Published in final edited form as:

J Mol Biol. 2009 July 24; 390(4): 760–774. doi:10.1016/j.jmb.2009.05.059.

INTERNAL REGULATORY INTERACTIONS DETERMINE DNA BINDING SPECIFICITY BY A HOX TRANSCRIPTION FACTOR

Ying Liu¹, Kathleen S. Matthews¹, and Sarah E. Bondos^{1,2}

¹Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77005

²Department of Molecular and Cellular Medicine, Texas A&M Health Science Center, College Station, TX 77843-1114

Summary

In developing bilaterans, the Hox transcription factor family regulates batteries of downstream genes to diversify serially repeated units. Given Hox homeodomains bind a wider array of DNA binding sites *in vitro* than are regulated by the full-length protein *in vivo*, regions outside the homeodomain must aid DNA site selection. Indeed, we find affinity for disparate DNA sequences varies less than 3-fold for the homeodomain isolated from the *Drosophila* Hox protein Ultrabithorax Ia (UbxHD), whereas for the full-length protein (UbxIa) affinity differs by more than 10-fold. The rank order of preferred DNA sequences also differs, further demonstrating distinct DNA binding preferences. The increased specificity of UbxIa can be partially attributed to the I1 region, which lies adjacent to the homeodomain and directly impacts binding energetics. Each of three segments within I1 – the Extradenticle-binding YPWM motif, the 6 amino acids immediately N-terminal to this motif, and the 8 amino acids abutting the YPWM C-terminus – uniquely contribute to DNA specificity. Combination of these regions synergistically modifies DNA binding to further enhance specificity. Intriguingly, the presence of the YPWM motif in UbxIa inhibits DNA binding only to Ubx•Extradenticle heterodimer binding sites, potentially functioning *in vivo* to prevent Ubx monomers from binding and misregulating heterodimer target genes. However, removal of the surrounding region allows the YPWM motif to also inhibit binding to Hox-only recognition sequences. Despite a modular domain design for Hox proteins, these results suggest that multiple Hox protein regions form a network of regulatory interactions that coordinate context- and gene-specific responses. Since most non-homeodomain regions are not conserved between Hox family members, these regulatory interactions have the potential to diversify binding by the highly homologous Hox homeodomains.

Keywords

Hox; DNA binding; homeodomain; YPWM; hexapeptide motif

© 2009 Elsevier Ltd. All rights reserved.

Address correspondence to: Sarah E. Bondos, Ph.D., 258 Reynolds Medical Building, Texas A&M Health Science Center, College Station, TX 77843-1114. Fax: 979-847-9481; SEBondos@medicine.tamhsc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

*This work was supported by grants from the Robert A. Welch Foundation (C-576) and the National Institutes of Health (GM22441) to K.S.M. We thank the members of the Matthews and Silberg laboratories at Rice University for helpful comments and suggestions. We also thank Joseph Mire (Texas A&M University) for assistance modeling the disordered YPWM-HD linker region. The UbxIa gene and the Exd and Hth constructs were the gifts of Dr. Ella Tour and Dr. Bill McGinnis, both from UCSD. The UbxHD expression vector was the gift of Dr. Philip Beachy from Johns Hopkins University.

INTRODUCTION

A central challenge in developmental biology is to link events in patterning and morphogenesis to the gene regulatory hierarchy that drives these processes. Because only a few proteins regulate tissue-specific gene transcription relative to the number of unique tissues that must be generated, each transcription factor active in development must uniquely function in multiple cellular contexts. Thus, this challenge is doubly complex, requiring identification of crucial genes and cis-acting regulatory sequences as well as elucidation of mechanisms that govern context-specific transcription factor function.

The Hox transcription factor family acts near the top of this regulatory hierarchy in all bilaterally symmetric animals. Each Hox protein specifies multiple structures – directing appendage and organ development as well as determining local attributes of tissues in all three germ layers.^{1–3} As transcription factors, the crucial contributions of Hox proteins are based on their ability to target different promoters in a tissue-specific manner.^{4–6} However, the DNA-binding specificity of Hox-DNA interactions is anticipated to be insufficient to distinguish cognate and non-cognate targets in even a single tissue.^{7–9} Consequently, we and others have postulated that regions of Hox proteins outside the homeodomain must influence DNA binding to generate the necessary level of sequence selection.^{7–11} Furthermore, modulation of non-homeodomain regulatory regions through post-translational modifications,¹² protein interactions,^{2, 13–15} or alternative splicing¹⁶ would provide an opportunity for tissue-specific regulation of target gene selection.

Unfortunately, identification of non-homeodomain regions capable of modifying binding is a long-standing challenge, hampered both by the predominance of intrinsic disorder outside the homeodomain¹⁰ and the tendency of full-length Hox proteins to aggregate.¹⁷ As a result, DNA binding studies have frequently been limited to the use of deletion mutants or unpurified proteins produced *in vitro*, preventing quantitative analysis of function in the native protein.^{11, 18} We have recently devised protocols to overcome these barriers for the full-length *Drosophila melanogaster* Hox protein Ultrabithorax Ia (UbxIa),^{10, 17} the dominant alternative splicing isoform expressed in the *Drosophila* embryo.¹⁶ By measuring equilibrium DNA binding by purified UbxIa of high activity under conditions where the DNA concentration is more than an order of magnitude less than the dissociation constant¹⁹ and excess non-specific DNA sequences are absent, we can accurately assess the DNA binding affinity of wild-type and mutant UbxIa, independent of the influence of other factors or competition with non-specific DNA interactions. Using this approach, we have discovered that most of the UbxIa protein sequence can influence the *affinity* of the homeodomain for the optimal Ubx DNA binding site.¹⁰ Both prediction algorithms and native state proteolysis experiments demonstrated that most of the sequences that modulate DNA-binding affinity in UbxIa are intrinsically disordered.¹⁰

Herein, we demonstrate directly for the first time that regions outside the homeodomain of a Hox protein also impact DNA binding *specificity* – altering affinity in a DNA sequence-dependent manner. DNA binding by full-length UbxIa and the Ubx homeodomain (UbxHD) was monitored for a variety of DNA sequences. Sites examined included the UbxHD optimal binding site²⁰ and representative Ubx genomic binding targets selected to vary a wide range of factors: binding site sequence, number and density, the presence of binding sites for the Hox-interacting protein Extradenticle (Exd), the transcriptional outcome of binding (activation versus repression), and the developmental stage and tissue in which the enhancer is active.^{18, 20–24} Although UbxHD binds tightly to all examined sequences, the affinity of full-length UbxIa varies by more than an order of magnitude with DNA sequence, indicating regions outside the homeodomain do improve specificity. We next sought to identify specificity

determinants – regions that either impact binding only to a subset of DNA sequences or have opposite effects on binding different DNA sequences. We focused on the I1 affinity-modulating region of Ubx,¹⁰ which contains an 18 amino acid sequence required for repression of *distalless* by Ubx *in vivo*.²⁵ We find removal of this sequence alters binding affinity in a DNA-sequence specific manner, demonstrating that the 18 amino acid region acts as a specificity determinant. We next divided this region into three sections: the N-terminal 6 amino acids, the central 4 amino acid Exd interaction motif (YPWM), and the intrinsically disordered C-terminal 8 amino acids, which contain a portion of the mI alternatively spliced microexon.^{10,14–16, 26–29} Mutation or deletion of each of these three sections has a differing impact on DNA specificity. Furthermore, combination of these regions produces new DNA binding specificity patterns. We propose the I1 region may act as a cellular “antenna”, integrating tissue-specific information provided by alternative splicing and protein interactions to direct UbxIa to the appropriate DNA binding sites.

RESULTS

Full-length UbxIa binds DNA with higher selectivity than UbxHD

Previously, we demonstrated that the full-length UbxIa splicing isoform (Fig. 1a) and its homeodomain have significantly different affinities and responses to pH when binding an optimal DNA sequence.¹⁰ Strikingly, most non-homeodomain regions of Ubx can modulate Ubx•DNA binding *affinity* (Fig. 1e), utilizing a variety of mechanisms.¹⁰ To determine whether non-homeodomain sequences also enhance Hox•DNA binding *specificity*, we compared the range of affinities observed for UbxIa and UbxHD binding to a variety of DNA sequences, including the UbxHD optimal binding site, termed 40AB,²⁰ and five representative Ubx genomic binding targets: Dll, Dpp, UA, A1, and Sal (Table 1• Fig. 2).^{18, 20–24} DNA binding sequences were selected to maximize the variety of (i) base sequences adjacent to the 5'-TAAT-3' consensus site or sites, (ii) binding site densities of the promoters regulated by Ubx *in vivo*, (iii) requirements for the Hox interacting protein Exd, (iv) transcriptional outcomes (activation versus repression), and (v) the cellular contexts in which Ubx binds the natural sequences, including the developmental stage, germ layer, and tissue. To facilitate analysis, small (< 90 bp) sections of these complex cis-acting regulatory regions were used for DNA binding experiments.

