Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Symposia on Quantitative Biology

Regulation and Products of the *Ubx* Domain of the Bithorax Complex

D.S. Hogness, H.D. Lipshitz, P.A. Beachy, et al.

Cold Spring Harb Symp Quant Biol 1985 50: 181-194
Access the most recent version at doi:10.1101/SQB.1985.050.01.024

References This article cites 26 articles, 4 of which can be accessed free at:
http://symposium.cshlp.org/content/50/181.refs.html

Article cited in:
http://symposium.cshlp.org/content/50/181#related-urls

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

To subscribe to *Cold Spring Harbor Symposia on Quantitative Biology* go to:
http://symposium.cshlp.org/subscriptions

Copyright © 1985 Cold Spring Harbor Laboratory Press

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

Regulation and Products of the *Ubx* Domain of the Bithorax Complex

D.S. HOGNESS, H.D. LIPSHITZ, P.A. BEACHY, D.A. PEATTIE, R.B. SAINT,* M. GOLDSCHMIDT-CLERMONT,† P.J. HARTE, E.R. GAVIS, AND S.L. HELFAND‡

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

The cluster of homeotic genes known as the bithorax complex (BX-C) (Lewis 1978) is divisible into three complementation groups, or functional domains (Sánchez-Herrero et al. 1985). Each domain provides the major developmental determinants for segment identity in one of three adjacent metameric regions that together extend from the anterior/posterior compartment boundary (a/p) of the second thoracic segment (T2) into the eighth abdominal segment (A8) (Fig. 1). Thus, the Ultrabithorax (*Ubx*) domain controls the T2p-A1a region (Lewis 1963, 1978, 1981, 1982; Morata and Kerridge 1981; Hayes et al. 1984; Struhl 1984; Sánchez-Herrero 1985 and this volume), and the abdominal-A (abd-A) and Abdominal-B (*Abd-B*) domains control the A1p-A4 and A5-A8 regions, respectively (Sánchez-Herrero et al. 1985).

These positional assignments of domain functions derive from transformations of segmental identities in the larval and adult epidermis caused by mutations in the respective domains. Recent evidence indicates that the BX-C also provides metameric identity functions for the central nervous system (CNS) and musculature—evidence derived both from mutational studies (Lawrence and Johnston 1984; Teugels and Ghysen 1985) and from the metameric distribution of BX-C RNAs (Akam 1983; Akam and Martinez-Arias 1985 and this volume; Harding et al. this volume) and proteins (White and Wilcox 1984; Beachy et al. 1985).

Present addresses: *Division of Entomology, CSIRO, P.O. Box 1700, Canberra City, ACT 2601, Australia; †Department of Molecular Biology, University of Geneva, 30 quai Ernest Ansermet, 1211 Geneva 4, Switzerland; ‡Department of Biology, Yale University, New Haven, Connecticut.

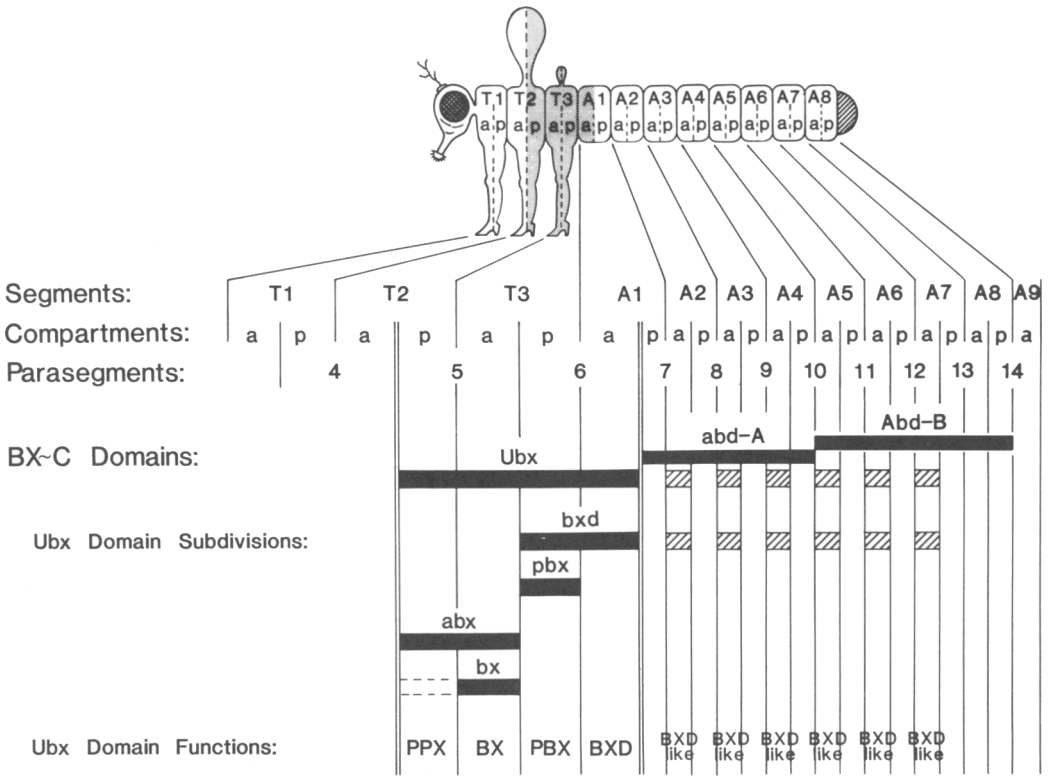

Figure 1. Mutations and functions of the BX-C domains. The diagram of the fly is in the style of Morata (Sánchez-Herrero et al. 1985). The stippled portion of the fly represents the region where the *Ubx* domain functions are the primary determinants of metameric identity.

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

Taken together, these observations indicate that the BX-C genes act as general transducers of positional information for the specification of metameric identities in quite different tissues. An understanding of the role played by the BX-C in metameric development therefore requires the solution of two basic problems: How is the expression of the BX-C genes regulated in a position-specific manner? How do the products of that expression specify the morphologies and functions that differentiate the metameric units within each of several different tissues?

A more precise molecular definition of these problems is required for their solution. To that end, we have concentrated on the molecular analysis of the *Ubx* domain, both because it is the best defined of the three BX-C domains and because we think it will serve as a model for the other two. The *Ubx* domain occupies the proximal or left one third of the BX-C DNA and contains two long transcription units (Fig. 2) (Bender et al. 1983; Beachy et al. 1985). In this paper we focus on the problem of how these two units can interact in *cis* so that one regulates the expression of the other in a position-specific manner.

Definition of the *cis*-Regulatory Problem

Genetic definition. The *Ubx* domain is defined by five classes of recessive mutations (Lewis 1963, 1978, 1981, 1982; Morata and Kerridge 1981; Hayes et al. 1984; Struhl 1984; Sánchez-Herrero et al., this volume). Four of these (anterobithorax [*abx*], bithorax [*bx*], bithoraxoid [*bxd*], and postbithorax [*pbx*]) serve to define the individual functions of the domain, whereas the fifth (*Ubx*), *cis*-inactivates the set of functions defined by the other four and thereby defines the genetic determinants of the set as a single functional domain.

The definition of these functions derives from the transformations of identity effected by these mutations in the T2p-A1a region of the larval and adult epidermis. These effects can best be described in parasegmental units, where a parasegment consists of the posterior compartment of a given segment plus the adjacent anterior compartment of the next segment (Fig. 1). According to the convention adopted by Martinez-Arias and Lawrence (1985), the region controlled by the *Ubx* domain therefore consists of parasegments 5 (T2p, T3a) and 6 (T3p, A1a). Four compartment-specific functions can be defined according to the rule that mutational inactivation of a given function transforms the compartment in which its expression is required for normal development to the identity of the homologous compartment in the anteriorly adjacent parasegment. These functions are referred to here as postprothorax (PPX), BX, PBX, and BXD and are required, respectively, for the normal development of T2p, T3a, T3p, and A1a (Fig. 1). (The PPX function was previously referred to as ABX [Beachy et al. 1985] and is changed here to conform with an earlier designation [Morata and Kerridge 1981] and with usage elsewhere in this volume [Akam and Martinez-Arias; Sánchez-Herrero et al.].)

The *abx* and *bxd* mutations effect a division of these four functions into two subsets required for the normal development of parasegments 5 and 6, respectively (Fig. 1). Thus, *abx* mutations induce transformations in both compartments of parasegment 5 and are therefore considered to inactivate both members of the parasegment 5 subset (PPX, BX), whereas *bxd* mutations are considered to inactivate both members of the parasegment 6 subset (PBX, BXD) because they transform both compartments of this parasegment (Lewis 1963, 1978, 1981, 1982; Sánchez-Herrero et al., this volume; Casanova et al. 1985). The division of each subset into two functions results from the more restricted transformations of the adult epidermis observed for *bx* and *pbx* mutations, which affect, respectively, T3a (inactivation of BX, but not PPX) and T3p (inactivation of PBX, but not BXD). *Ubx* mutations are considered to *cis*-inactivate both subsets because they transform parasegments 5 and 6 to the "ground state" identity of parasegment 4 and because they fail to complement any of the other four mutational classes (Lewis 1963, 1978; Morata and Kerridge 1981; Kerridge and Morata 1982; Hayes et al. 1984; Struhl 1984; Sánchez-Herrero et al. 1985).

