The bodies of many metazoa consist of a series of homologous segments or metameres which bear structures to reveal the shape and logic of some parts of the program other functions. The roles genes play in elaborating this common body plan are best understood in Drosophila specialized for feeding, locomotion, reproduction, and do not perturb axial polarity or segmental organization, clusters downstream of the transcription start sites, trinucleotide TAA or the related hexanucleotide TAA-TCG. These results suggest that the regulatory activities of Ubx are direct and are mediated by binding of Ubx proteins to promoter region sequences.

**Introduction**

The bodies of many metazoa consist of a series of homologous segments or metameres which bear structures specialized for feeding, locomotion, reproduction, and other functions. The roles genes play in elaborating this common body plan are best understood in Drosophila melanogaster, where extensive genetic studies have begun to reveal the shape and logic of some parts of the program underlying development. In the first phases of this program, the embryo acquires axial polarities and partitions into segments through the agency of several classes of maternally and zygotically acting genes (Nüsslein-Volhard and Wieschaus, 1980; Anderson and Nüsslein-Volhard, 1984; Schupbach and Wieschaus, 1986; Nüsslein-Volhard et al., 1987). Subsequent specialization of these segments for diverse functions requires the cytoplasmic action of a separate class of genes which includes all those within the bithorax complex (Lewis, 1978) and some of those within the Antennapedia complex (Wakimoto et al., 1981). The homeotic mutations by which these genes are identified do not perturb axial polarity or segmental organization, but produce transformations of metameric identity, with each gene's function associated with the identity of metameres in a particular anatomical region.

**Summary**

The homeotic gene Ultrabithorax (Ubx), located in the bithorax complex of Drosophila, encodes a family of closely related proteins that direct the developmental fates of posterior thoracic and anterior abdominal metameres. We have purified a member of the Ubx protein family from an overproducing E. coli strain and have shown that it is a sequence-specific DNA binding protein. The protein binds tightly to sequences near its own promoter and near the P1 promoter of Antennapedia (Antp), a homeotic gene Ubx is known to repress from genetic studies. The binding sites occur in clusters downstream of the transcription start sites, and far upstream at Antp P1. They range in size from 40 to 90 bp, and contain tandem repeats of the trinucleotide TAA or the related hexanucleotide TAA-TCG. These results suggest that the regulatory activities of Ubx are direct and are mediated by binding of Ubx proteins to promoter region sequences.

**An Ultrabithorax Protein Binds Sequences Near Its Own and the Antennapedia P1 Promoters**

Philip A. Beachy,* Mark A. Krasnow,†
Elizabeth R. Gavis,‡ and David S. Hogness
Department of Biochemistry
Stanford University School of Medicine
Stanford, California 94305

* Present address: Howard Hughes Medical Institute and Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205.
† The order of these authors is arbitrary.

Ultrabithorax (Ubx), a homeotic gene within the bithorax complex, specifies primarily the distinguishing features of parasegments 5 (PS5) and 6 (PS6), which together comprise a four-compartment region extending from the posterior compartment of the second thoracic through the anterior compartment of the first abdominal segment (T2p and T3a → PS5, and T3p and A1a → PS6; see Figure 1). The heart of the Ubx domain is an ~77 kb transcription unit that gives rise to a set of ~3.2 and ~4.3 kb mRNAs (Beachy et al., 1985; Hogness et al., 1985). Expressed early in embryogenesis and throughout the remainder of development, these mRNAs encode at least five closely related ~40 kd polypeptides (O'Connor et al., 1988; Kornfeld et al., unpublished data). All members of this protein family share extensive amino acid sequence encoded by the extreme 5' and 3' exons of the transcription unit. Differences between members derive from differences in internal RNA splicing patterns involving one 9-codon and two different 17-codon elements.

Although at least two other transcription units are located within the Ubx domain (Lipshitz et al., 1987), several considerations indicate that it is the Ubx family of proteins that executes the functions required for metameric specialization of PS5 and PS6. First, there is a general correspondence between the wild-type distribution of Ubx proteins and the regions of the animal affected in Ubx mutants (Akam, 1983; Beachy et al., 1985; White and Wilcox, 1984). Second, mutations that prevent production of the Ubx family of proteins fully inactivate the metameric identity functions of the domain. This group includes lesions which either grossly disrupt the Ubx transcription unit or delete short stretches of coding sequence common to all members of the Ubx protein family (Bender et al., 1983). Such mutations produce homeotic transformations of PS5 and PS6 to PS4 in homozygous larvae or mosaic adults (Lewis, 1963, 1978; Minana and Garcia-Bellido, 1982). Furthermore, cis regulatory mutations that leave the Ubx protein-coding potential intact do not inactivate all metameric identity functions of the domain, but only those in regions where Ubx protein expression is reduced or eliminated. For example, abx and bx mutations, which produce homeotic transformations in PS5, are associated with a reduction of Ubx protein levels in PS5, and similarly for bxd and pbx in PS6 (White and Wilcox, 1985; Beachy et al., 1985). Finally, mutations such as Cbx', which cause spatially inappropriate expression of Ubx proteins (White and Akam, 1985), result in a dominant homeotic transformation of the affected region. Thus Ubx proteins are not only necessary for specification of the PS5 and PS6 metameric identities, but are capable of promoting these developmental fates when abnormally expressed in other positions.