Equilibrium mobility shift analysis experiments (gel shifts) were used to assess DNA binding at protein concentrations near the dissociation constant, where binding is most sensitive to DNA sequence, and at DNA concentrations well below this range. Gel shifts allow separation and observation of complexes with different numbers of bound proteins. Equilibrium dissociation constants are reported as K_d for single DNA sites and as the apparent equilibrium dissociation constant (K_{app}) for DNA sequences with multiple binding sites. K_{app} reflects not only DNA affinity but also the availability of multiple binding sites and the additional energy provided by protein•protein interactions upon the creation of higher order complexes. However, binding by a single protein dominates at the point in the binding curve where the concentration of UbxIa is equal to K_{app} , although at higher concentrations multiple binding events are observed (Fig. 3a, b and Supplementary Materials Fig. 1).

Because binding experiments utilizing protein in crude cell extracts are not quantitative, and even very short truncations can dramatically impact the DNA binding properties of Ubx,¹⁰ our experiments were based on purified, full-length proteins. For each protein preparation, the purity, solubility, and activity were determined to permit confident assessment of small changes in DNA binding affinity. Wild-type and variant Ubx proteins are all untagged to preclude the possibility of a tag the protein's structure or function. The UbxIa splicing isoform was selected for these experiments because it is the most abundant isoform expressed during fly

development.^{16, 30} UbxIa contains all possible amino acid sequences except the 9 amino acid “b element” microexon, located between the YPWM motif and the mI microexon (Fig. 1).²⁸

We first compared the affinity of full-length UbxIa and its isolated homeodomain for the selected DNA binding targets (Fig. 3d–j, Table 2). UbxHD binds all target sequences with high affinity, varying only ~3-fold. Although this low level of sequence discrimination by UbxHD may contribute to target gene selection, it is generally believed insufficient to drive the level of reliable binding site discrimination required for *in vivo* function.^{7–9} Indeed, UbxHD binds equally well to its optimal site, 40AB, and the DII DNA, even though DII binding by Ubx *in vivo* requires interaction with the Exd protein.^{4,24,26,29} Full-length UbxIa affinity was lower than that of UbxHD for all DNAs investigated (Fig. 3d–j, Table 2). This lower affinity (higher K_d or K_{app}) is not due to an inability of multiple full-length UbxIa proteins to bind multi-site DNA sequences, since higher order complexes were also observed for the full-length protein at higher concentrations (Fig. 3c). Therefore, a net inhibition of binding by non-homeodomain regions, previously observed for the UbxHD optimal DNA binding sequence,¹⁰ may be a general feature of Ubx•DNA interactions.

In contrast to UbxHD, UbxIa binds these DNAs with affinities spanning more than an order of magnitude, demonstrating regions outside the homeodomain alter affinity in a manner dependent on DNA sequence. In fact, the ratio of UbxIa affinity to UbxHD affinity (Protein sensitivity^{FL-HD}, Table 2) varies ~10-fold with DNA sequence. Thus, non-homeodomain regions of UbxIa impact binding to some DNA sequences more than others. Significantly, the order of DNA sequences, when ranked by affinity, is also different for UbxHD and UbxIa. Therefore, the role of non-homeodomain sequences is more complex than to simply amplify pre-existing UbxHD sequence preferences.

An 18 amino acid sequence within the I1 region impacts DNA specificity

Our next step was to explore the non-homeodomain regions that modulate DNA binding specificity observed for full-length UbxIa compared to UbxHD. We have previously shown that the I1 region (Fig. 1e) reduces HD•DNA affinity ~2-fold.¹⁰ At the N-terminus of the I1 region is the YPWM motif, which forms a portion of the Exd interaction interface (Fig. 1b–e).^{15, 26–29} Much of the middle of the I1 region – corresponding to the alternatively spliced microexons¹⁶ – is predicted to be intrinsically disordered, consistent with its being hyper-susceptible to native state proteolysis.¹⁰ The C-terminus of the I1 region is also located adjacent to the N-terminal arm of UbxHD, which forms base-specific contacts with the minor groove.^{27,31–33} This location may allow the I1 region to impact DNA specificity, potentially in response to Exd interaction or alternative splicing. Indeed, deletion of 18 amino acids within I1, creating Ubx-18 (Fig. 1e), partially relieves the DNA binding inhibition caused by I1.¹⁰ Ubx-18 is also less able to regulate *distalless* transcription *in vivo*.²⁵ For these reasons, our search for determinants of DNA binding specificity initially utilized the Ubx-18 mutant. Although no effect of the mutation was observed for binding 40AB, Dpp, or UA, Ubx-18 binds roughly 2- to 3-fold better to DII, A1, and Sal DNA (Table 2, Fig 4a, UbxIa versus Ubx-18). Therefore, this 18 amino acid region impacts binding in a DNA sequence-specific manner. Indeed, the contribution of this region to specificity in the full-length protein is comparable to the level of sequence selection inherent to the homeodomain (~2.8 fold).

Internal deletions have the potential to induce structural rearrangements, which may, in turn, cause other protein regions to behave in a non-native manner. This possibility was particularly a concern for the Ubx-18 mutant, in which the type 1 reverse turn formed by the YPWM motif is removed.²⁷ However, this region is bounded by an intrinsically disordered region on the C-terminal side.¹⁰ A second large disordered region, just beyond the predicted activation domain α -helix,³⁴ is located N-terminal to the 18 amino acid deleted region.¹⁰ These disordered regions are likely to provide sufficient flexibility to mitigate any structural rearrangements due to

mutation. Indeed, Ubx-18 is both soluble and active *in vitro*¹⁰ and *in vivo*.²⁵ The circular dichroism spectra of UbxIa and Ubx-18 overlay (Fig. 4c), indicating that the 18 amino acid deletion has not caused any global structural rearrangements. Furthermore, affinity and activity are only marginally affected when binding the optimal DNA sequence, 40AB. Finally, both UbxIa and Ubx-18 cooperatively bind Dll DNA in conjunction with full-length Exd and Homothorax (Hth) (Fig. 5). Cooperative binding to this site is mediated by an alternate Exd interaction motif C-terminal to the homeodomain.²⁹ Therefore, the C-terminus of UbxIa is also unaffected by the 18 amino acid deletion (Fig. 5). Together, these data demonstrate UbxIa structure is not greatly altered due to internal deletion in this region. However, local structural perturbations beyond the detection limits of these assays could still be caused by the deletion.

The mechanism by which the 18 amino acid region impacts DNA specificity may involve intramolecular protein interactions (*e.g.*, with the homeodomain) as well as intermolecular Ubx•Ubx interactions on multi-site DNA sequences. To determine whether intermolecular interactions contribute to changes in DNA binding, we generated a new DNA oligonucleotide, Sal6, which only contains the middle of the 3 binding sites on the Sal DNA (Fig. 2, Table 1). UbxIa binds Sal and Sal6 with similar affinity, suggesting that single site binding dominates the UbxIa interaction with Sal (Table 3). However, Ubx-18 binds Sal6 with lower affinity than Sal. Consequently, Ubx-18 may prefer to interact with a different site on the Sal DNA, or Ubx•Ubx interactions may have a larger impact on Ubx-18 binding to Sal. In support of this second hypothesis, the affinity of Ubx-18 for Sal is higher than for the UbxIa•Sal interaction.

