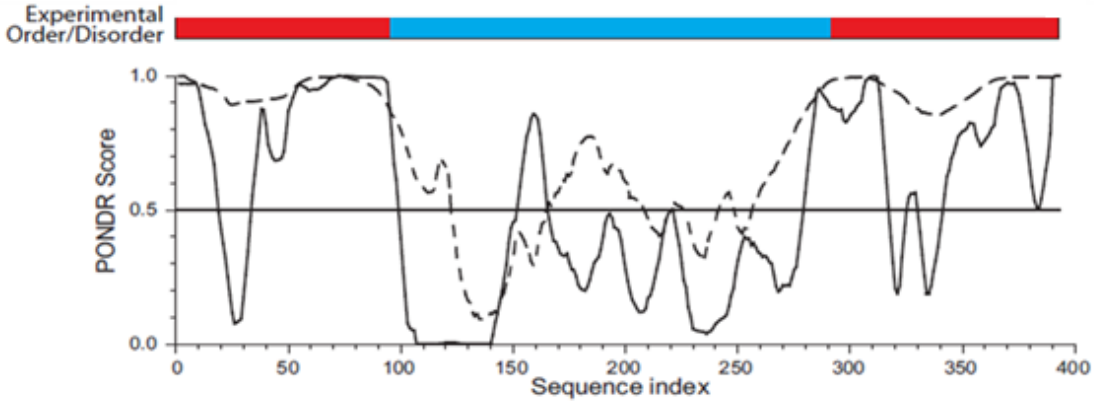

Figure 1.1 p53 protein disorder predictions. Two different prediction algorithms in PONDR predictors, VLXT and VSL2P, are shown as solid line and dashed line, respectively [20].

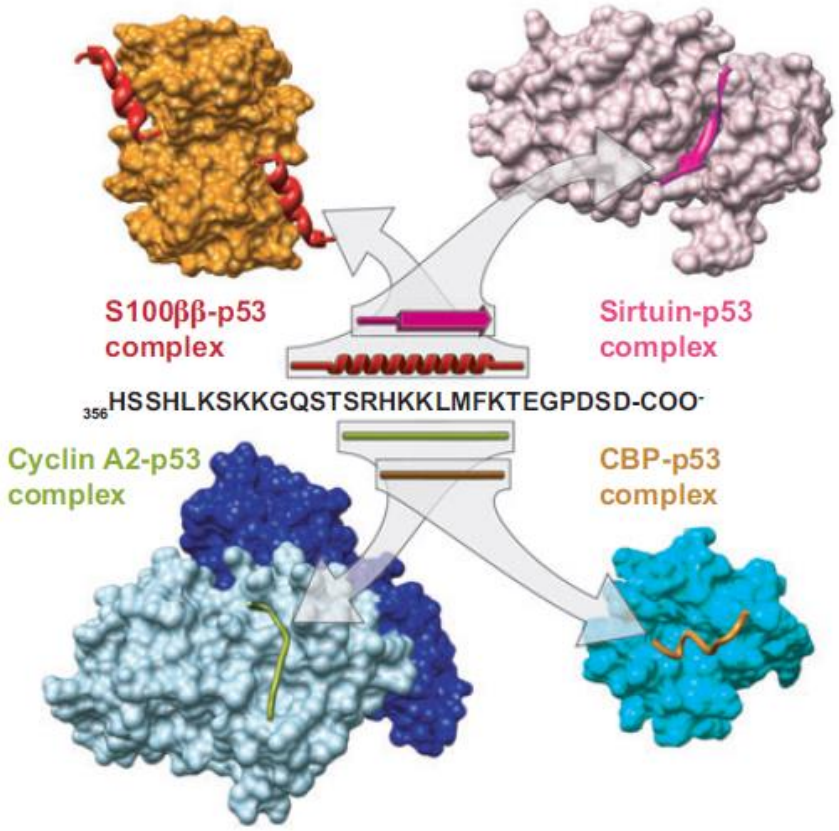

Figure 1.2 Conformational changes of single disordered regions in the C-terminal of p53. Primary, secondary, and quaternary structure of p53 disordered region complexes are depicted for each [18].

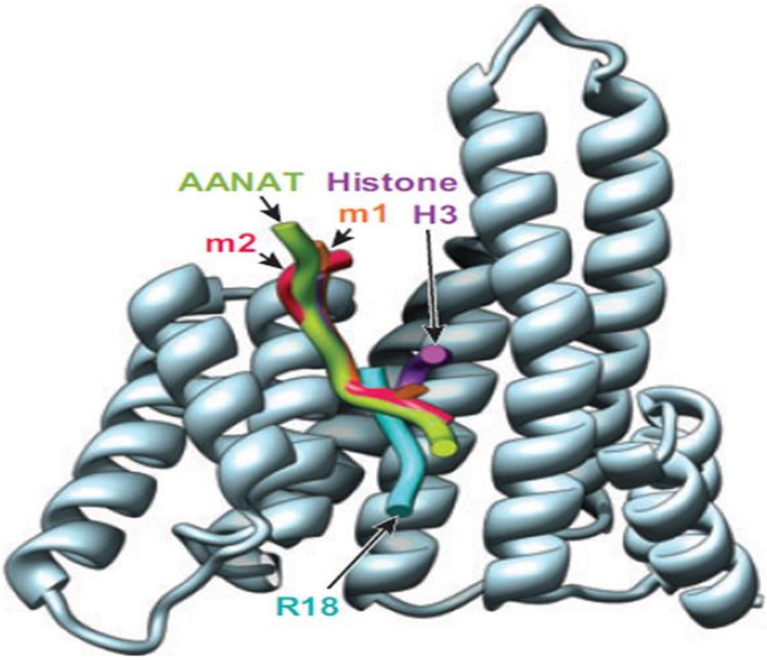

Figure 1.3 Alignment of multiple disordered sequences with single binding groove in 14-3-3ξ domain. This complex include peptide from histone H3, serotonin *N*-acetyltransferase (AANAT), a phage display-derived peptide (R18), and motif 1 (m1) and motif 2 (m2) peptides [20].

In addition to one disordered region interacting with multiple proteins, many disordered regions can also bind an individual structured region. A cell signaling protein with large disordered regions, 14-3-3 protein family, which is conserved among eukaryotic cells, interacts with more than 200 proteins such as kinases, proteases, and phosphatases to execute diverse regulatory functions including but not limited to signal transduction, apoptosis, cell cycle progression, DNA replication, and cell malignant transformation [37,38]. About 90% of 14-3-3 proteins interacting partners have at least one intrinsically disordered region [39]. **Figure 1.3** shows that five disordered region from other binding

sequences associated with single structured domain on 14-3-3. The wide array of partner proteins indicate that the intrinsically disordered characteristics of 14-3-3 protein indeed provides extensive binding ability for 14-3-3 [20].

Fuzzy complexes

Disordered proteins can also form transient and dynamic complexes with binding proteins. These complexes are considered as "fuzzy" complexes [40,41]. A stable protein binding complex requires the balance between the cost of entropy and gain of enthalpy which lead to the continuous search of the best docking interface and conformations. These conformations can be classified by the degree of dynamics of the protein complex [40]. For example, the association of Hepatitis-C Virus (HCV) non-structural protein 5A (NS5A) and VAPB, an RNA alternatively spliced isoforms of vesicle-associated membrane protein associated protein (VAP), form a fuzzy complex [42]. NS5A is involved in viral RNA replication with 2 large disordered regions including the domain 2 and domain 3 of NS5A. A dynamic complex is formed between the intrinsically disordered domain 3 of NS5A and the well-structured MSP domain of VAPB. Another good example is the dynamic protein complex formed by Prothymosin α (ProTα) and Kelch-like ECH-associated protein 1 (Keap1) [43]. Prothymosin α (ProTα) is an oncoprotein with most of peptides intrinsically disordered in unbound state. ProTα retains a high level of flexibility even in the bound state with Kelch, which is in agreement with the molecular dynamic (MD) simulations of the ProTα–Kelch complex [43].

Liquid-protein self-assembly

Multivalent interactions are involved in pivotal biological functions such as: extracellular matrix ligand-receptor interactions, intracellular signaling transduction, RNA metabolism and chromatin organization in the nucleus [44-46]. Furthermore, multivalent interaction could be the driving force in protein polymerization.

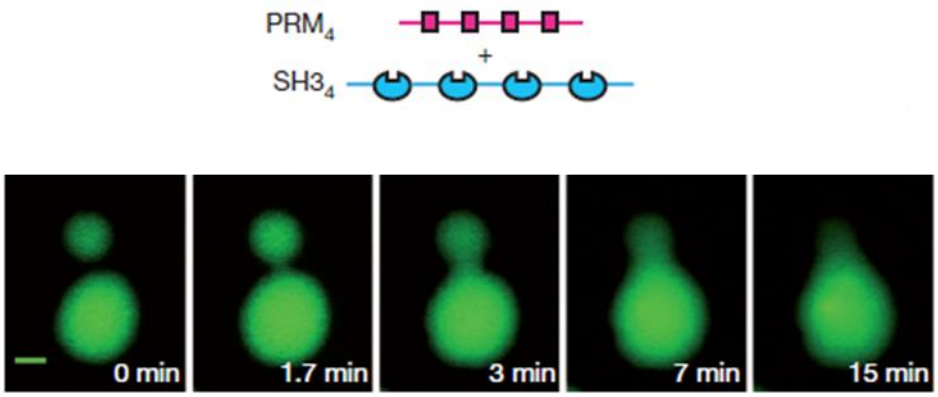

Figure 1.4 Time-lapse imaging of merging PRM_4-$SH3_4$ peptide droplets. The merging process of individual droplets composed of PRM_4 and $SH3_4$ peptides can be seen in the time-lapse pictures. Scale bar, 10 mm [47].

In polymer chemistry, valency and affinity of molecular control the phase transitions of monomeric species. Monomeric molecules have sharp phase transformation from small assembles to large gel like materials, known as the sol-gel transition. Similar interactions are also formed by intrinsically disordered protein interactions [44,47]. For example, SRC homology 3 (SH3) domain interact with its proline-rich motif (PRM) ligand generates hierarchal species in an concentration

dependent manner [44]. At critical concentration, SH3 and PRM ligand go through liquid-liquid de-mixing transition which generates micrometre-sized polymer droplets in buffer. Large scale polymerization can also be created by increasing concentration of monomeric SH3 peptide and PRM ligand for prolonged incubation (**Figure 1.4**). This observation may explain how intrinsically disordered proteins react to other proteins at extremely high concentrations and why they have modest affinity at interactions. In addition, the multivalent properties can also be used for protein with multiple partner proteins to execute the proper function at the right tissues and right time. The following review will provide introduce the concept of context-specific function using Hox protein family as an example, and Ubx protein, a Hox transcription factor, as the model protein for this research.

Context Specific Function

Homeotic genes regulate animal developmental function in multiple tissues and execute different functions dependent to tissue or/and developmental stage [48-50]. Mis-regulation of these genes likely causes serious consequences such as developmental deficiency which of represent the transformation of one structure at region of the body into another [51-56]. For example, ectopic expression of the Hox gene, Ubx, at the eye-antennal imaged disc of *Drosophila melanogaster* generates a pair of legs in place of the antennas are; however mis-expression of Ubx in the third thorax region causes the formation of extra wing (**Figure 1.5**) [57,58]. Therefore, in the eye-antennal imaginal disc, Ubx represses genes that generates antenna and activates genes required to

construct a leg. In contrast, in halteres, Ubx represses genes required to form the wing and activates genes required to form halteres. Thus Ubx regulates different groups of genes based on the tissue in which it is expressed.

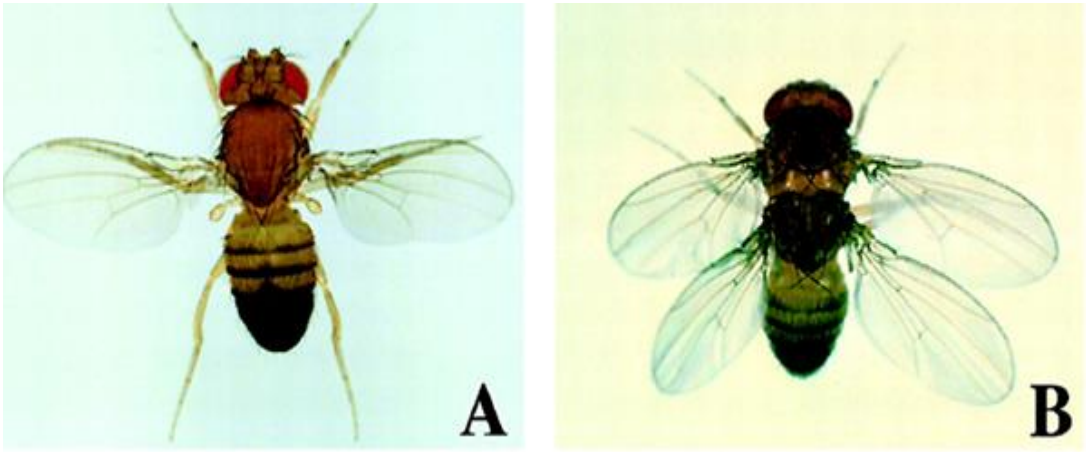

Figure 1.5 Phenotype of wild type and homeotic mutant *Drosophila melanogaster*. (A)Wild-type male *Drosophila melanogaster* have the haltere, small club-shaped organs located posterior to the wings at the 3rd thoracic segment. (B) The mis-expression of Ubx at the 3rd thoracic segment creates a mutant in an adult *Drosophila melanogaster* with extra pair of wing replaces the halters and 3rd thoracic segment [58].

Hox family as a model system for context specific studies

Hox proteins target a wide variety of genes in a stage specific manner such as: growth-factor-like molecules, membrane receptors, cell adhesion molecules, structural proteins and many other transcription factors [48,52,59,60]. Hox proteins can be found in bilateral organisms, in which the protein specifies the identity of the segment in bilaterally symmetric organisms [61]. The examples of how different tissues controlled

by evolutionally conserved Hox genes with functional overlap are shown in **Figure 1.6**. Mouse Hox 2.2 gene, an orthologue of Antp in mouse, effectively induces the homeotic transformation of *Drosophila* Antp in the thoratic segment in *Drosophila* embryos and larvae by Antp-Hox 2.2 chimera expression vector with a hybrid promoter. Mouse Hox 4B gene, an orthologue of Dfd is able to restore its orthologue Dfd's function in *Drosophila* embryos using a similar hybrid construct.

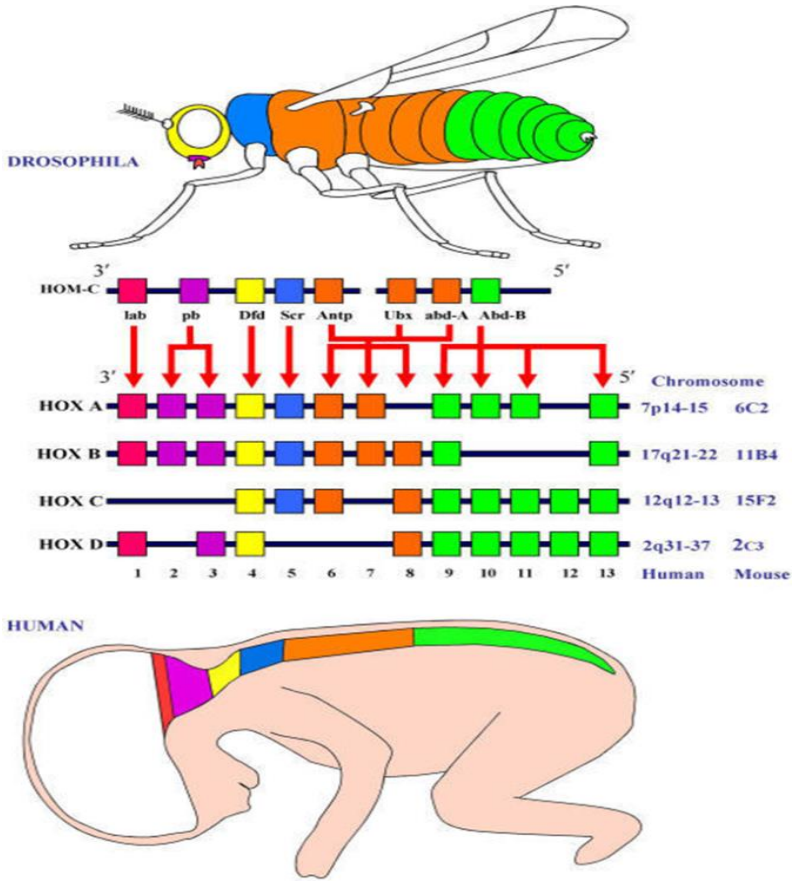

Figure 1.6 The four HOX gene clusters found in mammals are conserved from the *Drosophila* HOM-C complex in terms of nucleotide sequence and expression. During embryonic development, the genes (color blocks) are expressed in a pattern along the anterior to posterior axis that correlates with the position in the Hox gene cluster [62].

Hox protein cofactors and DNA binding specificity

Hox proteins recognize the 5'-TAAT-3' consensus sequence with relatively high DNA affinity [63-66]. The isolated DNA binding homeodomain of Hox proteins can bind the target DNA sequence with pM affinity [65,66]. Known for the low specificity due to the consensus DNA target sequence for most Hox proteins but requirement for a high degree of specificity *in vivo*, there are other factors in play of providing Hox protein with specificity in transcription regulation [67,68]. One of the possible mechanisms is through interacting with Hox cofactors [69,70]. A group of well-known Hox cofactors are the three amino acid loop extension (TALE) homeodomain proteins [71,72]. The TALE proteins have ability to binding DNA and they are evolutionally conserved in common model organisms. Two TALE proteins can be found in *Drosophila* include Extradenticle (Exd) and Homothorax (Hth). In contrast, there are four Exd-related proteins (Pbx1, Pbx2, Pbx3, Pbx4) and five Hth-related proteins (Meis1, Meis2, Meis3, Prep1, and Prep2) in the mouse [73]. Exd and Pbx proteins are also known as pre B cell (PBC) proteins.

Structural and *in vivo* studies have indicated a motif common to most Hox proteins, YPWM motif, makes direct contacts with the TALE motif in PBC homeodomain [71,74-79]. The Hox YPWM motif (also called Hexapeptide motif) forms a reverse turn and binds within a pocket formed by the 3^{rd} and 4^{th} helices of the divergent homeodomain of the Pbc protein[78]. In *Drosophila* Ubx, a motif C-terminal to the homeodomain mediates additional contacts with Exd [80]. Besides TALE family homeodomain proteins, the *Drosophila* Engrawolberiled (En) protein has also been

shown to be a Hox cofactor as well [81]. En bound cooperatively with both Ubx and Abd-A to a regulatory element from the Dll gene, and En input is required for Dll repression in the posterior compartments of the abdominal segments [81,82].

Ultrabithorax Molecular Functions and Features

The longest translational product of Ubx gene (Ubx Ib) has 389 amino acids, 60 of which compose the conserved homeodomain [79,83,84]. The regions of Ubx outside the homeodomain have significant biological functions such as transcription activation, protein interaction and transcription repression [Mann 1990, Galant 2002, Beachy 1993, Bondos 2004]. The primary function of the homeodomain is to interact with DNA. Previous studies have shown that Ubx homeodomain target DNA with consensus sequences of 5'- TAAT-3' with tight affinity as the value of $K_d = 7x10^{-11}$ M [84]. Ubx protein binds targeted DNA sequences cooperatively with each other [83]. Electrophoretic mobility shift assay (EMSA) using Ubx protein and oligonucleotides with 5'-TAAT-3' DNA sequence clusters have demonstrated the formation of multi-Ubx proteins composed complex by individual Ubx proteins [83]. Furthermore, the half-life of the Ubx DNA complex increase significantly for multi-site target DNAs. In addition to cooperative binding by individual Ubx proteins, like other Hox proteins, Ubx protein also binds DNA cooperatively with other cofactors for specific binding. For example, Exd protein facilitates the Ubx-DNA binding and vice versa to Hox-Exd composite DNA sites [85,86].

Ubx protein structure and intrinsically disordered feature

About 44% of the Ubx protein sequence is intrinsically disordered. The more structured homeodomain is composed of loops and helixes to form helix-loop-helix structure (**Figure 1.7**) [78]. Disordered regions are found throughout the remaining Ubx sequence with most of them situated at the transcription regulatory regions in Ubx [65]. These disordered regions have direct effect on Ubx-DNA binding affinity and specificity [66]. The full length Ubx protein has weaker DNA binding affinity than Ubx truncated mutants with homeodomain alone or mutants with shorter sequence surrounding the homeodomain region.

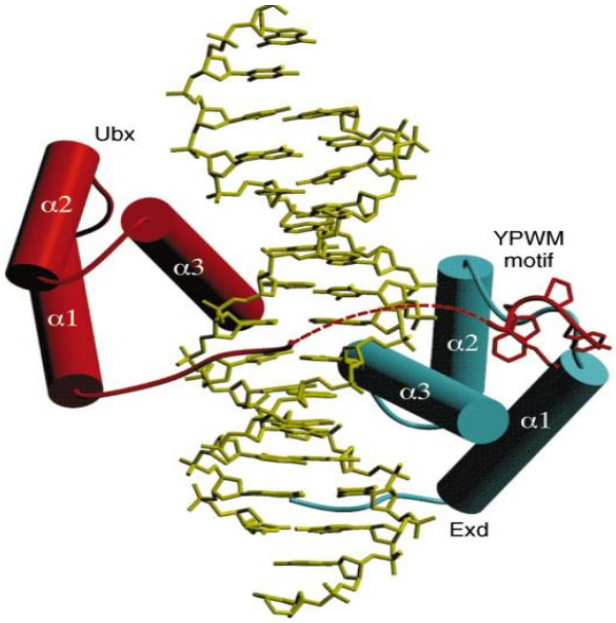

Figure 1.7 An overview of the Ubx -Exd-DNA ternary complex. The diagram shows the Ubx (red) and Exd (cyan) homeodomains binding in a tandem manner on opposite faces of the DNA (yellow). The dashed red line represents the disordered linker between the Ubx homeodomain and its YPWM motif [78].

mRNA alternatively spliced isoforms of Ubx

Immunohistochemical staining experiments using the isoform-specific antibodies demonstrated that Ubx isoforms la and lb are expressed primarily in the epidermis, mesoderm, and peripheral nervous system during embryonic development; whereas isoforms IIa and IIb are expressed primarily in the central nervous system and isoform IVa is expressed exclusively in the CNS (**Figure 1.8**) [87]. Class IIa and IVa RNAs and proteins increase in abundance during later stages with distinct expression patterns in the CNS. Ubx IIa and IVa RNA and protein expression decrease from the posterior to parasegment 6 (PS6) in *Drosophila*. In addition, there are strong differences in the level of expression of class IVa RNA and protein in different cells within a given neuromere, whereas IIa RNA and protein show a more uniform level of expression within each neuromere. These experiments confirm that individual neurons within each neuromere express quantitatively different combinations of Ubx IIa and Ubx IVa. In particular, detailed studies have shown that isoforms Ia and IVa differ quantitatively in their ability to direct development of the peripheral nervous system, Ia being able to carry out this function more effectively than IVa [88]. Large scale protein interactions not only important for executing biological function, it may also be utilized for material formation. The following sections will describe different type of protein materials.

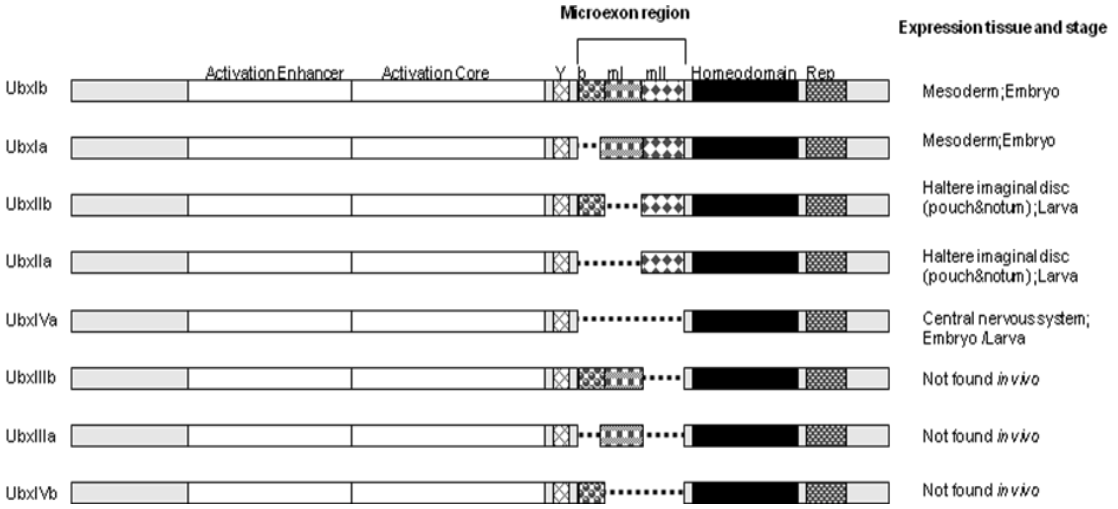

Figure 1.8 Schematic representations of natural and non-natural Ubx isoforms and expression domain *in vivo*. Dashed lines represent the missing microexon(s) in different isoforms [87,89].

Protein-Protein Interactions in Protein Materials

Protein-Protein interactions are not only involved for cellular regulatory functions but can also serve as a building block for cellular structure and tissue architectures. Many protein materials are cytocompatible, and biocompatible, thus could be developed as tissue engineering scaffold for regenerative medicine and other medical purposes [90-94]. A variety of protein materials are available for these applications, each with different properties. The following literatures review will provide overview for the properties and applications for protein based materials.

Elastin

Elastin proteins form elastic fibers in soft tissues, such as blood vessels, skin, and lung [95]. Due to the biocompatibility and cytocompatibility, elastin has been explored for the following applications including but not limited to cartilage and intervertebral

disc tissue engineering, vascular graft tissue engineering, ocular tissue engineering, and cell sheet engineering [96-99]. Elastin is formed from soluble precursor protein, tropoelastin (60-73 kDa) which is a type of extra cellular membrane proteins. Tropoelastin can be found in elastingenic cells such as smooth muscle cells, fibroblasts, and endothelial cells [100]. Varieties of elastin protein have been isolated from animal tissues or expressed in the precursor form using recombinant technology. Tropoelastins aggregate under physiological salt concentration at neutral pH and 37°C when the tropoelastin proteins reach the 10% weight/water concentration to form elastin [101]. Tropoelstin contains a nonpolar hydrophobic core with repeat motifs such as (GVGVP])n and (GXXPG)n, where X could be lysine or valine residues. The hydrophilic domains with high content of lysine residues mediate the crosslinking of elastin and stabilize the structure, and provide mechanical strength (**Figure 1.9**) [102,103].

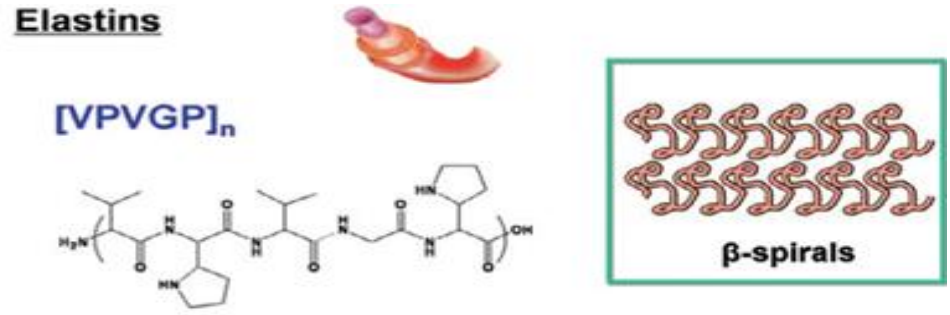

Figure 1.9 Elastins sequence and structure. Elastins sequence include repeats such as [GVGVP]n which can form β-spiral structure [104].

Collagens

Collagen is one of the most abundant proteins found in animals with up to 35% of whole body protein. There are 29 different types of collagen (TypeI- TypeVII) identified and each type of collagen has different functions primarily in mammalian (e.g. skin, bone, blood vessels, cartridge, and tendon) [105]. Collagen have been widely used for biomedical and tissue engineering purposes such as tissue engineering scaffolds, antibody production, drug delivery platform, and 3D cell culture scaffolds [106-108]. Collagen is composed of right-handed triple-helix fibril called tropocollagen. Individual tropocollagen fibril is formed by collagen α chain triple helix segment from either (GPX)n or (GXX) amino acids sequence repeats, where X could be any amino acid other than glycine, proline or hydroxyproline (**Figure 1.10**) [105,109]. Tropocollagen fibrils are intertwined to form a right-handed triple helical collagen bundle. Polymerization of collagen can be physical, enzymatic, and chemical through the modification of the amine and carboxyl groups to form covalent bonds between the tropocollagen fibrils.

Figure 1.10 Collagens sequence and structure. Collagens show sequence repeats such as (GPX)n and triple-helix structure [104].

Fibrin

Fibrin protein is used as the function of assists the blood clotting. The application of fibrin for medical purpose dates back to early 19th century for healing battleground [110]. Since then, it has been applied as a tissue engineering scaffold purposes including but not limited to: cardiac tissue engineering, bone reconstruction, muscle tissue engineering, skin engineering, ocular tissue engineering, and vascular tissue engineering [111,112]. The advantages of using fibrin as materials include biocompatible, biodegradable, mimic the extracellular matrix, and easy to transport the nutrition to cells. Since fibrin is isolated from animals and not purified as a recombinant protein, the possibility of transferring pathogens and endogenous material inconsistency often require modifications during material production [113].

Silk from silkworms and spider

Silk is composed of fibrous proteins and generated by constant pulling and spinning by silkworm from the posterior end glands of organisms [114]. Silk has been widely tested for a variety of applications such as: tissue engineering scaffolds, platforms for drug delivery and *in vitro* disease models [90,104,115,116]. Liquid- like silk proteins form β-sheet crystallite after dehydration, which happen after the liquid-like silk protein leaves the body of silkworm. These proteins are highly soluble *in vivo* compared with other proteins, which can reach a concentration of 30% over w/v without protein aggregation [117,118]. Silk protein can be stored in soluble liquid form at moderate temperature in silkworm for several weeks without premature formation of crystallites, which is a common problem for generating silk fibers *in vitro* [115]. Silk

crystallite formation can be induced by changing the physical and chemical environments. Changes of temperature, force, pH or addition of organic solvents can instigate the polymerization of silk fibroin proteins [115,119]. The primary structure of silk protein from *Bombyx mori* silkworms include the disordered repetitive sequence $GAGAGSGAAG[SG(AG)_2]_8Y$ with polar and charged flanking amino acids residues and an alanine rich hydrophobic core in the middle of peptide which from pleated β sheet (**Figure 1.11**) [104]. The formation of silk fibril starts begins with the micelle formation of multiple silk proteins. Dehydration process and sheer force enhance the hydrophobic association between individual silk fibroin which facilitates polymerization of silk fibroin micelle into silk fibers [114,120]. One of the most notable features for spider silk is the strength and toughness. For example, the spider silk is as strong as engineered steel with a strength to weight ratio at 1.1 GPa compared to 1.3 GPa from engineered steel. In addition, spider silk is tougher than Kevlar 81 fibers with a tested value at~ 165 k Jkg^{-1} compared to 33 k Jkg^{-1} for Kevlar 81 fibers [121].