The regulation of other genes has long been hypothesized as a mechanism by which single homeotic genes such as Ubx could specify the dramatic differences between metameres (Lewis, 1964). This suggestion is supported by the nuclear localization of Ubx proteins (Beachy...
An analogous repression of Ubx expression by the abdominal genes of the bithorax complex (Struhl and White, 1985) suggests that cross-regulation is an important common function of homeotic genes (in particular the repression of anteriorly acting genes by products of genes active in more posterior regions). Thus, while segmentation genes play an important early role in the establishment of homeotic gene expression (Ingham and Martinez-Arias, 1986; Scott and O'Farrell, 1986; Akam, 1987), subsequent refinement, and perhaps maintenance of these patterns, depends upon interactions between the homeotic genes themselves.

To determine the molecular basis of such regulatory interactions, we have initiated a study of the biochemical properties of Ubx proteins. Here, we report the purification of a Ubx protein produced in E. coli, and show that it interacts with DNA sequences near its own and the Antp P1 promoters. The binding sites are unusually large, and contain striking arrays of a repeating trinucleotide (TTA) or a related hexanucleotide repeat (TAATCG).

Results

High Level Expression of UBX lb

The UBX lb member of the Ubx protein family contains all three internal elements in addition to the common amino- and carboxy-terminal regions (Figure 2; Beachy et al., 1985). UBX lb was produced in E. coli and in two insect cell culture systems by introducing recombinant DNAs containing strong promoters and appropriate translational control signals fused to the UBX lb coding sequences. For the bacterial system, UBX lb coding sequences were inserted into the expression vector pAS-1 (Rosenberg et al., 1985) downstream of the bacteriophage lambda P1 promoter and the translational signals and three amino-terminal codons of the lambda cII gene (see Figure 2 and Experimental Procedures). Upon induction of the P1 promoter, E. coli cells harboring the recombinant plasmid accumulated a new species of the expected M, (~40 kd) at levels up to about 5% of total cellular protein (Figure 3A) depending on the host strain (see Experimental Procedures). This 40 kd protein was UBX lb since it was detected along with several less abundant forms of lower M, by immunoblotting with anti-Ubx antibodies (Figure 3B). The less abundant forms appear to be breakdown products of the larger protein since their presence was also dependent upon P1 promoter induction and their levels were reduced in mutant strains defective in protein degradation (see Experimental Procedures).

UBX lb was also expressed in two insect cell culture systems, one of which used the Drosophila hsp70 heat-inducible promoter to produce UBX lb in Drosophila Schneider line 2 cells (Riou and Rubin, 1985). The other entailed infection of the lepidopteran cell line S19 (from Spodoptera frugiperda) with a recombinant baculovirus containing UBX lb coding sequences under control of the promoter for the polyhedrin protein gene (Summers and Smith, 1986). In neither of these systems did levels of UBX lb protein comparable to those achieved in bacteria accumulate, although there was less degradation than in E. coli (Beachy, 1986; data not shown). These sources were useful for verifying that UBX lb from insect cells displays DNA binding properties similar to those of protein from E. coli, and in vitro studies of β-galactosidase hybrid proteins suggest a role for the homeodomain in DNA binding (Desplan et al., 1985).

More compelling evidence for the regulatory role of Ubx products comes from the distribution of Antennapedia (Antp) gene products in Ubx mutants. Antp normally specifies metameric fates in the anterior thorax and is maximally expressed there from two tandem promoters, P1 and P2 (reviewed in Scott, 1985). In Ubx mutants, however, Antp transcripts and proteins accumulate at abnormally high levels in the PS5 and PS6 portions of the thorax, thus accounting for the anteriorly directed homeotic transformations at these positions (Hafen et al., 1984; Carroll et al., 1986: see Figure 1).

Figure 1. Expression Patterns of Ubx and Antp in the Embryonic Central Nervous System

The Ultrabithorax (Ubx) domain of the bithorax complex supplies determinants of metameric identity for parasegments (PS) 5 and 6; Antennapedia (Antp), a homeotic gene within the Antennapedia complex, is primarily responsible for PS4 but also acts with Ubx in PS5 (Duncan, 1987). Ubx and Antp RNAs and proteins are expressed within the thoracic segments of the embryonic central nervous system at locations indicated by the lines labeled "UBX" and "ANTP" (Akam, 1983; Beachy et al., 1985; White and Wilcox, 1984; Hafen et al., 1983; Carroll et al., 1986). The line labeled "ANTP in Ubx" denotes Antp expression in embryos lacking wild-type Ubx function (Hafen et al., 1984; Carroll et al., 1986). Filled boxes indicate regions in which the proteins are expressed at high levels, dark stippling indicates intermediate levels, and light stippling indicates low levels of expression. Abbreviations: T1-T3, the three thoracic segments; A1-A8, the eight abdominal segments; a, anterior and p, posterior compartments (figure modified from Sanchez-Herrero et al., 1985 and Carroll et al., 1986).
DNA Binding Properties of a Purified Ubx Protein

1071

Figure 2. Construction of a Plasmid for High-Level Expression of UBX lb in E. coli

The structure of pQdm3712, a plasmid containing a Ubx lb cDNA, is shown in (A). Open reading frame lb (ORF lb) is indicated by the box; the three optional elements that distinguish the different Ubx ORFs and the homeodomain are indicated by the dark stippled boxes and the filled box, respectively. The 5' and 3' noncoding cDNA sequences are indicated by the thick line, and vector sequences by the thin line. The HindIII site just preceding the methionine initiation codon was converted to a BamHI site, and the BamHI fragment containing ORF lb was ligated into the BamHI site of the vector pAS-1. The resulting plasmid, pPL -UBX lb, shown in (B), contains ORF lb sequences downstream of the bacteriophage lambda P1 promoter and the ribosome binding site (r.b. site) and first three codons of the lambda cII gene.