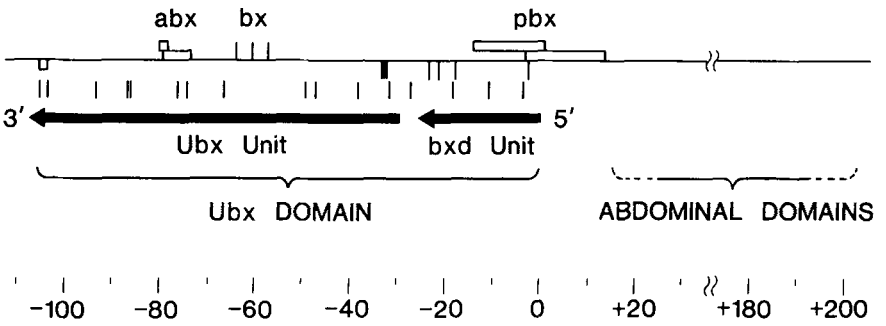

Figure 2. Molecular map of the *Ubx* domain. The positions of the breakpoints of the *Ubx* and the four *bxd* chromosomal rearrangements that define the *Ubx* and *bxd* regions are indicated by the vertical lines just above the solid arrows denoting the orientations and extents of the *Ubx* and *bxd* transcription units. The *Ubx* and *bxd* pseudopoint mutations are indicated just below the long horizontal line representing the chromosomal DNA. The *abx* and *pbx* deletions and the sites of the *bx* insertional mutations are indicated above this line. The scale is in kilobases. (Modified from Beachy et al. 1985.)

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

The above division of the parasegment 5 subset into the compartment-specific PPX and BX functions is supported by two additional lines of evidence. First, these two functions, or, more precisely, their genetic determinants, are dispensable after quite different times of development: PPX after only 7 hours of embryogenesis (Morata and Kerridge 1981) and BX after completion of embryonic and most of larval development (Lewis 1964; Morata and Garcia-Bellido 1976). Second, *abx* mutations are cold-sensitive with respect to the PPX, but not the BX, function (Casanova et al. 1985).

The mutational subdivision of this subset may not be as simple as we have supposed, however, since Casanova et al. (1985) have found that at low temperature (17°C), certain *bx* mutants exhibit a slightly defective PPX function in addition to their loss of the BX function. We have not attempted to account for such secondary complexities in the scheme given in Figure 1. Rather, this scheme is viewed as a useful first approximation that is most accurate with respect to the following conclusions that are particularly relevant to this paper. (1) The *Ubx* domain provides *multiple* identity functions for the T2p-A1a region. (2) These functions are divisible into two subsets that are specific for parasegments 5 and 6 and are inactivated by *abx* and *bxd* mutations, respectively. (3) Each subset consists of more than one function. (4) *Ubx* mutations inactivate in *cis* all functions of the domain. In addition, a set of functions that are expressed in the anterior portions of A2-A7 and inactivated by *Ubx* and *bxd* mutations are shown in Figure 1 (BXD-like functions) to account for the partial transformations induced by both mutations in these regions of the larval epidermis (Lewis 1978, 1981; Bender et al., this volume).

Molecular definition. Figure 2 shows a molecular map of the *Ubx* domain that emphasizes the relationship between the sites of mutations that define the domain (Bender et al. 1983) and two transcription units (~75 kb and ~25 kb) found within it (Beachy et al. 1985; R. Saint et al.; M. Goldschmidt-Clermont et al.; both in prep.). The sites of the *Ubx* and *bxd* mutations are not interspersed, rather, they occupy regions (*Ubx* and *bxd* regions) that are closely correlated with those defined by the 75-kb and 25-kb transcription units, which are therefore called the *Ubx* and *bxd* units respectively. The *abx* and *bx* mutations occupy closely linked but nonoverlapping regions within the *Ubx* unit, whereas the *pbx* mutations overlap the *bxd* unit.

The most curious aspect of this distribution is that mutations in either of two closely linked regions, each of which is closely correlated with a transcription unit, can *cis*-inactivate the parasegment 6 and BXD-like subsets of domain functions. These subsets therefore depend on the integrity of the DNA in each of two long regions and on the close linkage of these regions within the same DNA molecule. The dependence of the parasegment 6 functions on this close linkage is emphasized by the observation that these functions can be expressed only if the DNA spanning both regions is contiguous (Lewis 1978; Bender et al. 1983). In contrast, the *Ubx* unit is sufficient for the expression of the parasegment 5 functions defined by the *abx* and *bx* mutations located within that unit (Morata and Kerridge 1981; Bender et al. 1983). The *Ubx* unit therefore appears to be required for all domain functions, subsets of which also depend on the close linkage of the DNA in the *bxd* region.

The Domain Functions Depend on the *Ubx* Unit Because They Require Proteins Encoded by That Unit

An explanation for the dependence of all functions on the *Ubx* unit derives from an analysis of its RNAs (*Ubx* RNAs) and from the structure and location of the *Ubx* mutations. Figure 3 shows the three major size classes of the mature *Ubx* RNAs, along with their temporal profiles of expression. We have defined the structure of the 3.2-kb-size class by nucleotide sequence analyses of two cloned cDNAs and of homologous regions in the genomic DNA and by S1-protection and primer-extension analyses of its 5′ end (R. Saint et al., in prep.). These analyses indicate that the 3.2-kb RNAs have a common 5′ end and consist of four exons, of which the 5′ and 3′ exons account for 97% of the RNA sequence, the two internal 51-bp microexons accounting for the remainder. The two cDNA clones differ with respect to the donor splice site for the 5′ exon, one being 27 bp removed from the other. This heterogeneity in the 3.2-kb *Ubx* RNAs can be translated into a heterogeneity among the proteins encoded by the *Ubx* unit (*Ubx* proteins) because the 27 bp are in phase with a long open reading frame of 380 codons (ORF Ia; inclusion of the 27 bp yields ORF Ib of 389 codons) that is initiated by an AUG codon in the 5′ exon and terminated by a UAG codon in the 3′ exon (Fig. 4).

There is little doubt that these ORFs are translated to yield *Ubx* proteins in *Drosophila melanogaster* cells. The strongest evidence derives from experiments with antibodies directed against the *Ubx* determinants of β-galactosidase (β-Gal)-*Ubx* hybrid proteins encoded by a hybrid gene in which a sequence of codons from ORF Ia has been fused in-phase with the coding region of the *Escherichia coli lacZ* gene (White and Wilcox 1984; Beachy et al. 1985). These antibodies bind proteins that are present in wild-type but not in *Ubx* mutant embryos. These *Ubx* proteins exhibit metameric distributions in embryos and larval imaginal disks (see below) that are consistent with those observed for *Ubx* RNAs (Akam 1983; Akam and Martinez-Arias 1985), and they migrate in SDS gels with electrophoretic mobilities approximating those expected for these ORFs. The conclusion that these ORFs encode functional proteins is also supported by the following observations: The codon usage in these ORFs fits well with the codon frequencies observed for known *D. melanogaster* structural genes (Beachy et al. 1985); their 3′-exon regions contain a "homeo box" (Fig. 4) (McGinnis et al. 1984; Scott and Weiner 1984), the only one found in the *Ubx*

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

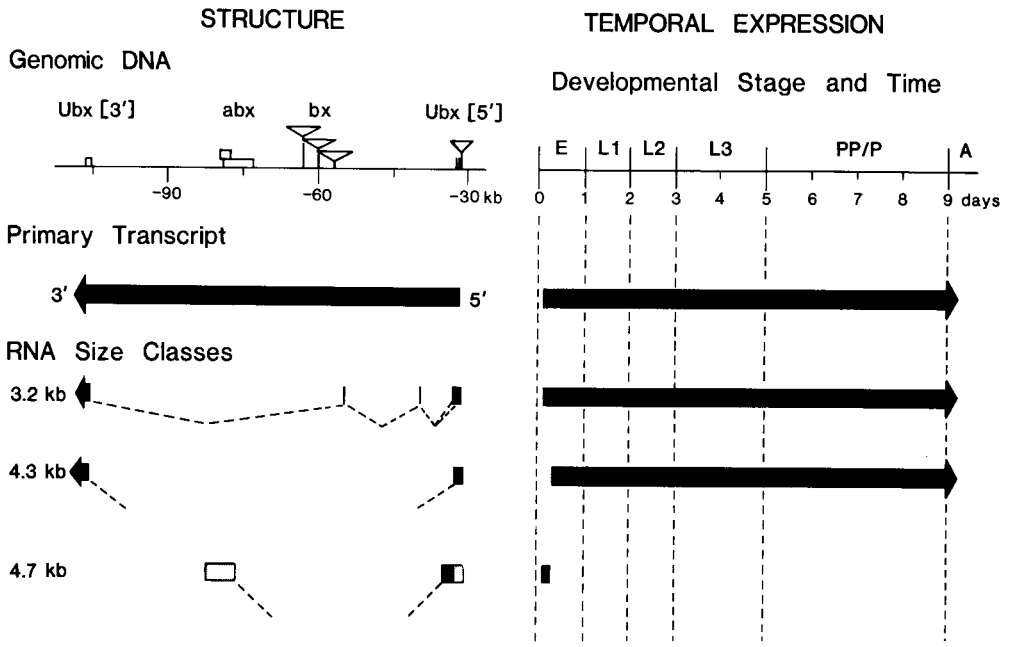

Figure 3. Structural and temporal expression of RNAs from the *Ubx* unit. The exon-containing regions are indicated by the filled boxes below the representation of the primary transcript. In the case of the 3.2-kb RNA size class, the dashed lines indicate the structure of the transcripts as defined by the cDNAs. Note the alternative donor splice site of the 5′ exon. The 3′ exon-containing region of the 4.3-kb size class can be seen to extend beyond that of the 3.2-kb class. Pseudopoint mutations in the genomic DNA are indicated by open boxes or vertical lines (deletions) and inverted triangles (transposable-element insertions). The gray-level of the boxes representing the 4.7-kb class indicates the relative strength of hybridization from these regions. The solid arrows and box on the right indicate the qualitative aspects of the temporal expression of each of the RNA size classes. (E) Embryonic; (L1, L2, L3) first, second, and third larval instar; (PP/P) prepupal and pupal stages; (A) adult.

domain (R. Saint et al., in prep.); and the 5′-exon sequence common to both ORFs is highly conserved in *D. melanogaster*, *D. funebris*, and *Musca domestica* (D. Wilde and M. Akam, pers. comm.; R. Saint et al., in prep.). Finally, the 3.2-kb RNAs in cytoplasmic extracts cosediment with polysomes of the appropriate size (D. Peattie, unpubl.).