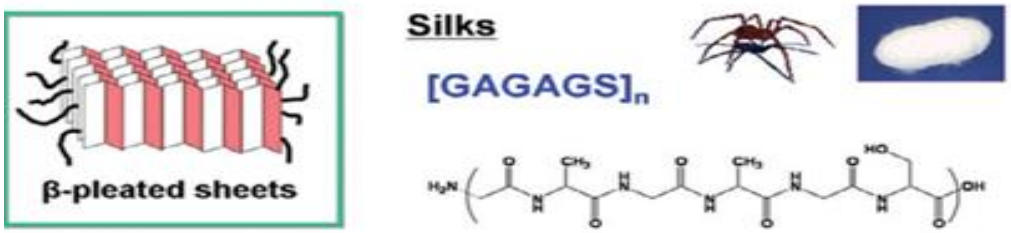

Figure 1.11 Spider and silkworm silks sequence and structure. Spider and silkworm silks show sequence repeats such as (GAGAGS)n or (AAAAA)n, and beta-pleased sheet crystal structure [104].

In addition of silkworms, spiders, bees, ants, and hornets can also generate silks with diverse structure and properties [122,123].

Ultrabithorax Protein Materials

Most protein materials are form by scaffold proteins which have inherited characteristic to support cellular structure. However, many proteins, other than scaffold or structure supporting protein, can also form materials. For example, Ultrabithorax (Ubx) protein, a transcription factor from Hox proteins family, can self-assemble into materials in vitro [124]. The following section will focus on illustrating Ubx self-assemble protein materials.

Ubx readily self-assemble into material and can be functionalized

Ubx proteins self-assemble and form hierarchy structure between air-liquid interfaces. Ubx protein does not need extreme condition such as high temperature, extreme pH or chemical reagents to facilitate the polymerization [124]. Ubx protein can form fibrils in phosphate based buffer at physiological condition such as room temperature and pH 7.4. Monomeric Ubx protein self- assembles into fibril 50 nm in diameter at air-liquid interface after 1 hour incubation under physiological relevant condition. Under suitable protein concentration (0.075 mg/mL and 0.6 mg/mL), Ubx fibrils can form films, sheets, and eventually ropes. In addition, Ubx can also form more complex structures such as bundles, lattices, and encapsulates [124]. Regarding the mechanical properties, individual Ubx protein self-assemble rope can be stretched to an additional 53% (±19%) of their original length before fracturing, which is more elastic

than silkworm silk, spider silk, and collagen [124]. Ubx materials can also tolerate 98°C, indicating the fibers are mechanical robust under extreme condition. Ubx materials can be functionalized using molecular biological approaches. Ubx fusion proteins can be created by standard molecular cloning. Several Ubx chimera proteins have been successfully created, including myoglobin, luciferase, mCherry, and EGFP [125].

Ubx is cytocompatible

Ubx-mCherry chimera fibers are not toxic to three different human cell lines including primary human aortic smooth muscle cells, primary human umbilical vein endothelial cells, and primary human brain vascular pericytes [126]. Ubx fibers can accommodate these three different cell types after prolong incubation without killing cells (**Figure 1.12**) [127].

Ubx materials is biocompatible

Ubx materials are biocompatible [127]. Ubx fibers attract very low levels of immune cells when they are implanted into mice. Ubx fibers are stable in host animal for prolong incubation. Western blotting and enzyme-linked immunosorbent assay (ELISA) indicated below detectable level of antibodies against Ubx in the sera from mice implanted with Ubx protein fibers (**Figure 1.13**). In additinn, Ubx fibers cultured with macrophages *in vitro* do not lyse or activate the macrophages, measured by TNF-α and NO secretion, which indicate low level of immunoresponce triggered by Ubx fibers . Finally, Ubx fibers do not induce hemolysis when co-incubated with red blood cells [127]. This dissertation will focus on the role of intrinsically disordered regions in

protein binding by Ubx and application of Ubx protein materials in establishing 3D cell culture. Studies, including validation of a high-throughput assay for identification of regions in Ubx protein, exploration for selection criteria for Ubx proteins and identification for domains in Ubx proteins, which interact with other proteins, as well as characterization of cell behaviors on biomaterials formed by self-assemble Ubx proteins will be discussed in this dissertation.

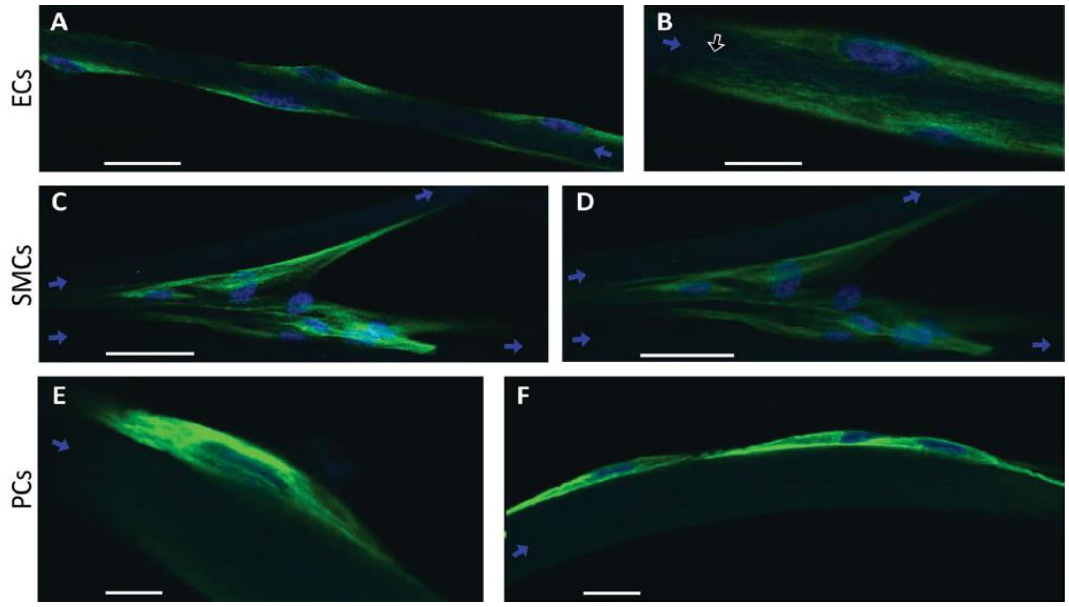

Figure 1.12 Ubx fibers accommodate cells regardless of the cell type. Ubx fibers are marked with blue arrows. Endothelial cells (ECs) (A and B), aortic smooth muscle cells (SMCs) (C and D), and brain vascular pericytes (PCs) (E and F) growing on plain Ubx fibers were immunostained for α- tubulin (green) and stained with DAPI (blue). In all cases, the cytoskeletons are normal and roughly aligned with Ubx fibers. Scale bars: (A) 20 μm; (B, E, and F) 10 μm; (C and D) 50 μm [126].

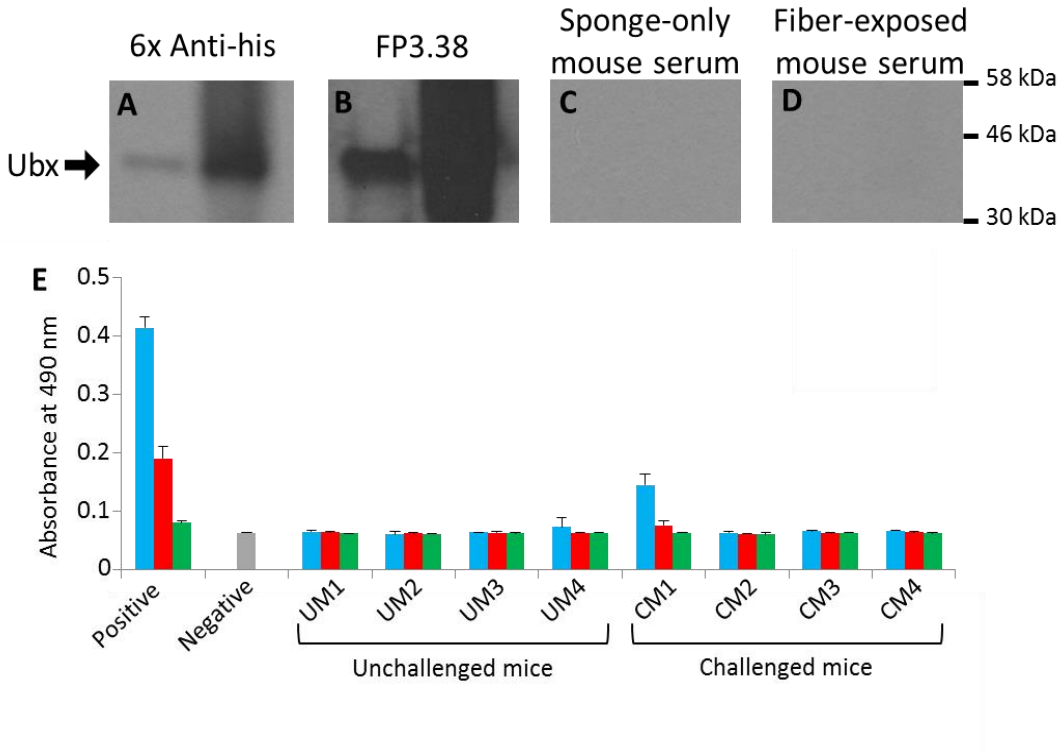

Figure 1.13 Ubx materials are stable in host animals implanted with Ubx materials. (A-D) Each panel shows a western blot using a different source for primary antibodies to determine whether antibodies are present that can detect Ubx. (E) ELISA assays show no detectable level of antibody in mice implanted with Ubx materials. The purified anti-homeodomain antibody (FP3.38) was used for positive control and purified donkey anti-mouse IgG HRP antibody was used for negative control. UM (1-4) represent unchallenged mouse and CM (1-4) represent challenged mouse serum. Dilution factors for primary antibody or mouse serum are 1:100 (blue), 1:1000 (red), and 1:10000 (green)[127].

CHAPTER II

MEDIA COMPOSITION INFLUENCES YEAST ONE- AND TWO-HYBRID RESULTS*

Introduction

The yeast two-hybrid system provides an efficient method for identifying novel protein-protein interactions in small-scale studies and high-throughput screens [128-130]. In this system, the first hybrid is composed of a bait protein fused to a DNA binding protein that recognizes DNA sequences upstream of a lacZ reporter gene. The second hybrid protein consists of a strong activation domain fused to a potential protein partner. Interaction between bait and partner proteins localizes the strong activation domain to the lacZ promoter, thus activating transcription of the lacZ reporter gene. Multiple cycles of lacZ transcription, translation, and β-galactosidase cleavage of X-gal generate visible quantities of the blue assay product, 5-bromo-4-chloro-3-hydroxyindole. This sensitive system detects interactions that may be difficult to observe by other means (e.g., co-immunoprecipitation) due to low abundance of the protein complex. The sensitivity of this method can also lead to several problems: (i) transformation efficiency affects signal strength and is highly dependent on technique [131], (ii) 25% to 50% of detected interactions are estimated to be false positives [132,133], and (iii) 55% to 90% of true protein interactions are not observed (false negatives) [129,133].

---

*Reprinted with permission from "Media Source Profoundly Influences the Outcome of Yeast One- and Two-Hybrid Experiments" by Liu, Y, Merchant, Z, Hsiao, H-C, Gonzalez, KL, Matthews, KS, and Bondos, SE., 2011. Journal of Biological Procedures Online 13, 6. Copyright 2011 by Biomedical Central Ltd. DOI 10.1186/1480-9222-13-6

Identifying sensitive parameters in yeast two-hybrid experiments could help address these difficulties. Herein, we report both the source of media components and the media preparation protocol can impact yeast one-hybrid and two-hybrid results.

In the yeast one-hybrid system, the bait protein, which includes a transcription activation domain, is fused to a DNA binding protein [128]. When the protein of interest is hybridized to a DNA binding domain, this assay can be used to assess the presence and strength of an activation domain [134]. Alternately, the protein of interest can be hybridized to a strong activation domain to screen for DNA binding [135]. The yeast one-hybrid method requires transformation of only one plasmid, resulting in improved transformation efficiency and fewer cellular manipulations[130]. This simpler experimental design reduces experimental variability, allowing focus on systematic issues that arise using this method.

Materials and Methods

Media were prepared according to the protocol in **Table 2.1** and **Table 2.2** with varying parameters as discussed below. EGY48-p8op-lacZ yeasts were transformed as previously described [132,134,136]. The genome of this reporter host strain (EGY48) includes a wild-type LEU2 gene under the control of a series of LexA operators. The reporter plasmid p8op-lacZ, which carries the lacZ gene under the control of LexA operators, had been stably transformed into this strain [134]. For each DNA tested, cells from a single transformation were divided between plates for each condition to ensure that commonly cited error sources, such as interactions with endogenous proteins, auto-

activation, and low transformation efficiencies [128-130], did not contribute to differences in results. The resulting blue color of yeast colonies, which reflects the level of transcription activation, was gauged with a standard color chart (**Figure 2.1**). Analysis by color chart yielded data comparable to analysis by enzyme assay. All experiments were repeated at least three times and yielded consistent results. Yeast nitrogen base sources were Sigma (Y1251-100G), Difco (233520), and Clontech (630421).

The procedures for making different yeast media plates are similar, but the identity of the dropout supplements, sugars, and antibiotic can vary. Since we have used the Clontech pLexA Matchmaker yeast two hybrid system, we must replace glucose with galactose and raffinose to induce chimera expression.

Table 2.1 Yeast media preparation protocol.

Step	Procedures
1	Add 300 ml water to a 1 L flask.
2	Using a stir plate, stir continuously without heat while adding: a. 0.85 g yeast nitrogen base without amino acids or ammonium sulfate (Difco #233520). The impact of using yeast nitrogen base from other companies is discussed in the accompanying paper. b. 2.5 g Ammonium sulfate (Mallinkrodt Chemicals, 7725) c. 10 g Galactose (Acros, 150611000) d. 5 g raffinose (Acros, 1956171000) e. 0.375 g –His/-Ura/-Trp dropout supplement (Clontech, 630424)
3	Bring the final volume to 445 ml.
4	Add 10.5 g agar (Fisher Scientific, BP1423). The new volume should be approximately 450 ml.
5	Autoclave at a setting of 1.27 kg/cm^2 at 121 ºC for 15-30 minutes. Clontech recommends 15 minutes, but we use 30 minutes without adverse effects.
6	Allow the flask to cool down. It is sufficiently cool when you can comfortably touch the bottom of the flask – where the liquid is located – to the inside of your arm (preferred method), or you can hold the flask in your hands for 30 seconds. If you proceed to step 7 too quickly, you will destroy the heat-sensitive reagents.
7	Using a sterile hood, add the following reagents while swirling the flask to prevent local cooling / solidification of the agar: f. 500 µl of 50 mg/ml kanamycin monosulfate (CRPI, K22000) for a final concentration of 50 µg/ml g. Turn off hood and room lights to add 0.4 ml of 50 mg/ml X-gal (Progene V3941). X-gal is light sensitive. If you make your own X-gal solution, use DMSO or DMF as a solvent. Either way, your X-gal solution should be clear. If you see traces of brown, it should be discarded. h. 50 ml 0.7M KP buffer*.

*The preparation of KP buffer is described in Table 2.2.

Table 2.2 KP* buffer preparation protocol.

Step	Procedures
1	Make 1 M K2HPO4 by dissolving 17.4 g of K2HPO4 (Sigma, P9666) to a final volume of 100 ml.
2	Make 1 M KH2PO4 by dissolving 13.6 g of KH2PO4 (Sigma, P9791) to a final volume of 100 ml.
3	Use a heated stir plate to dissolve both salts. Do not heat above 50 degrees. Do not heat after all of the salt is in solution.
4	Mix 61.5 ml of 1 M K2HPO4 with 28.5 ml KH2PO4.
5	Adjust the pH to 7.0 if necessary using the leftover 1 M salt solutions.
6	Dilute the resulting 1 M solution at pH=7.0 to 0.7 M.
7	Filter sterilize the resulting buffer using a 0.20 or 0.22 micrometer cutoff filter into an autoclaved flask or bottle.
8	KP buffer should be stored at 4 ºC used within one month and discarded if visibly contaminated.

*KP buffer is severed as a reagent to stabilize the pH in yeast media plates

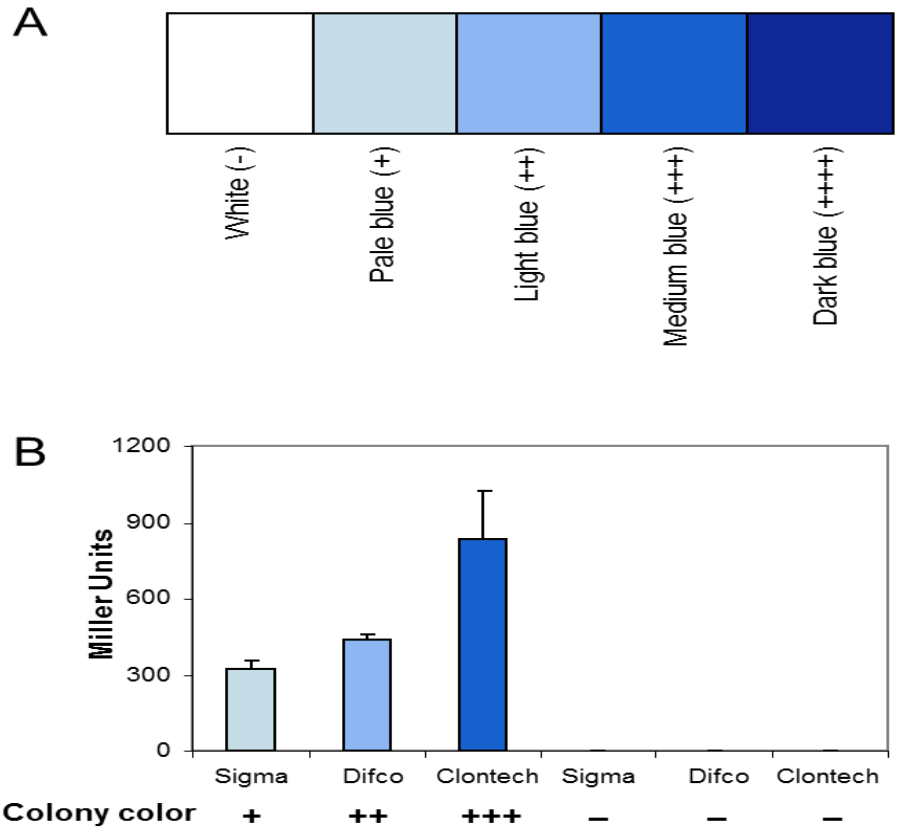

Figure 2.1 Evaluation of results by colony color. (A) Color chart used to standardize colony assessment. (B) Color analysis yields results comparable to quantitative assays for β-galactosidase activity, reported in Miller units. Miller units were calculated as specified in the Clontech yeast transformation protocol (PT3024-1).

Results

Effects of yeast nitrogen base source on yeast-one hybrid experiments

We first compared media in which the yeast nitrogen base (YNB) was produced by different companies and the pH was adjusted after sterilization to ensure consistency between media preparations. In general, media made with YNB from Sigma yielded the strongest signal in yeast one-hybrid experiments, whereas media made with Difco YNB yielded intermediate results, and media made with Clontech YNB produced the weakest signal (**Table 2.3, Conditions 1-3**). We began experiments using pLexA-Gal4, in which DNA encoding the LexA DNA binding domain is fused to DNA for the Gal4 transcription activation domain. Expression of this hybrid protein should produce a strong positive signal. At only 48 hours post-transformation, yeast colonies transformed with pLexA-Gal4 appeared light blue on media made with Sigma and Difco, whereas colonies on the Clontech media remained white. After 60 hours post-transformation, pLexA-Gal4 colonies on Sigma and Difco media were medium blue, and the colonies on the Clontech media were light blue. Although the results from Difco and Clontech media became indistinguishable after 72 hours, colonies on the Sigma plates were still darker blue. While the source of YNB made a consistent difference, using different lots from the same supplier had no impact on results (**Table 2.4**). To determine whether this media effect is unique to LexA-Gal4, we next tested LexA-Ubx, in which LexA is fused to the *Drosophila* Hox protein Ultrabithorax (Ubx), which activates transcription at a moderate level to yield medium blue colonies [134]. Colonies transformed with pLexA-Ubx

yielded a similar trend as pLexA-Gal4: colonies on media containing Sigma and Difco YNB were the first to show blue color (**Table 2.3, Conditions 1-3**).

Table 2.3 Yeast one-hybrid results are strongly dependent on media.

	Yeast Nitrogen Base	pH adjustment [1]	48 hrs observation			60 hrs observation			72 hrs observation		
			Gal4	*Ubx*	*Pro-2*	*Gal4*	*Ubx*	*Pro-2*	*Gal4*	*Ubx*	*Pro-2*
1	Sigma	+	++	–	–	+++	+	–	++++	++	++
2	Clontech	+	–	–	–	++	–	–	+++	++	++
3	Difco	+	++	–	–	+++	+	–	+++	++	++
4	Sigma	–	+++	–	–	++++	+++	++	++++	+++	++
5	Clontech	–	++	+	–	+++	++	+	+++	++	++
6	Difco	–	–	–	–	++	–	–	++++	+++	–

We examined six media conditions, defined by yeast nitrogen base supplier and pH adjustment post-autoclave. Difco media can yield false negative data (condition 6) and the other media preparations can yield false positives (conditions 1-5) compared to other transcription activation assay methods[134]. "pH" refers to the pH of the 0.7 M potassium phosphate buffer (KP); "Gal4" stands for pLexA-Gal4, the positive control plasmid expressing the DNA binding domain of LexA fused to the Gal4 transcription activation domain (Clontech). "Ubx" represents the pLexA-Ubx plasmid, "Pro-2" is the negative control, in which proline mutations disrupt transcription activation by the LexA-Ubx hybrid . The blue shade of colonies is reported here as symbols: "++++" for dark blue, "+++" for medium blue, "++" for light blue, "+" for pale blue, and "–" for white colonies. See Figure 1. for color standards. The pH after autoclaving but before addition of KP was 4.67, 5.11, and 4.31 for Sigma, Clontech, and Difco media, respectively. After addition of KP buffer post-autoclaving, the pH of all media was 6.7 ± 0.1. For conditions 1-3, the pH was adjusted to 7.0. Therefore pH differences are not the source of the supplier-based differences in results.

Table 2.4 Different lots media with no different results.

Lot number	Experiation date	48 hrs observation			60 hrs observation			72 hrs observation		
		Gal4	*Ubx*	*Pro-2*	*Gal4*	*Ubx*	*Pro-2*	*Gal4*	*Ubx*	*Pro-2*
144069	12-31-2004	−	–	–	++	−	−	++++	+++	−
7116446	10-30-2012	−	–	–	++	−	−	++++	+++	−
8190329	11-30-2012	−	–	–	++	−	−	++++	+++	−
9194329	6-30-2014	−	–	–	++	−	−	++++	+++	−

Yeast one-hybrid experiments yield consistent results using yeast nitrogen base from the same supplier (Difco), even with different lot numbers and expiration dates (the first one well past due). When preparing the media, no significant differences in pH were observed either before autoclaving or after adding KP buffer.

As a negative control, we used LexA-Pro2 in which LexA is fused to a Ubx variant containing mutations A223P/Q224P/T225P/A228P, which disrupt a predicted α-helix required for transcription activation [134]. Importantly, these proline mutants also prevent transcription activation by Ubx in a promoter-reporter assay in *Drosophila* S2 cells, corroborating the yeast results [134]. Whereas we expected colonies expressing LexA-Pro2 to be white, colonies grown on all three types of media were light blue after 72 hours. Consequently, the sensitivity of the signal to changes in media appears to depend on the strength of transcription activation: the strongest activator, LexA-Gal4, generated the largest differences, whereas results for the weakest activator, Pro2, were independent of YNB source for this media preparation method.

Based on our prior results, we expected to see a wide range of color: dark blue (++++) for LexA-Gal4, medium blue (+++) for LexA-Ubx, and white (-) for LexA-Pro2 [134]. However, the results for wild-type and mutant Ubx were similar at 72 hours post-transformation in all three conditions. Furthermore, colonies expressing LexA-Gal4 were

lighter than expected when grown on Clontech and Difco media. Thus, the range of signal strengths appears poor for Conditions 1-3 (**Table 2.3**).

Effects of adjusting media ph on yeast-one hybrid experiments

Strikingly, simply not readjusting the pH after sterilization resulted in amplification of the signal range. Comparing results for wild-type and mutant Ubx, a slight difference in signal was again observed in Sigma media which persisted to 72 hours (**Table 2.3, Condition 4**). However, using Difco YNB without pH re-adjustment not only increased the signal range at 72 hours but also matched prior results in yeast and *Drosophila* cell culture [134]. Specifically, for Condition 3 (Difco, no pH adjustment), the data ranged from light blue (++, Ubx and Pro-2) to medium blue (+++, Gal4), whereas for Condition 6 (Difco, pH adjustment), the data ranged from white (-, Pro-2) to dark blue (++++, Gal4).

Even without readjusting the pH, media using YNB sources did not replicate previous data in yeast and *Drosophila* cell culture. Consequently, media choice can influence the prevalence of false positives and false negatives. Expression of the LexA-Gal4 positive control correctly yielded a strong positive signal at 48 hours on Sigma media, but a false negative signal on Difco media. Likewise, although LexA-Ubx expression rapidly generates a positive signal on Sigma plates, expression of LexA-Ubx on Difco plates gives a false negative response at 60 hours post-transformation, although a positive signal eventually develops after longer incubations. Consequently, the researcher could miss this signal, since the pLexA-Gal4 positive control produces signal at 60 hours. Conversely, the Ubx mutant Pro-2, which should *not* activate transcription,

yields a false positive signal on Sigma media at after 60 hours incubation. Therefore, by not re-adjusting the pH after autoclaving, we observed a greater overall range of signal, greater media-dependent differences in signal, and greater differences in signal between transcription activators of different strengths.

Media effects on yeast-two hybrid experiments

Given that the source of the YNB had a profound impact on the results observed in a yeast one-hybrid assay, we reasoned that differences in media might partially account for the variations in data originating from different laboratories in yeast two-hybrid experiments [130,134]. Testing this possibility, we examined the impact of the media source in yeast two-hybrid experiments using a set of control plasmids, in which the T-Antigen-B42 activator chimera should bind a LexA-p53 chimera but not a LexA-Lamin chimera. Similar to yeast one-hybrid experiments, we found the results of yeast two-hybrid data were also dependent on media (**Table 2.5**).

At five days after transformation, only media produced with Clontech YNB yielded a positive signal, and at six days after transformation, colonies on Clontech media had the strongest signal and Sigma media consistently yielded the weakest signal. Surprisingly, this trend is opposite that observed for yeast one-hybrid experiments, suggesting either the two experimental methods are differentially sensitive, or the magnitude of the media effect is influenced by the proteins assayed.

Table 2.5 The source of yeast nitrogen base also impacts yeast two-hybrid results.

	Yeast Nitrogen Base	pH adjustment	5 days observation		6 days observation	
			p53	*LAM*	*p53*	*LAM*
1	Sigma	–	–	–	+	–
2	Clontech	–	+	–	+++	–
3	Difco	–	–	–	++	–

In this assay using control plasmids provided by Clontech, pB42AD-T-Antigen should bind pLexA-53, in which the p53 protein is fused to LexA, resulting in blue colonies [137]. Conversely, T-Antigen should not bind Lamin produced by pLexA-LAM [138].

Discussion

In summary, we have demonstrated that i) media components and preparation methods influence yeast one- and two-hybrid experiments, ii) for yeast one-hybrid experiments, this media sensitivity scales with increasing activation domain strength, and iii) media can determine the range of signal observed in yeast LacZ reporter assays. Such differences are sufficient to change data interpretation: experiments that should yield a negative result may appear positive, or vice versa. We observed media-dependent effects in four systems: yeast one-hybrid experiments using either the Gal4 activation domain, wild-type Ubx, or a Ubx mutant lacking activity, and finally yeast two-hybrid experiments reporting on the p53-T-antigen interaction. Therefore, this media sensitivity occurs across a wide range of yeast experiments and may contribute to differences in yeast two-hybrid results observed previously [129,137]. Our results for a single yeast nitrogen base source are consistent even when testing different lot numbers or chemicals differing in age by a decade, suggesting that the differences we observe

when the source of the yeast nitrogen base is changed are significant. Since each manufacturer reports the same yeast nitrogen base composition, these differences may be attributed to differences in contaminants of these media components. These experiments suggest that careful reporting of chemical sources and preparation methods for yeast media is essential to enable replication of experiments and comparison of data between studies. Finally, these media-dependent differences in results could be exploited to enhance and thus verify weak differences in signals from yeast assays.

CHAPTER III

THE INTRINSICALLY DISORDERED REGIONS OF THE *DROSOPHILA MELANOGASTER* HOX PROTEIN ULTRABITHORAX SELECT INTERACTING PROTEINS BASED ON PARTNER TOPOLOGY*

Introduction

Most biological processes are implemented and regulated by macromolecular complexes, in which proteins are major components. The function of an individual protein, therefore, is often determined by the identity and range of the proteins to which it binds [70,131,139]. Consequently, proteins must specifically and reliably bind the correct partners *in vivo* [65,66,140,141]. Interactions between structured proteins require complementary topologies that generate sufficient interfacial surface area [13,142-144] and complementary surface chemical groups capable of creating stable interprotein bonds [7,145]. Residues forming an interface between two structured proteins are often less dynamic relative to non-interfacial surface residues, even when the proteins are in the unbound state [142].

Intrinsically disordered proteins and protein regions are present in more than one third of protein complexes and are enriched in proteins with multiple partners [10,22,146-154].

As monomers, these proteins lack stable globular structures and rapidly interconvert among a large ensemble of conformations. Disordered protein monomers can sample structure present in the bound complex or be extremely dynamic with little detectable canonical structure [43,155,156]. The disordered region may fold to similar structures present in all interactions, or a single disordered region may adopt many different structures to bind protein partners with different topologies [157-161].

In contrast to the complementary interface formed by two folded proteins, a subset of disordered regions remain highly dynamic even when bound, either initially through an induced-fit binding mechanism [159,162] or as part of a heterogeneous final complex [40,163-165]. This structural heterogeneity in the complex has been proposed to be an essential component of fine-tuning the function of the complex [163] as well as maintaining the sensitivity of the complex to evolving cellular signals [166]. The extreme malleability of intrinsically disordered regions, even in the bound state, raises questions regarding the role of the structure and surface topology of the partner protein in these interactions. Indeed, disordered proteins bind more types of protein structures (folds) than do structured proteins [167].

