

Evolution of the activation domain in a Hox transcription factor

YING LIU^{1,2}, ANNIE HUANG², REBECCA M. BOOTH², GABRIELA GERALDO MENDES², ZABEENA MERCHANT¹,
KATHLEEN S. MATTHEWS¹ and SARAH E. BONDOS^{*,1,2}

¹Department of Biosciences, Rice University, Houston and

²Department of Molecular and Cellular Medicine, Texas A&M Health Science Center, College Station, TX, USA

ABSTRACT Linking changes in amino acid sequences to the evolution of transcription regulatory domains is often complicated by the low sequence complexity and high mutation rates of intrinsically disordered protein regions. For the Hox transcription factor Ultrabithorax (Ubx), conserved motifs distributed throughout the protein sequence enable direct comparison of specific protein regions, despite variations in the length and composition of the intervening sequences. In cell culture, the strength of transcription activation by *Drosophila melanogaster* Ubx correlates with the presence of a predicted helix within its activation domain. Curiously, this helix is not preserved in species more divergent than flies, suggesting the nature of transcription activation may have evolved. To determine whether this helix contributes to *Drosophila* Ubx function *in vivo*, wild-type and mutant proteins were ectopically expressed in the developing wing and the phenotypes evaluated. Helix mutations alter *Drosophila* Ubx activity in the developing wing, demonstrating its functional importance *in vivo*. The locations of activation domains in Ubx orthologues were identified by testing the ability of truncation mutants to activate transcription in yeast one-hybrid assays. In Ubx orthologues representing 540 million years of evolution, the ability to activate transcription varies substantially. The sequence and the location of the activation domains also differ. Consequently, analogous regions of Ubx orthologues change function over time, and may activate transcription in one species, but have no activity, or even inhibit transcription activation in another species. Unlike homeodomain-DNA binding, the nature of transcription activation by Ubx has substantially evolved.

KEY WORDS: *intrinsic disorder, natively unfolded, Ultrabithorax, Ubx, evo-devo*

Introduction

Evolution of the gene regulatory networks that act during development can create new morphological features and alter the body plan. Gene regulatory networks can evolve through changes in transcription factor binding sites (Villar *et al.*, 2014), alterations in groups of binding sites in cis-regulatory modules (Lonfat *et al.*, 2015; Luo and Baker 2015; Fish *et al.*, 2015), or evolution of transcription factor protein sequences (Galant and Carroll, 2002; Ronshuagen *et al.*, 2002; Cheatle Jarvela *et al.*, 2015). However, many transcription factors active in animal development are pleiotropic - regulating multiple genes in different tissues (Cheatle Jarvela and Hinman, 2015). Consequently, a beneficial change in transcription factor sequence in one context could have detrimental effects in another context (Chesmore *et al.*, 2016). The pleiotropy problem suggests that transcription factors are less likely to evolve

than the DNA sequences to which they bind.

Despite these constraints, several instances of transcription factor evolution have been reported (Sivanantharajah and Percival-Smith, 2015; Wagner and Lynch, 2008; Galant and Carroll 2002; Ronshaugen *et al.*, 2002). Linking evolutionary changes in transcription activation to variations in amino acid sequence is complicated by several challenges. First, because transcription factors operate in multiple tissues, a sequence within the transcription factor may be vital in one tissue and unnecessary in another tissue (differential pleiotropy, reviewed in Sivanantharajah and Percival-Smith, 2015). Second, although the changes in protein sequence should impact function *in vivo*, the collective action of

Abbreviations used in this paper: AfUbx, *Artemia franciscana* Ubx; AkUbx, *Akanthokara kaputensis* Ubx; HD, homeodomain; JcUbx, *Juonia coenia* Ubx; TcUbx, *Tribolium castaneum* Ubx; Ubx, Ultrabithorax.

*Address correspondence to: Sarah E. Bondos. 440 Reynolds Medical Building, Molecular and Cellular Medicine, TAMHSC, College Station, TX 77843-1114, USA. Tel: 979-436-0807. Fax: 979-847-9481. E-mail: sebondos@tamhsc.edu - web: <https://medicine.tamhsc.edu/faculty/bondos.html>

Submitted: 29 May, 2018; Accepted: 11 October, 2018.

many transcription factors acting on one gene can mask the impact of these changes (Villar *et al.*, 2014; Wagner and Lynch 2008). Indeed, the number of transcription factors that directly regulate a single target gene negatively correlates with evolutionary changes in the expression levels of that gene (Yang and Wittkopp 2017). Third, intrinsic disorder is often a vital part of transcription factor function (Bondos *et al.*, 2015; Triezenberg 1995; Ward *et al.*, 2004; Liu *et al.*, 2006; Staby *et al.*, 2017). Intrinsically disordered regions have low sequence complexity, a high mutation rate, and evolve rapidly via insertions and deletions (Light *et al.*, 2013; Brown *et al.*, 2002), features that complicate identifying and comparing analogous regions of protein sequence.

Many of these issues have been overcome for the *Drosophila* Hox transcription factor Ultrabithorax (Ubx), allowing examination of the evolution of its transcription activation domain. Hox transcription factors in general, and Ubx in particular, regulate many genes in a variety of tissues (Lonfat *et al.*, 2015; Grenier and Carroll 2000). Orthologues of Ubx have been identified that span 540 million years of evolution (Gallant and Carroll 2002, Ronshaugen *et al.*, 2002) and, when expressed in *Drosophila*, yield phenotypes distinct from those caused by *Drosophila* Ubx (DmUbx) expression (Grenier and Carroll, 2000; Galant and Carroll, 2002; Ronshaugen *et al.*, 2002). Therefore, in Ubx, changes in amino acid sequence have altered protein function. Importantly, conserved motifs distributed throughout the Ubx sequence can guide sequence analysis, allowing direct comparison of specific protein regions despite variations in the length and sequence of these regions.

Herein, we determine how evolutionary changes in the Ubx amino acid sequence have altered transcription activation. The transcription activation domain in Ubx is located in the N-terminal region of DmUbx (Fig. 1A) (Tan *et al.*, 2002). The activation domain can be subdivided into two regions, an “enhancer region” that increases the overall efficacy of the domain (aa 68-158), and a “core region” required for transcription activation (aa 159-242). Analysis of Ubx

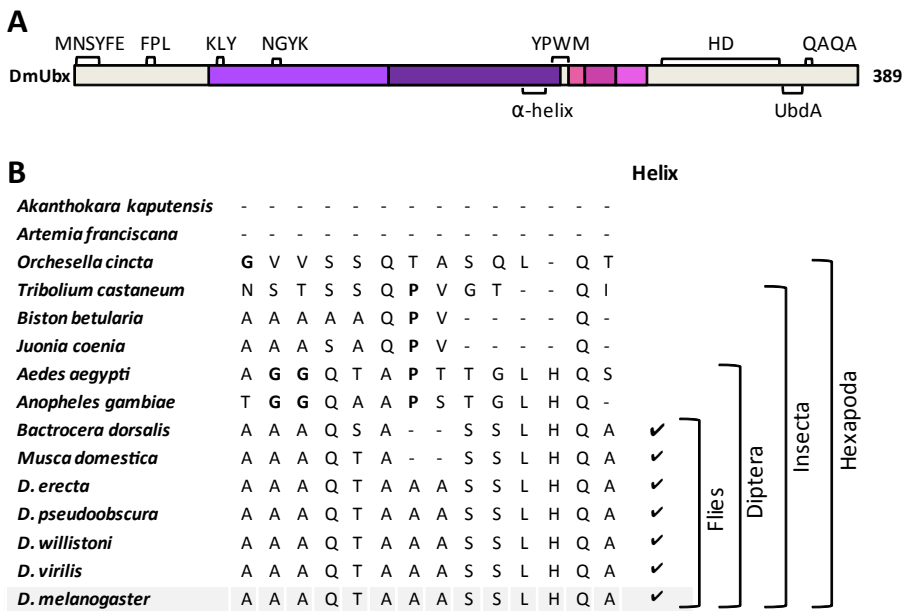
mutants revealed that the strength of activation correlates with the predicted presence of a helix within the core (aa 221-234), allowing generation of mutants that increase the strength of transcription activation or prevent transcription activation altogether by DmUbx (Tan *et al.*, 2002). Curiously, this helix is not preserved in species more divergent than flies, suggesting that either the helix is irrelevant for transcription activation *in vivo*, or the mechanism of Ubx transcription activation has substantially evolved.