The 4.3-kb-size class provides a likely additional source of mRNAs encoding *Ubx* proteins. They exhibit the same 5′ ends as the 3.2-kb mRNAs, and Northern

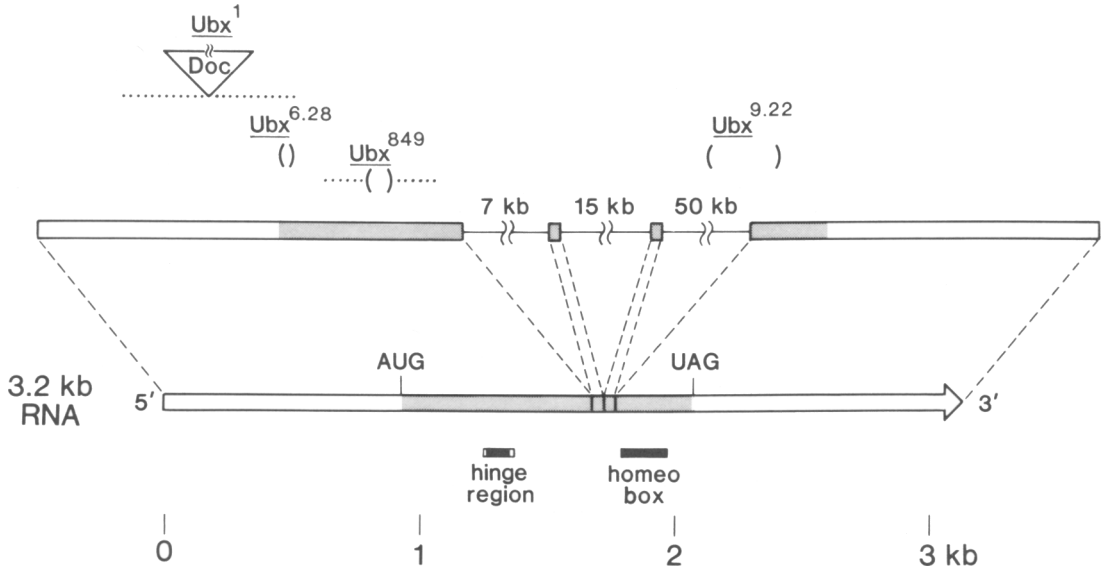

Figure 4. Structure and coding region of the *Ubx* 3.2-kb size class of RNAs. The orientation of this figure is opposite from that in Figs. 2, 3, 7, and 8. The locations of the *Ubx* pseudopoint mutations are shown above the horizontal bar representing the genomic DNA. The three introns are indicated by the thin lines. Stippled areas in the DNA and in the arrow below it representing the processed transcripts indicate the extent of the ORFs. The locations of the hinge region and homeo box are also shown. (Modified from Beachy et al. 1985.)

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

analyses with probes from the *Ubx* unit are indicative of strong sequence homologies between the 5′ and 3′ exons of the 3.2-kb mRNAs and the corresponding regions of the 4.3-kb RNAs; such analyses are not, however, sensitive enough to detect internal microexons. The two size classes differ in the times of their appearance during early embryogenesis and in their 3′ ends, that for the 4.3-kb class lying downsteam from that of the 3.2-kb class (Fig. 3) (R. Saint et al., in prep.). Although the exon structure of the 4.3-kb RNAs has not yet been determined by nucleotide sequence analyses of cloned cDNAs, the above characteristics suggest that they encode *Ubx* proteins with amino acid sequences akin to those encoded by ORF Ia and Ib.

What effects do the *Ubx* mutations have on the synthesis and/or structure of these two size classes of mRNAs and the proteins they encode? The *Ubx* mutations shown in Figure 2 form two classes. The "pseudopoint" class consists of three small deletions and a 4.3-kb insertion that are located within the *Ubx* unit near either of its ends. The other class consists of 12 "gross rearrangements" (translocations, inversions, and deficiencies), each with a single break within the *Ubx* unit. These breaks are scattered throughout the unit and are sealed by the fusion of the *Ubx* DNA with DNA from loci that are separated from the *Ubx* unit by large genomic distances.

Figure 4 shows the structure and distribution of the four pseudopoint mutations relative to the exon structure of the 3.2-kb mRNA and, more particularly, to its coding region (ORF Ia or Ib). Three of these mutations are deletions, one of which, $Ubx^{6.28}$, deletes only 32 bp from the coding region near its 5′ end (M. Akam, pers. comm.). The resulting shift in reading frame yields a stop codon shortly downstream from the deletion and should prevent the translation of all but a small fraction of ORF Ia and Ib. This conclusion has been verified by the observation that $Ubx^{6.28}$ homozygotes lack all *Ubx* proteins detectable in wild-type strains by anti-*Ubx* antibodies directed against that part of a β-Gal-*Ubx* hybrid protein encoded by codons 9–300 in ORF Ia (Beachy et al. 1985).

The other two pseudopoint deletions also eliminate sequences common to ORF Ia and Ib; Ubx^{849} deletes approximately 110 bp of the 5′ exon downstream from $Ubx^{6.28}$, and $Ubx^{9.22}$ deletes approximately 1.6 kb that includes part of the homeo box of ORF Ia and Ib and extends upstream to include the 3′-exon splice site and adjacent sequences in the last intron (Bender et al. 1983; R. Saint et al., in prep.; M. Akam, pers. comm.). Thus, all three pseudopoint deletions exhibit the common characteristic of altering the structure of the *Ubx* proteins encoded by ORF Ia and Ib. Furthermore, each will similarly alter any member of a family of *Ubx* proteins whose coding sequence begins with the AUG codon that initiates ORF Ia and Ib, extends downstream to include part or all of the approximately 110 bp deleted by Ubx^{849}, and by splicing, incorporates the homeo box. Such a family would thus be characterized by two constant regions that include, respectively, an amino-proximal sequence and the homeo box sequence. Members of this family would be distinguished by a variable region connecting the two constant regions (e.g., the proteins encoded by ORF Ia and Ib) and perhaps by another variable region downstream from the homeo box. Additional members of this family could be encoded by other ORFs derived from the 3.2-kb and/or the 4.3-kb mRNAs by differing patterns of splicing. Indeed, a hint of such additional members is given by the observation that SDS-gel electrophoresis of proteins extracted from embryos or haltere imaginal disks yields at least three bands capable of binding the anti-*Ubx* antibodies (White and Wilcox 1984; Beachy et al. 1985).

Because the only obvious common characteristic of these *Ubx* deletions is that each will alter the structure of all members of this family, and because each deletion is defective with respect to both the parasegment 5 and parasegment 6 functions (Kerridge and Morata 1982; Bender et al. 1983), we think that each of these functions requires one or more proteins from this family—a conclusion that explains why each function is dependent on the *Ubx* unit.

The structures of the other *Ubx* mutations shown in Figures 2 and 4 are consistent with this conclusion, since each can be interpreted as inhibiting the synthesis, or altering the structure, of all members of the above family of *Ubx* proteins. This is certainly the case for 10 of the 12 gross rearrangements whose breakpoints clearly lie between the 5′ end of the *Ubx* unit and its homeo box and may also be true for the two breakpoints nearest the 3′ end of the unit, although they have not been mapped at sufficient resolution to be sure (Bender et al. 1983; R. Saint et al., in prep.). The remaining mutation, Ubx^{1}, results from the insertion of a 4.3-kb transposable element called "Doc" that has been mapped to a region within the 5′ exon that overlaps adjacent parts of its untranslated and coding regions, as shown in Figure 4. Ubx^{1} homozygotes, like $Ubx^{6.28}$ homozygotes, lack *Ubx* proteins detectable by anti-*Ubx* antibody (S.L. Helfand, unpubl.). In sum, 14 of the 16 *Ubx* mutations clearly alter the structure or inhibit the synthesis of all members of the family, and the data for the other two are consistent with their having the same effect. We therefore propose that each *Ubx* mutation inactivates all domain functions by inactivating all proteins in this family.

This proposition does not exclude the possibility that one or more of the domain functions also require other products of the *Ubx* unit, particularly the 4.7-kb *Ubx* RNA (Fig. 3). This poly(A)$^-$ RNA has not been sufficiently characterized, however, to give more than a hint as to its possible functions. Northern analyses of this poly(A)$^-$ RNA indicate that it cannot encode proteins belonging to the above family since it lacks detectable sequences from the 3′ end of the *Ubx* unit, including the homeo box (R. Saint et al., in prep.). Although it contains some sequences from the 5′ exon of the 3.2-kb mRNA, it differs from this and the 4.3-kb mRNA in containing sequences just downstream from that exon, and preliminary evidence indicates that it contains sequences from a genomic region that overlaps the two

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

abx deletions (E. Gavis, unpubl.). The transient expression of the 4.7-kb RNA during embryogenesis indicates that its processing is controlled differently from that for the 3.2-kb and 4.3-kb mRNAs and suggests the possibility that it may act as an early determinant of metameric identity, as has been suggested previously in respect to parasegment 6 (Akam and Martinez-Arias 1985).

