model[10]) we can explain more than 80% of the variance in the observed $l = 2$–3 geoid. This agreement is unlikely to be due to chance.

There are certainly large heterogeneities in the upper mantle which might be expected *a priori* to mask the signal from the lower mantle. For example, several seismic studies have reported a degree 2 velocity anomaly in the upper mantle which correlates well with the degree 2 geoid[23–25]. It is remarkable that the fit of the geoid is so good, as we have neglected upper mantle contributions. The longest-wavelength dynamic response functions for our successful models (U10 or C10), however, are quite small in the upper mantle. The $l = 2$–3 geoid is not as sensitive to density anomalies there as it is to those in the lower mantle; the effects of heterogeneities in the transition zone are compensated almost totally by surface deformation.

The seismic results show that at degrees 2–3, geoid lows are associated with fast lower mantle whereas highs are associated with slow lower mantle. We suggest that these velocity variations are caused by temperature differences of convective origin. This interpretation is supported by a very high correlation ($r > 0.99$) between the degree 2 distribution of hotspots and the radially-averaged degree 2 lower mantle seismic velocity anomalies. This supports the hypotheses that geoid lows result from areas where subduction has cooled the mantle[19] and that shielding of the mantle by continents results in geoid highs and hotspots[18]. Mantle-wide flow provides a simple explanation for the fact that hotspot provinces at the surface tend to overly the long-wavelength expression of hot lower mantle. The relatively higher viscosity we find for the lower mantle may help to explain the long time scales inferred to be associated with these phenomena.

We have confined our analysis here to fairly simple viscosity distributions that, nonetheless, have allowed us to explain most of the variance in the longest-wavelength components of the geoid. The viscosity models presented are not unique; viscosity values should be regarded as weighted averages over the two layers used. We can provide even better matches to the geoid using more complicated viscosity models, including, for example, a low viscosity channel beneath a high viscosity layer representing a lithospheric lid. The best two-layer viscosity model used in fitting the seismic results to the residual geoid has a smaller viscosity contrast between the upper and lower mantle than the two-layer model used to isolate and remove slab effects[10], consistent with the lithosphere or upper mantle having a lower effective viscosity at plate boundaries than in plate interiors. It is possible, however, using a uniform composition model with four viscosity layers, to achieve a self-consistent spherically symmetric model for both the inferred lower mantle density anomalies and subducted slabs (B.H.H., in preparation).

Finally, although the agreement at long wavelengths is impressive, the agreement at $l \geq 4$ between the predictions of these models and the geoid is not good. This is in part the result of oversimplification in the flow models, which contain only two layers and which assume a spherically symmetric effective viscosity. On the other hand, the disagreement between the seismic methods at $l \geq 4$ leads to the hope that improved seismic data (such as will be provided by the proposed Global Seismic Array) and new techniques for its analysis, as well as including results for upper mantle heterogeneity, will yield an understanding of the geoid at shorter wavelengths. The insights into mantle dynamics provided by even these relatively coarse seismic studies are extremely valuable. Further cooperative studies by seismologists and geodynamicists should lead to even better constraints on the dynamics of convection in the Earth's mantle.

We thank D. L. Anderson, C. Chase, R. J. O'Connell and D. J. Stevenson for helpful comments. This work was supported by NASA grants NAG5-315, NSG-7610, NAS5-27226, NSF grant EAR-8317623, by the Alfred P. Sloan Foundation (B.H.H.), and by a Bantrell Postdoctoral Fellowship (R.P.C.). Contribution no. 4065, Division of Geological and Planetary Sciences, California Institute of Technology.

Received 2 July; accepted 29 October 1984.

1. Kaula, W. M. *J. geophys. Res.* **68**, 473–484 (1963).
2. Guier, W. H. *Nature* **200**, 124–125 (1963).
3. Dziewonski, A. M., Hager, B. H. & O'Connell, R. J. *J. geophys. Res.* **82**, 239–255 (1977).
4. Dziewonski, A. M. *J. geophys. Res.* **89**, 5929–5952 (1984).
5. Comer, R. P. & Clayton, R. W. *J. geophys. Res.* (submitted).
6. Clayton, R. W. & Comer, R. P. *J. geophys. Res.* (submitted).
7. Pekeris, C. L. *Mon. Not. R. astr. Soc. Geophys. Suppl.* **3**, 343–367 (1935).
8. Richards, M. A. & Hager, B. H. *J. geophys. Res.* **89**, 5987–6002 (1984).
9. Ricard, Y., Fleitout, L. & Froidevaux, C. *Annls Geophys.* **2**, 267–286 (1984).
10. Hager, B. H. *J. geophys. Res.* **89**, 6003–6015 (1984).
11. Ashby, M. F. & Verrall, R. A. *Phil. Trans. R. Soc.* A**288**, 59–95 (1977).
12. Richards, M. A. & Hager, B. H. *EOS* **65**, 1092 (1984).
13. Lerch, F. J., Klosko, S. M. & Patch, G. B. *LAGE05(GEM-L2)* (NASA Technical Memorandum, 84986, 1983).
14. Nakiboglu, S. M. *Phys. Earth planet. Inter.*, **28**, 302–311 (1982).
15. Kaula, W. M. *The Nature of the Solid Earth* (ed. Robertson, E. G.) 385–405 (McGraw-Hill, New York, 1972).
16. Chase, C. G. *Nature* **282**, 464–468 (1979).
17. Crough, S. T. & Jurdy, D. M. *Earth planet. Sci. Lett.* **48**, 15–22 (1980).
18. Anderson, D. L. *Nature* **297**, 391 (1982).
19. Chase, C. G. & Sprowl, D. R. *Earth planet. Sci. Lett.* **62**, 314–320 (1983).
20. Bass, J. D. & Anderson, D. L. *Geophys. Res. Lett.* **11**, 237–240 (1984).
21. Crough, S. T. *A Rev. Earth planet. Sci.* **11**, 165 (1983).
22. Malin, S. R. C. & Hide, R. *Phil. Trans. R. Soc.* A**306**, 281 (1982).
23. Masters, G., Jordan, T. H., Silver, P. G. & Gilbert, F. *Nature* **298**, 609–613 (1982).
24. Woodhouse, J. & Dziewonski, A. M. *J. geophys. Res.* **89**, 5953–5986 (1984).
25. Nakanishi, I. & Anderson, D. L. *Nature* **307**, 117–121 (1984).