Dissecting contributions within the -18 deleted region

The amino acids deleted in Ubx-18 include a partially conserved region at the N-terminus of the deletion, the highly conserved YPWM Exd-interaction motif, and the less conserved 8 amino acids C-terminal to the YPWM motif, which includes a portion of the mI microexon (Fig. 1d). To parse the effects of these three regions, each was abrogated in turn by mutation. The role of the hexapeptide motif was initially assessed by creating point mutants within the YPWM sequence (Fig. 1e). In the first mutant, the YPWM sequence was changed to YAAA, a version whose effects *in vivo* have previously been assessed.¹⁸ Unfortunately, YAAA aggregates when overexpressed in *E. coli* and could not be purified. These mutations are predicted to replace a reverse turn within an α -helix, thus significantly impacting global protein structure. In a second variant, the YPWM sequence was mutated to GPGG, eliminating intra-protein contacts made by this motif while simultaneously avoiding potential structural rearrangements. Indeed, the solubility and activity are similar to wild-type UbxIa. Furthermore, the circular dichroism spectra of UbxIa and GPGG are similar, although the ellipticity of GPGG is less negative below 205 nm, indicating a somewhat lower content of intrinsic disorder (Supplementary Materials Fig. 2).³⁵ GPGG binds within error limits with an affinity comparable to wild-type UbxIa for all tested DNA sequences excepting Dll, which is bound with ~3-fold higher affinity than UbxIa (Table 2, Fig. 4a). Therefore, the YPWM motif *only* inhibits binding by UbxIa to this composite DNA site.

To determine whether the YPWM motif is capable of inhibiting binding to other composite DNA sites, we measured binding by Ubx, Ubx-18, and GPGG to two additional composite DNA sequences. The Dppe4 sequence contains the e4 Hox•Exd binding site from the 812 bp enhancer used by Ubx to activate *dpp* expression in parasegment 7 of the embryonic mesoderm.³⁶ In Dppe4, the Exd binding site is farther away from the TAAT Ubx-binding site than in the Dll sequence (Table 1). The second DNA site tested is the murine HoxB1 enhancer “repeat 3” sequence that can be used by the Labial•Exd complex to repress reporter gene expression.²⁶ This sequence is similar to the consensus Hox•Exd site in which the Hox and Exd sites overlap, but the TAAT sequence is absent.⁶ For both DNA sequences, UbxIa bound with low affinity similar to Dll, and removal of the YPWM sequence either by deletion or by point mutations

improved DNA binding affinity (Fig. 4b, Table 3). Therefore, the YPWM motif inhibits binding to all composite Hox•Exd DNA binding sites tested, despite variations in binding site sequence and spacing.

Whereas the Ubx-18 and GPGG mutations similarly impact binding to composite sites, only Ubx-18 alters binding to the A1 and Sal Ubx multimer binding sites (Fig. 4a). Consequently, the amino acids flanking the YPWM motif must also contribute to specificity. To test this hypothesis, the N-terminal and C-terminal flanking regions of the YPWM motif were individually removed to form the Ubx-6 and Ubx-8 mutants, respectively (Fig. 1e). The circular dichroism spectra, solubility, and activity of these mutants confirm that the overall structure of the protein was not significantly altered by these deletions (Supplementary Materials Fig. 3). However, small differences in the CD spectra below 210 nm may reflect changes in the disorder content. Both mutants impact DNA binding specificity, although they have different effects. Relative to wild-type Ubx1a, Ubx-6 binds with somewhat higher affinity only to Dll (Table 2, Fig. 6a), a pattern that is similar to (but significantly weaker than) that observed for the GPGG mutant. Therefore, the role of these 6 amino acids may be to enhance the regulatory role of the YPWM motif in preventing monomer association with Ubx•Exd composite binding sites. In contrast, Ubx-8 binds less well to 40AB and Dpp, yet better to Dll (Table 2, Fig. 6a, Ubx1a versus Ubx-8). The ability of the C-terminal 8 amino acids to drive both positive and negative effects upon binding different DNA sequences suggests this region can impact DNA interactions via multiple mechanisms.

Given the YPWM motif and 6 amino acids N-terminal to this motif both alter DNA binding specificity, removing the 8 amino acids C-terminal to the YPWM motif could change specificity either by truncating the distance between the homeodomain and the YPWM motif and / or by independently acting as a specificity determinant. To distinguish these mechanisms, we generated a new mutant, termed Ubx8scramble, which rearranges the 8 amino acids C-terminal to the YPWM motif in such a way as to maintain a similar disorder and hydrophobicity profile (Fig. 1e, Fig. 6c,d).^{37,38} Binding of Ubx8scramble was measured for all DNA sequences in which Ubx-8 was different than Ubx1a. For each of these sequences, binding of Ubx8scramble closely matched that of Ubx-8 (Fig. 6b, Table 3). We conclude that the impact of the 8 amino acid region includes sequence-specific effects. Furthermore, the flexibility of this region appears to prevent major local disruptions from internal deletions.

The YPWM motif and its flanking regions interact in a complex manner to enhance DNA binding specificity

Surprisingly, neither mutation of the YPWM motif (Fig. 4a, Ubx1a versus GPGG), nor removal of either flanking region affects binding to A1 and Sal (Fig. 6a, Ubx1a versus Ubx-6 and Ubx-8), even though simultaneous removal of all three regions does alter affinity for these DNA sequences (Fig. 4a, Ubx1a versus Ubx-18). These results suggest these three segments also impact each other in response to the DNA sequence to create non-additive effects. To test this hypothesis, a third internal deletion mutant, Ubx-14, was constructed which removes both the N-terminal 6 amino acids and C-terminal 8 amino acids (Fig. 1e). Similar to the other mutants, simultaneous deletion of both regions did not impact the solubility or activity of the proteins. Circular dichroism experiments revealed only some loss of intrinsic disorder, similar to that observed for the GPGG mutant (Supplementary Materials Fig. 4). Both Ubx-8 and Ubx-14 bind more weakly than Ubx1a to 40AB and Dpp but more tightly to Dll. However, these effects are more pronounced in the Ubx-14 mutant (Fig. 6a,e).

The DNA binding specificity of a given Ubx1a variant was estimated by the variation in binding affinity for our six sample DNA sequences ("Fold variation", Table 2). UbxHD, Ubx-8, and Ubx-14 are much less sensitive to changes in the DNA sequence (affinities for each vary only ~ 2-fold) relative to Ubx-6 (~8-fold variation), or the GPGG mutant (~6-fold variation). The

wild-type UbxIa protein was best able to discriminate between the tested DNA sequences (~12-fold variation).

The YPWM motif and its surrounding regions combine to increase DNA binding specificity. The improvement in Dll binding caused by the 14 amino acid deletion (Fig. 6e, UbxIa versus Ubx-14) was similar in magnitude to the GPGG point mutations (Fig. 4a, UbxIa versus GPGG), but weaker than that observed upon removal of the entire region (Fig. 6e, UbxIa versus Ubx-18). Therefore, the YPWM motif and its flanking sequences likely work together to inhibit monomer binding to Hox•Exd composite sites. A more pronounced effect was observed for binding to A1 and Sal DNAs. The GPGG mutation and the 6, 8, and 14 amino acid deletions either inhibit or do not alter binding to Sal and A1 (Fig. 4a, Fig 6a). However, binding is significantly enhanced when all three regions are removed in the Ubx-18 mutant (Fig. 6e).

The underlying cause can be elucidated by comparing the impact of the YPWM motif in different Ubx sequence backgrounds (Fig. 4a, Fig. 6e). In the context of the full-length protein, mutation of YPWM to GPGG inhibits DNA binding only to Dll DNA (Fig. 4a). However, comparison of the Ubx-14 and Ubx-18 mutants reveals the YPWM motif, present only in Ubx-14, inhibits binding to every DNA sequence tested – even the optimal Ubx binding site – when the surrounding 14 amino acids are removed (Fig. 6e). Therefore, the region surrounding the YPWM motif controls the magnitude and sequence dependence of DNA inhibition by this motif. The inhibition of DNA binding by the YPWM motif in the Ubx-14 mutant appears to decrease variation in DNA binding affinity, and thus decrease DNA binding specificity, to levels even lower than observed for UbxHD (Table 2, Compare Fold variation for Ubx-14 and UbxHD).

Together, we find that each mutation or deletion in the 18 amino acid region impacts Ubx•DNA binding in a unique, sequence-dependent manner. The impact of this 18 amino acid region varies substantially with DNA sequence (*e.g.*, Protein sensitivity^{FL-18}, Table 2). The strongest effect was observed for binding Dpp (7-fold), whereas the UA and A1 DNA sequences are much less sensitive to changes in this region. We conclude multiple subregions of the I1 inhibiting segment form a network of regulatory interactions with each other and the DNA sequence to direct different aspects of specific binding.

