Cooperative interactions between paired domain and homeodomain

Susie Jun and Claude Desplan*

Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

*Author for correspondence (e-mail: desplan@rockvax.rockefeller.edu)

SUMMARY

The Pax proteins are a family of transcriptional regulators involved in many developmental processes in all higher eukaryotes. They are characterized by the presence of a paired domain (PD), a bipartite DNA binding domain composed of two helix-turn-helix (HTH) motifs, the PAI and RED domains. The PD is also often associated with a homeodomain (HD) which is itself able to form homo- and hetero-dimers on DNA. Many of these proteins therefore contain three HTH motifs each able to recognize DNA. However, all PDs recognize highly related DNA sequences, and most HDs also recognize almost identical sites. We show here that different Pax proteins use multiple combinations of their HTHs to recognize several types of target sites. For instance, the *Drosophila* Paired protein can bind, in vitro, exclusively through its PAI domain, or through a dimer of its HD, or through cooperative interaction between PAI domain and HD. However, *prd* function in vivo requires the synergistic action of both the PAI domain and the HD. Pax proteins with only a PD appear to require both PAI and RED domains, while a Pax-6 isoform and a new Pax protein, Lune, may rely on the RED domain and HD. We propose a model by which Pax proteins recognize different target genes in vivo through various combinations of their DNA binding domains, thus expanding their recognition repertoire.

Key words: cooperativity, DNA binding, homeodomain, paired domain, *Pax* genes

INTRODUCTION

The Pax proteins are a family of important developmental regulators that are characterized by an evolutionarily conserved DNA binding domain, the paired domain (PD; Bopp et al., 1986; Treisman et al., 1991). Seven *Pax* genes exist in *Drosophila*, including *paired* (*prd*; Frigerio et al., 1986), nine *Pax* genes are known in mouse and human (*Pax-1* to *Pax-9*) and other *Pax* genes are found in a variety of species from nematodes to vertebrates (Chalepakis et al., 1993; Stapleton et al., 1993; Wallin et al., 1993; Walther et al., 1991). About half of the *Pax* genes also encode a Paired-class homeodomain (HD), which suggests a potential interaction between the two DNA binding domains.

Pax gene expression is spatially and temporally restricted during development and in the adult. These genes are involved in a variety of developmental processes. Several of them are associated with mutant phenotypes (for review, see Strachan and Read, 1994). In particular, mutations in the *Pax-3* gene and a *Drosophila* homologue *paired*, result in the *Splotch* phenotype in mouse (Epstein et al., 1991a,b; Goulding et al., 1993), Waardenburg's syndrome in humans (Baldwin et al., 1992; Morell et al., 1992; Tassabehji et al., 1992; Tsukamoto et al., 1994), and a pair-rule phenotype in *Drosophila* (Kilchherr et al., 1986). Among the Pax proteins, Pax-6 best exemplifies the conservation of function through evolution. *Pax-6* mutations affect regulatory mechanisms of the developing eye, *Small eye* in mouse (Hill et al., 1991); *Aniridia* in humans (Glaser et al., 1992; Hanson et al., 1994; Jordan et al., 1992), and *eyeless* in *Drosophila* (Quiring et al., 1994). *Pax-6*, or *eyeless*, has been shown to be a master control gene for eye development in *Drosophila* (Halder et al., 1995).

The basic function of Pax proteins as regulators of transcription depends upon the intact function of the paired domain (Chalepakis et al., 1991; Treisman et al., 1991). Several mutant alleles of *Pax* genes which cause loss-of-function phenotypes encode a PD with missense mutations that affect DNA binding (for review see Xu et al., 1995). Pax proteins have also been shown to have oncogenic potential and the transforming ability of Pax-3 is dependent on the DNA binding function of the PD (Chalepakis et al., 1993; Maulbecker and Gruss, 1993).

Recent structural data (Xu et al., 1995) showed that the PD folds as two subdomains referred to hereafter as PAI (N-terminal) and RED (C-terminal) domains, respectively (PAI + RED = PD). Each subdomain contains a helix-turn-helix (HTH) motif and has the potential to bind to DNA. However, in the crystal structure of the PrdPD, only the PAI domain made contacts to the 15 bp binding site, which corresponds to the region footprinted in vitro with the whole PD (Treisman et al., 1991). Other PDs (Pax-1, Pax-5 and Pax-6) protect 24-28 bp (Chalepakis et al., 1991; Czerny et al., 1993; Epstein et al., 1994), and their RED domains (Pax-5 and Pax-6) have been shown to contribute to the overall binding of the PD, thus illustrating the bipartite nature of the PD.

Previous experiments studying the specific recognition of DNA by a variety of PDs (Czerny et al., 1993; Epstein et al.,

1994; Xu et al., 1995; this work) indicate that different classes of PDs recognize similar DNA sequences. This is consistent with the similar DNA binding specificities of other conserved DNA binding domains and leads to the general paradox: how do structurally similar domains, which recognize similar DNA sequences, achieve functional diversity in order to execute vastly different developmental programs? The fact that the PD is a bipartite DNA binding domain, and that it is often associated with a HD suggests that Pax proteins may use different combinations of DNA binding domains to achieve a variety of functions. Here, we show Pax proteins can use the two sub-domains of the PD in different ways to recognize distinct sets of specific sequences. We also show that the PD and the HD can act cooperatively to bind to DNA and that Pax proteins that also contain a HD may use combinations of the three modular HTH DNA binding motifs to expand their recognition repertoire.

MATERIALS AND METHODS

Cloning and protein preparation

The paired boxes of *paired* and *pax-8* were cloned by inserting PCR amplified DNA fragments into the *Eco*RI and *Bam*HI sites of pGEX2T (Pharmacia), into the *Nde*I and *Eco*RI sites of pAR3038, and into the *Nde*I and *Bam*HI sites of pET14b (Novagen). PCR primers were used to introduce appropriate restriction sites. 5′ PCR primers included the starting Met codon, and 3′ PCR primers included a stop codon. Similarly, a *prd* gene truncated after the homeobox was inserted into the *Nde*I and *Eco*RI sites of pET14b. Missense mutations which mutated His47 to Asn47 in PrdPD (PrdPDH47N) and the reciprocal mutation in Pax-6PD (Pax-6PDN47H) were made using the megaprimer method and were cloned into pET14b using standard methods.

GST fusion proteins were overexpressed by inducing with 0.4 mM IPTG (Sigma) in SF8 cells. Cells were lyzed in Y buffer (1× PBS, 1 mM EDTA, 1% Triton X-100, 1 mM PMSF, 0.1 mM benzamidine) and fusion proteins were purified using glutathione agarose beads. The protein slurry was stored at 4°C. Paired domain peptides were overexpressed by inducing with 0.4 mM IPTG in BL21 cells and extracts were prepared (Treisman et al., 1989) and stored at −80°C. HisTag purified peptides were prepared as specified in the pET manual (Novagen) using Chelating Sepharose Fast Flow beads (Pharmacia) and stored at −80°C. Purified PrdPD and Pax-6PD were generous gifts from Wenqing Xu and Carl Pabo's lab.

Random selection protocol

A library of random sequence oligos was synthesized and used as a template for primer extension with a purified 3′ primer to make a double-stranded library. In the first round of selection, 100 ng of the ds oligo library was mixed with 10 μl of an optimized concentration of GST-PD fusion protein attached to agarose beads at 4°C for an hour in 500 μl of SELEX buffer (20 mM Tris pH7.5, 10% glycerol, 100 mM KCl, 0.5 mM EDTA, 1 mg/ml BSA) and 1 mM DTT and 2 μg/ml poly dIdC. The mixture was centrifuged and the pellet was washed twice with 1 ml of SELEX buffer. The pellet was resuspended in 30 μl of water and boiled for 3 minutes and centrifuged. In the subsequent PCR reaction, 5 μl of the supernatant was used as a template with 45 μl of PCR buffer (20 mM Tris pH 7.5, 50 mM KCl, 2 mM MgCl$_2$, 0.2 mM dNTPs, 1 unit Taq polymerase) and 500 ng of each primer for 20 cycles of amplification of 94°C (60 seconds), 44°C (30 seconds), 72°C (30 seconds), followed by the addition of 1.2 μg of each primer and a final round of PCR 94°C (1 minute), 44°C (1 minute), 72°C (10 minutes), to double strand the oligos. In each of

the following rounds of selection, 10 μl of the previous round's PCR reaction were mixed with 10 μl of the protein slurry using the same conditions, followed by a PCR amplification of the purified pool. After the final round of selection, a fraction of the final PCR reaction was digested with the appropriate restriction enzyme. Individual oligos were cloned into pKSII and sequenced with the T7 primer using the Sanger dideoxy method. The sequences were aligned and a consensus binding sequence was derived from the alignment.

```
5′ R18B  5′ CGAGAATTCGCTGAGCTATG 3′
3′ R18B  5′ CGAGAATTCGCTGAGCTATG 3′

5′ R18C  5′ GCTGAATTCTAGGCATGG 3′
3′ R18C  5′ AGAGAATTCTCCATGCCT 3′

5′ R18D  5′ CGTGAATTCTAGACATGCAG 3′
3′ R18D  5′ GATGAATTCGGGTATGCAAG 3′

5′ R18E  5′ CGTGAATTCCCGATACGTTC 3′
3′ R18E  5′ GATGAATTCATCTGTACTGC 3′

5′ R30A  5′ CTAGGATCCGATACTCTC 3′
3′ R30A  5′ GCAGGATCCACTGCACTT 3′

5′ R30B  5′ CATGGATCCGCATTGATC 3′
3′ R30B  5′ GCGGGATCCCATTGCATC 3′

5′ R30C  5′ GTAGGATCCATACGAGTT 3′
3′ R30C  5′ GTAGGATCCTGTTGGTA 3′
```

```
R30A
5′ CTAGGATCCGATACTCTCAGA N30 CACAAGTGCAGTGGATCCTCG 3′
R30B
5′ CATGGATCCGCATTGATC N10 TCACGC N20 GATGCAATGGGATCCCGC 3′
R30C
5′ GTAGGATCCATACGAGTTCCG N10 GCCAC N20 GATGCAATGGGATCCCGC 3′

R18B
5′ CGAGAATTCGCTGAGCTATGTC N18 GTGCCATGTCACTGAATTCCGA 3′
R18C
5′ GCTGAATTCTAGGCATGGAGAG N18 GCGTAGGCATGGAGAATTCTCT 3′
R18D
5′ CGTGAATTCTAGACATGCAGTATG N18 TGCACTTGCATACCCGAATTCATC 3′
R18E
5′ CGTGAATTCCCGATACGTTCCAGT N18 GTATGCAGTACAGATGAATTCATC 3′
```