Segmental distribution of bithorax complex proteins during *Drosophila* development

Philip A. Beachy*, Stephen L. Helfand† & David S. Hogness*

Departments of *Biochemistry and †Biological Sciences, Stanford University School of Medicine, Stanford, California 94305, USA

The Ubx *and* bxd *transcription units comprise a single functional domain in the bithorax complex of* Drosophila melanogaster. *The segmental distributions and nuclear localization of proteins encoded by the* Ubx *unit have been determined by immunofluorescence staining with antibodies raised against a fusion protein containing* Ubx *coding sequences. Wild-type and mutant distributions are consistent with a model in which the protein-coding functions of the domain derive from the* Ubx *unit and are regulated by the* bxd *unit.*

THE metazoan genome faces a unique set of demands in its task of directing the complex process of multicellular development. Not only must the characteristics of individual cells be specified, but also the size, shape and other higher-order characteristics of the multicellular structures they constitute. In *Drosophila melanogaster*, the body plan of both larva and fly is characterized by a series of segments whose identities vary according to position along the anterior-to-posterior axis of the animal. The role genes play in the specification of these identities has been defined, in part, by homoeotic mutations that produce specific intersegmental transformations of identity by causing embryonic founder cells at one position to adopt the fates reserved normally

for those at another position. The genes defined by these mutations are thus presumed to mediate the developmental decisions that differentiate one segment from another.

The identities of the second and third thoracic segments (T2, T3) and all eight of the abdominal segments (A1–8) depend on a cluster of homoeotic genes known as the bithorax complex (BX-C)[1,2]. This complex is divided into three functional domains[3]. We concentrate here on the Ubx (Ultrabithorax) domain, which occupies the left or proximal one-third of the BX-C DNA (Fig. 1) and acts in the specification of the T2, T3 and A1 identities. More precisely, the multiple functions of the Ubx domain specify the identities of the four adjacent anterior (a) and posterior (p) compartments (T2p, T3a, T3p and A1a) that mark the anterior end of the segmental range of BX-C expression. The Ubx domain is defined by five classes of recessive loss-of-function mutations that can be divided into two groups[1,3–10]. One group consists of four classes (*abx*, *bx*, *bxd* and *pbx*) which serve to define individual, compartment-specific functions of the domain because each causes the homoeotic transformation of a different subset of the four compartments. By contrast, the fifth or *Ubx* class of mutations *cis*-inactivates all functions in the set defined by the other classes and thereby defines that set as a single functional domain; thus, *Ubx* mutations do not complement any of the other mutations and cause homoeotic transformations corresponding to the sum of those induced by the other classes.

Given this *cis*-linkage of domain functions, it is curious that the molecular mapping summarized in Fig. 1 reveals a division of the domain into two non-overlapping, commonly oriented transcription units: the ~75-kilobase (kb) *Ubx* unit, which is virtually coextensive with the region delineated by the sites of *Ubx* mutations and which includes the *abx* and *bx* loci, and the ~25-kb *bxd* unit, which is similarly coextensive with the *bxd* sites and overlaps the *pbx* mutations. The problem of how *Ubx* mutations can *cis*-inactivate all functions of the domain, while the other classes of mutations *cis*-inactivate specific subsets of these functions, is a fundamental aspect of the more general problem of how the bithorax complex specifies segmental identities. As a step towards a better definition of these problems we have begun an investigation of the proteins encoded by the *Ubx* unit.

Open reading fames in the *Ubx* unit

First, we summarize the relevant aspects of an analysis of the *Ubx* RNAs to be described elsewhere (Fig. 2 and R. B. Saint, P.A.B., D. Peattie, P. Harte and D.S.H., in preparation). Three major size classes of *Ubx* RNAs appear in the first 8 h of embryogenesis: a transient 4.7-kb poly(A)$^-$ RNA that disappears near the end of this time and two poly(A)$^+$ RNAs of 3.2 and 4.3 kb found at all subsequent stages of development. The 3.2-kb class first appears at ~3 h, slightly after the 4.7-kb and before the 4.3-kb RNAs. Figure 2 shows that the 3.2-kb RNAs contain a long AUG-initiated open reading frame (ORF) generated by two exons from the 5′ and 3′ ends of the *Ubx* unit and two internal 51-base pair (bp) micro-exons. Sequence analysis of two cDNA clones shows that either of two donor sites located 27-bp apart may be used to splice the 5′ exon to the first micro-exon. The resulting ORF Ia and ORF Ib contain, respectively, 380 and 389 codons in the same frame and have the capacity to encode two polypeptides of relative molecular mass (M_r) 39,000 and 40,000, which differ only by the insertion of nine amino acids between residues 247 and 248 of the shorter polypeptide.

Several indicators other than significant length suggest that these ORFs encode *Ubx* proteins: their codon usage agrees well with codon frequencies tabulated for known *D. melanogaster* structural genes (K. Burtis, unpublished); their 3′-exon regions contain a homoeo box (a ~180-bp conserved sequence found in several *D. melanogaster* homoeotic and segmentation genes, and in the genomes of many other species[11–14]); the 5′-exon region common to both ORFs is highly conserved among *D. melanogaster*, *Drosophila funebris* and *Musca domestica* (D.

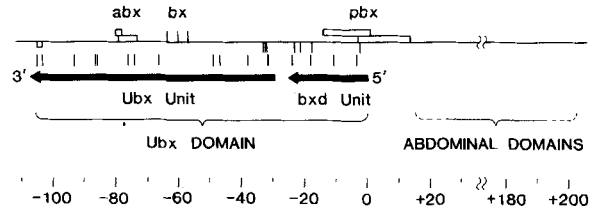

Fig. 1 Molecular map of the Ubx domain in the BX-C. The location of DNA lesions associated with recessive mutations in the *Ubx* domain are indicated in kb relative to the *Ubx* and *bxd* transcription units. The mutation sites are taken from Bender *et al.*[20] and from M. Akam and D.S.H. (unpublished data). Positions of the *Ubx* and *bxd* units are taken, respectively, from R. B. Saint, P.A.B., D. Peattie, P. Harte and D.S.H. (in preparation) and unpublished data of M. Goldschmidt-Clermont, D. Peattie, H. Lipshitz and D.S.H. The breakpoints of *Ubx* chromosomal rearrangements that split the BX-C are indicated by the vertical lines in the row just above the solid arrow denoting the orientation and extent of the *Ubx* unit. Four pseudo-point *Ubx* mutations (three small deletions and an insertion) are indicated by an open box (3′ end) and three closely spaced vertical slashes (5′ end) just below the long horizontal line representing the chromosomal DNA; they are also shown in Fig. 2 at higher resolution. The breakpoints of *bxd* rearrangements and the positions of *bxd* pseudo-point mutations are similarly indicated above the solid arrow delineating the *bxd* unit. The *abx* and *pbx* deletions are denoted by open boxes above the line whereas the sites of *bx* insertional mutations are indicated by vertical slashes.

