Development 121, 2799-2812 (1995)
Printed in Great Britain © The Company of Biologists Limited 1995

Homeotic response elements are tightly linked to tissue-specific elements in a transcriptional enhancer of the *teashirt* gene

Alison McCormick[1], Nathalie Coré[2], Stephen Kerridge[2] and Matthew P. Scott[1,*]

[1]Departments of Developmental Biology and Genetics, Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford CA 94305-5427, USA
[2]Laboratoire de Génétique et Physiologie du Développment, CNRS, Case 907, Parc Scientifique de Luminy, Université d'Aix-Marseille II, 1388 Marseille, Cedex 9, France

*Author for correspondence: email scott@cmgm.stanford.edu

SUMMARY

Along the anterior-posterior axis of animal embryos, the choice of cell fates, and the organization of morphogenesis, is regulated by transcription factors encoded by clustered homeotic or 'Hox' genes. Hox genes function in both epidermis and internal tissues by regulating the transcription of target genes in a position- and tissue-specific manner. Hox proteins can have distinct targets in different tissues; the mechanisms underlying tissue and homeotic protein specificity are unknown.Light may be shed by studying the organization of target gene enhancers. In flies, one of the target genes is *teashirt* (*tsh*), which encodes a zinc finger protein. *tsh* itself is a homeotic gene that controls trunk versus head development. We identified a *tsh* gene enhancer that is differentially activated by Hox proteins in epidermis and mesoderm. Sites where Antennapedia (Antp) and Ultrabithorax (Ubx) proteins bind in vitro were mapped within evolutionarily conserved sequences. Although Antp and Ubx bind to identical sites in vitro, Antp activates the *tsh* enhancer only in epidermis while Ubx activates the *tsh* enhancer in both epidermis and in somatic mesoderm. We show that the DNA elements driving tissue-specific transcriptional activation by Antp and Ubx are separable. Next to the homeotic protein-binding sites are extensive conserved sequences likely to control tissue activation by different homeodomain proteins. We propose that local interactions between homeotic proteins and other factors effect activation of targets in proper cell types.

Key words: transcriptional enhancer, homeotic, *teashirt*, *Antennapedia*, *Drosophila*

INTRODUCTION

Cascades of gene interactions are critical for the establishment of cell identity in multicellular eukaryotic organisms. In *Drosophila melanogaster*, early maternal and zygotic regulatory events ensure proper spatial patterns of transcription of the homeotic genes, which specify segment identity (Duncan, 1987; Kaufman et al., 1990; Morata, 1993). Each homeotic gene is expressed in a specific domain along the anterior-posterior axis, predominantly where cuticle defects are seen in embryos that lack homeotic gene function. Clusters of homeotic genes, collectively called Hox genes, control many aspects of development in animals as divergent as mice, flies and nematodes (McGinnis and Krumlauf, 1992; Krumlauf, 1994; Manak and Scott, 1994).

The products of Hox genes are transcription factors, which control the formation of segment-specific structures, tissues and morphologies by regulating the transcription of downstream 'target' genes. The identities and functions of Hox target genes must be learned to understand how homeotic proteins control animal development. In particular, the basis for differential regulation of target genes must be learned. Target genes can be activated by some homeotic proteins and repressed by others (reviewed in Andrew and Scott, 1992; Botas, 1993; Morata, 1993). Some targets are controlled by a single homeotic protein; some by more than one. Target genes may be activated by a particular homeotic protein in one tissue type but not another. The mechanistic basis of these different types of specificity: positive versus negative regulation, multiple homeotic proteins versus a single specific one and cell-type-specific regulation by homeotic proteins, is largely unknown.

Within homeotic proteins the most important determinant of homeotic specificity discovered to date is the DNA-binding domain, the homeodomain. This 60 amino acid structure is similar in all homeotic proteins but, outside the homeodomain, Hox proteins have little in common (Bürglin, 1994). Experiments with chimeric proteins reveal that sequences within the homeodomain (Gibson et al., 1990; Lin and McGinnis, 1992; Chan and Mann, 1993; Zeng et al., 1993) or the C-terminal tail (Mann and Hogness, 1990) are essential for specifying different segment identities. The homeotic proteins encoded by Antennapedia (Antp) and Ultrabithorax (Ubx) differ by only seven amino acids in the homeodomain, but control the spec-

ification of thoracic (Denell et al., 1981; Struhl, 1982; Martinez-Arias, 1986) versus abdominal segment identity (Lewis, 1978). How homeodomain sequence differences lead to different target gene selection in vivo is poorly understood.

Only a few homeotic response elements (HOMREs) have been characterized at the molecular level (Regulski et al., 1991; Appel and Sakonju, 1993; Zeng et al., 1994; Capovilla et al., 1994; Manak et al., 1994; Sun et al., 1995), although more candidates have been identified by genetic analysis. Most HOMREs contain a TAAT core, which in vitro is required for optimal binding by homeodomain proteins. Flanking sequences 3′ to the TAAT can influence the affinity that different proteins have for *cis* elements (Desplan et al., 1988; Ekker et al., 1991, 1994). The occurrence of multiple binding sites in a HOMRE can also influence DNA-binding specificity by facilitating cooperative homeotic protein interactions (Appel and Sakonju, 1993; Beachy et al., 1993).

Protein-protein interactions constitute a fundamental mechanism by which homeodomain protein-DNA binding specificity or transcriptional activation may be controlled (Zeng et al., 1994). For example Ubx acts in parallel with another homeodomain protein, extradenticle (exd), at specific sequences in the *decapentaplegic* (*dpp*) midgut enhancer (Rauskolb et al., 1993; Chan and Mann, 1994, Manak et al., 1994; Rauskolb and Wiechaus, 1994; Sun et al., 1995). The paucity of knowledge of HOMRE sequences and cofactors has been an obstacle to understanding how homeotic proteins control target gene selection in vivo.

A major goal in seeking target genes regulated by homeotic genes is to learn how homeotic transcription factors govern morphogenesis, a problem of great significance in understanding the development of all animals. Targets are most significant if they perform a morphogenetic function clearly related to the homeotic gene's function. The gene *teashirt* (*tsh*) has a complex and fascinating relationship to the clustered homeotic genes. Working in combination with the homeotic genes *Antp*, *Ubx*, *abdominalA* (*AbdA*) and *Abdominal B* (*AbdB*), *tsh* promotes trunk development and prevents head development (Fasano et al., 1991). *tsh* is activated by some homeotic genes (Röder et al., 1992; Mathies et al., 1994), as we describe further here. To what end do homeotic genes activate *tsh*? The higher Tsh levels may activate or repress the transcription of targets shared with the homeotic genes, thus modulating the effects of homeotic proteins. *tsh* alters the activities of homeotic genes, for example by determining whether *Scr* will direct a head or thoracic fate (Fasano et al., 1991; de Zulueta et al., 1994). *tsh* prevents *Scr* from inducing salivary gland development (Röder et al., 1992; Andrew et al., 1994).

We have identified and characterized a HOMRE within the *tsh* gene. We find that different homeotic proteins control *tsh* enhancer activity in different tissues. The homeotic binding sites are evolutionarily conserved and sequences immediately adjacent to homeotic binding sites are also conserved and are likely to be binding sites for other proteins that regulate homeotic specificity.

MATERIALS AND METHODS

Genetic strains

Antp^{w10}*/TM6B*, *Ubx*$^{6.28}$ e*/TM6B*, *Ki Antp*^{w10} *pp Ubx*$^{6.28}$ *e/TM6B* and *Scr*^{w17} *Antp*^{w10} *pp Ubx*$^{6.28}$ *Df(3R)109/TM6B* are null alleles used to test the effects of mutations in homeotic genes. Mutations are described in Lindsley and Zimm (1992). *Antp* and *Ubx* heat-shock strains used were Antp$^{gbb2\Delta a\text{-}1}$/Cyo and UbxIa 22/TM6B, respectively. *Df(2L)R6* was created by excision-hopping (N. Coré, A. McCormick, M. P. Scott and S. Kerridge, unpubllished data) and *Df(2L)ltx217* was created by X-ray mutagenesis by B. Wakimoto (unpublished). All transformant lines were created in *yw*^{67}.

Transformation constructs and germline transformation

tsh genomic DNA fragments, cloned into Bluescript (Stratagene) in either *Xho*I or *Eco*RI sites, were cloned directionally into the C4PlZ vector (S. Crews, unpublished) at *Not*/*Asp* sites 5′ to the *lacZ*-coding region. DNA was prepared by cesium chloride banding after alkaline lysis (Sambrook et al., 1989). 200 ng *tsh*C4PlZ DNA was mixed with 100 ng Δ2-3 integration defective vector (D. Rio, unpublished), then spun at 14000 for 20 minutes at 4°C to eliminate any particulate matter. This DNA was injected into 1 hour old *yw*^{67} embryos as has been previously described (Rubin and Spradling, 1982). Surviving adults were crossed to flies from the *yw*^{67} stock and transgenic offspring were identified by light yellow to red eye color generated from the w^+ minigene carried on the C4PlZ transformation vector. Multiple lines of each construct were established and chromosomal location was determined by standard genetic analysis using balancer chromosomes. Staining patterns were examined for a minimum of three lines for each construct reported here.

Embryo immunohistochemistry

Wild-type embryos were stained for the presence of Antp or Tsh protein using antibodies prepared as follows. Bacterially expressed Antp-β-gal fusion (Carroll et al., 1986) was gel-purified and injected into rats along with Ribi adjuvant (Ribi International) every two weeks. Bleeds were tested for Antp-reactive antibodies both on protein blots and by staining embryos. High titer serum was used at 1:1000 for embryo immunohistochemistry. Anti-Tsh antibodies were prepared as described in Mathies et al. (1994) and used at 1:500 for embryo immunohistochemistry. Anti-β-gal antibodies were prepared by co-injecting 200 μg β-gal protein (Sigma) and Ribi into rats on a bimonthly schedule as described (Mathies et al., 1994). Anti-β-gal antibodies were used at 1:500 in all experiments. Embryos were collected overnight on grape juice agar caps and dechorionated and fixed as previously described (Mann and Hogness, 1990). Incubations were made in PBT (1× PBS with 0.1% Triton X-100) as were all washes. β-gal and secondary antibodies were preadsorbed to wild-type 0-24 hour embryos at 1:50 before use. Anti-rat secondary antibodies conjugated with HRP (Horseradish peroxidase) were used for Antp or Tsh stains, and biotin-labeled antibodies were used for β-gal stains. An avidin-HRP ABC Elite kit (Vector labs) was used to detect biotin-conjugated rat secondary antibodies. All HRP reactions were carried out in the presence of 0.5 μg/ml DAB (Sigma) and 0.004% H_2O_2 and after 1-5 minutes stopped by several washes with PBT. Embryos were dehydrated with several ethanol washes and then preserved in methyl salicylate (Sigma).

Embryos destined for heat-shock treatments were collected for 4 hours, aged 2 hours, then heat-treated twice at 37°C for 30 minutes with a 90 minute recovery in between heat pulses and a 2-hour recovery after heat treatment.

All embryos were photographed using Kodak 64T color slide film on a Zeiss Axiophote photomicroscope. Color slides were scanned using a Nikon scanner to create a black and white digital image. Adobe Photoshop 3.0 was used to adjust contrast, image size and assemble composites.