In this paper, we demonstrate that helix mutants alter Ubx activity when expressed in the developing wing, demonstrating that this region is indeed functionally important *in vivo*. Despite lacking this helix, orthologues of DmUbx are all capable of activating transcription, some more effectively than DmUbx. As a first step toward elucidating evolutionary differences in the mechanism of transcription activation, we used a yeast one-hybrid assay to locate the transcription activation domain in Ubx orthologues. This assay is independent of other molecular functions, such as co-factor interactions, transcription repression, and DNA binding. Surprisingly, we found that the location of the activation domain within the protein has moved during evolution relative to the conserved motifs. Furthermore, sub-domain organization, amino acid enrichment, and even the role of specific protein segments have also changed. We conclude the nature of transcription activation by Ubx has substantially evolved.

Results

Testing the impact of Ubx activation domain α -helix mutations *in vivo*

A predicted helix was required for transcription activation by Ubx, both in yeast one-hybrid experiments, which are not dependent on DNA binding by Ubx, and promoter-reporter assays in culture of *Drosophila melanogaster* S2 cells, which require Ubx-DNA binding (Tan *et al.*, 2002). In these experiments, mutations that



to the predicted α -helical region. Only Ubx sequences isolated from flies are predicted to form a helix in this region. In addition to the non-*Drosophilid* species examined herein, listed are *Musca domestica* (common house fly), *Bactrocera dorsalis* (oriental fruit fly), *Anopheles gambiae* (African malaria mosquito), *Aedes aegypti* (yellow fever mosquito), *Biston betularia* (peppered moth), and *Orchesella cincta* (springtail).

Fig. 1. Evolution of the α -helix activation domain.

(A) *DmUbx* sequence schematic, showing the homeodomain (HD). The *ubx* mRNA is alternatively spliced, which alters the length of the amino acid sequence between the YPWM motif and the HD. Amino acid sequences included or excluded by alternative splicing are named *b*, *mI*, and *mII*, and indicated by boxes colored pink, dark pink, and light pink, respectively. The transcription activation core (purple box) and the transcription activation enhancing region (light purple box) are indicated. The position of a predicted α -helix that is required for transcription activation in cell culture assays is marked by a bracket. The MNSYFE motif is part of a larger sequence required for activation of the *tsh* gene in the head epidermis (Tour *et al.*, 2005), the YPWM motif is used in some contexts to bind the Hox co-factor Extradenticle (Johnson *et al.*, 1995) and inhibits DNA binding (Liu *et al.*, 2008). An alternate Exd-interaction motif is provided by the UbdA conserved motif. Finally, the QAQA motif confers the ability to repression transcription (Galant and Carroll, 2002; Ronshaugen *et al.*, 2002). **(B)** The amino acid sequence of Ubx orthologues corresponding

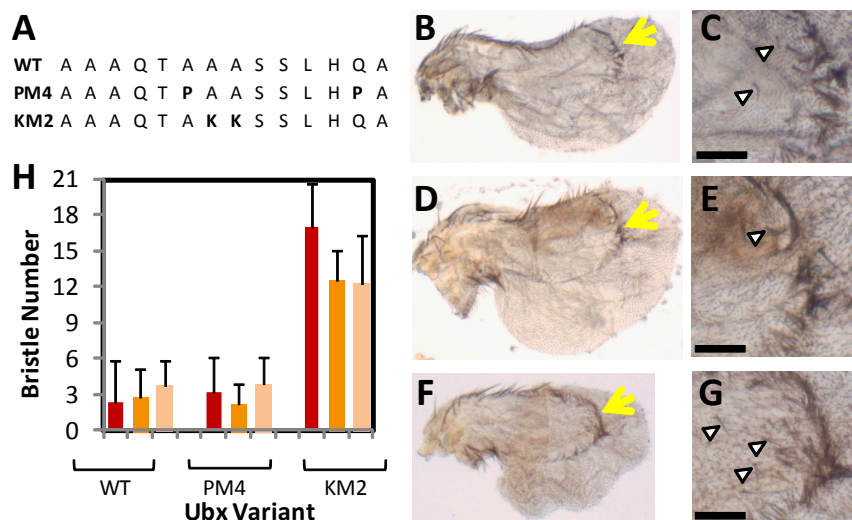


Fig. 2. Analysis of Ubx activation domain helix mutants in the developing *Drosophila* wing. (A) Sequence alignment of the activation domain predicted a-helix from DmUbx, with the two helix mutants tested *in vivo*. The prolines in the PM4 are expected to disrupt helix folding, and this mutant cannot activate transcription in cell culture (Tan *et al.*, 2002). The lysines in KM2 were designed to prevent alanines in the helix from forming hydrophobic interactions. This mutant is better able to activate transcription in cell culture. **(B-H)** Wild-type or mutant DmUbx was ectopically expressed in the third instar larval wing imaginal disc using the Gal4-MS1096 driver and fly lines with UAS-Ubx inserted at 2R-51D (red eyes), 3R-96E (orange eyes) and 2R-58A (light orange eyes). Adult wings are shown at 4x **(B,D,F)** and 10x **(C,E,G)**. The triple row of bristles marking the wing margin is indicated by a yellow arrow. Bristles inside the wing margin are indicated by white arrowheads. **(B,C)** Wild-type Ubx. **(D,E)** PM4 Ubx. **(F,G)** KM2 Ubx. **(H)** A graph depicting the number of bristles within the margin for the progeny of MS1096 virgin females mated with males harboring

the genes for wild-type, PM4, or KM2 DmUbx genes inserted into the genome at 2R-51D (red), 3R-96E (orange), 2R-58A (light orange) landing sites, respectively. Altering the Ubx gene insertion site had no impact on bristle number. However, the wings of all flies expressing the KM2 super-activating mutant displayed significantly more bristles than the wild-type or PM4 mutants inserted at the corresponding site ($p < 10^{-23}$).

increase the predicted helical propensity also increase the levels of transcription activation, whereas mutations that decrease helical propensity prevent transcription activation. These mutations neither impaired DNA binding nor transcription repression. While this predicted helix is conserved in *Drosophilids*, house flies (*Musca domestica*), and oriental fruit flies (*Bactrocera dorsalis*), the helix breaking amino acids proline and glycine are present even in the closely related mosquito Ubx sequences (*Anopheles gambiae* and *Aedes aegypti*) (Fig. 1B). Amino acids similar to those in the *Drosophila* helix are absent in more distantly related species. Thus, the vital role for this region in cell culture assays contrasts directly with its poor evolutionary conservation relative to other molecular functions such as DNA binding or transcription repression (Galant and Carroll, 2002; Ronshaugen *et al.*, 2002).

If this helix is truly important to the function of DmUbx, then mutation of this helix should also impact Ubx function *in vivo*. To test this hypothesis, wild-type, PM4, and KM2 DmUbx (Fig. 2A) were expressed in *Drosophila* larvae using the GAL4-UAS system (Duffy 2002). In the PM4 mutant, two amino acids are replaced by proline to prevent helix formation. This mutant was unable to activate transcription in cell culture assays (Tan *et al.*, 2002). The KM2 mutant, in which two alanines are replaced by lysine to alter the nature of a hydrophobic surface of the helix, is better able to activate transcription in S2 cells than wild-type Ubx (Tan *et al.*, 2002). Some aspects of Ubx function may be dependent on protein concentration. To ensure that a similar amount of each Ubx variant is produced in the developing larvae, the GAL4-Ubx DNA sequences were inserted into specific sites in the genome using phage ϕ C31 Integrase (Groth *et al.*, 2004). Three landing sites in the genome were tested, each of which yield a different eye color due the presence of the *white* gene in the inserted DNA: 2R-51D (red eyes), 3R-96E (orange eyes), and 2R-58A (light orange eyes).