DISCUSSION

Design of Hox•DNA binding experiments

The crucial role of Hox proteins in animal development, wound healing, stem cell differentiation, and carcinogenesis^{39–40} necessitates an understanding of Hox function, and in particular Hox•DNA binding, at the molecular level. Our results highlight two problems commonly encountered in assessing Hox•DNA interactions. First, Hox truncation mutants, in which the predominately disordered sequences are removed, are commonly used to avert aggregation.^{11, 15} However, we demonstrate that sequences outside the homeodomain can have dramatic effects on Hox•DNA binding affinity¹⁰ and specificity, requiring cautious interpretation of studies relying solely on truncation mutants. In particular, removal of only amino acids 2–19 from the N-terminus of UbxIa is sufficient to impact binding.¹⁰ Since the N-terminus is moderately conserved in both Ubx orthologues and Hox paralogues,^{25,41,42} even minor N-terminal truncations may alter binding in any Hox protein.

The second problem involves using an optimal DNA binding site to characterize transcription factor binding. Such sites are extremely useful for identifying potential genomic binding sites and elucidating differences between related transcription factors.^{20,43,44} However, the optimal DNA binding sequence masks the impact of non-homeodomain regions that alter binding to natural DNA targets (Protein sensitivity^{FL-HD}, Table 2). *In vivo*, these regions may regulate

binding through cooperative binding or cofactor interactions.^{21,31} Therefore, regulatory mechanisms potentially utilized *in vivo* may be overlooked when binding studies exclusively utilize the optimal DNA sequence. We have therefore examined an array of Ubx target sequences in analyzing the impact of the YPWM region.

In vivo roles for DNA specificity-regulating regions

Our search for specificity determinants outside the UbxIa homeodomain focused on an 18 amino acid sequence within the II region, which inhibits UbxIa•DNA binding.¹⁰ We find all mutations involving this region reduced the DNA binding specificity of UbxIa (Table 2, Fold variation). *In vivo*, Ubx-mediated repression of the *dll* and *antp* promoters is impaired by deletion of this region.²⁵ The increase in binding by Ubx-18 to the Dll Hox•Exd composite DNA may permit binding and misregulation of *dll* by Ubx monomers. Since this deletion does not appear to alter DNA binding to the A1 portion of the large *antp* promoter,²⁵ changes in Ubx-18 interactions with other Hox binding site clusters may alter *antp* expression (Fig. 2). Alternately, mutations in this region may also impact other aspects of UbxIa function such as protein interactions or transcription repression.^{13–15, 25, 26} Since this region participates in many molecular functions, it is not possible to unambiguously ascribe any single molecular event as the direct cause of a phenotype *in vivo*. However, in such cases *in vitro* studies can create a list of candidate functions that may, alone or in combination, contribute to a particular phenotype. Intriguingly, participation of a single protein region in multiple molecular events creates the potential for one interaction to regulate other molecular functions. Indeed, this scenario has already been suggested for the analogous region of a truncated Labial mutant: inhibition of DNA-binding by the Exd-interaction motif is relieved upon binding Exd.²⁶ While we cannot measure an accurate dissociation constant at the protein concentrations required to observe a Ubx•Exd•Hth supershift of Dll DNA, our approximation of an upper bound for the true K_d is comparable to the K_d for Ubx monomer binding Dll DNA. Consequently, for Ubx we cannot conclusively determine whether interaction with Exd is able to reduce the DNA binding inhibition imposed by the YPWM motif.

We have identified three smaller regions within this 18 amino acid sequence, each with separable effects, which combine in a complex manner to generate larger variations in DNA binding affinity. We demonstrate for the first time that DNA binding inhibition by the YPWM motif in UbxIa is limited to Hox•Exd composite DNA binding sites. Thus, one role of the YPWM motif may be to specifically prevent Hox monomers from binding composite DNA sites, thereby misregulating their associated genes.^{45,46} Since an alternate Exd interaction motif is present in Ubx,²⁹ the YPWM motif may, therefore, be free to also regulate DNA binding specificity in a subset of Ubx•Exd heterodimers.

The region surrounding the YPWM motif modulates both the strength and DNA sequence-dependence of YPWM-mediated binding inhibition. In addition, the amino acids N- and C-terminal to the YPWM motif independently impact DNA site selection, an effect that is dependent on the amino acid sequence for the C-terminal sub-region. Divergence – to different degrees – of these regions in Ubx orthologues suggests DNA site selection may have evolved and thus contributed to the emergence of distinct body morphologies (Fig. 1d).¹⁰

Other specificity determinants in UbxIa

A key question is whether the 18 amino acid region is the only specificity determinant in UbxIa. Since the impact of this region is DNA sequence-dependent, the sensitivity of each DNA sequence to non-homeodomain regions must be separately addressed. In general, the effects of the 18 amino acid region do not match the net impact of all non-homeodomain sequences (Protein sensitivity^{FL-18} versus Protein sensitivity^{FL-HD}, Table 2). For instance, of the DNAs examined, the *Dpp* sequence was the most sensitive to changes in the 18 amino acid region,

yet is the least sensitive to the net impact of all non-homeodomain sequences. Conversely, binding to the UA and A1 DNAs is relatively insensitive to the 18 amino acid region, but extremely sensitive to all non-homeodomain sequences. Therefore, although *in vitro* and *in vivo* data²⁵ suggest the importance of the 18 amino acid region in DNA site selection by full-length UbxIa, other specificity determinants must also be present.

Implications for the Hox paradox

Hox proteins must specifically regulate their target genes, despite the wide array of binding sites that meet the low sequence standards of their highly homologous homeodomains.^{7–9, 20,47,48} In contrast to other homeodomain families, most Hox homeodomains contain a “molecular code” that favors the 5′-TAAT-3′ sequence.^{32, 43, 44} We and others find that the Ubx homeodomain binds with less than 3-fold variation in DNA binding affinity to DNA sequences containing this core motif.²⁰ Although this variation may contribute to sequence selection *in vivo*, it is likely to be insufficient to reliably distinguish Hox and non-Hox targets. For example, under conditions where an obligatory gene target is 100% occupied by a Hox protein, sites with unfavorable flanking sequences would still be ~40% occupied and therefore subject to mis-regulation. Indeed, we observe high affinity (72 ± 3 pM) binding of the UbxHD monomer to the Dll DNA sequence, which should only be bound by Hox•Exd heterodimers *in vivo*.

Individual Hox proteins direct the development of multiple tissues. Consequently, Hox proteins must also be able to recognize the cellular context and respond by binding only the target genes appropriate for that tissue.^{5, 6} The 18 amino acid region in Ubx could provide several mechanisms to modulate DNA binding in response to contextual cues. First, the structure or energetics of this region could be altered by protein interactions.^{13, 14, 26} Second, during *Drosophila* development, five different Ubx isoforms are produced in a spatiotemporal-dependent manner by varying the microexon content at this location.¹⁶ The C-terminal portion of the 18 amino acid region both modulates Exd interaction and includes the first portion of the mI microexon.^{16, 28} Therefore, mRNA splicing may alter both DNA binding specificity and the balance between binding by Ubx monomers and Ubx•Exd heterodimers to direct functions specific to a particular tissue or developmental stage. Indeed, five of the eight *Drosophila* Hox proteins are alternatively spliced between the YPWM motif and the homeodomain, encompassing a portion of the 18 amino acid region, suggesting that this region may act as an “antenna” which senses and responds to cell context in many Hox proteins.

Indeed, context-specific alternative splicing of intrinsically disordered regions has been proposed to be a means to diversify protein function by including or excluding functional domains (*e.g.*, protein or nucleic acid binding domains) or regulatory regions in a tissue-dependent manner.⁴⁹ Here we have shown that the intrinsically disordered 8 amino acids C-terminal to the YPWM motif, which includes a portion of the mI microexon, act as a specificity determinant. This region also acts as a regulator of another regulatory region, modulating inhibition of binding by the YPWM motif. We hypothesize that such modulation is mediated by altering the distance between the YPWM motif to the homeodomain via alternative splicing, a novel mechanism for regulation by an intrinsically disordered region. If true, then the other microexons may have a similar role, a hypothesis currently under investigation in our laboratory.

For many proteins, the length of a flexible linker has previously been shown to impact the activity of its terminal functional domains. Reducing linker length by mutation to within 5–10 amino acids can alter both the activity and stability of the N- and C-terminal functional domains.^{50,51} Conversely, increasing linker length to 60–70 amino acids may render the functional domains insensitive to one another.^{52,53} In our case, the N-terminal domain is the short YPWM motif which regulates the activity of the C-terminal DNA binding homeodomain.