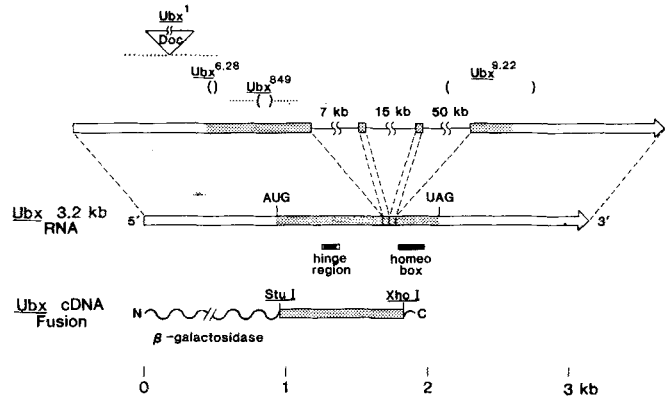

Fig. 2 The *Ubx* open reading frame used to construct the *Ubx/lacZ* fusion. In the upper line, which represents the genomic DNA of the *Ubx* unit, the four boxed regions denote exon sequences retained in the spliced mRNA. The mRNA contains a 380-codon ORF (Ia, in the stippled region), of which codons 9–300 were incorporated into the *Ubx/lacZ* fusion (bottom line). The locations of the *Ubx* pseudo-point mutations (ref. 20; M. Akam, personal communication; W. Bender, unpublished results; P.A.B. and D.S.H., unpublished) are indicated above the upper line: *Ubx*$^{6.28}$, *Ubx*$^{8.49}$ and *Ubx*$^{9.22}$ are associated, respectively, with deletions of ~32 bp, 110 bp and 1.6 kb; *Ubx*^{1} is associated with the insertion of a 'Doc' transposable element. The thin boxes indicate the positions of the homoeo box[11–14], and of the 'hinge region', which has been so named because of the many glycine codons. (Because of its high glycine content, the hinge region almost certainly exists within native *Ubx* proteins as a region devoid of secondary structure[34–36], perhaps functioning as a flexible connection between spatially distinct domains.) The glycine content of the entire hinge region is 73% (22 of 30 amino acid residues) and higher (89%, 17 of 19 amino acid residues) in the filled portion of the box. The gene fusion was constructed by standard methods[37], using the λgt11 expression vector[38]. The *Eco*RI linkers (BRL octamer) were selected to allow joining of the amino- and carboxy-terminal ends of the ORF Ia fragment (from *Ubx* cDNA pφ3602) in-frame with *lacZ* gene sequences. A recombinant phage, λLdm3751, was identified by its *lacZ*$^-$ phenotype, by hybridization to *Ubx* sequences and by production of a novel polypeptide of expected electrophoretic mobility in SDS gels on infection of *E. coli* host strain Y1089 (ref. 38; see Fig. 3).

Wilde and M. Akam, personal communication). In addition, Fig. 2 shows that of the four *Ubx* pseudo-point mutations which have been mapped, two ($Ubx^{6.28}$, Ubx^{849}) are short deletions in the 5′-exon region common to both ORFs; one ($Ubx^{9.22}$) deletes the entire 3′-exon region of each, and one (Ubx^{1}) consists of an insertion in or near their 5′-exon regions. The 12 other *Ubx* loci shown in Fig. 1 represent breakpoints of gross chromosomal rearrangements that separate the ends of the *Ubx* unit by large genomic distances; although these breakpoints are less informative than the pseudo-point mutations, they are consistent with a coding function for the ORFs because they should prevent formation of the appropriate mRNAs by interrupting transcription of the unit.

Working model for the Ubx domain

The location of the pseudo-point deletions and the conclusion that *Ubx* mutations inactivate all functions of the domain, suggest that each of these functions requires one or more proteins encoded by the *Ubx* unit and that these *Ubx* proteins form a family characterized by common amino-acid sequences in their amino- and carboxy-proximal regions. The members of the family would then be distinguished by differences in the amino-acid sequence connecting these two constant regions, exemplified by the two proteins encoded by ORF Ia and b. Other members of the family may be encoded by the 4.3-kb RNAs, which also contain sequences from both ends of the *Ubx* unit and exhibit the same 5′ initiation site as the 3.2-kb RNAs (R. B. Saint, P.A.B., D. Peattie, P. Harte and D.S.H., in preparation). Thus, we suppose that differential processing of the long *Ubx* primary transcript generates a set of mRNAs that encode the *Ubx* proteins and that specific subsets of these mRNAs are produced and translated in different compartments to provide the individual, compartment-specific functions of the domain. Certain predictions of this model are confirmed by the segmental distribution of the *Ubx* proteins described here.

Antibodies to *Ubx* proteins

The distributions examined here are those registered with a polyclonal antibody probe that should detect all members of the *Ubx* protein family because the antigen used to elicit these antibodies contains their common amino-proximal region. This antigen is a *Ubx*-β-galactosidase hybrid protein (fIa1) synthesized in *Escherichia coli* from a gene fusion consisting of codons 9-300 of ORF Ia inserted in-frame into the *E. coli lacZ* gene of the λgt11 expression vector, as indicated at the bottom of Fig. 2 and described in the Fig. 2 legend. To enrich for antibodies directed against *Ubx* determinants, the IgG fraction of serum from rabbits injected with purified fIa1 was affinity-purified by binding to fIa1, after which antibodies directed against β-galactosidase and other *E. coli* proteins were removed by adsorption (see Fig. 3 legend for details). Figure 3 shows that this antibody preparation reacts only with the *Ubx* determinants of fIa1 and of another *Ubx*-β-galactosidase hybrid (fIb1) containing the amino acid sequence encoded by codons 9-389 of ORF Ib.

Segmental distributions of *Ubx* proteins

We have used indirect immunofluorescence staining with these anti-*Ubx* antibodies to examine the prediction of the model that at least one member of the *Ubx* protein family will be expressed in each compartment whose identity depends on the functions of the Ubx domain. These functions are defined by the general rule that mutational inactivation of a given function transforms the compartment in which it is normally expressed to the identity of the homologous compartment in the anteriorly adjacent segment, where it is not normally expressed. Four functions are required to explain the compartmental transformations in the larval and adult cuticle caused by the five classes of mutations in the Ubx domain. These functions are referred to here as ABX, BX, PBX and BXD and are expressed, respectively, in the T2p, T3a, T3p and Ala compartments. The transformations caused by each mutational class can be explained if the following functional inactivations are assumed: *Ubx* mutations inactivate all functions; *abx* and *bxd* mutations inactivate the ABX, BX and BXD, PBX pairs of functions, respectively; *bx* and *pbx* mutations inactivate the BX and PBX functions, respectively (refs 1, 2, 4–10; G. Morata, personal communication).

According to the model, we therefore expect that *Ubx* proteins will be found in the cell lineages for T2p, T3a, T3p and Ala and will not be present in more anterior compartments, which are not affected by the five classes of mutation. Posterior to Ala, however, the situation is more complex. Although the primary BX-C determinants for segmental identities in the remainder of the abdomen derive from the other two BX-C domains[1,3], *Ubx* and *bxd* mutations cause subtle transformations of the larval cuticle in the anterior portions of A2–7, indicating that a BXD-related function is expressed in the presumed anterior compartments of these segments[1]. Thus, the model predicts that the cells of the anterior compartment lineages will also contain *Ubx* proteins.

