The paired box encodes a second DNA-binding domain in the Paired homeo domain protein

Jessica Treisman, Esther Harris, and Claude Desplan
Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10021 USA

The homeo box, which encodes the DNA-binding homeo domain, is a DNA sequence motif present in several Drosophila developmental genes; it has been used to identify many homologous genes involved in mammalian development. The paired box is another conserved sequence motif, first identified in the paired (prd) and gooseberry (gsb) Drosophila homeo domain genes. It encodes a 128-amino-acid domain, the paired domain, which has since been found in other fly and mouse gene products, in association with the homeo domain or in its absence. We show that the paired box of the prd gene encodes a DNA-binding activity, independent of the DNA-binding activity of the Paired (Prd) homeo domain and with a different sequence specificity. The amino-terminal region of the paired domain, including one of the three predicted α-helices, is necessary and sufficient for binding. We investigate the binding of the Prd protein to two sites in the even-skipped promoter, which are composed of overlapping sequences bound by the homeo domain and by the paired domain. We also show that a mutation in the paired box of Prd, corresponding to the mutation in the paired box of the mouse Pax-1 gene thought to cause the undulated skeletal phenotype, destroys the ability of the Prd protein to bind to the paired domain-specific site. This supports the view that the undulated phenotype results from the inactivation of the DNA-binding activity of the paired domain of Pax-1.

[Key Words: Homeo domain protein; Drosophila; paired box]

Received December 19, 1990; revised version accepted January 22, 1991.

Genetic studies in Drosophila have identified a large number of genes involved in embryonic development [Nüsslein-Volhard and Wieschaus 1980]. The products of many of these genes contain DNA-binding protein domains such as homeo domains [McGinnis et al. 1984a; Scott and Weiner 1984], zinc fingers, or helix-loop-helix motifs (for review, see Dressler and Gruss 1988), suggesting that they are transcription factors. A network of these regulators expressed in the zygote responds to maternal organizing factors to set up the body pattern of the embryo [for review, see Ingham 1988]. Several genes expressed during mammalian embryogenesis have homology to these developmental genes of the fly [Kessel and Gruss 1990]. Similar mechanisms of transcriptional regulation are therefore likely to be used during the development of many multicellular organisms. There are also other domains common to several Drosophila developmental genes that have no known function. Examples are the paired box, the PRD repeat, and the opa or M repeat [McGinnis et al. 1984b; Wharton et al. 1985; Bopp et al. 1986; Frigerio et al. 1986]. Some of these domains have also been used as probes to identify homologous genes that may function in vertebrate development [see Kessel and Gruss 1990].

The paired box was first described by Bopp et al. [1986] as a 128-amino-acid region of homology between the products of the Drosophila pair-rule gene paired (prd) and those of the two transcription units from the gooseberry (gsb) segment polarity locus. These three proteins also contain highly homologous homeo domains [Baumgartner et al. 1987]. Subsequently, two genes that contain a paired box motif but no homeo box were identified in Drosophila [Bopp et al. 1989]; the Pox meso and Pox neuro genes are expressed in the mesoderm and the nervous system, respectively, and are under the control of segmentation genes such as prd. Furthermore, antibody staining has shown that their products are localized to the nucleus, raising the possibility that they could act as transcriptional regulators [Bopp et al. 1989].

Structure predictions suggest that the domain encoded by the paired box, called the paired domain, contains three α-helices, of which the second and third are separated by 8 amino acids [Bopp et al. 1989; Burri et al. 1989]. However, there is no homology to either the helix-turn-helix or the helix-loop-helix motifs. Furthermore, the most conserved residues lie outside the helices, a feature also found in the homeo domain.

At least eight paired box genes (Pax1–Pax8) have been identified in the mouse genome [Deutsch et al. 1988; Dressler et al. 1990; Jostes et al. 1990; Plachov et al. 1990; C. Walther and P. Gruss, pers. comm.] and three in the human genome [HuP1, HuP2, and HuP48; Burri et al. 1989]. Three of the mouse genes also encode a homeo domain, whereas the others do not [Kessel and Gruss...
The set of paired box genes can be organized into three subclasses of related proteins: those containing a homeo box, those homologous to *Pox neuro*, and those homologous to *Pox meso*. Interestingly, it has recently been suggested that a point mutation in a conserved non-helical region of the paired box of the Pax-1 gene is responsible for the phenotype of the mouse *undulated* mutant (Balling et al. 1988). The phenotype of this mutant (distortion of the vertebral column) is consistent with the expression of Pax-1 in the intervertebral disks of the developing embryo.