During development, alternative mRNA splicing varies the distance between the YPWM motif from 8 to 51 amino acids,⁴⁹ roughly spanning the range in which functional domains appear to be sensitive to linker length. While this range is likely to be dependent on the size, function, and type of regulation used for each protein, the linker lengths separating other regulatory and functional modules also fall within this length range.^{54,55} As for Ubx, the ability of these distant motifs to regulate function is also dependent on the length and amino acid sequence.^{54,55}

The YPWM Exd-interaction motif forms a type I reverse turn and is located between two intrinsically disordered regions¹⁰ and thus has the characteristics of a linear motif, in which a short segment that mediates protein interactions is embedded in a larger disordered segment.⁵⁶ The less stringent sequence requirements for linear motifs relative to protein interaction surfaces in globular proteins enable more rapid evolution of protein interaction networks. In the case of Ubx, the YPWM linear motif is a conserved feature found in most Hox proteins in all bilaterally symmetric animals. However, the region surrounding this motif has evolved to not only be flexible, but also to regulate the function of both the YPWM motif and the homeodomain. Hence, linear motifs may also act to create functional diversity not by evolving the protein interaction motif but by using the disordered regions to modulate adjacent functional domains.

Variations in these DNA binding regulatory mechanisms may also generate functional diversity among Hox homologues. Members of the Hox protein family have nearly identical homeodomain sequences, and thus very similar DNA binding site preferences.^{7–9, 20, 43, 44} Despite this similarity, each Hox protein within an organism must regulate different genes *in vivo* to confer unique identities to otherwise similar regions. For instance, AbdA, but not Ubx, activates expression of *lh*, *tina-1*, *wg*, and *ndae1* in the *Drosophila* embryonic cardiac system.^{57–59} Since the sequence surrounding its YPWM motif is very different in AbdA and Ubx,⁶⁰ this region may diversify DNA site selection by their nearly identical homeodomains. Indeed, the region N-terminal to the homeodomain, including the 18 amino acid specificity-conferring sequence, is sufficient to dictate AbdA versus Ubx function in the visceral mesoderm.⁶⁰ The inability of this region to distinguish AbdA and Ubx functions in the epidermis further demonstrates the role of this region is context-dependent.

In summary, previous *in vivo* developmental biology studies established the 18 amino acid region alters Hox activity in a tissue-specific manner.^{25, 60} Furthermore, this region also distinguishes the roles of individual Hox proteins *in vivo*.⁶⁰ Our quantitative DNA binding data not only support these prior observations, but also provide a molecular mechanism by which the region surrounding the YPWM motif may diversify Hox function *in vivo* as well as respond to the cellular context by altering DNA site selection.

MATERIALS AND METHODS

Molecular Cloning of Hexapeptide Mutants

UbxIa and the Ubx-18 mutant genes, the gifts of Dr. Ella Tour and Dr. Bill McGinnis (UCSD), were sub-cloned into the pET3c expression vector. All other internal deletions and point mutations were constructed using the QuikChange site-directed mutagenesis kit (Stratagene).

Protein Expression and Purification

Wild-type UbxIa and UbxHD were expressed and purified as described previously.¹⁰ All point mutations and small internal deletions were examined in full-length, untagged UbxIa. The expression and purification procedures for all variants were the same as for wild-type UbxIa.

10

Full-length Exd, cloned into pET-9a, and histidine-tagged, full-length Homothorax (Hth), in pET-14b, were co-expressed in *E. coli* using the same procedures as for Ubx.¹⁰ The cell lysate was sonicated with twelve pulses at 15 s intervals using a VWR Bronson Scientific sonicator followed by centrifugation at $22,000 \times g$ for 20 min. Polyethyleneimine (50% W/V, 200 μ l) was added to the supernatant prior to centrifugation for 15 min. Protein samples were mixed with 4 mL nickel-nitrilotriacetic acid (Ni-NTA) agarose resin (slurry), which was pre-equilibrated with wash buffer, containing 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM imidazole, and 5% glucose. The protein mixture was loaded onto the column, washed with 50 mL wash buffer containing 20 mM imidazole, followed by 30 mL wash buffer containing 40 mM imidazole, and 20 mL wash buffer containing 80 mM imidazole. Protein was eluted with wash buffer containing 200 mM imidazole and analyzed by SDS-PAGE and western blots.

Aggregation Assays

The filter aggregation assay¹⁷ was used to examine solubility of purified wild-type and mutant UbxIa. The YAAA mutant, which could not be purified, formed insoluble inclusion bodies. UbxIa and all other UbxIa derivatives were soluble.

Electrophoretic Mobility Shift (Gel Shift) Analysis

Synthesized DNA binding oligonucleotides were annealed and labeled as previously described.¹⁰ Oligonucleotide sequences are listed in Table 1. The Dpp DNA binding sequence is located on chromosome 2L, at bases 2452271–2452228; the UA sequence is on 3R, at bases 12560064–12560101; the Sal sequence is on 2R, at bases 11454558–11454617; the A1 sequence is on 3R, at bases 2830985–2831025; and the Dll sequence is on 2R, at bases 20691027–20690984. Both protein activity and affinity experiments were performed using gel retardation assays, using the same binding buffer and gel composition reported previously.¹⁰ Activity assays measured the fraction of active protein using stoichiometric conditions, where $[\text{DNA}] \gg K_d$. All purified proteins were greater than 80% active. The affinity gel retardation assays were performed under equilibrium conditions, where $[\text{DNA}] \ll K_d$, generally 10^{-13} to 10^{-10} M depending on the protein utilized and DNA sequence.¹⁰ Since some DNA sequences used have multiple binding sites, the disappearance of free DNA was analyzed for consistency. Therefore, the *apparent* dissociation constant, K_{app} , rather than K_d , is reported for the multi-site target DNAs. Each measurement is corrected for protein activity and is the average of at least nine gel shift experiments, as reported previously.¹⁰

Exd/Hth Supershift Experiments

For gel super-shift experiments, purified UbxIa or Ubx-18 (concentrations ranging from 1×10^{-11} M to 1×10^{-9} M) with 5×10^{-8} M purified Exd•Hth and 10^{-12} M Dll oligonucleotide in 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 100 μ g/mL BSA, 5 mM DTT, and 10% glycerol were incubated for 30 min. at room temperature. Retardation gels contained 0.5% agarose and 3% polyacrylamide (37.5:1 acrylamide:bisacrylamide), 0.5X TBE (0.023 M Tris-borate, 0.5 mM EDTA, pH 8.0), and 3% glycerol. The gels were pre-run at constant current (150 V) while circulating 0.5X TBE buffer for at least 30 min. Samples were loaded at 300 V, and after 5–8 min. the voltage was reduced to 150 V for approximately three hours. Gels were blotted onto filter paper, dried on a vacuum dryer, and exposed to a FUJI phosphorimaging plate overnight. Images of protein•DNA complexes were scanned with a Fuji Imager and quantified using the MacBAS1000 2.0 program.

Circular Dichroism Experiments

Circular dichroism was used to assess structural changes due to mutation, whereas magnetic circular dichroism was used to determine protein concentrations by measuring tryptophan

concentration in an environmentally independent manner.⁶¹ Buffer conditions and experimental parameters were as previously described.¹⁰

