Transcription, Chromatin, and Epigenetics: Multiple Intrinsically Disordered Sequences Alter DNA Binding by the Homeodomain of the Drosophila Hox Protein Ultrabithorax

Ying Liu, Kathleen S. Matthews and Sarah E. Bondos
doi: 10.1074/jbc.M800375200 originally published online May 27, 2008

Access the most updated version of this article at doi: 10.1074/jbc.M800375200

Find articles, minireviews, Reflections and Classics on similar topics on the JBC Affinity Sites.

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2008/05/28/M800375200.DC1.html

This article cites 97 references, 33 of which can be accessed free at
http://www.jbc.org/content/283/30/20874.full.html#ref-list-1
Multiple Intrinsically Disordered Sequences Alter DNA Binding by the Homeodomain of the Drosophila Hox Protein Ultrabithorax*

Ying Liu†, Kathleen S. Matthews‡, and Sarah E. Bondos†§†§

From the †Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005 and the ‡Department of Molecular and Cellular Medicine, Texas A & M Health Science Center, College Station, Texas 77843-1114

Received for publication, January 15, 2008, and in revised form, April 21, 2008 Published, JBC Papers in Press, May 27, 2008, DOI 10.1074/jbc.M800375200

During animal development, distinct tissues, organs, and appendages are specified through differential gene transcription by Hox transcription factors. However, the conserved Hox homeodomains bind DNA with high affinity yet low specificity. We have therefore explored the structure of the Drosophila melanogaster Hox protein Ultrabithorax and the impact of its nonhomeodomain regions on DNA binding properties. Computational and experimental approaches identified several conserved, intrinsically disordered regions outside the homeodomain of Ultrabithorax that impact DNA binding by the homeodomain. Full-length Ultrabithorax bound to target DNA 2.5-fold weaker than its isolated homeodomain. Using N-terminal and C-terminal deletion mutants, we demonstrate that the YPWM region and the disordered microexons (termed the I1 region) inhibit DNA binding ~2-fold, whereas the disordered I2 region inhibits homeodomain-DNA interaction a further ~40-fold. Binding is restored almost to homeodomain affinity by the mostly disordered N-terminal 174 amino acids (R region) in a length-dependent manner. Both the I2 and R regions contain portions of the activation domain, functionally linking DNA binding and transcription regulation. Given that (i) the I1 region and a portion of the R region alter homeodomain-DNA binding as a function of pH and (ii) an internal deletion within I1 increases Ultrabithorax-DNA affinity, I1 must directly impact homeodomain-DNA interaction energetics. However, I2 appears to indirectly affect DNA binding in a manner countered by the N terminus. The amino acid sequences of I2 and much of the I1 and R regions vary significantly among Ultrabithorax orthologues, potentially diversifying Hox-DNA interactions.

Development of all bilaterally symmetric animals requires reliable temporal and spatial regulation of gene expression by members of the Hox protein family. Hox proteins are expressed in contiguous domains along the anterior-posterior axis, where they regulate region-specific differentiation, patterning, and proliferation (1–6). Misexpression of a Hox protein transforms one region into another, altering tissue and appendage fates. These dramatic phenotypes underscore the absolute requirement for specific and reliable Hox function in vivo.

To regulate development, individual Hox family members control transcription of unique batteries of downstream genes (6). However, the homeodomain of each Hox protein mediates DNA binding with high affinity but extremely low specificity (7–10). In Drosophila, the amino acid sequences of Hox homeodomains range from 73 to 98% identical (supplemental Fig. 1), with more similar homeodomain sequences derived from Hox proteins with adjacent expression domains (2, 11, 12). Although homeodomain-DNA contacts are largely conserved (13), the more dissimilar homeodomains exhibit some differences in DNA affinity and specificity (11, 14). For instance, the Ultrabithorax (Ubx) and Deformed homeodomains bind with a less than 3-fold difference in affinity to the Ubx optimal DNA site (15). This affinity difference is greatly amplified by altering the pH or osmotic strength (15, 16).

The DNA binding properties of Hox homeodomains with higher sequence identity appear indistinguishable. For example, the Ultrabithorax and Antennapedia homeodomains are 90% identical, including all DNA-contacting residues. Consequently, DNA affinity and specificity are also very similar (11, 17, 18). However, in vivo Ultrabithorax and Antennapedia regulate different genes and drive development of unique body structures; Drosophila midthoracic legs and wings are formed within the Antennapedia expression domain, whereas development of halteres and the posterior-most pair of thoracic legs from analogous tissues requires Ultrabithorax (2, 6, 19). This disparity between the absolute requirement for distinct Hox activities in vivo and the similarity of homeodomain-DNA recognition in vitro has been termed the “Hox paradox” (12).

This paradox is resolved, in part, through Hox interactions with other transcription factors, increasing specificity by requiring tandem Hox and partner DNA binding sites (20–24). Since the expression and activity of many Hox partners is limited to specific tissues, protein interactions potentially provide contextual information to Hox proteins as well as contribute to target site selection (22). However, a subset of Hox-regulated enhancers lack sites for known Hox partners. Thus, several laboratories, including ours, have been exploring the hypothesis that amino acid sequences outside the homeodomain contribute to differences in binding site selection by Hox proteins dur-
Intrinsically Disordered Regions Regulate Hox-DNA Binding

Intrinsically disordered regions of Hox proteins, which are known to partially overlap the transcription activation domain and modulate DNA binding by Hox proteins despite the highly conserved homeodomain sequence and structure (2, 11–13, 18), are potentially utilized for DisEMBL: Savitzky-Golay smoothing frame 8, minimum peak width 8, maximum join distance 4. The VXLT sub-domain sequences vary significantly between Hox proteins (6, 15, 16, 20, 25–28). These nonhomeodomain sequences vary significantly between Hox proteins (supplemental Fig. 2) and thus potentially permit distinct Hox functions in vivo. However, little is known about protein structure or function outside the homeodomain (18, 29), and very few quantitative studies have examined DNA binding by full-length Hox proteins (17, 30).