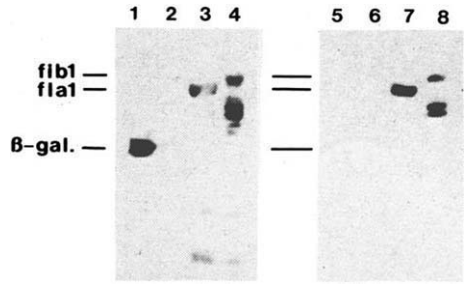

Fig. 3 Antibodies to *Ubx* determinants. A purified immunoglobulin fraction derived from hybrid protein-elicited antibodies reacts exclusively with *Ubx* determinants on SDS gel blots[39,40] of total protein extracts from *E. coli* lysogens. This fraction was prepared in two steps: (1) by affinity purification of antibody using the hybrid protein as ligand; (2) subsequent removal of β-galactosidase-reactive material (lanes 1–4, specificity of antibody binding after step 1; lanes 5–8, specificity after steps 1 and 2). Thus, in lanes 7 and 8, the hybrid proteins and their cellular breakdown products react, whereas in lanes 5 and 6 neither β-galactosidase nor other *E. coli* proteins are bound. Protein extracts in lanes 2 and 6 were from the $lacZ^{-}$ host cell, Y1089 (ref. 38). The remaining extracts were from induced lysogens of the following phage in the Y1089 host: lanes 1 and 5, the vector λgt11, carrying the $lacZ^{+}$ gene; lanes 3 and 7, λLdm3751; lanes 4 and 8, λLdm3752, a related onstruction incorporating codons 9-389 from ORF Ib (P.A.B. and D.S.H., unpublished data). The hybrid proteins encoded by λLdm3751 and λLdm3752 are labelled fIa1 and fIb1, respectively; their apparent M_rs are as predicted from the cDNA sequences (fIa1, 145,000; fIb1, 153,000).

Methods. After growth to $A_{600} = 0.5$ in L-broth, a lysogen of λLdm3751 (see Fig. 2) in Y1089 (ref. 38) was induced by a temperature shift (43°, 15 min) and addition of isopropyl thiogalactoside (IPTG; Sigma) to 5 mM. Cells were collected 45 min after IPTG induction and lysed by a freeze-thaw cycle following lysozyme digestion[41]. To the cleared lysate (supernatant of a 30 min, 200,000g centrifugation), solid ammonium sulphate was added to 35% of saturation at 0 °C; the resulting precipitate was resuspended and dialysed (Fraction I) and loaded in sample application buffer on 6% SDS gels[39]. After light staining with Coomassie Brilliant Blue, the band corresponding to the *Ubx*-β-galactosidase hybrid protein was excised, homogenized and emulsified with Freund's complete adjuvant. Primary immunization of virgin female New Zealand White rabbits was by subcutaneous (s.c.) injection of 250 μg (estimated by comparison with protein standards in Coomassie-stained gels) of hybrid protein per rabbit at ~20 sites. Booster immunizations were given 67 days after the primary immunizations by s.c. injection at ~10 sites with 100 μg of hybrid protein emulsified in Freund's incomplete adjuvant per rabbit. An IgG fraction was purified from serum collected 10 days after boosting[42] and was purified further using an affinity adsorbent made by immobilizing Fraction 1 proteins (the hybrid protein constitutes 10–20% of the total protein in Fraction 1) on derivatized beads (Affigel 10; Biorad)[43,44]. Antibodies directed against β-galactosidase or other *E. coli* proteins were removed by exhaustive treatment with an adsorbent similarly made with a protein fraction from a lysogen of λgt11 bearing no insert.

