Connectin, a target of homeotic gene control in Drosophila

ALEX P. GOULD* and ROBERT A. H. WHITE

Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK

*Present address: National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Summary

The homeotic genes of Drosophila encode transcription factors that specify morphological differences between segments. To identify the genes that they control, we developed a chromatin immunopurification approach designed to isolate in vivo binding sites for the products of the homeotic gene Ultrabithorax. Here, we report the analysis of one immunopurified binding site. This 110 bp fragment maps within a regulatory region of a gene under homeotic control, connectin. A 4 kb DNA fragment, including the immunopurified binding site, is sufficient to reproduce the appropriate homeotic control within a subset of the full tissue distribution of connectin. Analysis of the role of the 110 bp binding site indicates that it mediates transcriptional controls by Ultrabithorax and other homeotic genes. This is the first report of a functional in vivo binding site isolated using the chromatin immunopurification method. We also show that the protein product of the connectin gene is predicted to be a cell-surface molecule containing leucine-rich repeats. The protein, connectin, can mediate cell-cell adhesion thus suggesting a direct link between homeotic gene function and processes of cell-cell recognition.

Key words: homeotic gene, Ultrabithorax, leucine-rich repeat, Drosophila development, homeobox.

Introduction

Differences between metameres in Drosophila are controlled by a small group of homeotic selector genes, expressed and required in specific anatomical domains (Lewis, 1963; Garcia-Bellido, 1975; Lewis, 1978; Akam, 1987). For example, the Ultrabithorax gene (Ubx) is expressed predominantly in a domain extending from parasegment (PS)5 to PS13 and is required primarily in the third thoracic and first abdominal segments (Lewis, 1978; Akam and Martinez-Arias, 1985; White and Wilcox, 1985). As long ago as 1975, it was proposed that the homeotic loci select particular developmental pathways by regulating a battery of subordinate target genes and that these target genes are required to coordinate the building of structures characteristic of each segment (Garcia-Bellido, 1975). More recently, the wealth of molecular data on homeotic genes have demonstrated that they encode sequence-specific DNA-binding proteins (Beachy et al., 1988; Muller et al., 1988) capable of activating or repressing transcription from a variety of reporter gene constructs in vitro (Thali et al., 1988; Krasnow et al., 1989; Winslow et al., 1989). However, the in vivo target genes have been difficult to identify (Andrew and Scott, 1992).

The classical genetic approaches, which have been so successful in identifying the early pattern-forming genes in Drosophila development, have so far failed to identify genes that act immediately downstream of the homeotics in the developmental genetic hierarchy, although Kennison and Tamkun (1988) have isolated a number of loci interacting with dominant homeotic mutations. Recently, Wagner-Bernholtz and colleagues have used P-element-mediated enhancer detection to screen for regulatory elements that respond, either directly or indirectly, to ectopic expression of Antennapedia (Antp). Using this approach, a previously unidentified gene, spalt major, has been identified as a target of negative regulation by Antp (Wagner-Bernholz et al., 1991). In addition, a few promising candidate targets have been identified from amongst the many genes that have previously been cloned. Perhaps the best characterised so far is the decapentaplegic (dpp) gene, which plays a role in both dorsoventral and proximodistal pattern formation (St. Johnston and Gelbart, 1987). The protein encoded by dpp is related to a vertebrate growth factor and functions in the visceral mesoderm in Drosophila to induce the expression of another homeotic gene, labial, in the midgut endoderm (Immergluck et al., 1990; Panganiban et al., 1990). Genetic experiments show that the dpp gene is positively regulated by Ubx and negatively regulated by another homeotic gene, abdominal-A (Immergluck et al., 1990; Reuter et al., 1990). However, none of the approaches described so far have been able to demonstrate a direct interaction between homeotic proteins and the regulatory DNA of a candidate target gene.

In an effort to identify direct targets of homeotic regulation, we have developed an immunopurification approach that enriches for short chromatin fragments through their association with endogenous Ubx protein (Gould et al.,...
1990). In brief, embryonic nuclei are digested with a restriction enzyme and then lysed; the soluble chromatin is affinity-purified against a matrix containing antibodies to Ubx proteins; the DNA fragments from the immunopurified chromatin are then cloned. Using this method, at least two candidate Ubx in vivo binding sites were isolated. Both target sites (clones 35 and 48) are capable of highly specific binding to Ubx proteins in vitro and are located in genomic DNA close to transcription units regulated by Ubx in vivo. These two transcripts were therefore presented as good candidates for genes directly regulated by Ubx (Gould et al., 1990). We have continued the analysis of these candidates in order to ask, firstly, whether the immunopurified target sites identify regulatory elements mediating homeotic control and, secondly, what can be learnt of the developmental function of these putative target genes.