Figure 3. Expression of UBX lb in E. coli

E. coli strain AR68 harboring pPL -UBX lb was grown at 30°C to an ODso0 = 0.4, and transcription from PL was induced by the addition of 5 volumes of medium prewarmed to 66°C. Growth at 42°C was continued and aliquots were removed at the times indicated above each lane. Total protein (~10 μg) from each sample was analyzed by electrophoresis through 12% polyacrylamide gels containing SDS and either stained with Coomassie blue (A) or transferred to nitrocellulose and probed with the anti-Ubx monoclonal antibody FP3.38 (6). The position of the 40 kd UBX lb is indicated by a horizontal line, and the lower molecular weight forms revealed by antibody staining are indicated by a vertical line.

Purification of a Novel DNA Binding Activity from E. coli Cells Expressing UBX lb

The altered spatial distribution of Antp products observed in Ubx mutants suggests a role for Ubx products in the regulation of Antp; if regulation is direct, it might be mediated through binding of Ubx proteins to Antp DNA sequences. To test for such an interaction, E. coli extracts containing UBX lb were incubated with a 32P-labeled DNA fragment extending from -6 to +0.8 kb with respect to the Antp Pl transcription start site and filtered through nitrocellulose under conditions where protein–DNA complexes (but not free DNA) are retained. Even with unfractonated extracts the Antp fragment was retained at least 20-fold better than the 3H-labeled pUC8 control DNA (Figure 4A), whereas both fragments were retained at similar low levels by control extracts lacking UBX lb (Figure 4B).

The filter-binding assay was used to follow the DNA binding activity during fractionation of cell extracts, and the UBX lb polypeptide was monitored by polyacrylamide gel immunoblot. Substantial enrichment for UBX lb was achieved by ammonium sulfate fractionation, and homogeneous protein was obtained by chromatography through DEAE Sephacel, phosphocellulose, and hydroxylapatite, followed by gel filtration. Details of the purification are described in Experimental Procedures, and an electrophoretic analysis of samples from various stages in the purification is displayed in Figure 5A. Throughout purification, the bulk of the Antp DNA binding activity fractionated with the full-length UBX lb protein (Figures 5B and 5C), confirming that UBX lb was the new DNA binding activity.
Figure 4. An Antp DNA Binding Activity in Extracts from E. coli Expressing UBX lb
One fmol of a 7 kb 32P-labeled Antp DNA fragment from the P1 promoter region (~6.3 kb to +0.6 kb) and 29 fmol of a 7H-labeled linear pUC8 DNA were incubated for 10 min at 37°C under standard conditions with the indicated amounts of E. coli cell extract. After filtration through nitrocellulose, the amount of each fragment retained on the filter as a result of protein binding was determined by scintillation counting. Extracts were from nalidixic acid-induced cultures of E. coli AR120 harboring either (A) the UBX lb expression plasmid pPL-UBX lb or (B) its parent plasmid pAS-1.

Definition of UBX lb Binding Sites at Antp P1
To localize UBX lb binding sites more precisely, the 7 kb Antp DNA fragment was cleaved with various restriction endonucleases, end-labeled, and tested in the filter-binding assay. Fragments retained were identified by electrophoretic analysis following their elution from the filter with SDS. The results of an experiment using fragments generated by Hinf cleavage are shown in Figure 6A. The selectivity of fragment retention increased with salt concentration of the wash, and at approximately physiological ionic strength three Antp fragments were retained (lane 5). A 472 bp fragment located 6 kb upstream of the Antp transcription start site (H472) displayed the highest affinity for UBX lb, followed by a 543 bp fragment (H543) of moderate affinity located just downstream of the start site. The third fragment, H314, of lower affinity, is located ~2 kb upstream of the start site, between the other two (see Figure 6B).

Further definition of the binding sites in these fragments was obtained by nuclease protection experiments (Galas and Schmitz, 1979), in which sequences bound by UBX lb are protected from DNAase I digestion, thus leaving a gap or footprint in the pattern of nuclease cleavages. Selected examples of such experiments are shown in Figure 7, these data are summarized in Figure 9, and the positions of the protected regions with respect to the transcription start site are illustrated in Figure 8.

Figures 7A and 7B show that H472 and H543 each contain strong binding sites for UBX lb. Within H472, three regions were protected from nuclease digestion by purified UBX lb (Figure 7A), designated A-1, A-2, and A-3. A-1 is an ~40 bp region protected at low UBX lb concentrations (lane 5). At several-fold higher protein concentrations (lane 6), the protected region expanded to include the adjacent ~40 bp region, A-2, and a new protected region appeared ~50 bp upstream, at A-3. Corresponding regions were protected when the complementary strand was labeled (Figure 9A). Within H543, ~50 (A-A) and ~60 (A-B) nucleotide regions were protected (Figure 7B), located, respectively, 290 and 385 bp downstream of the Antp P1 transcription start site (Figure 8). H314, the fragment least efficiently retained in filter-binding experiments (Figure 6), showed only several diffuse, weakly protected regions at the highest concentrations of UBX lb tested (data not shown).