In this paper, we explore the importance of partner topology in protein interactions mediated by Ultrabithorax (Ubx), a *Drosophila melanogaster* Hox transcription factor. Ubx is composed of both structured and disordered regions [65,78,134]. Amino acids 1-102 of Ubx, herein termed Region 1, include a mixture of short structured elements interspersed with disordered sequences. Region 2 is a large disordered region, spanning amino acids 103 to 216 and including a portion of the

transcription activation domain [134]. A putative α-helix required for transcription activation is located in Region 3 [134]. Amino acids 250-303, termed Region 4, encompass intrinsically disordered, alternatively spliced microexons and the disordered N-terminal arm of the homeodomain. Finally, the C-terminal Region 5 includes the structured portion of the homeodomain. Based on native state proteolysis rates, the disordered regions of Ubx are significantly more exposed than the disordered regions of proteins that fold upon ligand or co-factor binding [65]. Moreover, Region 2 is extremely glycine rich (27%, including 13 contiguous glycines). Polyglycine peptides are compact, yet very dynamic, and lack stable intra-protein contacts [168,169]. Because the extent of monomer disorder correlates with the degree of disorder present in the bound state [162,170,171], the extremely dynamic disordered regions in Ubx are unlikely to fold into a stable structure upon partner protein binding.

Ubx is a "one-to-many" protein, in that it physically interacts with 39 known partner proteins with a wide variety of molecular functions [6,78,79,172,173]. This large number of partner proteins provides a sufficiently diverse sample to identify common traits that enable binding to Ubx. Several of these interactions have been validated *in vivo* [6,79,132]. Proteins that genetically interact with Ubx, unsupported by physical interaction data, were not included in this study since genetic interactions can arise from processes other than physical interaction between proteins. We found that specific folds are significantly enriched in Ubx-interacting proteins. Single domains of the partner protein that exhibit the selected fold are sufficient to bind Ubx. Interestingly, the intrinsically disordered regions of Ubx are necessary for these protein interactions.

Although partners bind all three disordered regions cooperatively, individual partner proteins rely on specific disordered regions to varying extents, creating opportunities for competition and collaboration in forming higher order complexes. Regions 1 and 2 are multiply phosphorylated, providing another mechanism to regulate partner binding *in vivo*. Partner binding also varies among Ubx isoforms arising from *ubx* mRNA splicing, providing a third regulatory mechanism. Interestingly, the preference of protein partners for specific Ubx isoforms correlates well with the topology of the partner protein. Thus, phosphorylation and alternative splicing, both tissue-specific processes, have the potential to regulate protein interactions. The regions of Ubx involved in partner binding also regulate DNA binding and include a portion of the transcription activation domain [65,66,134]. Linking different Ubx functions via intrinsically disordered regions has the potential to provide the specificity and reliability required for Hox activity *in vivo*.

Materials and Methods

Definition of intrinsically disordered regions of Ubx

Ubx disordered regions were defined by a combination of prediction algorithms and experimental assays. Disordered and structured regions were predicted using the average score from three programs, VLXT-PONDR, IUPRed, and DisEMBL (loops/coils) [65]. Predicted amino acid residues with an average prediction score ≥ 0.6 are designated disordered. A residue with an averaged prediction score between 0.4 and 0.6 was considered as uncertain and thus was not defined in this study. A residue with an

average prediction score ≤ 0.4 was considered structured. Native state proteolysis data [65] were used to refine the predicted boundaries of disordered and structured regions. Since successful proteolysis requires a minimum of five disordered amino acids on both sides of the severed bond, the regions designated as disordered were expanded at a few positions to include these sequences. The designations of structure and disorder agree with previous data on the locations of structure in the Ubx homeodomain, the partially structured nature of the HoxB1 FPWM motif in the absence of Pbx1 binding (analogous to the Ubx YPWM motif which was designated as "uncertain" by our analysis), and the location of a putative α-helix involved in transcription activation by Ubx [78,134]. Protein interface and molecular recognition features were predicted by the Anchor and MoRFpred algorithms, respectively [6,173].

Classification of the Drosophila interactome by fold

The *Drosophila melanogaster* large-scale yeast two-hybrid dataset [172] was used for this global analysis. The structural assignments, definitions, and evolutionary relationships listed in Flybase [174] and the Structural Classification of Proteins (SCOP) database version 1.65 release 3 were used to group the proteins by folds. SCOP merges computer algorithms and human curation to classify protein domains based on structural and evolutionary similarities. Interaction maps were generated and modified using Osprey 1.20 (http://biodata.mshri.on.ca/osprey/servlet/Index).

Databases built using Microsoft Access were used to construct the figures and tables in the Supporting Data, which can be accessed from http://rice.allgeek.net. Algorithms to analyze the raw protein interaction data were written using Windows

Visual Basic 6.0. The genome database was compiled from a list of all *Drosophila* genes downloaded from Flybase. If the Flybase reference for the corresponding protein had one or more assigned folds as defined by SCOP, then all potential fold-fold pairs were included in the database. Any structure assignments that were fragments of other folds, "not a true fold", or duplicates of other entries were eliminated. By this analysis, roughly one quarter of *Drosophila* proteins have an assigned fold. Each fold in multifold proteins was included in the genome database, accounting for 23% of the proteins, and was listed as an interacting fold for all interactions in which the multifold protein participates, yielding 63% of the total interactions examined. The interactome database contains previously defined interactions and includes the confidence score assigned to that interaction by Giot *et al.* [172]. Data fitting for the scale-free graph was completed using Igor Pro Version 4.02A (WaveMetrics).

Classification of Ubx protein interactions by fold

Proteins with assigned folds that physically interact with Ubx included data from Giot *et al.* [172], our laboratory [132,136], and other laboratories [79]. Proteins encoded by genes that only genetically interact with *ubx* were not included, because molecular events other than protein interactions can yield a genetic interaction. Folds within this protein list were identified as described above.

Yeast two-hybrid assays

Ubx deletion and truncation mutants were created using the QuikChange site-directed mutagenesis kit following manufacturer instructions (Agilent). Ubx variants were cloned into the pLexA plasmid (Clontech) between the EcoRI and BamHI

restriction enzyme sites. Ubx binding partners had previously been cloned into the pB42 vector [Bondos 2004, Bondos 2006]. DNA encoding the individual domains of Al (residues 81 to 142) and Arm (residues 155 to 273) were synthesized by Blue Heron Biotechnology Inc., USA.

Ubx variants and partner plasmids were co-transformed into EGY48 *Saccharomyces cerevisiae* already carrying the p8op-LacZ reporter plasmid (Clontech). In this process, 500 μl of an overnight liquid culture of yeast ($OD_{600\ nm} \approx 1.5$) was centrifuged, and the pellet was washed with 2 mM lithium acetate (Acros) and 100 mM dithiothreitol (DTT, Fisher Scientific). Cell pellets, re-suspended in 100 μl of transformation reaction mix, containing 2 mM lithium acetate, 50% polyethylene glycol (Sigma, MW3350), 10 μg/ml salmon sperm DNA (Sigma), and 100 mM DTT, were mixed with Ubx-pLexA plasmid and Ubx binding partner pLexA fusion (500 ng per plasmid). The resulting mixture was incubated at 46 °C for one hour and subsequently centrifuged. The pellet was re-suspended in sterile water and spread on a designated synthetic amino acids drop-out yeast medium agar (2%) plate containing 80 μg/ml of X-gal (Research Products International) following incubation for 5-6 days at 30 °C.

The blue or white color of the colonies provided an initial qualitative measure of binding. The results of this qualitative assay matched subsequent quantitative results using the Miller β-galactosidase reporter assay [175]. In this assay, an individual yeast colony was used to inoculate 5 ml of the designated synthetic amino acid drop-out yeast medium, then grown overnight at 30 °C with 250 rpm shaking to an $OD_{600nm} \approx 1.5$. β-Galactosidase liquid assays generally followed the Clontech Yeast Protocols Handbook

(Clontech). In brief, 2 ml of the overnight yeast culture were used to inoculate 8 ml of the $trp^-/his^-/ura^-$ drop-out yeast medium containing 10% galactose (Sigma) to activate the B42-partner chimera and 5% raffinose (Sigma) to provide a carbon source and incubated at 30 °C for 3-5 hr with 250 rpm shaking until the cells reached mid-log phase with $OD_{600\,nm} \approx 0.8$. To harvest the yeast culture, 1.5 ml was removed and centrifuged 10,000 x *g* for 30 seconds. Supernatant was discarded and the pellet was mixed thoroughly with 1.5 ml of Z Buffer (70 mM Na_2HPO_4, 40 mM $NaH_2PO_4 \bullet H_2O$, 10 mM KCl, 1.3 mM $MgSO_4$). After re-centrifugation and decanting the supernatant, the pellet was re-suspended in 300 μl of Z Buffer, divided into three 100 μl aliquots, frozen in liquid nitrogen for 1 minute, and incubated at 37 °C for 45 seconds. This freeze and thaw process was repeated two more times. To the cell lysate, 4 mg/ml of *ortho*-nitrophenyl-β-galactoside (ONPG, Sigma) in Z Buffer and 700 μl of 27% β-mercaptoethanol in Z buffer were added, followed by 30 °C incubation with mixing by inversion every 10 minutes. β-Galactosidase expression levels were assessed by enzymatic assays that spectroscopically measure generation of the β-galactosidase enzymatic product, *o*-nitrophenol (ONP), at 420 nm. When yellow color was visible, reactions were quenched by addition of 400 μl of 1 M Na_2CO_3. The elapsed time from the beginning of the reaction (ONPG addition) to the end of reaction (Na_2CO_3 addition) was recorded. The reaction mixture was centrifuged at 10,000 x *g* for 10 minutes. Supernatant was collected and $A_{420\,nm}$ was recorded. The results were reported in Miller units, the amount of β-galactosidase that hydrolyzes 1 μmol of ONPG to ONP per min per cell.

Miller units were calculated using the following formula:

$$Activity = \frac{(1000)(A_{420nm})}{(t)(V)(OD_{600nm})}$$

in which *t* is the elapsed time (in min) of incubation, *V* is 0.1 ml x dilution factor (5 for this protocol), $OD_{600\ nm}$ is the optical density of 1 ml induction culture before harvest measured at a wavelength of 600 nm, and $A_{420\ nm}$ is absorbance of 1 ml ONPG reaction product measured at 420 nm.

Western blotting

Extraction of yeast protein samples and their preparation for western blotting followed the Yeast Protocols Handbook (Clontech). Cells were lysed as described for yeast two-hybrid assays, and whole cell lysate was subsequently centrifuged at 10,000 x *g* for 10 minutes to remove cell debris and any insoluble Ubx. Proteins were separated by SDS-PAGE prior to western blotting with a 1:200 dilution of LexA murine monoclonal primary antibody (Santa Cruz Biotechnology) followed by a 1:5,000 dilution of IRDye 800CW Goat anti-Mouse IgG (H + L) secondary antibody (Li-Cor). Protein expression was visualized and quantified using an Odyssey infrared imaging system and software (Li-Cor).

Results

Ubx selects protein interactions based on partner topology

The *Drosophila* Hox protein Ubx is 44% intrinsically disordered, and binds many partner proteins [79,132,136,176]. However, the location and chemical nature of

most of these protein interfaces is unknown. To determine which physicochemical properties of partner proteins are important for mediating these interactions, we first examined the characteristics of Ubx-binding proteins (**Figure 3.1**).

Although some of the Ubx partner proteins form true interactions that alter Ubx function *in vivo* [132,136,171], other interactions have not been examined in flies. In addition, a few interactions are unlikely to be biologically relevant because the partner has a different sub-cellular localization and / or is involved in unrelated biological processes. However, binding by all partners results in similar reporter intensities in yeast two-hybrid assays, reflecting similar protein interaction affinities. Ubx is presumably able to bind the unlikely partners *in vitro* because these proteins share features recognized by Ubx when it binds the true partners. Therefore, including these unlikely partners in the analysis increases the occurrence of traits selected by Ubx while simultaneously decreasing the occurrence of traits commonly found in transcription factors but not specifically required for Ubx binding. Of the 39 known Ubx binding proteins [6,78,132,136,172,173], 34 have domains with assigned folds. We searched for properties common to these 34 Ubx partners. Ubx has a predicted net charge of + 7.3 at pH = 7.4. This strong positive charge is largely due to the DNA-binding homeodomain (+11 at pH = 7.4), the only large structured region within Ubx. Any proteins directly binding the homeodomain would be expected to have a compensating negative charge. Ubx partners have a surprisingly large range of predicted net charges at pH = 7.4, spanning +36 to -54 (**Figure 3.2**). Thus, net charge does not correlate with the ability to

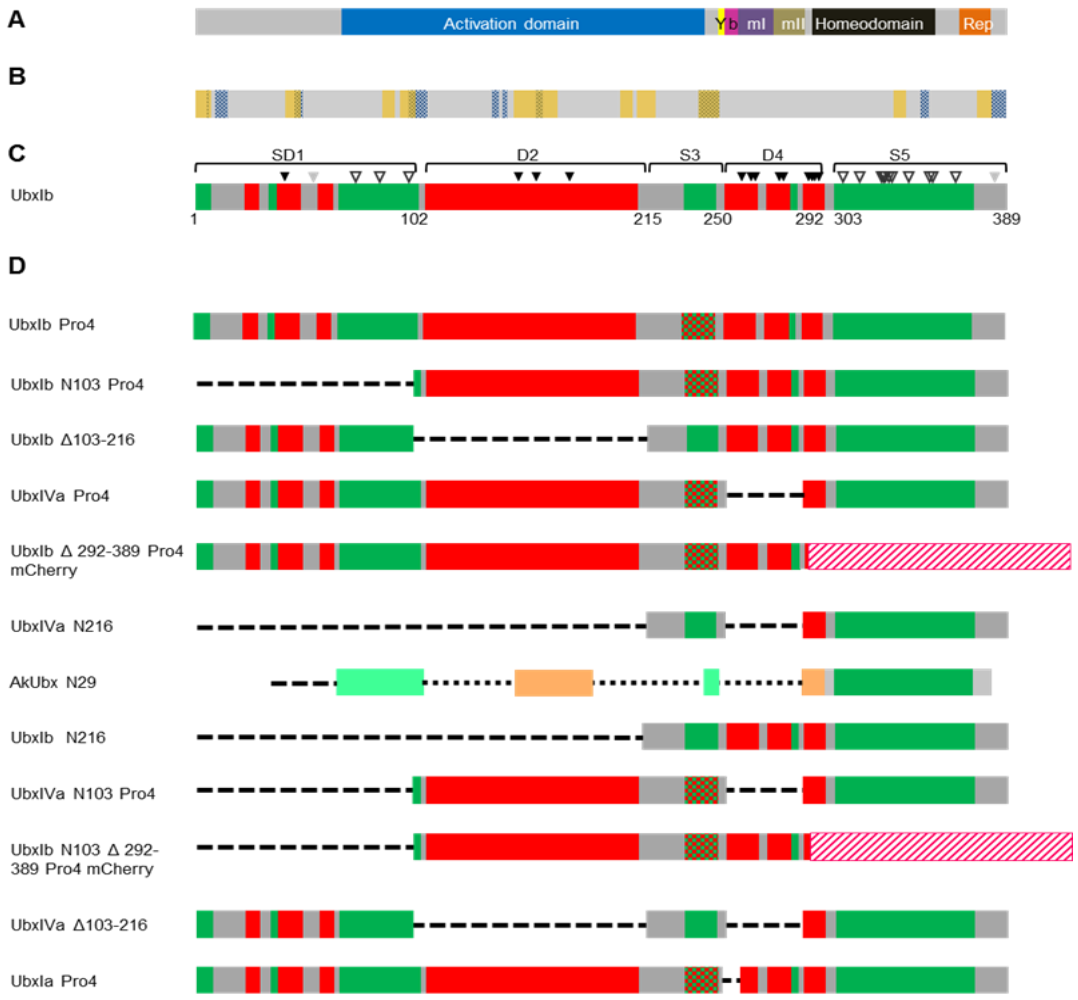

Figure 3.1 Location of structured and disordered regions in UbxIb, and design of Ubx variants. (A) A grey bar, representing the domain organization of the UbxIb transcription factor shows the position of its transcription activation domain (blue), YPWM Exd interaction motif (yellow), DNA-binding homeodomain (black), a partial transcription repression domain (orange), and protein regions encoded by three alternatively spliced microexons: the b element (pink), mI (purple), and mII (brown). (B) The location of predicted protein-interaction motifs in Ubx as predicted by ANCHOR (yellow stripes) and MoRFpred (blue stippled stripes). Regions predicted by both algorithms to be involved in protein interactions are marked with both yellow and blue. (C) A bar schematic depicting the positions of structured and intrinsically disordered regions in UbxIb. The boundaries were determined by a combination of computational and experimental approaches. The scores from three disorder prediction algorithms were averaged to identify structured (green) and disordered (red) regions. Native state proteolysis, in which only disordered segments can be cleaved by trypsin, was used to verify these assignments, and, where appropriate, slightly expanded the boundaries of the predicted disordered regions [65]. Sites cut by trypsin (▼), sites not cut by trypsin (▽), and sites that could not be definitively assigned (▼) are indicated. (D) Bar schematic for predicted protein interfaces and molecular recognition features (MoRFs) on Ubx peptide. The schematic bars show Anchor algorithm predicted Ubx- partner protein interfaces (orange bars) and MoRF algorithm predicted Ubx-partner protein interface (blue bars with pattern fill). (D) Bar schematics of Ubx truncation mutants and internal deletion mutants used in yeast two-hybrid assays to identify partner binding interfaces. UbxIb, UbxIa, and UbxIVa are isoforms created by alternative splicing *in vivo*. To prevent auto-activation, the activation domain was de-activated either by removal of amino acids 102 to 216 or by the Pro4 mutation, in which Ala and Glu are mutated to Pro at amino acids 226 and 233 (indicated by a red-green stipple), respectively, which should prevent formation of a predicted α-helix required for transcription activation [134]. In two variants, the structured C-terminus of the protein was replaced by mCherry, represented by a pink/white striped bar.

bind Ubx, suggesting that all partners are unlikely to exclusively bind the positively charged homeodomain. Topology is a key factor affecting interactions between structured proteins, and sorting protein interactions based on the folds of the interacting partners can yield useful information about the nature of the interactions. Using the terminology of the Structural Classification of Proteins (SCOP) hierarchical classification database [176] , analysis of Ubx partners at the level of protein folds reveals that 23 of the 34 Ubx binding partners contain one of just 7 different folds, out of the 1195 folds identified by SCOP (**Table 3.1, Table 3.2**). All of the selected folds in Ubx-interacting proteins are enriched relative to the frequency with which these folds occur in the *Drosophila* proteome (**Table 3.3**). However, this level of enrichment may not be specific to Ubx: some folds are more prevalent in the *Drosophila* interactome. To determine whether these folds are more likely to bind Ubx than a random protein, we compared the extent of fold enrichment among Ubx partners with data derived from a high-throughput yeast two-hybrid experiment on *Drosophila* proteins [172] (**Figure 3.3**). Grouping the high-throughput data by fold did not change the scale-free nature of the network (**Figure 3.4**). The DNA/RNA binding 3-helical bundle fold, the α-α superhelix fold, and the dsRNA binding motif fold occur more frequently among Ubx-interacting proteins than in the *Drosophila* interactome, indicating that the enrichment of these folds among Ubx partners is not an artifact of their increased propensities to bind proteins in general (**Table 3.3**). For Ubx and each protein in the *Drosophila* interactome, we also calculated the number of folds each protein binds (F) divided by the number of proteins each binds (I) (**Figure 3.5**). Proteins with an F/I ratio approaching 1 do not select

partners by topology, whereas proteins with a low F/I ratio are highly selective. Whereas Ubx has an F/I ratio of 0.61, approximately 90% of the proteins analyzed have a higher F/I ratio, indicating they are less selective than Ubx. Despite the fact that large regions within Ubx are disordered and presumably extremely dynamic, these results suggest that topology is an important criterion by which Ubx selects protein partners.

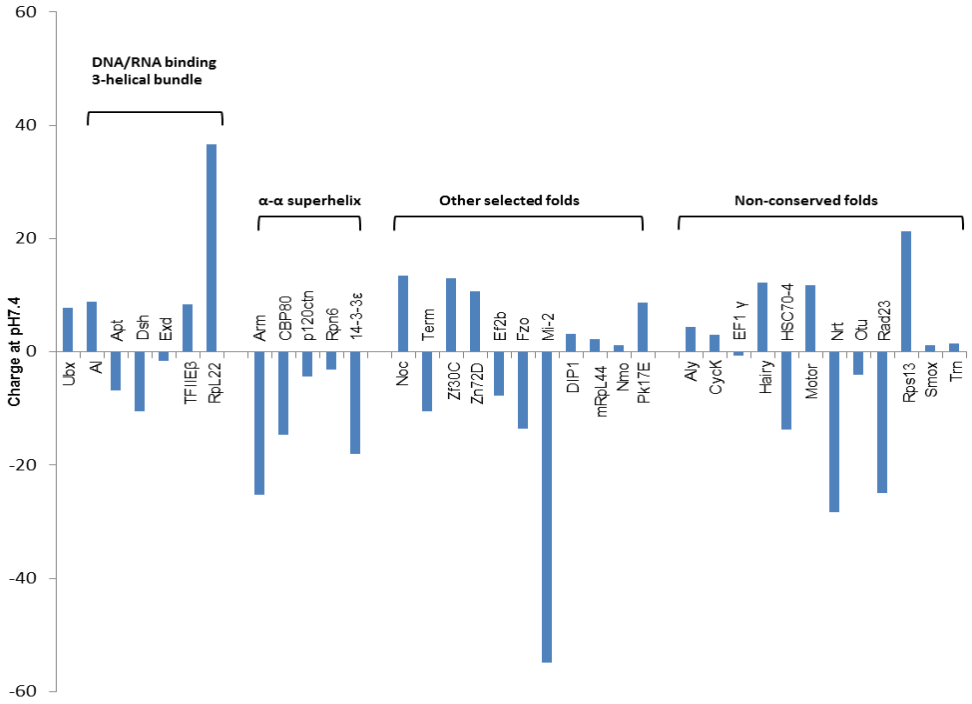

Figure 3.2 Ubx binds both positively and negatively charged proteins. The chart shows the predicted net charge at pH = 7.4 of Ubx and the subset of its partner proteins with known folds [79,132,136]. Abbreviations: Al, Aristaless; Aly, Always early; Apt, Apontic; Arm, Armadillo; CycK, Cyclin K; CBP80, Cap-binding protein 1; Dsh, Dishevelled; DIP1, Disconnected-interacting protein 1; Ef2b, Elongation factor 2b; EF1 γ, Elongation factor 1γ; Exd, Extradenticle; Fzo, Fuzzy onions; mRpL44, Mitochondrial ribosomal protein L44; HSC70-4, Heat shock protein cognate 4; Nmo, Nemo; Noc, No ocelli; Nrt, Neurotactin; p120ctn, Adherens junction protein p120; Otu, ovarian tumor; PK17E, Protein kinase-like 17E; RpL22, Ribosomal protein L22; Rpn6, Protease p44.5 subunit; Rps 13, Ribosomal protein S13; Smox, Smad on X; Term, terminus.

Table 3.1 Specific folds are enriched in Ubx-binding proteins.

Fold	Partner	Fold	Partner
DNA/RNA binding 3-helical bundle	RpL22	**P-loop containing NTP hydrolases**	EF2b
	Apt		Mi-2
	Al		Fzo
	Dsh		
	Ubx	**dsRBD-like**	DIP1
	Exd		mRpL44
α-α superhelix	Arm	**Ferridoxin-like**	EF2b
	Rpn6		Aly
	P120ctn		
	CBP80	**Protein kinase-like**	Nmo
	14-3-3ε		Pk17E
Zinc-Finger C2H2 and C2HC	Noc		
	Zf30C		
	Zn72D		
	Term		

Exd is a well-established Ubx binding protein [79], and Ubx cooperatively binds DNA [83]. All other Ubx binding partners were identified by yeast two-hybrid assays. Ubx binding partners were classified by the fold/shape according to SCOP. Folds with more than one partner were defined as "selected". The interactions with Term, Fzo, mRpL44, and Pk17E were reported by Giot *et al.* [172]. The remaining interactions were reported by Bondos *et al.* [132,136].

Table 3.2 Ubx partners with non-selected folds.

Fold	Partner	Fold	Partner
FYVE/PHD zinc finger	Mi-2	**Ribosomal protein S5 domain 2-like**	Ef2b
HLH-like	Hairy	**Ribonuclease H – like motif**	Hairy
Cyclin-like	CycK	**α/β-Hydrolases**	Neurotactin
XPC-binding domain	Rad23	**Chromo domain-like**	Mi-2
S15/NS1 RNA Binding Domain	RpS13	**Smad/FHA domain**	SMOX
β-Grasp	Rad23	**Reductase/isomerase/elongation factor common domain**	Ef2b
Smad MH1 domain	SMOX	**PDZ domain-like**	Dsh

A fold with only one partner was classified as a non-selected fold. Folds for Ubx binding partners were classified according to SCOP.

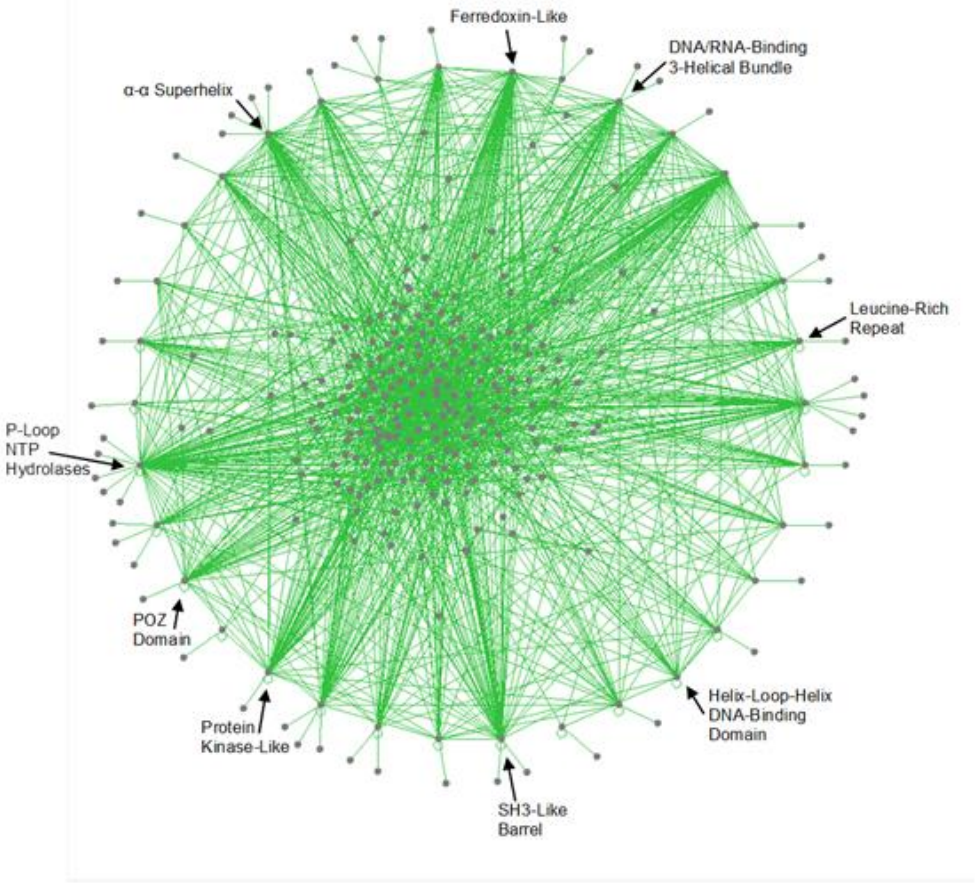

Figure 3.3 Maps of a large-scale *Drosophila melanogaster* yeast two-hybrid data parsed by fold, in which dots represents specific folds, and lines between dots depict interactions between the connected folds. All fold•fold interactions with a confidence score of at least 0.5 are shown. Intrafold interactions are depicted as loops which connect back to the originating node.

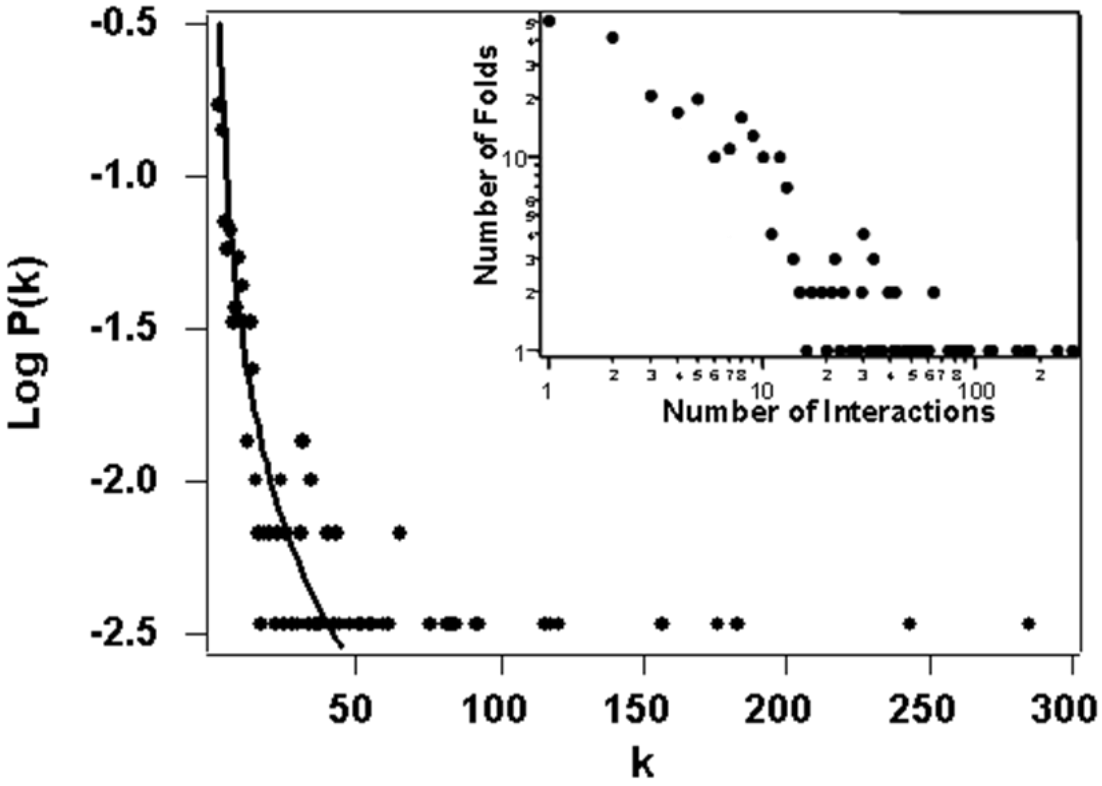

Figure 3.4 Probability distribution curves for *Drosophila* protein interactions from a large-scale yeast two-hybrid experiment parsed by fold. Data were fit to a truncated scale-free model. The scatter observed at high k is often observed in scale-free systems [172,177]. The similarity of these graphs to each other and with the protein data [172] indicates that grouping data by structure do not alter network character. Graphs depicting the number of superfamilies, proportional to P(k), that have k interactions is shown as an inset. Deviations from a straight line in these graphs are indicative of biological restrictions on highly interactive proteins within a scale-free network.

Table 3.3 Comparison of the occurrence of folds in the *Drosophila* proteome and interactome.

