Cell, Vol. 55, 1069–1081, December 23, 1988, Copyright © 1988 by Cell Press

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

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 TAATCG. 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 zygotic 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.

* 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 sequences 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*[1], 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

et al., 1985; White and Wilcox, 1984) and by the presence within each family member of the homeodomain (Beachy et al., 1985; O'Connor et al., 1988; Kornfeld et al., unpublished data), a 60 amino acid segment conserved among the products of homeotics and some segmentation genes. The C-terminal third of the homeodomain shows some similarity to the DNA binding domains of known transcriptional regulatory proteins in bacteria and yeast (Laughon and Scott, 1984; Shepherd et al., 1984), 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).

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 (TAA) or a related hexanucleotide repeat (TAATCG).

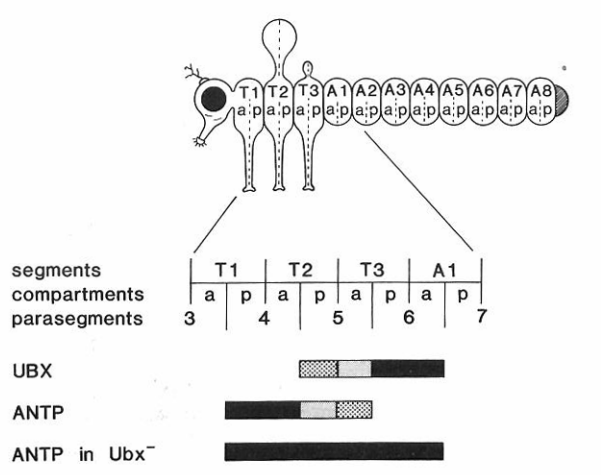


Figure 1. Expression Patterns of *Ubx* and *Antp* in the Embryonic Central Nervous System

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. Abbrevations: 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).

Results

High Level Expression of UBX Ib

The UBX Ib 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 Ib 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 Ib coding sequences. For the bacterial system, UBX Ib coding sequences were inserted into the expression vector pAS-1 (Rosenberg et al., 1983) downstream of the bacteriophage lambda P_L 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 P_L promoter, E. coli cells harboring the recombinant plasmid accumulated a new species of the expected M_r (~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 Ib since it was detected along with several less abundant forms of lower M_r 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 P_L promoter induction and their levels were reduced in mutant strains defective in protein degradation (see Experimental Procedures).

UBX Ib was also expressed in two insect cell culture systems, one of which used the Drosophila hsp70 heat-inducible promoter to produce UBX Ib in Drosophila Schneider line 2 cells (Rio and Rubin, 1985). The other entailed infection of the lepidopteran cell line Sf9 (from Spodoptera frugiperda) with a recombinant baculovirus containing UBX Ib 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 Ib 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 Ib from insect cells displays DNA binding properties similar to those of protein from E.

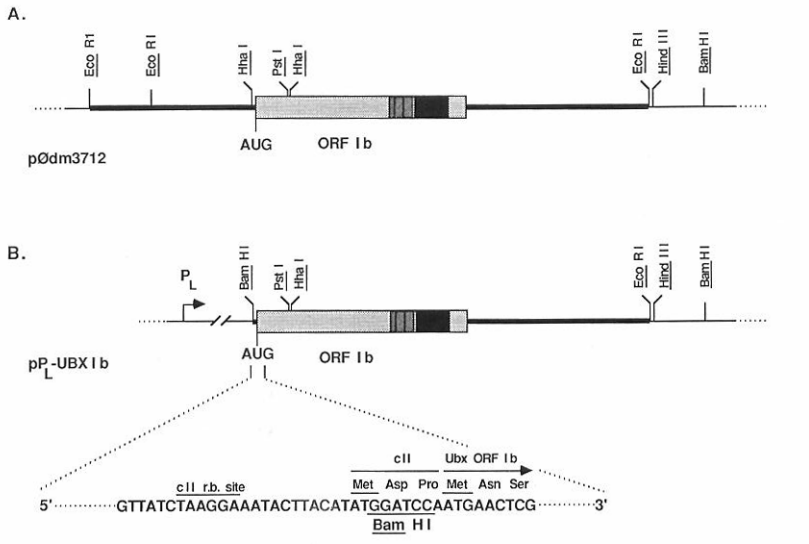


Figure 2. Construction of a Plasmid for High-Level Expression of UBX Ib in E. coli

The structure of pΦdm3712, a plasmid containing a *Ubx* Ib cDNA, is shown in (A). Open reading frame Ib (ORF Ib) 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 HhaI site just preceding the methionine initiation codon was converted to a BamHI site, and the BamHI fragment containing ORF Ib was ligated into the BamHI site of the vector pAS-1. The resulting plasmid, pP_L – UBX Ib, shown in (B), contains ORF Ib sequences downstream of the bacteriophage lambda P_L promoter and the ribosome binding site (r.b. site) and first three codons of the lambda cII gene.

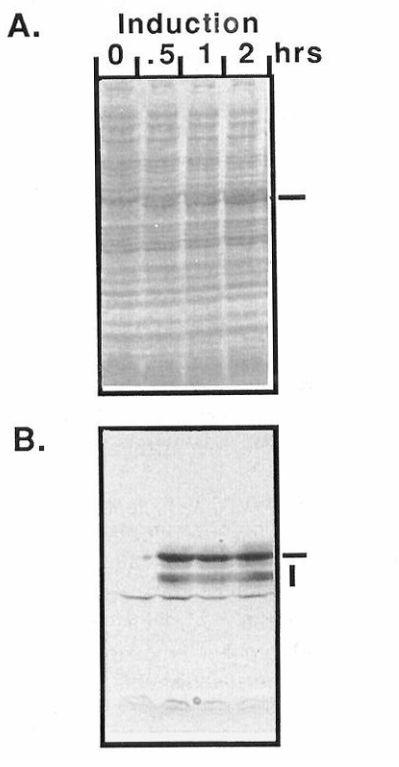


Figure 3. Expression of UBX Ib in E. coli

E. coli strain AR68 harboring pP_L – UBX Ib was grown at 30°C to an OD_{600} = 0.4, and transcription from P_L 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 (B). The position of the 40 kd UBX Ib is indicated by a horizontal line, and the lower molecular weight forms revealed by antibody staining are indicated by a vertical line.

coli (see below). Because of this concordance in biochemical characteristics, and because of the higher level of UBX Ib expression and the simplicity of bacterial culture, E. coli was chosen as the source for further purification and characterization of UBX Ib protein.

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

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 Ib were incubated with a ^{32}P-labeled DNA fragment extending from −6 to +0.8 kb with respect to the *Antp* P1 transcription start site and filtered through nitrocellulose under conditions where protein–DNA complexes (but not free DNA) are retained. Even with unfractionated extracts the *Antp* fragment was retained at least 20-fold better than the ^{3}H-labeled pUC8 control DNA (Figure 4A), whereas both fragments were retained at similar low levels by control extracts lacking UBX Ib (Figure 4B).

