Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8

Zbynek Kozmik¹, Thomas Czerny and Meinrad Busslinger²

Research Institute of Molecular Pathology, Dr Bohr-Gasse 7, A-1030 Vienna, Austria

¹Present address: Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo n.2, Prague, Czech Republic

²Corresponding author
e-mail: Busslinger@nt.imp.univie.ac.at

Z.Kozmik and T.Czerny contributed equally to this work

Transcription factors of the Pax family bind to their target genes via the paired domain which is known to be composed of two subdomains each recognizing distinct half-sites in adjacent major grooves of the DNA helix. We now demonstrate that the mammalian Pax8 gene gives rise, by alternative mRNA splicing, to a protein isoform containing an extra serine residue in the recognition α -helix 3 of the paired domain. This Pax8(S) protein does not interact with bipartite paired domain-binding sites, indicating that inactivation of the N-terminal DNA-binding motif severely restricts the sequence specificity of the paired domain. However, the Pax8(S) protein binds *in vitro* and *in vivo* to the 5aCON sequence which was previously identified as a high-affinity binding site for the Pax6(5a) splice variant carrying a 14-amino-acid insertion in the paired domain. The 5aCON sequence is shown to consist of four interdigitated 5' half-sites of the bipartite consensus sequence and is thus bound by four Pax8(S) molecules via the intact C-terminal DNA-binding motif of the paired domain. Together these data suggest that inactivation of the N-terminal region of the paired domain by alternative splicing is used *in vivo* to selectively target Pax transcription factors to gene regulatory regions containing highly specialized 5aCON-like sequences.

Keywords: alternative splicing/DNA binding/paired domain/Pax6/Pax8

Introduction

The Pax proteins constitute a family of important developmental regulators which are defined by the presence of a highly conserved DNA-binding motif of 128 amino acids, the so-called paired domain (Bopp *et al.*, 1986; Chalepakidis *et al.*, 1991; Treisman *et al.*, 1991). About half of the Pax proteins furthermore use a paired-type homeodomain as a second DNA-binding region. The mammalian genome contains nine Pax genes (Walther *et al.*, 1991; Stapleton *et al.*, 1993) which are expressed in spatially and temporally restricted patterns during development and in adult

tissues (reviewed by Stuart *et al.*, 1994). Several of these Pax genes have been implicated by genetic evidence in the control of early development, differentiation and oncogenesis, as they are frequently associated with mouse developmental mutants, human disease syndromes or chromosomal translocations in human tumours (reviewed by Strachan and Read, 1994; Stuart *et al.*, 1994; Busslinger *et al.*, 1996).

Some aspects of Pax gene function have been highly conserved in evolution, as is best exemplified by Pax6. Dysfunction of the Pax6 gene causes defects in eye development from *Drosophila* to man, as Pax6 mutations are responsible for the *Small eye* phenotype in the mouse (Hill *et al.*, 1991), aniridia and Peters' anomaly in humans (Ton *et al.*, 1991; Hanson *et al.*, 1994) and *eyeless* in *Drosophila* (Quiring *et al.*, 1994). Gain-of-function experiments in *Drosophila* furthermore demonstrated that Pax6 acts as a master control gene in eye morphogenesis (Halder *et al.*, 1995). During mouse ontogeny, the Pax6 gene is expressed, in addition to the developing eye, in the pancreas, nasal epithelia and several distinct regions of the CNS including the forebrain, hindbrain and spinal cord (Walther and Gruss, 1991; Grindley *et al.*, 1995; St-Onge *et al.*, 1997). Alternative splicing is known to generate two distinct Pax6 isoforms, referred to as Pax6 and Pax6(5a), which differ by the inclusion of 14 amino acids encoded by the additional exon 5a (Walther and Gruss, 1991; Glaser *et al.*, 1992). This 14-amino-acid insertion was shown to alter profoundly the DNA-binding activity of the Pax6 paired domain (Epstein *et al.*, 1994a). In contrast, the Pax8 gene is known to code for alternative splice products with different transactivation potentials (Kozmik *et al.*, 1993).

The Pax8 gene is expressed during mouse ontogeny in the developing CNS, thyroid gland, kidney and placenta (Plachov *et al.*, 1990; Kozmik *et al.*, 1993). Transcripts of this gene give rise to four alternative splice products (Pax8a to Pax8d) which differ in their C-terminal sequences while sharing common N-terminal regions including the paired domain. These Pax8 isoforms consequently bind DNA in an indistinguishable manner, but exhibit distinct transactivation properties (Kozmik *et al.*, 1993). Two additional isoforms, Pax8e and Pax8f, lack paired domain sequences altogether and are therefore expected to interact with DNA differently, if at all (Kozmik *et al.*, 1993). Interestingly, the alternative splicing of the six Pax8 mRNA isoforms is spatially and temporally regulated during mouse embryogenesis (Kozmik *et al.*, 1993).