The Distribution of *Ubx* Proteins Correlates with the Distribution of Domain Functions

The proposition that all functions of the *Ubx* domain require members of the *Ubx* protein family predicts the presence of one or more of these proteins in each metameric unit to which a domain function has been assigned by the mutational data (Fig. 1). We have tested this prediction by an immunofluorescence assay of the *Ubx* proteins present in the CNS of approximately 12-hour-old embryos and in the larval imaginal disks (Beachy et al. 1985). Because the polyclonal antibody used in these assays is directed against that part of a hybrid protein encoded by codons 9–300 of ORF Ia, it should detect any *Ubx* protein containing the amino-proximal constant region of the family. It does not, however, effectively bind to the homeo box constant region, since it does not detect proteins encoded by other homeotic genes containing that box, as witnessed by the observation, noted above, that Ubx^{1} and $Ubx^{6.28}$ homozygotes lack proteins detectable by this antibody. The *Ubx* proteins exhibit a strong nuclear localization in all tissues where they have been detected by this antibody. Close examination of antibody-labeled nuclei indicates that the *Ubx* proteins are localized within the nucleus rather than being bound to the nuclear membrane (Beachy et al. 1985), as do preliminary results, indicating that these proteins exhibit sequence-specific binding to DNA (P.A. Beachy, unpubl.).

Figure 5 shows the results obtained when the CNS of an approximately 12-hour-old wild-type embryo was labeled with this antibody (see Fig. 6 for a schematic of this labeling). The distribution of label is in good agreement with the distribution of the domain functions given in Figure 1 and, consequently, with the above prediction. The most intense labeling is in parasegment 6, where virtually all nuclei are labeled at the same intensity. A single wave of label extends anteriorly with a trough in T3a and a peak in T2p, specific subsets of nuclei in each of these presumptive compartments of parasegment 5 being labeled. Seven waves of decreasing amplitude of label extend posteriorly from parasegment 6, with troughs in the posterior parts of A1-A7 and peaks in the anterior parts of A2-A8, each wave again corresponding to a parasegment. The number and arrangement of the subset of nuclei labeled in each of these abdominal peaks are similar, with the exception of the last, where the nuclei in A8 are too faintly labeled for comparison. These results have been confirmed by White and Wilcox (1984), using a monoclonal antibody induced by a β-Gal-*Ubx* hybrid protein in which the *Ubx*-encoded part was somewhat shorter than that used here.

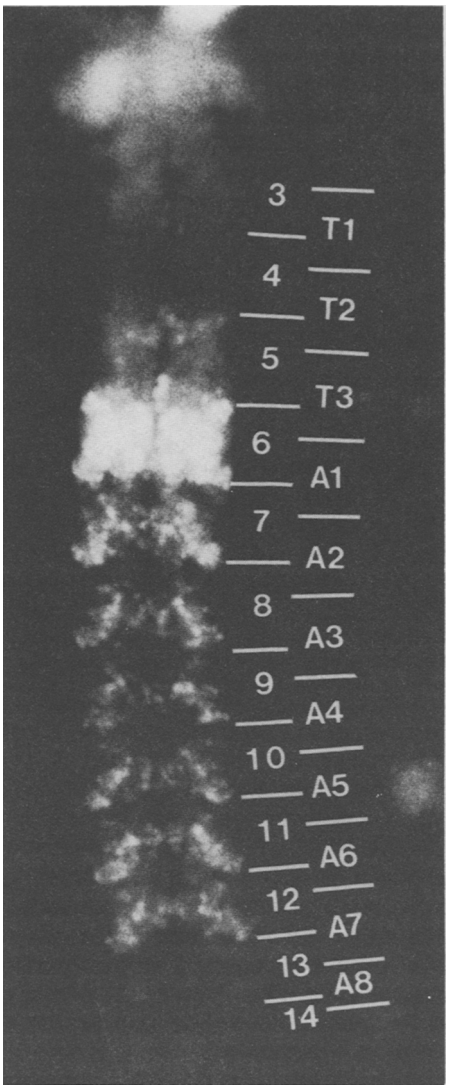

Figure 5. *Ubx* proteins in the CNS of an approximately 12-hr wild-type embryo. The parasegments and segments are indicated (for details, see text).

The distribution of *Ubx* proteins in the imaginal disks overlaps that for the mid-embryonic CNS at the qualitative level, although the disk distribution does not include A1-A8, since these adult epidermal segments arise from small larval histoblasts that were not examined (White and Wilcox 1984; Beachy et al. 1985). Both the T3a and T3p compartments of the third leg and haltere disks contain relatively high levels of the *Ubx* proteins in most if not all of their nuclei, although the levels are generally higher in the T3p than T3a nuclei of the third leg disk, which also exhibits localized heterogeneities along the proximo-distal axis. More anteriorly, the levels drop precipitously. In T2, the second leg disk exhibits localized low levels, and no *Ubx* proteins were detected in the wing disk proper, although

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

White and Wilcox (1984) found them in its peripodial membrane, which is of uncertain metameric origin. Similarly, no *Ubx* proteins were detected in T1 (first leg) and head (eye-antennal) disks, nor in the genital disk that contributes to the most posterior epidermal segment of the adult.

The low and high levels in the T2 and T3 disks, respectively, are consistent with clonal analyses which demonstrate that the *abx*$^+$ and *Ubx*$^+$ determinants of the PPX function of T2p are dispensable after a critical embryonic period, whereas the *bx*$^+$, *pbx*$^+$, and *Ubx*$^+$ determinants of the BX(T3a) and PBX(T3p) functions are not dispensable until after most of larval development (Lewis 1964; Morata and Garcia-Bellido 1976; Morata and Kerridge 1981; Kerridge and Morata 1982; Casanova et al. 1985). They are also consistent with the distribution of *Ubx* transcripts in the disks, as determined by in situ hybridization (Akam 1983) and by Northern analyses, which revealed the 3.2-kb and 4.3-kb mRNAs in isolated haltere disks but not in wing disks (R. Saint et al., in prep.). Akam (1983) also failed to detect *Ubx* transcripts in the wing disk by in situ hybridization, whereas low levels were obvious in the second leg disk. The presence of even low levels of *Ubx* transcripts and proteins in the second leg disk is curious at this late stage of larval development, particularly since temperature-shift experiments based on the cold sensitivity of PPX in *abx* mutants indicate that the PPX function is not required for second leg development after embryogenesis (Casanova et al. 1985). Although these *Ubx* products may represent residual ineffective expression of PPX proteins, their absence from the wing disk, whose development also requires PPX during embryogenesis (Lewis 1981, 1982; Kerridge and Morata 1982; Casanova et al. 1985), suggests that they may reflect another domain function not yet defined by mutation.

The *Ubx* protein and RNA distributions in the mid-embryonic CNS are also in good agreement (Akam 1983). Both these and the imaginal disk distributions fail, however, to resolve the important question of whether different members of the respective protein and mRNA families exhibit different distributions. Akam and Martinez-Arias (1985) have approached this problem by constructing hybridization probes that distinguish the 4.3-kb mRNAs from the sum of the mRNAs. Although these authors did not examine the mid-embryonic CNS (~12 hr), they did use these probes to determine that the 3.2-kb mRNA, but not the 4.3-kb mRNA, is expressed at 7–8 hours in the ectodermal cell layer containing CNS precursors with a pattern resembling that for the *Ubx* proteins in the mid-embryonic CNS; at a later stage, when the CNS is partially condensed, the probes for the 4.3-kb mRNA and for the sum of the two yield the same pattern. Hence, the only differential determinant for mRNA expression detected thus far is time (Fig. 3). The preparation of antibodies specific to individual proteins of the *Ubx* family may offer a more productive approach, particularly as such antibodies can be prepared on the basis of the nucleotide sequence of the respective cDNAs, e.g., antibodies directed against the sequence of nine amino acids that differentiates the proteins encoded by ORF Ia and Ib.

The *bxd* Unit Regulates the Metameric Distribution of *Ubx* Proteins

The conclusion that all domain functions, including those inactivated by mutations in the *bxd* region, require one or more proteins in the family encoded by the *Ubx* unit induces the proposition that the *bxd* region acts as a *cis*-regulator of the *Ubx* unit, particularly as the long *bxd* transcript does not appear to yield any mRNAs (see below). We have tested this proposition by examining the distribution of *Ubx* proteins in the mid-embryonic CNS of embryos homozygous for certain *bxd* mutations (Beachy et al. 1985). Teugels and Ghysen (1985) observed that extreme *bxd* mutations alter CNS development of both presumptive compartments of parasegment 6, in keeping with their effects on epidermal development. According to the above proposition, we should therefore expect such *bxd* mutations to alter the distribution of *Ubx* proteins in this parasegment and perhaps in more posterior parasegments where *bxd* mutations effect partial transformations in the larval epidermis (Lewis 1978, 1981; Bender et al., this volume).

Like the *Ubx* mutations, the *bxd* mutations divide into pseudopoint mutations and gross rearrangements with a breakpoint in the *bxd* region; the pseudopoint mutations that have been mapped each result from the insertion of a 7.3-kb transposable element of the "gypsy" family, although the orientation of insertion varies (Fig. 2) (Bender et al. 1983). We examined the *Ubx* protein distribution in homozygotes of two pseudopoint mutations (*bxd*^{1}, *bxd*^{55i}) located near the 3′ end of the *bxd* unit and one gross rearrangement, *bxd*^{113}, an inversion with a breakpoint near the center of the unit associated with a small deletion of DNA at or near the breakpoint (Fig. 7). This gross rearrangement exhibits a more extreme transformation of parasegment 6 in the adult epidermis than do the two pseudopoint mutations (Lewis 1981), one of which, *bxd*^{1}, was examined by Teugels and Ghysen (1985) and found to induce only minor alterations of the development of the CNS.