Deletion series constructs

The *tsh* full-length enhancer, a 1.3 kb genomic DNA blunt end-cloned into the *Eco*RV site in Bluescript (Stratagene) was used as a starting point for nested deletions as follows. 5′ deletions were created by cutting 10 μg plasmid DNA with *Asp*718 and *Apa*I (both sites are 5′ to the end of the *tsh* insert) and cleaved with 1 i.u. of T4 polymerase

in the absence of nucleotides for 1-15 minutes as described (Sambrook et al., 1989). Reactions were stopped on ice at one minute intervals after 2 minutes, and blunt ends were created by adding *Exo*III, T3 polymerase and 2.5 mM dNTPs for 30 minutes at 37°C. DNA from each time point was analyzed by agarose gel electrophoresis for the efficiency of 5′ cleavage, ligated and screened for appropriate size inserts, and sequenced by standard methods (Sanger et al., 1977). 3′ deletions were created essentially the same way except DNA was first cut with *Eco*RI and *Pst*I (both sites are 3′ to the end of the *tsh* insert). Construct Δ1-2 was prepared by cutting Δ1-5 (Fig. 6) with *Xmn*I, which cuts between boxes 2 and 3 and with *Asp*718, which cuts 5′ to conserved sequences and liberates a fragment of 220 bp containing boxes 1 and 2. This insert was cloned directionally into *Asp*718/*Eco*RV cut Bluescript. Construct Δ3-5 was generated by liberating a 280 bp *Xmn*I and *Bam*HI fragment of Δ1-5, which contains conserved boxes 3, 4 and 5, and cloned into *Bam*HI/*Eco*RV Bluescript. Δ1-4 was created by cutting Δ1-5 with *Eae*I, which lies between conserved boxes 4 and 5, and *Eag*I which lies 3′ to box 5 within the Bluescript polylinker, and religated to remove box 5. Constructs created in Bluescript were then cut with *Not*I/*Asp*718 and cloned directionally into the *Not*I/*Asp*718 site of C4PlZ.

In vitro transcription and translation of Antp protein

A *Bam*HI-*Bgl*II fragment of *Antp* cDNA in Bluescript KS+ vector (Winslow et al., 1989) was linearized with *Not*I and used to generate in vitro synthesized capped RNA as described (Urness and Thummel, 1990). The full-length transcript of 1.7 kb was checked for yield and integrity by formaldehyde gel electrophoresis (Sambrook et al., 1989). 1 μg of RNA (approximately 1 μl) was used as template in a 20 μl reaction of rabbit reticulocyte lysate plus [^{35}S]methionine (Amersham) exactly as described in the Promega protocol provided with the rabbit reticulocyte lysate. The products were separated by SDS-PAGE and detected by overnight autoradiography at −80°C with an intensifying screen. A major band of 66×10^3 M_r was evident, similar in size to Antp protein produced in tissue culture cells and therefore probably full length.

DNA-binding analysis with [^{35}S]-labeled Antp protein

Full-length [^{35}S]-labeled Antp protein was incubated with 100 ng of *Hae*III-digested plasmid DNA containing either *Eco*R1 or *Xho*I restriction fragment inserts subcloned from the *tsh* phage EMBL4-10b or EMBL3-6B, respectively (Fig. 2A). Reactions contained 1 μl of a freshly prepared reticulocyte lysate reaction, 1 μl phenol-extracted and ethanol-precipitated DNA, 1.5 μl 10× binding buffer (0.2 M Tris-HCl [pH7.4], 0.5 M KCl, 30 mM $MgCl_2$, 10 mM EDTA, 10 mM DTT and 1 μg of poly(dIdC). Incubations were carried out for 30 minutes at room temperature. After the addition of 5 μl loading buffer (1× binding buffer, 40% glycerol, 0.3% bromophenol blue and 0.3% xylene cyanol), reactions were loaded immediately onto a freshly poured 1.5 mm thick TPE 4% polyacrylamide gel. Gels were prerun for 2-3 hours and electrophoresed overnight at 25 mA at room temperature. 10× TPE was prepared fresh for gels and running buffer and consisted of 27 g Tris base, 5 ml of 5 M EDTA, 3.9 ml o-phosphoric acid, 6.5 ml HCl in a 250 ml final volume. After electrophoresis, gels were fixed in 10% methanol and 10% acetic acid, and the signals were enhanced by incubation in 50 ml Fluoro-Hance (Research Products International). Gels were dried and autoradiographed at −80°C with an intensifying screen for 2-3 days.

Bacterial expression of homeotic proteins for in vitro binding analyses

Partially purified full-length Antp protein was made from a PET3a vector as previously described (Hoey et al., 1988) and Antp homeodomain-GST (D. P. LeBrun, unpublished) was prepared and purified by Glutathione agarose chromatography as has been described (Hakes and Dixon, 1992). Affinity-purified Ubx protein was kindly provided by Brad Johnson (Johnson and Krasnow, 1992) and affinity-purified abdA protein was kindly provided by John Manak (Manak et al., 1994).

Dnase I protection analysis

End-labeled probes were generated by cutting Δ1-5 DNA with *Asp*718 (lanes 1-4) or *Bam*HI (lanes 5-8) filling in 5′ overhangs with [^{32}P]dATP as described (Sambrook et al., 1989). After labeling, DNA fragments were generated by cleaving DNA labeled at *Bam*HI with *Asp*718, and by cleaving DNA labeled at *Asp*718 with *Bam*HI. Fragments were electrophoresed in 1% agarose, isolated by glass purification (Qiagen) and specific activity was assessed at $4\text{-}6\times10^8$ cts/minute/μg. End-labeled probes were also generated from Δ2-5 constructs in a similar fashion. 1 ng of end labeled DNA (about 10,000 cts/minute) was mixed on ice with 20 ng Antp GST protein (except lane 15, which contained 60 ng Antp protein), Ubx or abdA protein with 1 μg poly(dIdC) (Promega) in Z buffer (Hoey et al., 1988) supplemented to 1 mg/ml Dnase-free fraction V BSA (New England Biolabs), and allowed to incubate for 10 minutes on ice. Dnase reactions were initiated by adding 3-5 ng Dnase I (Worthington) and proceeded for 5 minutes on ice. Reactions were stopped with high salt buffer containing tRNA as described (Jones et al., 1985). After phenol extraction, DNA was ethanol precipitated and resuspended in 4 μl formamide loading buffer (1× TBE, 90% deionized formamide, 0.1% bromophenol blue and 0.1% xylene cyanol). Maxam/Gilbert sequence reactions (Sambrook et al., 1989), prepared from the same DNA using a kit and instructions from Sigma, were run side by side with footprint reactions on an 6% acrylamide 8% urea 0.4 to 0.8mm wedge gel at 50 watts for 1-4 hours. Gels were dried and autoradiographed with intensifying screens at −80°C overnight.

RESULTS

The *teashirt* gene is regulated by Antp protein in vivo

In epidermis, the highest level of Tsh protein is in the cells that contain Antp protein. Wild-type stage 11 embryos were stained to detect Antp protein and Tsh protein. Antp protein is present in posterior PS3 (PS, parasegment; Martinez-Arias and Lawrence, 1985), all of PS4 and anterior PS5 in developing thoracic epidermis and mesoderm, and at high levels in the ventral nerve cord from PS3-13 in stage 11 embryos (Fig. 1A, bracket; Carroll et al., 1986; Wirz et al., 1986; stages according to Campos-Ortega and Hartenstein, 1985). Tsh expression in the epidermis, mesoderm and ventral nerve cord extends through all the cells of PS3 to PS13. The highest level of Tsh protein is in the epidermis of PS3, PS4 and anterior PS5 (Fig. 1B, bracket), while lower levels are detected in the mesoderm and ventral nerve cord in the same region, and in the abdomen (Fig. 1B; Fasano et al., 1991; Röder et al., 1992).

High level expression of the *tsh* gene in the thoracic epidermis requires *Antp* function (Röder et al., 1992; Fig. 1C). Tsh epidermal expression in the thorax is lowered in the absence of *Antp* function (Fig. 1C, bracket). Detectable levels of Tsh protein appear in PS3-13 prior to *Antp* expression (not shown), so initial *tsh* expression is activated by other regulators. Only after detectable Antp protein has accumulated (stage 9) is the level of Tsh protein augmented in thoracic cells.

tsh has binding sites for Antp protein in vitro

The activation of *tsh* by *Antp* may be a direct interaction, with Antp protein binding the *tsh* gene, or may be indirect, mediated by another regulatory molecule whose expression or function

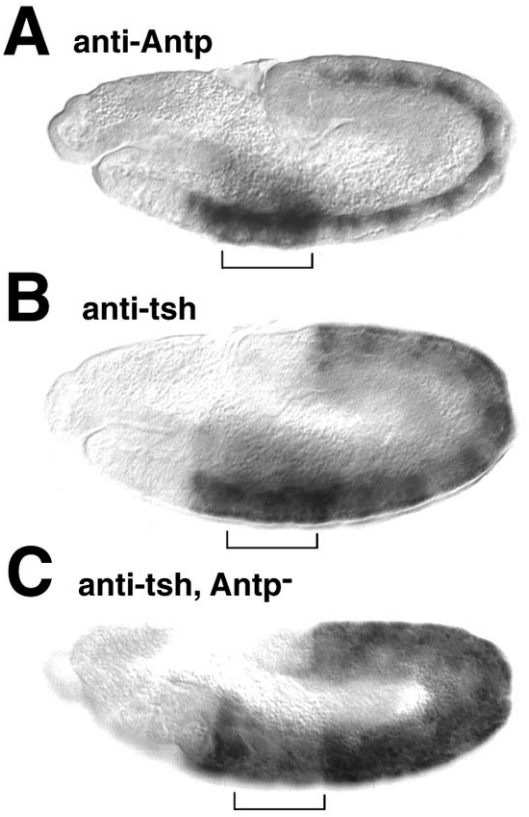

Fig. 1. Expression of *teashirt* requires Antp protein in vivo. **(**A-C) Wild-type embryos were incubated with polyclonal antibodies directed against Antp (A) or Tsh (B) and patterns of protein expression were visualized by immunolocalization with peroxidase-conjugated secondary antibodies. All embryos are oriented with anterior to the left, and brackets indicate the region of the embryo where Antp protein is expressed in the epidermis. (C) Tsh protein distribution in an *Antp*$^{-}$ background. The level in the thoracic epidermis is diminished compared to wild-type embryos. Homeotic mutant embryos were identified by counter staining with antibodies specific for Antp protein.

is activated by *Antp*. We scanned genomic DNA spanning the *tsh* gene (Fasano, 1990) for sites to which Antp binds in vitro (Fig. 2A). Various restriction fragments harboring the *tsh* gene were incubated with [^{35}S]-labeled Antp protein synthesized in vitro (Materials and Methods). Binding was detected by running the mixture of DNA and protein on a non-denaturing acrylamide gel and looking for restriction fragments capable of carrying the labeled protein into the gel (Fig. 2B; Hope and Struhl, 1985, Urness and Thummel, 1990). Seven fragments of *tsh* DNA (indicated in Fig. 2B) contain one or more binding sites for Antp protein. By testing smaller portions of each of these seven fragments, each Antp-binding site was located within 1 kb of DNA (data not shown).