The filter-binding assay was used to follow the DNA binding activity during fractionation of cell extracts, and the UBX Ib polypeptide was monitored by polyacrylamide gel immunoblots. Substantial enrichment for UBX Ib 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 Ib protein (Figures 5B and 5C), confirming that UBX Ib was the new DNA binding

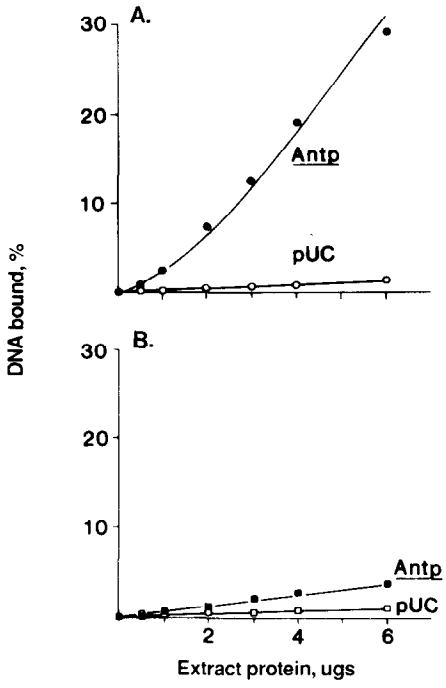


Figure 4. An *Antp* DNA Binding Activity in Extracts from E. coli Expressing UBX Ib

One fmol of a 7 kb ^{32}P-labeled *Antp* DNA fragment from the P1 promoter region (−6.3 kb to +0.8 kb) and 29 fmol of ^{3}H-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 Ib expression plasmid pP_L−UBX Ib or (B) its parent plasmid pAS-1.

species in the extracts. The apparent M_r of 78 kd for UBX Ib in gel filtration (Figure 5C and data not shown), which is twice the monomer molecular weight, suggests that UBX Ib exists in solution as a homodimer of 40 kd protomers.

Definition of UBX Ib Binding Sites at *Antp* P1

To localize UBX Ib 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 HinfI 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 Ib, 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 Ib 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 Ib. Within H472, three regions were protected from nuclease digestion by purified UBX Ib (Figure 7A), designated A-1, A-2, and A-3. A-1 is an ~40 bp region protected at low UBX Ib 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 Ib tested (data not shown).

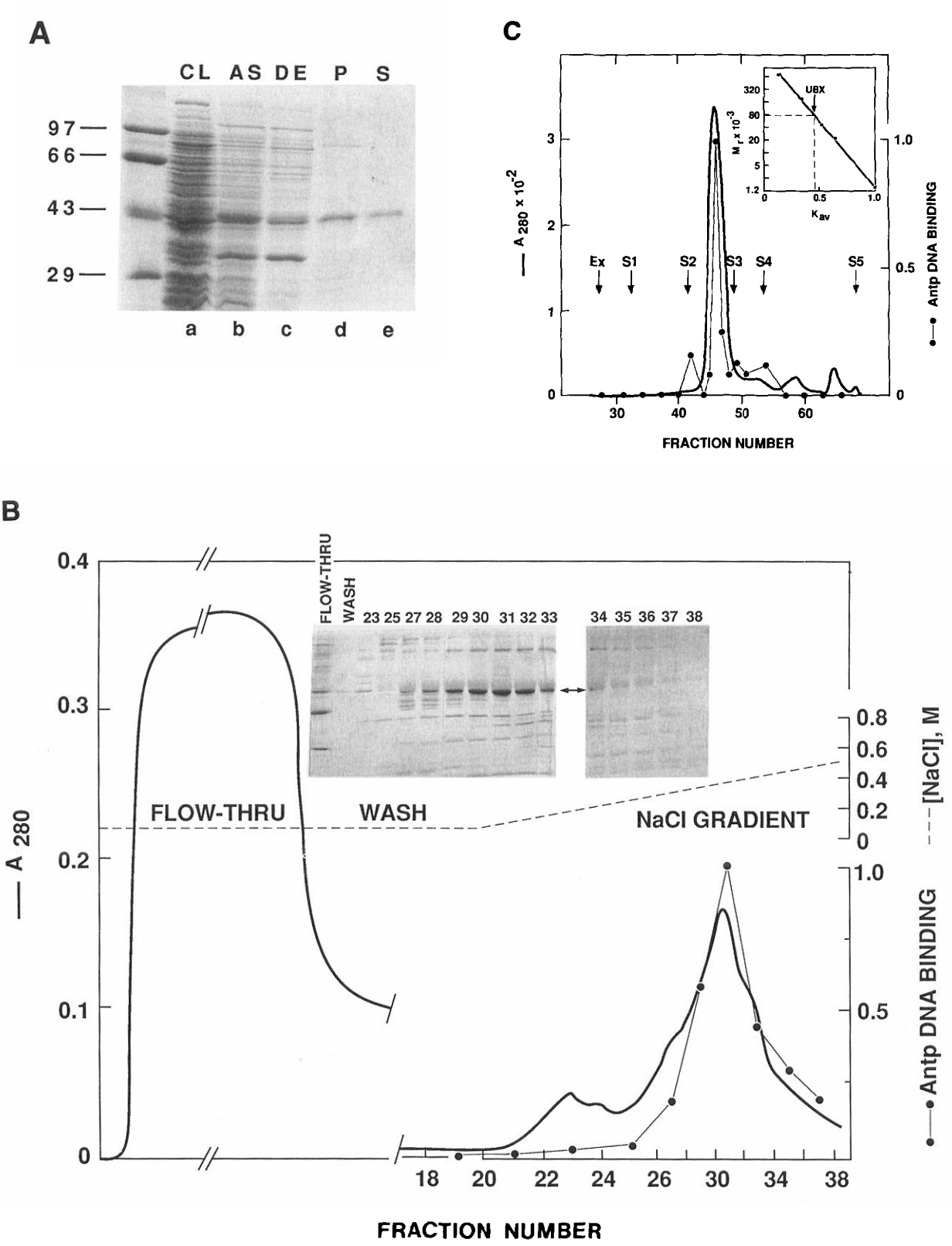
Figure 7 also shows that nuclear extracts prepared from Spodoptera cells infected with the UBX Ib recombinant baculovirus gave the same DNAase I protection pattern of H472 as protein purified from E. coli (compare lanes 8 and 9 with lanes 5 and 6 in A). Thus, by the criterion of DNA binding, UBX Ib proteins produced in E. coli or cultured insect cells are indistinguishable.

Figure 5. Purification of UBX Ib

UBX Ib was purified from a cleared lysate of E. coli AR120 harboring pP_L−UBX Ib by ammonium sulfate fractionation, chromatography through DEAE-Sephacel, phosphocellulose, and hydroxylapatite, and gel filtration.

(A) Electrophoretic analysis of UBX Ib purification fractions. Protein samples from various stages of the purification containing similar amounts of UBX Ib were electrophoresed along with molecular weight standards in an SDS–polyacrylamide gel and stained with Coomassie blue. CL: cleared lysate; AS: 20% ammonium sulfate pellet; DE: pooled fractions from the DEAE-cellulose column; P: pooled fractions from the phosphocellulose column; S: ~1 μg protein from the Superose 12 pool. The positions of the molecular weight standards (phosphorylase b, 97 kd; bovine serum albumin, 66 kd; ovalbumin, 43 kd; carbonic anhydrase, 29 kd) are indicated to the left.