Applying a combination of computational and experimental techniques to the Drosophila Hox protein Ubx, we have identified several evolutionarily conserved disordered domains. Despite their lack of structure, we show that these regions modulate DNA binding by the homeodomain. Differences in DNA binding by Ubx and its homeodomain (UbxHD) as a function of pH were exploited to search for regions that influence ionizable residues in the homeodomain and thus directly impact DNA binding. We show that affinity is modulated by the less conserved, disordered regions of Ubx, providing potential mechanisms to vary DNA binding by Hox proteins despite the highly conserved homeodomain sequence and structure (2, 11–13, 18). Furthermore, Ubx sequences that modulate DNA binding partially overlap the transcription activation domain and known protein interaction domains (20, 21, 31), potentially allowing mutual influence of these functions.

EXPERIMENTAL PROCEDURES

Structure and Disorder Predictions—The online prediction algorithms IUPred (32), PONDR (33), and DisEMBL 1.4 (34) were used to identify potentially disordered regions of Ubx. For IUPred, any score greater than 0.5 was considered disordered. The loops/coils definition and the following parameters were utilized for DisEMBL: Savitzky-Golay smoothing frame 8, minimum peak width 8, maximum join distance 4. The VXLT sub-domain searching for long disorder was applied in PONDR. Regions predicted to have secondary structure were identified using the nnpredict and GOR V (Garnier-Osguthorpe-Robson) algorithms (35, 36).

Construction of UbxIa, UbxIa Deletion Mutants, and UbxHD—The UbxIa gene was cloned into both pET3c and pET19b vectors (Novagen) by PCR with primers that introduce an NdeI site at the 5′-end and a BamHI site at the 3′-end following two stop codons. The UbxIa N-terminal truncation mutants were produced from pET19b-UbxIa by engineering a second NdeI restriction cutting site using the QuickChange site-directed mutagenesis kit (Stratagene), excising the intervening sequence by NdeI digestion, and ligating the remaining plasmid. The mutant that also had a C-terminal truncation (N235C24) was constructed by mutating amino acids 357 and 358 to stop codons in the N235 mutant.

Wild-type UbxHD was produced from the original pET3c-HD expression construct, which includes two copies of the UbxHD gene (17). To simplify mutagenesis, one of the homeodomain genes was first deleted. The three histidine residues in the remaining UbxHD were then mutated to lysine residues to form the HD-HK mutant. All gene sequences were confirmed by sequencing.

Protein Expression and Purification—UbxIb, UbxIa, and all mutants were expressed in Escherichia coli BL21(DE3)pLysS cells as described previously (37) with the following minor variations. Cells expressing UbxIb, UbxIa, and UbxIa deletion mutants were harvested 105 min after induction, whereas cells expressing UbxHD or HD-HK were collected 120 min after induction. Cells were resuspended in ~5 ml of collection buffer (500 mM NaCl, 50 mM NaH2PO4, pH 8.0, 5% glucose, one-twelfth tablet of Complete Proteinase Inhibitor Mixture (Roche Applied Science), 1 mM dithiothreitol (DTT) and frozen at −20°C. Untagged UbxIa and UbxIb were purified as described for UbxIb previously (37), with the exception that they were eluted from the phosphocellulose column with a 0.2–0.9 M NaCl gradient.

UbxIa and all UbxIa truncation mutations, excepting UbxIb, were expressed from the pET19b plasmid to provide a His10 tag at the UbxIa N terminus. A 1-liter cell pellet was thawed at room temperature and lysed in 10–15 ml of buffer containing 800 mM NaCl, 10 mM β-mercaptoethanol, one-quarter tablet of Complete Proteinase Inhibitor Mixture (Roche Applied Science), 5% glucose, 5 mM imidazole, and 50 mM NaH2PO4, pH 8.0. Released DNA in the cell lysate was digested with DNase I (20 mg/ml, 40 μl) prior to centrifugation at 25,000 × g for 20 min. The supernatant was mixed with polyethyleneimine (50% (w/v), 200 μl) and centrifuged. Nfic2+ -nitrilotriacetic acid-agarose resin (4 ml; Qiagen) was preequilibrated with wash buffer containing 500 mM NaCl, 10 mM β-mercaptoethanol, 5% glucose, 5 mM imidazole, and 50 mM NaH2PO4, pH 8.0, and gently rocked with cleared cell lysates at 4°C for 10 min before being poured into the column. Packed resin was washed with 80 ml of wash buffer containing 5 mM imidazole followed by 80 ml of wash buffer containing 20 mM imidazole, and finally 40 ml of wash buffer containing 80–100 mM imidazole. Protein was eluted with wash buffer containing 200 mM imidazole.

To purify UbxHD and HD-HK, a 1-liter cell pellet was lysed in 10–15 ml of 50 mM Tris-HCl, pH 8.0, 800 mM NaCl, 4 mM DTT, one-quarter tablet of complete proteinase inhibitor mixture, and 20 mM EDTA. To digest genomic DNA, 40 μl of 20 mg/ml DNase I and 1 ml of 1.6 M MgCl2 were added, followed by centrifugation at 25,000 × g for 20 min. PEI (50% (w/v), 200 μl) was added to the supernatant, and the sample was recentrifuged. Protein samples were diluted to 100 ml with buffer Z (10% glycerol, 1 mM DTT, 50 mM Tris-HCl, pH 8.0) and gently rocked with cleared cell lysates at 20°C. Untagged UbxIa and UbxIb were purified as described previously (37), with the exception that they were eluted from the phosphocellulose column with a 0.2–0.9 M NaCl gradient.

Aggregation Assay—The aggregation assays of N149 in protein storage buffer and Tris DNA binding buffer (see below) at pH 7.5 with no bovine serum albumin were performed as described previously (37).

Native State Proteolysis—Proteolysis experiments were similar to those described by Hubbard and Beynon (38). Each reac-

2 The abbreviation used is: DTT, dithiothreitol.
Intrinsically Disordered Regions Regulate Hox-DNA Binding

UbxIb Contains Evolutionarily Conserved Intrinsically Disordered Domains—Six isoforms are generated from the 
ubx gene in vivo by the selective inclusion of three microexons (44, 45). The UbxIb isoform contains all possible Ubx sequences and was therefore used to computationally and experimentally assess any possible structure (or lack thereof) outside of the homeodomain (44, 45). Indeed, the presence of intrinsic disorder is suggested by the extremely high glycine content (17%) of Ubx, generally around 80% and never less than 60%). Data from at least three experiments, using protein from at least two purifications, were averaged to generate each reported K_d.