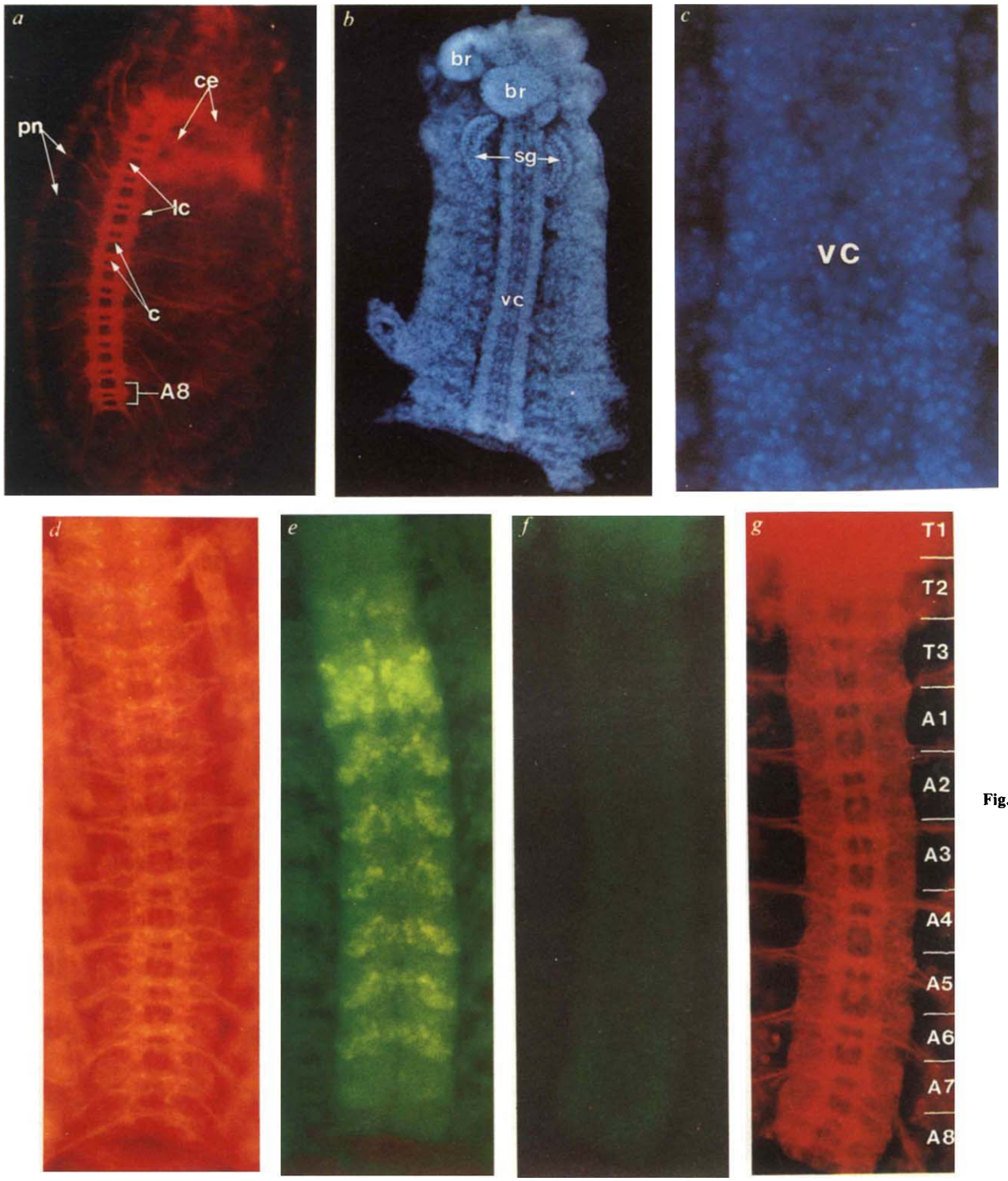

Fig. 4

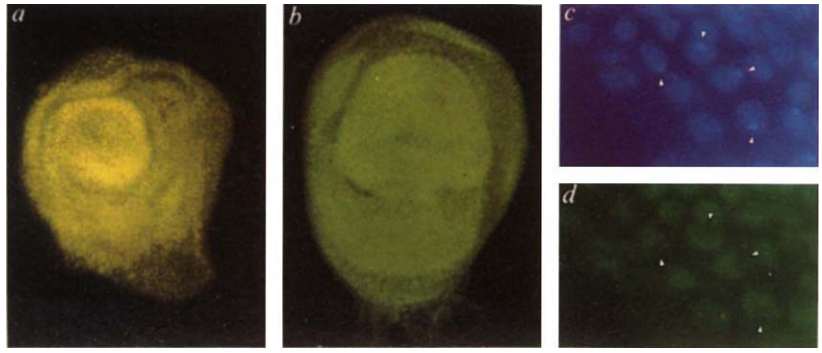

Fig. 5

We have examined the distribution of *Ubx* proteins in the central nervous system (CNS) at mid-embryogenesis and in the imaginal disks that are the precursors of the adult cuticular segments. The embryonic CNS is excellent for this purpose because it is easily accessible, expresses *Ubx* transcripts at relatively high concentrations[15] and its ventral nerve cord consists of readily distinguishable paired ganglia corresponding to T1–A8, as well as the more anterior pair of suboesophageal ganglia and the incomplete ninth abdominal ganglia (Fig. 4*a–c*, *g*). Figure 4*d*, *e* shows the simultaneous staining of a wild-type CNS with anti-tubulin and anti-*Ubx* antibodies, respectively. The anti-tubulin staining reveals the ladder-like axon scaffolding overlying the cell bodies of the nerve cord, whose paired rungs (commissures) provide the landmarks for identifying the segmental boundaries[16]. Using these landmarks, weak anti-*Ubx* staining is observed in the posterior portion of T2, followed by strong staining in posterior T3 and anterior A1, with the anterior portions of A2–7 exhibiting progressively weaker staining (see Fig. 6). Except for the very weak staining of anterior T3, the qualitative aspect of this pattern is that expected from the model, provided that the anti-*Ubx* staining is specific to *Ubx* proteins.

Figure 4*f*, *g* removes this proviso by showing that the anti-*Ubx* staining disappears in the embryonic CNS from a $Ubx^{6.28}$ homozygote. Preliminary sequence analysis (M. Akam, personal communication) indicates that the small $Ubx^{6.28}$ deletion near the 5′ end of ORF Ia and b (Fig. 2) shifts the reading frame to yield a stop codon downstream from the deletion. Clearly, the absence of protein interacting with the anti-*Ubx* antibodies is consistent with this sequence analysis and demonstrates that these antibodies react uniquely with the *Ubx* family of proteins.

Figure 5*a*, *b* shows that the haltere and third leg imaginal disks that form the T3 segment of the adult cuticle exhibit strong anti-*Ubx* staining. The wing and second leg disks of adult T2 exhibit little or no anti-*Ubx* staining, whereas the T1 first leg disk and the eye-antennal and genital disks are not stained significantly (data not shown). The results for the haltere and wing disks are consistent with the observation that the haltere (but not the wing) disk contains both the 3.2-kb and 4.3-kb size classes of *Ubx* RNAs (R. B. Saint, P.A.B., D. Peattie, P. Harte and D.S.H., in preparation). Furthermore, the pattern of anti-*Ubx* staining of both disks and the mid-embryonic CNS overlaps precisely the distribution of RNAs homologous to the 5′ exon of the *Ubx* unit, as registered by *in situ* hybridization[15], although the level of resolution for the RNA distributions is reduced considerably.

The sum of the CNS and disk distributions for *Ubx* proteins clearly fits the spatial localizations predicted by the model. Qualitative differences between the distributions are restricted to posterior T2 and anterior T3. The posterior T2 difference can be explained simply if it is assumed that ABX expression is required at the same time period in development of the adult cuticle and the CNS. Clonal analyses indicate that normal T2p development of the adult cuticle requires ABX function only during the first ~8 h of embryogenesis (ref. 6 and G. Morata, personal communication). As the CNS was assayed only a few hours after this critical period and the disks were assayed ~100 h later, the staining of posterior T2 in the embryonic CNS, but not the disks, may simply reflect the instability or dilution of the *Ubx* proteins and mRNAs associated with the ABX function.

The difference in staining of anterior T3 in the CNS and disks is of greater interest because it includes a staining difference between the two T3 disks. Haltere disk cells are stained strongly and uniformly with no anterior–posterior distinction. By contrast, cells of the third leg disk stain with markedly different intensities in defined regions, the most striking difference resulting from the weaker staining anterior compared with posterior cells. Thus, the anterior T3 staining increases from slight in the CNS to moderate and strong, respectively, in the third leg and haltere disks. This increase is correlated with the progressively more dorsal position of the precursors of these tissues on the dorsal–ventral axis of the blastoderm fate map[17], suggesting that the amount and/or combination of *Ubx* proteins expressed may be regulated according to position on this axis, just as it is regulated according to position on the anterior–posterior axis.

In this regard, it is of interest that *abx* mutations affect the development of posterior T2 and anterior T3 in both the CNS and the cuticle (ref. 18 and G. Morata, personal communication), whereas *bx* mutations produce little if any detectable effect in the CNS[18]. Thus, the dependence of BX function on bx^+ sequences observed in adult cuticle development does not obtain in the CNS, perhaps because of regulation along the dorsal–ventral axis such as that mentioned in the previous paragraph. By contrast, the ABX function defined for cuticle development could exist without modification in the CNS, as the requirement for abx^+ sequences is common to both.