We tested the ability of Ubx variants to transform the developing wing to a haltere fate, a well-established Ubx functional assay (Grenier and Carroll, 2000). Importantly, Exd, a Hox co-factor that can both activate and repress transcription, is not expressed in the developing wing pouch, and thus cannot mask changes in Ubx

function. Ubx regulates hundreds of genes to affect the wing to haltere transformation, roughly half of which are activated by Ubx (Pavlopoulos and Akam, 2006). Differences were observed when wild-type or variant Ubx was expressed in the pouch of the wing imaginal disc. An adult wing blade consists of two cell layers (dorsal and ventral faces), with the margin, located on the edge, studded with a triple row of bristles. In contrast, the haltere is much smaller, fluid-filled, and exhibits only a few large bristles on the ventral face. Expression of wild-type Ubx primarily on the dorsal side of the wing imaginal disc during larval development reduces the size of the appendage, especially on the dorsal side, and transforms the wing toward a haltere fate (Fig. 2 B,C) (Grenier and Carroll, 2000). Unlike wild type Ubx and the PM4 mutant, expression of the KM2 mutant in the developing wing created a structure with three lobes (Fig. 2F). When wild type Ubx is ectopically expressed, the triple row of bristles that normally forms along the wing margin is still largely intact, and additional haltere-like bristles appear on the dorsal face. Although expression of the PM4 mutant yields a similar phenotype (Fig. 2 D,E), expression of the more potent KM2 mutant dramatically increases the number of bristles (Fig. 2 F,G). This difference is statistically significant, and is not altered by the UAS-Ubx integration site used (Fig. 2H). Differences in protein expression created by using different landing sites for the Gal4-Ubx DNA did not impact experimental results. The PM4 mutant does not prevent transcription repression by Ubx (Tan *et al.*, 2002). Therefore, the lack of an effect by the PM4 mutant could be due to the nature of the Ubx-regulated genetic pathways that drives bristle formation in haltere development, since 9 of the 14 known Ubx target genes involved in bristle formation are repressed by Ubx (Galant *et al.*, 2002; Pavlopoulos and Akam, 2006). Increasing the efficacy of transcription activation by the KM2 mutant may reduce the ability to repress transcription of these genes.

The efficacy of transcription activation by Ubx has evolved

Observation of a mutant phenotype confirms that the predicted helix sequence is indeed important for DmUbx function *in vivo*. Consequently, the lack of a transcription activation helix in more

distant Ubx orthologues suggests that the nature of the Ubx activation domain must have evolved. As a first step towards elucidating these evolutionary changes, we used the yeast one-hybrid system to compare the strength of transcription activation by Ubx orthologues. In this assay, DNA encoding the LexA DNA binding domain is fused to the 5' end of the *ubx* gene. The yeast cells produce the corresponding LexA-Ubx fusion protein, which binds LexA binding sites to activate the *lacZ* reporter gene. One concern in using any assay with a fusion protein, such as the yeast one-hybrid assay, is attaching the heterologous protein (here, LexA) to the test protein (Ubx) could alter the function of the test protein, either by causing the test protein to mis-fold, or by sterically blocking ligand binding. However, yeast one-hybrid results using the LexA-Ubx fusion matched results from a promoter-reporter assay in *Drosophila* S2 cells, which does not require Ubx to be fused to another protein (Tan et al., 2002). Likewise, results from yeast two-hybrid experiments using LexA-Ubx matched data from phage display and GST-Ubx pulldown experiments, which rely on a different Ubx fusions, as well as for DNA binding gel supershift assays and *in vivo* genetic interactions using un-fused Ubx. Consequently, the LexA fusion is not altering Ubx activity in the studies below (Bondos et al., 2004, Bondos et al., 2006). In addition, differences in media composition were previously used to calibrate the sensitivity of the yeast two-hybrid assay to match results verified *in vivo* (Liu et al., 2011). These protocols were used in the assays reported herein.

Conversely, the yeast one-hybrid assay has an important advantage in studying Ubx - it reports solely on transcription activation. Much of the transcription activation domain also regulates DNA binding by the homeodomain (Liu et al., 2008), and is involved in protein interactions (Hsiao et al., 2014). Furthermore, the entire repression domain has not been located (Galant and Carroll, 2002; Ronshaugen et al., 2002). Therefore, mutations in this region have the potential to alter other Ubx molecular functions, which could mask the impact on transcription regulation in other assays.

Ubx orthologues were selected from the red flour beetle *Tribolium castaneum* (TcUbx), the butterfly *Juonia coenia* (JcUbx), the brine shrimp *Artemia franciscana* (AfUbx) and the velvet worm *Akanthokara kaputensis* (AkUbx). One of the challenges of this analysis is that the length of Ubx orthologues varies substantially, from 389 amino acids for DmUbx to 214 amino acids for AkUbx. Importantly, previously identified conserved motifs can be used as positional markers in the protein sequence, much of which is

intrinsically disordered and poorly conserved (Galant and Carroll, 2002; Ronshaugen et al., 2002; Liu et al., 2008). These conserved motifs allow comparison of the function of analogous regions of the protein despite the divergence in sequence composition and length.

The strength of Ubx activation was monitored by the intensity of the blue color generated by the *lacZ* reporter, benchmarked against a color standard (Fig. 3). The Pos-LexA positive control yields dark blue colonies, whereas LexA alone, which lacks an activation domain, yields white colonies. Despite the fact that Ubx proteins isolated from non-fly species lack a sequence that is critical for transcription activation in DmUbx, all of the Ubx orthologues were able to activate transcription. Therefore, the lack of the putative helix found in flies does not preclude transcription activation in other species. Interestingly, LexA-AkUbx yielded the darkest colonies and, therefore, AkUbx has the strongest activation domain. Colonies expressing LexA-DmUbx were medium blue. The light blue color of colonies expressing LexA-TcUbx and LexA-JcUbx reveals TcUbx and JcUbx have the weakest activation domains. Thus, strength of transcription activation appears to have evolved.

Locating transcription activation domains in Ubx orthologues

As a first step toward understanding the differences in transcription activation between Ubx orthologues, we used the yeast one-hybrid assay to locate the activation domain in each protein. For each Ubx orthologue, a series of truncation mutants were created to iteratively remove the conserved motifs and intervening regions without bisecting amino acid repeats. The truncation mutants allowed us to identify transcription activation core domains, defined as regions being necessary and/or sufficient for transcription activation either in the full-length Ubx orthologue or in a truncation mutant of that protein. Activation enhancer domains were defined as regions that improve the efficacy of transcription in the full-length or truncated protein, but cannot independently activate transcription in any variants of that Ubx orthologue. Finally, activation inhibition domains are regions that inhibit transcription activation. Truncation mutants are named based on the organism from which that Ubx sequence is derived, followed by the amino acids included in the mutant. For instance, Jc(14-253) is composed of amino acids 14-253 of *Juonia coenia* Ubx.