RESULTS

UbxIb Contains Evolutionarily Conserved Intrinsically Disordered Domains—Six isoforms are generated from the 
ubx gene in vivo by the selective inclusion of three microexons (44, 45). The UbxIb isoform contains all possible Ubx sequences and was therefore used to computationally and experimentally assess any possible structure (or lack thereof) outside of the homeodomain (44, 45). Indeed, the presence of intrinsic disorder is suggested by the extremely high glycine content (17%) of Ubx, compared with ∼7% in an average protein (46–49). The entire activation domain is especially glycine-rich (nearly 27% glycine), whereas the activation-enhancing region of the activation domain is even further enriched (32% glycine) (Fig. 1A). The range of conformations available to glycine increases the flexibility of glycine-containing polypeptides and tends to disrupt secondary structure (50–52). Consequently, glycine residues are often associated with loops or mobile regions that lack regular structure (53, 54).
Intrinsically Disordered Regions Regulate Hox-DNA Binding

Native state proteolysis experiments were used to determine whether these regions are, in fact, disordered. A protease will act primarily at unfolded, solvent-accessible peptides at least 10 amino acids long and containing its recognition sequence (38). Thus, unfolded or disordered regions can be identified by their extreme sensitivity to proteases relative to stable, folded proteins (55, 56). Three proteases with different target sequences were tested to probe different sites within the UbxIb sequence and to ensure that global results are not dependent on the optimal buffer of the protease. In order to monitor the progress of UbxIb proteolysis, the enzyme/substrate ratio had to be reduced to 1:1000 (w/w), a low ratio even for disordered proteins (17). Comparison of ellipticity per residue (Fig. 1F) clearly demonstrates that UbxHD has the higher stability of homeodomains (57, 58), together with the availability of a crystal structure for UbxHD (18), suggests that this region should be protected from proteases. Indeed, this residual band reacts with the anti-homeodomain antibody FP3.38 (59), demonstrating that the homeodomain’s structure protects it from these proteases (data not shown). To further test whether the entire homeodomain is resistant to proteases, wild-type UbxHD was purified and tested in the proteolysis assay (Fig. 1C). This domain was largely resistant to all of the proteases for the entire 30-min assay, although trypsin was able to cleave the N-terminal arm. Therefore, the Ubx regions most sensitive to proteolysis are located outside the homeodomain.

The low enzyme to substrate ratio required for UbxIb proteolysis experiments suggested that its intrinsically disordered regions are unusually flexible. To determine the extent of UbxIb flexibility relative to previously characterized proteins, Lac repressor (LacI) and apomyoglobin (ApoMb), were subjected to proteolysis under identical conditions (Fig. 1, D and E). The F helix is unfolded in ApoMb, and the hinge region and parts of the DNA-binding domain are unfolded in LacI in the absence of its nucleic acid ligand. At the low protease concentrations used for UbxIb, ApoMb was unaffected by proteolysis, despite being sensitive to both trypsin and chymotrypsin when the enzyme/substrate ratio is 10-fold higher (60). Likewise, the hinge linker region of LacI is considered extremely sensitive to proteolysis (61, 62). However, LacI was minimally cut only by Proteinase K, whereas UbxIb was largely degraded by all three enzymes. The density of potential cutting sites in both LacI and ApoMb is similar to or greater than that in the predicted disordered regions of UbxIb; therefore, their reduced proteolysis cannot be attributed to a lack of susceptible sites. Together, these results demonstrate that UbxIb is both predominantly disordered and far more flexible than either LacI or ApoMb.

CD was used to compare structure in UbxIb and UbxHD (Fig. 1F). Both proteins contain α-helices but not β-sheet, as demonstrated by negative ellipticity at 208 and 222 nm but not at 218 nm. This result was anticipated for UbxHD based on previous crystallography data (18). UbxIb appears to contain little nonhelical structure. The stronger ellipticity of UbxIb at 222 nm relative to 208 nm suggests the presence of coiled coil, a structural feature previously predicted to exist in all fly Hox proteins (17). Comparison of ellipticity per residue (Fig. 1F) clearly demonstrates that UbxHD has the higher percentage of structured residues. These data indicate that significant portions of UbxIb must be intrinsically disordered, a conclusion corroborated by negative ellipticity at 200 nm, a characteristic of disordered proteins (63). Both CD spectroscopy and native
Intrinsically Disordered Regions Regulate Hox-DNA Binding

state proteolysis indicate substantial intrinsically disordered regions within full-length UbxIb.

Three algorithms were used to predict the boundaries of intrinsically disordered domains based on the amino acid sequence characteristics unique to these regions (32–34, 64). PONDRI is a neural network algorithm trained to recognize features, including amino acid composition, complexity, charge, hydrophobicity, and flexibility, which signal a potential disorder domain (33). IUPRED identifies a reduction in the total interaction energy of all possible pairs of amino acids within a region, another characteristic of unfolded domains (32). DisEMBL detects regions likely to be coils or loops by comparing sequence with regions lacking coordinates in crystal structures (34, 64). Despite the fact that these programs search for different sequence characteristics, all programs agree on the presence of a large disordered region encompassing most of the activation domain, as well as smaller regions near the N terminus and between the YPWM motif and the homeodomain (Fig. 2, A and B). The latter region lacks coordinates in an x-ray crystal structure of a portion of Ubx, including the YPWM motif, a shortened linker region, and the homeodomain (18). Furthermore, several secondary structure prediction algorithms consistently predict structural elements only in regions not predicted to be disordered (31) (Fig. 2B).

To experimentally verify these boundaries, each gel band in the proteolysis assays was probed by Western blotting using a series of primary antibodies whose epitopes collectively span the Ubx sequence (39, 59) (Fig. 2C). By comparing the size of the band with the epitopes encompassed by that fragment, the location of its defining proteolytic sites could be identified. All but two of the potential trypsin target sites were either categorized as susceptible to or protected from proteases (Fig. 2C). Similar results were achieved with chymotrypsin (data not shown), indicating that the differences in buffer and pH required by these enzymes do not significantly alter the boundaries of the disordered domains. The coincidence of computational and experimental methods in identifying intrinsically disordered regions suggests that these results are not an artifact of protein handling. Thus, the specific, reliable, context-dependent functions of Hox proteins must be mediated by a protein that combines a low specificity DNA binding domain with highly flexible regions outside this domain.