Figure 7 also shows that nuclear extracts prepared from Spodoptera cells infected with the UBX lb recombinant baculovirus gave the same DNAase I protection pattern of H472 as protein purified from E. coli (compare lanes 8 and 9 with lanes 5 and 6 in A). Thus, by the criterion of DNA binding, UBX lb proteins produced in E. coli or cultured insect cells are indistinguishable.
flow-through fraction and 4 μl of the other indicated fractions is shown, and the predominant band (double-headed arrow) is the 40 kD Ubx protein. (C) Gel filtration of UBX lb. One-fourth of the hydroxyapatite pool (0.25 ml) was passed through a precalibrated Superose 12 HR 10/30 FPLC column, and 0.3 ml fractions were collected and assayed for DNA binding activity. The absorbance profile at 280 nm is shown by the thick line, and the arrows indicate the position of the calibration standards: Ex, exulin; S1, bovine thyroglobulin; S2, bovine gamma globulin; S3, chicken ovalbumin; S4, horse myoglobin; S5, vitamin B-12. The major absorbance peak, with a K\textsubscript{D} of 0.45, was UBX lb (see electrophoretic analysis above in A, lane S). The inset is a semilogarithmic plot of the K\textsubscript{D} of the standards and UBX lb vs. their relative molecular mass (M\textsubscript{r}).
UBX lb Binds Sequences Near Its Own Promoter

Segmentation genes play an early role in the activation of Ubx expression (Ingham and Martinez-Arias, 1986; Scott and O'Farrell, 1986; Akam, 1987), but these genes are, for the most part, not expressed beyond the embryonic period. Refinement and maintenance of the Ubx expression pattern during later stages of development may therefore require additional regulatory activities. An attractive mechanism for maintaining Ubx expression would be through an autoregulatory effect of a Ubx product creating a positive feedback loop. We therefore searched for UBX lb binding sites near the Ubx promoter.

Nuclease protection experiments with a 950 bp fragment (from -650 to +300 with respect to the Ubx transcription start site; Figure 7C) indeed revealed two UBX lb binding sites 40 and 220 bases downstream of the Ubx transcription start site, designated U-A and U-B, respectively (Figure 8). The U-B region is striking in its size (~90 bp), but the footprint contains several sites of enhanced nuclease cleavage (arrowheads in Figure 9E), which suggests that the region may accommodate more than one UBX lb molecule. U-A is about half the size of U-B, and also contains several positions where cleavage is enhanced by the binding of UBX lb (Figure 9D).

A Repeated Trinucleotide Common to UBX lb Binding Sites

Within each protected region are tandem arrays of a simple sequence element, the trinucleotide TAA. The most striking array is in A-I, which deviates from a perfect match at only 1 of 33 bp (Figure 9A, stippled region). Other protected regions depart from this pattern at many positions, but maintain the triplet rhythm throughout the site. At several binding sites these stretches of patterned sequence are punctuated by several nucleotides of arrhythmic sequence. Some of these breaks in pattern are associated with positions of unprotected or enhanced DNAase I cleavages, as is most evident in the middle of the U-A binding site, where the occurrence of several dG-dC base pairs coincides with enhanced cleavages on each strand (Figures 7C and 9D). The most striking deviation from the repeated TAA motif occurs at A-B (Figure 9C, stippled re-
Figure 7. DNAase I Protection Analysis of UBX lb Binding Sites

Binding reactions containing ~30 fmol of 32P 3’end-labeled DNA fragments and UBX lb protein as indicated were incubated under standard conditions, treated with DNAase I, and the digestion products visualized by autoradiography following electrophoresis in polyacrylamide/urea gels. Regions protected from DNAase I digestion by UBX lb protein are marked by the filled (fully protected regions) or striped (partially protected regions) boxes, and their designations are given to the right of each box. The scales to the left of each panel (except A) indicate the distance in nucleotides from the transcription start site, and were derived from reference ladders generated by base-specific chemical degradation (Maxam and Gilbert, 1977) or from a series of DNA size standards. In (A), the sites are ~6 kb upstream of the transcription start site, but the exact distance is unknown and the scale refers to the relative position within the fragment.

(A) The DNA was an EcoRI-HindIII fragment from pUC19-4331 containing H472 from the Antp P1 region (see Figure 6B). The fragment was labeled on the strand corresponding to the coding strand at the EcoRI site. Reactions contained no Ubx lb (lanes 3, 7, and 10), or 40, 80, and 120 ng purified UBX lb (lanes 4-6) or 1 and 4 µg of a nuclear extract from Spodoptera frugiperda cells expressing UBX lb (lanes 8 and 9). Lanes 1 and 2 display sequence ladders derived from the labeled fragment.