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

REFERENCES

1. Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature* 1978;276:565–570. [PubMed: 103000]
2. Bondos SE, Tan XX. Combinatorial transcriptional regulation: The interaction of transcription factors and cell signaling molecules with homeodomain proteins in *Drosophila* development. *Crit. Rev. Eukaryot. Gene Expr* 2001;11:145–171. [PubMed: 11693959]
3. Hughes CL, Kaufman TC. Hox genes and the evolution of the arthropod body plan. *Evol. Dev* 2002;4:459–499. [PubMed: 12492146]
4. Castelli-Gair J, Akam M. How the Hox gene *Ultrabithorax* specifies two different segments: The significance of spatial and temporal regulation within metameres. *Development* 1995;121:2973–2982. [PubMed: 7555723]
5. Graba Y, Aragnol D, Pradel J. *Drosophila* Hox complex downstream targets and the function of homeotic genes. *Bioassays* 1997;19:379–388.
6. Pearson JC, Lemons D, McGinnis W. Modulating Hox gene functions during animal body patterning. *Nature Rev. Genet* 2005;6:893–904. [PubMed: 16341070]
7. Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-Perez D, Affolter M, Otting G, Wüthrich K. Homeodomain-DNA recognition. *Cell* 1994;78:211–223. [PubMed: 8044836]
8. Hoey T, Levine M. Divergent homeobox proteins recognize similar DNA sequences in *Drosophila*. *Nature* 1988;332:858–861. [PubMed: 2895896]
9. Kalionis B, O'Farrell PH. A universal target sequence is bound *in vitro* by diverse homeodomains. *Mech. Dev* 1993;43:57–70. [PubMed: 7902124]
10. Liu Y, Matthews KS, Bondos SE. Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the *Drosophila* Hox protein Ultrabithorax. *J. Biol. Chem* 2008;283:20874–20887. [PubMed: 18508761]
11. Joshi R, Rohs R, Jain R, Sosinsky A, Crickmore MA, Jacob V, Aggarwal AK, Honig B, Mann RS. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 2007;131:530–543. [PubMed: 17981120]
12. Taghli-Lamalle O, Hsia C, Ronshaugen M, McGinnis W. Context-dependent regulation of Hox protein functions by CK2 phosphorylation sites. *Dev. Genes Evol* 2008;218:321–332. [PubMed: 18504607]
13. Bondos SE, Tan XX, Matthews KS. Physical and genetic interactions link Hox function with diverse transcription factors and cell signaling proteins. *Mol. Cell. Proteomics* 2006;5:824–834. [PubMed: 16455680]
14. Bondos SE, Catanese DJ Jr, Tan XX, Bicknell A, Li L, Matthews KS. Hox transcription factor Ultrabithorax Ib physically and genetically interacts with disconnected interacting protein 1, a double-stranded RNA-binding protein. *J. Biol. Chem* 2004;279:26433–26444. [PubMed: 15039447]
15. Chan SK, Jafe L, Capovilla M, Botas J, Mann RS. The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with Extradenticle, another homeoprotein. *Cell* 1994;78:603–615. [PubMed: 7915199]
16. López AJ, Artero RD, Pérez-Alonso M. Stage, tissue, and cell specific distribution of alternative *Ultrabithorax* mRNAs and protein isoforms in the *Drosophila* embryo. *Roux's Arch. Dev. Biol* 1996;205:450–459.
17. Bondos SE, Bicknell A. Detection and prevention of protein aggregation before, during, and after purification. *Anal. Bioch* 2003;316:223–231.

18. Galant R, Walsh CM, Carroll SB. Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development* 2002;129:3115–3126. [PubMed: 12070087]
19. Rippe K. Analysis of Protein-DNA binding at equilibrium. *B.I.F. Futura* 1997;12:20–26.
20. Ekker SC, Young KE, von Kessler DP, Beachy PA. Optimal DNA sequence recognition by the Ultrabithorax homeodomain of *Drosophila*. *EMBO J* 1991;10:1179–1186. [PubMed: 1673656]
21. Beachy PA, Varkey J, Young KE, von Kessler DP, Sun BI, Ekker SC. Cooperative binding of an Ultrabithorax homeodomain protein to nearby and distant DNA sites. *Mol. Cell. Biol* 1993;13:6941–6956. [PubMed: 8105373]
22. Krasnow MA, Saffman EE, Kornfeld K, Hogness DS. Transcriptional activation and repression by Ultrabithorax proteins in cultured *Drosophila* cells. *Cell* 1989;57:1031–1043. [PubMed: 2567632]
23. Gebelein B, Culi J, Ryoo HD, Zhang W, Mann RS. Specificity of *Distalless* repression and limb primordial development by Abdominal Hox proteins. *Dev Cell* 2002;3:487–498. [PubMed: 12408801]
24. Gebelein B, McKay DJ, Mann RS. Direct integration of Hox and segmentation gene inputs during *Drosophila* development. *Nature* 2004;431:653–659. [PubMed: 15470419]
25. Tour E, Hittinger CT, McGinnis W. Evolutionarily conserved domains required for activation and repression functions of the *Drosophila* Hox protein Ultrabithorax. *Development* 2005;132:52571–5281
26. Chan SK, Pöpperl H, Krumlauf R, Mann RS. An extradenticle-induced conformational change in a Hox protein overcomes an inhibitory function of the conserved hexapeptide motif. *EMBO J* 1996;15:2476–2487. [PubMed: 8665855]
27. Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK. Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature* 1999;397:714–719. [PubMed: 10067897]
28. Johnson FB, Parker E, Krasnow MA. 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* 1995;92:739–743. [PubMed: 7846045]
29. Merabet S, Saadaoui M, Sambrani N, Hudry B, Pradel J, Affolter M, Graba Y. A unique extradenticle recruitment mode in the *Drosophila* Hox protein Ultrabithorax. *Proc. Natl. Acad. Sci. USA* 2007;104:16946–16951. [PubMed: 17942685]
30. O'Connor MB, Binari R, Perkins LA, Bender W. Alternative RNA products from the *Ultrabithorax* domain of the bithorax complex. *EMBO J* 1988;7:435–445. [PubMed: 2452731]
31. Furukubo-Tokunaga K, Flister S, Gehring WJ. Functional specificity of the Antennapedia homeodomain. *Proc. Natl. Acad. Sci. USA* 1993;90:6360–6364. [PubMed: 8101003]
32. Damante G, Pellizzari L, Esposito G, Fogolari F, Viglino P, Fabbro D, Tell G, Formisano S, Di Lauro R. A molecular code dictates sequence-specific DNA recognition by homeodomains. *EMBO J* 1996;15:4992–5000. [PubMed: 8890172]
33. Frazee RW, Taylor JA, Tullius TD. Interchange of DNA-binding modes in the Deformed and Ultrabithorax homeodomains: A structural role for the N-terminal arm. *J. Mol. Biol* 2002;323:665–683. [PubMed: 12419257]
34. Tan XX, Bondos S, Li L, Matthews KS. Transcription activation by Ultrabithorax Ib protein requires a predicted α -helical region. *Biochemistry* 2002;41:2774–2785. [PubMed: 11851425]
35. Batra-Safferling R, Abarca-Heidemann K, Körschen HG, Tziatzios C, Stoldt M, Budyak I, Willbold D, Schwalbe H, Klein-Seetharaman J, Kaupp UB. Glutamic acid-rich proteins of rod photoreceptors are natively unfolded. *J. Biol. Chem* 2006;281:1449–1460. [PubMed: 16280326]
36. Sun B, Hursh DA, Jackson D, Beachy PA. Ultrabithorax protein is necessary but not sufficient for full activation of decapentaplegic expression in the visceral mesoderm. *EMBO J* 1995;14:520–535. [PubMed: 7859741]
37. Kyte J, Doolittle R. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol* 1982;157:105–132. [PubMed: 7108955]
38. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet* 2001;42:38–48. [PubMed: 11093259]
39. Iimura T, Pourquié O. Hox genes in time and space during vertebrate body formation. *Dev. Growth Differ* 2007;49:265–275. [PubMed: 17501904]