Our previous investigations of the *Drosophila* Paired (Prd) protein have shown that it is a complex protein able to recognize more than one DNA sequence (Treisman et al. 1989). We showed that the specificity of the Prd homeo domain could be dramatically affected by point mutations of amino acid 9 of its recognition helix. In the course of this analysis, we identified a DNA-binding activity in the Prd protein that was not mediated by the helix-turn-helix motif of the homeo domain. We now show that the paired domain is necessary and sufficient for this activity and thus constitutes a new type of DNA-binding domain. We also show that a DNA sequence from the *even-skipped* (*eve*) promoter, identified previously as a binding site for Prd (Hoey and Levine 1988), is composed of two juxtaposed sites, one bound by the homeo domain and the other by the paired domain. We show that a mixture of two mutant proteins, each with only one active domain, can completely protect the site, demonstrating that two molecules can simultaneously occupy the dual site.

Preliminary dissection of the paired domain shows that the first 90 amino-terminal amino acids are sufficient for binding. Furthermore, a mutation disrupting the putative second helix located between amino acids 80 and 88 does not affect binding, suggesting that the most conserved amino-terminal region contains the DNA-binding activity. We show that the first helical structure within this region is essential for binding. Finally, we show that a mutation in the Prd protein, corresponding to the point mutation in mouse Pax-1, which was proposed to lead to the *undulated* phenotype (Balling et al. 1988), destroys the ability of the paired domain to bind to DNA.

**Results**

**The paired domain is a DNA-binding domain**

Our previous work (Treisman et al. 1989) had shown that the Prd protein was able to bind to a consensus DNA sequence named Prd3 and that the homeo domain was sufficient to confer specific binding to this sequence. The Prd protein was also shown to bind to a site (*es*) in the *eve* promoter, but this binding did not require the presence of a functional recognition helix in the homeo domain. Nonconservative mutations at the three most conserved positions in the recognition helix (*L*, *W*, and *F*) did not lead to a loss of this second DNA-binding specificity in the protein M2.IWF, although they did prevent binding by the homeo domain to a site recognized by the amino acid at position 9 of the recognition helix.

We therefore suspected that another DNA-binding domain existed in Prd outside the recognition helix of the homeo domain. We made a series of deletions of the Prd protein to determine which parts of the molecule were necessary for this binding. Figure 1 describes the organization of the Prd protein, which contains the 128-amino-acid paired domain at its amino terminus, the 60-amino-acid homeo domain in the middle, and the PRD histidine-proline repeat at the carboxyl terminus (Bopp et al. 1986). It also shows the deletions used to define the two DNA-binding domains of Prd. Our analysis revealed that the binding to the *es* site was complex and that the two halves of this site were bound by different domains of Prd (Fig. 2A). Deleting the carboxy-terminal region of the protein, up to but not including the homeo domain, preserved binding to both halves of the site (PrdΔ, Fig. 2A). Deleting into the homeo domain abolished binding to the upper half of the site (Prd.N-RV). A deletion consisting only of the paired domain (Prd.N-Eag) still bound to the lower half of the site, as did a deletion that removes the carboxy-terminal helix of the paired domain (Prd.N-Sma). When we deleted into the paired domain from the amino terminus, leaving the rest of the protein intact, binding to the lower half of the site was lost while binding to the upper half of the site was preserved (Prd.Nco-C and Prd.Sma-C). This binding appears to be mediated by the homeo domain, since the same footprint pattern is seen with a deletion containing little more than the homeo domain (Prd.Sma-Bam) or with the Prd homeo domain in the context of the Fushi tarazu (*Ftz*) protein (*Ftz*-PrdHD, data not shown). In these experiments, binding to one half of the site serves as an internal control for activity of the protein extract and shows that failure to bind to the other half is not due to an inactive extract.

We also noted that another site (*es*$_4$) was protected by Prd proteins containing both the homeo domain and the paired domain [Prd and PrdΔ; Fig. 2A] but did not exhibit any protection by the deletions lacking the homeo domain [Prd.N-Sma, Prd.N-Eag, and Prd.N-RV]. This site showed only very faint protection by deletions lacking the paired domain [Prd.Nco-C and Prd.Sma-C]. Therefore, protection of *es*$_4$ required the presence of both the homeo domain and the paired domain and could be interpreted as an indication of cooperativity between the two domains. Comparison between the sequences protected at the *es*$_4$ and *es*$_5$ sites showed limited homology and suggested that *es*$_4$ contains sites weakly matching both the paired domain and the homeo domain sites in *es*$_5$ (Fig. 2B).