JcUbx contains 253 amino acids, much shorter than the 389 amino acid DmUbx. The large intrinsically disordered region found in DmUbx (amino acids 106 - 229) is significantly shortened in

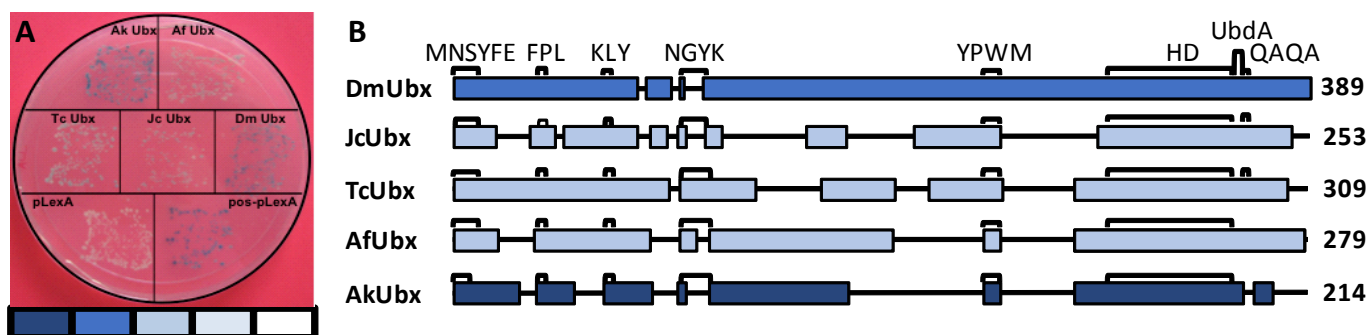
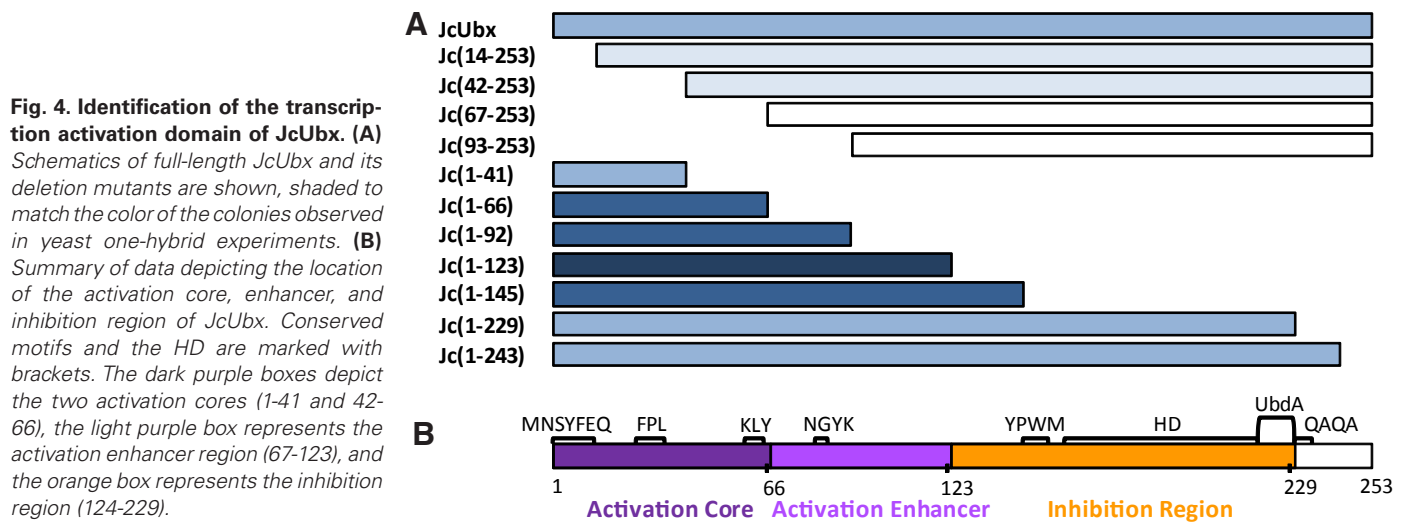


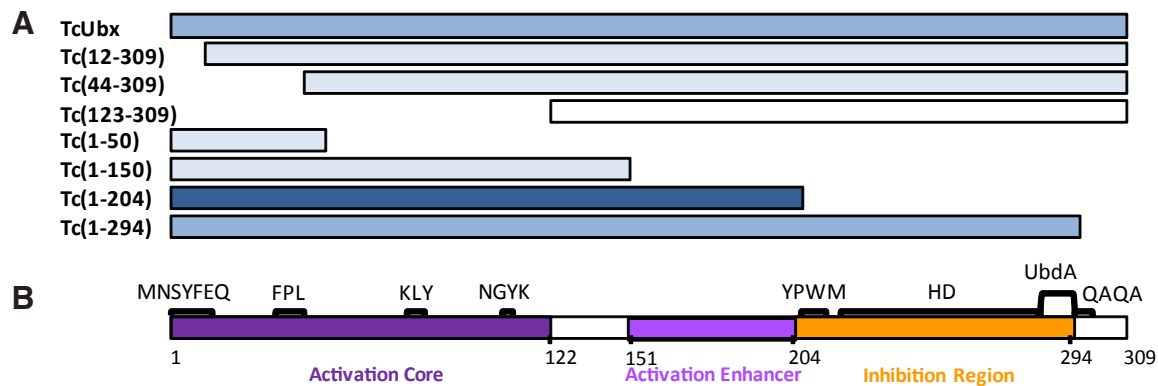
Fig. 3. Comparison of the strength of Ubx transcription activation domains. (A) Ubx orthologues yield different colored colonies in yeast one-hybrid assays, indicated different abilities to activate transcription. For consistency, colony color was gauged against a standard scale (below). The number of amino acids in each protein is indicated to the right of the schematics. (B) Alignment of Ubx orthologues, showing the position of conserved motifs (brackets) which are named for amino acid sequence, and the DNA binding homeodomain (HD). Gaps in the alignment are indicated by dashed black lines. The number of amino acids in each protein sequence is listed to the right.



JcUbx (Liu *et al.*, 2008), and the 13-amino acid poly glycine tract present in DmUbx (amino acids 111-123) is absent in JcUbx. Both proteins have an identical homeodomain amino acid sequence, a highly conserved C-terminus, and all of the identified conserved motifs. Unlike DmUbx, the N-terminus of JcUbx contributes to transcription activation (Fig. 4). Transcription activation levels were reduced by N-terminal truncations as small as 13 or 41 amino acids (Jc(14-253) and Jc(42-253)), and abolished upon loss of amino acids 42-66. Thus, amino acids 42-66 constitute part of the activation core domain because it is required for transcription activation in N-terminal JcUbx truncations. Conversely, removal of the C-terminal 14 amino acids had no impact on transcription activation (Jc(1-229) compared with Jc(1-243)). Further truncation that removes the YPWM motif and homeodomain actually *increases* levels of transcription activation (Jc(1-145)). Additional C-terminal truncation to amino acid 123 increased transcription activation still further. Therefore, amino acids 124-229 comprise the activation inhibition domain in JcUbx, which includes the YPWM motif, the homeodomain, and the UbdA motif. Further C-terminal truncations decrease transcription activation levels, with some activity detected in a fragment only containing amino acids 1-41. Because this small fragment can independently activate transcription activation, it is also part of the JcUbx activation core. Thus, in JcUbx amino acids 1-66, including the first 3 conserved motifs, form the activation core

domain. Because Jc(1-123) is better able to activate transcription compared to Jc(1-92), then the JcUbx activation enhancer domain is located between amino acids 93-123.

Like JcUbx, full-length TcUbx from the red flour beetle is a relatively weak activator of transcription. Once again, the N-terminus of the protein contributes to transcription activation, since Tc(12-309) produces lighter blue colonies than the full-length protein (Fig. 5). The N-terminus is required for activation, because colonies expressing Tc(123-309) do not activate transcription. However, the N-terminus alone in Tc(1-50) or Tc(1-150) is less able to activate transcription than full-length protein. Removing the very C-terminus of TcUbx, including the QAQA conserved motif, does not impact transcription activation levels relative to the full-length protein. Similar to JcUbx, a C-terminal truncation removing the homeodomain and YPWM motif increases the level of reporter gene expression beyond that of the full-length protein. Therefore, TcUbx amino acids 204-294 constitute the Activation Inhibition Region. A further truncation which leaves the first four conserved motifs intact substantially decreases the level of transcription regulation (Tc(1-150) versus Tc(1-204)). No TcUbx truncation mutant was able to produce dark blue colonies, as observed for Jc(1-204), suggesting that the transcription activation domain in TcUbx in the absence of an inhibition domain, is inherently weaker than the activation domain of the other orthologues. Because Tc(123-309) cannot activate transcription



enhancer, and inhibition region of TcUbx. Conserved motifs and the HD are marked with brackets. The transcription regulatory regions include the activation core (12-122, purple), the activation enhancer regions (1-11 and 151-204, light purple), and the inhibition region (205-294, orange).