(B) Phosphocellulose chromatography of UBX Ib. The diluted DEAE-Sephacel pool was loaded onto phosphocellulose, washed with the loading buffer, and eluted with an increasing linear gradient of NaCl in the loading buffer. The *Antp* DNA-binding activity of the indicated 2 ml (numbers 19–25) and 1 ml (numbers 26–38) fractions was determined by the nitrocellulose filter assay; flow-through and wash fractions had 0.15 and 0 U of binding activity, respectively. The thick line shows the absorbance profile at 280 nm. In the inset, a SDS–polyacrylamide gel analysis of 2 μl of the



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 Ib. One-fourth of the hydroxylapatite 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, excluded volume; S1, bovine thyroglobulin; S2, bovine gamma globulin; S3, chicken ovalbumin; S4, horse myoglobin; S5, vitamin B-12. The major absorbance peak, with a K_{av} of 0.45, was UBX Ib (see electrophoretic analysis above in A, lane S). The inset is a semilogarithmic plot of the K_{av} of the standards and UBX Ib vs. their relative molecular mass (M_r).

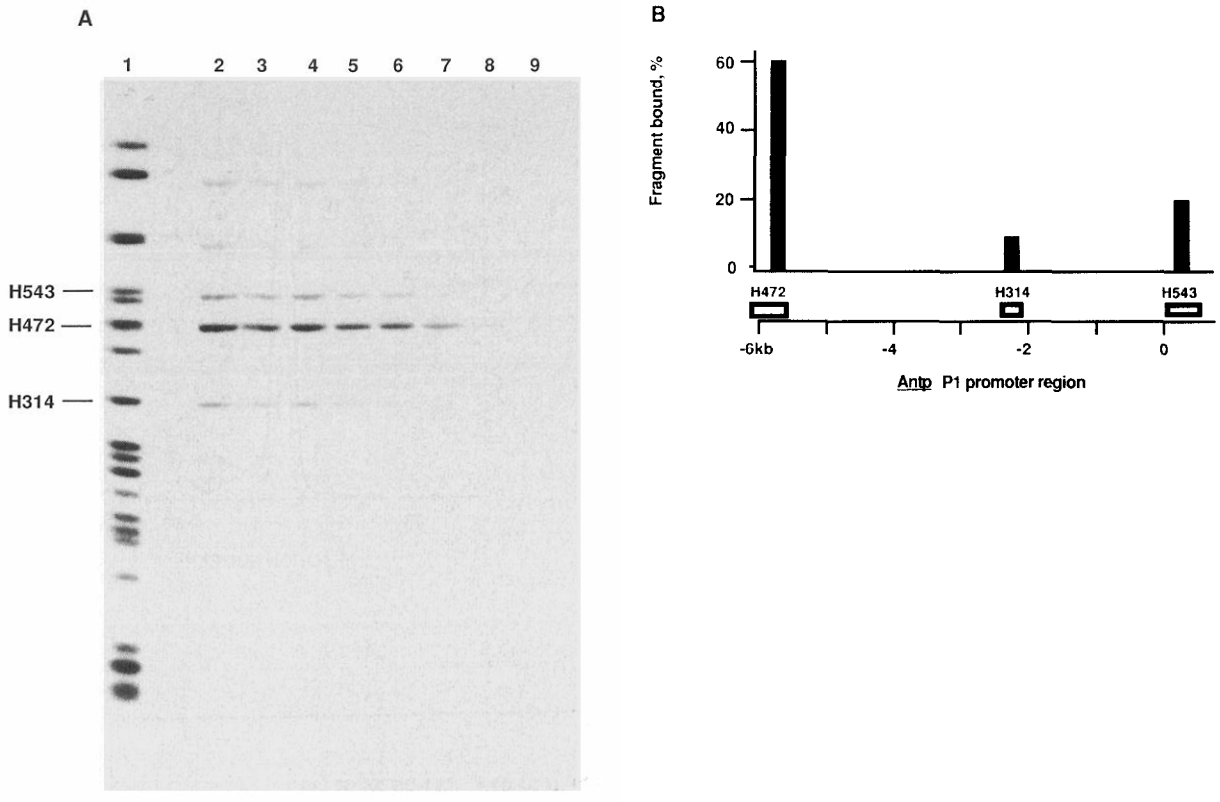


Figure 6. The Mapping of UBX Ib Binding Sites Near the *Antp* P1 Promoter

(A) Binding reactions were as described in the legend to Figure 4 except that the DNA fragments were ^{32}P-labeled HinfI digestion products (lane 1) of the 7 kb *Antp* DNA, and each reaction contained 3 μg of extract protein. Reactions were passed through nitrocellulose and the filters washed with buffer containing: 50 (lane 2), 100 (lane 3), 150 (lane 4), 200 (lane 5), 250 (lane 6), 300 (lane 7), 500 (lane 8), and 1000 (lane 9) mM NaCl. Bound fragments were eluted from the filters with SDS, and visualized by autoradiography after electrophoresis in agarose. The three highest affinity fragments are labeled to the left according to their sizes in base pairs.

(B) The relative amount of each fragment retained on the filter after the 200 mM NaCl wash (lane 5 in A) was determined by integration of peak areas after densitometric scanning of the autoradiogram. These levels are expressed as a percentage of total DNA bound, and the fragment positions are shown with respect to the *Antp* P1 transcription start site at coordinate 0.

UBX Ib 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 Ib 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 Ib 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 accomodate more than one UBX Ib 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 Ib (Figure 9D).

A Repeated Trinucleotide Common to UBX Ib Binding Sites

Within each protected region are tandem arrays of a simple sequence element, the trinucleotide TAA. The most striking array is in A-1, 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-