Under the conditions of these experiments, UBX lb was in molar excess relative to the H472 fragment, and the protein concentration required to produce half-maximal protection of the binding site may be considered an estimate of the dissociation coefficient. Assuming a native M, of 78 kd for UBX lb (Figure 5B), the reaction in lane 4 was ~10 nM in UBX lb, and the binding coefficient for A-1 can therefore be estimated at ~10−6. The actual affinity may be somewhat higher since the effect of the competitor poly d(C) is not taken into account and our UBX lb preparation may not be fully active.

(B) The reactions contained no UBX lb (−) or ~120 ng UBX lb (+), and the DNA was an EcoRI–SalI fragment from pUC 19–4330 containing H543 and labeled on the noncoding strand at the SalI site.

(C) The reactions contained 0 (lanes 3 and 6), 30 (lane 4), or 120 ng (lane 5) of UBX lb and the DNA was a HindIII–EcoRI fragment of pde5′–46 labeled on the coding strand at the HindIII site. This fragment contains sequences extending from −46 to +358 with respect to the Ubx transcription start site. Lanes 1 and 2 are sequence ladders derived from the labeled fragment.

(D) The reactions contained no UBX lb (−) or ~120 ng (+) and the DNA was a XhoI–PvuII fragment from pPB177 containing the cloned oligonucleotide (TAA)n labeled on the complementary strand at the XhoI site. The location of the repeated motif (in brackets) is shown in relation to the 21 nucleotide protected region. The dashed line indicates a weakly protected portion of the 21 nucleotide region, and the arrowhead indicates the position of an enhancement of DNAase I cleavage by UBX lb.
region), where, in the downstream portion of the binding site, TAA alternates with TCG to create the hexanucleotide motif TAATCG. This hexanucleotide is also embedded in some of the other binding site sequences, but nowhere in an extended array as at A-B, where in one part of the protected region only 3 of 36 nucleotides deviate from this pattern.

Although UBX lb binding sites are dA/dT-rich, other sequences of similar dA/dT composition were not bound in filter-binding and nuclease protection experiments. For example, just downstream of A-2, a 24 bp region containing 22 dA-dT base pairs is followed a few base pairs further downstream by a 22 bp region containing 21 dA-dT base pairs; neither sequence was protected from DNAase I cleavage by UBX lb, even at concentrations 10-fold higher than those necessary to protect A-1. These regions do not contain multiple repeats of the TAA motif, which suggests that nucleotide sequence and not simply comp-

Figure 8. Locations of UBX lb Binding Sites Near the Ubx and Antp P1 Promoters

The open boxes in the top portion of the figure show the locations of fragments H472 and H543 with respect to the Antp P1 transcription start site at coordinate 0. In the expanded diagrams below each open box, the solid and stippled boxes denote the locations of sequences fully or partially protected from DNAase I cleavage by UBX lb (see Figure 7). In the bottom half of the figure, two protected regions downstream of the Ubx transcription start site are indicated. The Antp and Ubx transcription start sites are indicated by large arrows at coordinate 0, and the small arrows show the predominant orientation (5' to 3') of the TAA repeats within each region (see Figure 9).

Figure 9. Sequences and DNAase I Protection Patterns of Naturally Occurring UBX lb Binding Sites

The results of the footprinting experiments shown in Figure 7 are summarized along with the results of similar experiments using the same fragments labeled on the complementary strands. Sequences of the binding sites and flanking regions (Laughon et al., 1988; Saari and Bienz, 1987; Wilde and Akam, 1987; Kornfeld et al., unpublished data; see Experimental Procedures) are shown. Thick lines indicate the extent of protection of each strand from DNAase I cleavage, and the dashed extensions indicate sequences protected only at higher relative levels of UBX lb. The circles denote unprotected or incompletely protected cleavage within a footprint; the arrows indicate positions of enhanced cleavage. Markings pertain to the P-O bond 3' of the indicated base except at A-A and A-B, where the assignments are accurate only to within a few nucleotides. Numberings at the first and last positions of the noncoding strand at A-A, A-B, U-A, and U-B are the nucleotide distances from the Antp P1 and Ubx transcription start sites.

The thin lines immediately above each sequence indicate positions in a footprint that conform to the TAA motif, the dots mark positions that do not match but which maintain the triplet rhythm. The regions of near-perfect TAA and TAATCG repeats in A-1 and A-B are stippled. The orientation of U-A has been inverted so that the TAA triplets appear on the top strand. (Because the A-3 protection falls in a region of few DNAase I cleavages [see Figure 7A], the boundaries of the UBX lb binding site cannot be precisely determined and these sequences were therefore not tabulated.)
position is critical for UBX lb binding, in contrast to several other dA/dT binding proteins (see Solomon et al., 1986, and references therein).

**UBX lb Binds to a Cloned Synthetic (TAA)₄ Repeat**

To determine whether the TAA motif is sufficient for UBX lb binding, we tested a cloned synthetic oligonucleotide containing four repeats of the TAA trinucleotide in DNAase I protection experiments. As shown in Figure 7D, UBX lb protected a 21 nucleotide region centered on the TAA repeat. Thus, four TAA repeats are sufficient for UBX lb binding, even outside their natural context. Since this was the minimum number of TAA repeats required for binding (P. A. Beachy and Doris von Kessler, unpublished data), most of the naturally occurring sites, where 40 or more base pairs are protected, probably accommodate several molecules of UBX lb. This suggestion is strongest at U-B, where several distinct arrays of the TAA motif are separated by short stretches of unpatterned sequence and the total protected region is four times the size of that for the (TAA)₄ sequence.

