Direct regulation of knot gene expression by Ultrabithorax and the evolution of cis-regulatory elements in Drosophila

Bradley M. Hersh and Sean B. Carroll

University of Wisconsin-Madison, Howard Hughes Medical Institute, 1525 Linden Drive, Madison, WI 53706, USA
*Author for correspondence (e-mail: sbcarrol@wisc.edu)

Accepted 2 February 2005

Development 132, 1567-1577
Published by The Company of Biologists 2005
doi:10.1242/dev.01737

Summary
The regulation of development by Hox proteins is important in the evolution of animal morphology, but how the regulatory sequences of Hox-regulated target genes function and evolve is unclear. To understand the regulatory organization and evolution of a Hox target gene, we have identified a wing-specific cis-regulatory element controlling the knot gene, which is expressed in the developing Drosophila wing but not the haltere. This regulatory element contains a single binding site that is crucial for activation by the transcription factor Cubitus interruptus (Ci), and a cluster of binding sites for repression by the Hox protein Ultrabithorax (UBX). The negative and positive control regions are physically separable, demonstrating that UBX does not repress by competing for occupancy of Ci-binding sites. Although knot expression is conserved among Drosophila species, this cluster of UBX binding sites is not. We isolated the knot wing cis-regulatory element from D. pseudoobscura, which contains a cluster of UBX-binding sites that is not homologous to the functionally defined D. melanogaster cluster. It is, however, homologous to a second D. melanogaster region containing a cluster of UBX sites that can also function as a repressor element. Thus, the knot regulatory region in D. melanogaster has two apparently functionally redundant blocks of sequences for repression by UBX, both of which are widely separated from activator sequences. This redundancy suggests that the complete evolutionary unit of regulatory control is larger than the minimal experimentally defined control element. The span of regulatory sequences upon which selection acts may, in general, be more expansive and less modular than functional studies of these elements have previously indicated.

Key words: Hox genes, Cis-regulatory element, Ultrabithorax, Drosophila, Evolution

Introduction
The Hox proteins are important regulatory molecules that shape the patterning of the anteroposterior axis in animal development, and changes in Hox expression pattern during evolution are associated with morphological modifications (Gellon and McGinnis, 1998). Despite marked differences in appearance, fruit flies, fish, humans and all other bilaterally symmetrical animals employ Hox proteins in the elaboration of this major body axis. The Hox transcription factors regulate downstream genes that may themselves have broad effects on morphology, as well as genes involved in terminal differentiation (Brody et al., 2002; Rozowski and Akam, 2002). Hox proteins can influence developmentally important processes such as apoptosis (Knosp et al., 2004; Lohmann et al., 2002), cell proliferation (Dolle et al., 1993; Salser and Kenyon, 1996) and cell fusion (Shemer and Podbilewicz, 2002). Despite these myriad effects, very few direct Hox-regulated target genes have been identified (Mann and Carroll, 2002). Furthermore, although individual Hox proteins are capable of either positively or negatively regulating target genes (Capovilla et al., 1994; Li and McGinnis, 1999; Vachon et al., 1992), it is not understood how these different activities are determined. To determine how different Hox-regulated target genes are controlled, how novel target genes are incorporated into regulatory networks and how changes in regulatory networks result in alterations in morphology, it is necessary to identify direct Hox-regulated target genes and characterize the regulatory elements that control them.

Within insects, the Hox gene Ultrabithorax (Ubx) is important for proper specification of the third thoracic segment. In Drosophila lacking Ultrabithorax function, the third thoracic segment is transformed to a second thoracic segment fate, resulting in complete duplication of the wing and mesonotum. In butterflies, clonal loss of Ubx also results in transformation of hindwing scales to a forewing pattern (Weatherbee et al., 1999). Thus, orthologous UBX proteins specify differences between forewings and hindwings in these two morphologically distinct contexts. A simple model postulates that Ubx modifies hindwing morphology by regulating different sets of downstream target genes in these insect orders (Weatherbee et al., 1999).

Several genes that are differentially expressed in the forewing and the haltere in Drosophila, and therefore are genetically downstream of Ubx, have been identified (Weatherbee et al., 1998). However, direct regulation of only one gene, spalt (sal), has been demonstrated (Galant et al., 2002). Through identification of additional UBX-regulated targets and characterization of their regulatory elements, we may determine sequence features that are required for UBX
regulation, and better understand how regulation by Hox proteins is integrated into a morphogenetic program, together with regulation by signaling pathways, other selector proteins and tissue-specific transcription factors.