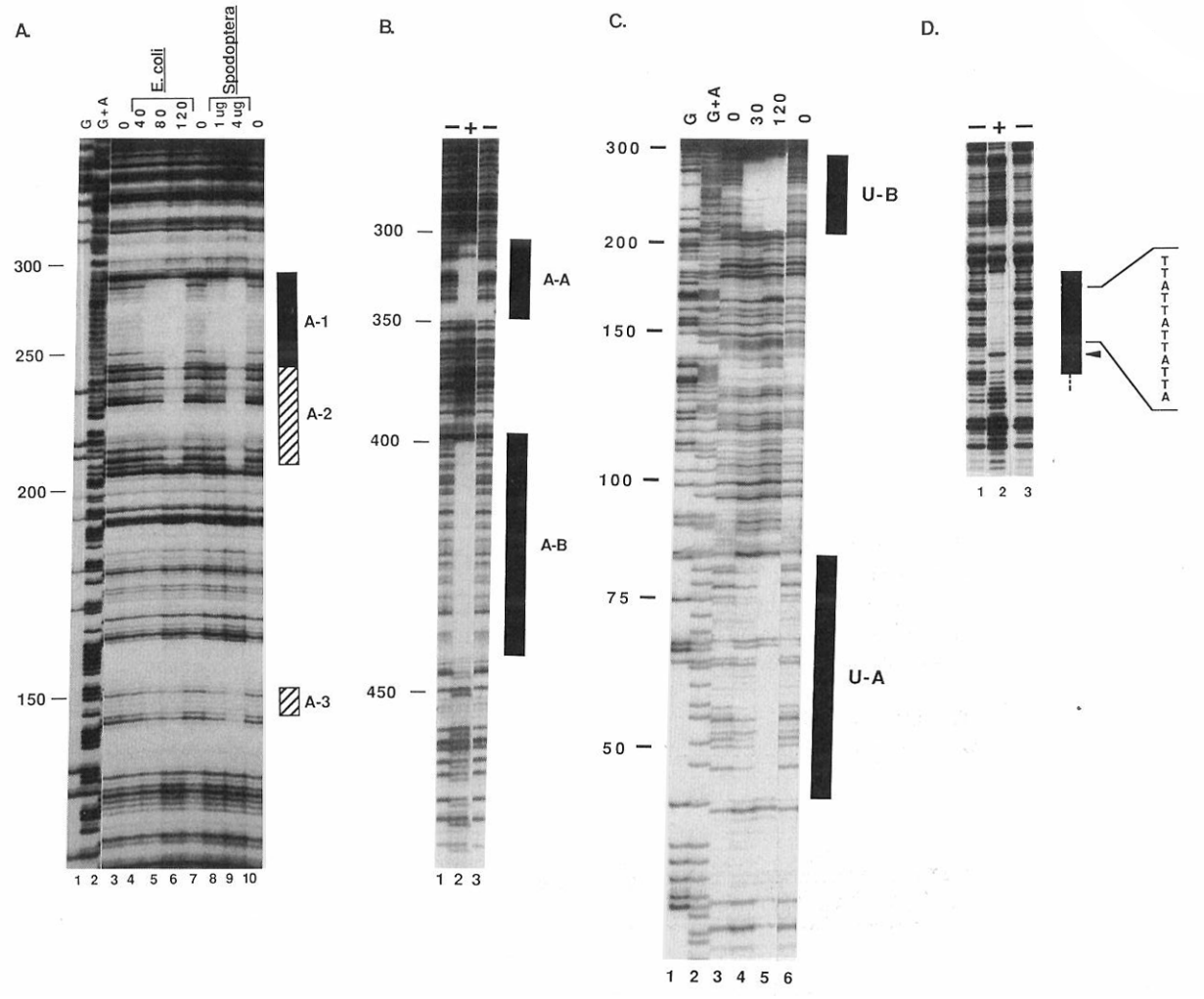


Figure 7. DNAase I Protection Analysis of UBX Ib Binding Sites

Binding reactions containing ~30 fmol of ^{32}P 3′ end–labeled DNA fragments and UBX Ib 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 Ib 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 Ib (lanes 3, 7, and 10), or 40, 80, and 120 ng purified UBX Ib (lanes 4–6), or 1 and 4 μg of a nuclear extract from Spodoptera frugiperda cells expressing UBX Ib (lanes 8 and 9). Lanes 1 and 2 display sequence ladders derived from the labeled fragment.

Under the conditions of these experiments, UBX Ib 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_r of 78 kd for UBX Ib (Figure 5B), the reaction in lane 4 was ~10 nM in UBX Ib, and the binding coefficient for A-1 can therefore be estimated at $\sim 10^{-8}$. The actual affinity may be somewhat higher since the effect of the competitor poly d(I-C) is not taken into account and our UBX Ib preparation may not be fully active.

(B) The reactions contained no UBX Ib (−) or ~120 ng UBX Ib (+), 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 Ib and the DNA was a HindIII–EcoRI fragment of pdel5′−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 Ib (−) or ~120 ng (+) and the DNA was a XhoI–PvuI fragment from pPB177 containing the cloned oligonucleotide $(TAA)_4$ 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 Ib.

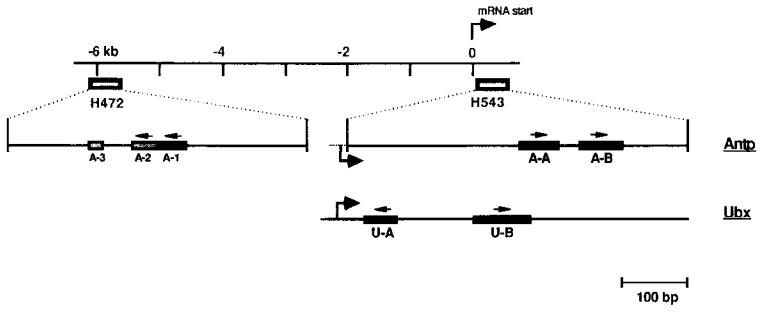


Figure 8. Locations of UBX Ib Binding Sites Near the *Ubx* and *Antp* P1 Promoters

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 Ib (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).

gion), 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 as extended an array as at A-B, where in one part of the protected region only 3 of 36 nucleotides deviate from this pattern.

Although UBX Ib 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 Ib, 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 com-

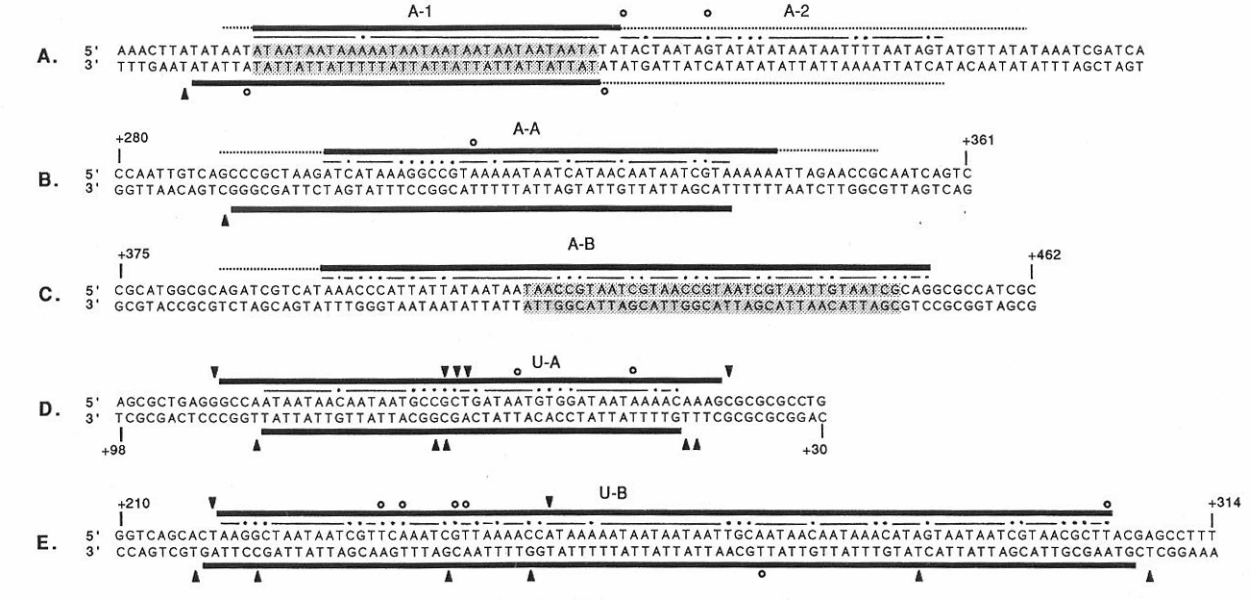


Figure 9. Sequences and DNAase I Protection Patterns of Naturally Occurring UBX Ib Binding Sites

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., 1986; 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 Ib. 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 of 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 Ib binding site cannot be precisely determined and these sequences were therefore not tabulated.)

position is critical for UBX Ib binding, in contrast to several other dA/dT binding proteins (see Solomon et al., 1986, and references therein).