Fold	**Frequency in *Drosophila* proteome**	**Frequency in *Drosophila* interactome**	**Frequency in Ubx partner list**	**P-value of enriched fold relative to *Drosophila* proteome**	**P-value of enriched fold relative to *Drosophila* interactome**
DNA/RNA binding 3-helical bundle	2.7%	8.4%	17.6%	P<0.05	0.05<P<0.1
α-α superhelix	3.4%	7.2%	14.7%	P<0.05	0.05<P<0.1
Zinc Finger C2H2 and C2HC	3.7%	11.7 %	11.8%	P<0.05	0.05<P
dsRBD-like	0.2%	0.9%	5.9%	P<0.05	P<0.05
Protein kinase-like	2.8%	5.6%	5.9%	0.05<P	0.05<P
p-loop containing NTP hydrolases	5.6%	8.4%	8.8%	0.05<P	0.05<P
Ferridoxin-like	2.8%	8.1%	5.9%	0.05<P	0.05<P

p-value of enriched fold relative to *Drosophila* proteome/interactome was generated using Chi-Squared test.

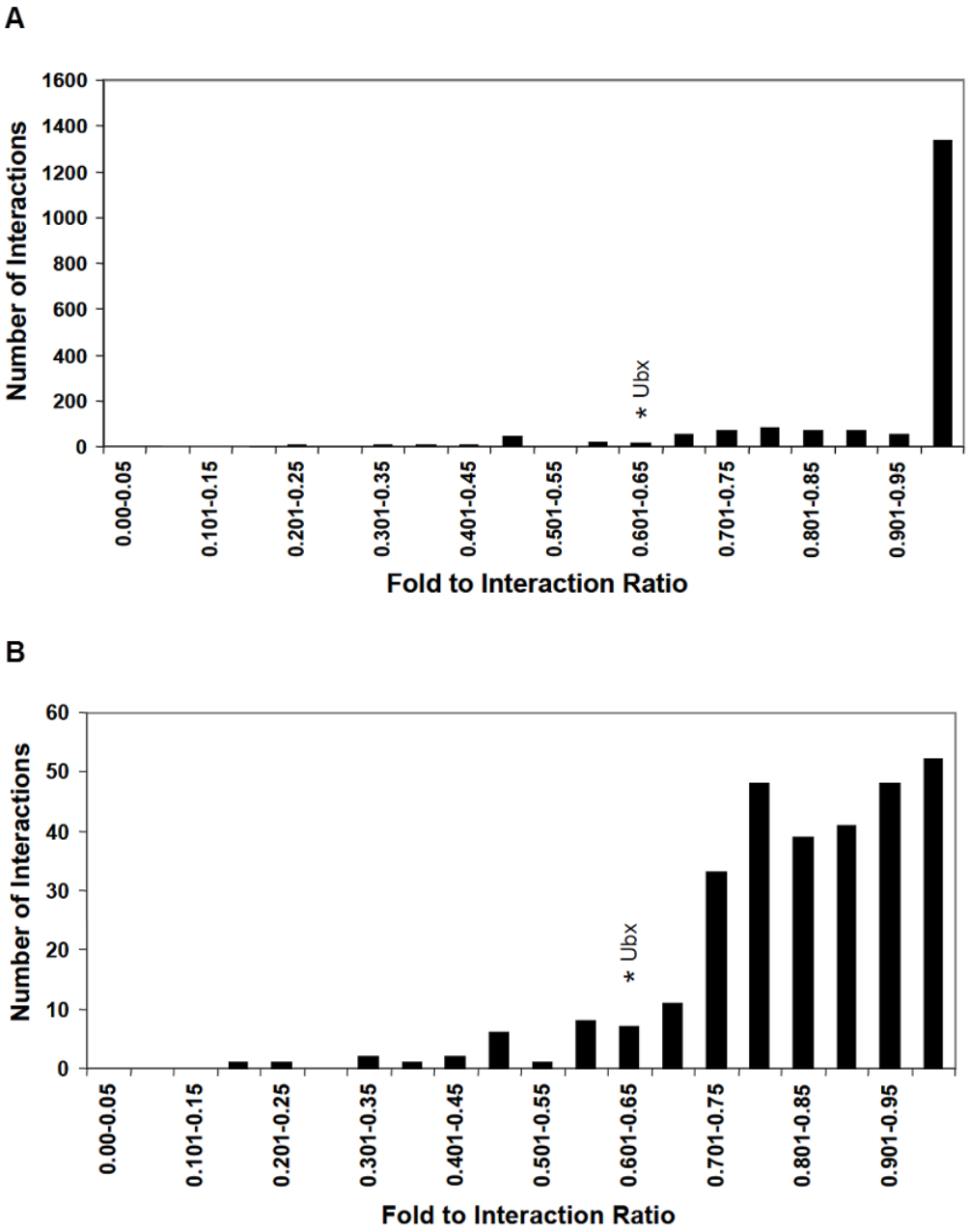

Figure 3.5 The distribution of the fold to interaction ratio (F/I). (A) All single domain proteins. (B) All single domain proteins with more than one partner. Proteins with a high ratio do not select protein partners by fold, whereas interactions with proteins with a low ratio have strong fold preferences. Ubx has an F/I ratio of 0.61, indicating a strong ability to select partners by fold.

Importantly, proteins unlikely to naturally bind Ubx can have the same fold as true Ubx partners. For example, DIP1 alters transcription regulation by Ubx in cell culture assays and inhibits Ubx function *in vivo* [136]. DIP1 has the same fold as mRpL44, a mitochondrial ribosomal protein that should not co-localize with Ubx *in vivo*. Therefore, even if some Ubx-protein interactions lack a biological role, they can still yield information regarding the physicochemical properties of partner proteins bound by Ubx *in vivo*. This phenomenon underscores the importance of partner topology in the selection of protein partners by Ubx.

The enrichment of particular folds among Ubx partners may be caused by Ubx preferring to bind the surface topologies created by these folds. Alternately, the types of proteins Ubx binds *in vivo*, transcription factors and cell signaling proteins, may be enriched in these folds (*e.g.*, a DNA/RNA binding 3-helical bundle fold). Consequently, the "selected folds" may be enriched among Ubx partners due to their cellular function rather than presentation of a binding interface on the surface of the selected fold. In order to determine whether the selected folds are sufficient to mediate Ubx interactions, we used yeast two-hybrid assays to probe whether Ubx interacts with the regions of partner proteins that correspond to the selected topology. We utilized the yeast-two hybrid method because (i) these assays do not interfere with Ubx binding to these partners, (ii) these assays do not rely on other Ubx functions, such as DNA binding or transcriptional regulation, iii) yeast two-hybrid assays allow quantitative comparison of the strength of binding, and (iv) many partners identified by yeast two-hybrid assays also alter Ubx function *in vivo* [79,132,136,178], demonstrating this method likely reflects

native protein interactions involving Ubx. We created two constructs: a single α-α superhelix domain from Arm (amino acids 155-273) and a DNA/RNA binding 3-helical bundle domain from Al (amino acids 81-142). We hypothesized that the individual domain in an Ubx partner is sufficient to interact with full-length Ubx without surrounding sequences. To prevent reporter gene activation by Ubx in the absence of partner binding, a full-length Ubx mutant (UbxIb Pro4) was used that is incapable of transcription activation [134]. Individual yeast two-hybrid experiments between these two isolated domains and UbxIb Pro4 [134], exhibit similar levels of reporter gene expression as for experiments in which UbxIb Pro4 binds the corresponding full-length partners (**Figure 3.6**). This result indicates that the α-α superhelix and DNA/RNA binding 3-helical bundle folds in these proteins are sufficient for Ubx interaction.

Our general approach is to remove either intrinsically disordered or structured sequences within Ubx (Figure 3.1) to assess whether these regions impact binding to protein partners. All of the Ubx mutants were carefully designed to minimize the impact on regions of Ubx structure that are well-folded. In the N216 and N103 Ubx truncation mutants, amino acids 2-215 (Regions 1 and 2) or 2-102 (Region 1) are removed, respectively (Figure 3.1). These variants have been successfully used for both *in vitro* DNA binding assays and yeast one- and two-hybrid experiments [65,134]. Indeed, both truncation mutants are soluble, active monomers capable of binding DNA with an

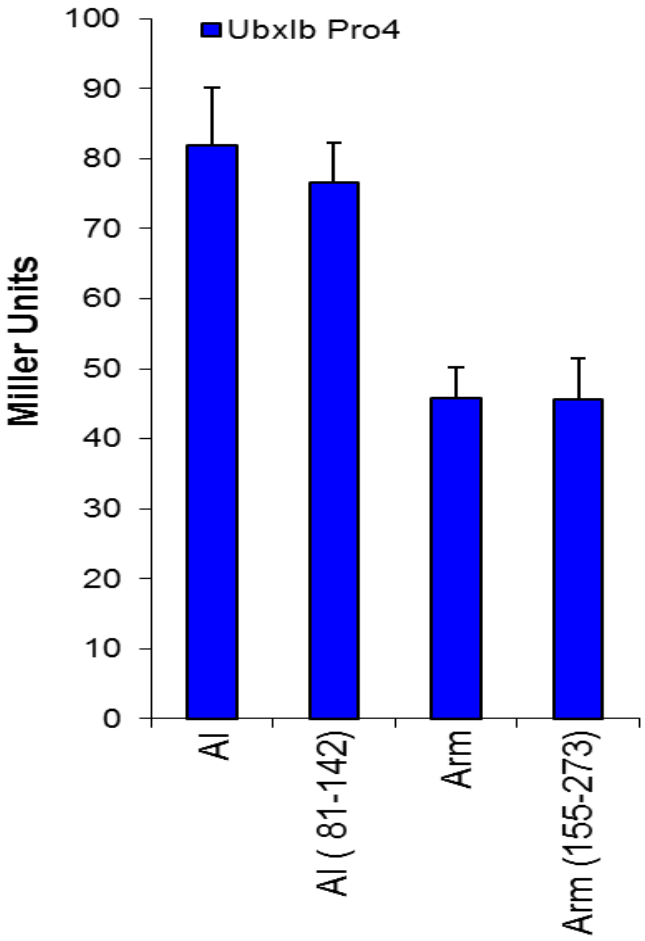

Figure 3.6 An individual partner domain is sufficient for Ubx binding. Full length Al and Arm have similar interaction strength as individual domains derived from Al (residues 81-142) and Arm (residues 155-273) with UbxIb Pro4. The intensity of the β-galactosidase reporter gene, reported as Miller Units, signal for each partner is similar to its respective single-domain variant.

affinity comparable to full-length Ubx [65]. Similarly, we made Ubx variants with an internal deletion (Δ103-216) which removes Region 2. Other Ubx mutants with internal deletions in this region are also soluble and capable of binding DNA [65]. Furthermore,

the length of this region is significantly reduced in natural Ubx orthologues (**Figure 3.7**) [179], consistent with observations that this internal deletion in *Drosophila* Ubx does not significantly disrupt the remaining Ubx structure. The C-terminal disordered region (Region 4) spans an alternatively spliced region of Ubx. The natural Ubx isoform UbxIVa removes nearly all (90%) of the intrinsic disorder in this region, and was used to assess the contribution of Region 4 to protein interactions.

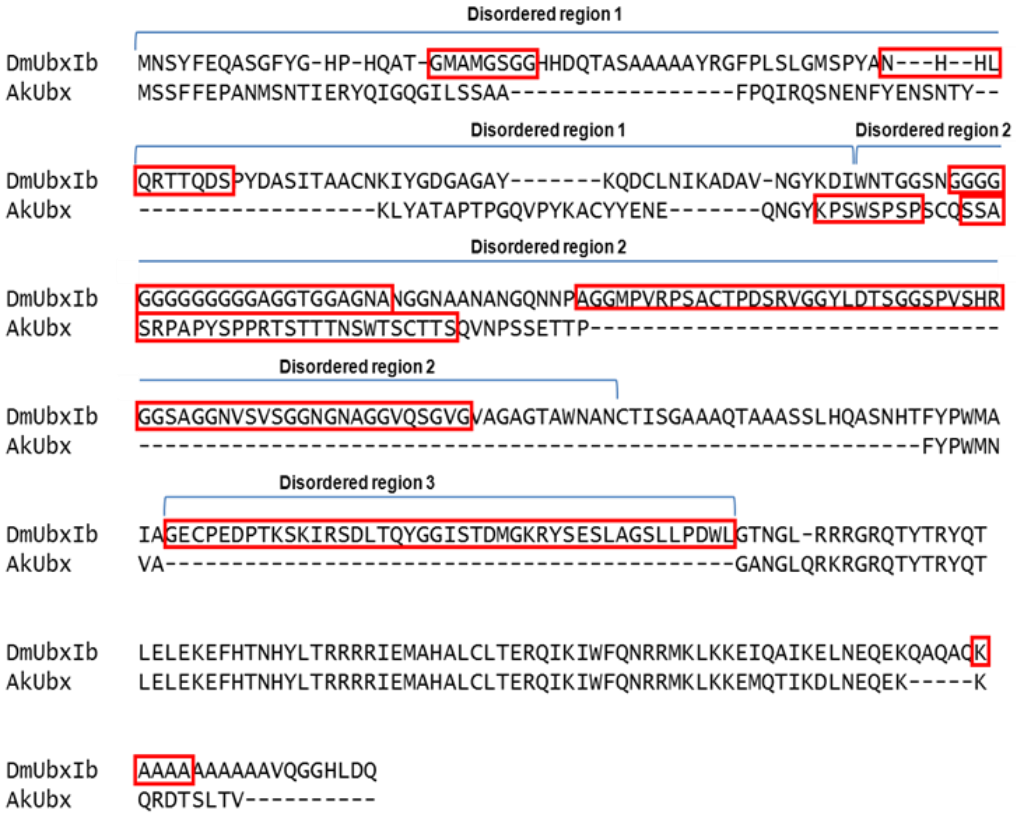

Figure 3.7 AkUbx, a Ubx orthologue with only one intrinsically disordered region, cannot bind *Drosophila* Ubx partners. Sequence alignment between *Akanthokara kaputensis* Ubx (AkUbx) and *Drosophila melanogaster* Ubx showing the locations of disordered residues (red boxes) and the three disordered regions (blue labels).

Because Ubx is fused to the LexA DNA-binding domain in the yeast two-hybrid assay, the transcription activation domain in Ubx was deactivated in each mutant to prevent the LexA-Ubx fusion from activating the reporter gene and generating false positive signals. This deactivation was accomplished either by removing a critical portion of the activation domain (amino acids 103-216) or by including the mutations A226P/Q233P, abbreviated as "Pro4", to unfold a putative α-helix required for transcription activation [134]. None of the Ubx variants in this study were able to activate transcription on their own, or bind products of the empty bait vector pB42 (**Figure 3.8**). Furthermore, the expression levels of all Ubx variants in yeast were similar, except the two Ubx fusion proteins in which the DNA-binding homeodomain was replaced with mCherry, which were expressed at much higher levels (**Figure 3.9**).

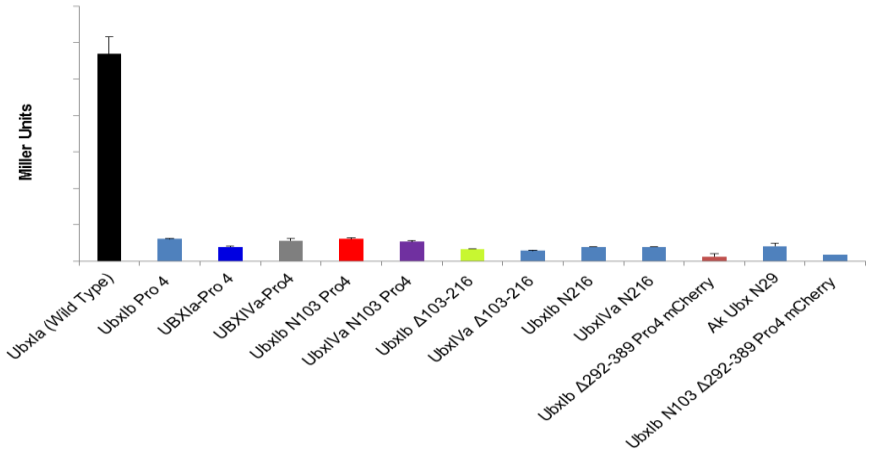

Figure 3.8 Ubx variants did not interact with B42 protein activation in the absence of Ubx partners. Yeast two-hybrid results for wild type full length Ubx or Ubx variants with truncation and/or Pro4 mutation showed no significant interaction with B42 protein activation domain from β-galactosidase reporter gene expression, listed as Miller Units.

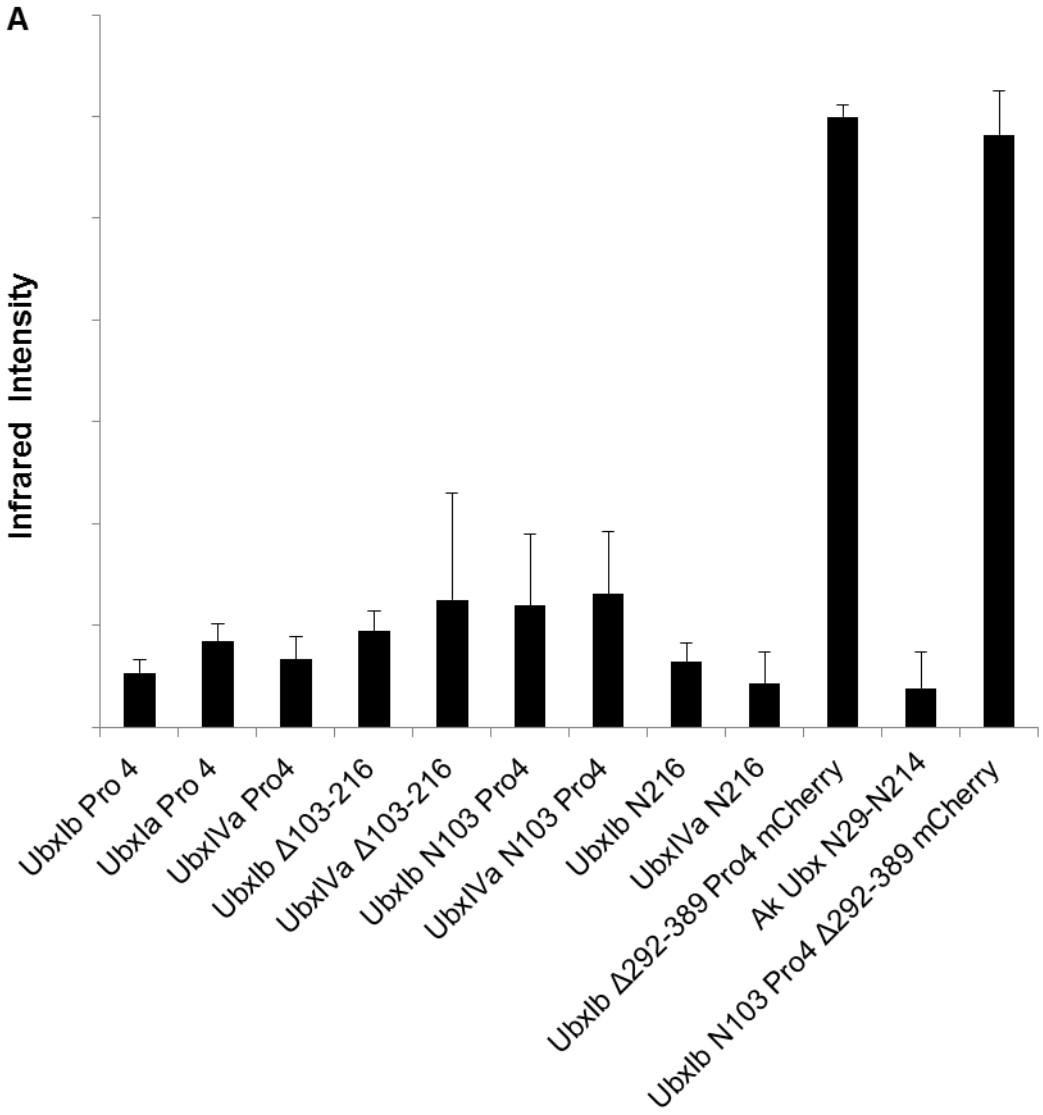

Figure 3.9 Ubx variants expression level does not correspond with partner interaction strength. (A) Quantitative Western blotting result for Ubx variants protein expression in yeast (Strain:EGY48 transformed with p8op-LacZ reporter plasmid). (B) Weak correlation between yeast two-hybrid result and Ubx variants protein expression without outliers (R^2=0.1403). Inset plot shows the influence of the two outliers (UbxIbN103 Pro4 Δ292-389 mCherry and UbxIb Δ292-389 Pro4 mCherry) on the correlation between yeast two-hybrid result and the Ubx variants protein expression.

To clarify which portions of the Ubx sequence are included or removed in each variant, the name of each Ubx variant in this text is introduced followed by a notation representing the Ubx sequences present in parentheses. We have divided the Ubx sequence into 5 regions (**Figure 3.1**). The number representing each region will be preceded by an S if the region is structured, a D if the region is disordered, and SD if that region contains both structured and disordered elements. Thus the sequence of full-length, wild-type Ubx would be depicted as (SD1,D2,S3,D4,S5). Regions that are missing or mutated in a particular variant are designated by 0. The UbxIb Pro4 mutant, in which the helix in region S3 has been destabilized by mutation to prevent transcription activation, would be notated as (SD1,D2,0,D4,S5).

We made a series of Ubx truncations or mutations to sequentially test whether each portion of the Ubx sequence contains a critical partner binding site (**Figure 3.10**). All data were compared with UbxIb Pro4 (SD1,D2,0,D4,S5), a full-length variant of Ubx which binds all partners but cannot activate the reporter gene in the absence of partner interaction. UbxIb N103 Pro4 (0,D2,0,D4,S5), in which the structured and disordered elements in Region 1 were removed, still bound the partner proteins, indicating Region 1 is dispensable for partner binding. UbxIb Δ103-216 (SD1,0,S3,D4,S5), which removes the intrinsically disordered Region 2, also bound some partners. The previously established ability of UbxIb with the Pro4 mutation (SD1,D2,0,D4,S5) to bind partners indicates that the helix in Region 3 cannot be responsible for partner binding [132,136]. Conversely, the Pro4 mutations are not required for partner binding, because partners bind UbxIb Δ103-216 (SD1,0,S3,D4,S5),

which retains the wild-type helix sequence in Region 3. UbxIVa Pro4 (SD1,D2,0,0,S5) binds partner proteins, even though the disordered Region 4 has been removed. Finally we created UbxIb Δ292-389 Pro4 mCherry (SD1,D2,0,D4,0), in which the structured C-terminus (Region 5) has been removed and replaced with the mCherry protein sequence. mCherry alone is unable to bind any of the Ubx partners (data not shown). However, UbxIb Δ292-389 Pro4 mCherry bound all partners, indicating that Region 5, which includes the DNA-binding homeodomain, is not necessary for partner binding. Collectively, these data indicate that more than one region of Ubx is required for protein interactions.

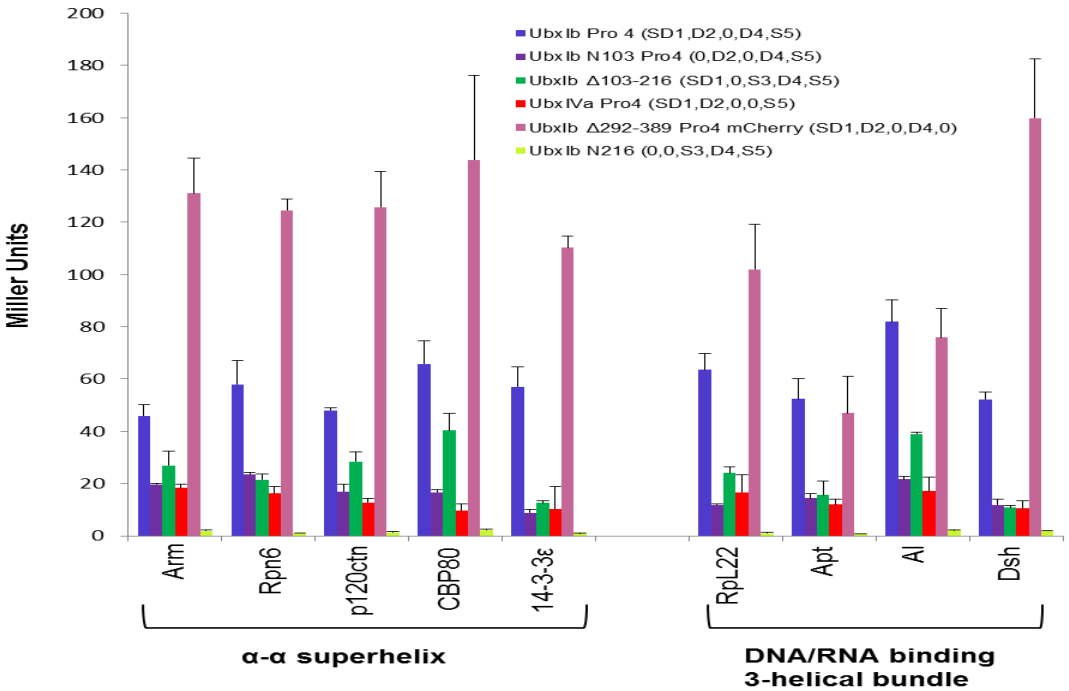

Figure 3.10 Partner proteins bind more than one region in Ubx. Yeast two-hybrid results for Ubx variants in which each region of Ubx has been sequentially mutated or deleted. Each of these variants retain some ability to bind Ubx relative to UbxIbN216 (0,0,S3,D4,S5). Partners are grouped based on the fold they have in common.

The intrinsically disordered regions in Ubx are necessary for protein interactions

The next step was to identify some portion of the Ubx protein that is necessary for partner binding. The MORF and ANCHOR algorithms both identify many short motifs in the intrinsically disordered regions of Ubx that have the potential to engage in protein interactions (**Figure 3.1B**), suggesting the intrinsically disordered regions may be collectively required for partner binding. In order to test this hypothesis, we compared binding by the structured versus disordered halves of the Ubx sequence. UbxIVa N216 (0,0,S3,0,S5) lacks all of the intrinsically disordered regions but retains two of the three regions containing structure. This mutant is based on the natural UbxIVa mRNA splicing isoform, which removes Region 3, and the N216 truncation, which removes Regions 1 and 2 (**Figure 3.1D**). The remainder of this Ubx variant is almost entirely structured (> 90%).

Conversely, UbxIb Δ292-389 Pro4 mCherry (SD1,D2,0,D4,0) retains all of the disordered regions, but lacks the Region 3 helix and the structured homeodomain in Region 5. UbxIVa N216 (0,0,S3,0,S5), which lacks intrinsically disordered sequences, was unable to bind all partner proteins, whereas UbxIb Δ292-389 Pro4 mCherry (SD1,D2,0,D4,0), which contains all of the intrinsically disordered sequences, bound all partners. In fact, this variant yielded an even more intense reporter signal than Ubx alone. Much of this elevated signal can be attributed to the increased expression level of UbxIb Δ292-389 Pro4 mCherry relative to the Ubx variants lacking mCherry (**Figure**

3.11). Thus Regions 1, 2, and 4, which include all of the intrinsically disordered regions in Ubx, are sufficient for partner binding.

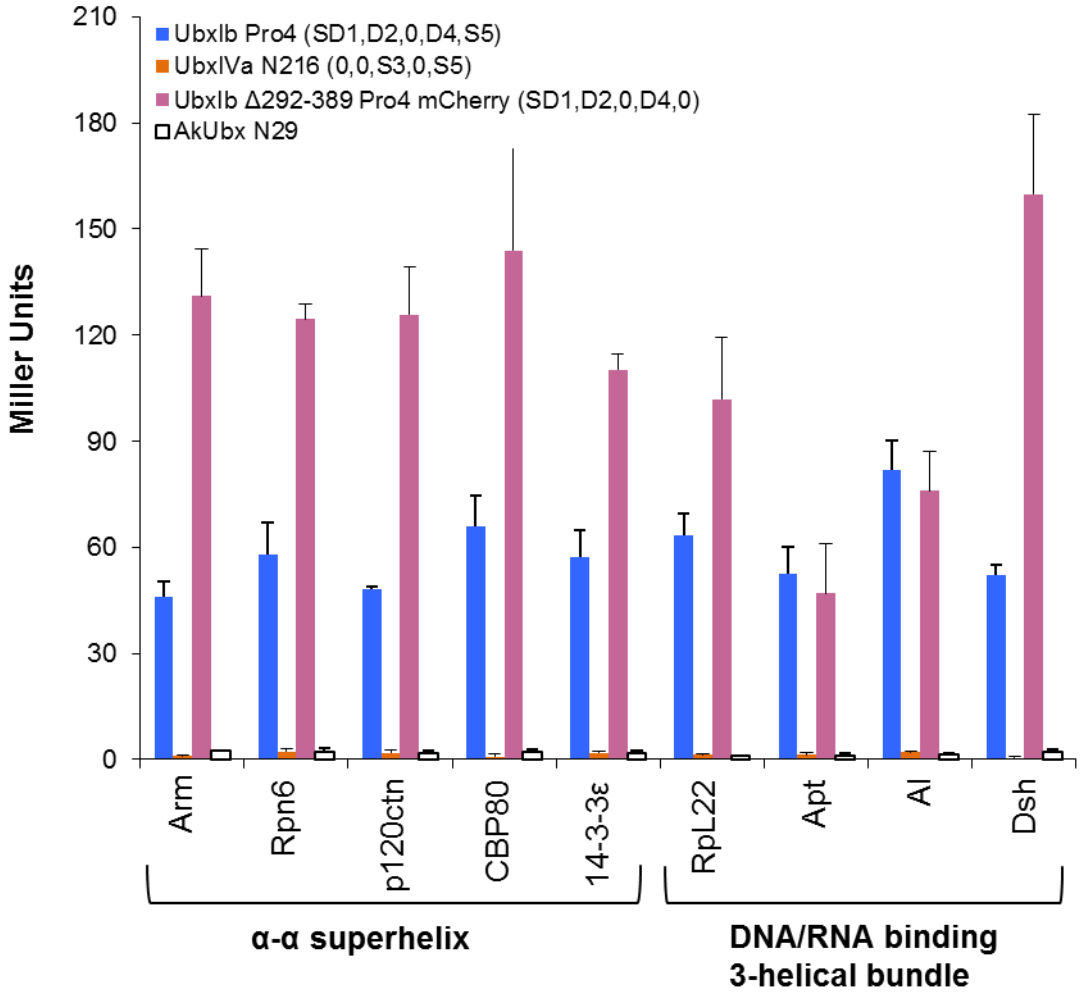

Figure 3.11 The intrinsically disordered regions in Ubx are necessary for protein interactions. Yeast two-hybrid indicates that Ubx variants, either lacking all disordered regions (UbxIVa N216) or all structured regions (UbxIb N103 Δ292-389 Pro4 mCherry), cannot bind partner proteins. Likewise, AkUbx, a primitive Ubx orthologue derived from *Acanthokara kaputensis*, naturally lacks most of the disordered sequences and is also unable to bind partner proteins.