Gel mobility shift assay

All oligos used in the gel shifts were gel purified and designed to contain 5′ GATC overhangs. The probes were labeled with Klenow and [α-^{32}P]dATP. In 20 μl, the protein was diluted with GS buffer (15mM Tris pH 7.5, 6.5% glycerol, 90 mM KCl, 0.7 mM EDTA, 0.2 mM DTT, 0.5 mg/ml BSA, 50 ng/μl poly dIdC, 0.5% NP40) and mixed with 100 pg of labeled probe for 20 minutes at RT. The mixture was loaded onto an 8% non-denaturing acrylamide (29:1, poly-acrylamide : bis-acrylamide) gel buffered in 0.25× TBE and electrophoresed at 12-15 V for 2-2.5 hours. The gel was exposed to a PhosphorImager (Molecular Dynamics) screen and/or film. Quantifications of the gel shift bands were made using the ImageQuant (Molecular Dynamics) program.

The following oligos (PH sites) were used to study the binding of the two domains, the paired domain and the homeodomain.

```
T PHO  5′ GATC TCTTC CAATTA GTCACGCTT GAGTG 3′
B PHO  5′ GATC CACTC AAGCGTGAC TAATTG GAAGA 3′

T PH2  5′ GATC CTTC CAATTA GA GTCACGCTT GAGT 3′
B PH2  5′ GATC ACTC AAGCGTGAC TC TAATTG GAAG 3′

T HP2  5′ GATC CTTC TAATTG GA GTCACGCTT GAGT 3′
B HP2  5′ GATC ACTC AAGCGTGAC TC CAATTA GAAG 3′

T PHX  5′ GATC TCTTC GAGCTA GTCACGCTT GAGTG 3′
B PHX  5′ GATC CACTC AAGCGTGAC TAGCTC GAAGA 3′

T PXH  5′ GATC TCTTC CAATTA CCTTGATCC GAGTG 3′
B PXH  5′ GATC CACTC GGATCAAGG TAATTG GAAGA 3′
```

The following oligos were labeled with polynucleotide kinase and [γ-^{32}P]ATP. The PrdWT sequence is the oligo used in the cocrystal-

lization experiment with the PrdPD. The other oligos are mutants of this oligo based on predictions from the crystal structure. The following oligos were gifts from Wenqing Xu and Carl Pabo.

```
T PrdWT      5′ AACGTCACGGTTGAC 3′
B PrdWT      5′ TTGTCAACCGTGACG 3′

T Mut1       5′ AACCTCACGGTTGAC 3′
B Mut1       5′ TTGTCAACCGTGAGG 3′

T Mut2       5′ AACGTCACAATTGAC 3′
B Mut2       5′ TTGTCAATTGTGACG 3′

T Mut3       5′ AACCTCACAATTGAC 3′
B Mut3       5′ TTGTCAATTGTGAGG 3′

T Mut4       5′ AAGCAGTGCCAACTG 3′
B Mut4       5′ TTCAGTTGGCACTGC 3′

T Mut5       5′ AACGTCACGATTGAC 3′
B Mut5       5′ TTGTCAATCGTGACG 3′

T PrdL       5′ AACGTCACGGTTGAC TGCATCGAGG 3′
B PrdL       5′ TTCCTCGATGCAGTC AACCGTGACG 3′

T PrdLF      5′ AACGTCACGGTTGAC ACGTAGCTCC 3′
B PrdLF      5′ TTGGAGCTACGTGTC AACCGTGACG 3′
```

The following oligos were obtained and purified from Operon, Inc.

```
T PAX6WT     5′ AACTTCACGGTTGAC 3′
B PAX6WT     5′ TTGTCAACCGTGAAG 3′
T PAX6L      5′ AACTTCACGGTTGAC TGCATCGAGG 3′
B PAX6L      5′ TTCCTCGATGCA GTCAACCGTGAAG 3′
```

The following oligos were used as binding sites to test the tethering requirement for PrdPD and PrdHD binding. The oligos were obtained and purified by Operon.

Quantification using the PhosphoImager program

For each set of protein dilutions bound to one probe, a rectangle was drawn around one representative band. Volume measurements were made at three positions within each lane. The first measurement, the background measurement, was made in the region between the free probe band and the DNA:protein complex band. The second measurement was taken of the free probe band, unbound measurement, and the third measurement was taken of the DNA:protein complex, bound measurement. The background measurement, was subtracted from both the unbound and bound measurements and these were respectively labeled unshifted and shifted. The ratios of the relative Kd of each concentration of each protein bound to two different probes were calculated as shown in the sample calculation.

Comparison of relative Kd for each concentration of PrdPD:

apparent $Kd = [P][D]/[PD]$ or $Kd(app) = [P]$ (unshifted/shifted) because $[P] >>> [D]$.

Ratio of $(1/Kd(X_1))/(1/Kd(X_2))$ at $[P]_N$

$$= \frac{1/([P]_{N,X1}(U/S)_{N,X1})}{1/([P]_{N,X2}(U/S)_{N,X2})} = \frac{([P]_{N,X2}(U/S)_{N,X2})}{([P]_{N,X1}(U/S)_{N,X1})} .$$

As $[P]_{N,X2} = [P]_{N,X1}$, then the ratio of apparent Kd for two probes at the same concentration of protein

$$= \frac{(U/S)_{N,X2}}{(U/S)_{N,X1}} .$$

Ratios of apparent Kd for two probes at the same concentration of protein were calculated for each set of serial dilutions of the protein as shown in a single gel shift experiment. The standard deviation was calculated based on points derived from a single experiment. Multiple trials of each experiments yielded similar values and all trends were observed as reported for each trial.

Measuring cooperativity of individual domains

$Kp = \{PD:DNA\}/\{PD\}\{DNA\}$
$Kh = \{PD:HD:DNA\}/\{PD\}\{HD:DNA\}$
$Kh/Kp = (\{PD:HD:DNA\}/\{HD:DNA\})/(\{PD:DNA\}/\{FP\})$
The two domains were said to be cooperative when the ratio of $Kh/Kp(PH0) > Kh/Kp(PH2)$.

RESULTS

'Optimal' binding sequences for paired domains

High affinity 'optimal' binding sites were selected from a library of oligos containing eighteen consecutive random positions using a modified version of the SELEX assay (Wilson et al., 1993). Fusion proteins of glutathione S-transferase (GST) and the paired domain (PD) of Paired (Prd) or Pax-8 were used to select oligos containing specific DNA sequences. After the last round of selection, individual oligos were cloned and sequenced. The sequences were aligned and the resulting consensus sequence represents the optimal binding sequence. Although biologically relevant DNA binding sites do not need to be optimal binding sites, these sites are useful tools because all potential protein-DNA interactions are emphasized.

The PDs can be placed into classes based on the comparison of their amino acid sequences, their overall genomic structure and the presence of other conserved domains in the molecule, such as a HD and an octapeptide (Noll, 1993). After nine rounds of selection, the PrdPD, which belongs to the same class as Pax-3, selected an optimal binding sequence of 14 bp (Fig. 1A, and Xu et al., 1995). The Pax-8PD, which belongs to a different class of PD, selected a highly related 16 bp binding sequence (Fig. 1B) after ten rounds of selection. Optimal binding sequences for the Pax-2 and Pax-5 PDs, which belong to the same class as Pax8PD, and for the Pax-6PD, which belongs to a third class, have been identified previously (Czerny et al., 1993; Epstein et al., 1994). The consensus sequences for all of these PDs are essentially identical (Fig. 1C), although the Pax-2/-5/-6 sequences are 5-6 bp longer. The sequence similarity among these selected binding sites confirms the prediction from the crystal structure that the different classes of PDs fold and bind to DNA in a very similar manner: In addition, this binding site, recognized by a monomer, is unexpectedly long, as compared to other DNA binding domains whose monomers typically recognize 3-6 bp binding sites.