**Identification of a paired domain-specific site**

We then isolated the paired domain and Prd homeo domain-binding sites by cloning synthetic oligonucleotides containing three repeats of either the sequence footprinted by the paired domain [lower half of the *es*$_5$ site, AGCACCGTTCGGCTC] or that footprinted by the homeo domain [upper half of the *es*$_5$ site, CACCGCACGGATTAG; Fig. 2B]. As expected, the fragment contain-
ing the paired domain sites (named PB3) was protected over the three PB sites by those proteins containing an intact paired domain [Prd.N-Eag, Prd.N-RV, PrdΔ, Prd; Fig. 3] but not by proteins in which amino-terminal parts of the paired domain have been removed [Prd.Nco-C, Prd.Sma-C, Fig. 3]. A shorter version of the paired domain lacking the carboxy-terminal helix also bound to the PB3 site [Prd.N-Sma]. The presence of the homeo domain was not necessary for binding, thus, the paired box encodes a DNA-binding domain that can function independently.

**A site in the eve promoter contains binding sites for both the homeo domain and the paired domain**

We also tested the ability of our mutant proteins to bind to the synthetic Prd homeo domain site [PHD3] defined from the es site [Fig. 2B]. Figure 4 shows that the PHD3 fragment was not bound by an isolated paired domain [Prd.N-Eag] but was bound by constructs containing the Prd homeo domain in the intact Prd protein, in deletions removing part or all of the paired domain and/or the carboxy-terminal part of the molecule [PrdΔ, Prd.Sma-Bam], or inserted into the Ftz protein [FtzPrdHD]. The PHD3 site is also recognized by the homeo domain proteins Ftz, Even-skipped [Eve] [Fig. 4], Orthodontic [Otd] [data not shown], and Zerknült [Zen; see Fig. 4 in Treisman et al. 1989]. The affinities of these protein extracts for the PHD site are comparable to their affinities for the appropriate control sequence [NP6 or Bcd3; data not shown]. However, its recognition by Prd is prevented by mutations at the conserved positions Ia, WT, and Fs, which destroy the activity of the homeo domain recognition helix [Prd.IWF; Fig. 4].

The es site was first defined by Hoey and Levine [1988] as a specific target site for the Eve protein in the eve promoter. es contains two tandem copies of the Eve consensus binding site TCAGCACCG [Hoey and Levine 1988]; the PHD site extends from the middle of one to the middle of the other, while the PB site extends from the middle of the second into a region not footprinted by Eve [see Fig. 2B]. When the consensus Eve site is isolated and oligomerized it can be bound by Eve but not by Prd [data not shown]; therefore, neither the Prd homeo domain nor the paired domain recognizes the same sequences as Eve in the es site.
Paired box encodes a DNA-binding domain

Figure 2. Binding of the Prd deletions to the \( e_5 \) site. (A) DNase I protection of the \( eve \) promoter sequence. Approximately 2 ng of DNA was incubated without protein (0) or with 5, 20, or 80 \( \mu \)g protein of the soluble fractions of bacterial extracts containing the indicated proteins (increasing protein concentration is shown by arrows). DNase I footprinting was carried out as described in Materials and methods. The \( e_4 \) and \( e_5 \) sites are indicated, and the division of the \( e_5 \) site into regions protected by the homeo domain (HD) and by the paired domain (Box) is also shown. (B) Sequences of the \( e_4 \) and \( e_5 \) sites, indicating the regions protected by the Prd homeo domain (PHD, double underline) and by the paired domain (PB, boxed). The Eve consensus sequence \([\text{Eve; Hoey and Levine 1988} \]) is also shown. Nucleotides identical in both the \( e_4 \) and \( e_5 \) PB sites are indicated by bold letters.

We also tested the ability of a related paired domain-containing protein to bind to the \( e_5 \), PB3, and PHD3 sites. The Gooseberry-proximal (Gsb-p) protein contains both a paired domain and a homeo domain highly related to those of Prd [Baumgartner et al. 1987]. An \( Escherichia coli \) extract containing the Gsb-p protein expressed from the T7 promoter (a gift from Y. Zhang and R. Holmgren) footprinted these sequences in a way that is indistinguishable from Prd [data not shown].