A *tsh* enhancer is active in epidermis and mesoderm

We tested whether any of the *tsh* DNA fragments bound by Antp in vitro act as an *Antp*-responsive enhancer in vivo. Each fragment was placed upstream of a promoter-*lacZ* fusion in a P element vector and introduced into the fly genome by germline transformation. Of the fragments tested, one has enhancer activity. This 1.3 kb DNA, hereafter referred to as the *tsh* enhancer (Fig. 2A), lies 16 kb downstream of the *tsh* coding exons (Fig. 2A; Fasano et al., 1991). In germband-extended embryos, the *tsh* enhancer drives expression of *lacZ* in ventral epidermis in posterior PS3 through PS13 in stripes spanning all but the anterior-most cells of each parasegment (Fig. 3A, arrow; Figs 3B, 4D). This pattern suggests activation by *Antp* and *Ubx* and two other homeotic proteins active in posterior abdomen, *abdominal A* (*abdA*) and *Abdominal B* (*AbdB*). Graded β-gal is seen in somatic mesoderm from PS6-12, at a high level in PS6 and lower level more posteriorly (Fig. 3A, arrow; Fig. 4D) like *Ubx* (Bienz et al., 1988; Carroll et al., 1988). Strong mesoderm expression is also evident in PS13 where *AbdB* is expressed (Kuziora and McGinnis, 1988; Celniker et al., 1989). In germband-shortened embryos, *tsh-lacZ* expression is maintained in ventral epidermis from posterior T1 to A8 (Fig. 3B,C), while mesoderm expression expands dorsally in A1-A8, and resolves into a complex

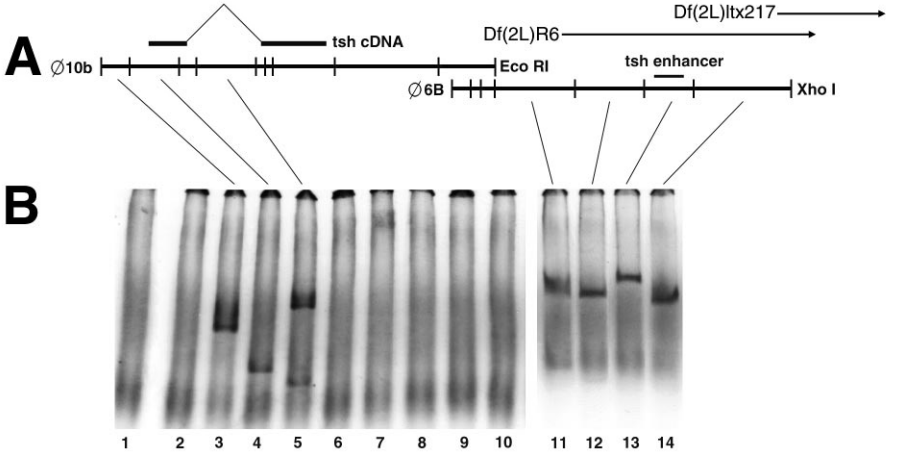

Fig. 2. Identifying Antp-binding sites within the *tsh* locus. (A) Two phage clones contain DNA spanning the *tsh* gene (Fasano et al., 1991), are indicated by either *Eco*RI (phage 10b) or *Xho*I (phage 6B) restriction fragments. The location of the *tsh* enhancer reported in this paper is indicated. Also indicated are the locations of two 2(L) deficiencies with breakpoints 5′ Df (2L)R6 or 3′ Df (2L) ltx217 to the *tsh*-enhancer. (B) Antp protein-binding sites within the *tsh* gene. Lane 1, incubation of unprogrammed rabbit reticulocyte lysate with *Hae*III-digested Bluescript (BS) plasmid DNA. Labeled Antp protein, produced by in vitro transcription of *Antp* cDNA and in vitro translation in the presence of [^{35}S]methionine (Materials and Methods), was incubated with *Hae*III-digested BS (lane 2) or with *Hae*III-digested plasmids containing either *Eco*RI (lanes 3-10) or *Xho*I inserts from the *tsh* locus as indicated (lanes 11-14). Unbound Antp protein remains in the wells, while protein-DNA complexes migrate into the native phosphate-buffered gel upon electrophoresis. Gels were enhanced (Materials and Methods) and autoradiographed at –80°C for one week.

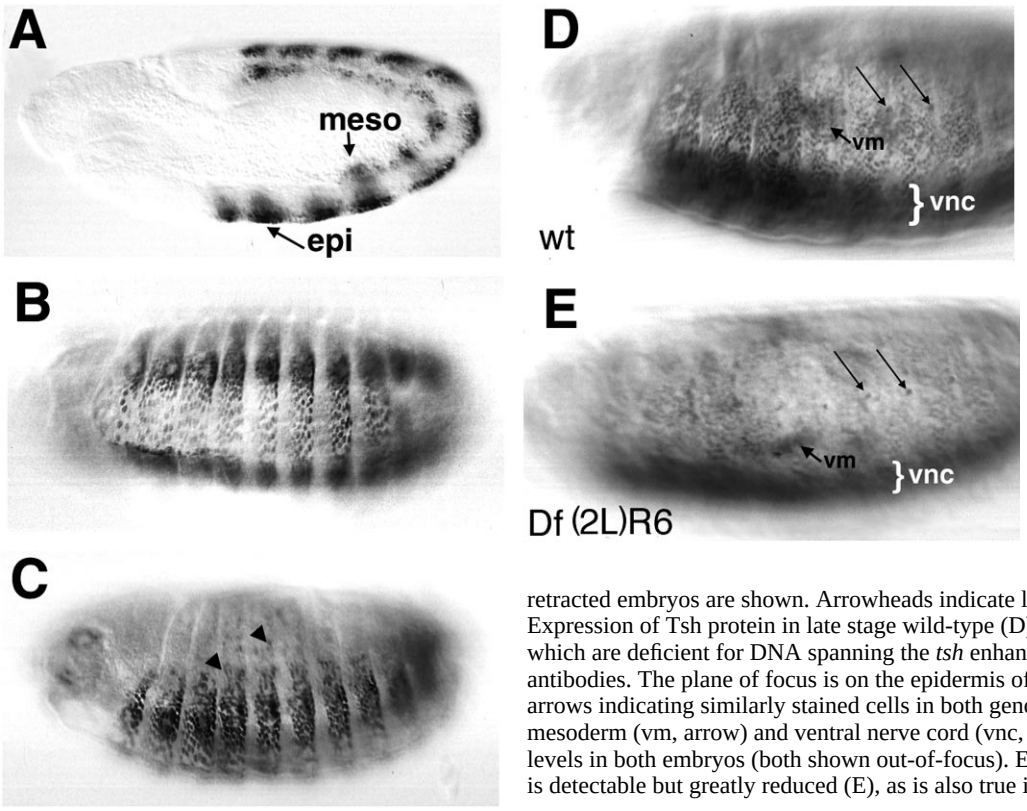

Fig. 3. The *tsh* enhancer drives β-gal expression in epidermis and mesoderm, and lies within a region critical for proper expression of the *tsh* gene. The 1.3kb *tsh* enhancer drives expression of a heterologous gene in a manner that reflects its activity within the context of the endogenous gene. Patterns of *tsh-lacZ* expression are visualized by the accumulation of β-gal protein. Wild-type transgenic embryos were incubated with polyclonal antibodies to β-gal and protein levels were visualized with Avidin-HRP/Biotin-conjugated secondary antibodies. Shown in A are germ-band-extended embryos expressing *tsh-lacZ*. An arrow indicates epidermis staining (epi) and mesoderm staining (meso). In B (ventral), C (lateral) germ-band-retracted embryos are shown. Arrowheads indicate lateral muscle staining in C. Expression of Tsh protein in late stage wild-type (D) or homozygous Df*(2L)R6* embryos which are deficient for DNA spanning the *tsh* enhancer(E) as visualized with anti-Tsh antibodies. The plane of focus is on the epidermis of similarly staged embryos, with arrows indicating similarly stained cells in both genotypes. Tsh expression in visceral mesoderm (vm, arrow) and ventral nerve cord (vnc, bracket) are expressed at similar levels in both embryos (both shown out-of-focus). Expression of Tsh epidermal staining is detectable but greatly reduced (E), as is also true in mesoderm (not shown).

pattern of lateral abdominal wall muscles (Fig. 3C, arrowheads).

Function of the *tsh* enhancer region in the normal *tsh* gene

An ideal test of the importance of the *tsh* enhancer in normal *tsh* function would require a precise deletion of the enhancer in the context of an otherwise normal *tsh* gene, but this is presently impossible. Instead, Tsh protein distribution was examined in embryos homozygous for a deficiency, *Df(2L)R6*, which removes the *tsh* enhancer, has a 5′ breakpoint just upstream of the *tsh* enhancer (Fig. 2A) and an 3′ end 38 kb downstream (N. Coré, A. McCormick, M. P. Scott and S. Kerridge, unpublished data). The effect of *Df(2LR6)* was compared to that of a largely overlapping deficiency, *Df(2L)ltx217* (Fig. 2A), which does not remove the *tsh* enhancer. The difference between the two deficiencies is a 10 kb region, including the *tsh* enhancer, which is required for proper expression of Tsh protein. *Df(2L)ltx217* homozygous embryos have normal Tsh expression, except that protein is not detected in PS3 (data not shown). In stage 15 embryos, Tsh protein levels in midgut and the nervous system are comparable to those of wild-type embryos (Fig. 3D), but Tsh protein is greatly reduced or absent in epidermis and in mesoderm in *Df(2L)R6* embryos (Fig. 3E). Thus the *tsh* enhancer lies within a region of the chromosome essential for epidermal and mesodermal expression, consistent with its ability to drive *lacZ* in abdominal somatic mesoderm, and thoracic and abdominal epidermis.

The *tsh* enhancer is regulated in vivo by homeotic proteins

Thoracic *tsh-lacZ* expression is dependent on *Antp* function in vivo, like the endogenous *tsh* gene. In *Antp* null embryos, epidermal expression of *tsh-lacZ* is lost in posterior PS3, all of PS4 and the anterior of PS5, exactly the places where *Antp* is expressed (Fig. 4A, bracket). Residual expression of the *tsh-lacZ* transgene is detectable in a small number of thoracic epidermal cells in the absence of *Antp* function (Fig. 4C). The high level of Antp in thoracic somatic mesoderm fails to activate the *tsh* enhancer in wild-type embryos, nor is the enhancer derepressed there in *Antp*$^{-}$ embryos.

The *tsh-lacZ* transgene is activated by the products of *Ubx*, *abdA* and *AbdB* homeotic genes in abdominal epidermis and mesoderm. In *Ubx* mutants, *Antp* becomes derepressed in the epidermis of the anterior abdomen (PS6), and increased *tsh-lacZ* expression is seen there (Fig. 4B, arrow). In *Ubx*$^{-}$ mutant embryos, *tsh-lacZ* is expressed in fewer cells in the posterior epidermis of every abdominal parasegment (Fig. 4B,C), where Ubx is normally expressed (White and Wilcox, 1985). *Ubx* is the only homeotic gene expressed in PS6-12 abdominal somatic mesoderm (Bienz et al., 1988; Carroll et al., 1988). The effect in abdominal mesoderm is even more dramatic; all *tsh-lacZ* expression in PS6-12 mesoderm is absent (Fig. 4B,C, arrowhead).