The knot gene is a candidate for direct UBX regulation in the haltere. Loss of knot function causes apposition of the L3 and L4 veins, and loss of the L3-L4 intervein region in the forewing (Mohler et al., 2000; Vervoort et al., 1999). knot is expressed at the anteroposterior compartment boundary in the developing wing, where it is activated by Cubitus interruptus (Ci) (Vervoort et al., 1999), the transcriptional effector of the Hedgehog signaling pathway. knot expression is absent from the haltere, and knot is expressed cell-autonomously in clones that overexpress UB in the wing (Galant et al., 2002). In addition, D-SRF (Drosophila serum response factor; bs – FlyBase), which is itself a target of UB repression (Weatherbee et al., 1998), requires knot for activation (Vervoort et al., 1999). Thus, lack of D-SRF expression in the sequence may be due either to direct action of UBX on D-SRF regulatory sequences or to UBX repression of knot, its activator.

The knot gene is also required during embryonic development for formation of embryonic muscle (Crozatier and Vincent, 1999) and several head structures (Crozatier et al., 1999; Seecoomar et al., 2000). knot is expressed in the lymph gland precursors, and is required for the development of lamellocytes, large cells that encapsulate foreign bodies, in response to parasitization (Crozatier et al., 2004). This multiplicity of functions suggests a multiplicity of regulatory elements that control knot expression in its various contexts.

We have identified a wing-specific regulatory element for the knot gene and demonstrate its direct regulation by the Hedgehog signaling pathway and the UBX Hox protein. We find that a minimal element for repression in the haltere is not conserved, but a second, apparently redundant, element is conserved, and is located more than 500 bp from the minimal region. This result suggests that UBX repression is distributed over a large regulatory region that may not have sharply bounded elements, as defined by sequence conservation. In addition, a second, novel UBX repression element appears to have evolved in the D. melanogaster lineage in the presence of a pre-existing functional element, suggesting that selection is acting on a larger region than the minimally defined regulatory module.

Materials and methods
knot reporter constructs
Initial constructs were generated by PCR amplification or restriction digest from Drosophila melanogaster genomic DNA P1 clone DS00158. DNA fragments located 5′ of knot were cloned into the hsp-lacZ-CaSpeR reporter plasmid (Nelson and Laughon, 1993) and verified by sequence analysis. Reporter plasmids were injected to generate transgenic fly lines. Smaller fragments were generated either by restriction digest or PCR amplification, and subsequent cloning into the lacZ reporter vector. Cloning details are available upon request. β-Galactosidase activity for reporter constructs was detected with rabbit anti-Kn antibody provided by Michèle Crozatier (Crozatier and Vincent, 1999). Engrailed protein was detected with mouse monoclonal antibody 4F11 provided by Nipam Patel (Patel et al., 1989).

Mutagenesis of Ci- and UBX-binding sites
Ci-binding sites were altered by PCR mutagenesis. Site Ci1047 was altered from TGGGTGGG to TGGTGGAGCA; site Ci1341 was altered from GCCGCCGCT to GCCGTCAGTC; site Ci680 was altered from TGTCGAGCC to TGTCGAGCC. UBX-binding sites were altered or deleted by PCR mutagenesis. UBX site 1 was altered from GCTAAATTG to GCTGGCTTTG; UBX site 2 was altered from AGAATTAGC to AGAGCTAGGC; UBX site 3 was altered from CCACATTATA to CCACGGCGGC. The entire sequence of UBX site 4 shown in Fig. 4C was deleted by PCR sewing. The sequence of site 1835-1840 was altered from AACATGT to GGCCTGT by PCR mutagenesis. UBX sites in aligned block 2 were altered using the Stratagene Quickchange Mutagenesis kit following the manufacturer’s instructions.

DNasel footprinting
The fragment knMel1599-1991 was amplified by PCR and cloned into pGEM-T-Easy (Promega). The plasmid was linearized with Spel, labeled with [32P]-dNTPs by fill-in with Klenow enzyme, and precipitated overnight with ammonium acetate and ethanol. The insert was digested from the linearized plasmid with EcoRI, agarose gel purified and extracted from the gel using a Quickpurge PCR Cleanup column (Qiagen). UBX binding was performed by incubating ~40,000 cpm of labeled DNA with 3.3, 10, 30, 60 and 90 ng of purified UBX homeodomain (gift of Phil Beachy) in 1× DNasel reaction buffer (40 mM Tris pH 8.0, 10 mM MgSO4, 1 mM CaCl2) for 30 minutes. For digestion, 5 µl DNasel (1:40 dilution of Promega RQ1 DNasel in DNase reaction buffer) was added; reactions were allowed to incubate for 2 minutes, and were stopped by addition of 140 µl 20 mM EGTA (pH 8.0). Reactions were extracted with phenol:chloroform, precipitated in ethanol, resuspended in formamide loading buffer and separated on an 8% polyacrylamide sequencing gel. G+A sequencing ladder was generated as previously described (Maxam and Gilbert, 1980).