Although no obvious changes in the *Ubx* protein distribution were observed in homozygotes for the weaker pseudopoint mutations, the *bxd*^{113} homozygote produced a striking change as shown in Figure 6. As expected, parasegments 6–12 exhibit conspicuous changes in their *Ubx* protein distributions. The level of the *Ubx* proteins in the parasegment 6 nuclei is markedly reduced from that in the wild type, as is the level in parasegments 7–12, which also suffer a regionally specific approximately twofold reduction in the number of nuclei containing detectable *Ubx* proteins. However, the effect of *bxd*^{113} is not limited to these parasegments; both presumptive compartments of parasegment 5 exhibit an increase in the number of nuclei containing

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

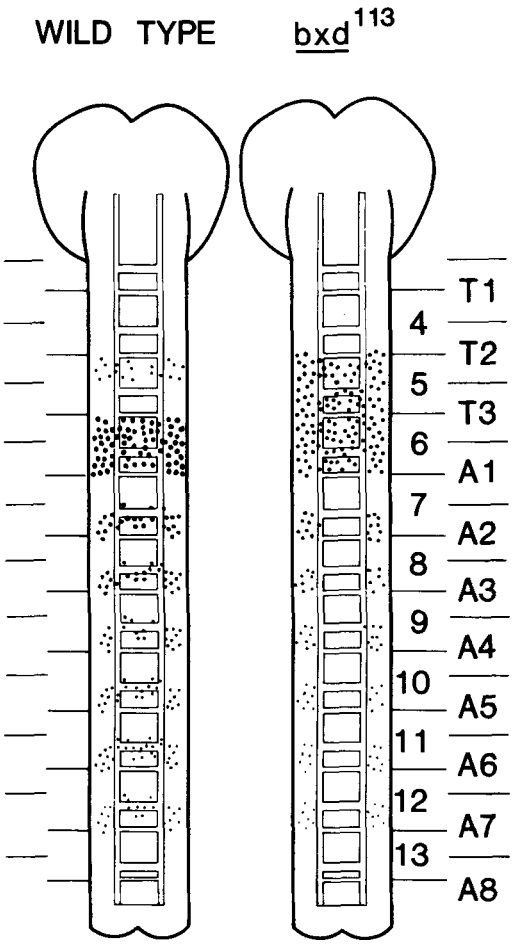

Figure 6. Spatial distribution of *Ubx* proteins in the CNS of wild-type and bxd^{113} mutant embryos. The parasegments and segments are indicated. The size of the dots correlates with the relative staining intensity of the nuclei, while the number of dots approximate the relative number of nuclei that are stained. In the wild type, the weak staining in T3a and A8 discussed in the text is not shown. (Modified from Beachy et al. 1985.)

Ubx proteins and in the amount of these proteins per nucleus. This increase appears to be coordinated with the decreased nuclear levels in parasegment 6 so that the two parasegments produce a uniform zone in which most of the nuclei contain equivalent amounts of the *Ubx* proteins. Evidently, position-specific controls on the expression of the *Ubx* unit are relaxed or overridden by the bxd^{113} mutation.

The change observed in parasegment 5 is curious since its epidermal identity does not appear to be affected by bxd^{113} or by *bxd* mutations generally, although they are known to modify the effects of other mutations in this region (Lewis 1955 and pers. comm.). Since other *bxd* gross rearrangements have recently been shown to cause similar changes in the *Ubx* protein distribution of parasegment 5 (S.L. Helfand, unpubl.), it is likely that this effect results from the loss of *bxd* DNA and not from the juxtaposition of novel sequences with the *bxd* DNA that remains linked to the *Ubx* unit. In any case, the important question for our immediate purpose is not whether, or how, a new distribution of *Ubx* proteins might cause a change in the morphological identity of a given parasegment in the CNS, but rather, whether *bxd* mutations can cause a change in the metameric distribution of *Ubx* proteins. Clearly, these mutations do cause such a change, and the fact that they do indicates that the *bxd* region regulates the expression of the adjacent *Ubx* unit in a position-specific manner.

Transcripts of the *bxd* Unit

Given that the *bxd* region acts as a *cis*-regulator of the *Ubx* unit, it is reasonable to suppose that its transcription plays a role in that regulation. This supposition implies that the transcripts of the *bxd* unit (*bxd* RNAs) function in another capacity than that of mRNAs. The peculiar properties of the early class of *bxd* RNAs described here indicate that they are not mRNAs, although the properties of the late *bxd* RNA strongly suggest it is an mRNA (M. Goldschmidt-Clermont et al., in prep.; H. Lipshitz; D. Peattie; both unpubl.).

The early class is first detected between 1.5 hours and 3 hours of embryogenesis and disappears by 6 hours (Fig. 7). Northern analyses with DNA probes specific for leftward transcription reveal two distinct sets of RNAs: a poly(A)$^+$ set with lengths in the 1.2 ± 0.1-kb range, and a heterogeneous smear of large poly(A)$^-$ RNAs that extends to lengths greater than 10 kb. The 1.2-kb poly(A)$^+$ RNAs are detectable with probes overlapping either end of the unit or a central region (Fig. 7), whereas the poly(A)$^-$ smear is detectable with all probes from the unit. No transcripts were detected with adjacent probes on the left or right or with probes specific for rightward transcription.

Members of the 1.2-kb poly(A)$^+$ set exhibit the same 5′ terminus and 5′ exon, as determined by S1-protection and primer-extension analyses at the nucleotide level. Similar S1-protection analyses covering the entire unit detect a central exon and a complex set of five exons near the 3′ end of the unit, of which four represent an overlapping series and the other lies downstream from this series without overlapping any of its members. The stoichiometry of the exon lengths suggests that the common 5′ exon is spliced to the central exon in all of the 1.2-kb RNAs, which vary according to which 3′ exon is spliced to the central exon. Figure 7 shows the structure of one of these RNAs in relation to sites of the eight *bxd* mutations given in Figure 2. This structure derives from the nucleotide sequence of a cloned cDNA that includes and extends from the 3′ poly(A) tail to within 50 nucleotides of the common 5′ terminus defined above and from sequence analyses of the *bxd* unit. The 1144 nucleotides in this RNA (exclusive of its poly[A] tail) derive from the common 5′ exon, the central exon, and the furthest downstream of the 3′ exons. Southern blot analysis of a second cloned cDNA establishes a structure like that of the first, except that its 3′ exon belongs to the set of overlapping exons located

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

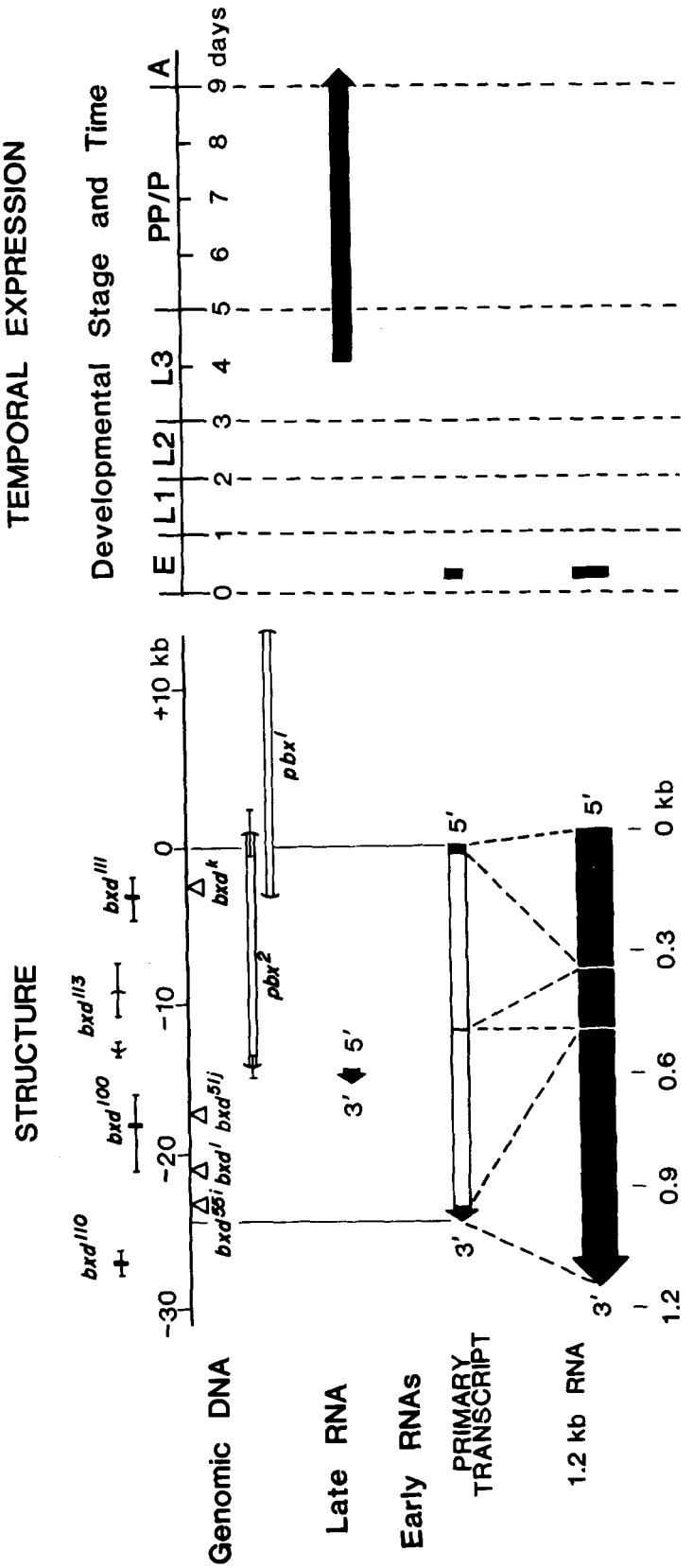

Figure 7. Structure and temporal expression of RNAs from the *bxd* unit. The extent of the *bxd* region DNA and the locations of the mutations discussed in the text are shown. The structures of the 0.8-kb late RNA and one of the 1.2-kb early RNAs are shown below. The exons contributing to the latter are shown as solid boxes within the arrow representing the early primary transcript.

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

just upstream of that in the first cDNA clone. Both cDNAs are therefore consistent with the prediction from the S1-protection data.