In mutant embryos lacking *Antp* and *Ubx* function, *tsh-lacZ* is expressed in the anterior epidermis of PS7-12 and in epidermis and mesoderm in PS13 (Fig. 4C). The heightened expression in PS6 seen in *Ubx* single mutants is gone, as expected, due to the lack of *Antp* (Fig. 4C, arrow). The residual abdominal expression, unaffected in *Antp*$^{-}$*Ubx*$^{-}$ embryos, is due to activation by *abdA* in PS7-12 abdominal epidermis and by *AbdB* in posterior abdominal epidermis and PS13 somatic mesoderm (data not shown). Therefore, four different

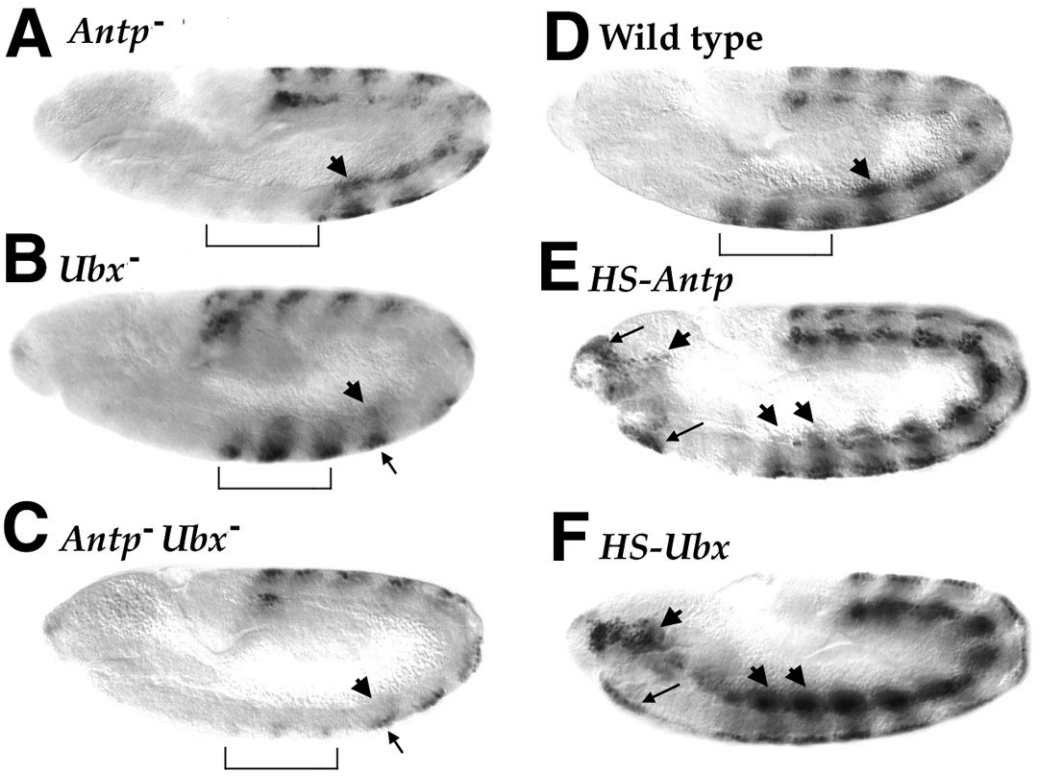

Fig. 4. The expression of the tsh enhancer in homeotic mutant embryos. (A) *Antp*$^{-}$ (B) *Ubx*$^{-}$ and (C) *Antp*$^{-}$*Ubx*$^{-}$ *tsh-lacZ* embryos stained for β-gal protein. Note the loss of expression in thoracic epidermis (bracket) in *Antp* mutant embryos (A and C, compared to D), and the loss of expression in abdominal mesoderm in *Ubx* mutant embryos (arrowhead, B and C). In a *Ubx* mutant embryo (C), *Antp* is derepressed in PS6, which leads to an increased level of *tsh-lacZ* epidermal expression, which is not observed in the *Antp*$^{-}$ *Ubx*$^{-}$ double mutant embryo (compare arrows, B and C). (D) A slightly later stage wild-type embryo expressing the full-length *tsh* enhancer in embryos that do not contain a heat-inducible construct, but have been heat treated. Heat-induced overexpression of Antp (E) and Ubx (F) causes expression of the *tsh* enhancer in anterior regions of the embryo in epidermis (arrows) and mesoderm (arrowheads).

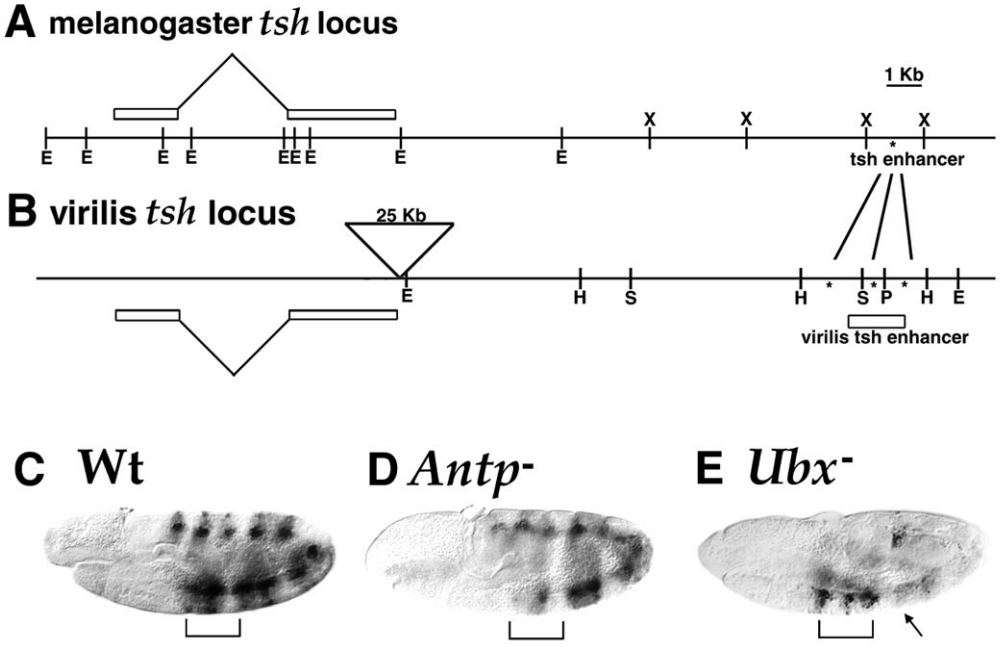

Fig. 5. Identification of the *D. virilis tsh* enhancer, and its enhancer activity in *D. melanogaster*. (A) A comparison of the *D. melanogaster* with (B) the *D. virilis* genomic regions spanning the *tsh* locus. An insertion of approximately 25 kb separates the *D. virilis* enhancer region from the coding exons compared to *D. melanogaster*, as is indicated. Restriction enzyme sites are indicated; *Eco*RI (E), *Hin*dIII (H), *Pst*I (P), *Sal*I (S), *Xho*I (X). A 2 kb region of the *virilis tsh* enhancer, as indicated schematically in B, was cloned into C4PLZ plasmid and introduced into *D. melanogaster* embryos by germline transformation. (C) The expression of β-gal driven by the *D. virilis tsh-lacZ* enhancer in wild-type *D. melanogaster* embryos. Arrows indicate expression in epidermis, arrowheads indicate expression in mesoderm and the bracket indicates the relative location of Antp protein. (D,E) *D. virilis tsh-lacZ* expression in *Antp* and *Ubx* mutant embryos, respectively. In *Antp* mutant embryos, residual expression is seen in somatic mesoderm, but is lost in epidermis. *Ubx* mutant embryos show a complete lack of epidermis and mesoderm staining in the abdomen. Unlike melanogaster *tsh-lacZ* expression, there is no expansion of epidermal expression into abdominal epidermis (arrow).

homeotic genes activate the *tsh* enhancer, four in the epidermis and two in the somatic mesoderm.

Regulation of the enhancer by homeotic gene products was further tested using heat-inducible homeotic gene expression. Overexpression of Antp by heat induction resulted in expression of the *tsh-lacZ* in more anterior locations. Most of the ectopic expression is in the epidermis, the tissue where *Antp* normally activates the *tsh* enhancer (Fig. 4E, arrows). Heat treatment of embryos without an inducible homeotic protein does not cause anterior epidermal *tsh-lacZ* expression

(Fig. 4D). Ectopic Antp production also induces low levels of expression in PS3-5 mesoderm and head mesoderm (4E, arrowheads). Overexpression of Ubx protein by heat induction also causes the *tsh* enhancer to be expressed in more anterior locations in two tissues where *Ubx* regulates the *tsh* enhancer: very strongly in mesodermal derivatives (Fig. 4F, arrowheads) and moderately in epidermis (Fig. 4F, arrow).

Sequence conservation of the *tsh* enhancer in two fly species

With the expectation that important regulatory sequences in the *tsh* enhancer have been conserved during evolution, we isolated the corresponding enhancer from *D. virilis*, a species that diverged from *D. melanogaster* approximately 60 million years ago (Beverley and Wilson, 1984). A *D. virilis* genomic library was screened at low stringency with the *tsh* cDNA and DNA fragments 3′ to the *tsh* exons. Five overlapping phages, which reveal a single gene, were recovered (Fig. 5A). A region of approximately 25 kb separates the *tsh* enhancer from the coding sequence in *D. virilis* (Fig. 5B) but is not present in the corresponding position in *D. melanogaster* (Fig. 5A). *D. virilis* DNA hybridizing to the *tsh* enhancer was located in a 2 kb fragment, which was then fused to the same promoter-*lacZ* gene used in the *D. melanogaster* experiments.

The *D. virilis tsh-lacZ* expression pattern in *D. melanogaster* embryos (Fig. 5C) is similar to, but distinct from, the pattern found with the *D. melanogaster* fragment (Fig. 3A). β-gal is observed in thoracic and abdominal epidermis (arrows), and in thoracic and abdominal somatic mesoderm (arrowheads). The *D. virilis* enhancer directs high level β-gal synthesis in thoracic somatic mesoderm, where the *D. melanogaster* enhancer is inactive, and lower levels in thoracic and abdominal epidermis than does the *D. melanogaster* enhancer (compare Fig. 5C with Figs 3A or 4D). In *Antp* mutant embryos expression of *D. virilis tsh-lacZ* is lost from thoracic epidermis and reduced in mesoderm (Fig. 5D). In *Ubx* mutant embryos, abdominal epidermis and mesoderm expression is lost (Fig. 5E). Thus *D. virilis* and *D. melanogaster* enhancers are similarly regulated by homeotic genes except in the thoracic mesoderm.