One of the two dual \( eve \) promoter sites can be occupied by two \textit{Paired} molecules

Our demonstration that the two domains of Prd can bind DNA independently raised the possibility that the PB and PHD halves of the \( e_5 \) site are bound by a single molecule of Prd. To investigate this question we determined whether it was possible to fit two molecules onto the \( e_5 \) site. Figure 5A shows that the pattern of protection by a mixture of extracts, each containing only one of the two domains, is the sum of the patterns of protection by each extract alone. For this experiment we used Prd.N-Sma, a deletion containing only the paired domain, and Prd.Sma-Bam, a deletion containing only the homeo domain. To generalize this result to the full-length proteins, we used mutants defective in binding by one domain or the other. M2.IWF carries mutations at the most conserved positions of the homeo domain recognition helix, rendering it inactive as a DNA-binding domain [see above]. Prd.Nco-C has the amino-terminal 46 amino...
acids of the paired domain deleted and can bind only through its homeo domain. A mixture of these two extracts fully protected the e5 site [Fig. 5A], showing that it is possible to pack two essentially full-length Prd proteins onto the e5 site. The e4 site is protected by full-length Prd or by PrdA, but only weakly by an isolated homeo domain and not at all by an isolated paired domain. In contrast to e5, e4 could not be fully protected by a mixture of either two isolated domains or two full-length proteins, suggesting that intramolecular cooperativity is required for binding to this site [Fig. 5B].

A subset of the paired domain is sufficient for DNA binding

As shown in Figure 6A, the paired domain contains three predicted α-helices [Bopp et al. 1989]. The deletion analysis presented in Figure 1 demonstrates that deletion of the third helix does not affect binding. To address the role of the other two helical regions, where most of the variations between classes of paired domains occur, we created mutations that should disrupt either helix 1 or helix 2 by mutating pairs of residues in the middle of each helix to the helix-breaking amino acids glycine and proline [Fig. 6A]. Mutation of helix 1 [amino acids 23–31] destroyed the binding of Prd to PB3 without affecting binding to PHD3. However, a protein with a mutation in helix 2 [amino acids 80–88] behaved like wild-type Prd, binding to both PB3 and PHD3 [Fig. 6B]. We conclude that helix 2 is not required for the DNA-binding function or correct folding of the paired domain.

A mutation in the paired domain mimicking the mouse undulated mutation destroys binding

It has been proposed that a glycine to serine change at position 15 in the sequence of the paired domain of the mouse Pax-1 gene is responsible for the phenotype of the undulated mutant strain [Balling et al. 1988]. Although it is not known whether undulated is a null allele of the Pax-1 locus, we tested whether a mutation similar to undulated in the paired domain of Prd would affect its DNA-binding capability. Using site-directed mutagenesis, we changed the conserved glycine at position 15 of the Prd paired domain to either a serine, as in undulated, or an aspartic acid, a more dramatic change [Fig. 6A]. We
Paired box encodes a DNA-binding domain

Figure 4. A homeo domain binding site with broad specificity. DNase I footprint of the PHD3 site. Protein concentrations and footprint procedures were as described for Fig. 2. The three repeats of the PHD site are indicated by arrows (5'-CACCGCA-CGATTAG-3').

tested the ability of these mutant proteins to bind to the PB3 sites and, as an internal control for the activity of the protein, to the PHD3 sites. Figure 6B shows that both mutations destroy the ability of Prd to bind to PB3, without affecting the PHD3 binding. Therefore, this very conserved glycine appears to be essential for binding, and it is likely that the undulated phenotype is caused by the loss of DNA-binding activity of the paired domain of Pax-1.

Discussion

The paired protein contains two DNA-binding domains

The results presented above demonstrate that the paired box encodes a protein domain able to bind DNA. This binding is autonomous, since a derivative of Prd containing only this domain will bind both to a subset of a site present in the eve promoter [lower half of e3] and to the oligomeric repeat of this sequence [PB3]. It was predicted that the paired domain would have an independent function, since it exists in some proteins in the absence of a homeo domain [Deutsch et al. 1988; Bopp et al. 1989; Dressler et al. 1990; Plachov et al. 1990]. Our results provide supporting evidence for the idea that the Drosophila Pox and mouse Pax proteins, which are localized to the nucleus, act as transcriptional regulators. The significance of the frequent association of the paired domain with the Prd homeo domain remains to be determined. We detect cooperativity between the two domains when binding to the e4 site present in the eve promoter but not to the nearby e5 site, which apparently has stronger binding sites for each individual domain. Thus, the e4 site could illustrate the potential for synergism of action between the Prd homeo domain and the paired domain. We do not believe that the e4 and e5 sites mediate any action of prd on the establishment or maintenance of eve stripes in the embryo, since these aspects of eve expression are normal in a prd mutant [Frasch and Levine 1987]. However, an influence of prd on eve expression in the nervous system has not yet been investigated. It is also possible that another, undiscovered, paired domain protein acts on the e4 and e5 sites in the embryo. We are in the process of testing the functions of the two domains of Prd in vivo by introducing mutated or deleted Prd protein-coding sequences into the embryo. We also plan to examine the promoters of genes controlled by prd (or the gsb genes) for binding sites for each of the two Prd DNA-binding domains. A functional analysis of these
sites will elucidate the actions of Prd and the transcriptional role of the paired domain.