Determinants of paired domain DNA binding specificity

The crystal structure has shown that all but one of the PrdPD amino acids that contact the DNA are conserved among all PDs (Xu et al., 1995), which explains the shared core recognition (Fig. 1C). Several of these residues make specific contacts to bases in the core motif in both the major and minor grooves, and mutant alleles of Pax proteins that contain missense mutations at these critical residues result in a loss of function phenotype. To further test the role of these contacts in the high affinity binding of the PrdPD to DNA, we introduced a series of mutations at critical nucleotides within the core motif of the binding site. Using a gel shift assay, we showed (Fig. 2A) that the binding affinity of the PrdPD is reduced by 65- or 75-fold respectively when either the first position of the core (contacted in the major groove by H47 of the recognition helix; Mut1) or
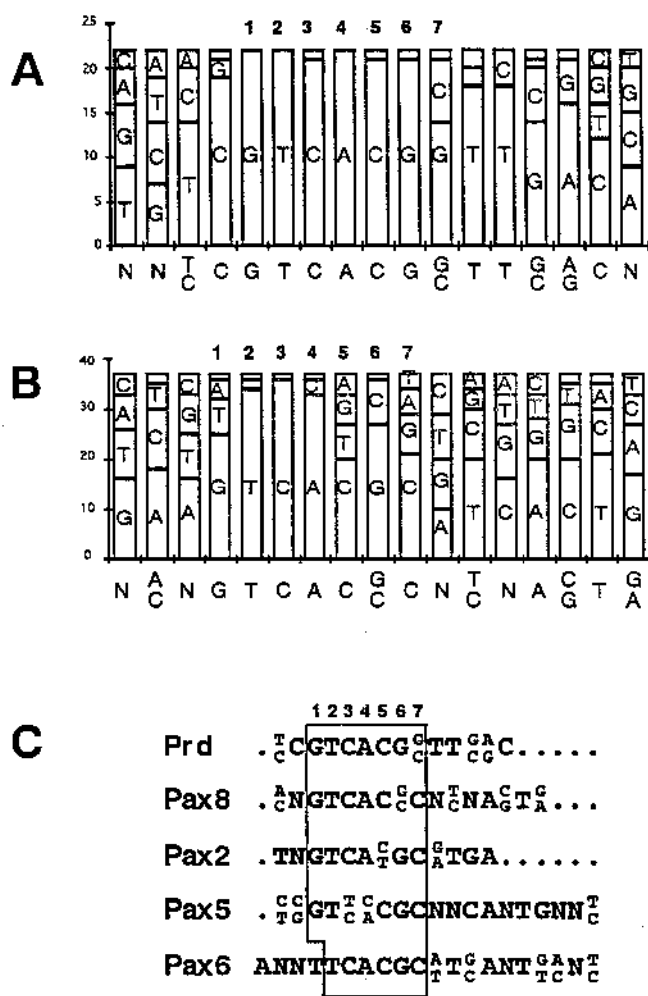
Fig. 1. Optimal PD binding sites. The alignment of the individual cloned sequences after the last round of selection are shown in a graph, a 'consensogram' (Wilson et al. 1993), where the height of each bar represents the number of occurrences of each base at each position along the consensus sequence, as written at the bottom, for (A) the PrdPD and (B) the Pax-8PD. (C) A comparison of the known consensus or optimal binding sites for various PDs. References for the selected binding sites: PrdPD and Pax-8PD, see above; Pax-2PD and Pax-6PD; (Epstein, et al., 1994), Pax-5 (Czerny, et al., 1994). The core motif is boxed. Position 1 of the core motif of the Pax-6PD binding site is a T instead of a G.

the seventh position of the core (contacted in the minor groove by N14 of the β-turn, Mut5) are mutated. The binding affinity to a site in which positions 6 and 7 (both contacted by N14) were mutated (Mut2) was decreased by an additional 10-fold over Mut5. PrdPD did not bind to a site that contains mutations at all three positions (1, 6, 7; Mut3) or to an unrelated oligo (Mut4). Thus, the conserved protein-DNA contacts to the core motif underlie the general mechanism of DNA sequence recognition by the PD and demonstrate the importance of the major and minor groove contacts to its overall binding potential.

Role of the RED domain

The long binding site for PrdPD is contacted exclusively by a monomer of the PAI domain through a set of conserved residues

(Xu et al., 1995). Pax-5 and Pax-6 recognize even longer sequences suggesting that the PAI and RED domains both contact DNA, and thus contribute to DNA binding, confirming the previous observation that they recognize specific DNA sequences (Czerny et al., 1993; Epstein et al., 1994). We therefore tested whether the PrdPD could use its RED domain to bind to an extended DNA binding site. However, a more stringent selection with the PrdPD failed to show any preference for DNA sequences flanking the 14 bp identified in the original selection experiments. We performed selection experiments with libraries of oligos containing 18 or 30 consecutive random positions and only identified the 14 bp site described above. We also used a library containing a core PD binding site (N10-TCACGC-N20), and another library with a core site modified at position 2 (N10-GCCAC-N20), a change which has been reported to induce RED domain binding to 3′ sequences (Czerny et al., 1993). These selections again, identified only the 14 bp site, including the fixed core sequence, or identified a new complete consensus sequence in the 20 randomized positions (data not shown). Thus, the RED domain of Prd, unlike that of Pax-5 and Pax-6, may only weakly affect the overall DNA binding potential of the entire PD.
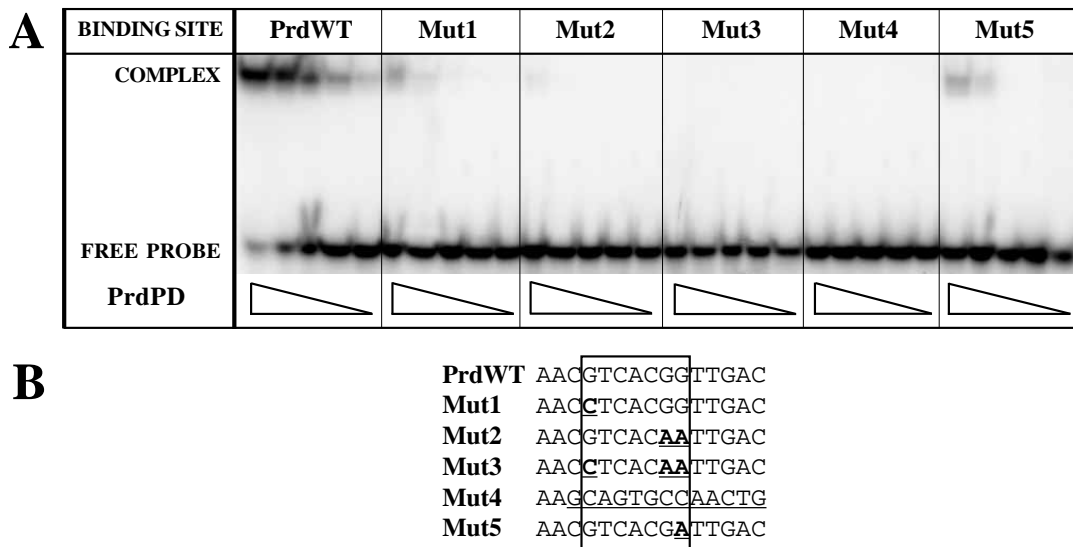
To examine the role of the RED domain in PD DNA binding more precisely, three binding sites, each 25 bp, were designed (Fig. 2D). PrdL contains the 15 bp binding site used in the crystallization study, followed, at the 3′ end, by the 10 bp flanking the core sequence in the selected Pax-6PD site. These 10 bp represent a potential binding site for the RED domain. Pax-6L is identical to PrdL except for a T→G substitution of the first position of the core motif. PrdLF contains the 15 bp PrdPD binding site, but with the following 10 bp randomized by reversing their sequence.

The PrdPD preferred PrdL over PrdLF by only 2.7-fold (Fig. 2A) (see Materials and Methods for quantification). It bound equally well to PrdLF, a 25mer, as to PrdWT, a 15mer used in the crystallization experiments (data not shown), suggesting that nonspecific DNA sequences 3′ to the core do not provide a docking platform for the RED domain of Prd. It also confirms that Prd binds primarily through the PAI domain. Although the RED domain may recognize specific DNA sequences 3′ to the selected binding sequence, it does not play a major role in the overall binding of the PrdPD.

Pax-6PD bound well to Pax-6L but bound neither to PrdLF (Fig. 2B) nor to PrdWT (15mer), nor to a 15mer PAX6 site identical to PrdWT with a T→G substitution at the first position of the core motif (Data not shown). This confirms previous observations (Epstein et al., 1994) that Pax-6 has a strict requirement for DNA sequences 3′ to the PD core sequences. These are recognized by the RED domain, which has a large contribution to Pax-6 binding.

We also tested Pax-8, belonging to the same class of PDs as Pax-5 and Pax-2, all of which are not associated with a HD. As predicted by the presence of H47 (see below), the Pax-8PD bound better to PrdL than to PAX6L by 6-fold, and better to PrdL than to PrdLF by 80-fold. It did not bind to PrdWT (15mer) (Fig. 3C). This confirms the observation that the RED domain of Pax-5, and thus also that of Pax-2 and Pax-8, plays an important role in DNA binding (Czerny et al., 1993).

These experiments demonstrate that the PD is composed of two HTH DNA binding motifs, and that these two domains can interact to recognize specific DNA sequences. They confirm previous observations that, unlike Prd, Pax-6 and Pax-2/-5/-8

A

BINDING SITE	PrdWT	Mut1	Mut2	Mut3	Mut4	Mut5
COMPLEX						
FREE PROBE						
PrdPD						

B

PrdWT	AAC GTCACGG TTGAC
Mut1	AAC **C**TCACGG TTGAC
Mut2	AAC GTCAC**AA** TTGAC
Mut3	AAC **C**TCAC**AA** TTGAC
Mut4	AA**G**CAGTGCC AACTG
Mut5	AAC GTCACG**A** TTGAC

Fig. 2. Specific protein-DNA interactions of the PD. DNA determinants. (A) EMSA of the PrdPD peptide with a series of probes. Decreasing concentrations of a four-fold serial dilution series (starting concentration = 2.9×10^{-6} M) of the purified PrdPD peptide bound to wild-type or mutated 15mer binding sites. (B) PrdWT is the binding site used in the crystallization experiments. Mutations (Mut1-5) were introduced into PrdWT at positions contacted by the PAI domain in the major (Mut1) or minor (Mut2 and Mut5) grooves, or both (Mut3). Mut4 is the reverse sequence of PrdWT and is not a binding site for the PrdPD.

rely on both the PAI and the RED domains for the overall sequence-specific DNA binding of the PD. Mutations that affect binding by the PAI domain (see below) do not affect the binding properties of the RED domain, suggesting that the two subdomains may be relatively independent.