**Discussion**

Using an overproducing E. coli strain as a high level source, we have purified a member of the Ubx family of proteins, UBX lb, and have shown that it is a sequence-specific DNA binding protein. High affinity binding sites for UBX lb occur in clusters just downstream of the Antp Pl and Ubx promoters and ~6 kb upstream of the Antp P1 promoter. The sequences within these binding sites hold in common arrays of the trinucleotide TAA (with a related hexanucleotide repeat, TAATCG, also present at one site), and the cloned synthetic oligonucleotide (TAA)₄ is sufficient for binding of UBX lb. The naturally occurring binding sites most likely accommodate several UBX lb molecules as they each contain more extensive triplet arrays than the synthetic sequence and give proportionately larger UBX lb footprints (40–80 vs. 21 bp).

Although previous evidence from mutant studies demonstrated a negative effect of Ubx products on accumulation of Antp transcripts and proteins in posterior thoracic regions (Hafen et al., 1984; Carroll et al., 1986), the biochemical role played by Ubx products was not addressed. Our results support the regulatory effect is direct and mediated by the high affinity interactions of Ubx proteins with DNA sequences near the Antp P1 promoter. In addition, a direct autogenous regulatory effect is suggested by the presence of UBX lb binding sites near the Ubx promoter.

The locations of UBX lb binding sites near promoter regions suggest that Ubx proteins affect either transcription initiation or elongation, since an effect on RNA stability or turnover would additionally require the ability of Ubx proteins to interact with RNA. Indeed, the clustering of binding sites with short distances between clusters is reminiscent of the arrangement of binding sites of known transcription factors such as the mammalian Sp1 protein (see McKnight and Tjian, 1984, and references therein) and the Drosophila heat shock transcription factor (Wiederroth et al., 1987). The functional significance of such clustering is unclear, but it may allow cooperative binding or synergism in regulatory effect of bound proteins. This might be particularly important for Ubx proteins, with an estimated equilibrium dissociation coefficient of ~10⁻⁶ (see legend to Figure 7), which falls near the low end of the range of affinities (10⁻⁹ to 10⁻¹² M) reported for purified mammalian and Drosophila transcription factors (enhancer binding protein, Johnson et al., 1987; CCAAT transcription factor, Jones et al., 1997; the adenovirus major late transcription factor, Chadosh et al., 1986; and the Drosophila heat shock transcription factor, Wu et al., 1988).

The location of some of the UBX lb binding sites downstream of the transcription start site would be unusual but not unprecedented for cis-acting regulatory elements. Downstream regulatory sequences have been reported near the adenovirus major late promoter (Mansour et al., 1986; Reinberg et al., 1987), within the α- and β-globin genes (Charnay et al., 1984; Wright et al., 1984), within the cellular gene encoding thymidine kinase (Merrill et al., 1984), and for immunoglobulin genes (Gilles et al., 1983; queen and Baltimore, 1983; Banerji et al., 1983). In addition, several Drosophila promoters contain important elements located immediately downstream of the transcription start site, C. Thummel and D. S. Hogness, unpublished data; Soeller et al., 1988) including the Ubx promoter (Biggin and Tjian, 1988). Such downstream cis-acting sequences are capable of mediating both negative (for the cellular tk gene) and positive (for the other genes) regulatory effects.

Evidence supporting the proposed transcriptional regulatory role for Ubx proteins has come from recent studies with cultured Drosophila cells. Although the Schneider line 2 cells used in these studies do not produce Ubx proteins endogenously (Beachy, 1986), Ubx proteins can be produced from introduced constructs. Transfected genes containing the Antp P1 or Ubx promoter region linked to a reporter gene responded negatively (Antp P1 promoter) and positively (Ubx promoter) to Ubx protein expression, thus demonstrating that in the same cellular background a Ubx protein can have opposite regulatory effects (M. A. Krasnow and D. S. Hogness, unpublished data).

One intriguing possibility is that the opposite effects on Antp P1 and Ubx promoter fusions are due to differences in arrangement of the UBX lb binding sites with respect to each other or to some common element such as the transcription start site. For example, the distances between (125 bp for Ubx, 32 bp for Antp P1) or the relative orientations of (inverted for Ubx, direct for Antp) nearby binding sites (see Figure 8) could influence the cooperative formation of structures with positive and negative regulatory values, respectively, for the Ubx and Antp P1 binding site clusters. Alternatively, the 40 nucleotide distance between the Ubx transcription start site and the first binding site may be permissive of a stimulatory interaction between bound UBX lb and RNA polymerase or another transcription factor, which the corresponding 290 nucleotide distance at Antp P1 does not permit. The importance of the Ubx binding site sequences we
have defined is emphasized by their requirement for ectodermal expression in embryos of Ubx promoter-lacZ fusion genes (Bienz et al., 1988), and by their conservation in another Drosophila species, D. funeris. Apart from protein coding sequences, two large regions of interspecies homology have been identified near the Ubx promoter (Wilde and Akam, 1987). One of these is the UBX lb binding site U-B, which conforms to a repeated TAA pattern at 53 of 85 bp (65%) and corresponds to a similar region in D. funeris that conforms at 66 of 80 bp (83%). The other region is more highly conserved and spans the transcription start site, beginning at −32 and extending downstream to +82, including the UBX lb binding site U-A. Within this 114 bp region there are only five differences between species, with none in U-A, suggesting an extreme functional constraint. If, as we suspect, this sequence is a cis-acting regulatory region, it must be filled to the point of overlap with target sites for regulatory molecules, possibly including some which interact or compete with Ubx proteins.