Among homeo domain proteins, the paired domain is found only in those with homeo domains homologous to that of Prd. Another restriction on its presence is suggested by the observation that in this group of proteins, all those that have a serine at position 9 of their homeo domain recognition helix also contain a paired domain. This is true of Prd itself and of the two Gsb proteins (Bopp et al. 1986). Extending these observations, we found that the only gene in a new collection of Drosophila Prd-type homeo domain proteins [B. Kalionis and P. O’Farrell, pers. comm.] that has a serine at position 9 also contains a paired domain (C. Desplan, unpubl.). In contrast, the Otd (Finkelstein et al. 1990) and frog Mix-I (Rosa 1989) proteins have homeo domains homologous to Prd but have a lysine and a glutamine at this position, respectively. They have different homeo domain DNA-binding specificities (Treisman et al. 1989) and do not contain a paired domain. The high degree of conservation of the homeo domains belonging to the Prd class may be due to their ability to interact with other protein domains, one of them being the paired domain itself. Serine at position 9 could be an example of an amino acid that maintains a structure and/or a DNA-binding specificity of the homeo domain necessary for its interaction with the paired domain. This potential cooperativity between DNA-binding domains may be of importance in vivo and is exemplified in vitro by the binding to the e4 site.

Structure of the paired domain

Secondary structure analysis of the paired domain predicts that it contains three α-helices, with 8 amino acids between the second and third helices (Bopp et al. 1989). Most of the sequence differences between different paired domains lie in the helical regions, particularly in the second helix. We show here that the third predicted helix can be deleted and the second predicted helix mutated without affecting DNA binding. The carboxy-terminal 50 amino acids are therefore not important for binding. There is often an intron separating this sequence from the first 70 amino acids (Burri et al. 1989; C. Walther and P. Gruss, pers. comm.), this reinforces the view that the two parts of the paired domain evolved as separable subdomains with distinct functions. The amino-terminal region comprises two highly conserved regions surrounding a putative helical region in which the divergence can be used to classify paired domain proteins (Burri et al. 1989). Because the helix is the most variable part of this structure, we propose that the conserved regions (including the region where the undulated mutation lies) may be required for proper folding of the domain, whereas the helical region may contact DNA and determine the specificity of sequence recognition.
Figure 6. Mutations affecting binding by the paired domain. (A) Structures of the proteins used. These were all truncated at the BamHI site. The three predicted α-helices within the paired domain are shown as hatched bars; mutations disrupting their structure are indicated by crosses (×). The conserved glycine (G) at position 15 is shown, as well as substitutions of aspartic acid (D) and serine (S) at this position. (+) Binding to PB3 or PHD3. (NA) Not applicable. (B) DNase I footprint of the PHD3 (left) and PB3 (right) sites. Protein concentrations and footprint procedures were as described for Fig. 2. The three repeats of the PHD and PB sequences are indicated by arrows.
Multiple DNA-binding mechanisms by the homeo domain

In the course of this analysis we have found the first sequence, PHD3, which is recognized by the Prd homeo domain in the context of the intact Prd protein. We previously identified a Prd protein-binding site called Prd3 [Treisman et al. 1989], but binding to this site was prevented by the presence of the carboxy-terminal part of Prd. Such inhibition is not found for binding to PHD3 [Fig. 4]. The PHD3 sequence is also bound by the Ftz, Otd, Zen, and Eve homeo domains, which differ from Prd at multiple positions in their recognition helices, including position 9. We have not yet determined which amino acids are involved in the recognition of this sequence. However, the mechanism of this recognition is clearly different from the model that we and others reported for Ftz, En, Eve, and Zen binding to the NP sequence [Desplan et al. 1988; Hoey et al. 1988] and Bcd binding to the hunchback promoter [Driever and Nüsslein Volhard 1989]. In these cases, binding specificity is mediated by amino acid 9 of the recognition helix [Hanes and Brent 1989; Treisman et al. 1989; Kissinger et al. 1990; Otting et al. 1990]. The Eve protein binds both to the PHD site in the e3 site and to the overlapping Eve consensus sequence TCAGCACC (Fig. 2B). It is not known what determines recognition of Eve sites by Eve. From studies using chimeric proteins exchanging domains of the Eve and Engrailed (En) proteins, Hoey et al. (1988) concluded that the binding of Eve to e3 was due both to the homeo domain and to the flanking regions. It is therefore also different from the mode of homeo domain DNA binding described above.