UBX Ib Binds to a Cloned Synthetic $(TAA)_4$ Repeat

To determine whether the TAA motif is sufficient for UBX Ib 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 Ib protected a 21 nucleotide region centered on the TAA repeat. Thus, four TAA repeats are sufficient for UBX Ib 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 Ib. 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 Ib, and have shown that it is a sequence-specific DNA binding protein. High affinity binding sites for UBX Ib occur in clusters just downstream of the *Antp* P1 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 Ib. The naturally occurring binding sites most likely accommodate several UBX Ib molecules as they each contain more extensive triplet arrays than the synthetic sequence and give proportionately larger UBX Ib footprints (40–90 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 suggest that 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 Ib binding sites near the *Ubx* promoter.

The locations of UBX Ib 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, 1986, and references therein) and the Drosophila heat shock transcription factor (Wiederrecht 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 $\sim 10^{-8}$ (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., 1987; the adenovirus major late transcription factor, Chodosh et al., 1986; and the Drosophila heat shock transcription factor, Wu et al., 1988).

The location of some of the UBX Ib 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 (Gillies 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 Ib 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 Ib 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. funebris. 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 Ib 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. funebris 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 Ib 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-*engrailed* 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 Ib 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, $(GXC)_4$ sequences are bound by the Adf-1 transcription factor, and other nuclear factors interact with nearby $(GCT)_4$ and $(TAATAC)_4$ 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 Ib (Figure 5B), since a dimeric structure in prokaryotic regulatory proteins has consistently been associated with binding site sequences with dyad symmetry (Pabo and Sauer, 1984). Such a symmetric structure for UBX Ib 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 Ib. In order to distinguish between these and other models, it will be critical to determine the elements essential for UBX Ib binding.

Experimental Procedures

Plasmids

The UBX Ib expression plasmid pP_L-UBX Ib was constructed by treating the 178 bp HhaI fragment from pφ3712 that contains the *Ubx* open reading frame (ORF) Ib initiation codon (Beachy, 1986; see Figure 2) with T4 DNA polymerase to remove the 3′ extensions (Challberg and Englund, 1980) 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 Ib was reconstructed in pUC8 by ligating the PstI–SalI fragment from pφ3712 into PstI–SalI-digested pPB130 to yield pPB118, in which noncoding sequences upstream of ORF Ib have been removed and the HhaI site immediately upstream of ORF Ib has been replaced by a BamHI site. The BamHI fragment containing the entire *Ubx* ORF Ib was ligated into the BamHI site of pAS-1 (Rosenberg et al., 1983) in the proper orientation for expression from the P_L promoter to yield pP_L-UBX Ib (Figure 2B). Two additional E. coli expression systems incorporating either the *trp–lac* hybrid promoter P_{tac} (Amann and Brosius, 1985) or the lambda P_R promoter (Queen, 1983) were also tested, but these accumulated UBX Ib at levels one-tenth or less that of the P_L system.

Plasmid pAc611-UBX Ib was constructed by ligating the 2.2 kb BamHI–EcoRI fragment of pPB118 to the large BamHI–EcoRI fragment of pAc611 (Summers and Smith, 1986), thus placing *Ubx* ORF Ib downstream of the Autographa californica nuclear polyhedrosis virus (AcNPV) polyhedrin promoter. After cotransfection of the Spodoptera frugiperda cell line Sf-9 (ATCC #CRL 1711) with pAc611-UBX Ib and wild-type viral DNA, the recombinant baculovirus AcNPV-UBX Ib containing the *Ubx* insert was identified by plaque hybridization and isolated as described by Summers and Smith (1986).

The plasmid p599 (kindly supplied by M. Scott), which was used in filter-binding studies, contains a genomic HindIII fragment extending from ~−6 kb to +0.8 kb with respect to the *Antp* P1 transcription start site (Laughon et al., 1986). Three HinfI fragments from p599 (H543, 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 H543 and H314 were previously reported by Laughon et al. (1986).

The *Ubx* promoter deletion plasmid, pdel5′−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 pUC13.

Plasmid pPB177 was constructed by annealing the oligonucleotides 5′GATCCTAATAATAATAAG and 5′GATCCTTATTATTATTAG and ligating them into the BamHI site of Bluescript KSM13+ (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 ^{32}P-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 6 were made from E. coli strain AR120 bearing pP_L-UBX Ib or its parent plasmid pAS-1. After growth at 37°C to $OD_{600}=0.5$, nalidixic acid was added to 40 μg/ml to induce P_L transcription and growth was con-

tinued for 4 hr. A cleared lysate was made by the lysozyme/heat method of Fuller and Kornberg (1983) except that the final lysis buffer contained 10% sucrose, 50 mM Tris–HCl (pH 8.0), 1 mM dithiothreitol, 20 mM Na_2EDTA, 20 mM spermidine–HCl, and ammonium sulfate at 2.5% of saturation.

For preparation of nuclear extracts from baculovirus-infected cells, Sf-9 cells were grown according to Summers and Smith (1986) in 300 ml TNM-FH medium supplemented with 10% fetal bovine serum (Irvine) in 500 ml spinner flasks at 27°C. At a cell density of 1.2×10^6 per ml, AcNPV-UBX Ib virions were added to an moi of ~3, and agitation was interrupted for 30 min to facilitate viral attachment. After 72 hr, cells were harvested and nuclear extracts were prepared by the procedure of Parker and Topol (1984) except that 1 mM dithiothreitol was included in all buffers and 0.3 g instead of 0.25 g ammonium sulfate was added per ml of cleared supernatant. UBX Ib accumulated in the nucleus to about 0.5% of total cell protein after infection, and the final preparation contained an equal proportion of a 41 kd *Ubx* protein and a 35 kd presumptive degradation product as determined by SDS–polyacrylamide gel analysis and immunoblotting with the anti-*Ubx* monoclonal antibody FP3.38 (White and Wilcox, 1984).