One concern is that the structured regions may contribute to binding in the full-length protein, but are mis-positioned by the absence of the disordered regions in the UbxIVaN216 (0,0,S3,0,S5) mutant. To test the latter possibility, we examined whether the Ubx partners could bind an orthologue of Ubx derived from the velvet worm *Akanthokara kaputensis* (AkUbx), an onychorphoran whose last common ancestor with *Drosophila* lived 540 million years ago. Hox proteins in this ancient organism only have very basic molecular functions, which are reflected in the relatively simple and repetitive body plan of the animal [21,179]. When expressed in *Drosophila*, AkUbx can replicate some, but not all, of the functions of *Drosophila* Ubx. Alignment of the Ubx and AkUbx sequences demonstrates that the disordered sequences in Regions 1 and 3 are absent in this ancient Ubx orthologue, and roughly half of the disordered sequences in Region 2 are missing (**Figure 3.7**). In contrast, the homeodomain and much of the structured portions of Region 1 are preserved. Therefore, by testing whether AkUbx can bind Ubx partners, we can use a native, folded Ubx orthologue to observe whether the loss of most of the intrinsically disordered regions prevents partner interaction. AkUbx showed little to no interaction with Ubx partners in the yeast two-hybrid assay (**Figure 3.11**). These results confirm that the disordered regions in Ubx are required for partner binding. Because no individual disordered region is solely responsible for partner interactions, we conclude that the intrinsically disordered regions in Ubx must cooperate to bind partner proteins. The requirement of multiple, non-contiguous disordered regions for partner interaction has been observed previously for other proteins [89,155].

Either Region 1 or Region 4 is required as a scaffold to position intrinsically disordered Ubx sequences

To try to identify a minimal region of Ubx required for protein interactions, we began with UbxIb N103 Pro4 (0,D2,0,D4,S5), a truncated variant which binds all partner proteins, and iteratively removed each remaining structured or disordered region (**Figure 3.12A**). UbxIb N216 (0,0,S3,D4,S5), which additionally removes the disordered Region 2, cannot bind any of the Ubx partners. Likewise, UbxIVa N103 Pro4 (0,D2,0,0,S5) which removes the disordered Region 4, cannot bind any of the Ubx partners. Finally, the structured C-terminus was removed in UbxIb N103 Δ292-389 Pro4 mCherry(0,D2,0,D4,0), which also cannot bind Ubx partners. Therefore Regions 2, 4, and 5 can be considered a minimal partner interaction region.

These data apparently conflict with data from the UbxIb Δ292-389 Pro4 mCherry (SD1,D2,0,D4,0) mutant, which also is able to bind all partners but lacks the S5 region in the minimal partner interaction region described above. Instead, this variant includes the SD1 region with mixed structure and disorder. Removal of the SD1 region to create UbxIb N103 Δ292-389 Pro4 mCherry (0,D2,0,D4,0) prevents binding to Ubx partners (**Figure 3.12B**). Therefore the UbxIb Δ292-389 Pro4 mCherry (SD1,D2,0,D4,0) variant constitutes a second minimal partner interaction region. The presence of two minimal partner interaction regions that are compatible with many Ubx-binding proteins may provide an opportunity for multiple partners to simultaneously bind Ubx. Inclusion of multiple binding sites has been observed for other disordered proteins [132]. The fact that both minimal partner binding regions are mainly composed of intrinsically

disordered sequences highlights the important role that disorder plays in interactions mediated by Ubx.

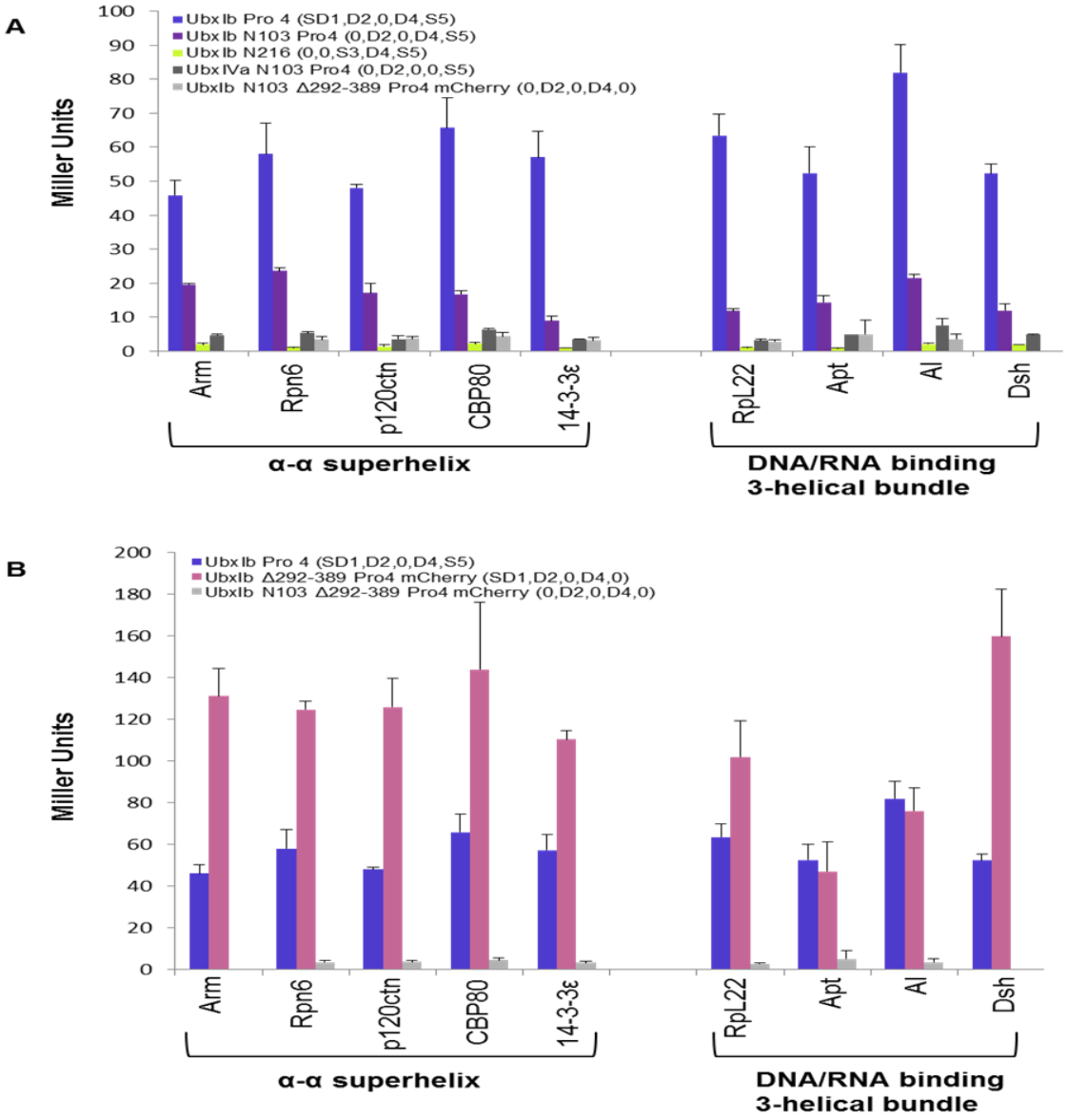

Figure 3.12 Defining minimal partner interaction domains. Analysis of yeast two-hybrid data using Ubx variants identifies two overlapping minimal partner interaction domains: UbxIb N103 Pro4 (0,D2,0,D4,S5) (Panel A) and UbxIb Δ292-389 Pro4 mCherry (SD1,D2,0,D4,0). Both minimal partner binding domains include the disordered Regions 2 and 4.

The intrinsically disordered regions in Ubx differentially contribute to partner binding

Although the disordered regions are required for partner binding, different Ubx partner proteins may best interact with a subset of the Ubx disordered domains. If so, then identifying which intrinsically disordered region within Ubx is preferred by partner proteins could provide clues regarding the functional outcome of each partner interaction. For example, a partner that bound the Ubx transcription activation domain might alter the balance between transcription activation and repression by Ubx [78].

Although the experiments described above suggest that the disordered regions are necessary for binding, they do not reveal which of the disordered regions are bound by partners. The most straightforward approach is to compare a Ubx variant with no disordered regions (UbxIVa N216) with a variant which includes just one of the disordered regions (Region 1, UbxIVa Δ103-216 (SD1,0,S3,0,S5); Region 2, UbxIVa N103 Pro4 (0,D2,0,0,S5); Region 4, UbxIb N216 (0,0,S3,D4,S5). However, little to no partner binding was observed for all three of these variants, indicating more than one disordered region must be present for any partner to bind, consistent with the identification of the minimal binding regions described above (**Figure 3.13**).

To test the strength of different cooperative units, we compared variants missing each of the three disordered regions in turn (Region 1 deleted, UbxIb N103 Pro4 (0,D2,S3,D4,S5); Region 2 deleted, UbxIb Δ103-216 (SD1,0,S3,D4,S5); Region 4 deleted, UbxIVa Pro4 (SD1,D2,S3,0,S5). As already discussed, each of these mutants is still able to bind Ubx partner proteins. However, partner affinity is reduced to different extents (**Figure 3.10**). Binding by 14-3-3ε, RpL22, Apt, and Dsh was equally affected

by removing Regions 1, 2, or 3. Since a large percentage (≥ 59%) of the signal was lost in each of these interactions, an interesting interpretation is that these proteins may simultaneously bind all three regions. For other partners, the magnitude of the reduction in protein interaction varies for the three regions. Whereas removing Regions 1 and 3 had a significant effect on binding all partners, for a subset of partners (*e.g.*, p120ctn, Al, and CBP80), removing Region 2 had less impact. The ability of these three variants to bind partner proteins does not appear to correlate with the topology of the partner.

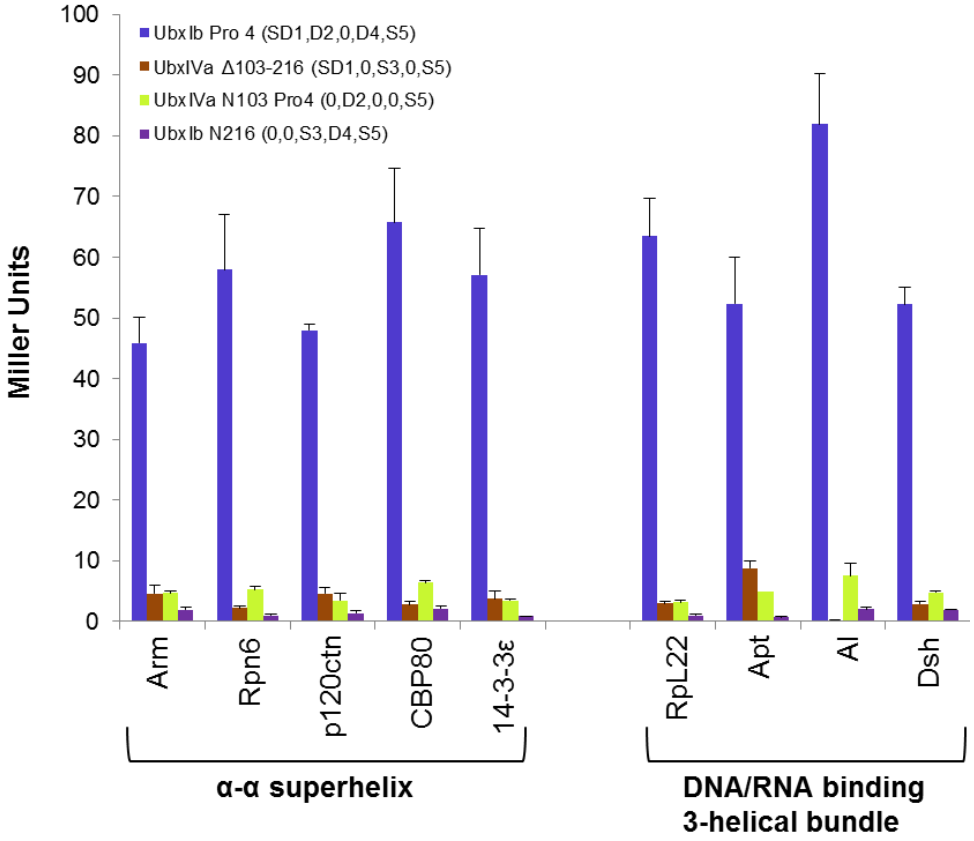

Figure 3.13 Ubx-interacting proteins cooperatively bind Regions 1, 2, and 3, all of which contain intrinsically disordered sequences. No single Ubx disordered region is sufficient to support partner binding, suggesting multiple disordered regions function as a cooperative unit.

Partners differentially interact with alternatively spliced isoforms of Ubx

Binding by all partners relies to some extent on contacts with Region 4, which contains sequences included in or excluded from Ubx by alternative mRNA splicing. Expression of Ubx splicing isoforms is regulated in a stage- and tissue-specific manner during *Drosophila* embryonic development [180]. Ubx isoforms are generated through differential inclusion of three different microexons in *ubx* mRNA, all of which code for protein sequences within Region 3: the b element, microexon I, and microexon II (**Figure 3.1**). Expression of these three splice variants elicits different phenotypes *in vivo* [2,88]. To determine the impact of alternative splicing on partner interactions, we compared the ability of UbxIb Pro4 (containing all three microexons), UbxIa Pro4 (containing the mI and mII microexons) and UbxIVa Pro4 (containing no microexons) to bind partner proteins.

Removal of all three microexons in the UbxIVa Pro4 variant reduces the ability of Ubx to bind all partners relative to UbxIb Pro4 (**Figure 3.14A**). This reduction ranges from 85% (CBP80) to 60% of binding lost (Arm). For some partners (RpL22, Apt, and Dsh), removal of only the 9-amino acid b element altered binding to the same extent as removing all three microexons, indicating these interactions are critically dependent on the presence of the b element. We cannot discern from these experiments whether the b element contributes key chemical groups required for interaction or simply lengthens the intrinsically disordered region to generate a sufficiently large binding interface. Partner affinity has also been linked to the dynamics of the disordered region [181]. Intriguingly, disorder prediction algorithms yield very different scores for different Ubx

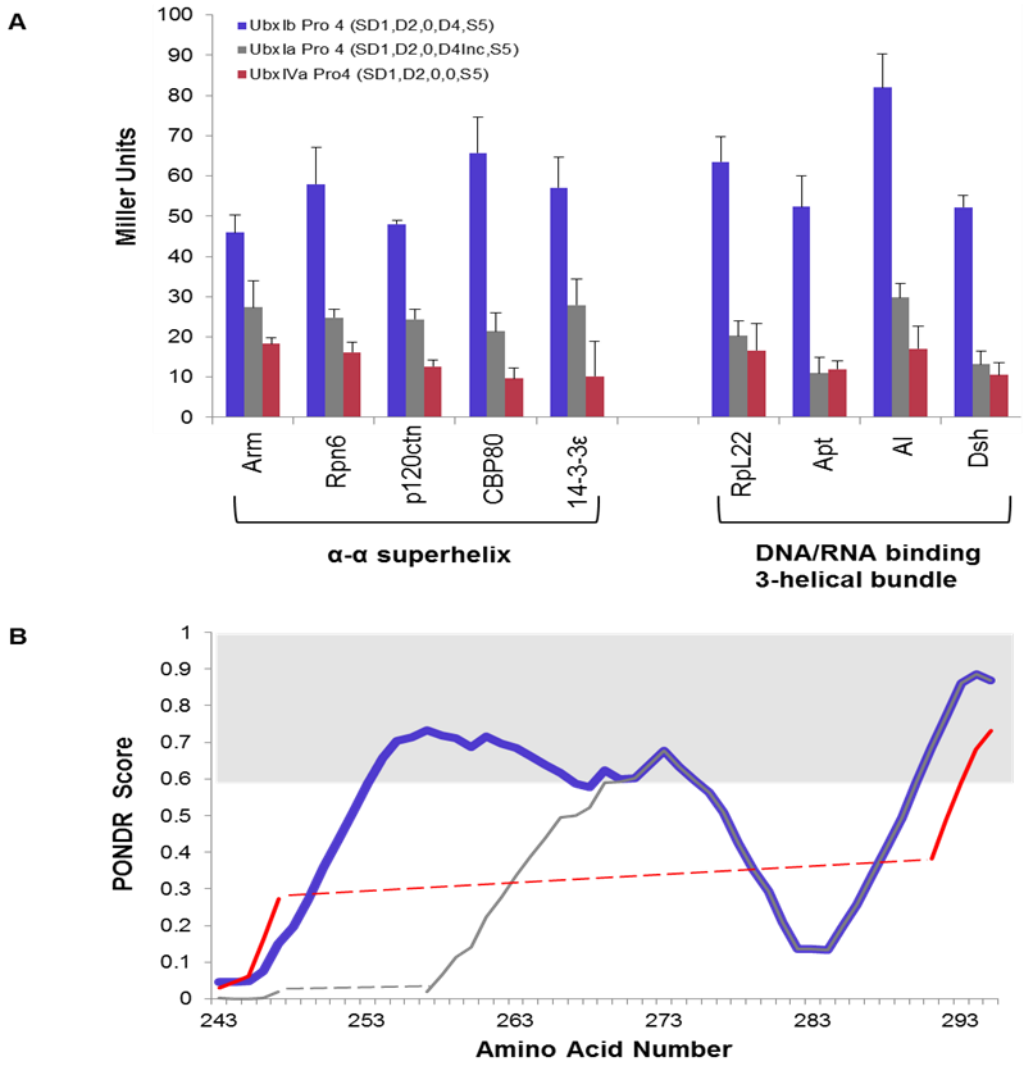

Figure 3.14 Partner topology determines Ubx splicing isoforms are differentially able to bind partner proteins which correlate with the predicted disorder profile of isoforms. (A) Whereas all partners with an α-α superhelix fold bind UbxIa better than UbxIVa, among partners with a DNA/RNA binding 3-helical bundle fold only Al binds these two Ubx isoforms differently. "Inc" denotes Region 3, the microexon region, and is incomplete in the UbxIa splicing isoform. The disordered regions remaining in each variant are listed in parentheses after the protein name. (B) Colored lines represent intrinsic disorder prediction scores for the microexon region for different Ubx splicing isoforms, generated using the PONDR VL-XT algorithm [182]. Blue line, UbxIb which has all 3 microexons; grey line, UbxIa which lacks the 9 amino acid b element; red line, UbxIVa, which lacks all 3 microexons. Dashed lines connect data across the microexon sequences removed in the shorter isoforms. The extent of predicted disorder (score >0.6, region shaded light grey) correlates with the ability to bind the tested partner proteins.

splicing isoforms (**Figure 3.14B**). These differences suggest that Ubx dynamics may influence Ubx-partner binding.

Partner topology generally correlates with partner affinity for different Ubx splicing isoforms. All proteins with an α-α superhelix fold bind UbxIa better than UbxIVa, whereas all but one protein (Al) with a DNA/RNA binding 3-helical bundle fold bind UbxIa and UbxIVa equally well (**Figure 3.14A**). This correlation reflects similarities in binding by partners with the same fold. Interestingly, the Eukaryotic Linear Motif (ELM) prediction algorithm revealed a 14-3-3ε binding motif in the mII microexon sequence [183-185], which may explain why 14-3-3ε binds UbxIa Pro4, but not UbxIVa Pro4, which is missing this motif. In general, the proteins with a strong isoform effect (UbxIb > UbxIa > UbxIVa) were all negatively charged (14-3-3□, Al, Arm, CBP80, p120ctn, and Rpn6), perhaps due to the position of the alternatively spliced microexons adjacent to the positively charged homeodomain. The previously characterized Ubx partner, Exd, also has a net negative charge and differentially binds Ubx isoforms [79]. Proteins that bind UbxIa and UbxIVa equally well can be either positively or negatively charged. Thus, although all partners bind disordered regions, the topology and charge of the partner protein correlate with their ability to bind different Ubx isoforms. Differences in the affinity of partners for Ubx isoforms create the potential for *ubx* mRNA splicing to regulate Ubx-partner interactions *in vivo*.

Discussion

We have demonstrated that partner topology is a key aspect of protein interactions formed by the intrinsically disordered regions of the *Drosophila* Hox protein Ubx. Greater than 60% of Ubx-binding proteins have a fold in common with at least one other Ubx partner, and Ubx binds the selected fold within these proteins. Other laboratories have also identified disordered proteins that bind multiple partner proteins with similar structures [157,186]. These partners were related proteins from the same protein family. In contrast, Ubx binds structurally similar, yet widely diverse proteins with very different chemical natures and molecular functions. Binding multiple partners with similar structures may reduce frustration in the Ubx-partner interface compared to interactions disordered proteins and an array of partner topologies [83].

A model for the role of structure in Ubx-partner binding

Many proteins that interact with intrinsically disordered proteins or regions bind a MORF, a short motif within a disordered region of a protein that often folds upon partner binding. In the case of Ubx, three large disordered regions all simultaneously contribute to partner binding. The fact that the topology of the partner protein is important suggests that the disordered regions may need to be positioned in a specific manner in order to maximize interactions with the partner protein. This model fits with our data on the role of Regions 1 (partially structured) and 5 (structured) in partner binding. Structure-containing region is not sufficient for partner binding and partner binding can occur in the absence of either region. The inability of AkUbx, a natural Ubx orthologue which lacks most of the disordered regions, to bind partners demonstrates

that the lack of binding is not an artifact induced by mis-positioning structured regions in Ubx mutants. However, either Region 1 or Region 5 must be present for the disordered regions in Ubx to bind partner proteins, suggesting either of these regions can correctly position the disordered domains for partner binding. This positioning may involve binding the disordered regions: the Ubx homeodomain, which is located in Region 5, has a DNA/RNA-binding 3-helix bundle fold, one of the two major folds selected by Ubx. The intrinsically disordered regions of Ubx directly interact with the homeodomain to alter its DNA binding affinity and specificity and with each other to enable cooperative DNA binding *in vivo* and materials formation *in vitro* [124,187].

Implications for Ubx function

The identification of partner-binding regions within Ubx, and the overlap of these regions with each other and with known functions or regulatory mechanisms, has important implications for regulating tissue-specific Ubx function *in vivo*. Whereas some partners bind all three regions to an equal extent (14-3-3ε, RpL22, Apt, and Dsh), other partners depend more heavily on Regions 1 and 3 for binding to Ubx (Arm, p120ctn,CBP80, and Al). Ubx partners reliant on the same regions of Ubx for binding may compete for binding to these regions. For partners that bind equally well to all three intrinsically disordered regions, the long length of these regions, may enable more than one partner to simultaneously bind Ubx. Indeed, other proteins with long disordered regions can act as a scaffold to simultaneously bind multiple partner proteins and create multi-functional complexes [188,189]. In the context of transcription regulation, using Ubx as a scaffold for constructing a multi-protein transcription factor complex allows

Ubx-mediated transcription regulation to respond to input from multiple protein systems [70]. The correct, tissue-specific regulatory complex would be stabilized by Ubx-DNA interaction, partner-DNA interactions, and partner-Ubx interactions.

All Ubx partners rely, to some extent, on Region 2 for binding. Since Region 2 includes critical sequences for transcription activation by Ubx [134], partner binding may modulate the ability of Ubx to activate transcription. Further, multiple phosphorylation sites exist within Regions 1 and 2 [190], suggesting that phosphorylation of this region *in vivo* has the potential to regulate Ubx activity by removing bound proteins, stabilizing protein interactions, and / or altering which proteins are bound to this region. Alternative splicing alters the ability of Ubx to bind partners, a regulatory mechanism used to regulate the other protein interactions [191]. Alternative splicing, combined with protein partner availability may also impact how Ubx selects DNA binding sites. Ubx binds three different categories of DNA sequences, defined by the protein interactions in which Ubx engages: i) multiple, closely spaced Hox binding sites that permit cooperative Ubx binding, ii) single or multiple Hox binding sites interspersed with binding sites for other transcription factors, or iii) Hox-Exd heterodimer binding sites (**Figure 3.15**). The partner binding preferences of each Ubx isoform, combined with the presence or absence of partners in the tissues in which that isoform is expressed, could determine which subset of DNA sequences are regulated by Ubx in each tissue. For example, the presence of the b element enhances binding by the partners examined in this study, but reduces binding by Exd, the general Hox cofactor in *Drosophila*. Thus, we would predict that UbxIa would preferentially bind

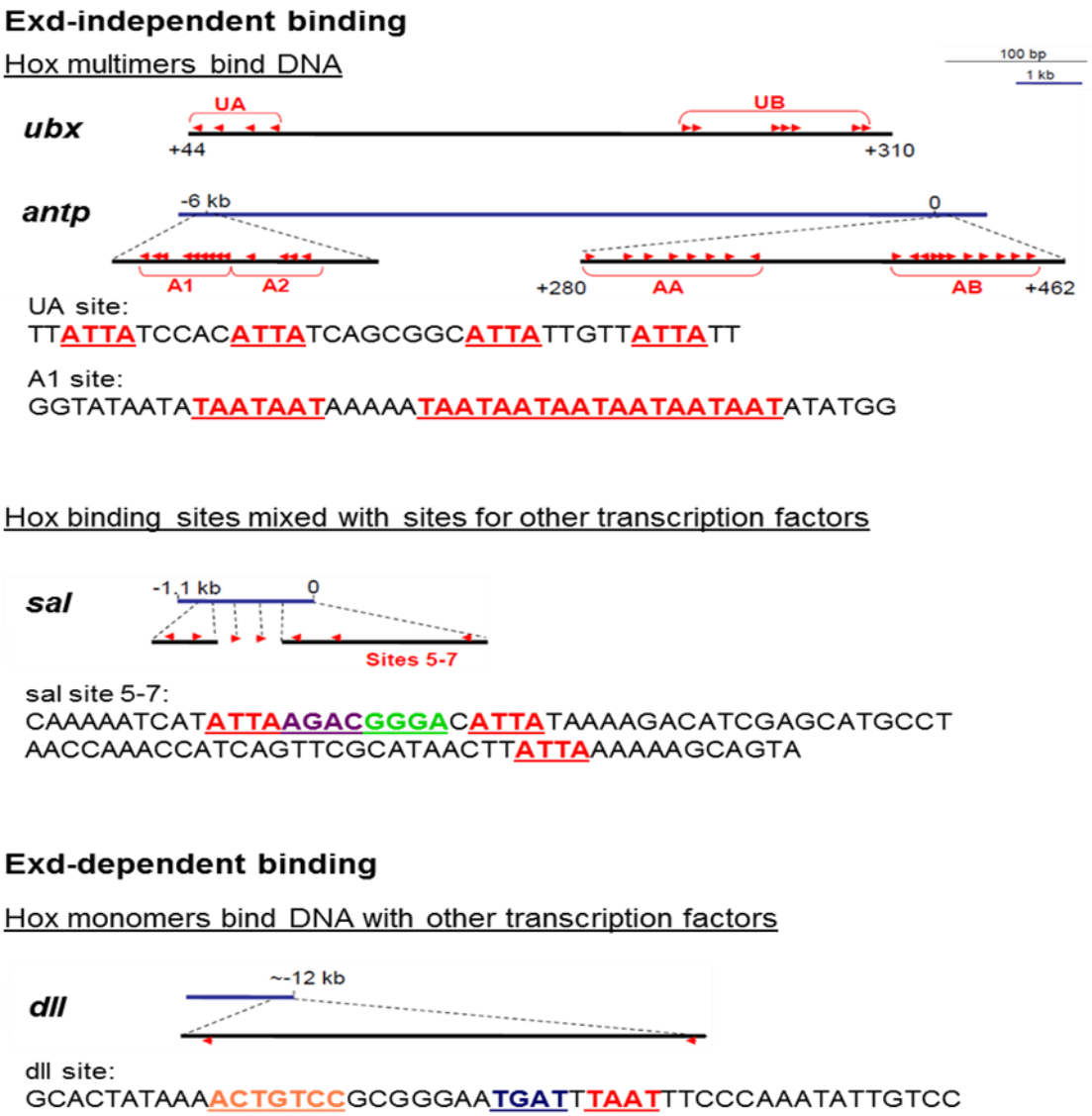

Figure 3.15 Ubx recognizes three categories of DNA binding sites. Ubx cooperatively binds multimers of Hox binding sites (TAAT/ATTA, red text), including enhancers for the *ubx* and *antp* genes [83]. Other transcription factors are not known to influence Ubx binding to these sites. In the second category, DNA binding sites for Ubx monomers are separated by DNA binding sites for other transcription factors (Medea, purple text, and Mad, green text) [60]. Regulation of the *sal* gene is coordinated by both Ubx and BMP signaling, which controls the activity of Medea and Mad. In the final category, Ubx binds DNA and regulates transcription in association with Exd (blue text) and Hth (orange text), general Hox co-factors [192,193]. The positions of the DNA sequences are marked in bp relative to the start of transcription.

Exd, and hence Hox-Exd heterodimer DNA binding sites, whereas UbxIb would preferentially interact with other transcription factors to regulate DNA sequences in which Ubx binding sites are interspersed with partner binding sites. Because these isoforms are expressed in the same tissues but not at the same levels [194], the relative concentrations of UbxIb and UbxIa may partition the available Ubx protein between genes regulated by Ubx-Exd heterodimers relative to genes regulated by Ubx in conjunction with other partner proteins. Likewise, the absence of partner proteins or the decreased affinity of partner proteins for a particular Ubx isoform, may direct Ubx to cooperatively bind DNA as homo-oligomers. Together, these mechanisms may contribute to isoform-specific differences in target gene recognition *in vivo* [2,88,195].

Finally, whereas the α-α superhelix partners bound specific disordered regions and Ubx isoforms better than others, the DNA/RNA binding 3-helical bundle fold partners tended to bind all three disordered regions equally well and bound UbxIa as well as UbxIVa. The reduced sequence specificity of DNA/RNA binding 3-helical proteins may reflect the fact that all of the disordered regions in Ubx evolved to interact with the Ubx homeodomain to regulate DNA binding [65,66]. Since the homeodomain has a DNA/RNA binding 3-helical fold, the homeodomain-interacting disordered regions can also bind other proteins with this same fold. This hypothesis predicts that protein interactions may enhance DNA binding by removing the inhibitory disordered regions from the surface of the Ubx homeodomain. Conversely, DNA binding may facilitate Ubx-partner interactions by making the disordered regions more available for partner interactions. This scenario provides a mechanism, consistent with its cellular

role, for Ubx to functionally integrate binding to a multiplicity of diverse protein partners and to DNA.

The sequence of intrinsically disordered regions evolves more rapidly than for structured regions [196,197], enabling incorporation of novel functions or binding modes. Indeed, the evolution of novel protein interaction motifs can change Hox function [179] or even dramatically transform a Hox protein into to a different class of transcription factor [198]. Based on our current knowledge, Ubx appears unlikely to interact with a subset of the proteins identified as binding partners for its natural *in vivo* function [132]. However, the ability of Ubx to bind functionally different proteins with similar structures may provide a mechanism to evolve novel Ubx functions. A new protein may be able to bind the disordered regions in Ubx based on its resemblance to an established Ubx partner, creating new modes of Ubx (or partner) regulation *in vivo*. As the Ubx sequence evolves, a specific motif for binding that partner may emerge, and with time eventually become an obligatory binding site. Examples of proteins at each of these stages may be found among the Ubx partner proteins. Most of the partners appear to recognize the disordered regions without any clear sequence or motif preferences, representing a relatively early stage in the evolution of partner binding. However, a 14-3-3ε interaction motif occurs in the mII element of Ubx [183-185]. Although the presence of this motif enhances 14-3-3ε binding, this protein still binds Ubx, albeit weakly, in the absence of this motif. Furthermore, the motif is located in a region of the Ubx protein for which inclusion depends on *ubx* mRNA splicing, allowing tissue-specific control of Ubx's affinity for 14-3-3ε.

Evolution of Hox function

In our model, the enhancement, but not obligatory reliance, of partner binding by a recognition motif represents an intermediate stage of partner evolution. Finally, Exd/Pbx is an ancient Hox protein partner required for many basic Hox functions. Although the disordered regions in Ubx influence Exd binding, Exd interactions are primarily dependent on specific motifs in the Ubx sequence [197]. Exd binds different motifs in Ubx to elicit different functional outcomes *in vivo* [199]. Thus Ubx-Exd interactions represent a highly evolved partner interaction.