Cooperativity between the paired domain and the Prd homeo domain

The Prd homeo domain and paired domain can bind to DNA when separated from each other. However, the protection pattern of the e3 site indicates that binding to adjacent weak sites may require both domains. The presence of both domains on two different protein molecules is not sufficient, suggesting that the cooperativity observed derives from a single molecule filling the two half-sites. This implies that the two domains can simultaneously occupy adjacent sites on the same piece of DNA. Although the biological significance of such an observation is not clear, other proteins with two DNA-binding domains have been described. The λ integrase protein has been shown to have two autonomous DNA-binding domains, which may act to link two distinct parts of the recognition site [Moitoso de Vargas et al. 1988]. Other molecules also have more than one DNA-binding specificity. The yeast HAP-1 protein is one of them, although a single domain appears to mediate both specificities [Pfeifer et al. 1987]. PRDII-BF1 is a protein containing two highly related zinc fingers that can independently recognize the same DNA sequence [Fan and Maniatis 1990]. Another protein with dual DNA-binding specificities that are not yet characterized is the SV40 enhancer-binding factor TEF-1 [Davidson et al. 1988].

A mutation in the paired domain

The mouse undulated mutation appears to be caused by a point mutation in the Pax-1 gene [Ballung et al. 1988]. The change from glycine to serine is in a conserved, non-helical amino-terminal part of the domain. Changing this glycine to a serine in the paired domain of Prd abolishes binding to the PB3 sites, as does changing it to aspartic acid. It is possible that these changes prevent folding into the correct structure. Indeed, using a structure prediction computer program [Lambert and Scheraga 1989a,b, 1990], we noted that the glycine at position 15 induces a turn in a β-sheet structure extending between residues 10 and 19. This turn would not be present in the structure predicted for the same protein with serine at position 15. This observation supports the model that the undulated mutation is caused by the glycine-to-serine change in Pax-1. The paired domain is the only recognizable DNA-binding domain in the Pax-1 molecule. Because Pax-1 is likely to be a transcription factor, abolishing its DNA binding would destroy the function of the protein. The functional role of the paired domain in Prd is less clear, as Prd also contains a homeo domain. Prd carries out multiple functions, such as regulating the segment polarity genes en and gsb-d, gsb-p, and wingless [Baumgartner et al. 1987; J.R. Mullen and S. DiNardo, in prep.] and the Pax meso and Pax neuro genes [Bopp et al. 1989]. This variety of actions may be achieved by targeting the protein to different sites through multiple DNA-binding motifs. Combinatorial mechanisms are also possible, especially in view of the existence of adjacent sites that bind both domains, such as the e3 and e6 sites in the eve promoter.

In conclusion, Prd is a complex protein made of several domains; the homeo domain and the paired domain are conserved in other proteins and in other species and are DNA-binding domains. We do not yet understand the significance of this double potential. The homeo domain itself can bind to DNA in several different ways. Binding to one class of sequences [Prd3, Treisman et al. 1989] is inhibited by another domain located in the carboxyl terminus of the molecule. Binding to the PHD3 sites is not inhibited by this domain but does not distinguish Prd from divergent homeo domain proteins such as Ftz, En, or Eve. Interaction between the paired domain and the homeo domain, leading to either cooperativity on one site or linking of two sites, would further increase the complexity. This need for interaction could explain the high degree of conservation of both the homeo domain and the paired domain.

Materials and methods

Plasmids

The pARPrd expression plasmid and the p273 plasmid containing part of the eve promoter were the gifts of T. Hoey and M. Levine [Hoey and Levine 1988]. The carboxy-terminal deletions were made by cloning the following fragments of pARPrd into the expression vector pAR3040 [Rosenberg et al. 1987], cut with Ndel and HindIII, and filled in at the HindIII site with the Kle-
now fragment of DNA polymerase I: a 351-bp Ndel–Smal fragment (Prd.N-Sma), a 453-bp Ndel–Eagl fragment filled in at the Eagl site with Klenow (Prd.N-Eag), and a 715-bp Ndel–EcoRV fragment (Prd.N-RV). The Ndel site is precisely at the initiator ATG. The construction of PrdΔ is described in Treisman et al. [1989]. Prd.N-Sma-Bam was created by the same deletion procedure, beginning with Prd.N-C. The amino-terminal deletions were made by cloning a 1687-bp Ncol fragment, filled in with Klenow, into pAR3038 [Rosenberg et al. 1987] cut with BamHI and filled in with Klenow (Prd.N-Neo-C), and by cloning a 1700-bp Smal–EcoRI fragment into pAR3040 cut with BamHI and EcoRI and filled in with Klenow at the BamHI site (Prd.Sma-C). In both of these cases the Prd sequences are fused in-frame to 12 amino acids of the T7 gene 10 protein [Rosenberg et al. 1987]. Construction of FzPrdHD is described in Treisman et al. [1989].