The nucleotide sequences of the 1.3 kb *D. melanogaster tsh* enhancer and of the 5 kb surrounding the *D. virilis tsh* enhancer were determined and aligned. Five blocks of sequence homology are evident within a 500 bp region of each enhancer (Fig. 6). Nucleotide identities for the *D. melanogaster* and *D. virilis tsh* enhancer sequences are shown in capital letters, and the five blocks are boxed and arbitrarily numbered 1-5 (Fig. 6). The five conserved regions are colinear, with similar spacing between the conserved boxes in both species. Of the 28 TAAT potential homeotic protein-binding site motifs within the *D. melanogaster tsh* enhancer only four are conserved in *D. virilis.*

Antp and Ubx bind similarly to conserved *tsh* enhancer sequences

In order to determine whether the conserved sites represent *cis* regulatory sequences through which homeotic proteins act, we examined protein-DNA interactions by Dnase I footprinting (Fig. 7B). Bacterially expressed Antp (A; lanes 2, 6, 10, 14,

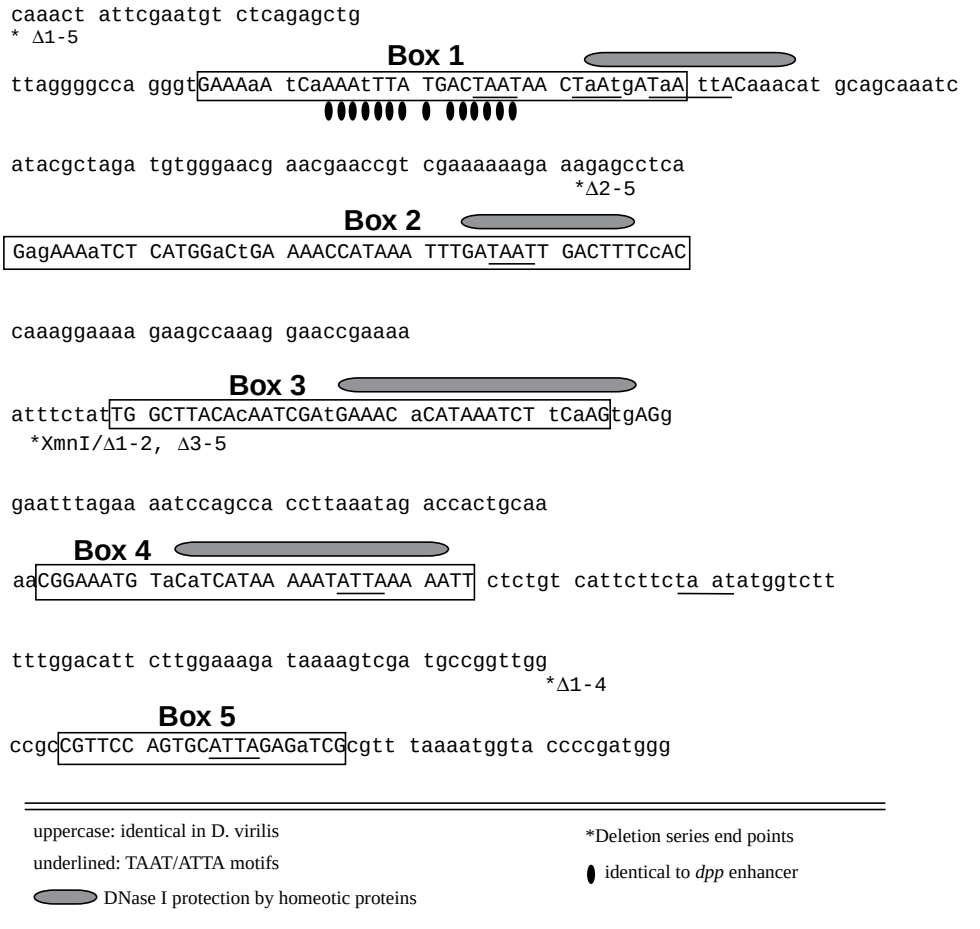

Fig. 6. Sequence conservation analysis indicate five important homeotic regulatory elements within the *tsh* enhancer**.** The sequence of the *tsh* enhancer conserved region; regions similar to the *D. virilis* DNA sequence is shown in capital letters and are boxed. The boundaries of homeodomain-binding sites are indicated above the sequence by filled ovals. TAAT motifs are underlined. In box 1, the nucleotides marked with dots are identical to a sequence within the homeotic response element in the *dpp* midgut enhancer (see Discussion). Important restriction sites and the positions of deletion series endpoints are indicated (asterisks).

15), Ubx (U; lanes 3, 7, 11, 16) and abdA (a; lanes 4, 8, 12), or the control protein bovine serum albumin (B; lanes 1, 5, 9, 13), was used to locate binding sites within the *tsh* enhancer. Probes used for this analysis are indicated in Fig. 7A.

Antp, Ubx and abdA bind strongly within boxes 1-3, weakly in box 4, and, despite the presence of a TAAT motif, not at all within box 5 (Fig. 6, diagram; Fig. 7, lanes 1-16). Antp protects box 1 with slightly different boundaries of protection than Ubx or abdA (compare lanes 2 with lanes 3 and 4). However, on the opposite strand, no such differences are observed (compare lane 6 with lanes 7 and 8). Box 2 protection by Antp, Ubx or abdA is nearly identical, but abdA has a different hypersensitive site on one strand (lane 4). Box 3 protection exhibits the most variability: abdA binds with lower affinity (lanes 8 and 12), while Ubx and Antp protect more strongly and generate a different pattern of hypersensitive sites (compare lanes 10 and 11). Although bound by Antp, abdA and Ubx at an AT-rich sequence, box 3 does not contain a TAAT motif. Box 4 binds Antp, Ubx and abdA with lower affinity than boxes 1, 2 or 3, with protection seen only as a reduction in the strength of a single hypersensitive site (lanes 10-12, and 14-16). Box 5 is not protected by Antp or Ubx (nor abdA; data not shown).

Homeotic protein-binding sites (ovals; Figs 6, 7) lie within conserved regions (boxes; Figs 6, 7). The conserved regions extend well beyond the nucleotides protected from Dnase I digestion by any of the homeotic proteins, suggesting that these may be binding sites for other regulators.

Conserved regions of the *tsh* enhancer control expression in different tissues, and are regulated by homeotic proteins in vivo

In order to examine the roles of the conserved boxes in responding to homeotic proteins and other regulators, deletions of DNA from either the 5′ or 3′ end of the 1.3 kb *D. melanogaster* enhancer were constructed and introduced into flies (Fig. 8). *tsh* enhancer expression pattern can be almost completely provided by a 500 bp region containing boxes 1-5 (Δ1-5, Fig. 8B).

Smaller fragments, from which some of the conserved boxes are absent, were tested for their ability to drive *lacZ* expression. Δ2-5 (Fig. 8C), which lacks box 1 but retains the others, has an expanded pattern in the epidermis compared to Δ1-5. At germ-band-extended stages of development more cells are stained in the anterior of PS3 (Fig. 8C), and in germ-band-shortened embryos substantially more *tsh-lacZ* expression is

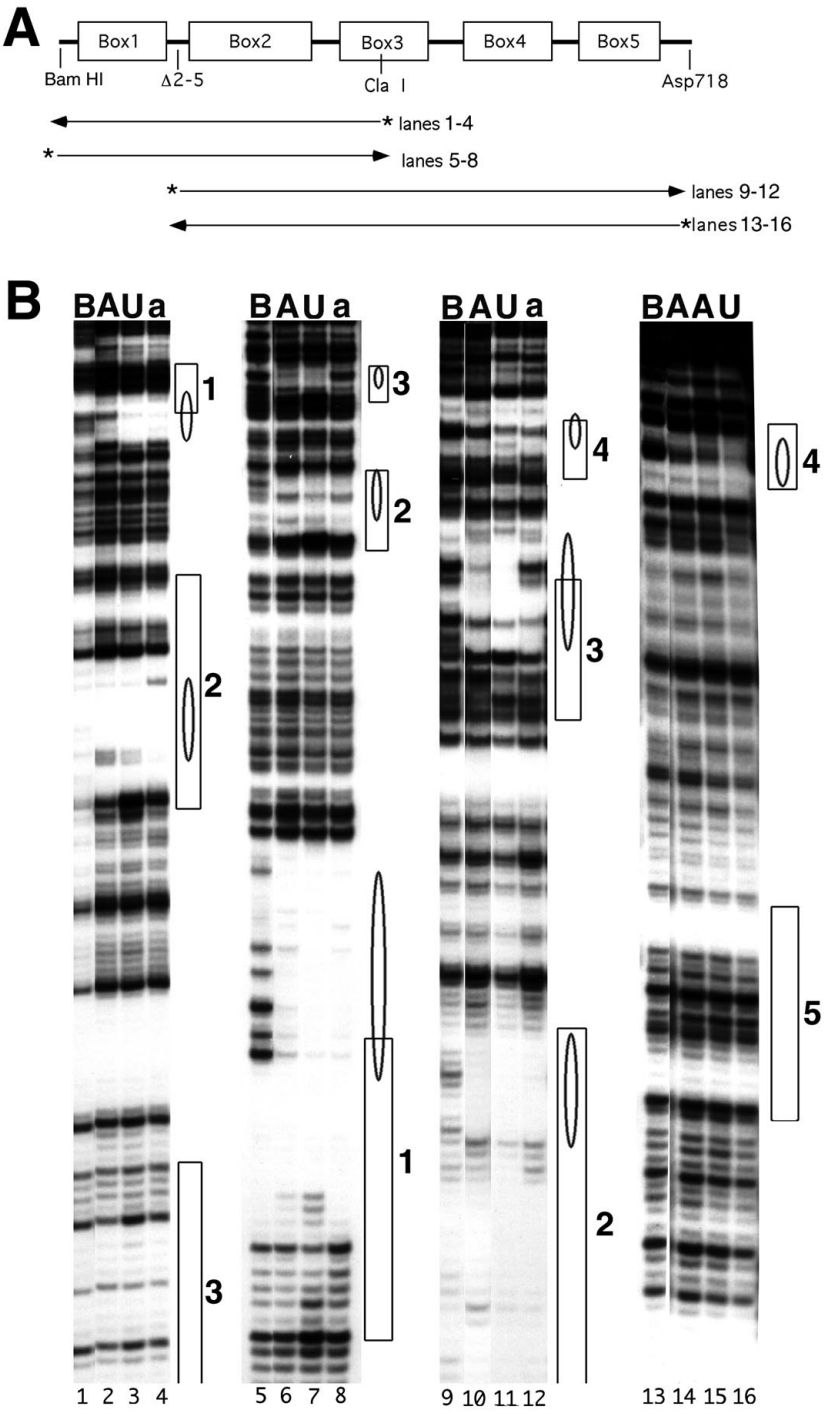

Fig. 7. Homeotic proteins bind directly to conserved sequences. (A) A diagram of the probes used in Dnase I analysis. Probes were end labeled with [^{32}P]dATP by fill in (as indicated by an asterisk), and purified after electrophoresis using glass powder. (B) Footprint analysis of the *tsh* enhancer. Antp homeodomain protein was expressed in bacteria as a GST fusion, and purified using GST-Agarose. Purified full-length Ubx and abdA proteins were expressed in bacteria, enriched by several column chromatographic steps and finally by DNA affinity chromatography. 40 fmole ^{32}P-end-labeled *tsh* enhancer DNA was incubated with 20 ng Antp (A); Ubx (U) and abdA (a) proteins for 10 minutes on ice (except lane 15 which was incubated with 60 ng of Antp protein), and digested for 5 minutes on ice with Dnase I at a final concentration of 5 ng/ml. Protein/DNA products were separated by phenol extraction, and DNA products were resolved by electrophoresis on a 5% acrlyamide/7 M urea gel, and visualized by autoradiography overnight. The positions of binding sites were determined by Maxam and Gilbert sequence reactions (not shown). Protected sites are indicated by ovals, and the relative positions of conserved DNA sequences are indicated by number and correspond to numbered boxes indicated in Fig. 6.