Amplification of Drosophila spp. knot regulatory sequences
Genomic DNA was isolated from additional species of Drosophila obtained from the Tucson Drosophila Stock Center: D. mauritiana, D. malerkotliana, D. biarmipes, D. pseudoobscura and D. viridis. Ten to fifteen flies suspended in homogenization buffer [10 mM Tris pH 7.8, 50 mM NaCl, 10 mM EDTA, 5% (w/v) sucrose] were crushed with a pestle. Lysis buffer (300 mM Tris pH 9.5, 100 mM EDTA, 0.625% SDS, 5% sucrose) and RNaseA (50 µg/ml final concentration) were added, and the homogenate was incubated at 70°C for 15 minutes. One-tenth volume sodium acetate was added and the mixture was incubated on ice 30 minutes. After pelleting debris, genomic DNA was extracted with phenol:chloroform, precipitated in ethanol and resuspended in TE. PCR amplification was performed with Turbo polymerase: forward primer 5′-GTCACCTTGATCCT-GCAATGG-3′; reverse primer 5′-GGATTTGCTTGGGGAATTG-3′. Amplified fragments were A-tailed and cloned into pGEM-T-Easy for sequencing. Sequence alignments were generated using CLUSTALW (Thompson et al., 1994) and then adjusted by hand.

Results
Identification of a wing-specific knot regulatory element
The knot gene is expressed in the developing Drosophila wing imaginal disc at the anteroposterior compartment boundary, but detected with rabbit anti-Kn antibody provided by Michèle Crozatier (Crozatier and Vincent, 1999). Engrailed protein was detected with mouse monoclonal antibody 4F11 provided by Nipam Patel (Patel et al., 1989).
is not expressed in the haltere imaginal disc (Fig. 1). Furthermore, knot expression is genetically downstream of Ubx, as overexpression of UBX in clones in the wing causes cell-autonomous loss of knot expression (Fig. 1C) (Galant et al., 2002). Because these features make knot a candidate for direct regulation by UBX, we set out to identify the regulatory element that controls knot expression in the wing.

One regulatory element of knot had been previously identified that drives reporter gene expression in the embryonic head and mesoderm. This element extends ~5 kb from the transcriptional start site of knot (Crozatier and Vincent, 1999). When a knot cDNA was placed under control of this element, embryonic lethality was rescued, but wing vein defects were not (Vervoort et al., 1999), indicating that the wing regulatory element is located elsewhere. The lesion underlying the wing-specific knot$^{ex2}$ allele, which is a complete translocation with a breakpoint 10-20 kb 5′ of the knot-coding region (Seecoomar et al., 2000), suggested the location of a wing-specific regulatory element.

Based on the location of the knot$^{ex2}$ lesion, we generated reporter constructs with genomic DNA from the region 5-20 kb 5′ of knot (Fig. 2A). We identified a 6.8 kb region of DNA ~15 kb 5′ of knot that drove expression of lacZ in a stripe at the anteroposterior compartment boundary in the wing imaginal disc (data not shown), consistent with the expression pattern of the Knot protein. No expression of lacZ was observed in the haltere, demonstrating that this large region accurately recapitulates the expression and regulation of the endogenous knot gene. To determine if both wing and haltere regulation was confined to a single region within this 6.8 kb, we further narrowed the activity to a 2.3 kb region that drove appropriate reporter gene expression (data not shown). All subsequent numbering of constructs is in reference to this 2.3 kb region, knMel1-2330.

Sequence conservation has been successfully employed in...
the identification of regulatory elements (Wasserman et al., 2000; Yuh et al., 2002). We attempted to use conservation to direct the further dissection of \( kn^{Mel701-2330} \) to define a minimal regulatory element. Based on several scattered blocks of sequence conservation between \( D. melanogaster \) and \( D. pseudoobscura \), we designed PCR primers to amplify sequence from a more distantly related fruit fly, \( D. virilis \). Three conserved blocks were shared between these three flies, and we split \( kn^{Mel701-2330} \) into two overlapping constructs (Fig. 2B), each containing two of the conserved blocks. Though both constructs included the central conserved block that contained several potential sites for regulators, only the 1.3 kb \( kn^{Mel701-1991} \) construct was capable of driving expression in a stripe in the wing (Fig. 2C), whereas the \( kn^{Mel224-1426} \) construct was only weakly expressed in a small spot at the intersection of the DV and AP axes in the wing (data not shown). Therefore, the 1.3 kb region accurately recapitulates the knot expression pattern (Fig. 2D) in both the wing and haltere, and must contain binding sites for the regulators that generate this pattern.