Fig. 4 (left) *Ubx* proteins in the CNS at mid-embryogenesis. *a–c*, Fluorescence photomicrographs showing the neuropil (*a*) and cell nuclei (*b*, *c*) of the 12-h CNS. *d–g*, Simultaneous staining with anti-tubulin and anti-*Ubx* antibodies of CNS dissections from Oregon-R wild-type and *Ubx* mutant embryos. The axon bundles of the neuropil are visualized by indirect immunofluorescence staining (*a*) of intact embryos[45] using an anti-tubulin antibody (supplied by T. Karr from a hybridoma cell line established by S. Blose). In *b* and *c*, the CNS has been exposed by dissection[16], and the nuclei are visualized with Hoechst 22358 (bisbenzidine), a fluorescent DNA-binding dye[45]. In *c*, individual nuclei can be distinguished as brightly fluorescing points in a field of more diffuse light, which comes from the nuclei above and below the plane of focus. The paired commissures of the neuropil can be used as landmarks to identify the ganglia of the ventral cord: three thoracic pairs and eight complete abdominal pairs (the ninth abdominal ganglia are incompletely formed[46]; E. Teugels and A. Ghysen, personal communication). In embryos at this stage of development, a slight longitudinal compression of the commissural ladder is apparent in A8, perhaps due to the onset of ventral cord condensation[47]. Abbreviations: br, brain; ce, circumoesophageal connectives; c, commissures; lc, longitudinal connectives; pn, peripheral nerves; sg, salivary glands; vc, ventral cord. Oregon-R embryos displayed consistently a characteristic staining pattern with both antibodies (*d*, *e*) and the anti-*Ubx* stain (*e*) proved to be nuclear on comparison with bisbenzidine staining patterns (see Fig. 6 for a schematic representation). Of the 38 embryos from the balanced heterozygous stock examined ($Ubx^{6.28}$/TMI; see Fig 2), 10 (26%) did not stain above background with anti-*Ubx* antibodies (*f*), although the neuropil (*g*) appeared normal (colour differences of *f* and *g* from *e* and *d* are due to the use of a different photographic emulsion). CNS dissections and antibody staining were as described in ref. 16, and secondary antibodies were from Sigma (fluorescein-conjugated goat anti-rabbit IgG) and Cappel (rhodamine-conjugated goat anti-mouse IgG). The anti-*Ubx* antibodies used for CNS experiments were prepared as for lanes 5–8 in Fig. 3, but in addition were treated to remove any antibodies cross-reacting with other *Drosophila* antigens. The adsorbent for this treatment consisted of de-vitellinized embryos[45] which had been fixed at a stage before 2 h after fertilization, the time at which *Ubx* RNAs are first expressed.

Fig. 5 (left) *Ubx* proteins in imaginal disks. *a*, *b*, Anti-*Ubx* staining of the T3 imaginal disks; *c*, *d*, nuclear localization of *Ubx* proteins. Anti-*Ubx* antibodies react strongly with cells of the haltere disk and the third leg disks of third-instar larvae (*a* and *b* respectively; fluorescence photomicrographs were taken under slightly different filter conditions), but only weakly, if at all, with cells of other imaginal disks. Whereas cells of the haltere disk stain uniformly, cells of the third leg disk stain with markedly different intensities in defined regions. The most striking difference results from the weaker staining of anterior compared with posterior cells, but intensity differences between proximal and distal cells are also observed. Anti-*Ubx* staining (*d*) can be seen to co-localize with bisbenzidine fluorescence (*c*) in a high-magnification view of a field of cells from a partially everted haltere disk. *Ubx* proteins are apparently excluded from a region of the nucleus, possibly the chromocentre, which stains very brightly with bisbenzidine (see arrowheads). Disks were stained as described in Zipursky *et al.*[48]. All photomicrographs in this figure Fig. 4 were taken under epi-fluorescence illumination with Zeiss optics.

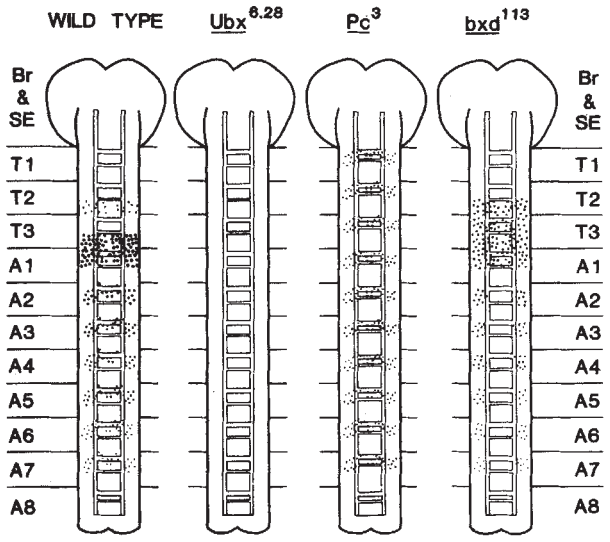

Fig. 6 Spatial distribution of *Ubx* proteins in the CNS of wild-type and mutant embryos. The results of anti-*Ubx* staining of CNS dissections of wild-type and mutant embryos are shown in schematic form. The number of dots in a given area is proportional to the number of nuclei stained and the relative staining intensities are indicated by dot size. In the wild-type CNS the most brightly staining nuclei are located in a region comprising posterior cells of the T3 ganglia (T3p) and anterior cells of the A1 ganglia (A1a). Most cells in this region are stained. More anteriorly, a subset of the cells in T2p is stained lightly, while the staining of T3a is so faint that it is not shown in the diagram. Posterior to the brightly staining T3p–A1a region, stained cells in segments A2–7 form a subset of the total complement of anterior cells. The number and arrangement of nuclei in these stained anterior patches is similar if not identical from segment to segment, although the intensity of staining decreases progressively in the posterior direction. (The extremely faint staining observed in A8 is not depicted.) Staining patterns of embryos collected from balanced heterozygous mutant stocks fell into two classes, of which the larger was indistinguishable from wild type and the smaller displayed the altered pattern shown in the figure. This variant class is presumed to represent the mutant homozygote as it generally comprised about a quarter of the embryos examined. Of the embryos collected from the $Ubx^{6.28}$/TM1 stock, 26% (10 of 38 individuals examined) did not stain above background levels, although the neuropil displayed no evident differences from the wild type (Fig. 4*g*). The variant classes from both *Pc* stocks examined (Pc^3/TM3 and Pc^2/TM1)[22,49] displayed similar differences in *Ubx* protein distribution and neuropil structure. These differences could be summarized as the preemption of the normal features of the thoracic and anterior abdominal segments by the features normally seen only in the most posterior abdominal segments. The proportion of abnormal individuals in the combined progeny from both *Pc* stocks was 28% (17 of 61). The variant class of the bxd^{113} stock examined (In(3)bxd^{113}/TM1)[11] displayed the modified staining pattern shown above, with the neuropil retaining its normal features; 10 individuals of the 31 examined (32%) from the bxd^{113} stock showed the abnormal pattern. Collection and staining of embryos was as described in Fig. 4 legend.

Regulation of the *Ubx* unit

One consequence of our model is that the *bxd* unit acts as a *cis*-regulator of the *Ubx* unit. This follows from the assumption that the individual functions of the *Ubx* unit, including those that are *cis*-inactivated by *bxd* mutations, are executed by specific combinations of *Ubx* proteins. The *bxd* unit delineated by these mutations would then *cis*-regulate the formation of particular members of the *Ubx* protein family, presumably by facilitating and/or inhibiting particular pathways by which the respective mRNAs are processed from the *Ubx* primary transcript. The idea that the *bxd* unit regulates the *Ubx* unit is also suggested by recent genetic evidence (ref. 19 and E. B. Lewis, personal communication).