DNA binding discrimination among PDs

A comparison of the optimal binding sequences for different classes of PDs (Fig. 1C) reveals a shared core motif, **G**TCACGC/G, for all PD binding sites. One notable exception is at the first position in the core which is a **T** for the Pax-6PD binding site which has the core motif, **T**TCACGC/G (Fig. 1C). The only residue that contacts the DNA and is not conserved is residue 47, which is H47 in all PDs except in the Pax-6PD which has an N47. In the crystal structure, H47 contacts the first position of the core in the major groove, and the identity of residue 47 can be correlated with the identity of position 1 in the core motif, G (H47) or T (N47). We predicted that the substitution of an N47 for H47 in Prd would change the preferred binding specificity of the PrdPD. We showed that a PrdPD H47N mutant preferred to bind to Pax-6L over PrdL, a long version of the Prd site which differs only at position 1 of the core (Fig. 3A). Similarly, the reciprocal mutation (N47H) in Pax-6PD, changed its preferred binding specificity to that of PrdPD (Fig. 3B).

PrdPD preferred PrdL over Pax-6L by 4.1-fold while the H47N mutant preferred Pax-6L over PrdL by 5.6-fold, showing a 23-fold difference in binding affinity overall. Pax-6PD preferred Pax-6L over PrdL by 2.7-fold while the reciprocal N47H mutant preferred PrdL over Pax-6L by 2.8-fold, showing an overall affinity difference of 7.5-fold (see Materials and Methods for quantification). Therefore, in the context of either Prd or Pax-6, residue 47 determines the preferred contact to position 1 of the core motif. The reason for the quantitative difference in the ability of each PD to discriminate between the two sites (23-fold in Prd and 7.5-fold in Pax-6) may depend on the influence of neighboring amino acids in the PD, or from contributions from the RED domain. A similar experiment performed with Pax-5 has identified three residues (including H47) which are important for discriminating between Pax-5 and Pax-6 binding sites (Czerny and Busslinger, 1995).

Interactions between paired domain and homeodomain

Many Pax proteins also contain a HD, a third HTH DNA binding domain. The HD found in these Pax proteins always belongs to the Prd class. It bears a specific S50 that is conserved among the Pax proteins. This conservation, along with that of PDs associated with HDs, suggests that the PD and the HD can interact. Unlike the PAI and RED domains which are always present together (see exceptions in the Discussion), the PD and HD are distinct and apparently independent DNA binding domains. Prd can activate transcription of a reporter gene in a transient transfection assay by binding to specific DNA sequences through either the PD or the HD. A full-length Prd protein which cannot bind through its HD, is still able to activate transcription through binding to PD sites, but not through HD sites. Similarly, a Prd mutant protein (G15S) whose PD does not bind DNA (Treisman et al., 1991) can activate transcription through binding to HD sites, but is no longer able to activate transcription through PD sites (G. Sheng and C. D., unpublished data). Yet, neither mutant protein can rescue *prd* function in vivo when placed under the control of the *prd* promoter, even when the two constructs are present in the same embryo, suggesting that both domains are required in the same molecule (Bertuccioli et al., 1996). This suggests that the PD and the HD must interact cooperatively to activate their targets and is consistent with in vitro data showing cooperative binding of PD and HD on some sites (Treisman et al., 1991).
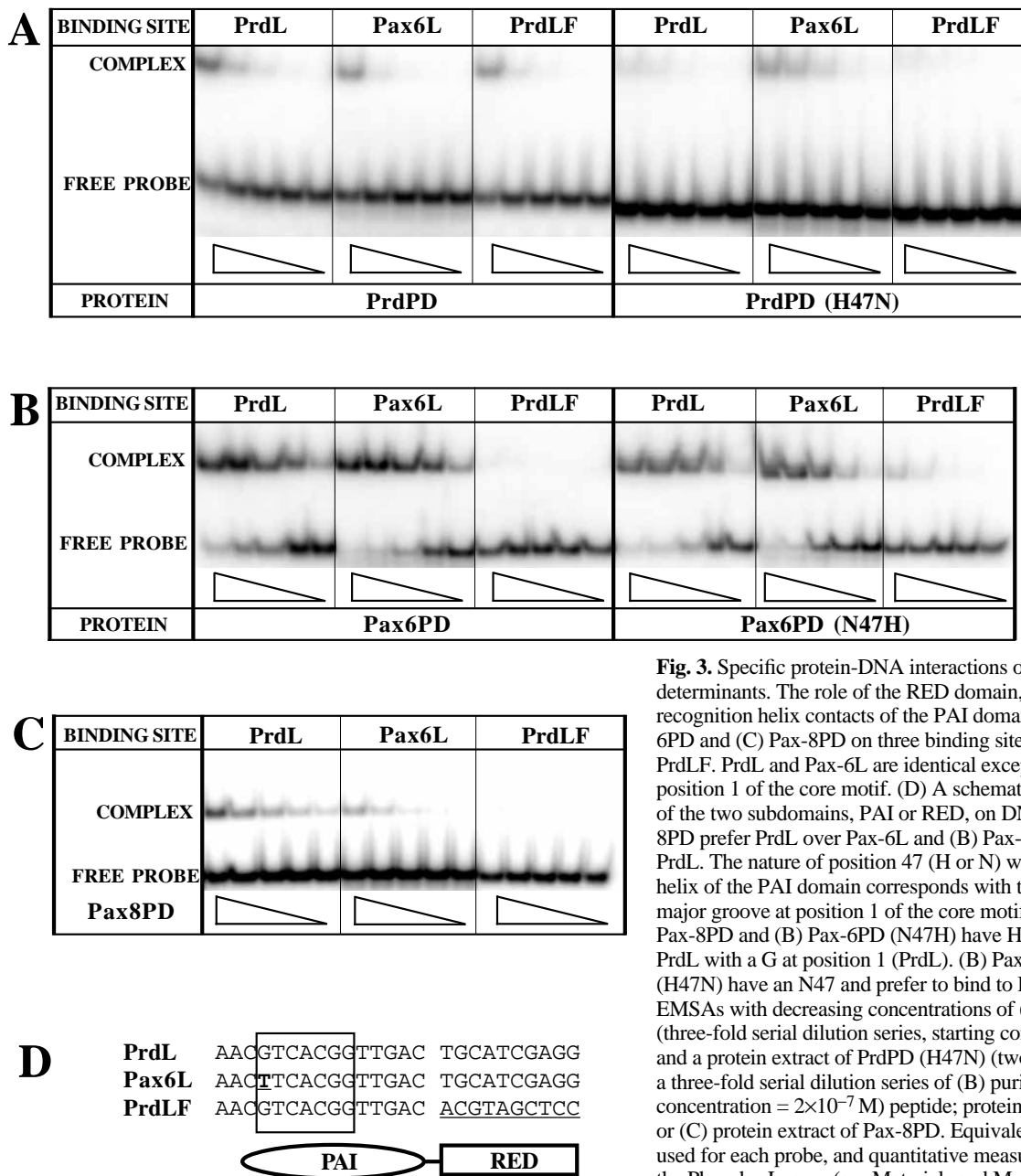
Fig. 3. Specific protein-DNA interactions of the PD. Protein determinants. The role of the RED domain, and the role of the recognition helix contacts of the PAI domain of (A) PrdPD, (B) Pax-6PD and (C) Pax-8PD on three binding sites (D) PrdL, Pax-6L and PrdLF. PrdL and Pax-6L are identical except for a single base pair at position 1 of the core motif. (D) A schematic illustrates the arrangement of the two subdomains, PAI or RED, on DNA. (A) PrdPD and (C) Pax-8PD prefer PrdL over Pax-6L and (B) Pax-6PD prefers Pax-6L over PrdL. The nature of position 47 (H or N) which lies in the recognition helix of the PAI domain corresponds with the base pair contacted in the major groove at position 1 of the core motif (G or T). (A) PrdPD, (C) Pax-8PD and (B) Pax-6PD (N47H) have H47 and prefer to bind to PrdL with a G at position 1 (PrdL). (B) Pax-6PD and (A) PrdPD (H47N) have an N47 and prefer to bind to Pax-6L with T at position 1. EMSAs with decreasing concentrations of (A) purified PrdPD peptide (three-fold serial dilution series, starting concentration = 7.1×10^{-7} M), and a protein extract of PrdPD (H47N) (two-fold serial dilution series), a three-fold serial dilution series of (B) purified Pax-6PD (starting concentration = 2×10^{-7} M) peptide; protein extract of Pax-6PD (N47H), or (C) protein extract of Pax-8PD. Equivalent amounts of protein were used for each probe, and quantitative measurements were made with the PhosphorImager (see Materials and Methods).

To identify a binding site on which the PD and HD could interact to recognize specific DNA sequences, a SELEX assay was performed using a Prd peptide spanning both PD and HD fused to GST. After nine rounds of selection, several classes of binding sites were selected. One class contained the binding sequence for the PrdPD. This suggests that the PD is not influenced by the HD. Another class contained the previously identified palindromic HD binding site, P2 (TAAT NN ATTA) (Wilson et al., 1993), confirming that the HD is able to cooperatively dimerize within the context of the Prd protein. A third class of selected sequences contained a site with a PD binding site abutted to a HD binding site (0 bp spacing, PH0) with a specific orientation (Figs 4A, 6). A few oligos contained a palindromic HD binding site (P2) abutted to a PD binding site

with the same specific orientation as PH0, suggesting that the HD can interact with the PD and other HDs simultaneously.
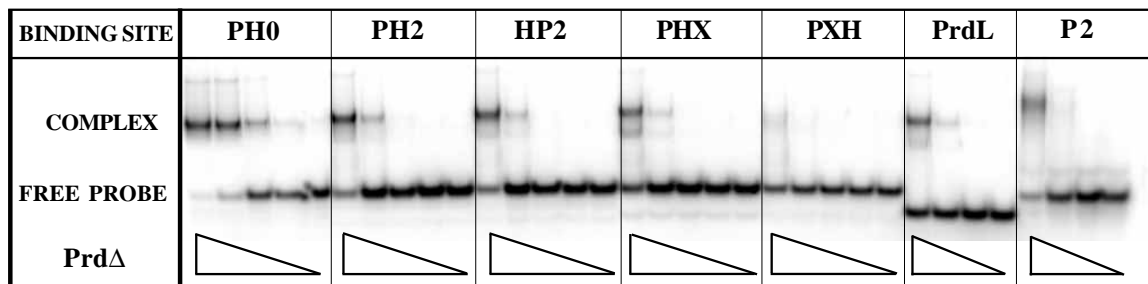
We compared the affinity of a Prd protein containing both the PD and HD for the PH0 site and for sites in which the spacing was altered, or in which one of the individual sites was mutated. We showed that the PD and the HD are able to cooperatively bind to the PH0 site and that the orientation and spacing of the individual sites are critical for the cooperativity (Fig. 4B). When the individual sites were spaced by 2 bp (PH2) instead of 0 bp (PH0), or when the orientation between the two sites was also inverted (HP2), the binding affinity decreased by 50-fold. Prd bound >100-fold better to the PH0 site than to the same oligo containing only one of the individual sites (PHX or PXH). It also preferred, by 20-fold, the PH0 site over the