The PHD3 and PB3 plasmids were obtained by synthesizing the oligonucleotides AGTGAGCCACAGTTTCATCCAG and their complementary strands, annealing them, and cloning into the Smal site of the pBSM13 + plasmid (Stratagene). The PARGsb-p plasmid was provided by Z. Yang and R. Holmgren.

Mutagenesis

Site-directed mutagenesis was performed as described in Treisman et al. [1989], where the creation of M2.IWF is also described. The sequences of the oligonucleotides used for mutagenesis were as follows:

Prd.IWF: ACGCAGGTTGCTGGCCTCCACCTGGTCGCGTGCCTCCGTGAG
Paired box GD15: TGGGCAAAGGACGATCGTTGATGAAAACTC
Paired box helix 2: GCGCTTGTACTCCTCGCCTGGGTTTTCGATTTCGGG
Paired box helix 1: ATCGGCGGCATCTCTCCGGGTTTAAGACGAATATT

Mutagenesis were as follows:

Site-directed mutagenesis was performed as described in Treisman et al. [1989]. The sequences of the oligonucleotides used for mutagenesis were as follows:

Prd.IWF: ACGCAGGTTGCTGGCCTCCACCTGGTCGCGTGCCTCCGTGAG
Paired box helix 1: ATCGGCGGCATCTCTCCGGGTTTAAGACGAATATT
Paired box helix 2: GCGCTTGTACTCCTCGCCTGGGTTTTCGATTTCGGG
Paired box helix 3: AGTGAGCCACAGTTTCATCCAG

and filled in with Klenow at the BamHI site (Prd.Sma-C). In both of these cases the Prd sequences are fused in-frame to 12 amino acids of the T7 gene 10 protein [Rosenberg et al. 1987]. Construction of FzPrdHD is described in Treisman et al. [1989].

References

Balling, R., U. Deutsch, and P. Gruss. 1988. Undulated, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of Pax-1. Cell 55: 531–535.

Baumgartner, S., D. Bopp, M. Burri, and M. Noll. 1987. Structure of two genes at the gooseberry locus related to the paired gene and their spatial expression during Drosophila embryogenesis. Genes & Dev. 1: 1247–1267.

Bopp, D., M. Burri, S. Baumgartner, G. Frigerio, and M. Noll. 1986. Conservation of a large protein domain in the segmentation gene paired and in functionally related genes in Drosophila. Cell 47: 1033–1049.

Bopp, D., E. Jamet, S. Baumgartner, M. Burri, and M. Noll. 1989. Conservation of the paired domain in metazoans and its structure in three isolated human genes. EMBO J. 8: 3447–3457.

Burri, M., Y. Tromvouikis, D. Bopp, G. Frigerio, and M. Noll. 1989. Conservation of the paired domain in metazoans and its structure in three isolated human genes. EMBO J. 8: 1183–1190.

Davidson, J., J.H. Xiao, R. Rosales, A. Staub, and P. Chambon. 1988. The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. Cell 54: 931–942.

Desplan, C., J. Theis, and P.H. O’Farrell. 1988. The sequence specificity of homeodomain-DNA interaction. Cell 54: 1081–1090.

Deutsch, U., G.R. Dressler, and P. Gruss. 1988. Pax-1, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. Cell 53: 617–625.

Dressler, G.R. and P. Gruss. 1988. Do multigene families regulate vertebrate development? Trends Genet. 4: 214–219.

Dressler, G.R., U. Deutsch, K. Chowdhury, H.O. Nornes, and P. Gruss. 1990. Pax-2, a new murine paired-box-containing gene, and its expression in the developing excretory system. Development 109: 787–795.

Driever, W. and C. Nüsslein-Volhard. 1989. The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. Nature 337: 138–143.

Fan, C.-M. and T. Maniatis. 1990. A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence. Genes & Dev. 4: 29–42.

Finkelstein, R., D. Smouse, T.M. Capacì, A.C. Spradling, and N. Perrimon. 1990. The orthodenticle gene encodes a novel homeo domain protein involved in the development of the Drosophila nervous system and occular visual structures. Genes & Dev. 4: 1516–1527.

Frasch, M. and M. Levine. 1987. Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in Drosophila. Genes & Dev. 1: 981–995.
Treisman et al.