**A single Ci binding site mediates activation in the wing**

Expression of the knot gene is dependent on Hedgehog (Hh) activity, and overexpression of Hh can trigger ectopic knot expression in the wing (Vervoort et al., 1999). The transcriptional effector of Hh signaling is the Cubitus interruptus (Ci) protein. Ci is a zinc-finger transcription factor of the Gli family, and binds a 9 bp consensus sequence (Von Ohlen et al., 1997). In the 1.3 kb \( kn^{Mel701-1991} \) fragment, we identified three potential Ci-binding sites that matched at least seven out of nine consensus residues and that were conserved in \( D. pseudoobscura \) (Fig. 3A). Two additional potential sites were present, but were not conserved in \( D. pseudoobscura \). We mutagenized the three conserved binding sites independently, converting a crucial guanine to an adenine (Zarkower and Hodgkin, 1993), and reintroduced the mutagenized element into flies. Changes at two of the three candidate sites had no effect on reporter gene expression (data not shown), whereas the mutation of site Ci1680 almost completely abolished reporter expression (Fig. 3B). Mutation of all three sites did not have a more severe effect than mutation of Ci1680 alone (data not shown). These results indicate that activation of the wing-specific enhancer element by Hh signaling is dependent primarily on a single Ci-binding site at position 1680 in the \( kn^{Mel701-1991} \) element.

**UBX repressor sites are physically separable from activator sites**

Because the \( kn^{Mel701-1991} \) element drives expression in the wing, but not the haltere, we postulated that this sequence integrates information from Hh signaling and the homeotic regulator, UBX. Therefore, we attempted to identify possible binding sites for UBX within this element. Isolated UBX homeodomain binds optimally in vitro to the sequence TTAATGG (Ekker et al., 1991), but binding sites in characterized UBX-responsive regulatory elements often are not exact matches to this optimal sequence (Capovilla et al., 1994; Galant et al., 2002; Vachon et al., 1992). Therefore, we searched for the TAAT core sequence commonly bound by homeodomain proteins. The \( kn^{Mel701-1991} \) fragment contains clusters of TAAT core sequences near both its 5’ and 3’ limits (Fig. 4A) that might mediate knot repression in the haltere. In addition, there is a single TAAT core sequence located within 10 bp of the crucial Ci-binding site and in a conserved block of sequence, suggesting that it may be important for repression by UBX.

To determine which TAAT sequences might be functionally important for UBX repression, we removed sequences from each end of \( kn^{Mel701-1991} \) and observed the effect on reporter gene expression in vivo. Removal of the 5’ end, with its small cluster of four core sequences, had no effect on expression. By contrast, removal of 156 bp from the 3’ end, including nine putative UBX-binding sites (\( kn^{Mel701-1835} \)), caused the reporter to be expressed at the AP compartment boundary in both the wing and the haltere (Fig. 4B). Therefore, \( kn^{Mel701-1991} \) does appear to be directly negatively regulated by UBX in the haltere, and removal of UBX-binding sites relieves repression in the haltere. In addition to the ectopic activation of expression in the haltere, we noted that the expression level in the wing is also elevated compared with \( kn^{Mel701-1991} \) (Fig. 2C, Fig. 4B), suggesting that additional repressor binding sites important for appropriate wing expression may have been removed in \( kn^{Mel701-1835} \). Importantly, the response to local spatial information within the wing field (encompassing both
the wing and haltere) was maintained, as expression was appropriately observed at the AP compartment boundary in both tissues. Because the single deletion preserved the response to spatial information within the dorsal appendage wing field but altered the response to spatial information along the anteroposterior axis, we suggest that activation by Ci and repression by UBX are mediated through physically separable sites within knot cis-regulatory sequences.

To identify which potential binding sites could be occupied by UBX in vitro, we performed DNaseI footprinting on a 392 bp fragment (knMel1599-1991) that includes the functional Ci site and the 156 bp required for repression in the haltere. This fragment is itself capable of driving expression in the wing, although at a significantly lower level than that driven by the full knMel701-1991, and is repressed in the haltere (data not shown). We identified four regions protected from DNaseI digestion by binding of UBX (Fig. 4C,D; sites 1-4). These four regions include all TAAT core sequences present in the 392 bp fragment (10 in total).