A

```
          TGTGTTCACTGCTTCCATAATCAATTAGTCACGGTCTGAG
          ACTTGTGATCTCATCGGGGTTATCAATTAGTCACGCTCTGAG
                    GATCCAATCAATTAGTCACGCCATTGCTCTAATCTTGCCGGGATG
                    GATCCAATCAATTAGTCACGCTTTATATGAGCTGGCAGGGGGATG
                    GATCGCCCAATTAGTCACGCTATAGGAACTTACCCTGATGGGATG
                    GATCCACCAATTAGTCACGCAATGGCATTACGATTTGTAGGATGC
                    GATCGCGCAATTAGTCACGCATGTTAACCCTCACCCTTAGGATG
                    GATCGCGTAATTAGTCACGCTCCTTCGATTAACTGAACGGGATG
          CTTGTGGAGCCTAATCGGTTATTAATTAGTGCCGCCTCTGAG
                    GATCTGCGAATTAGCCACGCATTAACCTGATTATCTTTGTGATG
           TTGTGAAAGCTGAATTCATAATCGATTAGTCACGCTCTGAG
           TGTGCTAATCTAATCACAAATCGATTAGTCTCGGTCTGAG
          TCTCAGAAGACGACGGTAATTGCGATTAGTCACGGTGCACA
          CTTGTGGTTGCCTGCGCTTAATCGATTAGTTCCGCCTCTGAG
                    GATCCACTGATTAGTCACGCCCTCGCGATTATACCCACCAGATGC
```
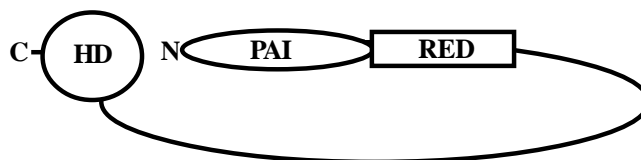
B

BINDING SITE	PH0	PH2	HP2	PHX	PXH	PrdL	P 2

COMPLEX

FREE PROBE

PrdΔ

```
PH0  GATCTCTTCCAATTA GTCACGCTTGAGTG
PH2  GATC CTTCCAATTAGAGTCACGCTTGAGT
HP2  GATC CTTCTAATTGGAGTCACGCTTGAGT
PHX  GATCTCTTCGAGCTA GTCACGCTTGAGTG
PXH  GATCTCTTCCAATTA CCTTGATCCGAGTG
```

C—(HD) N—(PAI   [RED])

Fig. 4. Site selection with Paired. (A) In a Prd protein containing the PD and HD, PrdΔ, the PD and HD can cooperatively bind to a specific sequence, PH0. Some of the sequences of individual clones of the fifth round of selection with R30B (N10 -TCACGC -N20) and the tenth round of selection with R30A (N30) are shown. The fixed core sequence of R30B oligo is shown in red, and PrdPD binding sites are blue. PrdPD binding sites in the R30A oligos are green. The palindromic HD binding motif (TAAT-NN-ATTA) is underlined. Fixed sequences in the flanking regions of the oligos are marked through with a line. The HD sequence of the PH sites is indicated in magenta. (B) Spacing and orientation of the half-sites are important for PD and HD cooperativity. Decreasing concentrations of a four-fold serial dilution series of PrdΔ, were bound to a series of PH binding sites. Equivalent amounts of protein were used for each probe. Quantitative measurements were made using the PhosphorImager. The PAI core motif is boxed and the individual PAI and HD sites are underlined. A schematic of the domains is aligned to the binding sites. The PH0 site contains each individual binding site abutted with a specific orientation. PH2 and HP2 have a spacing of two between the two individual sites. The relative orientation of the HD site and PD site are reversed in HP2. The PHX site and HPX site contain mutated HD and PD sites respectively. PrdL contains the PrdPD binding site, and P2 contains the palindromic HD binding site.

individual optimal PD site (PrdL) and HD site (P2, Fig. 4B) (see Materials and Methods for quantification). In addition, while the Prd protein bound to the P2 site as a dimer, a single Prd molecule bound to PH0 (Fig. 4B), consistent with the model that intramolecular cooperativity allows one Prd molecule to bind PH0 with high affinity. Similarly, only one molecule bound PHX and PXH sites, or sites in which cooperativity had been abolished (PH2 and HP2). This latter situation is likely due to tethering, as it is more likely for the second domain to bind than for a second molecule to enter the complex, even if no cooperativity exists between PD and HD.

### Tethered versus non-tethered interactions

By modeling the structure of the two domains as they bind to the PH0 site (Fig. 6), we observed that the recognition helices of each domain are on opposite sides of the DNA double helix.

The only possible contacts between PD and HD, which may explain the cooperative binding to the PH0 site, are between the second helix of the PAI domain and the N-terminal arm or extended conserved region of the HD. Alternatively, the observed cooperativity might be due to tethering similar to the POU-specific domain and POU HD (Klemm et al., 1994). It could also be due to changes in conformation of the protein or DNA that cannot be predicted from the model. It must be noted that, unlike the POU-specific domain and the POU HD, both the PAI domain and the HD are able to bind independently with high affinity to specific DNA sequences, and many proteins contain only a PD or a HD.

In order to test the importance of tethering, we used the indi-vidual PD and HD, which were produced separately (PrdPD; 128aa and PrdHD; 60aa; Fig. 5A). We mixed decreasing serial dilutions of the PD with two concentrations of the HD; a high concentration in which 50% of the PH0 and PH2 binding sites were shifted, and a low concentration in which neither site was shifted by the HD. Cooperativity was measured by comparing the supershift (i.e. the complex formed by binding of both the HD and the PD), obtained on the PH0 site and on the PH2 site (whose extra 2 bp disrupt cooperativity, see Fig. 4B). Cooperative DNA binding by the two domains was seen at both high and low concentrations of the HD, suggesting that the interaction occurs on the DNA and not in solution (Fig. 5A). The cooperativity observed (Fig. 5A) was weaker than when the

Fig. 5. Non-tethered PDs and HDs interact to bind cooperatively to DNA. EMSAs of individual PDs and HDs, produced separately and non-tethered, on PH or P6H sites. Decreasing three-fold serial dilutions of PDs were mixed with a constant concentration of HD. (A) Similar cooperative effects were observed at two concentrations of HD, a high concentration at which 50% of the probe was shifted, or a 50-fold lower concentration, at which the shift was barely detectable. (B) All peptide domains (PrdPD shown, others not shown) bind equally well to both probes. The protein mixture (PD and HD) was bound to PH sites with a spacing of 0 bp or 2 bp. A spacing of 2 bp disrupts cooperativity and this was measured by comparing binding to (F) PH0 over PH2 or P6H0 over P6H2. (A) Non-tethered PrdPD (starting concentration = 1.4×10^{-6} M) can interact with PrdHD (high = 2×10^{-7} M) to bind cooperatively to PH0, although the degree of cooperativity is about half of the tethered interaction (see Fig. 4). PDs and HDs from different molecules can interact to bind cooperatively to specific binding sites. (B) PrdPD (starting concentration = 1.7×10^{-6} M) can interact with EnHD (protein extract) to bind cooperatively to PH0 at a high concentration and a 50-fold lower concentration. (C) Pax-8PD (protein extract) or (D) Pax-6PD (starting concentration = 1.8×10^{-6} M) can interact with PrdHD to bind cooperatively to P6H0 over P6H2. The relative contribution of position 1 specificity is small as compared to that of the RED domain. (E) Pax-8PD (protein extract) can interact with EnHD to bind cooperatively. (F) Binding sites with individual HD and PD sites underlined. The schematic models the relative positions of the DNA binding domains. Note that although the HD used in these experiments contains a SQ50 (Wilson et al., 1995), this amino acid does not reside near the PD and is unlikely to play a direct role in potential protein-protein interactions (Fig. 6). This mutation also only weakly affects *prd* activity in a rescue assay in vivo (Bertuccioli et al., 1996).
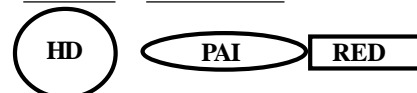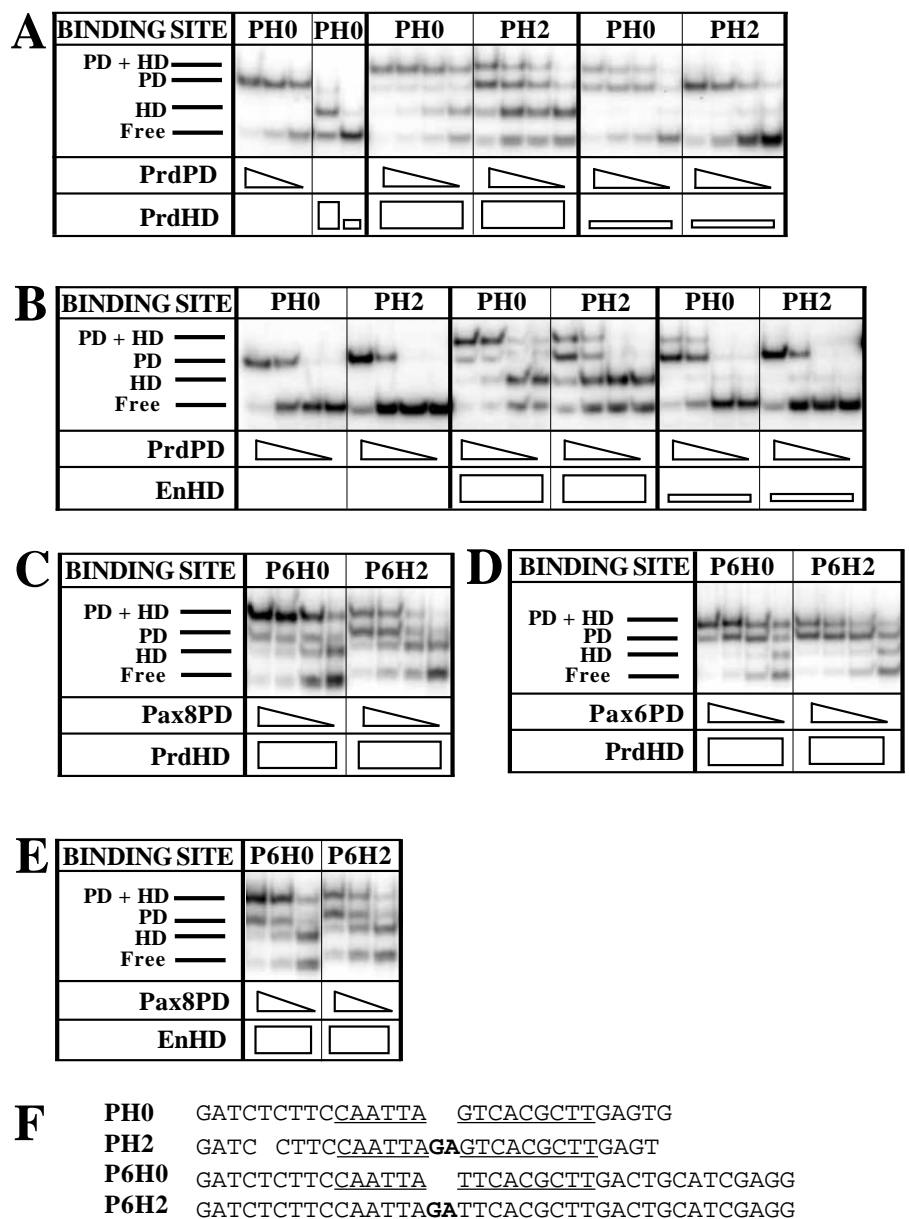
two domains were present on the same molecule (compare with Fig. 4B), and this difference indicates that tethering contributes about half of the cooperativity.