Indeed, Drosophila embryo extracts contain, in addition to U-A and U-B binding activities, factors that bind sequences adjacent to U-A (Biggin and Tjian, 1988). Also, engrailed protein gives a DNAase I protection pattern almost coincident with U-A (W. Soeller and T. Kornberg, personal communication), and a β-galactosidase-ensgrailed homeodomain hybrid protein protects A-A (C. Desplan and P. O'Farrell, personal communication). This may not be surprising, given the near identity of Ubx and engrailed proteins in the regions implicated in DNA binding, and we think it likely that the binding sites will prove to be targets for several and perhaps many regulatory proteins. In this respect, they may not be unlike steroid response elements (von der Ahe et al., 1985), CCAAT promoter elements (McKnight and Tjian, 1986; Dorn et al., 1985), and several enhancer sequences (Scheidereit et al., 1987; Baldwin and Sharp, 1988) which can interact with more than one binding protein.

The sequences of the UBX lb binding sites would appear to be unusual recognition elements in that their composition is largely restricted to two nucleotides organized in a tandemly arrayed triplet unit, TAA. Despite their unusual structure, these and other trinucleotide repeats may not be uncommon recognition elements, as they are present in promoter regions of many Drosophila genes (M. A. Krasnow, unpublished data), and some have been implicated as target sites for regulatory proteins. For example, at the distal and proximal Adh promoters, (GCT)₆ sequences are bound by the Adf-1 transcription factor, and other nuclear factors interact with nearby (GCT)₆ and (TATACT)₆ sequences (Heberlein et al., 1985). These and other trinucleotide repeats (Wharton et al., 1985; Haynes et al., 1987) may have some general property that favors their use as recognition elements.

The tandem TAA repeats are also somewhat surprising in light of the apparent dimeric structure of UBX lb (Figure 5B), since a dimeric structure in prokaryotic repressor proteins has consistently been associated with binding site sequences with dyad symmetry (Pabo and Sauer, 1984). Such a symmetric structure for UBX lb is still possible if it interacts only with the outer base pairs of each repeat unit, or if only one protomer of the dimer binds locally to the repeat. A more extreme model for site recognition is that the trinucleotide repeats adopt a novel conformation which is bound by UBX lb. In order to distinguish between these and other models, it will be critical to determine the elements essential for UBX lb binding.

**Experimental Procedures**

**Plasmids**

The UBX lb expression plasmid pPB130-UBX lb was constructed by treating the 178 bp Hhal fragment from pou3712 that contains the Ubx open reading frame (ORF) lb initiation codon (Beachy, 1986; see Figure 2) with T4 DNA polymerase to remove the 3′ extensions (Challberg and England, 1989) and ligating it into the filled ends of BamHI-digested pUC8 DNA (Messing and Vieira, 1982). This construct, pPB130, contains regenerated BamHI sites flanking the insert. The complete ORF lb was reconstructed in pUC6 by ligating the Psit-Sall fragment from pou3712 into Psit-Sall-digested pBPl30 to yield pPB118, in which noncoding sequences upstream of ORF lb have been removed and the Hhal site immediately upstream of ORF lb has been replaced by a BamHI site. The BamHI fragment containing the entire UBX ORF lb was ligated into the BamHI site of pBS-1 (Rosenberg et al., 1983) in the proper orientation for expression from the P₈ promoter to yield pPl – UBX lb (Figure 2B). Two additional E. coli expression systems incorporating either the trp-lac hybrid promoter P₈ (Lam and Broiusi, 1985) or the lambda P₈ promoter (Queen, 1983) were also tested, but these accumulated UBX lb at levels one-tenth or less that of the P₈ system.

Plasmid pAc611-UBX lb was constructed by ligating the 2.2 kb BamHI-EcoRI fragment of pPB18 to the large BamHI–EcoRI fragment of pAcels (Summers and Smith, 1986), thus placing UBX lb downstream of the Autographa californica nuclear polyhedrosis virus (AcNPV) polyhedrin promoter. After cotransfection of the Spodoptera frugiperda cell line SL-1 with pAc611-UBX lb and wild-type viral DNA, the recombinant baculuvirus AcNPV-UBX lb containing the UBX insert was identified by plaque hybridization and isolated as described by Summers and Smith (1986).

The plasmid pSl9 (kindly supplied by M. Scottt), which was used in filter-binding studies, contains a genomic HindIII fragment extending from −6 kb to +0 kb with respect to the Antp P₁ transcription start site (Laughon et al., 1986). Three HindIII fragments from pSl9 (H434, H472, and H314) were subcloned by filling their 5′ extensions and ligating each to the filled ends of XbaI-digested pUC19 (Yanisch-Perron et al., 1985) to generate, respectively, the plasmids pUC19-4330, pUC19-4331, and pUC19-4332. The sequence of H472 was determined by the dideoxy chain-termination method; the sequences of H434 and H314 were previously reported by Laughon et al. (1986).