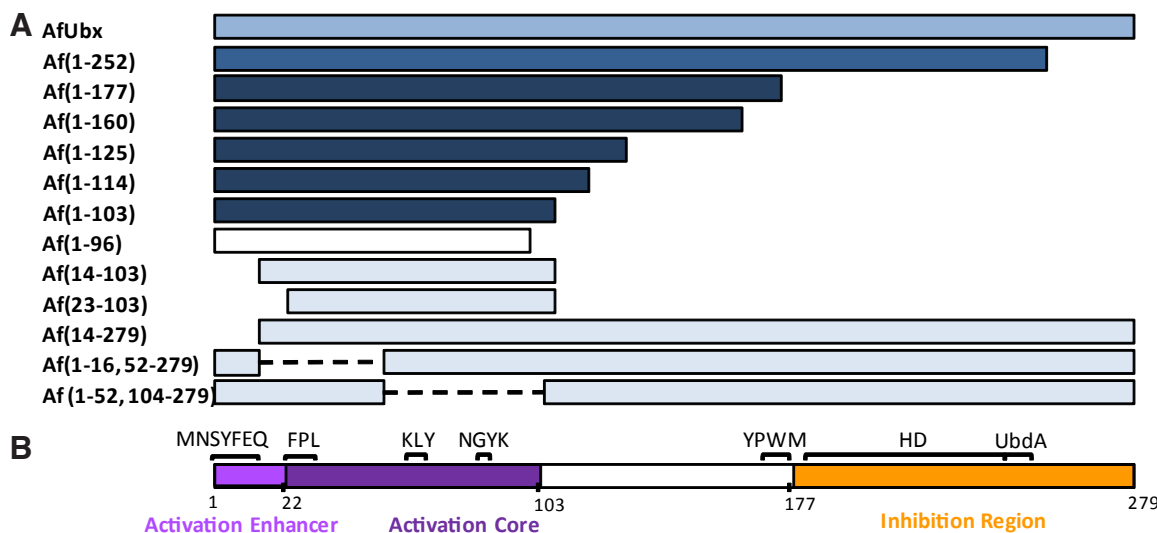


Fig. 6. Identification of the transcription activation domain of AfUbx. (A) Schematics of full-length AfUbx and its deletion mutants are shaded to match colony color from yeast one-hybrid experiments. (B) The location of the activation core (97-103, purple), enhancer (1-96, light purple), and inhibition regions (178-279, orange) of AfUbx. Conserved motifs and the HD are marked with brackets.

but Tc(44-309) can, the core activation domain includes residues 44-122. Furthermore, Tc(1-50) is able to activate low levels of transcription, thus this region is also part of the core activation domain. Consequently, the entire core activation domain in TcUbx spans residues 1 through 122. Inclusion of amino acids 151-204 further enhances transcription levels. Finally, the homeodomain and YPWM motif inhibit the activity of the transcription activation domain: Tc(1-204) is blue, whereas Tc(1-294) is light blue.

Analysis of AfUbx, derived from the brine shrimp *Artemia franciscana*, allows dissection of a Ubx activation domain in a crustacean. Both JcUbx and TcUbx have a QAQA motif near their C-termini, and removal of this motif had no effect on transcription levels. In contrast, AfUbx lacks this motif, and the corresponding truncation in AfUbx to remove the extreme C-terminus (amino acids 178-252, up to the UbdA motif) increased transcription activation. Removing the homeodomain further increased activation levels. Thus, the homeodomain and C-terminus of AfUbx constitute the activation inhibition region. Larger C-terminal truncations did not have a discernable effect until amino acids 96-103 were removed, which are required for transcription activation and therefore part of the activation core. Removing the N-terminal 14 amino acids reduced transcription activation levels (Af(14-279) vs. AfUbx, Fig. 6). Consequently, the N-terminus enhances transcription activation. Also truncating from the N-terminus (Af(14-103) and

Af(23-103)) reduces, but does not eliminate transcription activation levels compared to (Af(1-103)). Therefore, amino acids 1-22 form the activation enhancer domain, and amino acids 23-103, which can independently activate transcription, form the Activation Core Domain. The role of the AfUbx region between 14 and 96 were assessed using internal deletions. Removing amino acids 17-51 (Af(1-16,52-279)) or removing amino acids 52-101 (Af(1-51,103-279)) each reduced transcription activation relative to the full-length protein. Together, we conclude that the activation core in AfUbx spans amino acids 23-103 and the enhancer region is amino acids 1-22, while the inhibition region is located between amino acids 177-279.

The onychophoran (velvet worm) AkUbx was the most evolutionarily distant orthologue tested. Velvet worms are ancient animals with a simple body structure suggesting less evolved Hox function (Grenier and Carroll, 2000). The AkUbx protein is only 214 amino acids long, 55% of the length of DmUbx. The AkUbx sequence includes 5 of the 6 conserved motifs, and the intrinsically disordered regions are significantly reduced in length (Liu et al., 2008). Unlike the other Ubx orthologues, full-length AkUbx is a potent transcription activator. Removal of the N-terminal 12 amino acids reduced the protein's potency, and further truncation (29 amino acids, total) prevented transcription activation entirely (Fig. 7). Truncating the C-terminus up to residue 78 had no impact on transcription activation, whereas further truncation produced an inactive protein.

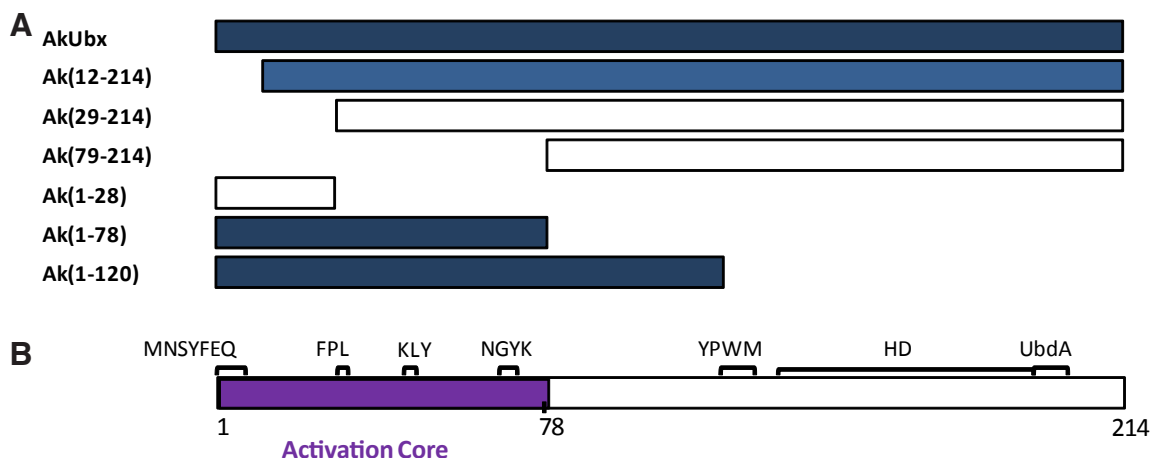


Fig. 7. Identification of the transcription activation domain of AkUbx. (A) Schematic structures of AkUbx and deletions shaded to match the color of colonies from yeast one-hybrid experiments. (B) Conserved motifs and the HD are marked with brackets on a schematic showing the position of the activation enhancer region (1-11) and the dark purple boxes represent the activation core region (12-78).