The PrdPD can also cooperatively interact with other HDs to bind to DNA. The Engrailed HD, which belongs to a different class than the PrdHD and is never associated with a PD, can also cooperatively bind with the PrdPD to the PH0 site, although more weakly than for the PrdHD (Fig. 5B). Similarly, the Pax-8PD, which is normally not associated with a HD, and Pax-6PD, which is associated with a HD, cooperatively bound with the PrdHD to a site containing the 25 bp Pax-6PD site and a TAAT HD site in the same orientation and spacing as in PH0 (P6H0 and P6H2; Fig. 5C,D). Finally, the Pax-8PD and the EnHD, both of which are found in proteins that are not associated with the other domain, are able to weakly interact to bind cooperatively (Fig. 5E) (see Materials and Methods for quantification). These experiments suggest that in addition to intramolecular interactions, the PD and the HD can potentially interact intermolecularly, which could be a mechanism to coordinate the regulation of specific promoters.

DISCUSSION

Sequence-specific DNA recognition by the PD

The PD is a bipartite DNA binding domain composed of two subdomains, the PAI and RED domains. Each contains a HTH motif that potentially can bind to DNA (Czerny et al., 1993; Epstein et al., 1994; Xu et al., 1995). The PAI domain contacts 13 bp, a sequence which is longer than previously defined binding sites for monomeric DNA binding domains. The residues of the PAI domain that contact the DNA, except for one, residue 47, are conserved among all PDs and as a consequence, all PDs bind to a stereotypical binding site. These contacts are required for high affinity sequence-specific binding by the PAI domain because mutations in either the protein or the DNA sequence significantly decrease binding affinity (see above and Xu et al., 1995). The single exception is residue 47 which determines the binding preference for either a G or a T at position 1 of the core motif. Reciprocal mutations at this position in Prd and Pax-6 lead to predicted switches in binding preference. However, these differences are small: reciprocal mutations at position 47 show a 23-fold discriminatory effect on binding affinity for Prd and only a 7-fold effect for Pax-6. How Pax proteins and other proteins with conserved DNA binding domains play different regulatory roles and have diverse functions while binding to functionally similar DNA binding sites remains a critical question. Why doesn't redundancy in recognition lead to redundancy in function?

Mechanisms involving intramolecular interactions between DNA binding domains

The POU domain (Klemm et al., 1994) is an example of a conserved DNA binding domain, the HD, associating with another DNA binding domain intramolecularly to recognize a different set of specific DNA sequences, leading to further functional specificity. Both domains contain a HTH motif that docks to DNA, yet both domains are required for high affinity recognition of individual binding sites for each domain separated by a specific spacing. In the structure of the complex (Klemm et al., 1994), the two domains do not contact each other physically, and the cooperativity appears to be mediated through the tethering of the two domains and through changes in DNA conformation (Klemm et al., 1994). Like the POU-specific domain and POU HD, both PAI and RED subdomains of the Pax-6 and Pax-2/-5/-8 PDs make major contributions towards the overall binding potential of the PD (Czerny et al., 1993; Epstein et al., 1994). However, in Prd, the PAI domain alone is able to bind to DNA with high affinity while the contribution of the RED domain is very weak. Furthermore, an in vivo rescue assay and an ectopic expression assay have both shown that the RED domain is dispensable for all *prd* functions (Bertuccioli et al., 1996; Cai et al., 1994).

Another mechanism for generating functional diversity is for multiple modular domains, such as the Zn finger domains, to interact intramolecularly (Rebar and Pabo, 1994). They interact with the DNA and with each other to modify their specific DNA recognition. Pax proteins contain multiple HTH motifs that are organized as modular DNA binding domains like the Zn fingers (Miller et al., 1985; Pavletich and Pabo, 1993). Multiple HTH domains may interact intramolecularly and 'mix-and-match' in different combinations to allow Pax proteins to recognize diverse sequences. By using these modular HTH domains in different combinations, the Pax proteins can bind in different modes and recognize a range of DNA sequences (Fig. 7).

However, although Zn finger proteins contain multiple modules, they achieve only a single binding specificity and thus bind by using the whole set of required fingers. In contrast, several different binding modes may coexist within the same Pax protein (Fig. 7). Each subdomain can bind alone, or may interact with another domain as a strategy for generating different specificities. In Prd, the PAI domain can bind independently to a 13 bp binding site. In Pax-6, an isoform contains an insertion in the recognition helix of the PAI domain which can no longer bind to DNA (Epstein et al., 1994). The PD is still able to bind to a distinct sequence through its RED domain. Alternatively, the PAI and RED domains of Pax-2/-5/-8 and Pax-6 both contribute significantly to DNA binding (Czerny et al., 1993, Epstein et al., 1994). In yet another mode, the PAI domain and the HD bind cooperatively to a specific site, PH0 (Fig. 7).

Based on this model, another possible mode may be represented by a new Pax protein, Lune, that contains only a RED domain (and no PAI domain), and a Prd-class HD (S.J and C.D. unpublished results). In Lune (and perhaps also in Pax-6 5a), the RED domain and the HD are predicted to interact to recognize specific DNA sequences (Fig. 7). In this way, the Pax proteins have found an innovative solution to the problem of generating functional diversity with conserved DNA binding domains.

Mechanisms involving intermolecular interactions between DNA binding domains

Another strategy used to generate functional diversity among proteins with similar DNA binding specificities is to interact with other DNA binding proteins. For example, the bHLH (Voronova and Baltimore, 1990), bZIP (Ellenberger et al., 1992), and nuclear receptor proteins (Luisi et al., 1991) use homo- and hetero-dimerization to modify their ability to recognize DNA sequences. The Prd-class HDs, including those