Here we report the molecular characterisation of the gene, connectin, found adjacent to clone 35 and we demonstrate that the short (110 bp) immunopurified fragment is an important component of a connectin regulatory element under homeotic gene control. Nose et al. (1992) have independently isolated this gene from a screen of enhancer trap candidates for genes directly regulated by Ubx (Nose et al., 1990). In brief, embryonic nuclei are digested with a restriction enzyme and then lysed; the soluble chromatin is affinity-purified against a matrix containing antibodies to Ubx proteins; the DNA fragments from the immunopurified chromatin are then cloned. Using this method, at least two candidate Ubx in vivo binding sites were isolated. Both target sites (clones 35 and 48) are capable of highly specific binding to Ubx proteins in vitro and are located in genomic DNA close to transcription units regulated by Ubx in vivo. These two transcripts were therefore presented as good candidates for genes directly regulated by Ubx (Gould et al., 1990). We have continued the analysis of these candidates in order to ask, firstly, whether the immunopurified target sites identify regulatory elements mediating homeotic control and, secondly, what can be learnt of the developmental function of these putative target genes.

Materials and methods

Fly strains

Fly strains were maintained at 25°C on cornmeal-agar medium and embryos were collected onto yeast-agar juice agar plates. Standard genetic markers are as described by Lindsley and Zimm (1985, 1990). Wild-type embryos were of the Oregon R strain. Homoeotic null mutations used are as follows: Df(3R) bxd 1000; a deficiency of the entire Ubx gene (Lewis, 1978), abd A M1 and Ubx abd B1 (Sanchez-Herrero et al., 1985) and AnnpN10 (Wakimoto and Kaufman, 1981). Recipients for P-element transformation were homozygous for cn and ry.

General DNA techniques and sequencing

Standard techniques used are as described by Sambrook et al. (1989). An EMBl 3 genomic library made from the iso-1 strain (gift of J. Tamkun) was used to accomplish a genomic walk of about 65 kb at the cytological region 64C. The initial isolation of clones from this region has been described previously (Gould et al., 1990).

A 10 kb SalI genomic fragment, detecting the embryonic transcript distribution, was used to screen a λgt10 cDNA library constructed from 3-12 hour embryonic poly(A)+ RNA. Subsequently the longest hybridising clone, gt35.8 (2.6 kb), was used to screen two independent size-selected plasmid cDNA libraries, one from 8-12 hour embryos and the other from 12-24 hour embryos (Brown and Kafatos, 1988); 40 primary isolates were characterised and all were found to be of one class but with differing 5’ and 3’ extents. Sequence analysis shows that the 5’ end of pNB35.8 is identical with that of two other independent cDNA clones, one isolated from a separate library. However, the 3’ extents of several clones sequenced were found to be unique and in no case was a poly(A) tract found. It therefore seems likely that additional (but short) 3’ message sequences are not included within pNB35.8.

cDNA and genomic clones were sequenced by the chain termination method (Sanger et al., 1977) using double-stranded templates and Sequenase (USB). Large regions were covered using nested deletions together with primer synthesis. Nested deletions of gt35.8 were created using the exonuclease III/taung bean nuclease system (Stratagene) and sequenced as single-stranded M13 templates using Klenow polymerase.

Sequences were analysed using the Staden (Staden, 1982) and GCG (Devereux et al., 1984) software.

Whole-mount in situ hybridisations using digoxigenin-labelled DNA probes were carried out according to Tautz and Pfeifle (1989) with some modifications (Gould et al., 1990). For hybridizations using RNA probes, the digoxigenin-labelled RNA probes were made according to manufacturers instructions (Boehringer-Mannheim, RNA labelling kit) and the in situ hybridization carried out as for the DNA probes except that the prehybridizations, hybridizations and post-hybridization washes were at 70°C.

Transformation constructs

P-element mediated germ-line transformation of cn ry embryos was according to Spradling (1986) but using pUCHsΔ2-3 (gift of Ernst Hafen) as a constitutive source of transposase. Regulatory regions of the connectin gene were cloned into the blunt XhoI site of pHZ50PL (Hiromi and Gehring, 1987).

The 10 kb SalI genomic fragment was inserted into pHBluescript (Stratagene). From this 10 kb subclone, a 4 kb EcoRI fragment (representing the 4 kb EcoRI-SalI genomic interval) was excised, blunted with Klenow polymerase, and then inserted into pHZ50PL to give the 4 kb construct. The 1 kb PstI-SalI genomic region was subcloned into pHBluescript and excised as a blunt SmaI-HincII fragment for insertion into pHZ50PL yielding the 1 kb construct.