FIGURE 2. Location of disorder in UbxIb. A–C, concurrence of computational and experimental approaches on the localization of disordered regions in UbxIb. A, the disorder prediction scores from PONDRI (blue), DisEMBL (pink), and IUPRED (yellow) are plotted against UbxIb amino acid number. Regions that consistently have scores greater than 0.5 (boundary marked by a dotted line) are predicted to be disordered. B, schematic of the UbxIb sequence with functional domains indicated as follows: transcriptional activation domain, divided into the requisite core region and the activation enhancing region (medium gray); the YPWM Exd interaction motif (Y; light gray), the three alternatively spliced microexons b, mI, and mII (dark gray), the homeodomain (black), and a partial repression domain (medium gray box near the C terminus) (31, 75, 76). Regions of UbxIb predicted to be disordered by all three algorithms are indicated by red lines above the schematic, whereas regions predicted to be ordered using the nnpredict and GOR V algorithms (35, 36) are indicated by light green lines. Note that the predicted structured and disordered domains do not overlap. C, to experimentally corroborate the predicted boundaries of the disordered segments, Ubx proteolytic products, separated by SDS-PAGE, were probed by Western blotting using primary monoclonal antibodies to identify different regions of the UbxIb sequence as follows: Ia2D.3 (N-terminal epitope indicated by a black line), Ia2D.7 (yellow line), J39.2 (turquoise line), 3.11F (blue line), 4A.1 (light gray line), and FP3.38 (C-terminal epitope marked by a purple line) (31, 58). Analysis of band size and epitope content allowed classification of most of the potential trypsin target sites in UbxIb. Scaled to the UbxIb schematic in B, filled circles indicate a site cut by trypsin, open circles mark sites protected from trypsin proteolysis by protein structure, and gray circles mark sites that could not be assessed with confidence. The theoretical analysis and experimental protease susceptibility data concur, indicating clearly that a large portion of the UbxIb amino acid sequence is intrinsically disordered.

To experimentally verify these boundaries, each gel band in the proteolysis assays was probed by Western blotting using a series of primary antibodies whose epitopes collectively span the Ubx sequence (39, 59) (Fig. 2C). By comparing the size of the band with the epitopes encompassed by that fragment, the location of its defining proteolytic sites could be identified. All but two of the potential trypsin target sites were either categorized as susceptible to or protected from proteases (Fig. 2C). Similar results were achieved with chymotrypsin (data not shown), indicating that the differences in buffer and pH required by these enzymes do not significantly alter the boundaries of the disordered domains. The coincidence of computational and experimental methods in identifying intrinsically disordered regions suggests that these results are not an artifact of protein handling. Thus, the specific, reliable, context-dependent functions of Hox proteins must be mediated by a protein that combines a low specificity DNA binding domain with highly flexible regions outside this domain.

Approaches to Quantitatively Measure DNA Binding by Full-length UbxIa and Its Variants—Intrinsic disorder is more prevalent among transcription factors than other types of proteins (65). Furthermore, transcription factors with a large number of DNA binding targets are more disordered than transcription factors with a limited repertoire, suggesting that disorder plays a regulatory role in target site selection (66, 67). Given that individual Hox proteins regulate a broad array of target genes (6, 68), these extensive disordered regions in Ubx potentially regulate DNA binding.

If regions outside the homeodomain modulate DNA binding, significant differences in UbxHD and Ubx interactions with DNA should be observable. Methods to quantitatively measure DNA binding affinity by full-length Ubx were consequently developed. DNA binding studies focus on the Ubx mRNA splicing isofrom la (henceforth referred to as “UbxIa"), to allow our in vitro measurements to be correlated with previous in vivo observations of gene regulation by wild-type UbxIa and analogous Ubx mutants. The UbxIa isofrom contains all possible Ubx sequences except for the 9-amino acid “b element” microexon (Fig. 2B) (45). Using a filter-based aggregation assay (37), we optimized buffer conditions and purification protocols to generate pure, active, full-length UbxIa, UbxIa deletions, and the Ubx homeodomain. We then measured the equilibrium dissociation constant, Kd, for each protein. The DNA concentration
guide point mutagenesis. Further, the presence of unstructured regions renders site-specific methods challenging. We therefore located regulatory regions by measuring DNA binding affinity for a series of UbxIa deletion mutants. This approach parallels many in vivo and in vitro studies of eukaryotic transcription factor function (69–74). UbxIa deletions were designed to iteratively remove sections roughly 40 amino acids in length (Fig. 3C). The deletion boundaries do not bisect potential structural or functional elements, including the homeodomain, the YPWM motif, a series of moderately conserved motifs (75, 76), or the alternatively spliced microexons (39, 44, 45). Successive removal of each section tests whether that region enhances, inhibits, or has no effect upon DNA binding.

UbxIa aggregation is sensitive to buffer conditions and protein handling, and aggregation is generally dependent on amino acid sequence (37, 77). Therefore, truncated UbxIa variants are likely to differentially respond to (and potentially aggregate in) the multiple buffer changes required for purification of untagged, full-length protein. Since an N-terminal histidine tag does not significantly alter either UbxIa-DNA affinity or pH response (supplemental Fig. 3 and Table 1), we added histidine tags to the deletion mutants to standardize purification conditions. Barring the initiator methionine, each deletion mutant is named for the terminus removed and the first amino acid present after the deleted sequence (Fig. 3C).

The aggregation propensity, activity (percentage of protein capable of binding DNA), and DNA binding affinity for each deletion mutant were determined. One deletion, N149, aggregated under multiple conditions and was therefore not investigated further. The remaining proteins were both soluble and active, indicating that the structure and DNA binding capacity of the homeodomain were not compromised.