Teugels and Ghysen[18] showed that extreme *bxd* mutations alter CNS development of posterior T3 and anterior A1, consistent with their effects on cuticular development. The strong anti-*Ubx* staining of these two regions (Figs 4*e*, 6) may then result from the combinations of *Ubx* proteins that execute the functions inactivated by these *bxd* mutations (that is, BXD and PBX). The staining of the anterior portions of A2–7 in the CNS may also be indicative of *Ubx* proteins with BXD-related functions, although this suggestion derives solely from observations of the subtle effects of *bxd* mutations on these segments in the larval cuticle[1]. To test these predictions, we have examined the distribution of *Ubx* proteins in the mid-embryonic CNS of the extreme bxd^{113} mutation (an inversion associated with a deletion of DNA at the breakpoint in the *bxd* unit) and of the weaker bxd^1 and bxd^{55i} mutations, each of which results from the insertion of a 'gypsy' transposable element in the *bxd* unit[20].

No obvious change in the *Ubx* protein distribution was observed in homozygotes for the weak *bxd* mutations, which is not surprising as Teugels and Ghysen observe only minor alterations in the CNS in a bxd^1 hemizygote[18]. By contrast, Fig. 6 shows that the extreme bxd^{113} mutation alters dramatically the segmental distribution of *Ubx* proteins. The expected regions are altered; the staining intensity of posterior T3 and anterior A1 cells is reduced markedly from that in wild type, as is the intensity in the anterior regions of A2–7, where the number of stained cells is also reduced by about twofold in a spatially specific manner. Both intensity and number of stained cells, however, is increased in posterior T2 and anterior T3; indeed, these two regions together with posterior T3 and anterior A1 form a block in which most cells are stained equivalently, as if the bxd^{113} mutation has relaxed a fine control over *Ubx* protein formation at a compartment or lower level. The change of pattern in posterior T2 and anterior T3 is not entirely unexpected, as in these regions, *bxd* mutations are known to modify the effects of other *Ubx*-related mutations (ref. 21 and E. B. Lewis, personal communication); in addition, changes in what is thought to be T3a of the CNS were observed by Teugels and Ghysen for the *bxd* mutations they examined[18]. Whatever the morphological transformations, the striking change in *Ubx* protein distribution caused by bxd^{113} strongly suggests that the *bxd* unit regulates expression of the *Ubx* unit.

Trans-regulation of *Ubx* protein distribution

The *Polycomb* gene (*Pc*) is a *trans*-regulator of BX-C expression located in the left arm of chromosome 3, far from the BX-C in the right arm[22]. Embryos lacking Pc^+ function suffer a massive homoeotic transformation by which all thoracic and abdominal cuticle segments exhibit identities characteristic of normal A8. Based on genetic evidence that this transformation requires the presence of the BX-C in the genome, Lewis[1] postulated that the Pc^+ product acts as a negative regulator of the expression of all genes in the BX-C. All BX-C genes were included in the scope of this regulator because he had proposed on other grounds that all BX-C genes are expressed in A8 (ref. 1); thus, it was argued that the transformation of all segments to A8 was equivalent to the activation of all BX-C genes in all segments by inactivation of the Pc^+ negative regulator.

We have tested this hypothesis by determining the distribution of *Ubx* proteins in the CNS of Pc^2 and Pc^3 homozygous embryos. Figure 6 shows that both the neuropil structure and the *Ubx* protein distribution are transformed in a manner closely approximating the cuticle transformation: the spacing between the paired commissures for the T1–A7 segments is reduced to the lesser spacing of wild-type A8 and the anti-*Ubx* staining pattern of segments T1–A7 appears intermediate between wild-type A7 and A8 (wild-type A8 staining is so faint that it is not noted in Fig. 6), although there is a gradual increase in intensity anteriorly. These results, however, are inconsistent with the simple hypothesis of a single *Pc* regulator for the BX-C. Clearly, other genes than *Pc* must function as regulators of the *Ubx* domain to account for the distribution observed in *Pc* mutants; indeed, there is a growing list of putative *trans*-regulators of

the BX-C[19,23-28] that may exhibit such a function and can be tested easily in the above manner. In addition, we are testing the possibility that other genes of the BX-C function in the abdomen to regulate expression of the *Ubx* domain. If these genes were themselves regulated by the Pc^+ product, they would be expressed ubiquitously in *Pc* mutants, resulting in the observed repetitive abdominal-like pattern of *Ubx* proteins.

Ubx proteins are nuclear proteins

Figure 5*c*, *d* shows that *Ubx* proteins are localized in the nuclei of the haltere disk cells. Nuclear localization was also observed for third leg disk and embryonic CNS cells and seems to be a general property of the *Ubx* proteins relevant both to hierarchical models of development (in which the BX-C products are postulated to directly regulate other genes[29,30]) and to speculation that homoeo boxes endow proteins with sequence-specific DNA-binding capabilities[31,32].

The homoeo box of the *Ubx* unit is located just downstream from the acceptor splice site for the 3′ exon region common to ORF Ia and b (Fig. 2). In our model for the *Ubx* protein family, the homoeo box is located in the carboxy-proximal constant region, near its boundary with the variable region. This location allows juxtaposition of the homoeo box with a variety of amino acid sequences that could modulate its function to create a diversity of related regulatory proteins. We emphasize that the function of the homoeo box need not be to bind DNA. For example, one can imagine equally that the variable region provides that function while the homoeo box provides a site by which *Ubx* proteins bind to a conserved protein, such as a histone, to create a perturbation in chromatin structure that has negative or positive regulatory value. Indeed, the *Ubx* proteins may regulate RNA processing rather than DNA transcription; it is known only that they accumulate in nuclei of segments whose identities are regulated by the Ubx domain.

Conclusions

The primary assumption of our working model for the Ubx domain is that all functions defined by mutations in the domain are executed by the proteins encoded by the *Ubx* transcription unit. That assumption leads to the prediction, confirmed qualitatively here, that the sum of the segmental distributions for these *Ubx* proteins will follow the sum of the segmental distributions for the functions of the domain. A second prediction of our model is that mutations in the other transcription unit of the domain, the *bxd* unit, will change the segmental distribution of *Ubx* proteins. This prediction has also been confirmed in a preliminary fashion; other *bxd* mutations, including their *pbx* relatives, must be tested, as must the prediction that this *bxd* effect is limited to *cis*-interactions.

A second assumption of the model is that a family of related proteins is encoded by the *Ubx* unit and that different combinations of these *Ubx* proteins execute the different functions of the domain, suggested by the observation that the *Ubx* unit produces a set of putative mRNAs, two of which contain ORFs that differ by a single step in their splicing pathways (ORF Ia and b). We have begun to test this part of the model by two methods: (1) the further definition of the *Ubx* RNAs via cDNA sequence analysis and the use of that sequence data to generate antibody probes specific to the different ORFs they may contain. For example, we have synthesized a peptide containing the nine-amino acid sequence that distinguishes ORF Ia from b and raised antibody against it for the identification of the ORF Ib protein. (2) We are performing a characterization of the proteins in *D. melanogaster* that react with the anti-*Ubx* antibodies. Although we have found multiple proteins in embryonic nuclear extracts with appropriate mobilities in SDS gels, the source of the apparent variation in M_r has not yet been determined. Similar observations were reported recently by White and Wilcox[33], who also obtained results similar to those reported here on the anti-*Ubx* staining of wild-type imaginal disks.