40. Anderson KM, Darweesh M, Jajah A, Tsui P, Guinan P, Rubenstein M. Does homeobox-related "positional" genomic information contribute to implantation of metastatic cancer cells at non-random sites? *Anticancer Res* 2007;27:2141–2153. [PubMed: 17695497]
41. Galant R, Carroll SB. Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* 2002;415:910–913. [PubMed: 11859369]
42. Ronshaugen M, McGinnis N, McGinnis W. Hox protein mutation and macroevolution of the insect body plan. *Nature* 2002;415:914–917. [PubMed: 11859370]
43. Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* 2008;133:1277–1289. [PubMed: 18585360]
44. Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML, Hughes TR. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* 2008;133:1266–1276. [PubMed: 18585359]
45. Pinsonneault J, Florence B, Vaessin H, McGinnis W. A model for *extradenticle* function as a switch that changes Hox proteins from repressors to activators. *EMBO J* 1997;16:2032–2042. [PubMed: 9155029]
46. Li X, Murre C, McGinnis W. Activity regulation of a Hox protein and a role for the homeodomain in inhibiting transcriptional activation. *EMBO J* 1999;18:198–211. [PubMed: 9878063]
47. Ekker SC, Jackson DG, von Kessler DP, Sun BI, Young KE, Beachy PA. The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. *EMBO J* 1994;13:3551–3560. [PubMed: 7914870]
48. Mastick GS, McKay R, Oligino T, Donovan K, López AJ. Identification of target genes regulated by homeotic proteins in *Drosophila melanogaster* through genetic selection of *Ultrabithorax* protein-binding sites in yeast. *Genetics* 1995;139:349–363. [PubMed: 7705635]
49. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl. Acad. Sci. USA* 2006;103:8390–8395. [PubMed: 16717195]
50. Bergeron LM, Gomez L, Whitehead TA, Clark DS. Self-regulating enzymes: design of an enzyme-chaperone chimera as a new approach to enzyme stabilization. *Biotechnol. Bioeng* 2009;102:1316–1322. [PubMed: 19170240]
51. Sorensen BR, Faga LA, Hultman R, Shea MA. An interdomain linker increases the thermostability and decreases the calcium affinity of the calmodulin N-domain. *Biochemistry* 2002;41:15–20. [PubMed: 11771998]
52. Shamoo Y, Abdul-Manan N, Williams KR. Multiple RNA binding domains (RBDs) just don't add up. *Nucl. Acids Res* 1995;23:825–728.
53. Evers TH, van Dongen EMWM, Faesen AC, Meijer EW, Merckx M. Quantitative understanding of the energy transfer between fluorescent proteins connected via flexible peptide linkers. *Biochemistry* 2006;45:13183–13192. [PubMed: 17073440]
54. Zhou HX. Quantitative relation between intermolecular and intramolecular binding of pro-rich peptides to SH3 domains. *Biophys. J* 2006;91:3170–3181. [PubMed: 16891373]
55. Mattison K, Oropeza R, Kenny LJ. The linker region plays an important role in the interdomain communication of the response regulator OmpR. *J. Biol. Chem* 2002;277:32714–32721. [PubMed: 12077136]
56. Fuxreiter M, Tompa P, Simon I. Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* 2007;23:950–956. [PubMed: 17387114]
57. Lovato TL, Nguyen TP, Molina MR, Cripps RM. The Hox gene Abdominal-A specifies heart cell fate in the *Drosophila* dorsal vessel. *Development* 2002;129:5019–5027. [PubMed: 12397110]
58. Monier B, Astier M, Sémériva M, Perrin L. Steroid-dependent modification of Hox function drives myocyte reprogramming in the *Drosophila* heart. *Development* 2005;132:5283–5293. [PubMed: 16284119]

59. Ryan KM, Hoshizaki DK, Cripps RM. Homeotic selector genes control the patterning of seven-up expressing cells in the *Drosophila* dorsal vessel. *Mech. Dev* 2005;122:1023–1033. [PubMed: 15922572]
60. Chauvet S, Merabet S, Bilder D, Scott MP, Pradel J, Graba Y. Distinct Hox protein sequences determine specificity in different tissues. *Proc. Natl. Acad. Sci. USA* 2000;97:4064–4069. [PubMed: 10737765]
61. Holmquist B, Vallee BL. Tryptophan quantitation by magnetic circular dichroism in native and modified proteins. *Biochemistry* 1973;12:4409–4417. [PubMed: 4750252]
62. Capovilla M, Brandt M, Botas J. Direct regulation of *decapentaplegic* by *Ultrabithorax* and its role in *Drosophila* midgut morphogenesis. *Cell* 1994;76:461–475. [PubMed: 7906203]

Fig. 1.

Homeodomain structure and Ubx sequence motifs. a. A schematic of the Ubx sequence showing mRNA splicing isoforms with the location of the homeodomain (HD, blue), the regions encoded by the b, mI, and mII microexons (shades of yellow/orange), and the YPWM motif (Y, blue), which may be used to bind the general Hox cofactor Exd.^{15, 26} The sequences corresponding to the alternatively spliced microexons are intrinsically disordered, as demonstrated by disorder prediction algorithms and extreme susceptibility to native state proteolysis.¹⁰ The DNA binding experiments in this work focus on UbxIa, the most abundant isoform in the *Drosophila* embryo.¹⁶ b,c. Two views of a double truncation mutant of UbxIVa, which contains none of the microexon-encoded sequences, spanning the YPWM motif through

the homeodomain (blue, side chains included for the YPWM motif) bound to a composite DNA binding site (black) and the Exd homeodomain (red).²⁷ The region linking the YPWM motif to the homeodomain is too disordered to be visible in the crystal structure. To demonstrate the size of the small UbxIVa linker relative to the distance between the YPWM motif and the homeodomain, a model of this loop is shown in yellow, in which each dot represents the alpha carbon of one amino acid in the UbxIVa isoform (the shortest of the isoforms). d. Alignment of a portion of Ubx orthologues, including the 18 amino acid sequence that enhances DNA binding specificity (marked by a heavy black line) and the region surrounding this crucial area. The sequences for butterfly *Juonia coenia* Ubx (JcUbx), beetle *Tribolium castaneum* Ubx (TcUbx), *Artemia franciscana* Ubx (AfUbx) and velvet worm *Akanthokara kaputensis* Ubx (AkUbx) are derived from references.^{41,42} Sequence homology to *Drosophila* UbxIa (Dm UbxIa, gray) decreases with distance from the YPWM motif. The variation of some sequences within this 18 amino acid region potentially reflects evolution of DNA binding specificity in Ubx orthologues. e. We focus on DNA binding specificity determinants within an 18 amino acid section located largely within the I1 domain. The sequence of this region is expanded to illustrate the UbxIa mutants used in this study. The bars above the UbxIa schematic indicate the positions of the DNA binding inhibitory domains (I1 and I2) and the domain that counters this inhibition to restore near-wild type high affinity DNA interaction (R).¹⁰

Exd-independent DNA binding sites

Exd-dependent DNA binding site

Fig. 2.

Schematics of enhancers containing the Ubx DNA binding sites. The numbers indicate the position of the DNA sequence used relative to the transcription start site for its associated gene. Ubx binding sites are marked by red right facing arrows (5'-TAAT-3' sites) and red left facing arrows (5'-ATTA-3' sites).^{18, 20-24} Atypical binding sites recognized by UbxIa are indicated by orange left facing arrows (5'-ATTT-3') and purple left facing arrows (5'-ATGT-3').⁶² Note that while UbxIa binds this Dpp sequence *in vitro*,²⁰ it is not known whether it is regulated by Ubx *in vivo*. The scale of the schematic is indicated by the 100 bp (black) and 1 kb (blue) scale bars. Because the large size of these enhancers precludes use of the entire enhancer in DNA binding experiments, only the sequences labeled and marked by a green box were utilized.

Fig. 3.

DNA binding by UbxHD and UbxIa. a,b. Gel shift of UbxHD (a) and UbxIa (b) binding UA DNA. Although multiple sites are bound by UbxHD and UbxIa, single protein binding dominates at the concentration of Ubx equal to the apparent dissociation constant, indicated by an arrow. Multiple binding by UbxHD at higher protein concentrations is consistent with previous reports.²¹ Gels for the remaining DNA sequences can be found in Supplementary Materials Fig. 1. c. All multiple site DNA sequences are bound many times by UbxHD and UbxIa. For each DNA examined, the sample in the left lane contains DNA only, whereas the sample in the right lane contains DNA and UbxIa at concentrations greater than K_d or K_{app} . Therefore, in addition to DNA affinity, UbxIa•UbxIa interactions and the availability of multiple sites also impact the apparent affinity reported. d. The dissociation constant, K_d , is reported for the 40AB and DII single site DNAs. The apparent dissociation constant, K_{app} , is reported for DNA sequences with multiple binding sites (Dpp, UA, A1, and Sal). e–j. DNA binding curves for UbxIa (grey) and UbxHD (black). To permit consistent analysis of binding

to single and multiple site DNAs, the disappearance of free DNA was monitored. Note the differences in affinity are generally more pronounced for the natural sites than for the optimal binding sequence, 40AB. Note also that the UbxIa curves are not significantly cooperative, indicating a relatively low level of protein•protein interaction.

Fig. 4.

The impact of the Ubx-18 and GPGG mutations on the function and structure of Ubx. a. Comparison of sequence-dependent DNA binding by Ubx-18 and the GPGG variant with UbxIa and UbxHD. Although both of these variations cause DNA sequence-specific effects on affinity, the major impact of the GPGG mutant is limited to enhancing affinity (lowering K_d) for DII DNA, which contains a single Hox•Exd composite sequence. The broader impact of the internal deletion in Ubx-18 suggests that sequences flanking the YPWM motif also modulate DNA binding specificity. b. UbxIa binds poorly to all Hox•Exd composite sequences examined; these sequences differ in DNA binding site sequences and spacing between binding sites. Removal of the YPWM motif, either by deletion or point mutations, improves binding

to a similar degree, suggesting inhibition of binding by the YPWM motif is a common feature of monomer binding to heterodimer sites. c. Circular dichroism of Ubx1a (open circles) and Ubx-18 (gray circles). The similarity of these spectra reveals that the 18 amino acid internal deletion does not cause any large-scale changes in Ubx structure.