We sequentially evaluated increasingly long UbxIa variants to determine which nonhomeodomain regions impact DNA affinity (Fig. 3D and Table 1). A truncation removing the N-terminal 235 amino acids binds with a 2-fold decrease in affinity compared with UbxHD. Since the N235 truncation contains additional sequences on both the N-terminal and C-terminal sides of the homeodomain, a second mutant was constructed (N235C24) that removes the additional C-terminal residues. The binding affinities of N235 and N235C24 are similar, suggesting the region between amino acid 235 and the start of the homeodomain, which includes the YPWM Exd interaction motif and the disordered microexons (Fig. 2B), can inhibit DNA binding 2-fold. We defined this region as a weak inhibition...
Intrinsically Disordered Regions Regulate Hox-DNA Binding

TABLE 1
DNA binding affinity and pH dependence of full-length Ubxla, Ubxla deletions, and UbxHD

The equilibrium dissociation constant ($K_d$) was measured in TBB (see “Experimental Procedures” for details). The reported $K_d$ values are the average from at least nine repeats, along with the S.D. value for all of the replicates. Ubxla and all of its variants listed below were soluble and active. The total number of charged residues at pH 7.5 in each deletion mutant is indicated.

<table>
<thead>
<tr>
<th>Ubx construct</th>
<th>$K_d$ pH 7.5</th>
<th>$K_d$ pH 6.0</th>
<th>$K_{6.0/7.5}$/K_{6.0/6.0}</th>
<th>Charged residues remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pH 7.5</td>
<td>pH 6.0</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>UbxHD</td>
<td>63 ± 24</td>
<td>3.5 ± 0.8</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>HD-HK</td>
<td>55 ± 12</td>
<td>3.2 ± 1.1</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>N216C24</td>
<td>140 ± 60</td>
<td>25 ± 4.5</td>
<td>5.6</td>
<td>15</td>
</tr>
<tr>
<td>N235</td>
<td>130 ± 49</td>
<td>17 ± 7.6</td>
<td>7.6</td>
<td>28</td>
</tr>
<tr>
<td>N196</td>
<td>190 ± 58</td>
<td>33 ± 0.9</td>
<td>5.8</td>
<td>36</td>
</tr>
<tr>
<td>N174</td>
<td>5200 ± 2000</td>
<td>770 ± 200</td>
<td>6.8</td>
<td>41</td>
</tr>
<tr>
<td>N139</td>
<td>3700 ± 2000</td>
<td>740 ± 310</td>
<td>5.0</td>
<td>44</td>
</tr>
<tr>
<td>N88</td>
<td>2400 ± 570</td>
<td>3500 ± 500</td>
<td>0.7</td>
<td>53</td>
</tr>
<tr>
<td>N49</td>
<td>2500 ± 220</td>
<td>2500 ± 160</td>
<td>1.0</td>
<td>61</td>
</tr>
<tr>
<td>N19</td>
<td>1500 ± 120</td>
<td>1400 ± 540</td>
<td>1.1</td>
<td>69</td>
</tr>
<tr>
<td>Ubxla (untagged)</td>
<td>160 ± 33</td>
<td>160 ± 70</td>
<td>1.0</td>
<td>70</td>
</tr>
<tr>
<td>His-Ubxla</td>
<td>180 ± 30</td>
<td>110 ± 23</td>
<td>1.6</td>
<td>70</td>
</tr>
<tr>
<td>Ubx-18</td>
<td>104 ± 50</td>
<td>98 ± 29</td>
<td>1.1</td>
<td>68</td>
</tr>
<tr>
<td>Ubx-42</td>
<td>368 ± 132</td>
<td>430 ± 72</td>
<td>0.9</td>
<td>65</td>
</tr>
</tbody>
</table>

* Amino acids 286–356.

region, I1. Our results for deletions of full-length Ubx are consistent with previous data demonstrating an impact of the YPWM region on DNA binding by truncated versions of the *Drosophila* Labial and Sex Combs Reduced proteins (25, 28).

The DNA binding affinities of both N216 and N235 are similar; therefore, amino acids 216–235 do not significantly impact binding. Indeed, multiple point mutations in this region did not harm the ability of full-length Ubx to regulate an enhancer-reporter construct, which relies upon Ubx-DNA binding, in *Drosophila* cell culture (31). In contrast, the DNA binding affinity of N174 is strikingly lower (~40-fold) than that of N235.

Strong inhibition of DNA binding by amino acids 174–216, which we define as inhibition region 2 (I2), is caused neither by aggregation nor lower protein activity. The I2 region is intrinsically disordered and is comprised of part of the core activation domain (31), potentially allowing cross-regulation of DNA binding and transcription regulation.

Iterative inclusion of amino acid sequences from 1 to 174 gradually improved DNA binding affinity, although amino acids 49–88 appear to have less impact. Intriguingly, this effect is roughly linear with the number of amino acids removed (Fig. 3D, $R^2 = 0.92$), although the error in measuring DNA binding for very short sequences may obscure nonlinear effects.

Given that much of this region is also disordered, the absence of structure may explain the linear correlation of function with sequence length. In general, we conclude that amino acids 1–174 restore most of the DNA binding affinity affected by the I1 and I2 regions. Thus, the majority of the Ubxla amino acid sequence either binds DNA or modulates DNA binding affinity.

Determining Mechanisms for DNA Binding Modulation by I1, I2, and R—There are at least two general mechanisms by which I1 and/or I2 could inhibit DNA binding in the N-terminal deletion mutants. In the first scenario, an inhibitory domain could prevent UbxHD-DNA contacts or destabilize UbxHD via native interactions with the homeodomain (Fig. 4A, R-independent mechanism). In this mechanism, the R domain would independently act on the homeodomain to restore the binding and/or stability of UbxHD. In contrast, in the second scenario, the inhibitory domain
prevents UbxHD-DNA interaction in a nonnative manner that occurs only in the absence of the R domain (Fig. 4A; R-dependent mechanism). R-dependent inhibition could occur through at least two further mechanisms. In the full-length protein, R may bind to I1 and/or I2 to prevent nonnative homeodomain interactions (Fig. 4B; inhibition by nonnative I-HD interaction). Alternatively, the disordered character of I1 and/or I2 could allow these regions to fluctuate throughout the space surrounding the homeodomain and thus inhibit the approach of DNA in deletion mutants. In this scenario, the R domain restricts the conformations available to I1 and/or I2 by interacting either with an inhibitory region or with sequences C-terminal to that inhibitory region (Fig. 4B; inhibition by conformational fluctuations).