CHAPTER IV

CELL BEHAVIOR STUDY IN 3D MODEL USING SELF-ASSEMBLE UBX PROTEIN MATERIALS

Introduction

Because cell culture is versatile, rapid, and cost-effective, it has been widely used for basic research and preclinical studies. Most cells are cultured systems on a 2-dimensional (2D) surface such as tissue culture flask, however, the *in vivo* environment is composed of 3-dimensional (3D) elements [200,201]. Indeed, cells often react differently in 3D and 2D culture systems [202-204]. With similar but more aggressive cellular cycle, cancer cells often show different responses between 2D and 3D culture including gene expression, proteolysis, and drug resistance [116,205,206]. For instance, expression of matrix metalloproteases (MMPs) was upregulated in 3D scaffolds compared with 2D flat surface culture with the same materials [207]. In addition, previous studies have shown that cancer cells are more resistant to anticancer drugs on a 3D substrate surface than a 2D substrate surface [208-211].

A variety of natural and chemically synthesized materials have been adopted from tissue engineering purpose to establish 3D cancer cell culture models *in vitro* [212-215]. The essential quality for materials suitable for tissue engineering purpose are also advantages in cancer cell cultures such as: robotic production, inert interactions with cells, appropriate properties and physical support to cells, and adjustable properties for various experimental designs [216]. However, few materials have been able to

accommodate these properties without significant compromise. For example, synthetic materials often lack cell binding moieties on the material surfaces and nature-derived materials often required sophisticated procedures for production and adjustable properties for various tasks [217].

In vivo, the *Drosophila melanogaster* Ultrabithorax (Ubx) protein functions as a transcription factor [57,218]. *In vitro,* monomeric Ubx protein self-assembles into nanoscale to microscale materials [124,219]. Importantly, Ubx materials are cytocompatible, biocompatible, and non-immunogenic [126,127]. Cells readily adhere to Ubx materials, which can serve as a substrate for cell cultures. In this study, we tested the compatibility between Ubx materials and three different isogenic human breast cell lines including MCF10A, MCF10AT, and MCF10CA-1a. The MCF10A cell line was derived from benign breast tissue from a woman with fibrocystic disease [220]. The premalignant MCF10 AT cell line, generated from T24 c-Ha-ras oncogene-transfected MCF10A cells, has a tendency for neoplastic progression [221]. The metastatic MCF10CA-1a cell line was derived from MCF10AT cells mouse xenografts [222]. Because MCF10A cell lines are isogenic, they provide a valuable opportunity for determining how tumorigenesis impact cell behavior in a 3D environment. We found that all MCF10A cell lines, either normal mammary gland cells (MCF10A), semi-tumorigenic cells (MCF10AT) or tumorigenic cells (MCF10CA-1a), survive on Ubx materials at various time points. Most interestingly, we also observed massive destruction of Ubx materials by MCF10CA-1a cell line but not MCF10A or MCF10AT. The possible mechanism behind the MCF10CA-1a specific Ubx materials destruction

was also investigated in this study. Differences in cytoskeleton structure between different cell types on a 2D surface and on 3D Ubx materials were also recorded. Our results indicate that Ubx materials have the potential for establishing *in vitro* 3D cell culture for cancer study.

Materials and Methods

Ubx Ia protein expression and purification

His-tagged Ubx Ia protein was expressed as previously described [124-127]. In brief, the plasmid, constructed in the pET19b expression vector and Ubx Ia cDNA, was transformed into BL21(DE3)-pLysS strain of *E.coli* (EMD Millipore). Ubx Ia protein expression was induced by adding 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, RPI) at the log phase of growth. The *E.coli* was allowed to express the protein at 25 °C for 2 hours with shaking at 250 rpm. Cells were harvested by centrifuged at 1,200 x *g* at 4 °C. Cell pellets, in aliquots corresponding to 2 L of bacterial culture, were stored at -20°C in 5 ml of PBS buffer. A frozen cell pellet was lysed in 15 mL of lysis buffer (Buffer G) containing 50 mM NaH_2PO4 (pH = 8.0), 5% glucose (w/v), and 500 mM NaCl. In addition, one ethylenediaminetetraacetic acid-free protease inhibitor tablet (Roche) and 1.2 mg/L DNase I were included to prevent proteolysis of monomeric Ubx proteins and facilitated the digestion of DNA in the cell lysates, respectively. After centrifuging for 30 minutes at 4°C, the supernatant was incubated with Ni-NTA resin (Thermal Scientific) pre-equilibrated with Buffer G for 15 minutes. The column was iteratively washed with Buffer G containing 20mM, 40mM, and 80mM imidazole

followed by the elution using Buffer G containing 200mM imidazole. The purified monomeric Ubx Ia proteins were stored 4°C until Ubx fiber production.

Production of Ubx fibers and bundles

Ubx Ia protein fibers were produced using the tray / buffer reservoir system [124,125]. To create Ubx bundles, inoculation loops were flattened with a hammer and subsequently wrapped in recently fibers immediately after drawing the fibers from a tray. Buffer G was applied to the inoculation loops to cover the with Ubx fibers with a thin film of buffer. Buffer G was removed slowly by pipetting from the area between the Ubx fibers and the inside edge of inoculation loops, thus forcing the Ubx fibers into contact, forming Ubx bundles. The Ubx bundles were dried in a biohazard hood before placement onto the chamber slide.

Cell culture

MCF10A and MCF10AT were cultured in sterile DMEM/F12 media containing 0.12 % sodium bicarbonate (w/v) (Sigma), 5 % heat inactivated fetal bovine serum (w/v) (Gibco), 10 µg/ml insulin (Sigma), 20 ng/ml epidermal growth factor (Gibco), 0.5 µg/ml hydrocortisone (Life technologies), 15 mM of 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, Caisson), and 17.5 mM glucose with pH = 7.4 and 5% CO_2 at 37 °C. MCF10CA-1a cells were cultured in DMEM/F12 media containing 0.12% sodium bicarbonate (w/v), 10% of horse bovine serum (w/v) (Gibco), 15 mM HEPES (Caisson), and 17.5 mM glucose with pH = 7.4, 5 % CO_2 at 37 °C. Cells were cultured to confluence then 20000 - 30000 cells per well were seeded in chamber slides (Ibidi) to maintain consistent confluence for each experiment. Inoculation loops,

each wrapped multiple times with a single Ubx fiber, were placed in separate wells on the chamber slide together with the cells. Sample wells for immunostaining were fixed with paraformaldehyde (4 % w/v) at designated time points.

Immunochemistry experiments

A freshly made 8% paraformaldehyde (w/v) stock in phosphate-buffered saline (PBS) was added to the existing culture media at a final concentration of 4%, and samples were fixed at room temperature for 2 hours in a chamber slide (Ibid). Fixed samples were washed three times for 10 minutes with 200 µl wash buffer 1 (25 mM Tris, 200 mM glycine). Samples were permeablized with 200 µl of Dulbecco's phosphate-buffered saline (DPBS, Life technologies), containing 0.5% Triton X-100 (Sigma), for 20 minutes at room temperature. Wells were aspirated and blocked at room temperature in 300 µl of blocking solution containing 0.1% Triton X-100, 1% BSA, and 5% goat serum for 1 hour. Primary antibodies raised against α-tubulin (Sigma) were diluted 1:50 in blocking solution and incubated in the wells for 1 hour at room temperature. Samples were washed 3 times for 10 minutes each in 200 µl of wash buffer 2 containing 0.1% Triton X-100 in DPBS and incubated with (1:200 dilution) goat anti-mouse Alexa 488 conjugated secondary antibodies (Life technologies) in blocking solution for 30 minutes at room temperature. Samples were washed 3 times in 300 µl DPBS containing 0.1% Triton X-100, 10 µM 40, 6-diamidino-2-phenylindole (DAPI, Molecular Probes) for 10 minutes and imaged immediately using confocal microscopy on a Nikon Eclipse Ti equipped with NIS Elements AR 4.10.01 software.

Live/dead cell viability assay

After incubation with Ubx fibers for the designated duration specified, the media was removed by vacuum and the wells were rinsed 6 times with 200 µl of DPBS containing 0.05 mM Mn^{2+} and 0.5 mM Mg^{2+} to remove excess serum in the media. DPBS solution containing 2 µM calcein acetoxymethylester and 4 µM ethidium homodimer-1 (Live /dead cell viability assay kit, Molecular Probes) was later added into the wells for cell staining. The live/dead cell viability assay reagents were incubated with cells for 10 minutes at room temperature and imaged using confocal and fluorescent microscopy on a Nikon Eclipse Ti equipped with NIS Elements AR 4.10.01 software. Images were taken in differential interference contrast (DIC) format or fluorescent format.

Proteolysis inhibition assay

MCF10CA-1a cells were seeded in a chamber slides together with 20000 cells/well in a chamber slide (Ibidi) with in medium in a final volume of 200 µl/well. A broad spectrum matrix metalloproteinase **(**MMP**)** inhibitor, GM6001, (Santa Cruz Biotechnology) was added into the medium to under a serial dilution 12 µM to 120 nM at the final concentration according to manufacturer recommendation to determine the optimal concentration. The final concentration of 12 µM was used in this study according to the results from serial dilutions. The total number of Ubx fibers was recorded immediately after the initial seeding the MCF10CA-1a cells. An individual Ubx fiber with the connection between two attaching points on individual inoculation loop is considered an intact fiber and vice versa. The percentage of intact Ubx fibers at a

designated time point versus total the number of intact Ubx fibers at the initial time point was calculated.

Results

Ubx fibers are compatible with MCF10A cells

In order to establish a sustainable 3D culture, the materials used for composing the scaffolds should not have any toxicity to cells. Previous study in our laboratory has shown that plain Ubx and mCherry-Ubx protein fusion materials are compatible with three different primary human cell lines including: umbilical vein ECs, brain vascular PCs, and aortic SMCs [126]. However it is not clear if Ubx materials will be toxic to mammary gland cells since there is an anti-tumorigenic sequence in the conserved homeodomain of Ubx [223-225]. In order to monitor the progression of the viability of the cells, we treated the cells with live/dead cell viability assay after cell incubated with Ubx fibers at various time points. No evidence of cell death was observed when MCF10A cells when incubated with Ubx fibers at various time points (**Figure 4.1**), either in close proximity to Ubx fibers (<10 μm), or far from Ubx fibers (>100 μm) at 24 hours and 48 hours (**Figure 4.1B-C, 4.1E-F**, **4.1H-I, 4.1K-L**). Both experimental groups showed over 97% survival rate. Therefore, Ubx materials are not toxic to these cells and they do not secret toxic substances into the media. Increased cell density of MCF10A cells on Ubx fibers at 12 hours after the initial seeding compared to 6 hours as observed (**Figure 4.1B-C**, **4.1E-F**). MCF10A cells attached to Ubx fiber with orientation

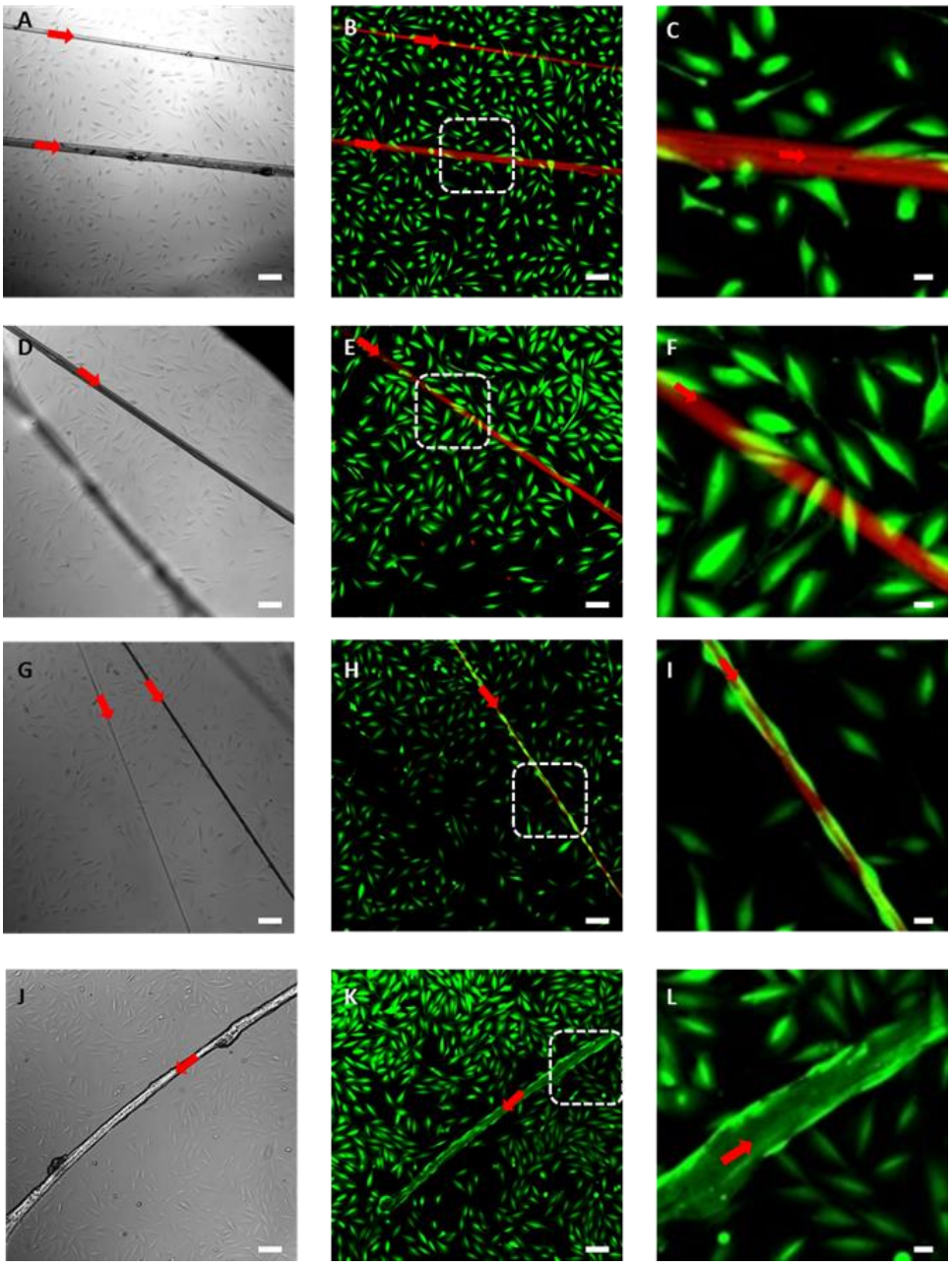

Figure 4.1 Ubx fibers are not toxic to MCF10A cells. DIC images were taken before live/dead staining at 6 (A), 12 (D), 24 (G), and 48 (J) hours, respectively, to show the relative position of Ubx fibers and cell population at different time points. Live/dead cell viability assays were performed to show the viability of MCF10A cells at 6 (B-C),12(E-F), 24(H-I), and 48 hours (K-L), respectively. Red arrows indicate Ubx fibers and live MCF10A cells are shown in green. The white dashed rectangular indicates area enlarged 2X for images with higher magnification. Scale bars equal 10 μm in panel C, F, I, and L. Scale bars equal 100 μm in all other panels.

aligned with the axis of Ubx fibers at 24 hours after the initial seeding (**Figure 4.1I-L**).

High density of MCF10A cells settled on Ubx fibers. Most notably, most MCF10A cells attached to both the top and sides of Ubx fibers. If MCF10A cells were forced to attach to Ubx fibers, attachment of the cell will only be seen at certain area and not in all dimensions of Ubx fibers. Our results indicate the self-motivated attachment to Ubx fibers from MCF10A cells instead of forced attachment during cell seeding.

Ubx materials are compatible with MCF10AT cells

Among MCF10A isogenic cell lines, MCF10AT cell was considered semi-tumorigenic due to the fact that 25% of MCF10AT cells will become malignant cells after prolonged culture and selection [221]. Due to the undetermined cell fate of MCF10AT cells, it will be useful to see the cellular responses of MCF10AT to Ubx materials for a comprehensive view of the progression in tumorigenesis. MCF10AT cells were seeded in a chamber slide containing designated medium and Ubx fibers. Cell viability was determined by staining the cells with live/dead cell viability assay and visualized using confocal and fluorescent microscopy at 6 hours, 12 hours, 24 hours, and 48 hours. Similar cellular responses as MCF10A cells were also observed from MCF10AT cells on Ubx fibers (**Figure 4.2**). MCF10AT cells did not show signs of cell death when incubated with Ubx fibers neither at shorter time points such as 6 hours (**Figure 4.2B, 4.2C**) and 12 hours (**Figure 4.2E, 4.2F**), nor at longer time points such as 24 hours (**Figure 4.2H, 4.2I**) and 48 hours (**Figure 4.2K, 4.2L**) . The average survival rate was over 98% on the Ubx fibers and cells in the same well on the bottom of chamber slide

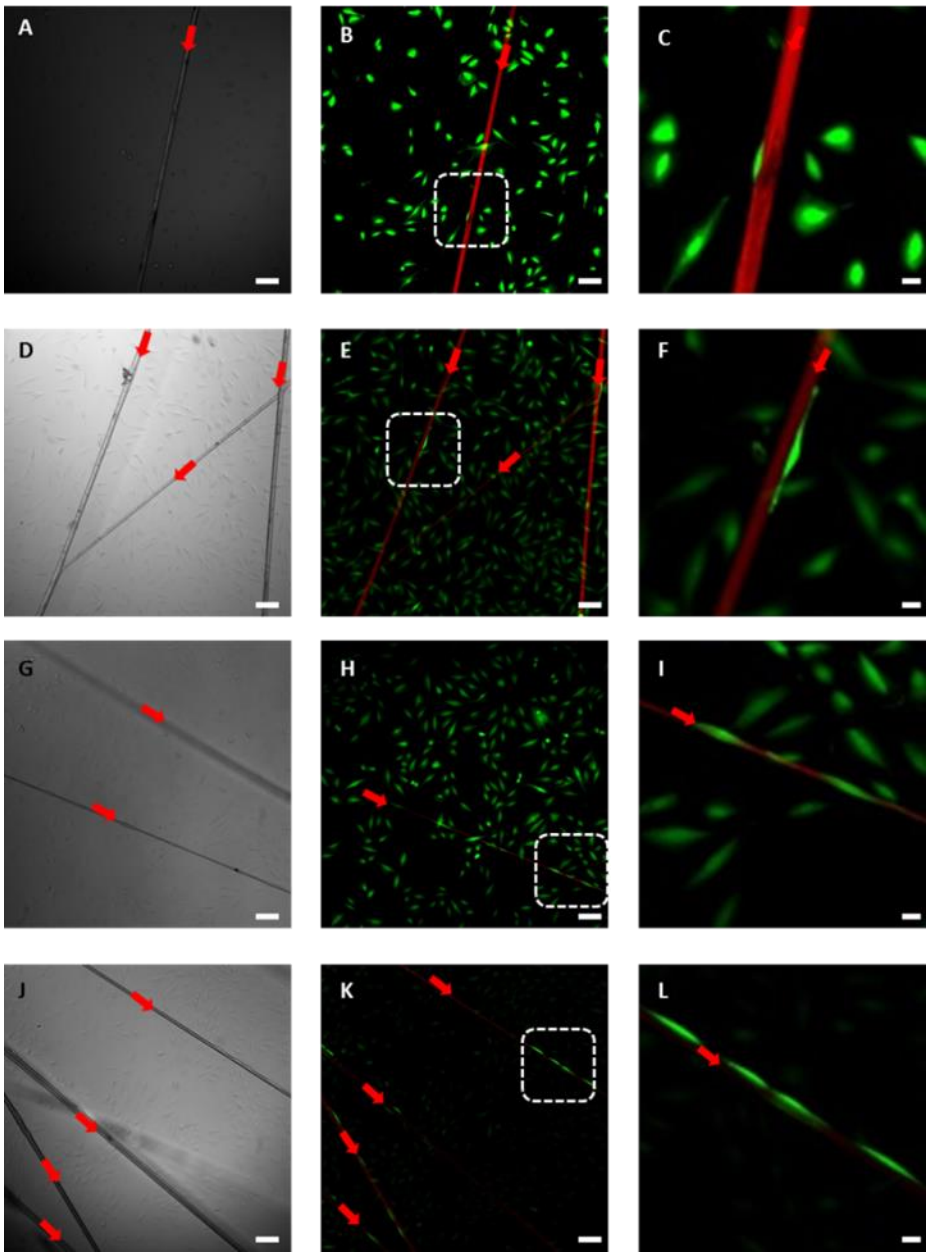

Figure 4.2 Ubx fibers are not toxic to MCF10AT cells for prolonged incubation. DIC images were taken before live/dead staining at 6 (A), 12 (D), 24 (G), and 48 (J) hours, respectively to show the relative position of Ubx fibers and cell population at different time points. Live/dead cell viability assays were performed to show the viability of cells at 6 (B-C), 12(E-F), 24(H-I), and 48 hours (K-L), respectively. Red arrows indicate Ubx fibers and live MCF10AT cells are shown in green. The white dashed rectangular indicates area enlarged 2X for images with higher magnification. Scale bars equal 10 μm in panel C, F, I, and L. Scale bars equal 100 μm in all other panels.

across all time points. There were more MCF10AT cells attached on Ubx fibers at longer time points then shorter time points (**Figure 4.2**). MCF10AT cells were also located on both the top and sides of Ubx fibers. Although, MCF10AT cells were aligned to the axis of Ubx fibers but MCF10AT cells showed more spindle type of morphology, which was different from MCF10A cells on Ubx fibers.

Ubx materials are compatible with MCF10CA-1a cells

Previous experiments have shown the cytocompatibility of Ubx fibers to both MCF10A and MCF10AT, which represent normal and semi-tumorigenic mammary gland cell, respectively, in the whole spectrum of MCF10A isogenic cell lines. Next, we tested the compatibility between MCF10CA-1 and Ubx fibers. There were no conclusive signs of cell death on Ubx fibers for MCF10CA-1a cells with direct contact or indirect contact with Ubx fibers at various time points (**Figure 4.3**). MCF10CA-1a cells were alive on Ubx fibers at 48 hours after the initial seeding (**Figure 4.3J-L**). Physical attachment of MCF10CA-1a cells was observed at 6 hours (**Figure 4.3A-C**). MCF10CA-1a cells covered most of the surface on the Ubx fibers at 24 hours (**Figure 4.3G-H, 4.3L**). Compared to MCF10A and MCF10AT cells, MCF10CA-1a cells exhibit two unusual behaviors: MCF10CA-1a broke Ubx fibers and they have an unusual morphology at later time points. Ubx fibers started to rupture at 12 hours (**Figure 4.3D-F**) and all Ubx fibers had broken into fragments with size of 10-20 μm in length by 48 hours (**Figure 4.3J-L**) when incubated with MCF10CA-1a. During this study, we noticed some correlations between the diameter of Ubx fibers and the time required for Ubx fibers to rupture. The working hypothesis was that Ubx fibers with larger diameter

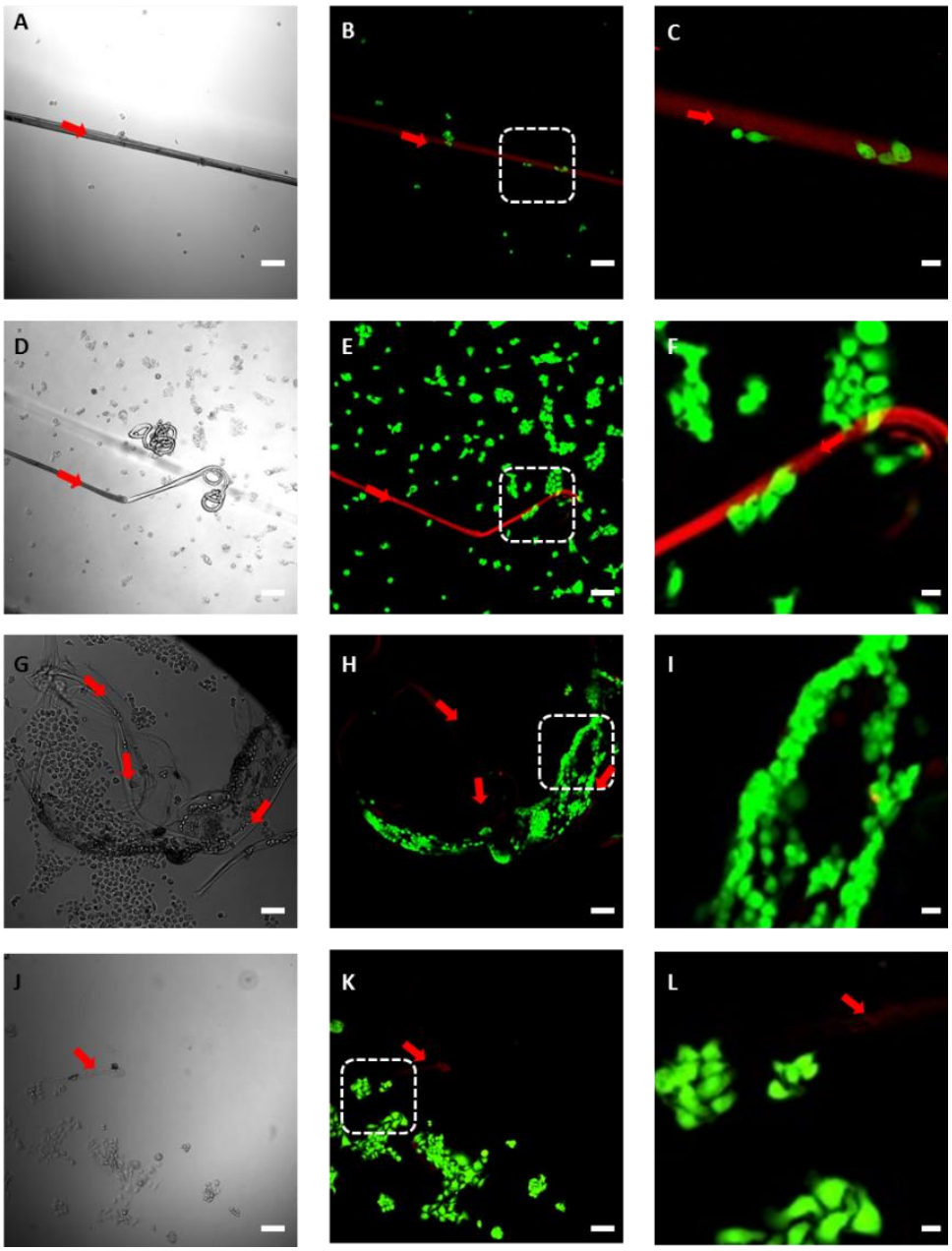

Figure 4.3 Ubx fibers are not toxic to MCF10CA-1a cells for prolonged incubation. DIC images were taken before live/dead staining at 6 (A), 12 (D), 24 (G), and 48 (J) hours, respectively to show the relative position and cell population at different time points. Live/dead viability assays were performed to show the viability of cells at 6 (B-C), 12(E-F), 24(H-I), and 48 hours (K-L), respectively. Red arrows indicate Ubx fibers and live MCF10CA-1a cells are shown in green. The white dashed rectangular indicates area enlarged 2X for images with higher magnification. Scale bars equal 10 μm in panel C, F, I, and L. Scale bars equal 100 μm in all other panels.

required more time to rupture compared with Ubx fibers with smaller diameter. We decided to fuse individual Ubx fibers into Ubx bundles [124]. The bundles, with their wider diameters, should enhance the durability of Ubx materials in the presence of MCF10CA-1a cells (**Figure 4.4**). As expected, the Ubx bundles were mostly intact until later time points (**Figure 4.5**), which provided the opportunity to qualify the response from MCF10CA-1a on Ubx bundles. The MCF10CA-1a cells started attaching to Ubx bundles at 6 hours and into later time points. There were no sign of cell death for cells in the contact with Ubx bundles at all of time points neither those cells surrounding Ubx bundles without direct contact.

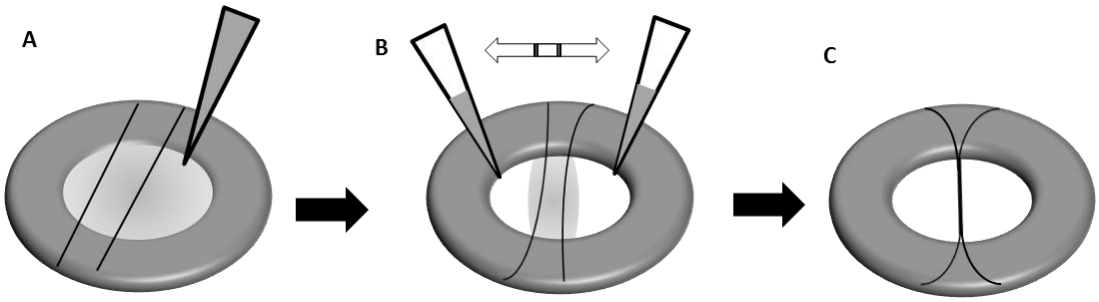

Figure 4.4 Ubx bundle production process. (A) Ubx fibers (black lines) were attached to the plastic loop (grey oval shape object). Buffer G (area with grey gradient) was applied on the plastic loop with Ubx fibers using pipette (grey triangle). (B) Buffer G was removed by pipetting on both sides of Ubx fibers to create surface tension between the two Ubx fibers. (C) Several Ubx fibers were bound as a bundle after the removal of Buffer G.