TABLE 1

**AMINO ACID CONTENT OF ACTIVATION CORE
AND ENHANCER DOMAINS IN UBX ORTHOLOGUES**

Ubx	Ala	Asn	Gly	His	Pro	Glu	Ser	Thr	Val	Tyr
Dm	16	10	27	1.7	3.4	2.9	8.6	5.1	5.1	2.9
Jc	14	24	18	3.3	7.3	8.1	7.3	1.6	2.4	7.3
Tc	13	6.3	9.7	7.4	5.7	8	18	2.8	5.1	5.1
Af	12	6.8	5.8	2.9	9.7	9.7	6.8	5.8	0	7.8
Ak	7.7	10.3	5.1	0	9	7.7	11.5	5.1	1.3	10.3

Enriched amino acids relative to the composition of an average protein are in bold (Tourasse and Li, 2000).

Consequently, the AkUbx core activation domain is residues 1-78, the smallest fragment capable of activating transcription. No region capable of inhibiting transcription was identified for this protein by us or by others (Galant and Carroll 2002).

Discussion

The transcription activation domain in the Hox protein Ultrathorax has evolved significantly. First, only Ubx derived from flies possess a sequence predicted to form an α -helix. This sequence is absolutely required for transcription activation by DmUbx in yeast and in *Drosophila* cell culture (Tan *et al.*, 2002) and contributes to Ubx activity *in vivo*. Second, comparison of five Ubx orthologues spanning 540 million years of evolution (Ronshaugen *et al.*, 2002, Gallant and Carroll, 2002) reveals that the strength of the activation domains has evolved (Fig. 8). Third, the locations of both the N- and C-termini of the activation domain differ among the orthologues. All orthologues except DmUbx include the N-terminus of the protein in their activation domain. The location of the C-terminus of the activation domain is much more variable. The activation domains of DmUbx and TcUbx end near the YPWM motif. The JcUbx and AkUbx activation domains extend just past the NGYK motif, while the AfUbx activation domain includes more of the sequence between the NGYK and YPWM motifs. Fourth, the location of the activation core domain relative to the activation enhancer domain also varies considerably. In DmUbx, and AfUbx, the activation core is

located C-terminal to the activation enhancing region. In contrast, the core of JcUbx, TcUbx, and AkUbx includes the N-terminus of each protein. Furthermore, AkUbx, derived from the least complex organism, lacks an activation enhancer domain. Consequently, the activation core of different orthologues encompasses different conserved motifs.

A fifth aspect that varies significantly is the presence and position of a domain that inhibits transcription activation. This region could either act by directly inhibiting the function of the activation domain, or it could act as a separate transcription repression domain that competes with transcription activation. For instance, in promoter-reporter assays in *Drosophila* S2 cells, the DNA sequence of the promoter determines whether Ubx activates or represses the reporter gene (Winslow *et al.*, 1989; Saffman and Krasnow, 1994). However, a Ubx mutant incapable of transcription activation is able to repress transcription using an activating promoter (Tan *et al.*, 2002). Only JcUbx, TcUbx, and AfUbx have regions capable of inhibiting transcription activation. In TcUbx, this region only encompasses the YPWM motif, homeodomain, and UbdA motif. In JcUbx, the activation inhibition domain also includes sequences N-terminal to the YPWM motif. In AfUbx, the activation inhibition domain lacks the YPWM motif but does include the C-terminus of the protein. Finally, AkUbx, like DmUbx, lacks an activation inhibition domain.

Because activation domains generally lack stable structure (Tan *et al.*, 2002), they are typically categorized based on amino acids enriched in their sequence. Indeed, the nature of the enriched amino acids has been proposed to correlate with mechanism of transcription activation (Treizenberg 1995; Courey and Jia 2001). Table 1 displays the percentages of amino acids that are either enriched in one of the studied proteins, or for which the amount of that amino acid varies substantially. The composition of the activation domains differs significantly between the selected proteins. No two proteins are enriched in the same set of amino acids. Conversely, no amino acid is enriched in all proteins.

Interestingly, the activation inhibition domain does not include the QAQA motif in any protein. The QAQA motif can inhibit transcription repression by DmUbx in cell culture and *in vivo* (Galant and Carroll 2002; Ronshaugen *et al.*, 2002). Our data support the idea

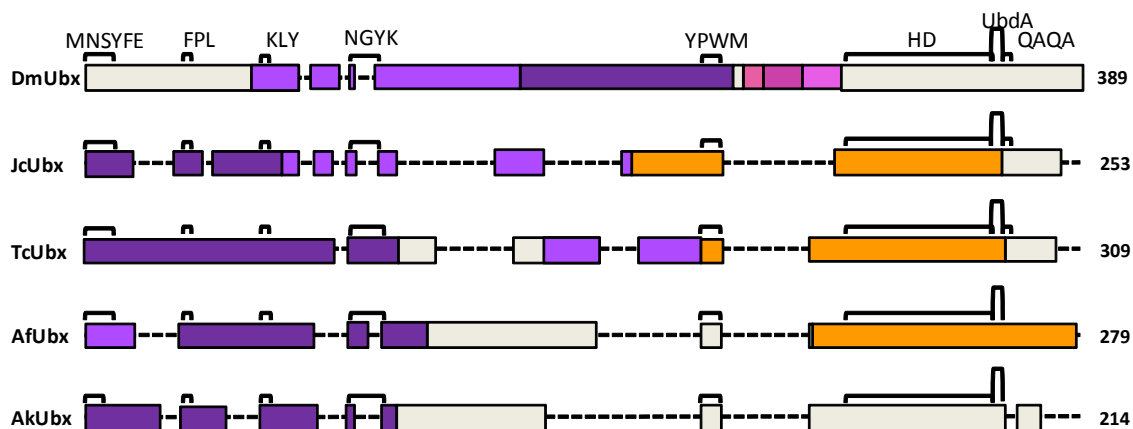


Fig. 8. Schematic comparison of transcription activation domains within the Ubx orthologues. Conserved motifs and the HD are marked with brackets and dashed lines represent missing sequences. Dark purple bars represent activation cores, light purple bars represent activation enhancers, orange bars represent inhibition regions, and gray bars represent sequences that do not regulate transcription activation. Note that the lengths of these proteins differ substantially. HD is identical in length in each of these proteins.

that the QAQA motif represses transcription by directly interacting with co-repressors rather than by inhibiting transcription activation. In contrast, the homeodomain inhibits transcription in other Hox proteins, including Deformed and HoxA7 (Li *et al.*, 1999; Schnabel and Abate-Shen 1996), and is included in the activation inhibition domains of JcUbx, TcUbx, and AfUbx. Thus the transcription repression previously observed by the homeodomain may result, at least in part, from inhibition of transcription activation.

These experiments highlight the critical importance of context in evaluating Hox function. First, as the domain organization of a transcription factor becomes more complex, the role of a region of the protein sequence can change as regions of the protein are added or removed. For example, the activation core of JcUbx is divisible. Amino acids 42-66 behave as a core, enabling transcription activation in the Jc(42-253) variant. However, in Jc(1-66), this same region behaves as only an enhancer, because amino acids 1-41 are separately able to mediate transcription activation. Analysis of domain organization using gene regulation assays *in vivo* is made even more complex by the inclusion of interacting transcription factors. A region of Ubx that appears to activate transcription may only recruit another transcription factor that actually activates transcription. Comparing results from multiple assays can resolve these issues, and reveal the extent to which a single transcription factor can adapt its function to regulate different genes in different tissues.

In many proteins, the rapid evolution of intrinsically disordered sequences combined with the potential for insertions and deletions complicates linking evolutionary changes in function to protein sequence. The presence of conserved motifs in Ubx orthologues allowed subdivision of the activation domain, such that analogous regions of the protein sequence could be compared between Ubx orthologues. Aligning these regions reveals that the role of a single sub-region of the activation domain can vary. For instance, the region between the MNSYFE and the FPL motifs functions as an activation core in JcUbx and AkUbx, an activation enhancer in AfUbx, and has no role in DmUbx. Likewise, the sequence N-terminal to the YPWM motif can be part of the activation core domain (DmUbx), enhancer domain (TcUbx), inhibition domain (JcUbx), or have no effect (AfUbx and AkUbx).