Precise deletion of clone 35 sequences (a 110 bp HaeIII fragment; Gould et al., 1990) was accomplished from a 0.7 kb PstI-SalI genomic subclone using the Inverse Polymerase Chain Reaction with Pfu DNA polymerase (Stratagene) at high stringency. Primers were chosen to generate an additional GC dinucleotide at the deletion junction, thus converting the two HaeIII half-sites into a diagnostic ApaI site. Sequencing confirmed that no inadvertant mutations had been introduced. A 0.4 kb SacI-SpeI fragment, containing the deletion of clone 35, was exchanged with the wild-type 0.5 kb SacI-SpeI fragment in both the 4 kb and 1 kb constructs to give the 4 kb-35 and 1 kb-35 constructs, respectively.

X-gal staining

X-gal staining was performed as described (Protocol 76; Ashburner, 1989b), except that 5% glutaraldehyde in PBS was used as the fixative. After the staining reaction, embryos were mounted in PBS, 10% glycerol and 0.1% Triton X-100 and photographed using differential-interference or bright-field optics.

Results

Cloning and sequencing of the connectin gene

The immunopurified 110 bp clone 35 was used as a probe to initiate a genomic walk at the cytological location 64C. Screening by in situ hybridization with genomic probes from this region revealed a transcription unit under homeotic gene control (Gould et al., 1990). To characterize this 35 transcription unit, we isolated many cDNAs covering the cloned interval, the longest being approximately 4.5 kb in length (pNB35.8). This is close to the size of the major polyadenylated RNA species (approximately
5 kb) detected on blots of 0-24 hour embryonic RNA using probes from this region (data not shown). Using this cDNA as a probe in embryo whole-mount in situ hybridisation, the same transcript distribution was obtained as had been seen previously with genomic probes.

To define the 5' and 3' limits of any open reading frame (ORF) present, several different cDNA clones were transcribed and translated in vitro. This indicated that both gt35.8 (2.6 kb) and pNB35.8 (4.5 kb) cDNAs encode a similarly sized polypeptide of about 72×10^3 M_r (data not shown). Sequencing of the gt35.8 cDNA revealed an ORF of 2046 nucleotides (ORF:35), encoding a polypeptide of 682 amino acids (76×10^3 M_r; Fig. 2). The codon usage found within this ORF:35 agrees well with that predicted for a Drosophila gene (Ashburner, 1989a). The second methionine codon of this reading frame is predicted to be slightly more favourable as a translation start than the first methionine codon, lying 27 nucleotides upstream (Cavener, 1987). Hydrophathy analysis (Fig. 3A) indicates that both the amino and carboxy termini are predicted to be highly hydrophobic. The amino-terminal hydrophobic region of 26 amino acids is characteristic of a signal sequence (von Heijne, 1986). There is no likely transmembrane domain but the carboxy-terminal hydrophobic region of 25 residues is consistent with the presence of a glycosyl-phosphatidyl inositol linkage to the cell membrane (GPI-anchoring; Ferguson and Williams, 1988). The existence of a GPI anchor has been confirmed by the phosphatidylinositol specific-phospholipase C-induced release of this protein from transformed Drosophila S2 cells expressing this ORF:35 (L. Meadows, A. P. G., R. A. H. W., unpublished results).

The most striking feature of the predicted polypeptide is the presence of 10 copies in tandem of a 24 amino acid leucine-rich repeat (LRR; Fig. 3B,C). This motif has a characteristic spacing of hydrophobic residues, in particular leucine and is found in many evolutionarily and functionally diverse proteins (Takahashi et al., 1985; Kataoka et al., 1985; Titani et al., 1987; Fisher et al., 1989; Reinke et al., 1988; Hashimoto et al., 1988; Rothberg et al., 1990). The consensus motif for the LRRs of ORF:35 is shown in Fig. 3B. In several molecules containing blocks of LRRs, there are also conserved flanking regions containing cysteine clusters (Lopez et al., 1987; Hickey et al., 1989; Schneider and Schweiger, 1991). Mutations affecting cysteine residues in a carboxy-terminal flanking region of the Drosophila LRR-protein Toll give rise to dominant Toll mutations (Schneider et al., 1991). ORF:35 shows only very limited homology to these conserved flanking sequences; however, there are cysteines at positions 100, 110, 112 and 126 followed by the sequence DTT occurring just on the amino-terminal side (137-139) of the LRR block and ICDRCL on the carboxy-terminal side (409-413). The other notable feature of the sequence of ORF:35 is that the amino-terminal region preceding the LRR block is serine-rich.

Comparison of the sequence of ORF:35 with that from a gene isolated by Nose et al. (1992) revealed identity and we refer to this gene as connectin.