Fig. 5.

Both UbxIa and Ubx-18 cooperatively bind the Dll DNA sequence with an Exd•Hth heterodimer. Although minimal DNA binding is observed by isolated UbxIa / Ubx-18 at these low concentrations, combination of these proteins permits facile formation of a quaternary complex (Ubx, Exd, Hth, and DNA), as revealed by the presence of a strong supershifted band (marked by U•E•H•D).

Fig. 6.

The impact of the Ubx-6, Ubx-8, and Ubx-14 mutations on DNA binding by Ubx. a. Comparison of the DNA specificity of the Ubx-6 and Ubx-8 internal deletion mutants with UbxIa and UbxHD. The Ubx-6 data recapitulate the DNA binding profile of the GPGG mutant, although with weaker effects. The-8 internal deletion enhances binding to DII yet inhibits 40AB binding relative to UbxIa. b. Deletion (Ub-8) and rearrangement (Ub8scramble) of the 8 amino acids C-terminal to the YPWM motif similarly impact DNA binding, indicating deletion mutants can be effectively used to probe the function of this region. c. PONDR VX-LT intrinsic disorder predictions are similar for UbxIa and Ub8scramble, suggesting the dynamics of this disordered region are not greatly altered. d. The hydrophobicity profile of this region is also

similar for Ubx1a and Ubx8scramble. Consequently, the differences in binding observed in panel b are likely due to changes in side chain interactions between this region and the rest of the protein. e. Comparison of DNA affinities for Ubx-14 relative to Ubx-18, Ubx1a, and UbxHD. Relative to Ubx1a, the -14 amino acid deletion largely recapitulates the effects of the -8 amino acid deletion. Comparison of Ubx-14 and Ubx-18 reveals that, in the absence of the 14 surrounding amino acids, the YPWM motif can inhibit binding to all DNA sequences tested. In contrast, the YPWM motif in the full-length protein can only inhibit binding to Hox•Exd composite sequences (Fig. 4 a,b).

Table 1
DNA binding sequences utilized in gel shift experiments

DNA sequences were selected to maximize diversity of binding sites, considering site number and density, the germ layer, tissue, and stage in which Ubx binds *in vivo*, whether Ubx activates or represses transcription once bound to these sequences, and finally the requirement for Exd interaction. Hox binding sites are underlined and in bold text. Bases protected by Exd binding are shaded dark grey with white letters, and an Hth binding site is shaded light grey with black letters. Although the Dpp sequence has not been shown to be regulated by Ubx *in vivo*, it was included because its multiple sites more closely resemble the Ubx optimal sequence than the other genome-derived sequences.

Oligo	Top Strand Sequence	Reference	Stage / Tissue	Trans.	Exd
40AB	5'-CCGGGCTGCACATGGTTAAATGGCCAGTCCACGCG TAGATC-3'	10 ^a , 20 ^a , 47	N/A ^d	N/A	N
Dll	5'-GCACTATAAACTGTCGCCGGGAAATGATTAAATT TCCCAAAATATTGTC-3'	23 ^a , 24	Embryonic ectoderm	R	Y
Dpp	5'-GGTAATGGTAATAGTAACGGTAAATGATAAATGG TAATGGTAATGG-3'	20	N/A	N/A	N
UA	5'-GGAATAATAACAAATAATGCCGCTGATAAATGTGGA TAATAAGG-3'	21	Embryonic mesoderm	A ^b	N
Sal	5'-TACTGCTTTTAAATAAGTTATGCGAACTGATGGT TTGGTTAGGCATGCTCGATGCTTTTATAAATGTGCCC GTCTTAATATGATTTTGTG-3'	18	Larval haltere imaginal disc	R ^c	N
Al	5'-GGTATAATATAATAATAAAAA TAATAATAATAATAATAATAATATATGG-3'	22	Embryonic mesoderm	R	N
Dppe4	5'-GTAGCTTCATCGATAAAACTTAATGGCCACTATA AC-3'	36	Embryonic mesoderm	A	Y
repeat 3	5'-GGGGTGAATGGATGGGGCGCTG-3'	26	N/A	N/A	Y
Sal6	5'-GATGTCCTTTTATAATGTGCCCGCTCTG-3'	18	Larval haltere imaginal disc	R ^c	N

^aNot applicable
^bTranscription activation
^cTranscription repression

Table 2
Equilibrium binding data for Ubx1a and its variants to the 6 DNAs used to examine sequence specificity
Equilibrium DNA binding constants (K_d or K_{app} , for single and multiple binding sites, respectively) are listed in pM for each protein and DNA combination. The fold variation column reports the largest affinity constant divided by the smallest affinity constant for each protein examined. Fold variation thus reflects DNA specificity, the ability of each protein to discriminate between binding sites. For each DNA sequence, protein sensitivity reports the largest affinity constant divided by the smallest affinity constant, reflecting the variation in protein binding elicited by that DNA sequence. Protein sensitivity^{FL-HD} only includes data for full-length Ubx1a and UbxHD. The degree of variation thus reflects the net impact of all non-homeodomain regions on DNA affinity. Protein sensitivity^{FL-18} includes data for all proteins except UbxHD, and thus reports on the impact of changes in the 18 amino acid region on DNA affinity.

	DNA Binding Affinity in pM						Fold variation
	Optimal 40AB	Composite DII	Multiple				
			Dpp	UA	A1	Sal	
UbxHD	63 ± 24	72 ± 3	76 ± 45	79 ± 23	28 ± 3	39 ± 10	2.8
Ubx1a	160 ± 33	1900 ± 280	280 ± 59	473 ± 48	680 ± 55	590 ± 150	12
GPGG	120 ± 48	720 ± 130	105 ± 140	461 ± 79	530 ± 150	570 ± 98	6.5
Ubx-18	100 ± 50	530 ± 70	250 ± 83	378 ± 130	360 ± 46	220 ± 59	5.3
Ubx-14	560 ± 20	1000 ± 170	760 ± 270	574 ± 10	610 ± 17	920 ± 260	1.8
Ubx-8	420 ± 50	1000 ± 180	490 ± 160	453 ± 89	550 ± 82	790 ± 66	2.4
Ubx-6	170 ± 34	1400 ± 170	310 ± 73	506 ± 120	690 ± 230	600 ± 190	8.2
Protein sensitivity ^{FL-HD}	2.5	26.5	3.7	6.0	24	15	
Protein sensitivity ^{FL-18}	5.4	3.5	7.2	1.5	1.7	4.2	

Table 3
Equilibrium binding data to composite and multiple sequences

Binding to the DNA sequences Sal (3 Hox sites) and Sal6 (middle site only) illustrates the role of protein-protein interactions in binding is dependent on both the DNA the protein sequences (italics text). Dissociation constants for the YPWM motif mutants GPGG and Ubx-18 binding to additional Hox-Exd composite DNA sites demonstrate that the YPWM motif generally represses binding to Hox•Exd binding sites (shaded text). Finally, binding of the Ubx-8 and Ubx8scramble proteins are similar, demonstrating internal deletions and mutations within this region have similar effect (bold text). All parameters are as described for Table 2.

	DNA Binding Affinity in pM						Fold variation
	Optimal Ubx40AB	Composite		Multiple			
		DII	Dppe4	repeat 3	Dpp	Sal	
UbxHD	63 ± 24	72 ± 3	48 ± 10	42 ± 5.9			
UbxIa	160 ± 33	1900 ± 280	1500 ± 340	2000 ± 540	280 ± 59	590 ± 150	660 ± 75
GPGG	120 ± 48	720 ± 130	800 ± 72	880 ± 140			
Ubx-18	100 ± 50	530 ± 70	620 ± 130	800 ± 130		220 ± 59	500 ± 110
Ubx-8	420 ± 50	1000 ± 180			490 ± 160	790 ± 66	
Ubx8scramble	540 ± 52	930 ± 250			590 ± 190	800 ± 210	
Protein sensitivity ^{FL-HD}	2.5	27	32	48			
							1.7