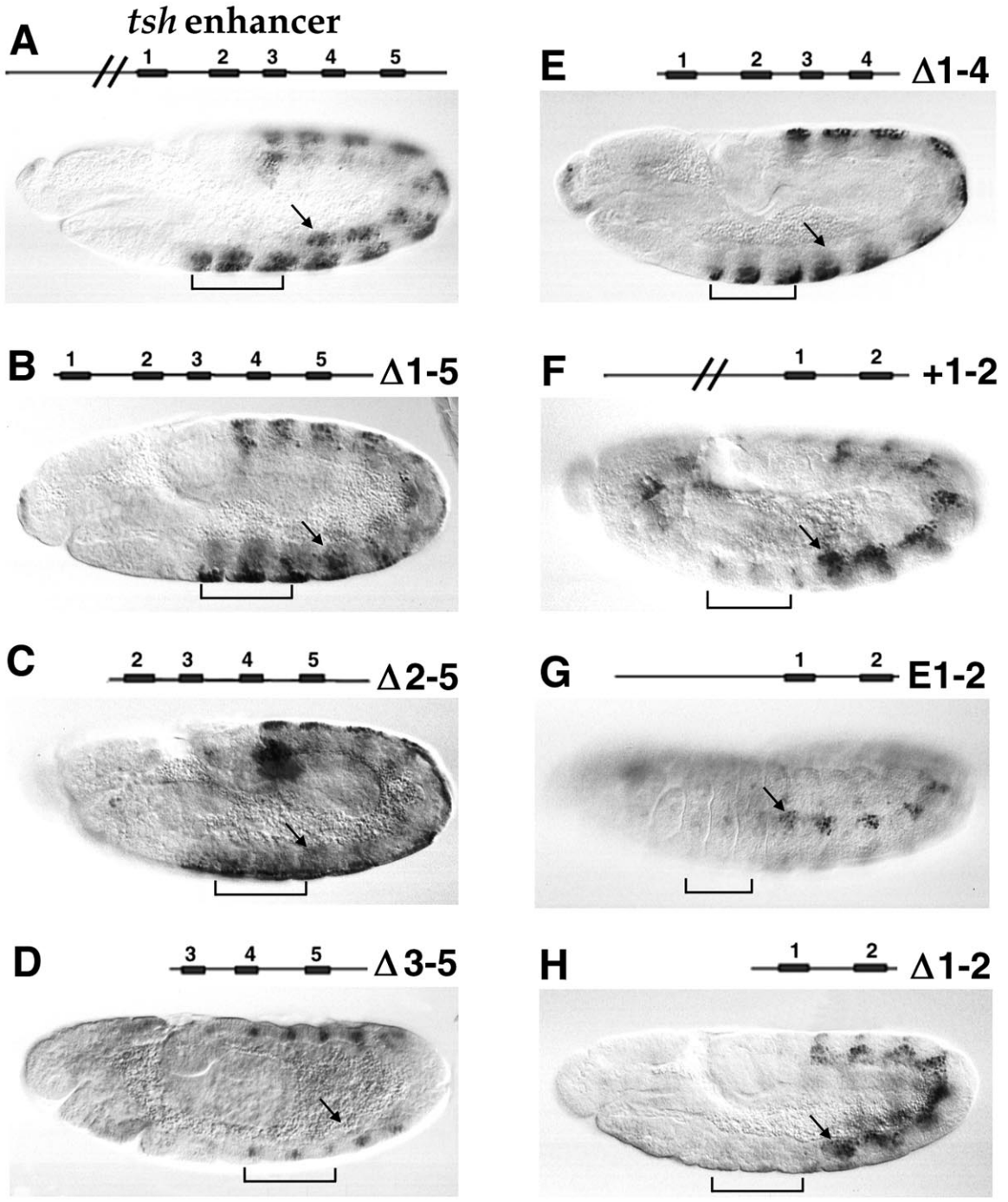

Fig. 8. Deletion analyses indicate that conserved regions are important for regulation by *Antp* and *Ubx* in different tissues. Schematic of the deletion constructs tested in vivo is shown above each panel. DNA was placed 5′ to a promoter *lacZ* fusion in the construct C4PLZ (see Materials and Methods). (A-H) Embryos transgenic for each of the constructs shown were stained for the presence of β-gal protein. Construct Δ1-5 (B), which lacks sequences from the 5′ end of the *tsh* enhancer that are not obviously conserved in the *D. virilis* enhancer, drives *lacZ* in a pattern similar to that of the full-length enhancer in epidermis and mesoderm. The anterior boundary of *tsh* enhancer activity still coincides with that of *Antp* expression (brackets). Expression in abdominal epidermis is slightly reduced, and some elevation of β-gal in somatic mesoderm (arrows) is also evident (compare A to B). DNA spanning the non-conserved sequence upstream of the 0.5 kb conserved region gives no expression as an 800 bp fragment (data not shown), but a smaller fragment of just 300 bp at the 5′ end generates β-gal expression in all mesodermal cells (data not shown).

evident in lateral and dorsal cells (not shown). However, Δ2-5 lacks all somatic mesoderm expression, the part of the pattern attributable exclusively to *Ubx*. This suggests that conserved box 1 is critical for both mesoderm-specific enhancer activity and dorsal/anterior boundary regulation in epidermis. Deletion of both box 1 and box 2 (construct Δ3-5; Fig. 8D), leaves only

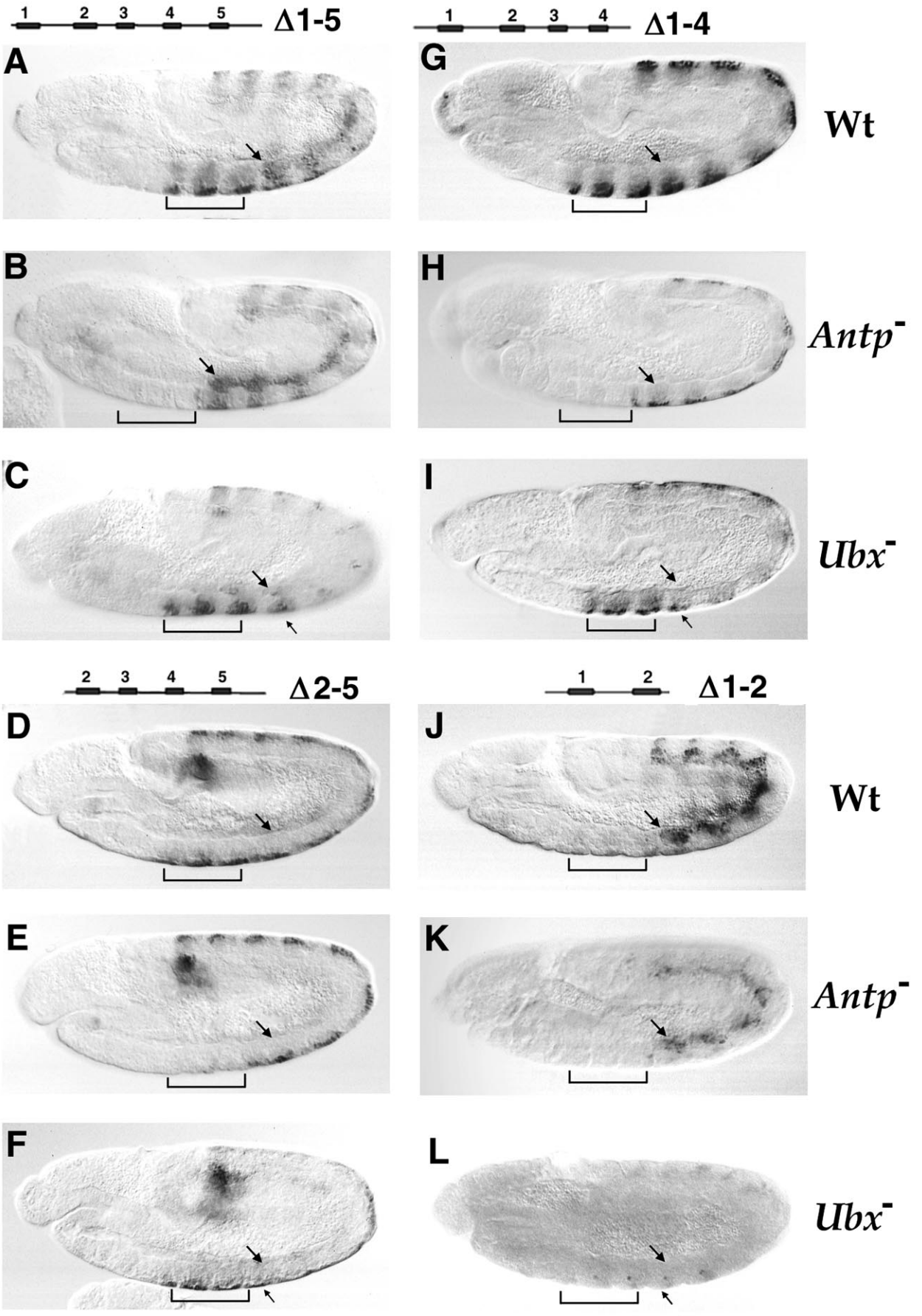
1 2 3 4 5
Δ1-5
A
B
C
1 2 3 4
Δ1-4
G
H
I
Wt
Antp-
Ubx-
2 3 4 5
Δ2-5
D
E
F
1 2
Δ1-2
J
K
L
Wt
Antp-
Ubx-

Fig. 9. *Antp* and *Ubx* regulate smaller *tsh* enhancer fragments in a tissue-specific manner. Schematic of the deletion constructs tested in vivo is shown above each panel. Expression of β-gal in wild-type embryos (A,D,G,J) is compared to patterns of expression in *Antp* (B,E,J,K) or *Ubx* (C,F,I,L) mutant embryos. Brackets indicate the position of Antp localization, arrows the location of somatic mesoderm. In C, F, L and I an additional arrow marks the position of derepressed epidermal expression which is due to the presence of Antp in PS6. The embryo pictured in L is highly overstained (20 minute reaction time compared to 2-5 minutes) to reveal β-gal expression in a few mesodermal cells.

a severely reduced level of expression in a single row of cells at the posterior of every parasegment, in the dorsal but not ventral region. These results indicate the importance of box 1 for mesoderm and of box 2 for epidermis expression.

Construct +Δ1-2, which lacks boxes 3, 4 and 5 but has all of the 5′ enhancer sequence joined to boxes 1 and 2, has mesoderm expression but no epidermal expression (Fig. 8F). A short piece containing boxes 1 and 2, together with some 5′ sequence, is also sufficient for mesoderm expression (EΔ1-2, Fig. 8G). Boxes 1 and 2 with no upstream sequence (Fig. 8H, Δ1-2) drive very strong mesoderm expression with no epidermal expression. Curiously, the addition of boxes 3 and 4 to boxes 1 and 2 blocks mesoderm expression (compare Δ1-4, Fig. 8E, with Δ1-2, Fig. 8G). Box 5 is required for mesoderm activity, but only if boxes 3 and 4 are present. Construct Δ1-4, which lacks the 3′-most box 5 (Fig. 8E), shows a dramatic loss of mesoderm expression while levels of expression in trunk epidermis are comparable to expression from the full-length or box 1-5 *tsh* enhancer (Fig. 8A). In summary, box 2 drives strong epidermal activation in conjunction with boxes 3, 4 and 5, but has no epidermal enhancer activity in combination with box 1. The combination of *tsh* enhancer elements strongly influences their function (summarized in Fig. 10).