paired domain   **Homeodomain**   **PRD repeat**
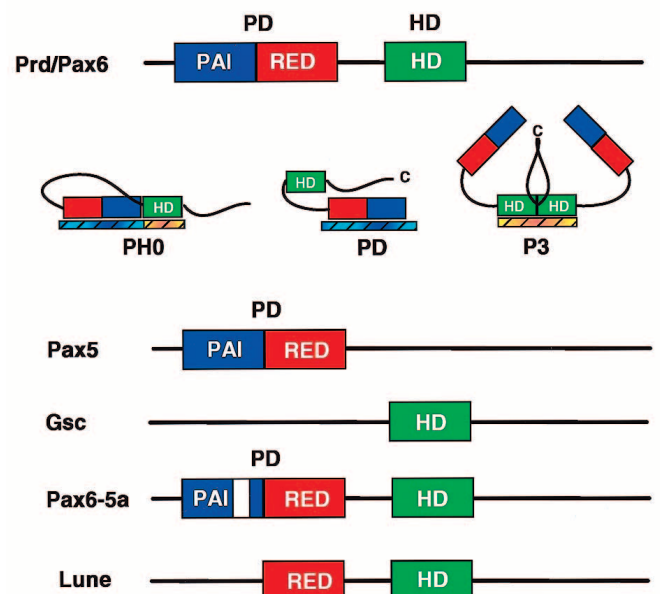
PAI   RED
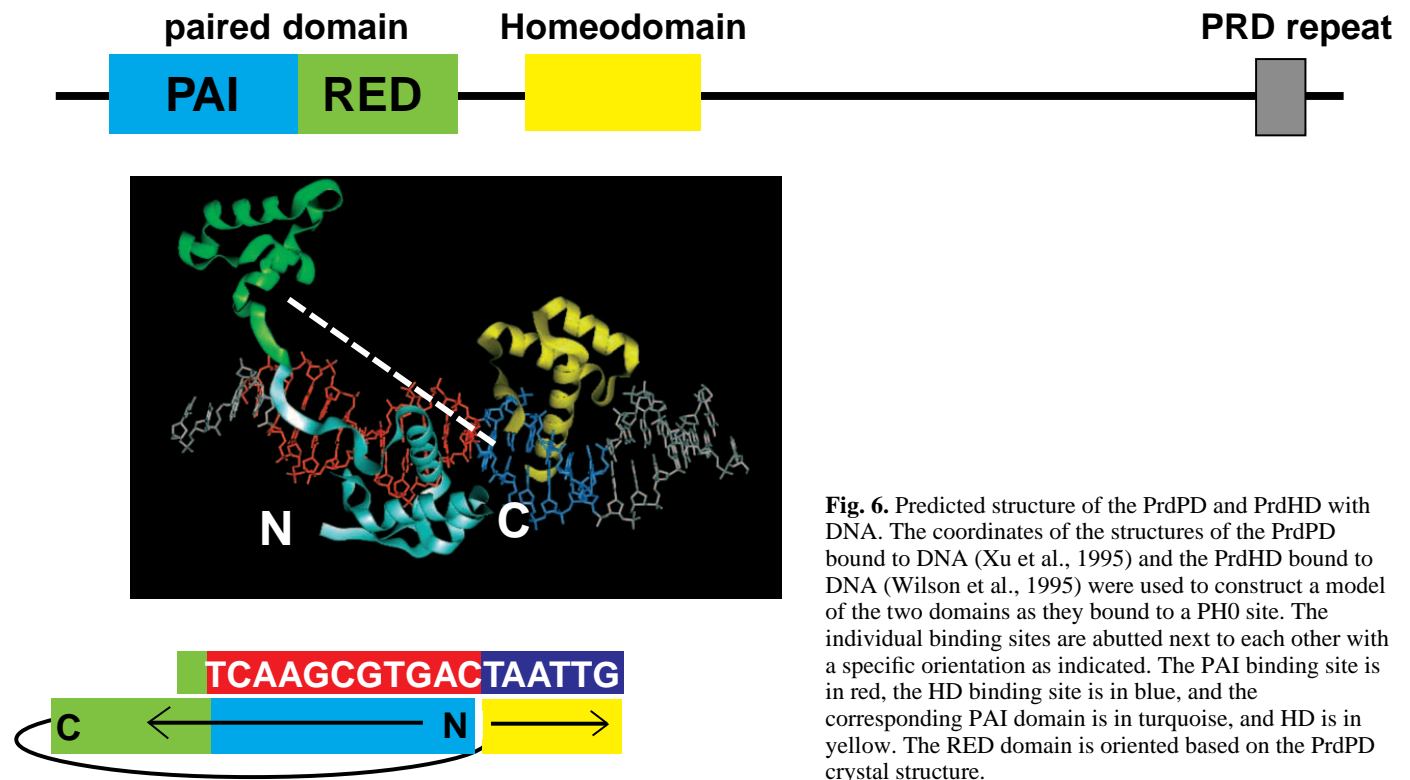
Fig. 6. Predicted structure of the PrdPD and PrdHD with DNA. The coordinates of the structures of the PrdPD bound to DNA (Xu et al., 1995) and the PrdHD bound to DNA (Wilson et al., 1995) were used to construct a model of the two domains as they bound to a PH0 site. The individual binding sites are abutted next to each other with a specific orientation as indicated. The PAI binding site is in red, the HD binding site is in blue, and the corresponding PAI domain is in turquoise, and HD is in yellow. The RED domain is oriented based on the PrdPD crystal structure.

TCAAGCGTGACTAATTG

present in Pax proteins, can form cooperatively homo- or hetero-dimers (Wilson et al., 1993, 1995; see Gsc Fig. 7). A variation of this strategy is to associate with cofactors. In *Drosophila*, Ultrabithorax and other HOX proteins interact with the cofactor Extradenticle to allow them to have different regulatory roles on different promoters (Chan et al., 1994; van Dijk and Murre, 1994; Chang et al., 1994).

We have shown that the PD and the HD do not need to be present in the same molecule in order to interact to bind cooperatively to DNA. We have also shown that this cooperativity can occur between PDs and HDs of different classes and that PDs from molecules that are not associated with HDs can cooperatively bind to DNA and interact with the HD of other Pax proteins or other HD proteins. However, at least in the context of Prd, intermolecular interactions may not be sufficient to sustain biological function. This still suggests a

possible mechanism of regulation in which PD proteins and HD proteins could interact intermolecularly to recognize specific DNA sequences and regulate transcription. It must be noted, however, that an in vivo rescue assay has shown that the Prd protein requires both a functional PAI domain and HD in the same molecule to function and activate its target genes (Bertuccioli et al., 1996). In an ectopic assay in vivo, however, some transcomplementation between a PD mutant and a HD mutant protein could be observed (Miskiewicz et al., 1996).

Fig. 7. Models of the different modes of DNA binding by the Pax proteins. Pax proteins that contain both a PD and a HD, have three distinct HTH domains which bind to DNA. Using different combinations of their HTH domains, Pax proteins can recognize distinct types of binding sites. The PAI and RED domains of Pax-6 interact to recognize PD sites (Epstein et al., 1994). The PAI and HD of Prd interact to recognize PH0 sites (see above), and HDs of two Prd-class proteins can dimerize to recognize P3 sites. Other examples of domain-domain interactions are modeled. Pax-5 requires both its PAI and RED domains to bind to PD sites. HDs found in proteins without PDs can dimerize to recognize P3 sites. Pax-6-5a contains a nonfunctional PAI domain and recognizes a fourth type of site through its RED domain. A speculative model of RED and HD interactions are illustrated by Lune, a newly identified protein, which lacks a PAI domain altogether and may recognize a new type of DNA sequence.

Functional PD and HD binding sites

Thus, the Paired protein, and likely all Pax proteins, can use combinations of their HTH domains through both intramolecular and intermolecular interactions to bind to a variety of DNA sequences. However, although interactions among the HTH motifs of the Pax proteins can be shown in vitro, few functional binding sites for such modes of binding have been identified in vivo. Sites for the Pax-5 PD have been identified in the promoters of B cell specific genes and they resemble the optimal PAI/RED binding sites obtained for this class of PD molecules (Czerny et al., 1993, this work). A palindromic sequence (P3) which is the optimal binding site for the binding of homodimers of Prd-class HDs is found to be highly conserved in the promoters of most opsin genes, from flies to human (RCS1 sequence), and is required for opsin expression in *Drosophila* (Fortini and Rubin, 1990; Mismer and Rubin, 1989). This sequence is likely to mediate the function of Pax-6, the master regulator of eye development encoded by the *eyeless* gene of *Drosophila* (Halder et al., 1995; Wilson and Desplan, 1995). Interestingly, in *C. elegans,* the same Pax-6 gene is found as two transcription units containing either the whole Pax-6 region (PD + HD), or only the HD, and corresponding to two distinct genetic functions encoded by the *vab-3* and *mab-18* genes, respectively (Chisholm and Horvitz, 1995; Zhang and Emmons, 1995). It is possible that the function of Pax-6 in regulating opsin gene expression is only mediated by dimerization through the HD, and not through the PD. However, binding sites for the PD of Pax-6 have also been identified in the crystallin promoters (Cvekl et al., 1995; Richardson et al., 1995), suggesting that these two domains act together similarly to the POU domain.

The type of sites recognized by cooperative binding by the PD and HD have not yet been identified. However, in vivo results with combinations of Prd mutant proteins indicate that both the PD and the HD are required on the same molecule to mediate *prd* function. Activation of the *prd* target segment polarity genes, *engrailed*, *wingless* and *gooseberry*, does not occur in the absence of wild type *prd*, even when two Prd rescue constructs bearing mutations affecting either the PAI domain or the HD are present in the same cell. This suggests that Prd may principally act by binding through a combination of its PAI domain and HD (The RED domain is dispensable for in vivo *prd* function; Bertuccioli et al., 1996; Cai et al., 1994). Consistent with this view, a site closely resembling the PH0 site has been found in a 'late promoter element' of the pair-rule gene *even-skipped (eve)*. This short element is able to mediate late activation of *eve* in single cell wide stripes resembling *engrailed* expression. Upon mutation of this site, this *prd*-dependent expression is lost. Mutations that affect the spacing between the two halves of the PH0 site, or either the PD or HD half site also lead to a dramatic decrease in expression (Fujioka et al., 1996). Thus, a PH0 site has been shown to be essential to mediate *prd* function in vivo. Finally, no functional sites bearing resemblance to the Pax-6 5a binding site (RED domain) have been found yet. We predict that binding sites for Pax-6 5a as well as binding sites for the Lune protein, which do not contain a functional PAI domain, will require the combined activity of the RED domain and the HD. These sites, whose geometry is unknown, may be difficult to identify because their length will likely allow multiple mismatches.

Conclusion

We propose that Pax proteins have modular HTH domains and use them in different combinations to recognize a variety of specific DNA sequences. In addition, Pax proteins and HD proteins can interact both intramolecularly and intermolecularly to regulate transcription. Using different combinations of HTH DNA binding domains, which otherwise have similar DNA binding specificities, Pax proteins can thus modulate the binding specificities of the individual HTHs. By doing this, they can achieve several distinct functional specificities and hence participate in several different developmental pathways.