Frigerio, G., M. Burri, D. Bopp, S. Baumgartner, and M. Noll. 1986. Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. Cell 47: 735–746.

Hanes, S.D. and R. Brent. 1989. DNA specificity of the Bicoid activator protein is determined by homeodomain recognition helix residue 9. Cell 57: 1275–1283.

Hoey, T. and M. Levine. 1988. Divergent homeo box proteins recognize similar DNA sequences in Drosophila. Nature 332: 858–861.

Hoey, T., R. Warrior, J. Manak, and M. Levine. 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.

Ingham, P.W. 1988. The molecular genetics of embryonic pattern formation in Drosophila. Nature 335: 25–34.

Jostes, B., C. Walther, and P. Gruss. 1990. The murine paired box gene, Pax 7, is expressed specifically during the development of the nervous and muscular system. Mech. Dev. 33: 27–38.

Kessel, M. and P. Gruss. 1990. Murine developmental control genes. Science 249: 374–379.

Kissinger, C.R., B. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo. 1990. Crystal structure of an engrailed homeo-domain-DNA complex at 2.8Å resolution: A framework for understanding homeodomain-DNA interactions. Cell 63: 579–590.

Lambert, M.H. and H.A. Scheraga. 1989a. Pattern recognition in the prediction of protein structure. I. Calculation of tripeptide conformational probabilities from the amino acid sequence. J. Comput. Chem. 10: 770–797.

——. 1989b. Pattern recognition in the prediction of protein structure. II. Chain conformation from a probability-directed search procedure. J. Comput. Chem. 10: 798–816.

——. 1990. Pattern recognition in the prediction of protein structure. III. An importance-sampling minimization procedure. J. Comput. Chem. 10: 817–831.

McGinnis, W., R.L. Garber, J. Wirz, A. Kuroiwa, and W.J. Gehring. 1984a. A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell 37: 403–408.

McGinnis, W., M.S. Levine, E. Hafen, A. Kuroiwa, and W.J. Gehring. 1984b. A conserved DNA sequence in homeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature 308: 428–433.

Moitoso de Vargas, L., C.A. Pargellis, N.M. Hasan, E.W. Bushman, and A. Landy. 1988. Autonomous DNA binding domains of lambda integrase recognize two different sequence families. Cell 54: 923–929.

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

Otting, G., Y.Q. Qian, M. Billeter, M. Muller, M. Affolter, W.J. Gehring, and K. Wuthrich. 1990. Protein–DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. 9: 3085–3092.

Otting, G., Y.Q. Qian, M. Billeter, M. Muller, M. Affolter, W.J. Gehring, and K. Wuthrich. 1990. Protein–DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. 9: 3085–3092.

Pfeifer, K., T. Prezant, and L. Guarente. 1987. Yeast HAP1 activator binds to two upstream activation sites of different sequence. Cell 49: 19–27.

Plachov, D., K. Chowdhury, C. Walther, D. Simon, J.-L. Guenet, and P. Gruss. 1990. Pax-8, a murine paired box gene expressed in the developing excretory system and thyroid gland. Development 110: 643–651.

Rosa, F.M. 1989. Mix-1, a homeobox mRNA inducible by mesoderm inducers, is expressed mostly in the presumptive endodermal cells of Xenopus embryos. Cell 57: 965–974.

Rosenberg, A.H., B.N. Lade, D.-S. Chu, S.-W. Lin, J.J. Dunn, and F.W. Studier. 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56: 125–135.

Scott, M.P. and A.J. Weiner. 1984. Structural relationships among genes that control development: Sequence homology between the Antennapedia, Ultrabithorax and fushi tarazu loci of Drosophila. Proc. Natl. Acad. Sci. 81: 4115–4119.

Treisman, J., P. Gönzcy, M. Vashishtha, E. Harris, and C. Desplan. 1989. A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell 59: 553–562.

Wharton, K.A., B. Yedvobnick, V.G. Finnerty, and S. Artavanis-Tsakonas. 1985. Opa: A novel family of transcribed repeats shared by the Notch locus and other developmentally regulated loci in D. melanogaster. Cell 40: 55–62.
The paired box encodes a second DNA-binding domain in the paired homeo domain protein.

J Treisman, E Harris and C Desplan

*Genes Dev.* 1991, 5:
Access the most recent version at doi:10.1101/gad.5.4.594

**References**

This article cites 37 articles, 9 of which can be accessed free at: [http://genesdev.cshlp.org/content/5/4/594.full.html#ref-list-1](http://genesdev.cshlp.org/content/5/4/594.full.html#ref-list-1)

**License**

**Email Alerting Service**

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