Although UBX site 1 is located only 4 bp from the Ci-binding site, it is still present in the knMel701-1835 construct that is derepressed in the haltere, so this site alone is not sufficient to mediate repression by UBX. To determine whether this site is necessary for repression by UBX, we mutated UBX site 1 (knMel701-1991UBX1KO) and did not observe any derepression of reporter gene expression in the haltere (Fig. 5A). Therefore, UBX Site 1, unlike individual UBX-binding sites in the spalt enhancer (Galant et al., 2002), does not appear to contribute significantly to repression of this element by UBX. Of the other regions protected by UBX, the largest spans six TAAT core sequences and 24 bp of sequence, and is located ~250 bp from the Ci binding site. Therefore, the DNA sequences necessary for repression in the haltere appears to be comprised of multiple, functional UBX-binding sites that do not overlap with the activating Ci-binding site. This organization suggests that UBX does not repress knot in the haltere by competing for activator binding sites.

**Role of UBX binding sites and additional regulators**

Individual UBX-binding sites can additively contribute to repression in the haltere of the sal wing-specific regulatory element (Galant et al., 2002). To determine how individual UBX binding sites in the knot element contribute to repression in the haltere, we independently mutated TAAT core sequences in UBX site 1 (Fig. 5A), UBX sites 2 and 3 (data not shown), and UBX site 4 (data not shown) in knMel701-1991, and introduced this construct (knMel701-1991KO) into flies. Elimination of all UBX sequences resulted in de-repression in the haltere (Fig. 5B), demonstrating that some combination of these sites is required for repression in vivo. However, we noted that the level of expression of this construct was lower than that observed in the deletion construct, knMel701-1835 (Fig. 4B, Fig. 5B). This difference was not expected and suggested the presence of additional regulatory sequences that contribute to repression in the haltere.
To determine where additional potential regulatory sequences are located, we restored sequence 3′ of the knMel701-1835 construct. Addition of 43 bp (knMel701-1878) was sufficient to partially restore repression in the haltere (Fig. 5C), suggesting the additional regulatory information was contained within this region. Deletion of this block of sequence (knMel701-1991Δ) resulted in very weak, inconsistent de-repression in the haltere. By contrast, point mutations introduced at positions 1834-1837 (knMel701-1991mut), the boundary of the derepressed knMel701-1835 construct, resulted in consistent, partial, de-repression (Fig. 5D). As this position is not a UBX site, this result suggests that at least one transcription factor acts in addition to UBX to repress knot in the haltere through this regulatory element. Mutation of both positions 1834-1837 and all UBX TAAT core sequences (knMel701-1991Kmut) resulted in full de-repression in the haltere through these sequences located between knMel1835-1991. The DNA sequence at knMel1834-1837 does not clearly match any binding sites archived in transcription factor databases, and as yet we do not know the identity of the factor that may act with UBX to repress knot in the haltere.

Identification of a functional repressor element in D. pseudoobscura

To understand how UBX-regulated target gene networks evolve, it is crucial to determine how UBX regulation of individual target genes evolves. We combined our dissection of the knot wing regulatory element with comparative genomics within Drosophila to establish how UBX-responsive regulatory sequences in knot have evolved. We compared the 156 bp knot repressor element from D. melanogaster to D. pseudoobscura sequence, and did not observe either significant sequence conservation or a comparable cluster of potential UBX-binding sites in D. pseudoobscura. Because the expression pattern of knot is the same between these two species (Fig. 1), these significant sequence differences suggest that regulation by UBX is mediated through different regulatory sequences in D. pseudoobscura. Therefore, we attempted to identify a functional regulatory element from D. pseudoobscura that could regulate reporter expression in the appropriate pattern.

Using blocks of sequence identity as relational anchor points, we amplified a fragment from D. pseudoobscura (knMel1-1935) that roughly corresponded to the knMel1-2330 D. melanogaster fragment (Fig. 6A). We introduced this fragment into D. melanogaster and found that it could properly drive expression in the wing while repressing expression in the haltere (Fig. 7A). The knMel1-1935 construct contained at its 3′ end a cluster of 12 TAAT UBX core binding sites. To determine if this region is important for repression by UBX in D. pseudoobscura, we generated a truncation of knMel1-1935 that eliminated the TAAT core sequences. This knMel1-1643 construct appropriately drove expression in the wing, but now also drove haltere expression (Fig. 7B). Therefore, the region containing these putative UBX-binding sites acts as a repressor element in the haltere.