Perhaps the most novel characteristic of the Ubx domain is that its *cis*-interacting functions depend on two non-overlapping, commonly oriented transcription units; the most novel aspect of our model is that we assign the protein-coding functions of the domain to the unit containing the homoeo box and ascribe *cis*-regulatory functions to the other. The possibility that this novel organization occurs in other homoeotic loci is suggested by recent evidence that each of the other two domains of the BX-C contains a single homoeo box and multiple transcription units (F. Karch and W. Bender, personal communication; S. Sakonju and D.S.H., unpublished experiments).

We thank Corey Goodman for advice and equipment, Tim Karr for advice and antibodies and Michael Akam, Deborah Wilde, Welcome Bender, Francois Karch, Alain Ghysen, Erik Teugels, Ed Lewis and Gines Morata for communication of unpublished results. We also thank members of the Stanford Department of Biochemistry for advice and help, particularly, Howard Lipshitz and Renato Paro for mutant flies and Robert Baldwin and Peter Kim for discussions about protein structure. We also thank Jeremy Nathans and Howard Lipshitz for comments on the manuscript. This work was supported by a NIH grant to D.S.H.; P.A.B. was supported by a NSF graduate fellowship and a NIH training grant and S.L.H. by a NIH postdoctoral fellowship.

Received 31 December 1984; accepted 17 January 1985.

1. Lewis, E. B. *Nature* **276**, 565–570 (1978).
2. Lawrence, P. A. & Morata, G. *Cell* **35**, 595–601 (1983).
3. Sanchez-Herrero, E., Vernos, I., Marco, R. & Morata, G. *Nature* **313**, 108–113 (1985).
4. Lewis, E. B. *Am. Zool.* **3**, 33–56 (1963).
5. Lewis, E. B. in *Developmental Biology Using Purified Genes* (eds Brown, D. D. & Fox, C. F.) 189–208 (Academic, New York, 1981).
6. Morata, G. & Kerridge, S. *Nature* **290**, 778–781 (1981).
7. Kerridge, S. & Morata, G. *J. Embryol. exp. Morph.* **68**, 211–234 (1982).
8. Minana, F. J. & Garcia-Bellido, A. *Wilhelm Roux Arch dev. Biol.* **190**, 331–334 (1982).
9. Hayes, P. H., Sato, T. & Denell, R. E. *Proc. natn. Acad. Sci. U.S.A.* **81**, 545–549 (1984).
10. Struhl, G. *Nature* **308**, 454–457 (1984).
11. McGinnis, W. *et al. Nature* **308**, 428–433 (1984).
12. Scott, M. P. & Weiner, A. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4115–4119 (1984).
13. McGinnis, W. *et al. Cell* **37**, 403–408 (1984).
14. Carrasco, A. E. *et al. Cell* **37**, 409–414 (1984).
15. Akam, M. *EMBO J.* **2**, 2075–2084 (1983).
16. Thomas, J. B., Bastiani, M. J., Bate, M. & Goodman, C. S. *Nature* **310**, 203–207 (1984).
17. Garcia-Bellido, A. & Ripoll, P. in *Genetic Mosaics and Cell Differentiation* (ed. Gehring, W. J.) 119–156 (Springer, Berlin, 1979).
18. Teugels, E. & Ghysen, A. *Nature* (submitted).
19. Ingham, P. W. *Cell* **37**, 815–823 (1984).
20. Bender, W. *et al. Science* **221**, 23–29 (1983).
21. Lewis, E. B. *Am. Nat.* **89**, 73–89 (1955).
22. Lindsley, D. L. & Grell, E. H. *Genetic Variations of* Drosophila melanogaster (Carnegie Instn Wash. Publ. 627, 1969).
23. Ingham, P. & Whittle, R. *Molec. gen. Genet.* **179**, 607–614 (1980).
24. Capdevila, M. P. & Garcia-Bellido, A. *Wilhelm Roux Arch dev. Biol.* **190**, 339–350 (1981).
25. Struhl, G. *Nature* **293**, 36–41 (1981).
26. Duncan, I. *Genetics* **102**, 49–70 (1982).
27. Duncan, I. & Lewis, E. B. in *Developmental Order: Its Origin and Regulation* (ed. Subtelny, S.) 533–554 (Liss, New York, 1982).
28. Ingham, P. W. *Nature* **306**, 591–593 (1983).
29. Garcia-Bellido, A. *Ciba Fdn Symp.* **29**, 161–182 (1975).
30. Garcia-Bellido, A. *Am. Zool.* **17**, 613–629 (1977).
31. Laughon, A. & Scott, M. P. *Nature* **310**, 25–31 (1984).
32. Shepherd, J. C. W. *et al. Nature* **310**, 70–71 (1984).
33. White, R. A. H. & Wilcox, M. *Cell* **39**, 163–171 (1984).
34. Ananthanarayanan, V. S., Andreatta, R. H., Poland, D. & Scheraga, H. A. *Macromolecules* **4**, 417–424 (1971).
35. Scheraga, H. A. *Pure appl. Chem.* **50**, 315–324 (1978).
36. Schulz, G. E. & Schirmer, R. H. *Principles of Protein Structure* (Springer, New York, 1979).
37. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
38. Young, R. A. & Davis, R. W. *Science* **222**, 778–782 (1984).
39. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
40. Burnette, W. N. *Analyt. Biochem.* **112**, 195–203 (1981).
41. Fuller, R. S., Kaguni, J. M. & Kornberg, A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7370–7374 (1981).
42. Van der Zeijst, B. *et al. J. Virol.* **48**, 249–261 (1983).
43. Ikeda, Y. & Steiner, M. *J. biol. Chem.* **251**, 6135–6141 (1976).
44. Kristiansen, I. in *Affinity Chromatography* (ed. Hoffman-Ostenhoff, O.) 191–206 (Pergamon, Oxford, 1978).
45. Mitchison, T. J. & Sedat, J. *Devl Biol.* **99**, 361–264 (1983).
46. Hartenstein, V. & Campos-Ortega, J. A. *Wilhelm Roux Arch dev. Biol.* **193**, 308–325 (1984).
47. Kankel, D. R. *et al.* in *The Genetics and Biology of* Drosophila (eds Ashburner, M. & Wright, T. R. F.) 295–369 (Academic, London, 1980).
48. Zipursky, S. L., Venkatesh, T. R., Teplow, D. B. & Benzer, S. *Cell* **36**, 15–26 (1984).
49. Lewis, E. B. *Dros. Inf. Serv.* **55**, 207–208 (1980).