Nitrocellulose Filter-Binding Assays

Standard binding reactions were incubated for 10 min at 37°C and contained in 0.5 ml: 1 fmol of HindIII-digested and ^{32}P-labeled p599 DNA, 29 fmol of HindIII-digested and ^{3}H-labeled pUC8 DNA, extract protein as indicated in the figure legends, 50 mM NaCl, 10 mM KCl, 10 mM $MgOAc_2$, 10 mM Tris–HCl (pH 7.4), 0.1 mM dithiothreitol, bovine serum albumin at 50 μg/ml (Fraction V, Miles), sonicated salmon sperm DNA at 48 μg/ml, and 5% dimethylsulfoxide. The 5000-fold mass excess of unlabeled salmon sperm DNA was included to reduce interference by endogenous DNA-binding proteins and exonucleases. Reaction products were filtered through pre-wetted nitrocellulose filters (Millipore, 0.45 μm, 24 mm dia) at a flow rate of 5 ml/min and washed immediately with 1.5 ml of wash buffer (150 mM NaCl, 10 mM KCl, 10 mM $MgOAc_2$, 10 mM Tris–HCl [pH 7.4], 5% dimethylsulfoxide). DNA retained was determined by liquid scintillation counting of dried filters. In Figure 5, *Antp* DNA binding of the column fractions was taken as the percentage of the ^{32}P counts retained on the filters less the percentage of ^{3}H counts retained, and all values were normalized to the fraction with maximal binding. In all experiments, less than 1% of each fragment was retained in the absence of added protein. In the experiment shown in Figure 6, the NaCl concentration of the wash buffer was varied as detailed in the figure legend, and DNA was eluted from the filters by two successive 30 min incubations in 0.2 ml of elution buffer (300 mM NaCl, 5 mM Na_2EDTA, 10 mM Tris–HCl [pH 8.0], 0.2% SDS, 25 μg/ml tRNA). Eluted DNA was concentrated by ethanol precipitation, and electrophoresed in a 3.0% NuSieve (FMC) agarose gel before exposure to pre-flashed XAR film (Kodak).

DNAase I Protection Assays

The standard reaction contained, in 50 μl: ~30 fmol of ^{32}P-labeled DNA fragment, 1 μg annealed poly d(I-C), 50 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 0.1 mM Na_2EDTA, 10 mM Tris–HCl (pH 7.4), 0.1 mM dithiothreitol, and UBX Ib as indicated in the figure legends. After incubation at 22°C–24°C for 10 min, 50 μl of diluent (10 mM $CaCl_2$, 20 mM $MgCl_2$) was added, followed immediately by 4 ng of DNAase I (Worthington). After 0.5 min, reactions were terminated by addition of 0.3 ml of stop buffer (300 mM NaCl, 20 mM Na_2EDTA, 50 mM Tris–HCl [pH 8.0], 1% sarkosyl, 33 μg/ml E. coli tRNA). Following extraction with 0.4 ml of phenol:chloroform (1:1), nucleic acids were concentrated by ethanol precipitation, electrophoresed in 5%, 6%, or 8% polyacrylamide gels containing 8 M urea, and detected by autoradiography. Reference ladders were generated by the chemical sequencing method of Maxam and Gilbert (1977).

UBX Ib Purification

To determine the optimal source of UBX Ib for purification, pP_L–UBX Ib was introduced into several E. coli strains harboring defective lambda prophages encoding the wild-type (cI) or thermolabile (cI857) repressor of the P_L promoter. Expression of UBX Ib from P_L was induced by nalidixic acid treatment to inactivate cI (Mott et al., 1985), or heat treatment at 42°C to inactivate cI857. Although UBX Ib accumulation was greatest (see Figure 3) after heat induction of E. coli AR68 (kindly supplied by A. Shatzman), a strain defective in protein degradation because of the htpR165 allele (Baker et al., 1984), most of the UBX Ib was insoluble. E. coli AR120 (Mott et al., 1985) accumulated soluble UBX Ib to ~1% 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.

Mass culture of AR120 cells containing pP_L–UBX Ib was in 100 l of MIM (Mott et al., 1985) at 37°C to an $OD_{600} = 1.0$. Nalidixic acid was added to 40 μg/ml and the incubation continued for 4 hr. Cells were harvested and resuspended at $OD_{600} = 300$ in Buffer X (10% (w/v) sucrose, 50 mM Tris–HCl [pH 7.4]), frozen in liquid nitrogen, and stored at −70°C.

Throughout purification, UBX Ib was monitored by immunoblotting (Burnette, 1981) of 12% polyacrylamide gels containing SDS (Laemmli, 1970) using the FP3.38 antibody and detection with horseradish peroxidase immunochemistry (Vector Labs). DNA binding activity was determined by the nitrocellulose filter assay. Because UBX Ib is poorly soluble at low ionic strengths, precipitation was minimized by keeping protein solutions dilute and using only short periods of dialysis.

Five hundred grams of cell paste was thawed, diluted with an equal volume of Buffer X, and brought to final concentrations of 400 mM NaCl, 10 mM Na_2EDTA, 2 mM dithiothreitol, and 0.25 mM phenylmethylsulfonyl fluoride. All subsequent steps were carried out at 4°C. Lysozyme was added to 0.4 mg/ml, and after 45 min Brij 58 was added to a final concentration of 0.15% to lyse the cells. After 15 min, the extract was centrifuged at 150,000 × g for 40 min and the supernatant (900 ml) was gradually brought to 0.5% polyethyleneimine to precipitate nucleic acids. The mixture was stirred for 20 min and the precipitate removed by centrifugation at 7,500 × g for 15 min. The supernatant was brought to 20% saturation with solid ammonium sulfate, and stirred for 1.5 hr. The precipitate was collected by centrifugation (22,000 × g for 30 min), frozen in liquid nitrogen, and stored at −70°C. Although only ~20% of UBX Ib was recovered, this low concentration of ammonium sulfate was used because it resulted in a 20-fold or better enrichment.

The ammonium sulfate precipitate was thawed, resuspended to a final volume of 10 ml in Buffer Y (5% glycerol, 2 mM dithiothreitol, 1 mM Na_2EDTA, and 50 mM Tris–HCl (pH 7.4)) containing 25 mM NaCl, and dialyzed for 2 hr against the resuspension buffer. Particulate matter was removed by centrifugation at 1500 × g for 10 min, and the supernatant was applied to a DEAE Sephacel column (1.8 cm^2 × 5.5 cm) equilibrated with the resuspension buffer. After a 60 ml buffer wash, the column was developed with a 100 ml linear NaCl gradient (0.025–1.0 M NaCl) in Buffer Y. Fractions containing UBX Ib (240–340 mM NaCl) were pooled (12.5 ml), diluted with 18 ml Buffer Z (5% glycerol, 0.5 mM dithiothreitol, 0.1mM Na_2EDTA, 25 mM sodium phosphate [pH 6.8]), and dialyzed for 1.2 hr against 100 mM NaCl in Buffer Z. Precipitated material was removed by centrifugation, and the supernatant (30 ml) was applied to a phosphocellulose column (0.8 cm^2 × 5 cm). After a 30 ml wash, the column was developed with a 40 ml linear gradient (0.1–1.0 M NaCl) in Buffer Z. Active fractions (370–410 mM NaCl) containing full-length UBX Ib were pooled (3 ml) and diluted to a final concentration of 100 mM NaCl with Buffer Z. The diluted material was applied to a hydroxylapatite column (0.4 cm^2 × 2.5 cm), and the column was washed with 10 ml of 100 mM NaCl in Buffer Z. UBX Ib was eluted with one column volume of 400 mM NaCl in Buffer Z, and the pooled material was adjusted to a final concentration of 800 mM NaCl. One-fourth of the hydroxylapatite pool (0.25 ml) was applied to a Superose 12 HR 10/30 (Pharmacia) gel filtration column (calibrated with standards ranging from 13.5 to 670 kd; BioRad) and active fractions containing the full-length protein (K_{av} = 0.45) were pooled (0.6 ml) and stored at −70°C.