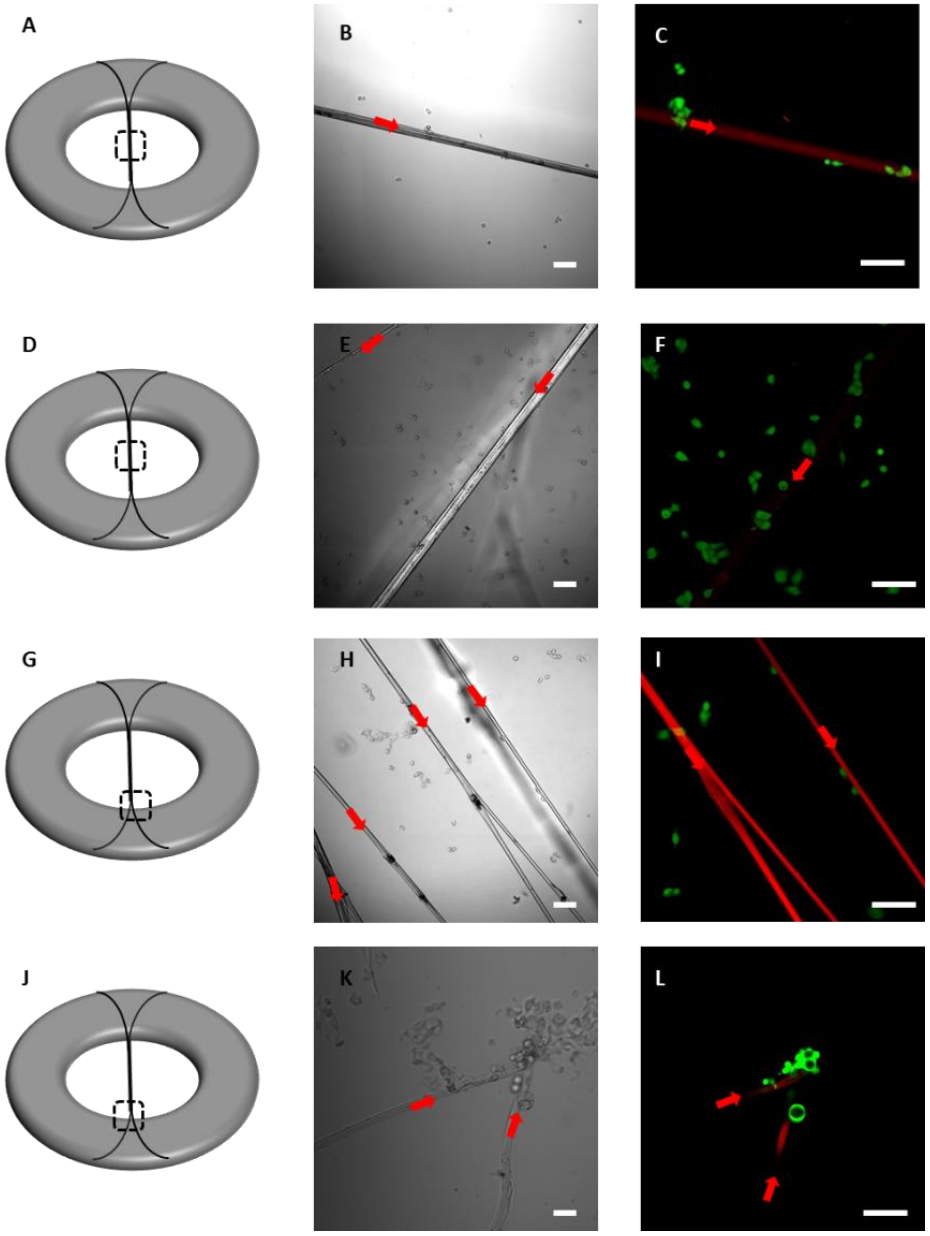

Figure 4.5 Ubx bundles can accommodate MCF10CA-1a cells for prolonged incubation. (A, D, G, J) The relative locations on Ubx bundles for confocal images are indicated by black dashed rectangular. DIC images of MCF10CA-1a cells with live/dead staining at 6 (B), 12 (E), 24 (H), and 48 (K) hours after the initial seeding, respectively, show the relative position and cell population at different time points. Live/dead viability assays were performed to show the viability of cells at 6 (C), 12(F), 24(I), and 48 hours (L), respectively. Red arrows represent Ubx bundles. Scale bars equal 100 μm in all panels.

Matrix metalloproteinases (MMPs) inhibitor prevents MCF10CA-1a specific Ubx materials destruction

The destruction of Ubx fibers and Ubx bundles by MCF10CA-1a cells triggered us to probe the possible mechanisms behind the phenomena. We compared all the images taken at different time points from three different cell lines tested in this study. We have found that only MCF10CA-1a cells are capable of breaking Ubx fibers, regardless the size of the fiber (**Figure 4.6**). At 6 hours, individual MCF10CA-1a cells were seen at areas where Ubx fibers were partially destroyed (**Figure 4.6A-B**).

Live/dead cell viability assays showed that MCF10CA-1a cells extend into the cracked regions on Ubx fibers. At 12 hours, Ubx fibers were more damaged in areas where MCF10CA-1a cells were situated. In some cases, sections of the Ubx fibers appeared to be cut away (**Figure 4.6C-D**).

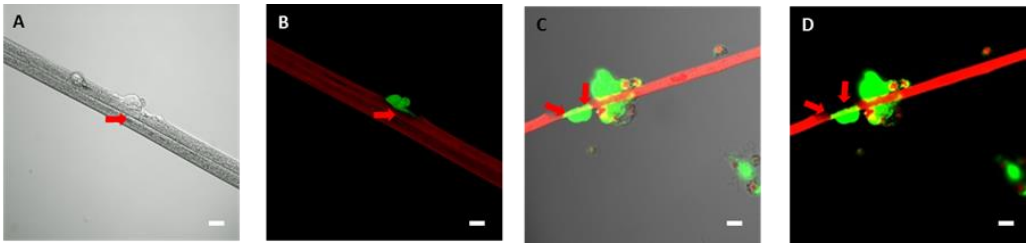

Figure 4.6 Ubx fibers were destructed by MCF10CA-1a cells. (A,C) DIC and fluorescence merged images show the topography of Ubx fibers and relative location of MCF10CA-1a cells on Ubx fibers. (A-B) MCF10CA-1a cells (green) situated at the Ubx fibers (red arrows) at 6 hours after the initial seeing. Ubx fibers were partially destructed (red arrows). (C-D) MCF10CA-1a cells (green) situated at the Ubx fibers (red linear structure) at 12 hours after the initial seeing. Ubx fibers were increasingly destructed (red arrows). Scale bars equal 10 μm in all panels.

This damage could potentially be caused by mechanical forces exerted by the cells on Ubx materials and/or by proteases secreted by MCF10CA-1a cells. MMPs are upregulated in MCF10CA series cells compared with MCT10A [226,227], indicating a possible link between the destruction of Ubx fibers and MMPs. Indeed, structural integrity of Ubx fibers was preserved at various time points when GM6001, a broad spectrum MMP inhibitor, was present in the media at concentrations of 12 μM (**Figure 4.7**).

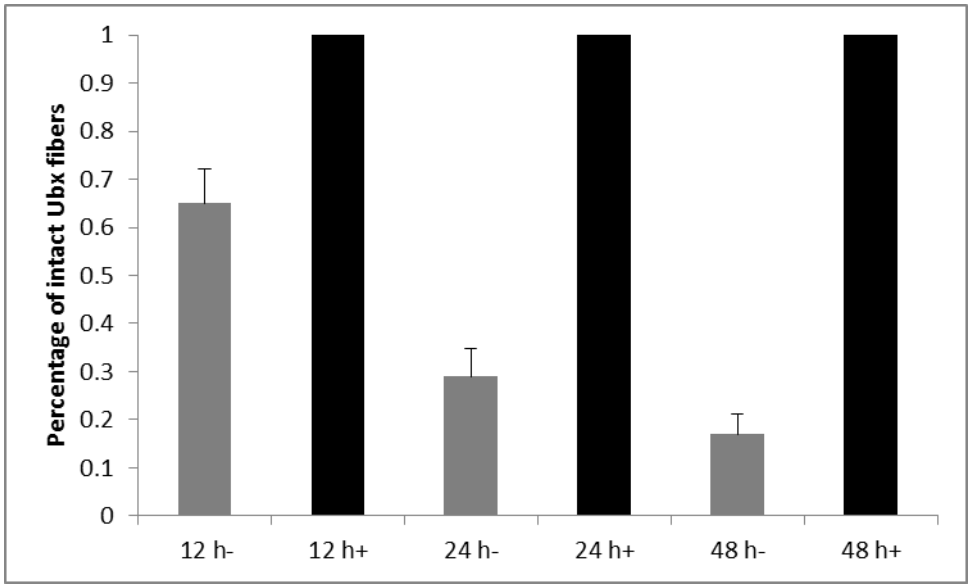

Figure 4.7 MMPs inhibitor prevents MCF10CA-1a related Ubx materials destruction. Time course experiments with broad spectrum MMP inhibitor show inhibition of MCF10CA-1a related Ubx fibers destruction. "-" represent data collected at different time points (X axis) without the addition of MMPs inhibitor. "+" represent data collected at different time points (X axis) under the presence of MMPs inhibitor.

It is possible that viability of MCF10CA-1a cells were compromised by GM6001. Live/dead cell viability assays were performed and have shown alive MCF10CA-1a cells on Ubx fibers under the presence of GM6001 at various time points (**Figure 4.8**). Therefore, the GM60001 did not impede the viability of MCF10CA-1a cells in this experiment.

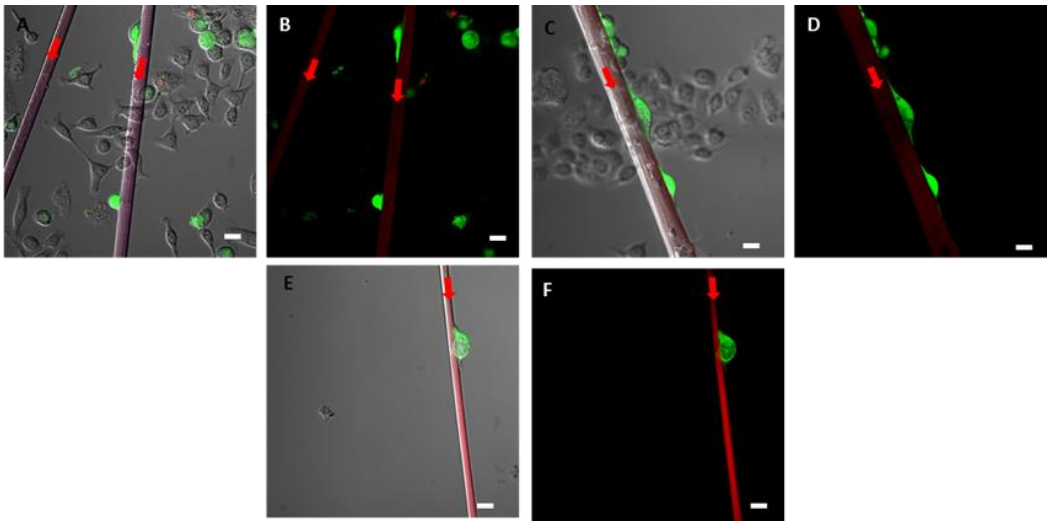

Figure 4.8 GM6001**,** a broad spectrum MMPs inhibitor, preserves structural integrity of Ubx fibers and do not kill MCF10CA-1a cells. (A, C, E) DIC and fluoresce merged images of MC10CA-1a cells (green) and Ubx fibers (red arrows). (B, D, F) Fluorescent images show alive MCF10CA-1a cells (green) on intact Ubx fibers (red arrows). (A, B) Live MCF10CA-1a cells on intact Ubx fibers at 12 hours after the initial seeding in the presence of GM6001. (C, D) Live MCF10CA-1a cells on intact Ubx fibers at 24 hours after the initial seeding in the presence of GM6001. (E, F) Live MCF10CA-1a cells on intact Ubx fibers at 48 hours after the initial seeding in the presence of GM6001.

Differential cytoskeletal topography between different cell lines and surface dimensions

The integrity of the cytoskeleton is a good indicator of healthy cells. Due to the limitation of live/dead cell viability assay for the visualization of cytoskeleton, addition assay may be required for revealing the cellular structures. Therefore, we applied immunochemistry techniques to reveal the morphology of cytoskeleton of different cell lines on Ubx materials, to have a comprehensive visualization of cellular using the α-tublin. MCF10A cells showed differential cytoskeletal topography on different surfaces at 48 hours after initial seeding. MCF10A cells cultured in tissue culture vessels showed expended geometrical arrangement of α-tublin (**Figure 4.9A**). However, MCF10A cells cultured on Ubx fibers have showed more confined α-tublin structure. In addition, there was visible alignment between the α-tublin and the axis of Ubx fibers (**Figure 4.9B**). MCF10AT cells, attached to different surfaces, also showed differences in cytoskeletal architecture. For MCF10AT cells attached to Ubx fibers, the α-tubulin aligned with the axis of Ubx fibers. Interestingly, some of the MCF10AT cells showed twisted α-tublin along the axis of the Ubx fibers, which could be contributed by microenvironment on the surface of Ubx fibers. However, MCF10AT cells on the chamber slide exhibited more disperse α-tubulin structures compared to MCF10AT cells on Ubx fibers (**Figure 4.9C-D**).

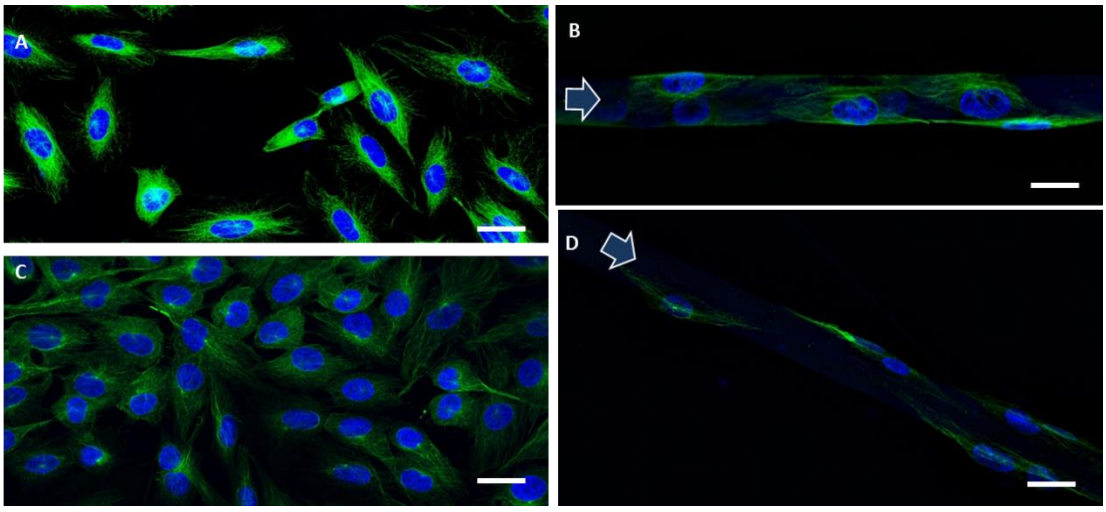

Figure 4.9 Discrepancy in cytoskeletal organization in MCF10A and MCF10AT cells cultured on different substrates. (A) MCF10A cells cultured on tissue culture vessels stained for α-tublin (green) and nuclei (blue). (B) MCF10A cells cultured on Ubx fibers (blue arrows) stained for α-tublin (green) and nuclei (blue). (C) MCF10AT cells cultured on tissue culture vessels stained for α-tublin (green) and nuclei (blue). (D) MCF10AT cells cultured on the Ubx fibers (blue arrows) stained for α-tublin (green) and nuclei (blue).

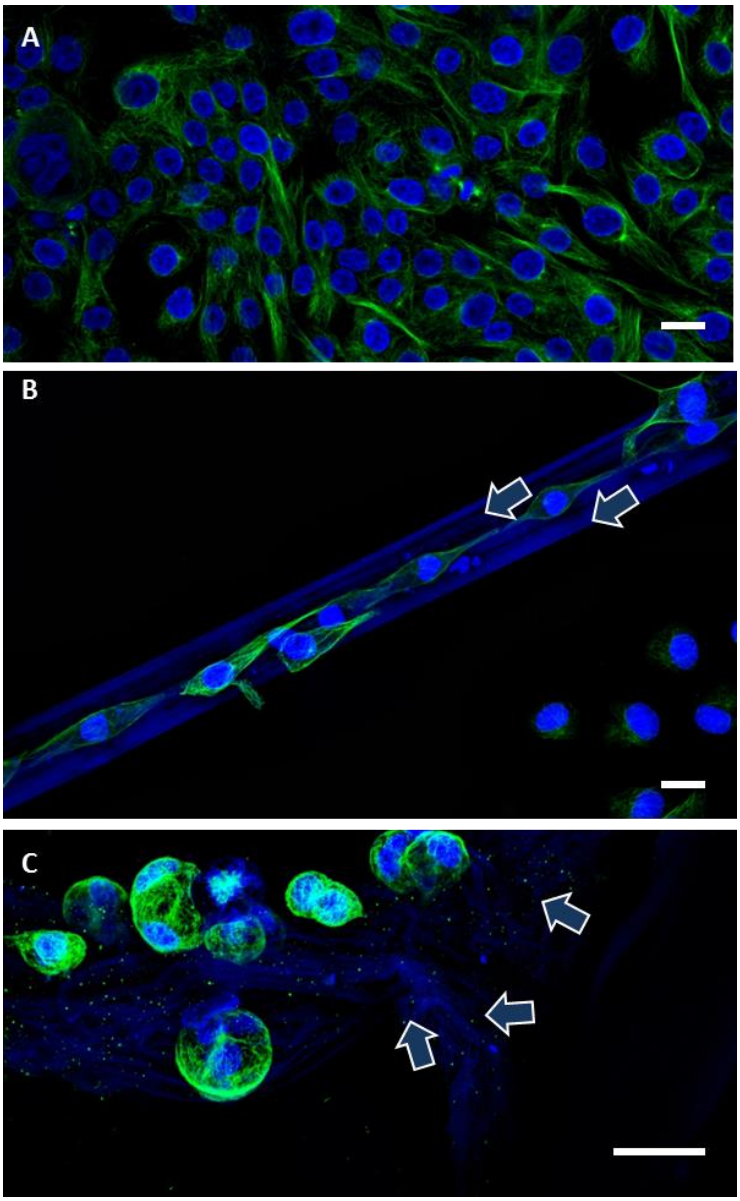

Figure 4.10 Discrepancy in cytoskeletal organization in MCF10CA-1a cells on different substrates. (A) MCF10CA-1a cells cultured on tissue culture vessels stained for α-tublin (green) and nuclei (blue). (B) MCF10CA-1a cells cultured on the Ubx fibers (blue arrows) stained for α-tublin (green) and nuclei (blue) at 24 hours after the initial seeding. (C) MCF10CA-1a cells cultured on the Ubx fibers stained for α-tublin (green) and nuclei (blue) at 48 hours after the initial seeding. Scale bars equal 10 μm in all panels.

Disrupted cytoskeleton in MCF10CA-1a cells after prolonged incubation with Ubx fibers

In contrast to MCF10A and MCF10AT cells, MCF10CA-1a showed more dramatic changes of cytoskeleton on different substrate surfaces (**Figure 4.10**). MCF10CA-1a cells cultured on chamber slides showed clearly defined more α-tublin topography at 48 hours after the initial seeding (**Figure 4.10A**). On the other hand, MCF10CA-1a cultured on Ubx fibers showed distinguishable changes of α-tublin topography corresponding to the structure of Ubx fibers (**Figure 4.10B-C**). The MCF10CA-1a cells on Ubx fibers had divided nuclei and were less compact than MCF10CA-1a cells cultured on tissue culture vessels 48 hours.

Discussion

In this study, we used the combination of confocal imaging and various assays to validate the possibility of using Ubx protein based materials for establishing 3D culture for characterizing carcinogenesis. Our results indicated that Ubx materials were cytocompatible for various isogenic breast mammary gland cell lines regardless of the degree of tumorigenesis. Therefore the anti-tumorigenic sequence in the Ubx protein is not active in Ubx materials. Similarly, anti-bacterial peptides linked to silk were active in silk monomer but not silk materials [228]. Furthermore, MCF10CA-1a cells attached on Ubx materials have shown signs of structural destruction that was not observed for MCF10A or MCF10AT cells. We also observed Ubx materials were destroyed by

MCF10CA-1a cells, caused, in part, by MMPs. In addition, there was discrepancy in cytoskeletal arrangement for all cell lines on tissue culture flask and Ubx materials.

The properties of surface substrate often influence cellular response [200,202,203]. For instance, the surface topography on the substrate can be influential to cellular responses [200,202,203,229]. Cell adhesion in 3D environments is often different in focal and fibrillar adhesions characterized on 2D substrates which could reflect in the composition cytoskeletal components and morphologies [201,230]. As an example, epithelial cells have a larger spreading area on a 2D than on a 3D substrate [231]. Besides the cell morphology, cells cultured in 3D substrate show enhanced cell biological activities than 2D substrate [200,201,232]. For example, human foreskin fibroblasts have faster attachment, migration rate, and proliferation on a 3D substrate than a 2D substrate [200]. Up-regulation of transforming growth factor β-1 (TGFβ-1) along with other mesenchymal regulatory genes indicating an epithelial and mesenchymal transition represented are observed in 3D polymer scaffolds [232]. The discrepancy of cytoskeletal structure among different MCF10A cell lines on the bottom of chamber slide and 3D Ubx materials indicate that cells from individual MCF10A cell lines may have adopted the topographical representation of cell-substrate interface. A variety of cells have shown this type of behavior on surface substrate. For example various breast cancer cells can align with the axis of the material scaffolds. H605, MCF 7, and MDA-MB-231 cells show aligned actins and nucleus with polymer scaffolds [232].

The MCF10CA-1a specific Ubx materials destruction in this study could be caused by proteolysis. Metastatic cancer cells often express high level of proteases during invasion [233-235]. Previous studies have shown MC10CA cell lines have up-regulated proteases such as MMPs at transcriptional and translational level. MCF10CA-cl and MCF10CA-1a cell lines have up-regulated MMP-2 and MMP-9 expression which contribute to the metastatic state of the cells [191,192]. The preservation of Ubx fiber integrity under the presence of broad spectrum MMPs inhibitor indicates the essential role of MMPs for MCF10CA-1a cells to secure the topographical location on Ubx fiber, which is a common phenomenon for cancer invasion [233,236].

In addition to proteolysis, the mechanical force from MCF10CA-1a cells can also contribute to the structural destruction of Ubx fibers. Cells apply physical stress on substrate, an important factor which influences cell development. Cancer cells are especially sensitive to mechanical forces [237,238]. Mechanical force and proteolysis are both essential components of metastasis. However, some studies have shown that both mechanical forces and proteolysis are required for metastasis *in vitro* but some have ruled out the significance of proteolysis [239-242]. Our proteolytic study indirectly indicates the importance of proteolysis in Ubx fibers destruction but we could not rule out the role of mechanical force in this event. In addition, it is not clear whether other cellular mechanisms involved in cell attachment were affected by the presence of the broad spectrum MMPs inhibitor. Since the mechanical force is generated by the cytoskeleton elements, the disruption of cytoskeleton proteins under the presence of MMPs inhibitor may have direct impacts to the application of mechanical force from

MCF10CA-1a cells to Ubx fiber. A comprehensive proteomic profiling for cytoskelctome may be helpful to reveal the role of cellular signaling associated with focal adhesion [243,244].

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

In this study, we have made several discoveries. First of all, we have discovered media sources and preparation methods can cause inconsistency in yeast- based assays. By testing different media base from different companies, we have found that media component from different sources has strong influence on the results generated by yeast one- hybrid and yeast-two hybrid assay. Secondly, we have proved the similarity of topological features for Ubx binding proteins. By using bioinformatics and yeast two-hybrid assay, we discovered that most of Ubx binding proteins can be classified into several different shapes and 7 of them are enriched with Ubx partner proteins. Particularly, α-α superhelix and DNA-RNA three helical bundle have most of Ubx binding proteins under these two folds. We also showed that each disordered region in the Ubx protein has different weights on the Ubx-partner protein interactions and they are coordinated during interactions. For instance, the N-terminal regions and mRNA alternatively spliced microexon region in Ubx peptide are critical for Ubx to interact with other proteins. The data from yeast two-hybrid assays using truncated Ubx peptides missing the N-terminal regions showed weaker interactions with other binding proteins. Besides, Ubx isoforms composed by different combinations of individual microexon regions show different degree of interaction strength with binding proteins. Third of all, we have explored the possibility of using Ubx protein based materials for establishing 3D cancer model. The experimental results showed that Ubx materials are

not toxic to three types of mammary gland cells tested in this study including MCF10A, MCF10AT, and MCF10CA-1a which has different degrees of tumorigenesis. We also observed MCF10CA-1a specific destruction on Ubx materials which can be stopped by the presence of broad spectrum MMPs proteases inhibitor.

For future directions, they can be categorized into three different subgroups due to the fundamental differences of my dissertation projects.

For the media component, we have learned that the nitrogen-based media from different sources cause the discrepancy. However, it is important to dissect which particular ingredient (s) has/have the most impact on the yeast assays. It is possible to apply the mass spectrometry to analyze the chemical components in the media formulation from different sources and compare the data with bench experiments.

For the phenomenon of topological selections observed in Ubx protein-partner interactions, it will be the ultimate goal to generate structural data on individual Ubx-partner protein binding complexes using structural based techniques, such as nuclear magnetic resonance (NMR) or X-ray crystallography. Since intrinsically disordered proteins often regulate important biological functions and participate in protein-protein interactions, the findings will serve as a great model for understanding other possible biophysical features for other intrinsically disordered protein interactions.

For the further application aspect of Ubx biomaterial project, a comprehensive profiling of proteases and cytoskeleton proteins at genomic and proteomic level for cells incubated with Ubx protein materials will provide a comprehensive view of the cellular responses at molecular level. In addition, this information can be used to pinpoint the

most possible proteases contributing to the MCF10CA-1a specific Ubx material destruction which will provide valuable information for facilitating the development of therapeutic strategy for treating cancer patients.

REFERENCES

1. Molkentin JD (2000) The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. Journal of Biological Chemistry 275: 38949-38952.

2. Reed H, Hoare T, Thomsen S, Weaver T, White RAH, et al. (2010) Alternative splicing modulates Ubx protein function in *Drosophila melanogaster*. Genetics 184: 745-758.

3. Iannello RC, Kola I, Dahl HH (1993) Temporal and tissue-specific interactions involving novel transcription factors and the proximal promoter of the mouse Pdha-2 gene. Journal of Biological Chemistry 268: 22581-22590.

4. Witek M, Snook A, Lin J, Blomain E, Xiang B, et al. (2014) A novel CDX2 isoform regulates alternative splicing. PLoS ONE 9: e104293.

5. Keskin O, Gursoy A, Ma B, Nussinov R (2008) Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chemical Reviews 108: 1225-1244.

6. Meszaros B, Simon I, Dosztanyi Z (2009) Prediction of protein binding regions in disordered proteins. PLoS Computational Biology 5: 18.

7. Janin J, Chothia C (1990) The structure of protein-protein recognition sites. Journal of Biological Chemistry 265: 16027-16030.

8. Liu J, Perumal N, Oldfield C, Su E, Uversky V, et al. (2006) Intrinsic disorder in transcription factors. Biochemistry 45: 6873-6888.

9. Dunker AK, Uversky V (2008) Signal transduction via unstructured protein conduits. Nature Chemical Biology 4: 229-230.

10. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. Journal of Molecular Biology 293: 321-331.

11. Dunker AK, Cortese M, Romero P, Iakoucheva L, Uversky V (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS Journal 272: 5129-5148.

12. Spyrakis F, BidonChanal A, Barril X, Luque FJ (2011) Protein flexibility and ligand recognition: Challenges for molecular modeling. Current Topics in Medicinal Chemistry 11: 192-210.

13. Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America 93: 13-20.

14. Jimenez R, Salazar G, Baldridge KK, Romesberg FE (2003) Flexibility and molecular recognition in the immune system. Proceedings of the National Academy of Sciences of the United States of America 100: 92-97.

15. Corrada D, Morra G, Colombo G (2013) Investigating allostery in molecular recognition: Insights from a computational study of multiple antibody-antigen complexes. Journal of Physical Chemistry B 117: 535-552.

16. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, et al. (2006) Intrinsic disorder in transcription factors. Biochemistry 45: 6873-6888.

17. Uversky V, DavÃƒ V, Iakoucheva L, Malaney P, Metallo S, et al. (2014) Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. Chemical Reviews 114: 6844-6879.

18. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: Introducing the D(2) concept. Annual Review of Biophysics. Palo Alto: Annual Reviews. pp. 215-246.

19. Mitrea D, Yoon M-K, Ou L, Kriwacki R (2012) Disorder-function relationships for the cell cycle regulatory proteins p21 and p27. Biological Chemistry 393: 259-274.

20. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, et al. (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics 9: S1.

21. Bustos D (2012) The role of protein disorder in the 14-3-3 interaction network. Molecular Biosystems 8: 178-184.

22. Haynes C, Oldfield C, Ji F, Klitgord N, Cusick M, et al. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Computational Biology 2: e100.

23. Buljan M, Chalancon G, Dunker AK, Bateman A, Balaji S, et al. (2013) Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. Current Opinion in Structural Biology 23: 443-450.

24. Mohan A, Oldfield C, Radivojac P, Vacic V, Cortese M, et al. (2006) Analysis of molecular recognition features (MoRFs). Journal of Molecular Biology 362: 1043-1059.

25. Hollstein M, Sidransky D, Vogelstein B, Harris CC (1991) p53 mutations in human cancers. Science 253: 49-53.

26. Levine AJ (1997) p53, the cellular gatekeeper for growth and division. Cell 88: 323-331.

27. Sherr CJ, McCormick F (2002) The RB and p53 pathways in cancer. Cancer Cell 2: 103-112.

28. Miller LD, Smeds J, George J, Vega VB, Vergara L, et al. (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proceedings of the National Academy of Sciences of the United States of America 102: 13550-13555.

29. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, et al. (2002) Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene 21: 7435-7451.

30. Ko Y, Abel J, Harth V, Brode P, Antony C, et al. (2001) Association of CYP1B1 codon 432 mutant allele in head and neck squamous cell cancer is reflected by somatic mutations of p53 in tumor tissue. Cancer Research 61: 4398-4404.

31. Tokumaru Y, Yamashita K, Osada M, Nomoto S, Sun DI, et al. (2004) Inverse correlation between cyclin A1 hypermethylation and p53 mutation in head and

neck cancer identified by reversal of epigenetic silencing. Cancer Research 64: 5982-5987.

32. Ahrendt S, Hu Y, Buta M, McDermott M, Benoit N, et al. (2003) p53 mutations and survival in stage I non-small-cell lung cancer: results of a prospective study. Journal of the National Cancer Institute 95: 961-970.

33. Pollack IF, Finkelstein SD, Woods J, Burnham J, Holmes EJ, et al. (2002) Expression of p53 and prognosis in children with malignant gliomas. New England Journal of Medicine 346: 420-427.

34. Lowe ED, Tews I, Cheng KY, Brown NR, Gul S, et al. (2002) Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. Biochemistry 41: 15625-15634.

35. Avalos JL, Celic I, Muhammad S, Cosgrove MS, Boeke JD, et al. (2002) Structure of a Sir2 enzyme bound to an acetylated p53 peptide. Molecular Cell 10: 523-535.

36. Mujtaba S, He Y, Zeng L, Yan S, Plotnikova O, et al. (2004) Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. Molecular Cell 13: 251-263.

37. Fu HA, Subramanian RR, Masters SC (2000) 14-3-3 proteins: Structure, function, and regulation. Annual Review of Pharmacology and Toxicology 40: 617-647.

38. Meek SEM, Lane W, Piwnica Worms H (2004) Comprehensive proteomic analysis of interphase and mitotic 14-3-3-binding proteins. Journal of Biological Chemistry 279: 32046-32054.

39. Collins M, Yu L, Campuzano I, Grant SGN, Choudhary J (2008) Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder. Molecular & Cellular Proteomics 7: 1331-1348.

40. Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends in Biochemical Sciences 33: 2-8.

41. Fuxreiter M, Tompa P (2012) Fuzzy complexes: A more stochastic view of protein function. In: Fuxreiter M, Tompa P, editors. Fuzziness: Structural Disorder in Protein Complexes. New York: Springer. pp. 1-14.

42. Gupta G, Qin H, Song J (2012) Intrinsically unstructured domain 3 of hepatitis C Virus NS5A forms a "fuzzy complex" with VAPB-MSP domain which carries ALS-causing mutations. PLoS ONE 7: e39261.

43. Khan H, Cino E, Brickenden A, Fan J, Yang D, et al. (2013) Fuzzy complex formation between the intrinsically disordered prothymosin α and the Kelch domain of Keap1 involved in the oxidative stress response. Journal of Molecular Biology 425: 1011-1027.