One reason for doubting that the 1.2-kb *bxd* RNAs are mRNAs is that the above 1144-nucleotide sequence does not contain any significantly long AUG-initiated ORFs. The longest ORF contains only 46 codons and is initiated in the 3′ exon, some 630 nucleotides from the 5′ end. Furthermore, it does not fit the codon usage frequencies for known structural genes in *D. melanogaster*, as do the ORFs in the *Ubx* mRNAs. Another reason for this doubt is that the 1.2-kb *bxd* RNAs exhibit an anomalous bimodal sedimentation pattern when embryonic extracts are zone-sedimented in a sucrose gradient. One mode overlaps polysomes small enough to represent the translation of a 46-codon ORF, whereas the other sediments as rapidly as the polysomes that translate the 3.2-kb *Ubx* mRNAs, whose ORFs would occupy the entire sequence of the above *bxd* RNA. It is difficult to account for this sedimentation pattern if the 1.2-kb *bxd* RNAs are mRNAs, particularly as preliminary sequence data indicate that the overlapping series of 3′ exons also lack ORFs of appreciable length. The sedimentation data suggest instead that these *bxd* RNAs may be packaged as RNP particles of two different-size classes.

The effects of four *bxd* gypsy insertion mutations (Fig. 7) on the 1.2-kb RNAs emphasize the questionable nature of their function. Northern analyses of the RNA in 0–6-hour embryos homozygous for bxd^{K}, bxd^{51j}, or bxd^{55i} indicate that neither the amount nor the lengths of the 1.2-kb RNAs are significantly altered by these mutations. In contrast, the bxd^{1} insertion causes a conspicuous increase in their lengths to yield a peak value of about 1.4 kb, again without change in amount. However, a spontaneous revertant of bxd^{1} (bxd^{1rev}) (Lewis stock 3-125), in which the gypsy element is replaced by a 0.5-kb insert presumed to be a gypsy terminal repeat (Bender et al. 1983), produces the same length distribution as bxd^{1}, although its adult phenotype is wild type. Taken together, these results indicate either that the 1.2-kb *bxd* RNAs are not required for the expression of the functions inactivated by gypsy insertions or that these insertions modify other characteristics of these RNAs, such as their spatial distribution, that have not been measured. In the latter case, one would also have to assume that the altered *bxd* RNAs in bxd^{1rev} retain the wild-type activity.

The late *bxd* RNA appears to be transcribed from a small central portion of this unit (Fig. 7) during a period extending from mid-third larval instar through the pupal and into the adult phases. Between the early and late periods, there is a quiescent period when no transcripts are detected by Northern analyses with probes covering the entire *bxd* unit. A single approximately 0.8-kb poly(A)$^+$ RNA is detected during the late period with a probe from the center of the unit that is specific for leftward transcription. None of the other probes from the unit detect this late RNA, nor do any of the probes detect the smear of high-molecular-weight poly(A)$^-$ RNAs that characterize the early class. Furthermore, S1-protection experiments with DNA probes covering the central region to which the 0.8-kb RNA sequences were localized by Northern analyses yield a single exon of approximately 630 nucleotides located at about −14 kb on the molecular map (Fig. 7). These data strongly suggest that the late primary transcript is not spliced and is initiated and terminated within the second intron for the 1.2-kb early RNA.

We have found the 0.8-kb RNA in the abdomen, but little or none in the thorax of dissected pupae. Although this result needs confirmation at the higher resolution afforded by in situ hybridization, it suggests that the 0.8-kb RNA may play a role in the BXD function of A1a of the adult. More specifically, it may encode a protein that acts in cooperation with proteins of the *Ubx* family to execute this BXD function—a possibility that is enhanced by cDNA sequence data demonstrating that the 0.8-kb RNA contains a significantly long AUG-initiated ORF that exhibits a good fit with the codon usage frequencies of known *D. melanogaster* structural genes (H. Lipshitz, unpubl.). In this respect, it should be emphasized that the preceding arguments demand only that *Ubx* proteins are necessary for the BXD function, but do not demand that these proteins are sufficient for this function. In the case imagined above, the adult epidermal BXD function would require a protein encoded by the late *bxd* RNAs in addition to the requisite *Ubx* protein(s).

Sequence Complementarity between the *Ubx* and Early *bxd* Transcripts

The peculiar properties of the early 1.2-kb *bxd* RNAs induced consideration of the possibility that the primary transcript of the *bxd* unit might be the *cis*-regulatory element by which the expression of the *Ubx* unit is controlled. More specifically, we imagined that the primary transcripts of the adjacent units might form a complex that would facilitate or inhibit the processing pathways of the *Ubx* transcript and thereby modulate the ratios among the processed *Ubx* RNAs.

We therefore used Southern hybridization techniques at low stringency to search the DNA of the two units for homologous sequences that might facilitate a specific interaction between their transcripts (H. Lipshitz, unpubl.). Two regions in the *bxd* unit (X^b and Y^b) were found that contain sequences homologous to those in two regions of the *Ubx* unit (X^U and Y^U, respectively; Fig. 8). The regions are intronic and lie in complementary order within the units (5′-Y^UX^U-3′ and 3′-Y^bX^b-5′). Interaction between the two transcripts could therefore occur in an inverted linear order by base-pairing in both regions if their transcript sequences are complementary. We have shown that this is the case for the X homology region (Fig. 8) (H. Lipshitz, unpubl.) but have not yet examined the Y region in this detail.

The nucleotide sequence of the 400-bp X^U region was determined and compared to that of the larger X^b region, whose sequence was known from previous studies (D. Peattie and L. Prestidge, unpubl.). The only sig-

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

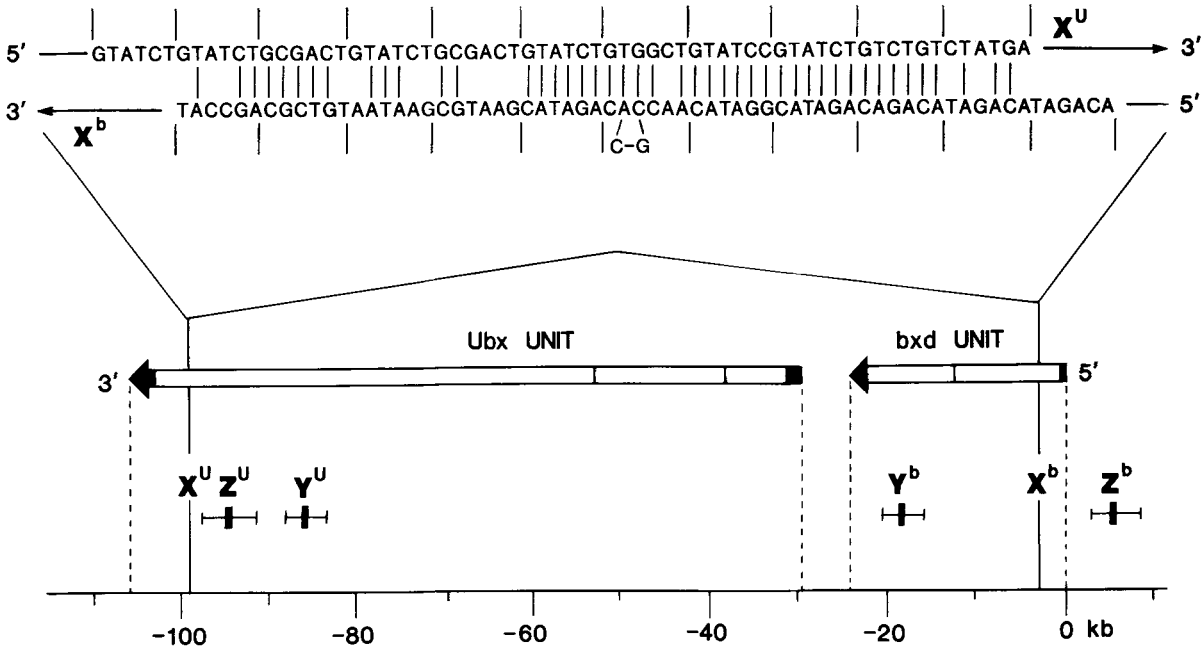

Figure 8. Homology between the *Ubx* and *bxd* regions. Homologies X, Y, and Z (see text) are shown as vertical bars above the horizontal line representing the genomic DNA of the *Ubx* domain. The extent of uncertainty in the location of these homologies is indicated. Note that only the X^b and Y^b homologies are within the *bxd* unit. The arrows represent the *Ubx* and *bxd* primary transcripts with the solid portions indicating the exons contributing to the 3.2-kb *Ubx* and 1.2-kb *bxd* RNAs. Details of the X^U and X^b homology are shown at the top. In each case, the sequence of the sense strand of the DNA is shown and base complementarity is indicated. The hexanucleotide repeats are also indicated.

nificant homology between the two is represented by an inverted repeat that provides the sequence complementarity between the two transcripts. In the configuration shown in Figure 8, the frequency of base pairing is 77% for a 57-base sequence, and within this region, 91% and 100% pairing is observed for 32- and 19-base sequences, respectively. Other configurations of significant pairing are possible because the sequences consist of an array of imperfectly repeated hexanucleotides. The consensus sequence for the 11 hexanucleotides in the sense strand of X^U is 5′-$G_{10}T_9A_7T_8C_9T_9$-3′, whereas that for the complementary 10 hexanucleotides in X^b is 3′-$C_6A_8T_5A_8G_5A_6$-5′, where the subscripts indicate the number of times the consensus base appears in the 11 or 10 hexanucleotides, respectively. (X^b is interrupted at its center by an octanucleotide.) Sequences homologous to those in the arrays appear to be present at a dozen or so other euchromatic loci, as assayed by whole-genome Southern hybridization with probes containing the arrays and by in situ hybridization to polytene chromosomes with genomic DNA clones identified by these probes.