REFERENCES

Baldwin, C. T., Hoth, C. F., Amos, J. A., da Silva, E. and Milunsky, A. (1992). An exonic mutation in the HuP2 paired domain gene causes Waardenburg's syndrome. *Nature* **355**, 637-638.

Bertuccioli, C., Fasano, L., Jun, S., Wang, S., Sheng, G. and Desplan, C. (1996). In vivo requirement for the paired domain and homeodomain of the *paired* segmentation gene product. *Development* **122**, 2673-2685.

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

Cai, J., Lan, Y., Appel, L. F. and Weir, M. (1994). Dissection of the Drosophila Paired protein: Functional requirements for conserved motifs. *Mech. Dev.* **47**, 139-150.

Chalepakis, G., Fritsch, R., Fickenscher, H., Deutsch, U., Goulding, M. and Gruss, P. (1991). The molecular basis of the undulated/Pax-1 mutation. *Cell* **66**, 873-884.

Chalepakis, G., Stoykova, A., Wijnholds, J., Tremblay, P. and Gruss, P. (1993). Pax: gene regulators in the developing nervous system. *J. Neurobiol.* **24**, 1367-1384.

Chan, S. K., Jaffe, L., Capovilla, M., Botas, J. and Mann, R. S. (1994). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell* **78**, 603-615.

Chang, C., Shen, W., Rozenfeld, S., Lawrence, H. J., Largman, C. and Cleary, M. L. (1995). Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev.* **9**, 663-674.

Chisholm, A. D. and Horvitz, H. R. (1995). Patterning of the Caenorhabditis elegans head region by the Pax-6 family member vab-3. *Nature* **376**, 52-55.

Cvekl, A., Sax, C. M., Li, X., McDermott, J. B. and Piatigorsky, J. (1995). Pax-6 and lens-specific transcription of the chicken delta 1-crystallin gene. *Proc. Natl. Acad. Sci. USA* **92**, 4681-4685.

Czerny, T. and Busslinger, M. (1995). DNA-binding and transactivation properties of Pax-6: three amino acids in the Paired Domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). *Molec. Cell. Biol.* **15**, 2858-2871.

Czerny, T., Schaffner, G. and Busslinger, M. (1993). DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev.* **7**, 2048-2061.

Ellenberger, T. E., Brandl, C. J., Struhl, K. and Harrison, S. C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* **71**, 1223-1237.

Epstein, D. J., Malo, D., Vekemans, M. and Gros, P. (1991a). Molecular characterization of a deletion encompassing the splotch mutation on mouse chromosome 1. *Genomics* **10**, 89-93.

Epstein, D. J., Vekemans, M. and Gros, P. (1991b). *splotch* (Sp2H), a Mutation Affecting Development of the Mouse Neural Tube, Shows a Deletion within the Paired Homeodomain of *Pax-3*. *Cell* **67**, 767-774.

Epstein, J., Cai, J., Glaser, T., Jepeal, L. and Maas, R. (1994a). Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J. Biol. Chem.* **269**, 8355-8361.

Epstein, J. A., Glaser, T., Cai, J., Jepeal, L., Walton, D. S. and Maas, R. L. (1994b). Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. *Genes Dev.* **8**, 2022-2034.

Fortini, M. E. and Rubin, G. M. (1990). Analysis of cis-acting requirements of the Rh3 and Rh4 genes reveals a bipartite organization to rhodopsin promoters in Drosophila melanogaster. *Genes Dev.* **4**, 444-463.

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

Fujioka, M., Miskiewicz, P., Raj, L., Gulledge, A., Weir M. and Goto, T. (1996). *Drosophila* paired regulates late *even-skipped* expression through a composite binding site for the paired domain and the homeodomain. *Development* **122 (9),** 000-000.

Glaser, T., Walton, D. S. and Maas, R. L. (1992). Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene. *Nature Genetics* **2**, 232-239.

Goulding, M., Sterrer, S., Fleming, J., Balling, R., Nadeau, J., Moore, K. J., Brown, S. D., Steel, K. P. and Gruss, P. (1993). Analysis of the Pax-3 gene in the mouse mutant splotch. *Genomics* **17**, 355-363.

Halder, G., Callaerts, P. and Gehring, W. J. (1995). Induction of Ectopic Eyes by Targeted Expression of the eyeless Gene in *Drosophila. Science* **267**, 1788-1792.

Hanson, I. M., Fletcher, J. M., Jordan, T., Brown, A., Taylor, D., Adams, R. J., Punnett, H. H. and van Heyningen, H. (1994). Mutations at the PAX6 locus are found in heterogeneous anterior segment malformations including Peters' anomaly. *Nature Genetics* **6**, 168-173.

Hill, R. E., Favor, J., Hogan, B. L. M., Ton, C. C. T., Saunders, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. D. and van Heyningen, H. (1991). Mouse *Small* eye results from mutations in a paired-like homeobox-containing gene. *Nature* **354**, 522-525.

Jordan, T., Hanson, I., Zaletayev, D., Hodgson, S., Prosser, J., Seawright, A., Hastie, N. and van Heyningen, H. (1992). The human PAX6 gene is mutated in two patients with aniridia. *Nature Genetics* **1**, 328-332.

Kilchherr, F., Baumgartern, S., Bopp, D., Frei, E. and Noll, A. M. (1986). Isolation of the paired gene of Drosophila and its spatial expression during early embryogenesis. *Nature* **321**, 493-499.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W. and Pabo, C. O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* **77**, 21-32.

Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R. and Sigler, P. B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497-505.

Maulbecker, C. C. and Gruss, P. (1993). The oncogenic potential of Pax genes. *EMBO J.* **12**, 2361-2367.

Miller, J., McLachlan, A. D. and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J.* **4**, 1609-1514.

Miskiewicz, P., Morrissey, D., Lan, Y., Raj, L., Kessler, S., Fujioka, M., Goto, T. and Weir, M. (1996). Both the paired domain and homeodomain are required for in vivo function of *Drosophila* Paired. *Development* **122**, 2709-2718.

Mismer, D. and Rubin, G. M. (1989). Definition of cis-acting elements regulating expression of the Drosophila melanogaster ninaE opsin gene by oligonucleotide-directed mutagenesis. *Genetics* **121**, 77-87.

Morell, R., Friedman, T. B., Moeljopawiro, S., Hartono, Soewito and Asher, J. J. (1992). A frameshift mutation in the HuP2 paired domain of the probable human homolog of murine Pax-3 is responsible for Waardenburg syndrome type 1 in an Indonesian family. *Human Molec. Genet.* **1**, 243-247.

Noll, M. (1993). Evolution and role of Pax genes. *Curr. Opin. Genet. Dev.* **3**, 595-605.

Pavletich, N. P. and Pabo, C. O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* **261**, 1701-1707.

Quiring, R., Walldorf, U., Kloter, U. and Gehring, W. J. (1994). Homology of the eyeless Gene of Drosophila to the Small eye Gene in mice and Aniridia in humans. *Science* **265**, 785-789.

Rebar, E. J. and Pabo, C. O. (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671-673.

Richardson, J., Cvekl, A. and Wistow, G. (1995). Pax-6 is essential for lens-specific expression of zeta-crystallin. *Proc. Natl. Acad. Sci. USA* **92**, 4676-4680.

Stapleton, P., Weith, A., Urbanek, P., Kozmik, Z. and Busslinger, M. (1993). Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX-9. *Nature Genetics* **3**, 292-298.

Strachan, T. and Read, A. P. (1994). Pax genes. *Curr. Opin. Genet.* Dev. **4**, 427-438.

Tassabehji, M., Read, A. P., Newton, V. E., Harris, R., Balling, R., Gruss, P. and Strachan, T. (1992). Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature* **355**, 635-636.

Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. and Desplan, C. (1989). A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell* **59**, 553-562.

Treisman, J., Harris, E. and Desplan, C. (1991). The Paired Box encodes a second DNA-binding domain in the Paired homeodomain protein. *Genes Dev.* **5**, 594-604.

Tsukamoto, K., Nakamura, Y. and Niikawa, N. (1994). Isolation of two isoforms of the PAX3 gene transcripts and their tissue-specific alternative expression in human adult tissues. *Human Genet.* **93**, 270-274.

van Dijk, M. A. and Murre, C. (1994). Extradenticle raises the DNA binding specificity of homeotic selector gene products. *Cell* **78**, 617-624.

Voronova, A. and Baltimore, D. (1990). Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proc. Natl. Acad. Sci. USA* **87**, 4722-4726.

Wallin, J., Mizutani, Y., Imai, K., Miyashita, N., Moriwaki, K., Taniguchi, M., Koseki, H. and Balling, R. (1993). A new Pax gene, Pax-9, maps to mouse chromosome 12. *Mammalian Genome* **4**, 354-358.

Walther, C., Guenet, J. L., Simon, D., Deutsch, U., Jostes, B., Goulding, M. D., Plachov, D., Balling, R. and Gruss, P. (1991). Pax: a murine multigene family of paired box-containing genes. *Genomics* **11**, 424-434.

Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. and Desplan, C. (1993). Cooperative dimerization of paired class homeodomains on DNA. *Genes Dev.* **7**, 2120-2134.

Wilson, D. S. and Desplan, C. (1995). Homeodomain Proteins: Cooperating to be different. *Current Biol.* **5**, 32-34.

Wilson, D. S., Guenther, B., Desplan, C. and Kuriyan, J. (1995). High resolution crystal structure of a Paired (Pax) class cooperative homeodomain dimer on DNA. *Cell* **82***, 709-719

Xu, W., Rould, M., Jun, S., Desplan, C. and Pabo, C. (1995). Crystal structure of a paired domain-DNA complex at 2.5A resolution reveals structural basis for Pax developmental mutations. *Cell* **80**, 639-650.

Zhang, Y. and Emmons, S. W. (1995). Specification of sense-organ identity by a Caenorhabditis elegans Pax-6 homologue. *Nature* **376**, 55-59.