Although the activation domain is always located N-terminal to the homeodomain, comprising up to 60% of the protein sequence, differences in the position, sequence, length and sub-domain organization of the activation domain were observed. Both the inherent capacity of the tested transcription activation domains are different, and the extent to which this capacity is offset by the inhibitor regions or variations in potency of transcription repression domains also varies (Galant and Carroll, 2002; Ronshaugen *et al.*, 2002) also changes. These molecular differences combine to vary the ability of different Ubx orthologues to activate transcription. Evolutionary changes in the sequence and function of the Ubx transcription repression domain have also been reported (Galant and Carroll 2002; Ronshaugen *et al.*, 2002). These changes are surprising, given the general organization of these proteins, e.g., the presence and arrangement of conserved motifs and the homeodomain, is generally preserved. In addition, other Hox proteins appear to be constructed along similar lines, with the activation domain embedded in a long disordered region N-terminal to the homeodomain and YPWM motif (reviewed in Tan *et al.*, 2002). We conclude that Hox transcription regulatory domains can be highly

variable, despite conservation of other aspects of Hox sequence organization and function.

Many mechanisms of transcription factor evolution have been proposed in which domains evolve as discrete modules, including gene duplication & divergence, domain shuffling, and insertions/deletions (indels) (Cheatle Jarvela and Hinman 2015, Dunker *et al.*, 2015; Light *et al.*, 2013; Niklas *et al.*, 2015; Srivastava *et al.*, 2018; Staby *et al.*, 2017, Wagner and Lynch 2008; Nnamani *et al.*, 2016). These mechanisms enable functions to be added to or removed from a transcription factor without catastrophic failure of the protein, thus avoiding negative pleiotropic effects. This work demonstrates the profound impact that such changes, distributed throughout a functional domain, can have on the function of that domain.

Materials and Methods

Ectopic expression of Ubx variants in Drosophila

Genes encoding wild type Ubx, PM4 Ubx, or KM2 Ubx were cloned into the plasmid pUASTattB between the *EcoRI* and *XbaI* restriction sites. ϕ C31 integrase-mediated site-specific transgenesis (Bischof *et al.*, 2007) was used to insert each *ubx* variant gene into three AttP landing sites in the *Drosophila melanogaster* genome: Bloomington Stock lines: 24483 (red eye color, attP position 2R-51D), 24487 (orange eye color, attP position 3R-96E), and 24484 (light orange eye color, attP position 3R-96E) by Best Gene, Inc. These efforts created a total of 9 *Drosophila* lines (3 *ubx* variants x 3 insertion sites), and allowed us to determine whether the DNA insertion site impacted assay results. Each of these lines were crossed with the MS1096-Gal4 line (Bloomington Stock Center) to express the corresponding Ubx variant in third instar larval wing imaginal disc. All crosses to express Ubx used virgin MS1096-Gal4 females and males carrying UAS-Ubx responder inserted into the specified AttP genomic site. The significance of differences in bristle number was assessed by univariate ANOVA.

Yeast one-hybrid assays

DNA sequences for the Ubx orthologues were kindly provided by Dr. Bill McGinnis (UCSD) and Dr. Sean Carroll (University of Wisconsin). These genes were individually inserted into the vector pLexA (Clontech Laboratories, Inc.) between the *EcoRI* site (5' end) and the *BamHI* site (3' end), placing the Ubx orthologue in the correct reading frame relative to the LexA DNA-binding domain.

Yeast were transformed with the pLexA plasmid DNA, encoding the Ubx orthologue as specified, following protocols in the Clontech Yeast Protocols Handbook. In brief, yeast EGY48 strain cells carrying the reporter plasmid p8op-*lacZ* were incubated at 30 °C in liquid YPD, containing 1.0% yeast extract, 2.0% peptone, 2.0% glucose, pH 7.0, and grown to an absorbance at 600 nm of around 1.0 before transformation. SD plates lacking uracil and histidine and containing 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal) were prepared using yeast nitrogen base purchased from Difco, and the pH was not adjusted after autoclaving (Hsiao *et al.*, 2014). Plasmid DNA (500 ng) encoding the LexA alone or the LexA fusion protein was transformed into yeast, plated on SD ura-his- X-gal plates and incubated at 30 °C until visible colonies had formed (roughly 3 days).

Transcription activation of the *lacZ* reporter gene was determined by monitoring β -galactosidase production by blue/white colony assessment. Blue colonies indicate transcription activation by the Ubx fragment in the construct, and the relative shade of blue reflects transcription activation strength. Full-length Ubx_{lb}, which is known to activate transcription in this assay (Tan *et al.*, 2002; Hsiao *et al.*, 2014) and pos-pLexA (Clontech Laboratories, Inc.) were used as positive controls. A Ubx mutant (PM4) which disrupts transcription activation, and empty pLexA without a transcription activation domain served as a negative control. Colony color was assessed by comparison to a standard color series (Fig. 1).

Acknowledgements

The authors thank Sean Carroll and Bill McGinnis for DNAs encoding *Ubx* orthologues. G.G.M. was funded by a scholarship from the Brazilian National Council for Scientific and Technological Development (CNPq).

Competing interests statement

The authors declare no competing financial interests.

Author contributions

Y.L., K.S.M., and S.E.B. conceived the project. S.E.B. designed and Y.L., A.H., Z.M. performed experiments. Y.L., A.H., Z.M., R.M.B., G.G.M., and S.E.B. analyzed data. Y.L., R.M.B., G.G.M., S.E.B. wrote the manuscript.