Puzzling Aspects of Early *bxd* Transcription

Mutational analyses place certain constraints on the above processing model, or any model based on early *bxd* transcription. Since clonal analyses demonstrate that the bxd^+ DNA is not dispensable until the end of larval development (Morata and Garcia-Bellido 1976), it is clear that such models can account for only a part of all bxd^+ DNA functions. The putative *bxd* mRNA appearing late in development may account for some of this late requirement for bxd^+ DNA, but probably not all. Among the multiple domain functions associated with the *bxd* region, one might then expect that disruption of early *bxd* transcription would inactivate only a subset. Consequently, only functions that are commonly inactivated by mutations that disrupt early *bxd* transcription could possibly belong to the subset.

Among the four gross rearrangements shown in Figure 7, three (bxd^{100}, bxd^{113}, bxd^{111}) result from breaks that clearly map within the *bxd* unit and therefore will disrupt its early transcription. Direct evidence that the pbx^1 and pbx^2 deletions prevent early *bxd* transcription derives from our finding that 0–6-hour embryos of the pbx^1/pbx^2 heterozygote do not contain *bxd* RNAs detectable by Northern analysis (M. Goldschmidt-Clermont et al., in prep.). This result was expected from the mapping data for pbx^1, but not necessarily from that for pbx^2, since the data for this deletion are not sufficiently precise to be sure it includes the 5′ terminus of the *bxd* unit (Fig. 7).

A comparison of the phenotypes of this group of five mutations indicates that the BXD and BXD-like functions required for early larval A1-A7 development and the BXD function for adult A1a epidermal adult development do not require early *bxd* transcription because they are not commonly inactivated by these mutations. The PBX function for adult T3p epidermal development is commonly inactivated and consequently could require this transcription. Thus, the two rearrangements with breakpoints nearest the *Ubx* unit (bxd^{100}, bxd^{113}) inactivate the largest number of func-

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

tions. Each exhibits the T3p(PBX) and A1a(BXD) transformations of the adult epidermis and the three early larval transformations, namely, (1) appearance of thoracic ventral pits on A1-A7 (BXD and BXD-like), (2) partially developed thoracic Keilin organs on A1 (BXD), and (3) transformation of the A1a ventral setal belts to the T3a identity (BXD) (Lewis 1981). The *bxd*^{111} rearrangement exhibits the same transformations, with the exception of that for the larval ventral setal belts and a reduced adult A1a transformation represented by a lower frequency of extra legs in A1 (Lewis 1981 and pers. comm.). Finally, *pbx*^{1} exhibits only the adult T3p transformation (*pbx*^{2} transformations are similarly limited in the adult but have apparently not been examined in larvae; E.B. Lewis, pers. comm.).

Models based on early *bxd* transcription are further constrained by the metameric distribution of *bxd* transcription during its transient expression between approximately 2.5 and 6 hours of embryogenesis. The processing model is additionally restricted by the overlap of the *Ubx* transcription distribution with that of *bxd* transcription during this period, since this model requires coincident transcription of both units.

Detailed in situ hybridization analyses of the *Ubx* distribution at late cellular blastoderm (i.e., during the period of *bxd* transcription) indicate that virtually all transcripts (~95%) are restricted to parasegments 6–12, half of these forming a high, narrow peak at parasegment 6 (Akam and Martinez-Arias 1985). This distribution was obtained with a probe from the 5′ end of the unit that registers all nascent transcripts, as well as the two processed RNAs present at this time (3.2-kb and 4.7-kb RNAs). Given that the time required to transcribe the long *Ubx* unit (~75 min) represents a large fraction of the period of transcription initiation preceding the time at which this distribution was measured, we think that it is determined primarily by the nascent transcripts. Distributions for the early *bxd* transcript are reported in this volume (Akam and Martinez-Arias). These distributions were generated with a cDNA probe to the 1.2-kb *bxd* RNA shown in Figure 7. For this reason and because of the shorter length of the *bxd* unit, such distributions are probably not dominated by the *bxd* nascent transcripts. These distributions are relatively constant over time and do not exhibit a peak at parasegment 6. To the contrary, the level of *bxd* transcripts increases posteriorly from a low, probably insignificant, level in parasegment 6 to exhibit significant overlap with the *Ubx* distribution only in the seventh and more posterior parasegments.

This is not what would be expected were early *bxd* transcription required for the PBX function, the only allowed candidate from the mutational analyses, particularly since PBX is expressed only in the anterior portion of parasegment 6 (i.e., T3p). Given the apparent absence of *bxd* transcription in parasegment 6 of the wild type, how could the general loss of *bxd* transcription caused by the *pbx*^{1} deletion account for the inactivation of the parasegment 6 PBX function? Such a disruption of *bxd* transcription by *pbx*^{1} would be expected to have effects in more posterior parasegments where the levels are normally high, yet *pbx*^{1} has no effect on development in this region. The mutation and distribution data therefore argue against the relevance of early *bxd* transcription to the *cis*-regulatory functions of the *bxd* region. One can work around this argument by assumptions regarding, for example, the bias of the *bxd* probe for the processed *bxd* RNAs, but these are ad hoc assumptions that have no simple, attractive basis. Evidently, the interest in the striking sequence complementarity between the transcripts of the two units (Fig. 8) should be switched to an interest in the inverted repeat between the *bxd* and *Ubx* DNAs and its possible function.

Transcription Control by the *bxd* Unit and Concluding Remarks

The existing data on the early and late transcripts of the *bxd* region invite the surprising conclusion that the early RNAs, with all their intriguing properties, are not required for *cis*-regulation of the *Ubx* unit. A corollary to this conclusion is that the *cis*-regulatory elements of the *bxd* region consist of the *bxd* DNA. We assume that these DNA elements regulate the transcription of the adjacent *Ubx* unit; we call them transcription-control elements to avoid implication that they have the properties of enhancers or silencers. Given the different combinations of functions inactivated by the different *bxd* rearrangements and *pbx* mutations shown in Figure 7, one might expect that the *bxd* region is divisible into transcription-control elements, or combinations of these elements, that can be differentially regulated.

A striking correlation between the position of the breakpoints of *bxd* gross rearrangements and their functional inactivations is presented in this volume by Bender et al. in their study of an extensive set of rearrangements with breaks that extend from almost −30 kb to approximately +10 kb on the molecular map. That correlation can be approximated by the rule that the greater the length of *bxd* DNA between the 5′ end of the *Ubx* unit and breakpoint, the lower the level of inactivation—where level of inactivation is loosely defined by a mixture of degree of inactivation of a given function and the number of functions affected. Thus, breaks in the region of *bxd*^{110} and *bxd*^{100}, which lie closest to *Ubx* (Fig. 7), exhibit severe inactivation of all functions (see above), breaks lying at the other end of the *bxd* region inactivate functions to a lesser degree or not at all, and breaks in the central region caused graded inactivations, particularly of the A1a ventral setal band of early larvae. Curiously, the larval ventral pit transformation of A1-A7 is retained by all rearrangements, although it is absent in *pbx*^{1}, as noted earlier. A similar correlation between phenotypic effect and breakpoint position has been observed for rearrangements over an approximately 50-kb region in the achaete-scute gene complex of *D. melanogaster* by Campuzano et al. (1985).

It is not clear from the morphogenetic effects of these rearrangements whether the *bxd* region can be

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

simply divided into regions specific to the domain functions, particularly if the effects of the *pbx* deletions are to be accounted. Of course, it is a long molecular jump from a change in arrangement of the *bxd* DNA to a change in the arrangement of the setae in a ventral setal band. To shorten that jump, we have been examining the effects of this array of *bxd* rearrangements on the amount and arrangement of *Ubx* proteins in the mid-embryonic CNS, using the immunofluorescent techniques described earlier (Fig. 5). Although these studies are in the initial stage, the differences of the effects of rearrangements near the *Ubx* unit (bxd^{110}, bxd^{100}) and in the middle of the *bxd* region (bxd^{113}, bxd^{106}) on *Ubx* protein expression are striking (Beachy et al. 1985; S.L. Helfand, unpubl.). The differences between wild type and bxd^{113}, which appears to have a somewhat stronger effect on *Ubx* protein expression than bxd^{106}, are shown in Figure 6 and have been discussed. In the bxd^{110} homozygote (and bxd^{100}, which is similar to it), parasegments 5 and 6 also form a block of similarly stained nuclei (i.e., with the anti-*Ubx* antibody) but at a much lower level than in bxd^{113} and bxd^{106}, while the abdominal pattern of staining is more dramatically changed. No nuclei are stained in A7 and A8, and only a few nuclei are stained in A2-A6. Thus, the loss of the *bxd* DNA that is present in bxd^{113} (and bxd^{106}), but not in bxd^{100} (Fig. 7), results in a decrease of *Ubx* protein expression and, perhaps as a consequence of this decrease, a change in the abdominal pattern of nuclei containing *Ubx* proteins. Similarly, the loss of *bxd* DNA to the right of the bxd^{113} appears to cause the aberrant expression of *Ubx* proteins in parasegment 5 (particularly in T3a), as well as a decreased expression of these proteins in more posterior parasegments.

These interpretations of the data depend on the assumption that the imported DNA fused to the *bxd* DNA that remains linked to the *Ubx* unit has little or no effect, a reasonable assumption in view of the similar effects observed for different members of a given pair (i.e., bxd^{110}, bxd^{100}, or bxd^{113}, bxd^{106}), and the ordered morphological effects observed by Bender et al. (this volume). The similarity between the effects of bxd^{110} and bxd^{100} is not surprising, since rearrangements with breakpoints more proximal to the *Ubx* unit than the most proximal transcription-control element will have the same effect as those that break that element, as they will separate the *bxd* transcription-control region from the *Ubx* unit by long genomic distances, thus preventing its *cis*-regulatory functions. Consequently, it is difficult to define the *Ubx* proximal end of the transcription-control region with these rearrangements.