44. Pawson T, Nash P (2003) Assembly of cell regulatory systems through protein interaction domains. Science 300: 445-452.

45. Lunde B, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. Nature Reviews Molecular Cell Biology 8: 479-490.

46. Ruthenburg A, Li H, Patel D, Allis CD (2007) Multivalent engagement of chromatin modifications by linked binding modules. Nature Reviews Molecular Cell Biology 8: 983-994.

47. Li P, Banjade S, Cheng H-C, Kim S, Chen B, et al. (2012) Phase transitions in the assembly of multivalent signalling proteins. Nature 483: 336-340.

48. Pavlopoulos A, Akam M (2011) Hox gene Ultrabithorax regulates distinct sets of target genes at successive stages of *Drosophila* haltere morphogenesis. Proceedings of the National Academy of Sciences of the United States of America 108: 2855-2860.

49. So A, Chaivorapol C, Bolton E, Li H, Yamamoto K (2007) Determinants of cell- and gene-specific transcriptional regulation by the glucocorticoid receptor. PLoS Genetics 3: e94.

50. Nicolaides N, Galata Z, Kino T, Chrousos G, Charmandari E (2010) The human glucocorticoid receptor: molecular basis of biologic function. Steroids 75: 1-12.

51. Akhtar M, Holmgren C, Göndör A, Vesterlund M, Kanduri C, et al. (2012) Cell type and context-specific function of PLAG1 for IGF2 P3 promoter activity. International Journal of Oncology 41: 1959-1966.

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

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

54. O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA, et al. (2001) The Polycomb-group gene Ezh2 is required for early mouse development. Molecular and Cellular Biology 21: 4330-4336.

55. Roberts DJ, Johnson RL, Burke AC, Nelson CE, Morgan BA, et al. (1995) Sonic hedgehog is an endodermal signal inducing BMP-4 and HOX genes during induction and regionalization of the chick hindgut. Development 121: 3163-3174.

56. Gellon G, McGinnis W (1998) Shaping animal body plans in development and evolution by modulation of Hox expression patterns. Bioessays 20: 116-125.

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

58. Duncan I, Montgomery G (2002) E. B. Lewis and the bithorax complex: part I. Genetics 160: 1265-1272.

59. Laney JD, Biggin MD (1997) Zeste-mediated activation by an enhancer is independent of cooperative DNA binding *in vivo*. Proceedings of the National Academy of Sciences of the United States of America 94: 3602-3604.

60. Vachon G, Cohen B, Pfeifle C, McGuffin ME, Botas J, et al. (1992) Homeotic genes of the Bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene Distal-less. Cell 71: 437-450.

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

62. Lappin TRJ, Grier D, Thompson A, Halliday H (2006) HOX genes: seductive science, mysterious mechanisms. Ulster Medical Journal 75: 23-31.

63. Damante G, Pellizzari L, Esposito G, Fogolari F, Viglino P, et al. (1996) A molecular code dictates sequence-specific DNA recognition by homeodomains. EMBO Journal 15: 4992-5000.

64. Noyes M, Christensen R, Wakabayashi A, Stormo G, Brodsky M, et al. (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell 133: 1277-1289.

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

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

67. Mann RS, Chan SK (1996) Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. Trends Genet 12: 258-262.

68. 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.

69. Merabet S, Kambris Z, Capovilla M, Berenger H, Pradel J, et al. (2003) The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions. Developmental Cell 4: 761-768.

70. Bondos SE, 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: 145-171.

71. 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.

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

73. Mukherjee K, Burglin TR (2007) Comprehensive analysis of animal TALE homeobox genes: New conserved motifs and cases of accelerated evolution. Journal of Molecular Evolution 65: 137-153.

74. Chang CP, Shen WF, Rozenfeld S, Lawrence HJ, Largman C, et al. (1995) Pbx proteins display hexapeptide-dependent cooperative DNA-binding with a subset of hox proteins. Genes & Development 9: 663-674.

75. Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, et al. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. Cell 131: 530-543.

76. LaRonde-LeBlanc NA, Wolberger C (2003) Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes and Development 17: 2060-2072.

77. Lu Q, Kamps MP (1996) Structural determinants within Pbx1 that mediate cooperative DNA binding with pentapeptide-containing Hox proteins: Proposal for a model of a Pbx1-Hox-DNA complex. Molecular and Cellular Biology 16: 1632-1640.

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

79. 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.

80. Merabet S, Saadaoui M, Sambrani N, Hudry B, Pradel J, et al. (2007) A unique Extradenticle recruitment mode in the *Drosophila* Hox protein Ultrabithorax. Proceedings of the National Academy of Sciences of the United States of America 104: 16946-16951.

81. Gebelein B, McKay DJ, Mann RS (2004) Direct integration of Hox and segmentation gene inputs during *Drosophila* development. Nature 431: 653-659.

82. Gebelein B, Mann R (2007) Compartmental modulation of abdominal Hox expression by engrailed and sloppy-paired patterns the fly ectoderm. Developmental Biology 308: 593-605.

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

84. Ekker SC, Young KE, Vonkessler DP, Beachy PA (1991) Optimal DNA-sequence recognition by the Ultrabithorax homeodomain of *Drosophila*. Embo Journal 10: 1179-1186.

85. van Dijk MA, Murre C (1994) extradenticle raises the DNA binding specificity of homeotic selector gene products. Cell 78: 617-624.

86. 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: 603-615.

87. 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: 9924-9928.

88. Subramaniam V, Bomze HM, López AJ (1994) Functional differences between Ultrabithorax protein isoforms in *Drosophila melanogaster*: evidence from elimination, substitution and ectopic expression of specific isoforms. Genetics 136: 979-991.

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

90. Altman GH, Diaz F, Jakuba C, Calabro T, Horan RL, et al. (2003) Silk-based biomaterials. Biomaterials 24: 401-416.

91. Drury JL, Mooney DJ (2003) Hydrogels for tissue engineering: scaffold design variables and applications. Biomaterials 24: 4337-4351.

92. Hutmacher DW (2000) Scaffolds in tissue engineering bone and cartilage. Biomaterials 21: 2529-2543.

93. Karageorgiou V, Kaplan D (2005) Porosity of 3D biomaterial scaffolds and osteogenesis. Biomaterials 26: 5474-5491.

94. Lutolf MP, Hubbell JA (2005) Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. Nature Biotechnology 23: 47-55.

95. Almine JF, Bax DV, Mithieux SM, Nivison-Smith L, Rnjak J, et al. (2010) Elastin-based materials. Chemical Society Reviews 39: 3371-3379.

96. Betre H, Chilkoti A, Setton LA, Ieee, Ieee, et al. (2002) A two-step chondrocyte recovery system based on thermally sensitive elastin-like polypeptide scaffolds for cartilage tissue engineering. Second Joint Embs-Bmes Conference 2002, Vols 1-3, Conference Proceedings: Bioengineering - Integrative Methodologies, New Technologies. New York: Ieee. pp. 829-830.

97. Welsh ER, Tirrell DA (2000) Engineering the extracellular matrix: A novel approach to polymeric biomaterials. I. Control of the physical properties of artificial protein matrices designed to support adhesion of vascular endothelial cells. Biomacromolecules 1: 23-30.

98. Girotti A, Reguera J, Rodriguez-Cabello JC, Arias FJ, Alonso M, et al. (2004) Design and bioproduction of a recombinant multi(bio)functional elastin-like protein polymer containing cell adhesion sequences for tissue engineering purposes. Journal of Materials Science-Materials in Medicine 15: 479-484.

99. Swierczewska M, Hajicharalambous CS, Janorkar AV, Megeed Z, Yarmush ML, et al. (2008) Cellular response to nanoscale elastin-like polypeptide polyelectrolyte multilayers. Acta Biomaterialia 4: 827-837.

100. Li DY, Brooke B, Davis EC, Mecham RP, Sorensen LK, et al. (1998) Elastin is an essential determinant of arterial morphogenesis. Nature 393: 276-280.

101. Cox BA, Starcher BC, Urry DW (1974) Coacervation of tropoelastin results in fiber formation. Journal of Biological Chemistry 249: 997-998.

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

103. Wise SG, Mithieux SM, Weiss AS (2009) Engineered tropoelastin and elastin-based biomaterials. In: McPherson A, editor. Advances in Protein Chemistry and Structural Biology, Vol 78. pp. 1-24.

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

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

106. Lee CH, Singla A, Lee Y (2001) Biomedical applications of collagen. International Journal of Pharmaceutics 221: 1-22.

107. Hoffman AS (2002) Hydrogels for biomedical applications. Advanced Drug Delivery Reviews 54: 3-12.

108. Olsen D, Yang CL, Bodo M, Chang R, Leigh S, et al. (2003) Recombinant collagen and gelatin for drug delivery. Advanced Drug Delivery Reviews 55: 1547-1567.

109. Brodsky B, Baum J (2008) Structural biology - Modelling collagen diseases. Nature 453: 998-999.

110. de la Puente P, Ludeñab D (2014) Cell culture in autologous fibrin scaffolds for applications in tissue engineering. Experimental Cell Research 322: 1-11.

111. Weber SC, Kauffman JI, Parise C, Weber SJ, Katz SD (2013) Platelet-rich fibrin matrix in the management of arthroscopic repair of the rotator cuff a prospective, randomized, double-blinded study. American Journal of Sports Medicine 41: 263-270.

112. Jockenhoevel S, Zund G, Hoerstrup SP, Chalabi K, Sachweh JS, et al. (2001) Fibrin gel-advantages of a new scaffold in cardiovascular tissue engineering. European Journal of Cardio-Thoracic Surgery 19: 424-430.

113. Carriel V, Garzon I, Jimenez JM, Oliveira AC, Arias-Santiago S, et al. (2012) Epithelial and stromal developmental patterns in a novel substitute of the human skin generated with fibrin-agarose biomaterials. Cells Tissues Organs 196: 1-12.

114. Jin HJ, Kaplan DL (2003) Mechanism of silk processing in insects and spiders. Nature 424: 1057-1061.

115. Hu X, Lu Q, Sun L, Cebe P, Wang X, et al. (2010) Biomaterials from ultrasonication-induced silk fibroin-hyaluronic acid hydrogels. Biomacromolecules 11: 3178-3188.

116. Talukdar S, Mandal M, Hutmacher DW, Russell PJ, Soekmadji C, et al. (2011) Engineered silk fibroin protein 3D matrices for *in vitro* tumor model. Biomaterials 32: 2149-2159.

117. Omenetto FG, Kaplan DL (2010) New opportunities for an ancient material. Science 329: 528-531.

118. Hu XA, Lu QA, Sun L, Cebe P, Wang XQ, et al. (2010) Biomaterials from ultrasonication-induced silk fibroin-hyaluronic acid hydrogels. Biomacromolecules 11: 3178-3188.

119. Hu X, Shmelev K, Sun L, Gil ES, Park SH, et al. (2011) Regulation of silk material structure by temperature-controlled water vapor annealing. Biomacromolecules 12: 1686-1696.

120. Sutherland TD, Young JH, Weisman S, Hayashi CY, Merritt DJ (2010) Insect silk: One name, many materials. Annual Review of Entomology. pp. 171-188.

121. Vollrath F, Knight DP (2001) Liquid crystalline spinning of spider silk. Nature 410: 541-548.

122. Sutherland TD, Church JS, Hu XA, Huson MG, Kaplan DL, et al. (2011) Single honeybee silk protein mimics properties of multi-protein silk. PLoS One 6: e16489.

123. Weisman S, Haritos VS, Church JS, Huson MG, Mudie ST, et al. (2010) Honeybee silk: Recombinant protein production, assembly and fiber spinning. Biomaterials 31: 2695-2700.

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

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

126. 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: 97-104.

127. Patterson JL, Arenas-Gamboa AM, Wang T-Y, Hsiao H-C, Howell DW, et al. (2014) Materials composed of the *Drosophila* Hox protein Ultrabithorax are biocompatible and nonimmunogenic. Journal of Biomedical Materials Research Part A: doi: 10.1002/jbm.a.35295.

128. Chen G, Shin JA (2008) AhR/Arnt : XRE interaction: Turning false negatives into true positives in the modified yeast one-hybrid assay. Analytical Biochemistry 382: 101-106.

129. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, et al. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences of the United States of America 98: 4569-4574.

130. Luban J, Goff SP (1995) The yeast two-hybrid system for studying protein-protein interactions. Current Opinion in Biotechnology 6: 59-64.

131. Tompa P (2005) The interplay between structure and function in intrinsically unstructured proteins. FEBS Letters 579: 3346-3354.

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

133. Huang H, Jedynak B, Bader J (2007) Where have all the interactions gone? Estimating the coverage of two-hybrid protein interaction maps. PLoS Computational Biology 3: e214.

134. Tan XX, Bondos S, Li LK, Matthews KS (2002) Transcription activation by ultrabithorax Ib protein requires a predicted alpha-helical region. Biochemistry 41: 2774-2785.

135. Herrmann F, Garriga Canut M, Baumstark R, Fajardo Sanchez E, Cotterell J, et al. (2011) p53 Gene repair with zinc finger nucleases optimised by yeast 1-hybrid and validated by Solexa sequencing. PLoS ONE 6: e20913.

136. Bondos SE, Catanese DJ, Tan XX, Bicknell A, Li LK, 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: 26433-26444.

137. Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. Nature 402: 86-90.

138. Iwabuchi K, Li B, Bartel P, Fields S (1993) Use of the 2-hybrid system to identify the domain of p53 involved in oligomerization. Oncogene 8: 1693-1696.

139. Auerbach D, Thaminy S, Hottiger MO, Stagljar I (2002) The post-genomic era of interactive proteomics: Facts and perspectives. Proteomics 2: 611-623.

140. Bellay J, Han SJ, Michaut M, Kim T, Costanzo M, et al. (2011) Bringing order to protein disorder through comparative genomics and genetic interactions. Genome Biology 12: 15.

141. Thomas J, Ramakrishnan N, Bailey-Kellogg C (2009) Graphical models of protein-protein interaction specificity from correlated mutations and interaction data. Proteins-Structure Function and Bioinformatics 76: 911-929.

142. Zen A, Micheletti C, Keskin O, Nussinov R (2010) Comparing interfacial dynamics in protein-protein complexes: an elastic network approach. BMC Structural Biology 10: 13.

143. JimÃƒnez J (2005) Does structural and chemical divergence play a role in precluding undesirable protein interactions? Proteins 59: 757-764.

144. Lawrence MC, Colman PM (1993) Shape complementarity at protein/protein interfaces. Journal of Molecular Biology 234: 946-950.

145. Chothia C, Janin J (1975) Principles of protein-protein recognition. Nature 256: 705-708.

146. Yura K, Hayward S (2009) The interwinding nature of protein-protein interfaces and its implication for protein complex formation. Bioinformatics 25: 3108-3113.

147. Ball KA, Phillips A, Nerenberg P, Fawzi N, Wemmer D, et al. (2011) Homogeneous and heterogeneous tertiary structure ensembles of amyloid peptides. Biochemistry 50: 7612-7628.

148. Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa P (2006) Disorder and sequence repeats in hub proteins and their implications for network evolution. Journal of Proteome Research 5: 2985-2995.

149. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets - The roles of intrinsic disorder in protein interaction networks. Febs Journal 272: 5129-5148.

150. Ekman D, Light S, Bjorklund AK, Elofsson A (2006) What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae? Genome Biology 7: 13.

151. Fisher CK, Huang A, Stultz CM (2010) Modeling intrinsically disordered proteins with bayesian statistics. Journal of the American Chemical Society 132: 14919-14927.

152. Patil A, Nakamura H (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. FEBS Letters 580: 2041-2045.

153. Patil A, Kinoshita K, Nakamura H (2010) Hub promiscuity in protein-protein interaction networks. International Journal of Molecular Sciences 11: 1930-1943.

154. Singh GP, Ganapathi M, Dash D (2007) Role of intrinsic disorder in transient interactions of hub proteins. Proteins-Structure Function and Bioinformatics 66: 761-765.

155. Marsh JA, Dancheck B, Ragusa MJ, Allaire M, Forman-Kay JD, et al. (2010) Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. Structure 18: 1094-1103.

156. Receveur Brechot V, Durand D (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. Current Protein & Peptide Science 13: 55-75.

157. Hsu WL, Oldfield CJ, Xue B, Meng JW, Huang F, et al. (2013) Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. Protein Science 22: 258-273.

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

159. Ganguly D, Zhang WH, Chen JH (2012) Synergistic folding of two intrinsically disordered proteins: searching for conformational selection. Molecular Biosystems 8: 198-209.

160. James LC, Roversi P, Tawfik DS (2003) Antibody multispecificity mediated by conformational diversity. Science 299: 1362-1367.

161. Uversky VN (2011) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. Chemical Society Reviews 40: 1623-1634.

162. Rogers J, Steward A, Clarke J (2013) Folding and binding of an intrinsically disordered protein: fast, but not 'diffusion-limited'. Journal of the American Chemical Society 135: 1415-1422.

163. Nagulapalli M, Parigi G, Yuan J, Gsponer J, Deraos G, et al. (2012) Recognition pliability is coupled to structural heterogeneity: A Calmodulin intrinsically disordered binding region complex. Structure 20: 522-533.

164. Fuxreiter M, Simon I, Bondos S (2011) Dynamic protein-DNA recognition: beyond what can be seen. Trends in Biochemical Sciences 36: 415-423.

165. Mittag T, Kay LE, Forman-Kay JD (2010) Protein dynamics and conformational disorder in molecular recognition. Journal of Molecular Recognition 23: 105-116.

166. Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447: 1021-U1011.

167. Clarke D, Bhardwaj N, Gerstein MB (2012) Novel insights through the integration of structural and functional genomics data with protein networks. Journal of Structural Biology 179: 320-326.

168. Tran HT, Mao A, Pappu RV (2008) Role of backbone - Solvent interactions in determining conformational equilibria of intrinsically disordered proteins. Journal of the American Chemical Society 130: 7380-7392.

169. Vitalis A, Wang XL, Pappu RV (2008) Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. Journal of Molecular Biology 384: 279-297.

170. Marsh JA, Teichmann SA, Forman-Kay JD (2012) Probing the diverse landscape of protein flexibility and binding. Current Opinion in Structural Biology 22: 643-650.

171. Zhou HX (2010) From induced fit to conformational selection: A continuum of binding mechanism controlled by the timescale of conformational transitions. Biophysical Journal 98: 15-17.

172. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, et al. (2003) A protein interaction map of *Drosophila melanogaster*. Science 302: 1727-1736.

173. Dosztanyi Z, Meszaros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 25: 2745-2746.

174. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) Scop - A structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology 247: 536-540.

175. Boube M, Hudry B, Immarigeon C, Carrier Y, Bernat-Fabre S, et al. (2014) *Drosophila melanogaster* Hox transcription factors access the RNA polymerase II machinery through direct homeodomain binding to a conserved motif of mediator subunit Med19. PLoS Genetics 10: e1004303.

176. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJP, et al. (2008) Data growth and its impact on the SCOP database: new developments. Nucleic Acids Research 36: 419-425.

177. Yook SH, Oltvai ZN, Barabasi AL (2004) Functional and topological characterization of protein interaction networks. Proteomics 4: 928-942.

178. Galant R, Walsh CM, Carroll SB (2002) Hox repression of a target gene: Extradenticle-independent, additive action through multiple monomer binding sites. Development 129: 3115-3126.

179. Grenier JK, Carroll SB (2000) Functional evolution of the Ultrabithorax protein. Proceedings of the National Academy of Sciences of the United States of America 97: 704-709.

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

181. Bustos DM, Iglesias AA (2006) Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins. Proteins-Structure Function and Bioinformatics 63: 35-42.

182. Romero P, Obradovic Z, Li XH, Garner EC, Brown CJ, et al. (2001) Sequence complexity of disordered protein. Proteins-Structure Function and Genetics 42: 38-48.

183. Gould CM, Diella F, Via A, Puntervoll P, Gemund C, et al. (2010) ELM: The status of the 2010 eukaryotic linear motif resource. Nucleic Acids Research 38: 167-180.

184. Panni S, Montecchi-Palazzi L, Kiemer L, Cabibbo A, Paoluzi S, et al. (2011) Combining peptide recognition specificity and context information for the prediction of the 14-3-3-mediated interactome in *S. cerevisiae* and *H. sapiens*. Proteomics 11: 128-143.

185. Puntervoll P, Linding R, Gemund C, Chabanis-Davidson S, Mattingsdal M, et al. (2003) ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. Nucleic Acids Research 31: 3625-3630.

186. Jemth P, Mu X, Engstrom A, Dogan J (2014) A frustrated binding interface for intrinsically disordered proteins. Journal of Biological Chemistry 289: 5528-5533.

187. Narayan V, Halada P, Hernychova L, Chong YP, Zakova J, et al. (2011) A multiprotein binding interface in an intrinsically disordered region of the tumor suppressor protein interferon regulatory factor-1. Journal of Biological Chemistry 286: 14291-14303.

188. Morgan JL, Song YJ, Barbar E (2011) Structural dynamics and multiregion interactions in dynein-dynactin recognition. Journal of Biological Chemistry 286: 39349-39359.

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

190. Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, et al. (2012) Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Molecular Cell 46: 871-883.

191. Ellis JD, Barrios-Rodiles M, Colak R, Irimia M, Kim T, et al. (2012) Tissue-specific alternative splicing remodels protein-protein interaction networks. Molecular Cell 46: 884-892.

192. Gebelein B, Culi J, Ryoo H, Zhang W, Mann R (2002) Specificity of Distalless repression and limb primordia development by abdominal Hox proteins. Developmental cell 3: 487-498.

193. Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 286: 509-512.

194. Brown CJ, Johnson AK, Daughdrill GW (2010) Comparing models of evolution for ordered and disordered proteins. Molecular Biology and Evolution 27: 609-621.

195. Butz M, Kast P, Hilvert D (2014) Affinity maturation of a computationally designed binding protein affords a functional but disordered polypeptide. Journal of Structural Biology 185: 168-177.

196. Light S, Sagit R, Sachenkova O, Ekman D, Elofsson A (2013) Protein expansion is primarily due to indels in intrinsically disordered regions. Molecular Biology and Evolution 30: 2645-2653.

197. Saadaoui M, Merabet S, Litim-Mecheri I, Arbeille E, Sambrani N, et al. (2011) Selection of distinct Hox-Extradenticle interaction modes fine-tunes Hox protein activity. Proceedings of the National Academy of Sciences of the United States of America 108: 2276-2281.

198. Walsh CM, Carroll SB (2007) Collaboration between Smads and a Hox protein in target gene repression. Development 134: 3585-3592.

199. Lohr U, Pick L (2005) Cofactor-interaction motifs and the cooption of a homeotic hox protein into the segmentation pathway of *Drosophila melanogaster*. Current Biology 15: 643-649.

200. Cukierman E, Pankov R, Stevens DR, Yamada KM (2001) Taking cell-matrix adhesions to the third dimension. Science 294: 1708-1712.

201. Yamada KM, Cukierman E (2007) Modeling tissue morphogenesis and cancer in 3D. Cell 130: 601-610.

202. Luo Y, Wang CM, Hossain M, Qiao Y, Ma LY, et al. (2012) Three-dimensional microtissue assay for high-throughput cytotoxicity of nanoparticles. Analytical Chemistry 84: 6731-6738.

203. Ziolkowska K, Kwapiszewski R, Brzozka Z (2011) Microfluidic devices as tools for mimicking the *in vivo* environment. New Journal of Chemistry 35: 979-990.

204. Pampaloni F, Reynaud E, Stelzer EHK (2007) The third dimension bridges the gap between cell culture and live tissue. Nature Reviews Molecular Cell Biology 8: 839-845.

205. Delcommenne M, Streuli CH (1995) Control of integrin expression by extracellular matrix. Journal of Biological Chemistry 270: 26794-26801.

206. Ghosh S, Spagnoli G, Martin I, Ploegert S, Demougin P, et al. (2005) Three-dimensional culture of melanoma cells profoundly affects gene expression profile: a high density oligonucleotide array study. Journal of Cellular Physiology 204: 522-531.

207. Mishra D, Sakamoto J, Thrall M, Baird B, Blackmon S, et al. (2012) Human lung cancer cells grown in an ex vivo 3D lung model produce matrix metalloproteinases not produced in 2D culture. PLoS ONE 7: e45308.

208. Yamaguchi Y, Deng DW, Sato Y, Hou YT, Watanabe R, et al. (2013) Silicate fiber-based 3D cell culture system for anticancer drug screening. Anticancer Research 33: 5301-5309.

209. Dunne LW, Huang Z, Meng WX, Fan XJ, Zhang NY, et al. (2014) Human decellularized adipose tissue scaffold as a model for breast cancer cell growth and drug treatments. Biomaterials 35: 4940-4949.

210. Girard YK, Wang CY, Ravi S, Howell MC, Mallela J, et al. (2013) A 3D fibrous scaffold inducing tumoroids: A platform for anticancer drug development. PLoS ONE 8: e75345.

211. Loessner D, Stok K, Lutolf M, Hutmacher D, Clements J, et al. (2010) Bioengineered 3D platform to explore cell-ECM interactions and drug resistance of epithelial ovarian cancer cells. Biomaterials 31: 8494-8506.

212. Feng X, Tonnesen MG, Mousa SA, Clark RAF (2013) Fibrin and collagen differentially but synergistically regulate sprout angiogenesis of human dermal microvascular endothelial cells in 3-dimensional matrix. International Journal of Cell Biology 2013: 11.

213. Liu J, Tan Y, Zhang H, Zhang Y, Xu P, et al. (2012) Soft fibrin gels promote selection and growth of tumorigenic cells. Nature Materials 11: 734-741.

214. Cross V, Zheng Y, Won Choi N, Verbridge S, Sutermaster B, et al. (2010) Dense type I collagen matrices that support cellular remodeling and microfabrication for studies of tumor angiogenesis and vasculogenesis *in vitro*. Biomaterials 31: 8596-8607.

215. Sieh S, Taubenberger A, Rizzi S, Sadowski M, Lehman M, et al. (2012) Phenotypic characterization of prostate cancer LNCaP cells cultured within a bioengineered microenvironment. PLoS ONE 7: e40217.

216. Chwalek K, Bray LJ, Werner C (2014) Tissue-engineered 3D tumor angiogenesis models: Potential technologies for anti-cancer drug discovery. Advanced Drug Delivery Reviews 79-80c: 30-39.

217. Chwalek K, Bray LJ, Werner C Tissue-engineered 3D tumor angiogenesis models: Potential technologies for anti-cancer drug discovery. Advanced Drug Delivery Reviews.

218. Appel B, Sakonju S (1993) Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products UBX and ABD-A in *Drosophila* embryos. Embo j 12: 1099-1109.

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

220. Soule HD, Maloney TM, Wolman SR, Peterson WD, Brenz R, et al. (1990) Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Research 50: 6075-6086.

221. Dawson PJ, Wolman SR, Tait L, Heppner GH, Miller FR (1996) MCF10AT: A model for the evolution of cancer from proliferative breast disease. American Journal of Pathology 148: 313-319.

222. Santner SJ, Dawson PJ, Tait L, Soule HD, Eliason J, et al. (2001) Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells. Breast Cancer Research and Treatment 65: 101-110.

223. Derossi D, Calvet S, Trembleau A, Brunissen A, Chassaing G, et al. (1996) Cell internalization of the third helix of the antennapedia homeodomain is receptor-independent. Journal of Biological Chemistry 271: 18188-18193.

224. Derossi D, Chassaing G, Prochiantz A (1998) Trojan peptides: the penetratin system for intracellular delivery. Trends in Cell Biology 8: 84-87.

225. Thoren PEG, Persson D, Karlsson M, Norden B (2000) The Antennapedia peptide penetratin translocates across lipid bilayers - the first direct observation. FEBS Letters 482: 265-268.

226. Marella N, Malyavantham K, Wang J, Matsui S-i, Liang P, et al. (2009) Cytogenetic and cDNA microarray expression analysis of MCF10 human breast cancer progression cell lines. Cancer Research 69: 5946-5953.

227. Mbeunkui F, Metge B, Shevde L, Pannell L (2007) Identification of differentially secreted biomarkers using LC-MS/MS in isogenic cell lines representing a progression of breast cancer. Journal of Proteome Research 6: 2993-3002.

228. Gomes SC, Leonor IB, Mano JF, Reis RL, Kaplan DL (2011) Antimicrobial functionalized genetically engineered spider silk. Biomaterials 32: 4255-4266.

229. Provenzano PP, Inman DR, Eliceiri KW, Keely PJ (2009) Matrix density-induced mechanoregulation of breast cell phenotype, signaling and gene expression through a FAK-ERK linkage. Oncogene 28: 4326-4343.

230. Paszek MJ, Zahir N, Johnson KR, Lakins JN, Rozenberg GI, et al. (2005) Tensional homeostasis and the malignant phenotype. Cancer Cell 8: 241-254.

231. Kraning Rush C, Reinhart King C (2012) Controlling matrix stiffness and topography for the study of tumor cell migration. Cell Adhesion and Migration 6: 274-279.

232. Saha S, Duan XR, Wu LY, Lo PK, Chen HX, et al. (2012) Electrospun fibrous scaffolds promote breast cancer cell alignment and epithelial-mesenchymal transition. Abstracts of Papers of the American Chemical Society 243: 1.

233. Kessenbrock K, Plaks V, Werb Z (2010) Matrix metalloproteinases: Regulators of the tumor microenvironment. Cell 141: 52-67.

234. Bogenrieder T, Herlyn M (2003) Axis of evil: molecular mechanisms of cancer metastasis. Oncogene 22: 6524-6536.

235. Shiomi T, Okada Y (2003) MT1-MMP and MMP-7 in invasion and metastasis of human cancers. Cancer and Metastasis Reviews 22: 145-152.

236. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: The next generation. Cell 144: 646-674.

237. Butcher DT, Alliston T, Weaver VM (2009) A tense situation: forcing tumour progression. Nature Reviews Cancer 9: 108-122.

238. Mierke CT (2014) The fundamental role of mechanical properties in the progression of cancer disease and inflammation. Reports on Progress in Physics 77.

239. Friedl P, Wolf K (2003) Tumour-cell invasion and migration: diversity and escape mechanisms. Nature Reviews Cancer 3: 362-374.

240. Friedl P, Wolf K (2010) Plasticity of cell migration: a multiscale tuning model. The Journal of Cell Biology 188: 11-19.

241. Wyckoff J, Pinner S, Gschmeissner S, Condeelis J, Sahai E (2006) ROCK- and myosin-dependent matrix deformation enables protease-independent tumor-cell invasion *in vivo*. Current Biology 16: 1515-1523.

242. Sabeh F, Shimizu Hirota R, Weiss S (2009) Protease-dependent versus -independent cancer cell invasion programs: three-dimensional amoeboid movement revisited. The Journal of Cell Biology 185: 11-19.

243. Wang YC, Klemke RL (2011) Proteomics method for identification of pseudopodium phosphotyrosine proteins. In: Shimaoka M, editor. Integrin and Cell Adhesion Molecules: Methods and Protocols. Totowa: Humana Press Inc. pp. 349-365.

244. Klemke RL, Jiang XN, Choi S, Kelber JA (2013) Proteomic and biochemical methods to study the cytoskeletome. In: Coutts AS, editor. Adhesion Protein Protocols, 3rd Edition. Totowa: Humana Press Inc. pp. 203-218.