### The expression of the connectin gene

Before considering the regulation of the connectin gene, we first discuss its expression pattern in more detail than has been presented previously (Gould et al., 1990). Transcripts are first detected by in situ hybridization in the extended germ band (stage 10; stages according to Campos-Ortega and Hartenstein, 1985) in a subset of the splanchnopleura, the cell layer that will subsequently give rise to the visceral mesoderm (Fig. 4A). There is expression in the presumptive foregut and hindgut mesoderm and in a segmentally repeated subset of the presumptive midgut visceral mesoderm. The labelling in the mesoderm around the foregut and hindgut persists but the expression in the midgut visceral mesoderm subsets narrows to small segmentally repeated patches by the end of stage 11 (Fig. 4B) and disappears after germ band retraction. In late stage 11, connectin RNA appears in the ventral nervous system and in the somatic mesoderm (Fig. 4B). The labelling in the ventral nervous system is initially in two longitudinal bands of cells with little segment-to-segment modulation in intensity. In contrast, the initial expression in the somatic mesoderm is highly modulated, being strong in the three thoracic segments and relatively weak in abdominal segments. There

---

**Fig. 1.** Organisation of the 35 transcription unit. The positions of restriction sites for HindIII (H), EcoRI (R), and SalI (S) are indicated above a line representing the genomic walk (scale above line in kb). Restriction fragments hybridising to cDNA clone pNB35.8 (see text and Materials and methods) are shown as solid bars beneath the walk with the 5' to 3' direction indicated. The location of the 110 bp clone 35 used to initiate this genomic walk is indicated, as are the extents of the lambda phage covering the cloned region. The position of the EcoRI-SalI fragment used in the 4 kb transformation construct is also shown.
Fig. 2. Nucleotide and predicted amino acid sequences of the 35 cDNAs. The sequence of nucleotides 1 to 343 is from cDNA pNB35.8, whilst nucleotides 344 (arrowhead) to 2973 are from the overlapping clone gt35.8 (see Materials and methods). Conceptual translation is shown for the N-terminal 367 amino acids with the potential N-linked glycosylation sites underlined. Amino and carboxy terminal amino acid sequences are compared with those of connectin, a developmentally expressed, trans-synaptic adhesion glycoprotein in Drosophila 

are differences, too, between the thoracic segments with the labelling in T1 being slightly weaker and having a different arrangement of labelled cells (Fig. 4C,D). By stage 14 (Fig. 4E,F), the intensity differences between segments in the somatic mesoderm are less pronounced but there are differences in the arrangement of mesodermal cell groups between thoracic and abdominal segments. From stage 12 onwards, the initially uniform labelling is succeeded by a more modulated pattern. Differences between the thoracic and abdominal neuromeres appear at stage 14 and, by stage 16, there is strong labelling in the anterior central nervous system, including the head ganglia. Labelling is also seen associated with the gnathal sense organs from stage 11 onwards.

R. A. H. W., unpublished observations; see also Nose et al. (1992). The labelling in the mesoderm declines shortly after stage 14. In the ventral nervous system, the initially uniform labelling is succeeded by a more modulated pattern. Differences between the thoracic and abdominal neuromeres appear at stage 14 and, by stage 16, there is strong labelling in the anterior central nervous system, including the head ganglia. Labelling is also seen associated with the gnathal sense organs from stage 11 onwards.
Hydropathy plot using the method of Kyte and Doolittle (1982). Significant regions of hydrophobicity (scoring positively) are indicated by black boxes. Shaded boxes connect in amino acids boxed. Bracketed numbers to the right indicate the protein precursor. Amino (N) and carboxy (C) terminal (B) Alignment of individual LRRs showing highly conserved N-linked glycosylation sites.

**Fig. 3.** Analysis of the **connectin** protein precursor. (A) Hydrophathy plot using the method of Kyte and Doolittle (1982). Significant regions of hydrophobicity (scoring positively) are found at both the amino and carboxy termini. (B) Alignment of individual LRRs showing highly conserved amino acids boxed. Bracketed numbers to the right indicate the position of each 24 amino acid LRR within the protein precursor. A consensus sequence for the 10 LRRs is shown below. (C) Schematic representation of the amino acid sequence of the protein precursor. Amino (N) and carboxy (C) terminal hydrophobic regions are indicated by black boxes. Shaded boxes show the position of individual LRRs and asterisks mark potential N-linked glycosylation sites.