The hydroxylapatite and Superose 12 pools contained traces of UBX Ib breakdown products, but these amounted to less than 10% of total protein. UBX Ib concentrations were determined by comparison with protein standards in Coomassie-stained gels.

Acknowledgments

We thank J. Bermingham and M. Scott for supplying the *Antp* DNA clones and sequence information prior to publication, R. White and M. Wilcox for the hybridoma cell line FP3.38, M. Biggin for a useful *Ubx* promoter deletion plasmid, A. Shatzman and M. Rosenberg for E. coli

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 P. Englar and S. 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

Akam, M. E. (1983). The location of *Ultrabithorax* transcripts in Drosophila tissue sections. EMBO J. *2*, 2075–2084.

Akam, M. E. (1987). The molecular basis for metameric pattern in the Drosophila embryo. Development *101*, 1–22.

Amann, E., and Brosius, J. (1985). ATG vectors for regulated high-level expression of cloned genes in *Escherichia coli*. Gene *40*, 183–190.

Anderson, K. V., and Nüsslein-Volhard, C. (1984). Genetic analysis of dorsal-ventral embryonic pattern in Drosophila. In Pattern Formation, G. M. Malacinski and S. V. Bryant, eds. (New York: Macmillan), pp. 269–289.

Baker, T. A., Grossman, A. D., and Gross, C. A. (1984). A gene regulating the heat shock response in *E. coli* also affects proteolysis. Proc. Natl. Acad. Sci. USA *81*, 6779–6783.

Baldwin, A. S., and Sharp, P. A. (1988). Two factors, NF-kB and H2TF1, interact with a single regulatory sequence in the class 1 MHC promoter. Proc. Natl. Acad. Sci. USA *85*, 723–727.

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell *33*, 729–740.

Beachy, P. A. (1986). The *Ubx* domain in the bithorax complex of Drosophila. Ph. D. thesis, Stanford University, Stanford, California.

Beachy, P. A., Helfand, S. L., and Hogness, D. S. (1985). Segmental distribution of bithorax complex proteins during Drosophila development. Nature *313*, 545–551.

Bender, W., Akam, M., Karch, F., Beachy, P. A., Peifer, M., Spierer, P., Lewis, E. B., and Hogness, D. S. (1983). Molecular genetics of the bithorax complex in *Drosophila melanogaster*. Science *221*, 23–29.

Bienz, M., Saari, G., Tremml, G., Müller, J., Züst, B., and Lawrence, P. A. (1988). Differential regulation of *Ultrabithorax* in two germ layers of Drosophila. Cell *53*, 567–576.

Biggin, M. D., and Tjian, R. (1988). Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. Cell *53*, 699–711.

Burnette, W. N. (1981). "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated Protein A. Anal. Biochem. *112*, 195- 203.

Carroll, S. B., Laymon, R. A., McCutcheon, M. A., Riley, P. D., and Scott, M. P. (1986). The localization and regulation of *Antennapedia* protein expression in Drosophila embryos. Cell *47*, 113–122.

Challberg, M. D., and Englund, P. T. (1980). Specific labelling of 3′ termini with T4 DNA polymerase. Meth. Enzymol. *65*, 39–43.

Charnay, P., Treisman, R., Mellon, P., Chao, M., Axel, R., and Maniatis, T. (1984). Differences in human α- and β-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell *38*, 251–263.

Chodosh, L. A., Carthew, R. W., and Sharp, P. A. (1986). A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol. Cell. Biol. *6*, 4723–4733.

Desplan, C., Theis, J., and O'Farrell, P. H. (1985). The Drosophila developmental gene, *engrailed*, encodes a sequence-specific DNA binding activity. Nature *318*, 630–635.

Dorn, A., Bollekens, J., Staub, A., Benoist, C., and Mathis, D. (1987). A multiplicity of CCAAT box-binding proteins. Cell *50*, 863–872.

Duncan, I. (1987). The bithorax complex. Annu. Rev. Genet. *21*, 285-319.

Fuller, R. S., and Kornberg, A. (1983). Purified dnaA protein in the initiation of replication at the *Escherichia coli* chromosomal origin of replication. Proc. Natl. Acad. Sci. USA *80*, 5817–5821.

Galas, D. J., and Schmitz, A. (1978). DNAase footprinting: a simple method for the detection of protein-DNA binding specificity. Nucl. Acids Res. *5*, 3157–3170.

Gillies, S. D., Morrison, S. L., Oi, V. T., and Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell *33*, 717–728.

Hafen, E., Levine, M., Garber, R. L., and Gehring, W. J. (1983). An improved *in situ* hybridization method for the detection of cellular RNAs in Drosophila tissue sections and its application for localizing transcripts of the homeotic *Antennapedia* gene complex. EMBO J. *2*, 617–623.

Hafen, E., Levine, M., and Gehring, W. J. (1984). Regulation of *Antennapedia* transcript distribution by the bithorax complex in Drosophila. Nature *309*, 287–289.

Haynes, S. R., Rebbert, M. L., Mozer, B. A., Forquignon, F., and David, I. B. (1987). Pen repeat sequences are GGN clusters and encode a glycine-rich domain in a Drosophila cDNA homologous to the rat helix destabilizing protein. Proc. Natl. Acad. Sci. USA *84*, 1918–1923.

Heberlein, U., England, B., and Tjian, R. (1988). Characterization of Drosophila transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. Cell *41*, 965–977.

Hogness, D. S., Lipshitz, H. D., Beachy, P. A., Peattie, D. A., Saint, R. B., Goldschmidt-Clermont, M., Harte, P. J., Gavis, E. R., and Helfand, S. L. (1985). Regulation and products of the *Ubx* domain of the bithorax complex. Cold Spring Harbor Symp. Quant. Biol. *50*, 181–194.

Ingham, P. W., and Martinez-Arias, A. (1986). The correct activation of Antennapedia and bithorax complex genes requires *fushi tarazu* gene. Nature *324*, 592–597.

Johnson, P. F., Landschulz, W. H., Graves, B. J., and McKnight, S. C. (1987). Identification of a rat liver protein that binds to the enhancer core element of three animal viruses. Genes Dev. *1*, 133–146.

Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J., and Tjian, R. (1987). A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell *48*, 79–89.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

Laughon, A., and Scott, M. P. (1984). Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins. Nature *310*, 25–31.

Laughon, A., Boulet, A. M., Bermingham, J. R. J., Laymon, R. A., and Scott, M. P. (1986). Structures of transcripts from the homeotic *Antennapedia* gene of *Drosophila melanogaster*: two promoters control the major protein-coding region. Mol. Cell. Biol. *6*, 4676–4689.

Lewis, E. B. (1963). Genes and developmental pathways. Am. Zool. *3*, 33–56.