Selected deletion series constructs were tested in *Antp*$^{-}$ and *Ubx*$^{-}$ genetic backgrounds (Fig. 9A-L). Each construct responds to homeotic mutations much as the parent construct does. The pattern of expression of Δ1–2, which is expressed only in mesoderm, is unaffected in an *Antp* mutant embryo (Fig. 9K). In a *Ubx* mutant embryo, Δ1-2 gives expression in a few cells which are only seen after extensive horseradish peroxidase incubation (Fig. 8L). In germband-retracted embryos, expression is seen in 4-5 mesodermal cells in every abdominal hemi-segment (data not shown).

DISCUSSION

We have described regulation in vivo of a *tsh* enhancer by the homeotic genes *Antp*, *Ubx*, *abdA* and *AbdB*. This is the first well-characterized HOMRE regulated by Antp protein. The enhancer is organized as a composite, with smaller control elements responding to specific homeotic proteins in specific tissues. Antp and Ubx, which have similar homeodomains, activate *lacZ* transcription by acting through different HOMREs within the enhancer. The *tsh* enhancer's arrangement of *cis* elements, in tissue-specific blocks, is revealing how an enhancer acquires regulatory specificity. The results strongly predict the existence of tissue-specific regulatory factors, or factors differentially active in different tissues,

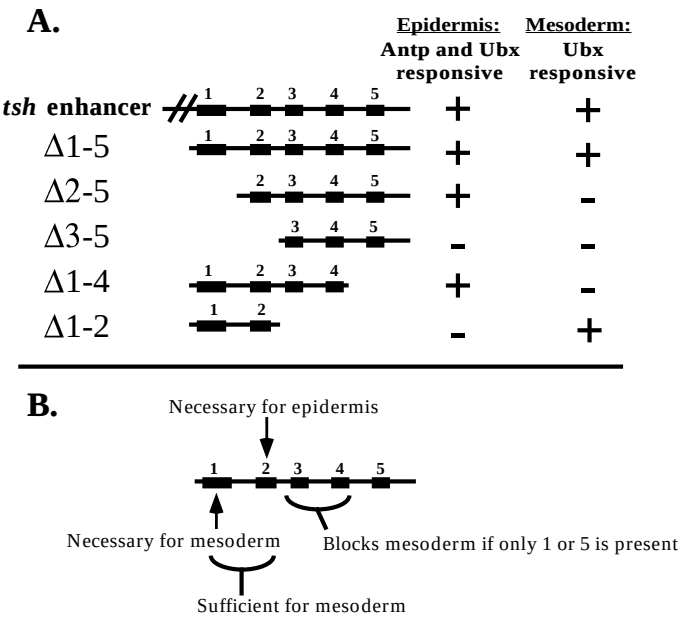

Fig. 10. Summary of results. (A) Expression of the *tsh* enhancer and fragments of it. The dependence of expression on Ubx and Antp is inferred for some of the deletions. (B) Requirements for epidermal and mesodermal expression.

which closely interact with homeotic proteins on the enhancer. We have described the conserved binding sites upon which these factors act.

Tissue-specific homeotic protein action and the organization of the *tsh* enhancer

Antp and Ubx proteins are both present in somatic mesoderm cells (Antp in the thorax and Ubx in the abdomen) but only *Ubx* activates the *tsh* enhancer in somatic mesoderm under normal conditions. In the epidermis, however, Antp is fully capable of activating the *tsh* enhancer.

The results of the deletion analysis show that Box 1, or a sequence near it, is required for mesoderm expression and activation by *Ubx*, and that Box 2, or a sequence near it, is required for epidermis activation by homeotic proteins (Fig. 10). It is striking that homeotic responsive elements within the enhancer are tightly linked to DNA elements that also control tissue specificity. A different arrangement is easy to imagine, with the sequences driving transcriptional activation in mesoderm or in epidermis in one region of the DNA and the sequences responsive to homeotic proteins located in another region of the DNA. The tight linkage of tissue specificity and HOMREs is also observed with a 45 bp fragment of the *Ubx*-responsive *dpp* midgut enhancer (Manak et al., 1994), with the *Deformed* (*Dfd*) autoregulatory HOMRE (Zeng et al., 1994), and with the *Antp* P2 promoter, which is repressed in a tissue-specific manner by *Ubx* and *abd A* (Appel and Sakonju, 1993). In no previous case have two tissue-specific elements been found tightly linked as in the *tsh* enhancer.

Evidence is growing for a general principle of homeotic-responsive enhancer organization: homeotic protein-binding sites are immediately adjacent to elements important for tissue-specific regulation, rather than tissue-specific responsiveness being due to one enhancer and homeotic responsiveness to

another. The tight linkage between tissue specificity and homeotic response elements is one mechanism by which homeotic proteins can regulate distinct target genes in different tissues.

A high level of homeotic protein can overcome certain tissue-specific limits to expression

Ectopic expression of either Antp or Ubx (Fig. 4) can activate the enhancer in thoracic somatic mesoderm, indicating that a high level of either protein is sufficient to overcome the block to activation in those cells. Normally, Ubx is not made in thoracic cells, so in the heat-shock experiments Ubx is merely doing in the thoracic cells what it normally does in the abdomen. Antp, in contrast, acquires a novel ability. The ability of overexpressed protein to take on properties normally seen with another protein has many precedents in yeast, where genes expressed from high copy number plasmids can complement mutations in different genes. There is also a precedent with homeotic genes: overexpressed Ubx can transform antennae into legs just as Antp does (Mann et al., 1990).

Overexpressed homeotic proteins are incapable of activating the *tsh* enhancer in the nervous system or visceral mesoderm. The enhancer may require a factor present only in the somatic mesoderm and epidermis, or be repressed by factors present in other tissues. In part of the posterior head overexpressed Antp fails to activate the *tsh* enhancer, either in epidermis or mesoderm, and HS-*Ubx* activates less strongly in mesoderm, so other factors may block action of homeotic proteins in these cells.

DNA binding in vitro does not explain homeotic protein specificity

The *tsh* enhancer is likely to be directly regulated by bound homeotic proteins. The genetic data are fully consistent with this possibility, as the enhancer responds both to loss of function and gain of function of *Antp* and *Ubx*. Further support is provided by the binding of homeotic proteins in vitro to evolutionarily conserved sequences.

The difference between in vitro binding and in vivo activity of homeotic proteins is most clearly demonstrated by boxes 1 and 2. Boxes 1 and 2 are both bound by Antp and Ubx proteins, yet display tissue and homeotic protein specificity in vivo. Antp and Ubx bind box 1 and 2 sequences similar to consensus binding sites defined by extensive in vitro analyses (Ekker et al., 1991, 1994; Laughon, 1991) and found in other target gene HOMREs (Appel and Sakonju, 1993; Capovilla et al., 1994; Manak et al., 1994) Of the three closely linked TAAT motifs in box 1, two have the sequence TAATAG while the other, like the TAAT motif in box 2, has the sequence TAATAC (Fig. 6). In vitro TAATAG and TAATAC are near-optimal for Ubx or Antp binding.

Given the behavior of boxes 1 and 2 in vivo, box 1 and box 2 DNAs might be expected to bind Antp and Ubx proteins with different affinities in vitro. For instance, Ubx might bind with higher affinity to box 1 than to box 2, or Ubx might bind with higher affinity than Antp to box 1 DNA, and with affinity similar to Antp on box 2 DNA. Gel mobility shift analyses confirm that Antp and Ubx bind to conserved box 1 and 2 DNAs with nearly identical binding affinity (data not shown), suggesting that binding affinities measured in vitro do not explain activity differences observed in vivo.

In the context of Δ1-4 the mesodermal enhancer box 1 is silent. This may be due to interaction among proteins bound to epidermal enhancer elements (presumably boxes 2, 3 and 4), which prevent activation by *Ubx* in mesoderm while still allowing *Ubx* to activate in epidermis. The opposite phenomenon is seen in the context of Δ1-2: box 2, which normally contributes to transcriptional activation of *tsh-lacZ* in epidermis, is now silent. Multimerization of HOMREs has already been proposed as a possible mechanism to regulate target genes (Appel and Sakonju, 1993; Beachy et al., 1993), but in this case the combination of elements may limit where activation occurs or box 2 alone is not sufficient to drive epidermal expression.

Models for specificity of homeotic protein actions in vivo

The binding behavior of Antp and Ubx in vitro does not explain the tissue-specific differences in their activities in vivo, presumably due to the need for accessory proteins. A combination of homeotic protein and tissue-specific factor, *independently* binding the enhancer, could lead to the observed transcription regulation. For example a mesoderm-specific factor could work in parallel to Ubx to activate transcription through box 1. Similarly in box 2 either Antp in the thoracic epidermis or Ubx in the abdominal epidermis could work in parallel with an epidermis-specific factor. In each case the homeotic and other proteins would bind independently and present a surface to the transcription machinery which successfully activates transcription.

A second model invokes a tissue-specific factor, or a factor active only in one tissue, that would bind cooperatively in a complex that includes the homeotic protein. In this case the ability of the homeotic protein to activate the enhancer in a particular tissue type would be due to compatibility with a tissue-specific factor for binding. For example Antp would not interact efficiently with a somatic mesoderm factor bound to box 1, while Ubx would. Overproduction of Antp could lead to mesoderm activation, despite the poor compatibility with the other factor, because the high level of protein forces occupancy of the unfavorable site. Cofactor proteins could also act as selective competitors, blocking the binding of one homeotic protein but not another.

Possible roles of conserved sequences outside homeodomain-binding sites

The close linkage of sequences conserved for at least 60 million years to the sites where homeotic proteins bind in vitro suggests that the binding of cofactor proteins is interdependent with homeotic protein binding, either by cooperation, synergism or by competition. The large blocks of conserved sequence adjacent to the homeodomain binding sites are dramatic. Within box 2, more than 50 bp are nearly perfectly conserved; 30 of the 50 bp are not protected by homeotic proteins from Dnase I digestion.

Box 1 contains a DNA sequence similar to the *Ubx*-regulated *dpp* midgut enhancer (Capovilla et al., 1994; Manak et al., 1994; Sun et al., 1995), one of the few well characterized HOMREs. The nucleotides in common are indicated by dots below box 1 in Fig. 6. Both the *dpp* enhancer and the *tsh* enhancer fragments containing box 1 are regulated by *Ubx*, but

they differ in tissue specificity. The *tsh* enhancer box 1 is required for somatic mesodermal activation, while the *dpp* enhancer is activated by *Ubx* in visceral mesoderm. The *tsh* enhancer sequence resembles a part of the *dpp* enhancer ('e3' exd-binding site; Sun et al., 1995) thought to bind Ubx and exd proteins. The *tsh* enhancer may also be regulated by a combination of *Ubx* and *exd* bound to box 1. If exd is required for activation in mesoderm through box 1 enhancer sequences, and Antp does not efficiently interact with exd (Johnson et al., 1995), this may explain why only higher than normal levels of Antp can activate the *tsh* enhancer in the mesoderm.