The linkage between the *Ubx* unit and the Y^b homology region is not broken by bxd^{113}, but may be in bxd^{100}; consequently, Y^b may be relevant to the difference of *Ubx* protein expression between these two rearrangements (Figs. 7 and 8). The X^b inverted repeat is also located in what appears to be a critical part of the *bxd* transcription-control region, as does another region of homology (Z^b; Fig. 8) located in the rightmost part of the *bxd* region. Its counterpart in the *Ubx* unit (Z^U) is located between X^U and Y^U. The position of these homology regions in the *Ubx* unit induces the question of whether members of the homology pairs may become coupled by proteins specific to each pair, thereby producing a chromatin loop that includes most of the *Ubx* unit and, most importantly, its 5′ end and upstream sequences. Such a coupling might alter the coiling properties of the chromatin within the loop to inhibit or facilitate *Ubx* transcription, different combinations of the coupling proteins in different cells thereby effecting cell-specific differential control of *Ubx* expression.

Finally, two unresolved questions deserve comment. The first is whether the different functions of the *Ubx* domain depend on different combinations of the *Ubx* proteins, as proposed by Beachy et al. (1985), and the second concerns the relevance of the early transcription of the *bxd* unit. The presumption that the *cis*-regulatory function of the *bxd* DNA consists in the control of *Ubx* transcription affects the first question only in denying a *cis*-regulatory role of that DNA in the regulation of the different processing pathways for the *Ubx* mRNAs. Clearly, different combinations of *trans*-regulatory factors for these pathways in the different metameres that require the *bxd* DNA for their development could lead to different combinations of *Ubx* proteins in those metameres. It is equally clear, however, that there is no need to assume differences in *Ubx* protein combinations to account for the different domain functions. For example, if the combination of *Ubx* proteins resulting from *Ubx* transcription were invariant, the different domain functions could result from the cooperative action of those proteins with others that are differently distributed among the metameres. There is, indeed, a rapidly expanding list of genes whose products may have *trans*-regulatory effects on the functions of the *Ubx* domain, including Polycomb (*Pc*) (Lewis 1978; Beachy et al. 1985), extra sex combs (*esc*) (Struhl 1981), Polycomb-like (*Pcl*) (Duncan 1982), super sex combs (*scx*), trithorax (*trx*) and certain segmentation genes (Ingham, this volume), and several others (Jürgens 1985) that probably include the *abd-A* and *Abd-B* domains of the BX-C. However, in none of these cases is it known whether the observed effect is direct or indirect, and as yet they are of little help in resolving the above question. Nor, as we have seen, are there adequate data regarding the distribution of the different *Ubx* mRNAs and proteins to resolve this question.

The early *bxd* transcription unit remains an enigma, despite a considerable knowledge of its products and its temporal and spatial expression during development. It would be most puzzling if a transcription unit with the properties of the *bxd* unit were not relevant to any BX-C function. Recall that it is an exceptionally long unit by *Drosophila* standards, that it occupies a majority of the *bxd* region, that its primary transcript is multiply processed in a specific manner, that its expression is limited to a crucial time during development, and that this expression is restricted to

Downloaded from symposium.cshlp.org on March 28, 2012 - Published by Cold Spring Harbor Laboratory Press

metameres whose identities are specified by the BX-C. Indeed, we think that it is relevant, but not to the *cis*-regulatory functions of the *bxd* DNA. Rather, we suppose that early *bxd* transcription, as well as the late transcription, yields *trans*-acting products and that the defects in the *cis*-regulatory functions caused by mutations in the *bxd* region so dominate the phenotype that defects in the *trans*-functions resulting from both early and late *bxd* transcriptions are not visible. Thus, if a mutation adversely affects the expression of the required *Ubx* proteins via a *cis*-regulatory defect, its effect on the expression of the *trans*-acting product of *bxd* transcription will be inconsequential to the phenotype. Clearly, we need mutations that affect the expression of these *trans*-acting products but not the *cis*-regulatory function. These may be difficult to obtain since their target size is apt to be small. However, the effect of such *trans*-acting products may be detectable when P-element-mediated insertions of the *bxd* DNA are tested against the pbx^2 deletion, which exhibits a lesser *cis*-regulatory defect than other mutations in the *bxd* region (Lewis 1982). If so, alterations in the *bxd* DNA could be made in vitro and tested in vivo. We are proceeding along these lines with the constructions of appropriate P-element cosmids (Steller and Pirotta 1985) with which we also plan to test the effects of the *bxd* DNA on the transcription of the *Ubx* unit.

ACKNOWLEDGMENTS

We thank Kenneth Burtis for his generous help in the analysis of the nucleotide sequence data and Ed Lewis, Welcome Bender, and Michael Akam for sharing their unpublished data and for many rewarding discussions. This work was supported by grants from the National Institutes of Health and the National Science Foundation to D.S.H. and by fellowships from the Helen Hay Whitney Foundation (H.D.L. and D.A.P.), the Jane Coffin Childs Foundation (M.G.-C. and P.J.H.), the National Institutes of Health (P.A.B., E.R.G., P.J.H., and S.L.H.), Damon Runyon–Walter Winchell Cancer Fund (R.A.S.), and the National Science Foundation (P.A.B.). Computer resources used in these studies included the BIONET™ National Computer Resource for Molecular Biology.

REFERENCES

AKAM, M.E. 1983. The location of Ultrabithorax transcripts in *Drosophila* tissue sections. *EMBO J.* **2:** 2075.

AKAM, M.E. and A. MARTINEZ-ARIAS. 1985. The distribution of Ultrabithorax transcripts in *Drosophila* embryos. *EMBO J.* **4:** 1689.

BEACHY, P.A., S.L. HELFAND, and D.S. HOGNESS. 1985. Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature* **313:** 545.

BENDER, W., M. AKAM, F. KARCH, P.A. BEACHY, M. PFEIFER, P. SPIERER, E.B. LEWIS, and D.S. HOGNESS. 1983. Molecular genetics of the bithorax complex in *Drosophila melanogaster*. *Science* **221:** 23.

CAMPUZANO, S., L. CARRAMOLINO, C.V. CABRERA, M. RUIZ-GOMEZ, R. ZILLARES, A. BORONAT, and J. MODELELL. 1985. Molecular genetics of the achaete-scute gene complex of *D. melanogaster*. *Cell* **40:** 327.

CASANOVA, J., E. SANCHEZ-HERRERO, and G. MORATA. 1985. Prothoracic transformation and functional structure of the *Ultrabithorax* gene in *Drosophila*. *Cell* **39:** 663.

DUNCAN, I.M. 1982. Polycomblike: A gene that appears to be required for the normal expression of the bithorax and antennapedia complexes of *Drosophila melanogaster*. *Genetics* **102:** 49.

HAYES, P.H., T. SATO, and R.E. DENNELL. 1984. Homoeosis in *Drosophila*: The Ultrabithorax larval syndrome. *Proc. Natl. Acad. Sci.* **81:** 545.

JÜRGENS, G. 1985. A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature* **316:** 153.

KERRIDGE, S. and G. MORATA. 1982. Developmental effects of some newly induced Ultrabithorax alleles of *Drosophila*. *J. Embryol. Exp. Morphol.* **68:** 211.

LAWRENCE, P.A. and P. JOHNSTON. 1984. The genetic specification of pattern in a *Drosophila* muscle. *Cell* **36:** 775.

LEWIS, E.B. 1955. Some aspects of position pseudoallelism. *Am. Nat.* **89:** 73.

———. 1963. Genes and developmental pathways. *Am. Zool.* **3:** 33.

———. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* **276:** 565.

———. 1981. Developmental genetics of the bithorax complex in *Drosophila*. *ICN-UCLA Symp. Mol. Cell. Biol.* **23:** 189.

———. 1982. Control of body segment differentiation in *Drosophila* by the bithorax gene complex. In *Embryonic development: Genes and cells* (ed. M. Burger), p. 269. A.R. Liss, New York.

———. 1984. Genetic control and regulation of developmental pathways. In *Role of chromosomes in development* (ed. M. Locke), p. 231. Academic Press, New York.

MARTINEZ-ARIAS, A. and P.A. LAWRENCE. 1985. Parasegments and compartments in the *Drosophila* embryo. *Nature* **313:** 639.

MCGINNIS, W., M.S. LEVINE, E. HAFEN, A. KUROIWA, and W.J. GEHRING. 1984. A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* **308:** 428.

MORATA, G. and A. GARCIA-BELLIDO. 1976. Developmental analysis of some mutants of the bithorax system of *Drosophila*. *Wilhelm Roux's Arch. Dev. Biol.* **179:** 125.

MORATA, G. and S. KERRIDGE. 1981. Sequential functions of the bithorax complex of *Drosophila*. *Nature* **290:** 778.

SÁNCHEZ-HERRERO, E., I. VERNÓS, R. MARCO, and G. MORATA. 1985. Genetic organization of *Drosophila* bithorax complex. *Nature* **313:** 108.

SCOTT, M.P. and A.J. WEINER. 1984. Structural relationships among genes that control development: Sequence homology between the Antennapedia, Ultrabithorax and fushi tarazu loci of *Drosophila*. *Proc. Natl. Acad. Sci.* **81:** 4115.

STELLER, H. and V. PIROTTA. 1985. A transposable P vector that confers selectable G418 resistance to *Drosophila* larvae. *EMBO J.* **4:** 167.

STRUHL, G. 1981. A gene product required for correct initiation of segmental determination in *Drosophila*. *Nature* **293;** 36.

———. 1984. Splitting the bithorax complex of *Drosophila*. *Nature* **308:** 454.

TEUGELS, E. and A. GHYSEN. 1985. The domains of action of Bithorax genes in *Drosophila* central nervous system. *Nature* **314:** 558.

WHITE, R.A.H. and M. WILCOX. 1984. Protein products of the Bithorax complex in *Drosophila*. *Cell* **39:** 163.