**Transcription driven by a 4 kb region encompassing clone 35**

The **connectin** gene was isolated as a transcription unit neighbouring the immunopurified clone 35, a candidate in vivo binding site for **Ubx** protein. We have previously shown that **connectin** expression is under the control of **Ubx** and other homeotic genes (Gould et al., 1990). If the immunopurification has identified a functional **Ubx** protein binding site, it should be possible to demonstrate that the 110 bp clone 35 sequence mediates **Ubx**-dependent regulation of **connectin**. The approach used was first to ask whether the immunopurified fragment had led us to a **connectin** regulatory element under homeotic gene control. We considered it unlikely that the 110 bp sequence would function efficiently by itself as an enhancer element and so larger genomic restriction fragments spanning the clone 35 region were tested for their ability to drive expression of a lacZ reporter gene from an hsp70 minimal promoter in transformed embryos (Hiromi and Gehring, 1987; Fig. 5). In Fig. 6, we show that the 4 kb **SalI-EcoRI** fragment encompassing clone 35 is sufficient to give a consistent pattern of lacZ expression corresponding to a subset of the total **connectin** expression pattern (compare Fig. 4D-F and Fig. 6A-C). (We note that this element failed to give the same staining pattern when placed upstream of the hsp70 promoter in the opposite orientation and therefore does not behave like a true enhancer.) High levels of reporter gene expression were detected, from the extended germ band stage onwards, in small groups of cells in the gnathal and thoracic segments and in a posterior segment, probably A9 (Fig. 6A-C). The pattern of gnathal staining correlates with the developing gnathal sense organs, which express **connectin** RNA, and the thoracic labelling correlates with the ventrolateral **connectin** expression in the somatic mesoderm described above for stage 13 embryos (Fig. 4E,F). One high-expressing line, 4.40, also showed some abdominal labelling (Fig. 6A) which correlated well with the positions of the ventral and lateral clusters of **connectin** expression in the abdominal somatic mesoderm. In marked contrast to the endogenous **connectin** RNA distribution, no staining is seen in the central nervous system (Fig. 6B). Staining in gnathal segments was first detectable in stage 10 embryos whilst the thoracic and A9 labelling was not observed until slightly later at stage 11. Thus the 4 kb construct gives a subset of the complete **connectin** pattern with predominant expression in gnathal and thoracic segments.

**Homeotic regulation of the 4 kb region**

The clear differences in expression between thoracic and abdominal segments observed with the 4 kb construct suggested that the 110 bp immunopurified sequence had led us to a regulatory element under homeotic gene control. The reduced levels of expression of the 4 kb construct in abdominal segments A1-A8 suggest a down regulation by **Ubx** and the abdominal homeotic genes. In order to test this, we examined the expression in embryos homozygous for **Ubx** or **abd-A** null mutations and also in embryos homozygous for the **Ubx** **abd-A** double mutant combination. In **Ubx** mutants, expression is derepressed in the abdominal segments, A1 and A2 (Fig. 6D). This is consistent with the requirement for **Ubx** function in both A1 and A2 somatic musculature (Hooper, 1986). In **abd-A** mutants, there were no consistent changes in the pattern of expression of the 4 kb construct (data not shown) but, in the **Ubx-****abd-A** double mutant, the derepression is more dramatic than that seen in embryos mutant for **Ubx** alone: high levels of expression are now seen in A3 and, in late embryos, intermediate levels of expression are detected further posteriorly (Fig. 6E). This indicates that, in the wild type, both **Ubx** and **abd-A** repress this construct. The explanation for the lack of effect of the **abd-A** single mutant on the expression pattern is likely to be transregulation between **abd-A** and **Ubx**. The **abd-A** gene normally represses **Ubx** expression in the posterior abdominal segments and, in **abd-A** mutants, **Ubx** expression in the abdomen is derepressed (Struhl and White, 1985). Thus, in the **abd-A** single mutant, high levels of **Ubx** expression are
Fig. 4. Expression of *connectin* RNA in wild-type embryos. (A) Stage 11; lateral view. Labelling is seen in segmentally repeated subsets of the splanchnopleura, which will form the visceral mesoderm of the midgut, and in the visceral mesoderm of the foregut (fgm) and hindgut (out of focus). (B) Later stage 11; lateral view. Strong labelling is seen in the visceral mesoderm of the foregut (fgm) and hindgut (hgm) and the arrowheads indicate two of the segmentally repeated subsets of labelling in the presumptive visceral mesoderm of the midgut. The predominant labelling in the three thoracic segments T1, T2 and T3 is in the somatic mesoderm. There is also weaker labelling in cells of the somatic mesoderm in the abdominal segments. The arrow indicates labelling in one of the longitudinal tracts of cells in the ventral nervous system. (C) Stage 12; lateral view focussing on the somatic mesoderm. Labelling is predominantly in specific cell clusters in the thoracic somatic mesoderm. (D) Stage 12; ventral view. The two longitudinal tracts of cells in the ventral nervous system are stained with uniform intensity along the anteroposterior axis. In the somatic mesoderm, there is a clear difference in labelling intensity between the thoracic and abdominal segments. Also, T1 is weaker than T2/T3 and shows a different arrangement of labelled cells. The arrow indicates labelling associated with gnathal sense organ precursors. (E) Stage 14; dissected preparation. The labelling in the ventral nervous system (vns) shows the beginnings of segment-to-segment modulation with the thoracic neuromeres showing slightly higher intensity labelling than the abdominal. In the somatic mesoderm, the intensity differences between thoracic and abdominal segments are less pronounced than at earlier stages. There are clear differences in the arrangement of the labelled cell clusters between thoracic and abdominal segments. (F) Stage 14; lateral view focussing on the somatic mesoderm showing the positions of the labelled cell clusters. The labelling in the T1 somatic mesoderm is rather obscured by out-of-focus labelling in the foregut visceral mesoderm. (G) Stage 16; labelling is predominantly in the brain lobes and in the thoracic neuromeres of the ventral nervous system. The arrows indicate labelling associated with head and gnathal sense organs. DNA probes were used for the in situ hybridizations in A, E and G; RNA probes were used for the others.
In order to examine the regulatory role of sequences more immediately surrounding clone 35, we next tested a 1 kb PstI-Sall fragment in which the 110 bp clone 35 is centrally located (Fig. 5). Like the 4 kb fragment, this 1 kb fragment was placed upstream of the hsp70 minimal promoter in both orientations. Consistent staining, related to the connectin gene expression pattern, was obtained only when the 1 kb element was in the same orientation as the functioning 4 kb element. Expression is confined to a subset of the 4 kb pattern with most labelling in the gnathal region (Fig. 7A). Compared to the 4 kb pattern, the staining observed in T1-T3 and in A9 appears much reduced in intensity relative to the gnathal expression (and is completely absent in several transformant lines). Thus, the 1 kb element lacks some of the sequences required for the high levels of expression in T1-T3 and A9 that are characteristic of the 4 kb construct and presumably these sequences reside in the 3 kb remainder fragment.