These mechanisms can potentially be distinguished by elucidating intramolecular interactions involving I1, I2, R, and the homeodomain. However, direct inquiry of intramolecular interactions in UbxIa is difficult, since a full-length protein structure is unavailable to guide interpretation. Any direct interaction between UbxIa-derived peptides would need to be interpreted with extreme caution, since these potential regulatory regions exhibit significant intrinsic disorder, and amino acids 1–216, encompassing both I2 and R, promote UbxIa aggregation.3

For these reasons, we have distinguished between these potential regulatory mechanisms by detecting the consequences of intramolecular interactions rather than probing for the interactions themselves. The general mechanisms (Fig. 4) differ in the dependence of the inhibitory domain on the presence of the R domain. Therefore, internal deletion mutants were created to determine whether I1 and I2 repress DNA binding in the presence of the R region. The I1 internal deletion mutant, Ubx-18, is missing 18 amino acids, including the YPWM motif and its surrounding residues. The gain in affinity caused by this internal deletion (1.4-fold; UbxIa versus Ubx-18) is similar to the gain caused by deleting this region from the N terminus (2.1-fold; N235C24 versus HD). Therefore, the effect of I2 on UbxHD-DNA interaction is clearly dependent on the R domain.

Similarly, we created a second internal deletion mutant, Ubx-42, to remove the more strongly inhibiting I2 domain (amino acids 174–216). In contrast to the I1 results, removal of I2 did not improve DNA binding affinity. In fact, binding was actually less effective than in the wild-type protein, suggesting that nonnative interactions between R and the homeodomain occur in the absence of I2. Therefore, the effect of I2 on UbxHD-DNA interaction is clearly dependent on the R domain.

The second experiment probed whether I1 and I2 directly impact key amino acids in the UbxHD-DNA interface by monitoring DNA affinity as a function of pH. DNA binding by UbxHD is strongly dependent on pH, peaking at pH 6.0 with a roughly 20-fold higher affinity than at pH 7.5 (Fig. 5, A and B, and Table 2). Intriguingly, mutation of the three homeodomain histidine residues to lysine does not alter this behavior.

3 A. Greer and S. Bondos, unpublished observations.
Intrinsically Disordered Regions Regulate Hox-DNA Binding

TABLE 2
pH dependence of DNA binding by full-length UbxIa and UbxHD

<table>
<thead>
<tr>
<th>pH 5.0</th>
<th>pH 6.0</th>
<th>pH 7.0</th>
<th>pH 7.5</th>
<th>pH 8.0</th>
<th>pH 9.0</th>
<th>pH 10.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>UbxIa</td>
<td>540 ± 240</td>
<td>160 ± 70</td>
<td>260 ± 130</td>
<td>160 ± 33</td>
<td>110 ± 40</td>
<td>120 ± 40</td>
</tr>
<tr>
<td>UbxHD*</td>
<td>25 ± 2.6</td>
<td>2.0 ± 0.3</td>
<td>32 ± 3.8</td>
<td>50 ± 2.1</td>
<td>63 ± 5.5</td>
<td>79 ± 8.2</td>
</tr>
</tbody>
</table>

*The binding affinities of UbxHD measured at pH 5.0–10.0 were from Li et al. (15). Comparable binding affinities at pH 6.0 and 7.5 were reproduced by us (Table 1).

suggesting that ionization of these histidines cannot account for the increase in DNA binding at pH 6.0 (Fig. 5, B and C, and Table 1). Therefore, other ionizable groups with shifted pKa values may contribute to both UbxHD-DNA interaction and the pH effect. This shift in pKa is expected to be dependent on and extremely sensitive to the local chemical environment. If an inhibitory region interacts directly with the homeodomain to prevent key contacts or alter homeodomain dynamics (Fig. 4B), then the pKa of these sensitive ionizable amino acids should be altered (inhibition via nonnative I-HD interaction). In contrast, if the approach of DNA to the homeodomain is affected by R-dependent conformational fluctuations, inhibition by this mechanism should not alter the microenvironment surrounding the ionizable amino acids in the homeodomain.

To determine whether non-homeodomain regions might impact ionizable groups, we first determined whether DNA binding by full-length UbxIa was affected over the broad pH range in a manner different from UbxHD (Fig. 5 and Table 2). Any change would suggest that at least one of the homeodomain’s ionizable amino acids energetically or chemically interacts with nonhomeodomain regions. Indeed, although the pH profile of UbxHD-DNA affinity reaches a maximum at pH 6.0, Ubx-DNA affinity remains constant from pH 10 to pH ~6.0, below which the affinity gradually weakens. Therefore, UbxIa regions outside the homeodomain mask or dominate the pH sensitivity of the homeodomain.

The region mitigating the pH response was located by assessing the ratio of DNA binding affinities for the UbxIa deletion mutant series at pH 7.5 and 6.0, thus normalizing for inherent changes in DNA affinity. The Kd(7.5)/Kd(6.0) ratio for homeodomain was 18, whereas that for full-length UbxIa was 1. In contrast, the UbxIa variants parse into two groups, each with ratios different from UbxHD (Table 1 and Fig. 5). N19, N49, and N88, like full-length UbxIa, do not respond to pH. This lack of pH response does not correlate with binding affinity, which is dramatically decreased for N49 and N88. The affinity ratio was 5 for N139, indicating that amino acids 88–139 only partially restrict the homeodomain’s response to pH. This moderate level of pH dependence (Kd(7.5)/Kd(6.0) = 5–8) was also observed for deletions with low DNA affinity (N139 and N174) as well as those with high affinity (N216 and N235). Therefore, the pH effect reports on a different aspect of DNA binding than affinity modulation.

The second significant difference in affinity ratios occurs between UbxHD and N235, which includes the I1 region (N-terminal to UbxHD), and an additional 24 residues C-terminal to the homeodomain. These C-terminal 24 amino acids in UbxIa do not modulate DNA binding, since their removal, creating the N235C24 mutant, alters neither the binding affinity nor the pH ratio (Table 1).

Surprisingly, the partial I1 internal deletion mutant, Ubx-18, has a pKa ratio similar to that of full-length Ubx. Therefore, either residues 88–139 within the R region are sufficient to mask the influence of I1, or a portion of I1 not deleted in the Ubx-18 mutant is responsible for ameliorating the homeodomain’s pH effect.