The Ubx promoter deletion plasmid, pAcSl5 – 46 (kindly supplied by M. Biggin), contains genomic sequences that extend from an artificially introduced HindIII site at −46 to a naturally occurring EcoRI site at +358 with respect to the UBX transcription start site, cloned into the vector pAc611.

Plasmid pBP777 was constructed by annealing the oligonucleotides 5′GATCCCTAATAGATAAAAG and 5′GATCTCTATTATGAT and ligating them into the BamHI site of Bluescript KS1+ (Stratagene Cloning Systems). The sequence of the cloned oligonucleotide was verified by the dideoxy chain-termination method (Sanger et al., 1977).

**Labeling and Purification of DNA Fragments**

Restriction fragments used in filter binding and DNAase I protection experiments were labeled by filling recessed 3′ ends with a 3H- or 32P-labeled dNTP using the large fragment of E. coli DNA polymerase I. For subcloning and DNAase I protection experiments, restriction fragments were purified from agarose gels.

**Preparation of Cell Extracts**

The extracts used for the experiments shown in Figures 4 and 5 were made from E. coli strain AR120 bearing pPB1-UBX lb or its parent plasmid pAC-l. After growth at 2°C to 0.5 OD₆₀₀, nalidixic acid was added to 40 μg/ml to induce P₈ transcription and growth was con-
Nitrocellulose Filter-Binding Assays

DNAase I Protection Assays

UBX lb Purification

Acknowledgments

We thank J. Berringham and M. Scott for supplying the Antp DNA clones and sequence information prior to publication, R. White and M. Wilusz for the hybrid kinase cell line FP330, M. Dugus for a useful UBX lb promoter deletion plasmid, A. Shatzman and M. Rosenberg for E. coli AR88 (kindly supplied by A. Shatzman), a strain defective in protein degradation because of the htpR185 allele (Baker et al., 1984), most of the UBX lb was insoluble. E. coli AR120 (Mott et al., 1985) accumulated soluble UBX lb to ~7% of total cell protein after nalidixic acid treatment and was used for large-scale production and for the experiments shown in Figures 4 and 6.

DNA Binding Properties of a Purified Ubx Protein

Nitrocellulose filter binding was carried out as described by Siegel et al. (1982) except that the nuclear extracts were prepared by the method of Boulton et al. (1981) and the Ubx protein was monitored by immunoblotting with the anti-Ubx monoclonal antibody FP3.8 (White and Wilcox, 1984).

DNAase I was used and the Ubx protein was monitored by immunoblotting with the anti-Ubx monoclonal antibody FP3.8 (White and Wilcox, 1984).

ACKNOWLEDGMENTS

We thank J. Berringham and M. Scott for supplying the Antp DNA clones and sequence information prior to publication, R. White and M. Wilusz for the hybrid kinase cell line FP330, M. Dugus for a useful UBX lb promoter deletion plasmid, A. Shatzman and M. Rosenberg for E. coli AR88 (kindly supplied by A. Shatzman), a strain defective in protein degradation because of the htpR185 allele (Baker et al., 1984), most of the UBX lb was insoluble. E. coli AR120 (Mott et al., 1985) accumulated soluble UBX lb to ~7% of total cell protein after nalidixic acid treatment and was used for large-scale production and for the experiments shown in Figures 4 and 6.

DNA Binding Properties of a Purified Ubx Protein

Nitrocellulose filter binding was carried out as described by Siegel et al. (1982) except that the nuclear extracts were prepared by the method of Boulton et al. (1981) and the Ubx protein was monitored by immunoblotting with the anti-Ubx monoclonal antibody FP3.8 (White and Wilcox, 1984).

ACKNOWLEDGMENTS

We thank J. Berringham and M. Scott for supplying the Antp DNA clones and sequence information prior to publication, R. White and M. Wilusz for the hybrid kinase cell line FP330, M. Dugus for a useful UBX lb promoter deletion plasmid, A. Shatzman and M. Rosenberg for E. coli AR88 (kindly supplied by A. Shatzman), a strain defective in protein degradation because of the htpR185 allele (Baker et al., 1984), most of the UBX lb was insoluble. E. coli AR120 (Mott et al., 1985) accumulated soluble UBX lb to ~7% of total cell protein after nalidixic acid treatment and was used for large-scale production and for the experiments shown in Figures 4 and 6.
strains and vectors. Verne Luckow and Max Summers for baculovirus vectors and advice on their use. D. von Kessler for technical assistance. M. Sepanski for oligonucleotide synthesis. T. Fields for photography, and T. Engler and G. Whitaker for help in preparing the manuscript. We are grateful to our many colleagues at Stanford and the Carnegie Institution for help and advice, and to A. Fire and A. Spradling for critical readings of the manuscript.

P. A. B. was a DuPont fellow of the Life Sciences Research Foundation and a fellow of the Alfred P. Sloan Foundation. M. A. K. was a fellow of the Helen Hay Whitney Foundation and is a Lucille P. Markey Scholar in Biomedical Science, and E. R. G. is a trainee of the Medical Scientist Training Program. This work was supported by a National Institutes of Health grant to D. S. H.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 18, 1988.

References


DNA Binding Properties of a Purified Ubx Protein

1081


Notes Added In Proof