References

- BISCHOF J, MAEDA R K, HEDIGER M, KARCH F, BASLER K (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific ϕ C31 integrases. *Proc Natl Acad Sci USA* 104: 3312-3317.
- BONDOS S E, CATANESE D J Jr, TAN X-X, BICKNELL A, LI L, MATTHEWS K S (2004). Hox transcription factor Ultrabithorax I β physically and genetically interacts with Disconnected Interacting Protein 1, a double-stranded RNA-binding protein. *J Biol Chem* 279: 26433-26444.
- BONDOS S E, SWINT-KRUSE L, MATTHEWS K S (2015). Flexibility and disorder in gene regulation: LacI/GalR and Hox proteins. *J Biol Chem* 290: 24669-24677.
- BONDOS S E, TAN X-X, MATTHEWS K S (2006). Physical and genetic interactions link Hox function with diverse transcription factors and cell signaling proteins. *Mol Cell Proteomics* 5: 824-834.
- BROWN C J, TAKAYAMA S, CAMPEN A M, VISE P, MARSHALL T W, OLDFIELD C J, WILLIAMS C J, DUNKER A K (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol* 55: 104-110.
- CHEATLE JARVELA A M, HINMAN V F (2015). Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. *EvoDevo* 6: 3.
- CHESMORE K N, BARTLETT J, CHENG C, WILLIAMS S M (2016). Complex patterns of association between pleiotropy and transcription factor evolution. *Genome Biol Evol* 8: 3159-3170.
- COUREY A J, JIA S (2001). Transcriptional repression: the long and the short of it. *Genes Dev* 15: 2786-2796.
- DUFFY J B (2002). GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife. *Genesis* 34: 1-15.
- DUNKERAK, BONDOS S E, HUANG F, OLDFIELD C J (2015). Intrinsically disordered proteins and multicellular organisms. *Sem Cell Dev Biol* 37: 44-45.
- FISH A, CHEN L, CAPRA J A (2015). Gene regulatory enhancers with evolutionarily conserved activity are more pleiotropic than those with species-specific activity. *Genome Biol Evol* 9: 2615-2625.
- GALANT R, CARROLL S B (2002). Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* 415: 911-913.
- GALANT R, WALSH C W, CARROLL S B (2002). Hox repression of a target gene: Extradenticle-independent, additive action through multiple monomer binding sites. *Development* 129: 3115-3126.
- GRENIER J K, CARROLL S B (2000). Functional evolution of the Ultrabithorax protein. *Proc Natl Acad Sci USA* 97: 704-709.
- GROTH A C, FISH M, NUSSE R, CALOS M P (2004). Construction of transgenic *Drosophila* by using the site-specific integrase from Phase ϕ C31. *Genetics* 116: 1775-1782.
- HSIAO H-C, GONZALEZ K L, CATANESE D J JR., JORDY K E, MATTHEWS K S, BONDOS S E (2014). The intrinsically disordered regions of the *Drosophila melanogaster* Hox protein Ultrabithorax select interacting proteins based on partner topology. *PLoS ONE* 9: e108217.
- JOHNSON F B, PARKER E, KRASNOW M A (1995). Extradenticle protein is a selective cofactor for the *Drosophila* homeotics: Role of the homeodomain and YPWM amino acid motif in the interaction. *Proc. Natl. Acad. Sci. USA* 92: 739-743.
- LI X, MURRE C, MCGINNIS W (1999). Activity regulation of a Hox protein and a role for the homeodomain in inhibiting transcriptional activation. *EMBO J* 18: 198-211.
- LIGHT S, SAGIT R, SACHENKOVA O, EKMAN D, ELOFSSON A (2013). Protein expansion is primarily due to indels in intrinsically disordered regions. *Mol. Biol. Evol.* 30: 2645-2653.
- LIU Y, MATTHEWS K S, BONDOS S E (2008). Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the *Drosophila* Hox protein Ultrabithorax. *J Biol Chem* 283: 20874-20887.
- LIU Y, MERCHANT Z, HSAIO H-C, GONZALEZ K L, MATTHEWS K S, BONDOS S E (2011). Media composition influences yeast one- and two-hybrid results. *Biol Proced Online* 13: 6.
- LIU J, PERUMAL N B, OLDFIELD C J, SU E W, UVERSKY V N, DUNKER A K (2006). Intrinsic disorder in transcription factors. *Biochemistry* 45: 6873-6888.
- LONFAT N, MONTAVON T, DARBELLAY F, GITTO S, DUBOULE D (2015). Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science* 346: 1004-1006.
- LUO S D, BAKER B S (2015). Constraints on the evolution of a doublesex target gene arising from doublesex's pleiotropic deployment. *Proc Natl Acad Sci USA* 112: E852-E861.
- NIKLAS K J, BONDOS S E, DUNKER A K, NEWMAN S A (2015). Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modification. *Front Cell Dev Biol* 3: 8.
- NNAMANI M C, GANGULY S, ERKENBRACK E M, LYNCH V L, MIZOUE L S, TONG Y, DARLING H L, FUXREITER M, MEILER J, WAGNER G P (2016). A derived allosteric switch underlies the evolution of conditional cooperativity between HOXA11 and FOXO1. *Cell Reports* 15: 2097-2108.
- PAVLOPOULOS A, AKAM M (2006). Hox gene Ultrabithorax regulates distinct sets of target genes at successive stages of *Drosophila* haltere morphogenesis. *Proc. Natl. Acad. Sci. USA* 108: 2855-2860.
- RONSHAUGEN M, MCGINNIS N, MCGINNIS W (2002). Hox protein mutation and macroevolution of the insect body plan. *Nature* 415: 914-917.
- SAFMAN E E, KRASNOW M A (1994). A differential response element for the homeotic Antennapedia P1 promoter of *Drosophila*. *Proc Natl Acad Sci USA* 91: 7420-7424.
- SCHNABEL C A, ABATE-SHEN C (1996). Repression by HoxA7 is mediated by the homeodomain and the modulatory action of its N-terminal-arm residues. *Mol Cell Biol* 16: 2678-2688.
- SIVANANTHARAJAH L, PERCIVAL-SMITH A (2015). Differential pleiotropy and HOX functional organization. *Dev Biol* 398: 1-10.
- SRIVASTAVA A, KUMAR A S, MISHRA R K (2018). Vertebrate GAF/ThPOK: emerging functions in chromatin architecture and transcriptional regulation. *Cell Life Sci* 75: 623-633.
- STABY L, O'SHEA C, WILLEMOËS M, THEISEN F, KRAGELUND B B, SKRIVER K (2017). Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem J* 474: 2509-2532.
- TAN X-X, BONDOS S, LI L, MATTHEWS K S (2002). Transcription activity by Ultrabithorax I β protein requires a predicted α -helical region. *Biochemistry* 41: 2774-2785.
- TOUR A, HITTINGER C T, MCGINNIS W (2005). Evolutionarily conserved domains required for activation and repression functions of the *Drosophila* Hox protein Ultrabithorax. *Development* 132: 5271-5281.
- TOURASSE N J, LI W-H (2000). Selective constraints, amino acid composition, and the rate of protein evolution. *Mol Biol Evol* 17: 656-664.
- TRIEZENBERG S J (1995). Structure and function of transcriptional activation domains. *Curr Opin Genet Dev* 5: 190-196.
- VILLAR D, FLICEK P, ODOM D T (2014). Evolution of transcription factor binding in metazoans—mechanisms and functional implications. *Nat Rev Genet* 15: 221-233.
- WAGNER G P, LYNCH V J (2008). The gene regulatory logic of transcription factor evolution. *Trends Ecol Evol* 23: 377-385.
- WARD J J, SODHI J S, MCGUFFIN L J, BUXTON B F, JONES D T (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337: 635-645.
- WINSLOW G M, HAYASHI S, KRASNOW M, HOGNESS D S, SCOTT M P (1989). Transcriptional activation by the *Antennapedia* and *fushi tarazu* proteins in cultured *Drosophila* cells. *Cell* 57: 1017-1030.
- YANG B, WITTKOPP P J (2017). Structure of the transcriptional regulatory network correlates with regulatory divergence in *Drosophila*. *Mol Biol Evol* 34: 1351-1362.

Further Related Reading, published previously in the *Int. J. Dev. Biol.*

Gdf11/Smad signalling and Cdx proteins cooperate to activate the Hoxc8 early enhancer in HepG2 cells

Stephen J. Gaunt
Int. J. Dev. Biol. (2017) 61: 427-432
<https://doi.org/10.1387/ijdb.170066sg>

ADHFe1: a novel enzyme involved in retinoic acid-dependent Hox activation

Yehuda Shabtai, Natalie Shukrun and Abraham Fainsod
Int. J. Dev. Biol. (2017) 61: 303-310
<https://doi.org/10.1387/ijdb.160252af>

The significance of Hox gene collinearity

Stephen J. Gaunt
Int. J. Dev. Biol. (2015) 59: 159-170
<https://doi.org/10.1387/ijdb.150223sg>

Synergistic action in P19 pluripotent cells of retinoic acid and Wnt3a on Cdx1 enhancer elements

Stephen J. Gaunt and Yu-Lee Paul
Int. J. Dev. Biol. (2014) 58: 307-314
<https://doi.org/10.1387/ijdb.140003sg>

The Parahox gene Pdx1 is required to maintain positional identity in the adult foregut

Andrew M. Holland, Sonia Garcia, Gaetano Naselli, Raymond J. MacDonald and Leonard C. Harrison
Int. J. Dev. Biol. (2013) 57: 391-398
<https://doi.org/10.1387/ijdb.120048ah>

Origins of Cdx1 regulatory elements suggest roles in vertebrate evolution

Stephen J. Gaunt and Yu-Lee Paul
Int. J. Dev. Biol. (2011) 55: 93-98
<https://doi.org/10.1387/ijdb.103252sg>

Direct control of Hoxd1 and Irx3 expression by Wnt/beta-catenin signaling during anteroposterior patterning of the neural axis in *Xenopus*

Sylvie Janssens, Tinneke Denayer, Tom Deroo, Frans Van Roy and Kris Vleminckx
Int. J. Dev. Biol. (2010) 54: 1435-1442
<https://doi.org/10.1387/ijdb.092985sj>



5 yr ISI Impact Factor (2016) = 2.421