Interestingly, this functional cluster of UBX-binding sites is conserved between D. pseudoobscura and D. melanogaster, and is located ~500 bp 3′ of the knMel1835-1991 sequence necessary for repression, just 3′ of the limit of the 6.8 kb fragment we originally isolated that contains the functional D. melanogaster knot regulatory element. Therefore, the knot regulatory region in D. melanogaster could potentially contain two sets of functional repressor input sites. To determine whether this second, conserved block can also function to repress the D. melanogaster knot regulatory element, we attached the D. melanogaster sequence to the de-repressed knMel701-1835 construct. Addition of 222 nucleotides (knMel2499-2722), homologous to the D. pseudoobscura sequence necessary for repression, to knMel701-1835 (to generate knMelcomposite) restored repression in the haltere (Fig. 7C). Therefore, D. pseudoobscura has a single element (located between knMel1643-1935) that represses expression of knot in the haltere, and this element is shared with D. melanogaster. However, D. melanogaster possesses a second element (located between knMel1835-1991), not shared with D.
Evolution of knot cis-regulatory sequences

We next sought to determine whether UBX-binding sites in the knMel2499-2722 conserved element are sufficient to repress reporter expression, or whether this element also requires the action of a collaborating repressor. We mutated all UBX core binding sites in this sequence and attached the mutated knMel2499-2722KO sequence to the de-repressed knMel701-1835 (generating knMelcompositeKO). Whereas mutation of UBX sites alone in knMel701-1991KO did not fully de-repress in the haltere, mutation of UBX sites in knMelcompositeKO was sufficient for complete de-repression in the haltere (Fig. 7D). Thus, the knMel2499-2722 and knMel1835-1991 repressor elements appear to be organized differently – the former with input only from UBX, and the latter with input from UBX and an additional trans-acting factor.

Does the presence of two elements in D. melanogaster indicate the acquisition of a new element in this lineage or the loss of an element in D. pseudoobscura? To analyze the distribution of these two regulatory elements in other Drosophila species: D. melanogaster, D. mauritiana, D. biarmipes, D. malerkotliana and D. pseudoobscura. (A) Schematic of a 1.9 kb fragment of D. pseudoobscura DNA, indicating the position of putative Ci-binding sites in blue and UBX TAAT core sequences in brown. This fragment (knPse1-1935) and a derivative that truncates ~300 nucleotides from the 3' end (knPse1-1643) were cloned from D. pseudoobscura and injected into D. melanogaster.

(B) Alignment of knot enhancer sequences from five Drosophila species: D. melanogaster, D. mauritiana, D. biarmipes, D. malerkotliana and D. pseudoobscura. (Above) Aligned block 1, which contains the footprinted UBX binding sites (red), the functional Ci binding site (blue) and the mutated site at the boundary of the knMel701-1835 construct (purple). (Below) Aligned block 2, with conserved TAAT sequences indicated in red. Arrows in aligned block 2 indicate the boundary of the knMel701-1835 construct (purple). (Below) Aligned block 3, with conserved UBX sites indicated in green. Arrows in aligned block 3 indicate the boundary of the knMelcomposite construct (purple).

Fig. 6. Functional UBX binding sites are not conserved within Drosophila (A) Schematic of a 1.9 kb fragment of D. pseudoobscura DNA, indicating the position of putative Ci-binding sites in blue and UBX TAAT core sequences in brown. This fragment (knPse1-1935) and a derivative that truncates ~300 nucleotides from the 3' end (knPse1-1643) were cloned from D. pseudoobscura and injected into D. melanogaster.

(B) Alignment of knot enhancer sequences from five Drosophila species: D. melanogaster, D. mauritiana, D. biarmipes, D. malerkotliana and D. pseudoobscura.
In Drosophila, we amplified the knot regulatory region from three additional Drosophila species – D. mauritiana, D. biarmipes and D. malerkotliana – phylogenetically intermediate between D. melanogaster and D. pseudoobscura (Schawaroch, 2002). All three species have sequence similar to knPse1-1935 (Fig. 6B), but also possess sequence similar to knMel1835-1991 in varying degrees. For example, the core TAAT of UBX site 3 is shared by all three additional species (though sequence surrounding the core is non-identical), whereas UBX site 2 is found only in D. mauritiana. The most interesting pattern is observed for UBX site 4. D. malerkotliana has only a single core UBX sequence conserved with D. melanogaster, D. biarmipes has two conserved core sequences and two additional core sequences that are unique, and D. mauritiana has five of the six core sequences present in D. melanogaster. Therefore, in this sample of five drosophilid species, we observe the pattern of an apparent accretion of UBX-binding sites in this region in the evolution of the D. melanogaster lineage.

**Discussion**

We have identified a wing-specific cis-regulatory element for the gene knot. This regulatory element is activated in the wing by direct input from Ci and is repressed in the haltere by direct input from UBX. The regulatory sequences governing activation and repression are physically separable, and the repression element was found not to be shared with D. pseudoobscura. We identified a distinct functional repression element in D. pseudoobscura that is shared with D. melanogaster, indicating that the entire knot wing regulatory region in D. melanogaster contains two apparently redundant repressor elements. One element appears to have been acquired in the course of the evolution of the D. melanogaster lineage. Our results suggest that complete functional cis-regulatory elements, the units of function that selection is operating upon, may be larger and more diffuse than the minimal functional sequences typically defined by molecular dissection.