We were surprised to observe an additional pattern of expression in the visceral mesoderm of the midgut with the 1 kb construct that is not observed with the 4 kb construct. As described above, endogenous connectin RNA is expressed in a segmentally repeated subset of the visceral mesoderm. However, whilst the expression of connectin RNA in the presumptive visceral mesoderm of the midgut is seen transiently from stage 11-13, β-galactosidase activity from the 1 kb construct does not appear until stage 15. The visceral mesoderm expression from the 1 kb construct is restricted to the developing second midgut constriction (Fig. 7B,C) and corresponds to a subset of parasegment 7. This is particularly interesting as Ubx is expressed in PS7 in the visceral mesoderm and Ubx function is required for the formation of the second midgut constriction (Bienz et al., 1988; Bienz and Tremml, 1988). The visceral mesoderm expression of the 1 kb construct is also dependent on Ubx as it is abolished in Ubx mutant embryos, whilst the gnathal pattern remains unaffected (Fig. 7D).

Deletion of clone 35 sequences from 4 kb and 1 kb regulatory regions
To test directly the requirement for the clone 35 sequence in the regulation conferred by the 4 kb or the 1 kb elements, constructs containing a precise deletion of this 110 bp region but similar in every other respect were made (called 4 kb-35 and 1 kb-35 constructs).

With the 4 kb-35 construct, the gnathal expression is little affected, but the expression in T1-T3 is weak and variable (Fig. 8A) indicating that the 110 bp region is required for the normal high levels of expression in T1-T3. As described above, comparison between the expression from the 4 kb and 1 kb constructs led to the conclusion that sequences outside the 1 kb region are also required for high levels of thoracic expression. Taken together, these observations imply that multiple sites of control are involved.

The 1 kb-35 construct gave an expression pattern identical to the 1 kb construct within the gnathal segments. However, whereas the 1 kb construct gave expression in the visceral mesoderm, staining in this tissue was completely absent in the 1 kb-35 construct (Fig. 8B). As the visceral mesoderm expression from the intact 1 kb element is also absent in a Ubx− genetic background, we conclude that clone 35 sequences are required in vivo to mediate Ubx-dependent activation in PS7 of the visceral mesoderm.