The ancient origins of the clustered Hox genes, and their parallel activities in controlling anterior-posterior development in diverse animals, make possible the existence of principles of HOMRE organization. Here we have suggested that one of those principles may be protein-protein interactions which confer tissue specificity and, in some cases, homeotic protein specificity as well.

We thank Laura Mathies for antibodies, the Bloomington, Indiana Drosophila stock center for efficiently helping with our many requests, and Steve Crews for unpublished plasmid vectors. We are grateful to David Clayton, Jerry Crabtree, David Kingsley, Roel Nusse and Mark Krasnow for their critiques of the manuscript. This research was supported by NIH Grant no. 18163 to M. P. S. and by a Jane Coffin Childs Postdoctoral Fellowship to A. M. We also gratefully acknowledge a McCormick Fellowship to A. M. from Stanford University. M. P. S. is an Investigator of the Howard Hughes Medical Institute.

REFERENCES

Andrew, D. J., Horner, M. A., Petitt, M. G., Smolik, S. M. and Scott, M. P. (1994). Setting limits on homeotic gene function: Restraint of *Sex combs reduced* activity by *teashirt* and other homeotic genes. *EMBO J.***13,** 1132-44

Andrew D. J. and Scott, M. P. (1992). Downstream of the homeotic genes. *New Biologist* **4**, 5-15.

Appel, B. and Sakonju, S. (1993). Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products Ubx and Abd-A in Drosophila embryos. *EMBO J.* **12**, 1099-1109

Beachy, P.A., Varkey, J., Young, K. E., von Kessler, D. P., Sun, B. I. and Ekker, S. C. (1993). Cooperative binding of an *Ultrabithorax* homeodomain protein to nearby and distant DNA sites. *Molec. Cell. Biol.* **13**, 6941-56

Beverley, S. M. and Wilson, A. C. (1984). Molecular Evolution in Drosophila and the higher Diptera. II. A time scale for fly evolution. *J. Mol. Evol.* **21**, 1-13

Bienz, M., Saari, G., Tremml, G., Müller, J., Zust, B. and Lawrence, P. A. (1988). Differential regulation of *Ultrabithorax* in two germ layers of *Drosophila*. *Cell* **53**, 567-76

Botas, J. (1993). Control of morphogenesis and differentiation by HOM/Hox genes. *Curr. Opinion in Cell Biol.* **5**, 1015-1022

Bürglin, T. R. (1994). The homeodomain phylum: a comprehensive survey of homeodomain sequences and associated motifs. *Guidebook to the Homeobox Genes.* (ed. D. Duboule). Sambrook & Tooze Publication, Oxford Press.

Campos-Ortega, J. A. and Hartenstein, V. (1985). *The Embryonic Development of Drosophila melanogaster*. Berlin: Springer-Verlag.

Capovilla, M., Brandt, M. and Botas, J. (1994). Direct regulation of *decapentaplegic* by *Ultrabithorax* and its role in midgut morphogenesis. *Cell* **76**, 461-475

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

Carroll, S. B., DiNardo, S., O'Farrell, P. H., White, R. A. and Scott, M. P. (1988). Temporal and spatial relationships between segmentation and homeotic gene expression in Drosophila embryos: distributions of the *fushi tarazu*, *engrailed*, *Sex combs reduced*, *Antennapedia*, and *Ultrabithorax* proteins. *Genes Dev.* **2**, 350-60

Celniker, S. E., Keelan, D. J. and Lewis, E. B. (1989). The molecular genetics of the bithorax complex of *Drosophila*:: characterization of the products of the *Abdominal-B* domain. *Genes Dev.* **3**, 1424-1436

Chan, S. K. and Mann, R. S. (1993). The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. *Genes Dev.* **7**, 796-811

Denell, R. E., Hummels, K. R., Wakimoto, B. T. and Kaufman, T. C. (1981). Developmental studies of lethality associated with the Antennapedia gene complex in *Drosophila melanogaster*. *Dev. Biol.* **81**, 43-50

Desplan, C., Theis, J. and O'Farrell, P. H. (1988). The sequence specificity of homeodomain-DNA interaction. *Cell* **54**, 1081-1090

de Zulueta, P., Alexandre, E., Jacq, B. and Kerridge, S. (1994). Homeotic complex and *teashirt* genes co-operate to establish trunk segmental identities in *Drosophila*. *Development* **120**, 2287-2296

Duncan, I. M. (1987). The bithorax complex. *Ann.Rev.Genett.* **21**, 285-319

Ekker, S. C., Young, K. E., Von Kessler, D. P. and Beachy, P. A. (1991). Optimal DNA sequence recognition by the *Ultrabithorax* homeodomain of *Drosophila*. *EMBO J.***10**, 1179-1186

Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E. and Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. *EMBO J.* **13**, 3551-3560

Fasano, L., Röder, L., Coré, N., Alexandre, E., Vola, C., Jacq, B. and Kerridge, S. (1991). The gene *teashirt* is required for the development of Drosophila embryonic trunk segments and encodes a protein with widely spaced zinc fingers. *Cell* **64**, 63-79

Gibson, G., Schier, A., LeMotte, P. and Gehring, W. J. (1990). The specificities of *Sex combs reduced* and *Antennapedia* are defined by a distinct portion of each protein that includes the homeodomain. *Cell* **62**, 1087-1103

Hakes, D. J. and Dixon, J. E. (1992). New vectors for high level expression of recombinant proteins in bacteria. *Analytical Biochemistry* **202**, 293-299

Hoey, T., Warrior, R., Manak, J. and Levine, M. (1988). DNA-binding activities of the *Drosophila melanogaster even skipped* protein are mediated by its homeodomain and influenced by protein context. *Mol. Cell. Biol.* **8**, 4598-4607

Hope, I. A. and Struhl, K. (1985). GCN4 protein, synthesized in vitro, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthethic genes in yeast. *Cell* **43**, 177-188

Johnson, F. B. and Krasnow, M. A. (1992). Differential regulation of transcription preinitiation complex assembly by activator and repressor homeodomain proteins. *Genes Dev.* **6**, 2177-2189

Johnson F. B., Parker, E. and Krasnow, M. A. (1995). Extradenticle protein is a selective cofactor for the Drosophila homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. *Proc. Nat. Acad. Sci. USA* **92**, 739-43.

Jones, K. A., Yamamoto, K. R. and Tjian, R. (1985). Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. *Cell* **42**, 559-572

Kaufman, T. C., Seeger, M. A. and Olsen, G. (1990). Molecular and genetic organization of the Antennapedia gene complex of *Drosophila melanogaster*. *Adv. Genet.* **27**, 309-362

Krumlauf, R. (1994). *Hox* genes in vertebrate development. *Cell* **78**, 191-201

Kuziora, M. A. and McGinnis, W. (1988). Different transcripts of the Drosophila Abd-B gene correlate with distinct genetic sub-functions. *EMBO J.* **7**, 3233-3244

Laughon A. (1991). DNA binding specificity of homeodomains. *Biochemistry* **30**, 11357-67

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

Lin, L. and McGinnis, W. (1992). Mapping functional specificity in the Dfd and Ubx homeo domains. *Genes Dev.* **6**, 1071-1081

Lindsley, D. L. and Zimm, G. G. (1992). *The genome of Drosophila melanogaster*. NY: Academic Press.

Manak, J., Mathies, L. and Scott, M. P. (1994). Regulation of a *decapentaplegic* midgut enhancer by homeotic proteins: *Development* **120**, 3605-3619

Manak, J. R. and Scott, M. P. (1994). A class act: Conservation of homeodomain protein functions. *Development* **1994 Supplement.** 61-71.

Mann, R. S. and Hogness, D. S. (1990). Functional dissection of Ultrabithorax proteins in D. melanogaster. *Cell* **60**, 597-610

Martinez-Arias, A. (1986). The *Antennapedia* gene is required and expressed in parasegments 4 and 5 of the Drosophila embryo. *EMBO J.* **5**, 135-141

Martinez-Arias, A. and Lawrence, P. A. (1985). Parasegments and compartments in the *Drosophila* embryo. *Nature* **313**, 639-642

Mathies, L. D., Kerridge, S. and Scott, M. P. (1994). Role of the *teashirt* gene

in *Drosophila* midgut morphogenesis: secreted proteins mediate the action of homeotic genes. *EMBO J.* **120**, 2799-2809

McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* **68**, 283-302

Morata, G. (1993). Homeotic genes of Drosophila. *Curr. Opinion in Genet.Devel.* **3**, 606-614

Rauskolb, C., Peifer, M. and Wieschaus, E. (1993). *extradenticle*, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene *pbx1*. *Cell* **74**, 1011-1112

Rauskolb, C. and Wiechaus, E. (1994). Coordinate regulation of downstream genes by *extradenticle* and the homeotic selector proteins. *EMBO J.* **13**, 3561-3569

Regulski, M., Dessain, S., McGinnis, N. and McGinnis, W. (1991). High-affinity binding sites for the Deformed protein are required for the function of an autoregulatory enhancer of the Deformed gene. *Genes Dev.* **5**, 278-286

Röder, L., Vola, C. and Kerridge, S. (1992). The role of the *teashirt* gene in trunk segmental identity in Drosophila. *Development* **115**, 1017-1033

Rubin, G. M. and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. *Science* **218**, 348-53

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Press.

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

Struhl, G. (1982). Genes controlling segmental specification in the Drosophila thorax. *Proc. Natl. Acad. Sci. USA* **79**, 7380-4

Sun, B., Hursh, D. A., Jackson, D. and Beachy, P. A. (1995). Ultrabithorax protein is necessary but not sufficient for full activation of *decapentaplegic* expression in the visceral mesoderm. *EMBO J.* **14**, 520-535

Urness, L. D. and Thummel, C. S. (1990). Molecular interactions within the ecdysone regulatory hierarchy: DNA binding properties of the Drosophila ecdysone-inducible E74A protein. *Cell* **63**, 47-61

White, R. A. H. and Wilcox, M. (1985). Distribution of *Ultrabithorax* proteins in *Drosophila*. *EMBO J.* **4**, 2035-2043

Winslow, G. M., Hayashi, S., Krasnow, M., Hogness, D. S. and Scott, M. P. (1989). Transcriptional activation by the *Antennapedia* and *fushi tarazu* proteins in cultured Drosophila cells. *Cell* **57**, 1017-1030

Wirz, J., Fessler, L. and Gehring, W. J. (1986). Localization of the *Antennapedia* protein in the Drosophila embryo and imaginal discs. *EMBO J.* **5**, 3327-3334

Zeng, C., Pinsonneault, J., Gellon, G., McGinnis, N. and McGinnis, W. (1994). Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in *Drosophila* embryos. *EMBO J.* **13**, 2362-2377

Zeng, W., Andrew, D. J., Mathies, L. D., Horner, M. A. and Scott, M. P. (1993). Ectopic expression and function of the *Antp* and *Scr* homeotic genes: The N terminus of the homeodomain is critical to functional specificity. *Development* **118**, 339-352

(Accepted 16 May 1995)