We conclude that the I1 region (amino acids 235–286) has multiple roles: binding affinity inhibition and altered pH response, suggesting that I1 directly inhibits homeodomain-DNA interaction by altering homeodomain structure or binding energetics by a mechanism independent of the N terminus. I1 is adjacent to the N-terminal arm of the homeodomain, a portion of which is also disordered and thus unresolved in the crystal structure of the Ubx-Exd-DNA complex (18). The remainder of the N-terminal arm makes key specific DNA contacts in the minor groove (78, 79). Furthermore, differences in binding between the Ubx and Dfd homeodomains can be traced to this region (15, 80, 81). We hypothesize the I1 region impedes DNA binding by propagating disorder along the backbone to the N-terminal arm. The I1 region could impact homeodomain energetics by becoming ordered upon DNA binding, a phenomenon previously observed in deletion mutants of other Hox proteins (18, 28).

In contrast, the I2 domain does not contribute to the relative pH insensitivity of UbxIa-DNA binding, since Ubx variants shorter than N174 but longer than the HD all have the same affinity ratio. Furthermore, I2 requires the absence of the N terminus to inhibit binding by the homeodomain, again suggesting that I2 does not directly impair the homeodomain in the native UbxIa monomer. Although the possibility remains that I2 may interact with the homeodomain using an alternate surface that does not perturb the ionizable groups with shifted pKa values, a region able to elicit a 40-fold effect on affinity would be unlikely to interact with the homeodomain without impacting these very sensitive ionizable amino acids. Therefore, an indirect mechanism of DNA binding inhibition, in which I2 sterically impedes DNA binding due to rapid structural fluctuations around the homeodomain, appears the most probable explanation for its behavior. This conclusion is supported by the identification of intrinsic disorder throughout most of I2. Likewise, the R region may abrogate the negative effects of I2 on DNA binding via intramolecular interactions with I2 or a region C-terminal to I2, thus restricting the flexibility of this potentially inhibitory region. In support of the latter hypothesis, residues 88–139 of R do alter the pH response of DNA binding and thus directly impact the structure or energetics of the home-
odomain. In the absence of I2, the R region, which is also predominantly disordered, may similarly impede DNA approach, thus accounting for the reduction in DNA affinity by the Ubx-I2 internal deletion mutant. Finally, DNA binding or protein interactions may further restrict or even abrogate conformational fluctuations in the I2 and R regions.

**DISCUSSION**

**Implications for Interpreting DNA-binding Experiments**

Despite the importance of the Hox protein family, quantitative DNA binding data on full-length Hox proteins are sparse. Many of the difficulties we initially encountered (purifying soluble, active, monomeric full-length protein and generating high quality, quantifiable gel shift results under nonstoichiometric conditions) can be attributed to the presence of large, fairly hydrophobic, disordered regions. Such regions are predicted to occur in most Hox proteins (Fig. 6) and may contribute significantly to the limited success of biophysical studies within this family. As a result, many laboratories have turned to alternative methods to qualitatively assess binding, including using unpurified proteins produced in vitro, deleting disordered regions, or using DNA and protein concentrations above the range applicable for quantitative measurement of affinity (28). Determination of binding constants from experiments in which the DNA concentration approaches or exceeds $K_d$ distorts the measured value, potentially by orders of magnitude, from the actual equilibrium constant. We have overcome Hox solubility and handling difficulties to provide a truly quantitative analysis of DNA binding by a full-length Hox protein.

Although transcription factors are typically viewed as loosely linked, independently acting functional domains, we have demonstrated that sequences outside the homeodomain can impact DNA binding affinity by nearly 2 orders of magnitude. Therefore, DNA binding experiments relying solely on Hox deletion mutants should be interpreted with extreme caution, since the deleted regions have the potential to profoundly impact DNA affinity as well as protein structure, activity, and stability. Thus, the 100-fold higher $K_d$ values reported for a deletion mutant of the Sex Combs Reduced *Drosophila* Hox protein (28) relative to full-length Ubx may result from the use of truncation mutants and/or the impact of DNA concentrations inappropriate for quantitative analysis.

**Roles for Intrinsic Disorder in Hox Proteins**

Much of the nonhomeodomain sequence of UbxIb is intrinsically disordered. We have demonstrated that this disorder is important for Ubx function, leading to the expectation that this feature should be conserved across multiple species. Indeed, we find that a small disordered region predicted near the N terminus is conserved in insects and crustaceans (last common ancestor 400 million years ago), and a very large disordered region near the middle of the protein is additionally predicted to be conserved in arthropods (540 million years of evolution) (Fig. 7) (75, 76), although the corresponding amino acid sequences of Hox orthologues are surprisingly divergent. For comparison, humans last shared an ancestor with chimpanzees ∼5–7 million years ago and with fish ∼450 million years ago (82). This conservation suggests that the intrinsically disordered regions N-terminal to the YPWM motif must be crucial for Ubx function. Consistent with this hypothesis, we also find disordered regions in the other seven *Drosophila* Hox proteins (Fig. 6).

Therefore, sequence conservation in Ubx occurs at three levels of stringency: (i) the highly homologous homeodomain, (ii) the more weakly conserved motifs, and (iii) the preservation of disordered character but not amino acid sequence. This conservation may provide some balance in Hox protein evolution, preserving regulation and patterning in crucial core structures (e.g., the central nervous system, muscle, and gut) while still permitting evolution of new morphologies and body plans.

In UbxIb, much of the observed disorder is found in the microexon region and the activation domain. Activation domains are frequently intrinsically disordered, a feature that may aid their interaction with general transcription factors (83, 84). Likewise, the frequent disorder in alternatively spliced regions is also likely to be necessary for protein interactions (85). Therefore, multiple protein interactions relying on disorder may be required for Ubx to regulate transcription as well as to communicate with other cell-regulatory systems. In support of this hypothesis, Ubx has been found to bind a surprisingly high number of proteins in yeast and in vitro, many of which can modify Ubx function in vivo (20, 21).

Ordering of intrinsically disordered regions is often observed upon ligand or protein interactions (56, 67). Conformational transitions can also transfer information between regions.
within a protein, allowing one molecular function to influence another (67). Thus, our observation of disordered regions in Ubx influencing DNA binding opens new avenues for dissecting both external and internal mechanisms for regulating DNA binding.