**Discussion**

In this paper, we have shown that the immunopurified clone 35, which was isolated as a potential in vivo Ubx protein binding site, appears to be an important component of a regulatory element under homeotic gene control. This regulatory element is part of the control sequences regulating the expression of connectin, a gene encoding a cell-surface member of the LRR family.
The presence of multiple tandem copies of the LRR motif in the connectin molecule does not immediately assign the molecule to a specific functional class. LRRs were first identified in the leucine-rich alpha 2-glycoprotein (Takahashi et al., 1985) and have since been found in a diverse set of intracellular, membrane and extracellular proteins including yeast adenylate cyclase (Kataoka et al., 1985), the von Willebrand factor receptor (Titani et al., 1987) and extracellular matrix proteoglycans (Fisher et al., 1989). There is evidence that the LRRs are involved in protein-protein interactions. LRR peptides appear to adopt an amphipathic β-sheet presenting polar and apolar potential interaction surfaces (Gay et al., 1991; Krantz et al., 1991).

Deletion analysis of the LRR region in adenylate cyclase implicates LRRs in complex formation with RAS proteins (Field et al., 1990). Also, several cell surface LRR molecules, including the Drosophila molecules chaoptin and Toll protein, share the common property of mediating cell-cell adhesion (Krantz and Zipursky, 1990; Keith and Gay, 1990). Chaoptin shows additional similarity to connectin as it is also linked to the membrane via a GPI-anchor (Krantz and Zipursky, 1990). Mutants in chaoptin disrupt the organization of photoreceptor membranes consistent with a role for chaoptin as a cell adhesion molecule required for normal photoreceptor morphogenesis. Expression of connectin in tissue culture cells reveals that it can also function as a homophilic cell-cell adhesion molecule (Nose et al., 1992;
1171

Homeotic target gene

L. Meadows, A. P. G. and R. A. H. W., unpublished results). This supports the connection between cell surface LRR-containing molecules and adhesion, and further implicates the LRR sequences themselves in protein-protein interactions.

Regulation of the connectin gene by multiple homeotic loci

Previously, we have demonstrated that connectin RNA expression is regulated in the central nervous system by Ubx and the abdominal homeotic genes (Gould et al., 1990). Here we have examined the control of a 4 kb connectin regulatory element that directs expression predominantly in the somatic mesoderm and not in the ventral nervous system. We have shown that this 4 kb region, located at least 45 kb 3’ of the transcription start site, is sufficient to drive a subset of the connectin expression pattern that is under the control of multiple homeotic genes, including Ubx. Genetic experiments have shown that Antp is the homeotic gene required to activate expression of the 4 kb construct in T2 and T3. However, in Antp− mutants,
expression remains in gnathal segments, in T1 and in A9. The expression in T1 is likely to be dependent on *Sex-combs reduced* (*Scr*), whilst *Deformed, labial* and *pro-boscipedia* may be required for the more anterior activation (Mahaffey and Kaufman, 1988).

The 4 kb construct is strongly expressed in the thoracic segments but shows very little expression in abdominal segments. We have shown that *Ubx* and *abd-A* are required for this repression in the abdomen. However, the possibility exists that this regulation could be mediated through the influence of *Ubx* and *abd-A* on the expression of *Antp*. Both *Ubx* and *abd-A* normally repress *Antp* expression in the abdominal segments (Hafen et al., 1984). Thus, in *Ubx* or *Ubx* *abd-A* mutants, *Antp* expression in the abdomen is derepressed and these higher levels of *Antp* protein might directly enhance expression from the 4 kb construct. However, studies of transformed cuticular phenotypes induced by ectopic *Antp* expression indicate that *Antp* is not able to influence cuticle phenotype in regions expressing *Ubx* or the abdominal homeotic genes (Gibson and Gehring, 1988). This may suggest that *Ubx* and *abd-A* exert a dominant direct repression of target genes, blocking their activation by *Antp*. This phenomenon has been called 'phenotypic suppression' (Gonzalez-Reyes and Morata, 1990) and it is likely to be important for our understanding of homeotic gene function. The analysis of the molecular basis of phenotypic suppression has been impeded by the lack of identified target genes. Now, with the identification of specific target gene regulatory elements, such as the *connectin* 4 kb element, we can hope to clarify this phenomenon.

Deletion of either the clone 35 sequence or of the 3 kb *Pstl-EcoRI* fragment leads to a similar reduction in the intensity of staining in T1-T3. Therefore, it is probable that there are at least two separable elements required for high levels of expression in this region. Their requirement in T2 and T3 indicates that elements within the 3 kb fragment and also within the 110 bp clone 35 both respond to *Antp* regulation. Another homeotic gene, most probably *Scr* (Mahaffey and Kaufman, 1988), may activate expression from these elements in T1.