**Mechanism of UBX repression**

Owing to their low DNA-binding specificity and paucity of known direct targets, mechanisms for the selection of specific target genes by Hox proteins remain to be fully explained. Much work has focused on the role of co-factors in increasing the binding specificity of their Hox partners. When Hox proteins interact with PBC and MEIS proteins, represented in Drosophila by EXD and HTH (Chan et al., 1997; Gebelein et al., 2002; Ryoo and Mann, 1999; Ryoo et al., 1999), the resulting compound-binding sites are of sufficient size and information content so as not to appear by random chance at high frequency in the genome. However, neither EXD nor HTH are necessary for development of the haltere, so the action of UBX in this tissue must be independent of these co-factors (Azpiazu and Morata, 1998; Azpiazu and Morata, 2000). Repression of spalt gene expression by UBX in the haltere depends upon multiple individual UBX monomer-binding sites, (Galant et al., 2002) rather than compound binding sites.
In addition, a DNA sequence that binds neither Hox proteins nor Hox-PBC dimers determines specificity of Deformed or Labial regulation of a Deformed autoregulatory element, but the identity of this co-factor is unknown (Li et al., 1999).

Our functional analysis of the knot regulatory element is consistent with UBX repression occurring through monomer sites. UBX-binding sites in the sal1.1 and knot minimal enhancers cannot be aligned beyond the TAAT core, and so neither suggest the role of a common DNA binding co-factor. However, mutation of the identified UBX binding sites alone did not result in full de-repression of the knot minimal element in the haltere. Rather, full de-repression required mutation of additional sites not bound by UBX. This sequence may bind either a bona fide co-repressor that interacts with UBX to repress target genes or a protein that independently, but additively, contributes to repression. Because kn\textsuperscript{Dmel}701-1835 drives a higher level of expression in the wing than kn\textsuperscript{Dmel}701-1991, this putative repressor may act in both the wing and the haltere.

Analyses of both the sal and knot regulatory regions suggest that UBX may be a weak repressor that requires the collaboration of other factors, which may act in the wing and haltere to regulate other features of these tissues, in order to mediate full repression. However, as mutation of UBX sites alone in kn\textsuperscript{Dmel}2499-2722 is sufficient for de-repression, UBX may, in some contexts, be able to mediate full regulatory activity on its own. Flexibility in the organization of UBX-responsive enhancers may be due to the unsystematic, undesigned assembly of regulatory elements during evolution.

Regulatory elements that are cobbled together, incorporating binding sites for multiple collaborating transcription factors to take advantage of an existing landscape of developmental regulators, appear to be common. In the developing Drosophila embryo, both UBX and ABD-A repress the target gene Distalless (Dll) in abdominal segments, limiting leg development to the thoracic segments (Gebelein et al., 2002; Vachon et al., 1992). Repression of Dll also requires the action of the compartment-specific regulators, Engrailed and Sloppy-paired (Gebelein et al., 2004), in collaboration with the Hox proteins. In addition, the Hox protein Labial interacts with the Decapentaplegic (Dpp) signaling pathway to direct appropriate expression of the lab550 autoregulatory enhancer element in the Drosophila embryo (Marty et al., 2001), and Abdominal-A similarly collaborates with Dpp signaling to regulate wingless expression (Grienenberger et al., 2003). Collaboration may be a common requirement for Hox-regulated enhancers. Thus, rather than being highly potent regulators, Hox proteins may be weak regulators that employ a variety of collaborative factors in order to perform their function. The ability of Hox proteins to act as either repressors or activators of target genes may be regulated by interactions with different collaborators (Li and McGinnis, 1999; Li et al., 1999).

Furthermore, the weak activity of UBX and the potential requirement for collaborators for Hox repression of target genes may help to explain why it has not been possible to impart UBX regulation to a naïve cis-regulatory element by the addition of UBX monomer binding sites. Extensive efforts in this laboratory have placed multiple UBX-binding sites in various positions in cis-regulatory elements active in the wing and haltere, but with no effect (C. M. Walsh and S.B.C., unpublished). The separability of Ci activator binding sites from UBX repressor binding sites in the knot regulatory element demonstrates that in this enhancer UBX does not repress by direct competition for activator binding sites, and suggests that distance of UBX-binding sites from activator binding sites is not the cause for this failure. If UBX is such a weak repressor that UBX-binding sites alone, even in multiple copies, are not sufficient to impart repression, then the proximity of binding sites for collaborating repressor proteins may be a crucial determinant.