Lewis, E. B. (1964). Genetic control and regulation of developmental pathways. In The Role of Chromosomes in Development, M. Locke, ed. (New York: Academic Press), pp. 231–252.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Lipshitz, H. D., Peattie, D. A., and Hogness, D. S. (1987). Novel transcripts from the *Ultrabithorax* domain of the bithorax complex. Genes Dev. *1*, 307–322.

Mansour, S. L., Grodzicker, T., and Tjian, R. (1986). Downstream sequences affect transcription initiation from the adenovirus major late promoter. Mol. Cell. Biol. *6*, 2684–2694.

Maxam, A., and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA *74*, 560–564.

McKnight, S., and Tjian, R. (1986). Transcriptional selectivity of viral genes in mammalian cells. Cell *46*, 795–805.

Merrill, G. F., Hauschka, S. D., and McKnight, S. L. (1984). *tk* enzyme expression in differentiating muscle cells is regulated through an internal segment of the cellular *tk* gene. Mol. Cell. Biol. *4*, 1777–1784.

Messing, J., and Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. Gene *19*, 269–276.

Minana, F. J., and Garcia-Bellido, A. (1982). Preblastoderm mosaics of mutants of the bithorax complex. Roux's Arch. Dev. Biol. *191*, 331–334.

Mott, J. E., Grant, R. A., Ho, Y.-S., and Platt, T. (1985). Maximizing gene expression from plasmid vectors containing the P_L promoter: strategies for overproducing transcription termination factor. Proc. Natl. Acad. Sci. USA *82*, 88–92.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature *287*, 795–801.

Nüsslein-Volhard, C., Frohnhöfer, H. G., and Lehmann, R. (1987). Determination of anteroposterior polarity in Drosophila. Science *238*, 1675–1681.

O'Connor, M. B., Binari, R., Perkins, L. A., and Bender, W. (1988). Alternative RNA products from the *Ultrabithorax* domain of the bithorax complex. EMBO J. *7*, 435–445.

Pabo, C., and Sauer, R. T. (1984). Protein-DNA recognition. Annu. Rev. Biochem. *53*, 293–321.

Queen, C. (1983). A vector that uses phage signals for efficient synthesis of proteins in *E. coli*. J. Mol. Appl. Gen. *2*, 1–10.

Queen, C., and Baltimore, D. (1983). Immunoglobulin gene transcription is activated by downstream sequence elements. Cell *33*, 741–748.

Reinberg, D., Horikoshi, M., and Roeder, R. G. (1987). Factors involved in specific transcription in mammalian RNA polymerase II. J. Biol. Chem. *262*, 3322–3330.

Rio, D. C., and Rubin, G. M. (1985). Transformation of cultured *Drosophila melanogaster* cells with a dominant selectable marker. Mol. Cell. Biol. *5*, 1833–1838.

Rosenberg, M., Ho, Y.-S., and Shatzman, A. (1983). The use of pKC30 and its derivatives for controlled expression of genes. Meth. Enzymol. *101*, 123–138.

Saari, G., and Bienz, M. (1987). The structure of the *Ultrabithorax* promoter of *Drosophila melanogaster*. EMBO J. *6*, 1775–1779.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Scheidereit, C., Heguy, A., and Roeder, R. G. (1987). Identification and purification of a human lymphoid-specific octamer-binding protein (OTF-2) that activates transcription of an immunoglobulin promoter in vitro. Cell *51*, 783–793.

Schupbach, T., and Wieschaus, E. (1986). Maternal-effect mutations altering the anterior–posterior pattern of the Drosophila embryo. Roux's Arch. Dev. Biol. *195*, 302–317.

Scott, M. P. (1985). Molecules and puzzles from the *Antennapedia* homeotic gene complex of Drosophila. Trends Genet. *1*, 74–80.

Scott, M. P., and O'Farrell, P. A. (1986). Spatial programming of gene expression in early Drosophila embryogenesis. Annu. Rev. Cell Biol. *2*, 49–80.

Shepherd, J. C. W., McGinnis, E., Carrasco, A. E., De Robertis, E. M., and Gehring, W. J. (1984). Fly and frog homoeo domains show homologies with yeast mating type regulatory proteins. Nature *310*, 70–71.

Soeller, W. C., Poole, S. G., and Kornberg, T. (1988). *In vitro* transcription of the Drosophila *engrailed* gene. Genes Dev. *2*, 68–81.

Solomon, M. J., Strauss, F., and Varshavsky, A. (1986). A mammalian high mobility group protein recognizes any stretch of six A·base pairs in duplex DNA. Proc. Natl. Acad. Sci. USA *83*, 1276–1280.

Struhl, G., and White, R. A. H. (1985). Regulation of the *Ultrabithorax* gene of Drosophila by other bithorax complex genes. Cell *43*, 507–519.

Summers, M. D., and Smith, G. E. (1986). A manual of methods for Baculovirus vectors and insect cell culture procedures. Texas Agricultural Station Bulletin, No. 15555.

von der Ahe, D., Janich, S., Scheidereit, C., Renkawitz, R., Schultz, G., and Beato, M. (1985). Glucocorticoid and progesterone receptors bind to the same sites in two hormonally regulated promoters. Nature *313*, 706–709.

Wakimoto, B. T., and Kaufman, T. C. (1981). Analysis of larval segmentation in lethal genotypes associated with the Antennapedia gene complex in *Drosophila melanogaster*. Dev. Biol. *81*, 51–64.

Wharton, K. A., Yedvobnick, B., Finnerty, V. G., and Artavanis-Tsakonas, S. (1985) *opa*: a novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in D. melanogaster. Cell *40*, 55–62.

White, R. A. H., and Akam, M. E. (1985). *Contrabithorax* mutations cause inappropriate expression of *Ultrabithorax* production in Drosophila. Nature *318*, 567–569.

White, R. A. H., and Wilcox, M. (1984). Protein products of the bithorax complex in Drosophila. Cell *39*, 163–171.

White, R. A. H., and Wilcox, M. (1985). Regulation of the distribution of *Ultrabithorax* proteins in Drosophila. Nature *318*, 563–567.

Wiederrecht, G., Shuey, D. J., Kibbe, W. A., and Parker, C.S. (1987). The Saccharomyces and Drosophila heat shock transcription factors are identical in size and DNA binding properties. Cell *48*, 507–515.

Wilde, C. D., and Akam, M. (1987). Conserved sequence elements in the 5′ region of the *Ultrabithorax* transcription unit. EMBO J. *6*, 1393–1401.

Wright, S., Rosenthal, A., Flavell, R., and Grosveld, F. (1984). DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell *38*, 265–273.

Wu, C., Wilson, S., Walker, B., Dawid, I., Paisley, T., Zimarino, V., and Ueda, H. (1987). Purification and properties of Drosophila heat shock activator protein. Science *238*, 1247–1253.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene *33*, 103–119.

Notes Added in Proof

Kornfeld et al., unpublished data, is now in press: Kornfeld, K., Saint, R. B., Beachy, P. A., Harte, P. J., Peattie, D. A., and Hogness, D. S., Genes Dev., 1989.

C. Desplan and P. O'Farrell, personal communication, is also in press: Desplan, C., Theis, J., and O'Farrell, P. H., Cell *54*, 1081–1090, 1988.