With the 1 kb construct, we observed expression in PS7 of the visceral mesoderm, although this is not seen with the 4 kb construct. There may be two possible explanations for this; either repressing elements are present in the deleted 3 kb region or the visceral mesoderm expression element in the 1 kb construct may not function correctly when placed too far upstream of the hsp 70 minimal promoter. The *connectin* RNA is expressed in repeating subsets of the developing visceral mesoderm including PS7, whereas the 1 kb construct appears to give only the PS7-specific subset of this pattern. However, we first detect expression from this construct only after *connectin* RNA levels in this region have declined (compare Figs 4B and 7B). Hence, in this particular instance, the 1 kb construct does not faithfully reproduce the timing of onset of *connectin* expression. Nevertheless, the PS7 domain does allow a specific connection to be made between the 110 bp immunopurified fragment and *Ubx* control as expression is dependent on the presence of both the clone 35 sequence and *Ubx* function. Given that clone 35 was originally immunopurified using an antibody directed against *Ubx* proteins and given that it interacts directly with *Ubx* protein in vitro (Gould et al., 1990), it seems likely that this 110 bp region mediates direct regulation by *Ubx*.

For the 1 kb construct, deletion of the 110 bp fragment has the same effect on expression in the visceral mesoderm as the removal of *Ubx* gene function. This is what would be expected if the 110 bp sequence identifies a specific *Ubx*-dependent regulatory site. However, for the 4 kb construct, a *Ubx* mutation results in increased expression in the abdominal somatic mesoderm whereas deletion of the 110 bp results in reduced thoracic expression. Our interpretation of these different effects is that the deletion of the 110 bp fragment removes, not only sites required for *Ubx* and *abd-A* repression, but also the sites required for activation by *Antp* and *Scr*.

The DNA sequence of the 110 bp region contains six copies of TAAT, the core homeodomain binding sequence (Ekker et al., 1991; Kissinger et al., 1990). Using mutagenised versions of the 1 kb construct, it should now be possible to assay each of the TAAT cores for a direct and functional interaction with *Ubx*, *abd-A*, *Antp* and *Scr*. As the 110 bp sequence appears to be recognised by multiple homeotic loci, it will be interesting to determine whether the different homeotic gene products act on separate binding sites or whether they compete for the same sites.

In summary, the immunopurification procedure has successfully led us to a regulatory element under homeotic gene control. We are currently investigating whether other binding sites from the immunopurified DNA library are also associated with elements regulated by homeotic genes. In principle, this method should be generally applicable to searches for in vivo targets of characterised transcription factors or DNA-binding proteins.

**Properties of genes regulated by the homeotic loci**

Garcia-Bellido proposed that the homeotic genes would exert their effects on morphogenesis by controlling the activities of a group of cytodifferentiation genes. These cytodifferentiation genes would specify a number of cellular characteristics including, mitotic rate, preferential mitotic orientations, cell affinities and final cellular differentiation (Garcia-Bellido, 1975). The isolation of *connectin* as a homeotic target gene fits well with this scheme. *Connectin* is a cell surface molecule capable of mediating cell-cell adhesion and the analysis of its protein distribution in detail reveals that it is expressed in a subset of motoneurons and also in their specific target muscle cells both prior to and after the formation of neuromuscular junctions (Nose et al., 1992). Thus it is tempting to speculate that connectin is a cell-adhesion or recognition molecule involved in the development of specific neuromuscular connections. Interestingly, recent studies in *C. elegans* have implicated an *Antp* homologue *mab-5*, in the control of cell surface molecules that affect region-specific migratory cell behaviours in the nervous system (Salser and Kenyon, 1992). It has also been suggested that the neural cell adhesion molecule (N-CAM) is likely to be regulated by mouse Hox-gene products (Jones et al., 1992).

Our observation that *connectin* is likely to be regulated directly by homeotic genes is interesting with regard to the overall command chain of developmental gene controls.
The complex pattern of homeotic gene expression is largely the result of a hierarchy of regulation by transcription factors extending from bicoid, through gap genes, pair-rule genes and segment polarity genes (reviewed Ingham, 1988). It had always been possible that several more layers of transcriptional control might necessarily lie between the homeotic selector genes and their cytodifferentiation targets. Here, with connectin, this appears not to be the case and suggests that at least some genes involved in the common processes of morphogenesis are directly under the transcriptional control of homeotic genes.

We thank Nick Brown and Sarah Bray for guidance with cDNA library screening, Shashidhara L. S. for the RNA-probe connectin in situ hybridizations and Peter Lawrence for help with the transformations. We are grateful to Sarah Bray, Lisa Meadows and David Strutt for their comments on the manuscript. This work was supported by the MRC and the Wellcome Trust.

References


Fisher, L. W., Termine, J. D. and Young, M. F. (1989). Deduced protein sequence of bone small proteoglycan I (Biglycan) shows homology with proteoglycan II (Decorin) and several nonconnective tissue proteins in a variety of species. J. Biol. Chem 264, 4571-4576.


Homeotic target gene 1173


(Accepted 21 September 1992)