**Potential Roles of the DNA Binding Regulatory Regions in Vivo**—Although I2 and R together have little net impact on binding in vitro, these regions may nonetheless be relevant for transcription factor function in vivo. Protein interactions could alter the regulatory balance between I2 and R via a variety of mechanisms, including binding I2 and further restricting its dynamics or binding the R region and thereby freeing I2. Indeed, removal of the N-terminal 19 amino acids has dramatic effects on transcription activation of the *decapentaplegic* gene in the embryonic midgut (72). Therefore, the I2 and R regions also potentially influence animal development by altering DNA binding in vivo.

Multiple in vivo studies have identified tissue-specific effects caused by mutation in and around the YPWM motif in Hox proteins (72, 86, 87). These phenotypes were initially ascribed to an impact on tissue-specific protein interactions (88, 89). In particular, many studies focused on Exd, a Hox-interacting protein known to bind the YPWM motif. Thus, the impact of the YPWM region on DNA occupancy by Hox proteins in the presence of Exd was anticipated to be indirect, mediated by altered partner interactions (88). However, the Ubx-18 internal deletion mutant lacking the YPWM motif is less able to repress *dll*, an Exd dependent target, and *antp*, an Exd independent target in vivo (72). Given that an alternate Exd binding motif is primarily used by Ubx for *dll* repression (4), these data suggest the YPWM region impacts functions other than Exd interaction (25, 72, 90). Indeed, research from the Mann laboratory has established that the YPWM region impacts interaction with Hox-Exd composite DNA binding sites by truncation mutants.

**FIGURE 7. Conservation of the disordered regions in Ubx.** An amino acid sequence alignment is shown for Ubx derived from the fruit fly *D. melanogaster* (DmUbx), the mosquito *Anopheles gambiae* (AgUbx), the beetle *Tribolium castaneum* (TcUbx), the butterfly *Juonia coenia* (JcUbx), the shrimp *Artemia franciscana* (AfUbx), and the Onychophoran velvet worm *Acanthokara kaputensis* (AkUbx). The consensus results from the disorder prediction algorithms PONDR, IUPred, and DisEMBL are marked for each Ubx orthologue by red boxes. Identical and conserved residues are marked by black and gray shading, respectively, and previously identified conserved motifs (75, 76) are identified by dashed lines. The YPWM motif and homeodomain are marked by solid lines.
of Labial and Sex Combs Reduced, two other Drosophila Hox proteins (25, 28). We find that regulation of DNA binding by the I1/YPWM region is not an artifact of the severe N-terminal truncations used in the previous binding studies (25, 28).

Hox genes located within the bithorax complex, including *ubx*, have different roles *in vivo* than the more anteriorly expressed *antp* complex Hox genes, in which the role of the YPWM region has previously been examined (25, 28, 91). Despite these functional differences and the known variations in the conformation and function the I1/YPWM within the Hox protein family (92), we find that the YPWM region also modulates DNA-binding affinity in Ubx (25, 28). Our quantitative results also provide the first evidence that the I1/YPWM region impacts binding of a Hox protein to a monomer consensus binding sequence and Hox multimers sequences as well as its previously described role in modulating binding to Hox-Exd composite DNA binding sites (25, 28, 92). This observation suggests changes in DNA binding arising from mutations in the YPWM region may contribute to *in vivo* loss of regulation for any Hox binding sequence. Indeed, this effect has been observed for regulation of both Exd-dependent and Exd-independent targets by the Ubx-18 mutant (72). Although the YPWM region impacts both types of target DNAs, the interactions mediated by the I1/YPWM region at the atomic level may differ for monomer and composite DNA binding sites (92).

**Implications for Transcription Factors and the Hox Paradox**—The necessary balance we note between binding inhibition by the I2 region and affinity restoration by the R region in Ubx probably exists in other transcription factors as well. Intrinsically disordered domains are frequently present in transcription factors (56, 93), especially in transcription regulatory domains (84), DNA binding domains (94), and regions responsive to cell signaling (95). Failure to restrict the conformations of these regions could also inhibit DNA binding in the same indirect, nonspecific manner by which I2 appears to inhibit binding by the homeodomain. Therefore, we anticipate that most transcription factors incorporate some mechanism to restrict this motion and thus permit DNA binding.

The Hox paradox stems from the wide variety of similar DNA target sequences bound with almost identical affinity by Hox homeodomains (11, 42) and the extreme conservation of HD-DNA interactions within the Hox family (12, 13). These features raise several important questions regarding Hox function *in vivo*. (i) How do Hox proteins select a specific target when they are capable of binding a very wide array of DNA sequences? (ii) How does a single Hox protein bind to different targets in different tissues to drive context-specific gene regulatory cascades? (iii) How do Hox family members from the same organism function differently to generate different Hox-dependent morphologies? (iv) How do orthologues of a single Hox protein vary their function to specify distinct body plans?

Unlike the homeodomain, the nonhomeodomain regions that alter DNA binding have diverse amino acid sequences across this family. These differences provide a means to vary DNA recognition between different Hox members and thereby drive tissue-specific, Hox protein-specific, or orthologue-specific functions. Significant portions of these sequences contain conserved, intrinsically disordered regions. Although similar investigations of Ubx orthologues will be required to determine whether conserved disorder or variant amino acid sequence dominates homeodomain-DNA interactions, it is already clear that the complex regulatory effects observed for UbxIa create the requisite opportunity for disordered regions to diversify DNA binding by Hox proteins. Furthermore, the DNA binding regulatory regions overlap domains associated with other Hox functions (Fig. 8). Thus, these regulatory mechanisms potentially coordinate DNA binding with activities such as transcription activation or other effects mediated by protein interactions. Indeed, multiple groups have demonstrated that DNA binding alters other Hox molecular functions (69, 96). Our identification of disordered regions that modify homeodomain-DNA interactions provides a critical step in moving toward resolution of the Hox paradox at the molecular level.

**Acknowledgments**—We thank the members of the Matthews and Silberg laboratories for helpful comments and suggestions and Dr. Liskin-Berg for her critical reading of the manuscript. The UbxIa gene and the Ubx-18 mutant were the gifts of Dr. Ella Tour and Dr. Bill McGinnis (both from University of California, San Diego), and the UbxIb and homeodomain expression constructs were gifts of Dr. Philip Beachy (The Johns Hopkins University). Anti-Ubx antibodies were the gift of A. Javier López (Carnegie Mellon University).

**REFERENCES**

7. Gehring, W. J., Qian, Y., Q., Biller, M., Furukubo-Tokunaga, K., Schier,