**Conservation, redundancy, and the unit of selection in cis-regulatory elements**

To better understand how UBX regulates morphology, we would ideally like to know all target genes on which it acts and the DNA regulatory sequences through which it exerts this control. Characterization of these regulatory sequences would elucidate the rules governing transcriptional regulation and how modification of regulatory sequences can occur during evolution. Our knowledge of the organizational constraints on regulatory sequences and how evolution operates within those constraints to maintain enhancer function is limited. Several analyses indicate that sequence within regulatory elements can vary even when function is maintained (Hancock et al., 1999; Ludwig et al., 1998; McGregor et al., 2001; Shaw et al., 2002). Nevertheless, sequence conservation between related organisms can successfully identify regulatory sequences in some lineages (Wasserman et al., 2000; Yuh et al., 2002). However, 98% of non-exonic multi-species conserved sequences within mammals do not correspond to known regulatory elements (Thomas et al., 2003). We can either suppose that these sequences primarily represent regulatory elements yet to be functionally characterized or that sequence conservation alone is not an indicator of regulatory function.

Our dissection of the knot regulatory region provides examples of apparently redundant binding sites for individual transcription factors, apparently redundant functional repressor elements, and sequence conservation without obvious biological function. For example, of three conserved putative Ci-binding sites, each contained within larger blocks of sequence conservation, only Ci1680 is necessary for activation of knot in the wing field. This observation suggests several possible interpretations. First, the other Ci sites may be functioning in a different context – a different tissue, for example – than examined in our assay, and selection has maintained these sites for that additional role. However, even the large 6.8 kb knot regulatory fragment did not appear to drive lacZ reporter expression in a limited set of additional tissues surveyed (data not shown), so we do not have any positive evidence supporting its role elsewhere in development. Similarly, the conserved blocks could represent binding sites for other factors, with conservation a consequence of maintaining those regulatory sites rather than the Ci sites. Next, it is possible that the evolutionary distance between D. melanogaster and D. pseudoobscura is not appropriate for addressing the relationship between sequence conservation and functional consequence. However, as this distance is approximately equivalent to the distance of the human-mouse comparison, we must then infer significant differences in the dynamics of sequence evolution within these two lineages. Finally, the additional Ci sites may contribute to regulation in the context of the wing to a degree that we are unable to detect,
but that purifying selection does act upon, and it is this view that we favor.

The apparent redundancy of UBX repressor elements in the D. melanogaster knot regulatory region also requires explanation. The accretion of UBX sites in the knot regulatory region in our sample of species phylogenetically intermediate to D. pseudoobscura and D. melanogaster suggests that a novel UBX-responsive element has evolved. Given the presence of a pre-existing, functional sequence that is maintained in both D. melanogaster and D. pseudoobscura, how has selection maintained the conserved element and allowed expansion of the novel element? Dissection of the eve stripe 2 regulatory element in both D. melanogaster and D. pseudoobscura demonstrated that compensatory evolution could lead to turnover of individual binding sites, resulting in a regulatory element with conserved function in the absence of sequence conservation (Ludwig et al., 1998; Ludwig et al., 2000). However, compensatory evolution that maintains repression of knot in the haltere does not seem to be the solution, as the downstream element is still present and therefore presumably capable of repressing knot expression. We conclude that the minimal element we identified in our functional assay is not necessarily identical to the functional unit upon which selection acts. That is, selection can detect and select for organismal-level effects of regulatory changes that are not obvious in our functional assay. Therefore, minimal functional regulatory elements defined by molecular dissection may not reflect the full, complete enhancers that selection has built. Rather than being sharply bounded and discrete, regulatory elements may be more diffuse collections of transcription factor inputs.

From an evolutionary standpoint, such a diffuse, flexible regulatory architecture seems a necessity. If particular precise arrangements of transcription factor binding sites are required to produce a transcriptional output, the probability of evolving a novel functional regulatory element by point mutation is exceptionally low. If, instead, a weak regulator, as UBX appears to be, collaborates with a factor already operating on an enhancer, then a novel output may be generated that may be reinforced by selection. This reinforcement may eventually result in a more precise, optimized arrangement of binding sites and a more robust regulatory output.

We thank J. Hooper for the GST-Ci plasmid; M. Crozatier, N. Patel and R. White for antibodies; P. Beachy for purified UBX homeodomain; R. Galant and C. Walsh for sharing unpublished results; J. Yoder and B. Williams for helpful comments on the manuscript; and K. Vaccaro for technical assistance. B.H. is supported by NIH NRSA F32GM65737 and S.B.C. is an Investigator of the Howard Hughes Medical Institute.

References


Evolution of knot cis-regulatory sequences