The paired domain is a bipartite DNA-binding motif, as was first revealed by our mutational analyses of Pax5 (BSAP) and its recognition sequences (Czerny *et al.*, 1993). According to this model, the paired domain is composed of two subdomains which bind to two distinct

half-sites in adjacent major grooves of the DNA helix. By analysing naturally occurring target sequences, we arrived at a consensus recognition sequence for Pax5 in which the 5' half-site was shown to interact with the C-terminal region of the paired domain (Czerny *et al.*, 1993). X-ray crystallographic analysis of the paired domain–DNA complex recently confirmed this bipartite model by demonstrating that each of the two subdomains is composed of a helix–turn–helix (HTH) motif resembling the structure of the homeodomain (Xu *et al.*, 1995). Interestingly, the paired domain of the *Drosophila* Paired protein relies entirely on the N-terminal subdomain for DNA binding (Jun and Desplan, 1996) in agreement with the observation that the C-terminal motif is dispensable for normal *paired* function *in vivo* (Cai *et al.*, 1994; Bertuccioli *et al.*, 1996). In contrast, the paired domain of Pax6 depends on both subdomains for DNA sequence recognition similar to all members of the Pax2/5/8 family (Epstein *et al.*, 1994b; Czerny and Busslinger, 1995). The bipartite consensus sequences defined for these two classes of Pax proteins are very similar, as they primarily differ in only one nucleotide position (Epstein *et al.*, 1994b; Czerny and Busslinger, 1995). The differential recognition of this divergent position by Pax5 and Pax6 is achieved by three distinct amino acid residues which are located within a heptapeptide sequence in the N-terminal part of the paired domain (Czerny and Busslinger, 1995; Jun and Desplan, 1996). This heptapeptide sequence is disrupted by the alternatively spliced 14-amino-acid insertion in Pax6(5a), thus inactivating the DNA-binding function of the N-terminal HTH motif of the paired domain (Epstein *et al.*, 1994a). By a binding site selection assay, Epstein *et al.* (1994a) identified a unique recognition sequence, termed 5aCON, for the extended paired domain of Pax6(5a) and therefore hypothesized that the exon 5a insertion acts as a molecular toggle to unmask the DNA-binding potential of the C-terminal subdomain.

Here we show that alternative mRNA splicing also modifies the paired domain of Pax8 by insertion of an extra serine residue into the recognition helix of the N-terminal subdomain. The generation of this Pax8(S) mRNA isoform is not regulated during mouse ontogeny and is uncoupled from the alternative splicing events affecting the C-terminal sequences. The Pax8(S) protein behaves like the insertion version Pax6(5a) in that it selectively binds *in vitro* and *in vivo* to the 5aCON site rather than to bipartite paired domain recognition sequences. The 5aCON sequence is shown to consist of a unique assembly of four consensus sites for the C-terminal DNA-binding motif of the paired domain. These data demonstrate therefore that inactivation of the N-terminal region of the paired domain by alternative splicing is a more general phenomenon which results in severe restriction of the DNA-binding specificity of the paired domain and thus in selective targeting of Pax proteins to gene regulatory regions containing 5aCON-like sequences.

Results

A novel Pax8 isoform with a serine insertion in the paired domain fails to bind to bipartite recognition sequences

We have previously described the cloning of distinct Pax8 isoforms (Pax8a–f) which differ in their C-terminal

sequences due to alternative splicing (Kozmik *et al.*, 1993). While characterizing these isoforms in transiently transfected COP-8 fibroblasts, we realized that 30% of all cDNAs (cloned in an expression vector) gave rise to Pax8 proteins which were unable to bind to the paired domain recognition sequence of the histone H2A-2.1 gene in electrophoretic mobility shift assays (EMSA) (Figure 1C; also Z.Kozmik, unpublished data). Sequence analysis of the corresponding cDNAs revealed the presence of three extra nucleotides (TAG) in the paired box, thus resulting in the insertion of an additional serine residue (Figure 1A). We therefore refer to this novel isoform as Pax8(S). Comparison of the Pax8(S) cDNA sequence with the exon–intron structure of the human *PAX8* gene (Kozmik *et al.*, 1993) indicated that the trinucleotide TAG originates from the 3' junction of intron 2 due to the presence of an alternative 3' splice site which precedes the normal splice junction by three nucleotides (Figure 1A). For evolutionary comparison, the corresponding intron 2 of the murine *Pax8* gene was amplified by PCR from genomic DNA and sequenced (see Materials and methods). As shown in Figure 1A, the sequences surrounding the splice junctions of intron 2 have been completely conserved between the human and murine *Pax8* genes in agreement with the fact that Pax8(S) cDNAs were isolated from both species. Moreover, the Pax8(S) splice form was found in combination with all the various alternative splices affecting the C-terminus of Pax8 (Figure 1C). These data indicate therefore that the N- and C-terminal splice events occur independently of each other.

The serine insertion in Pax8(S) is reminiscent of the Pax6(5a) splice variant which contains an insertion of 14 amino acids in the N-terminal part of the paired domain of Pax6 (Walther and Gruss, 1991; Glaser *et al.*, 1992; Püschel *et al.*, 1992; Dozier *et al.*, 1993). These extra 14 amino acids are encoded by an additional exon (named 5a) which is included, by alternative splicing, into the paired domain-coding sequences of the Pax6 mRNA (Figure 1B). This 14-amino-acid insertion was shown to prevent the Pax6(5a) isoform from binding to a Pax6 consensus recognition sequence (Epstein *et al.*, 1994a). For this reason we decided to compare directly the DNA-binding properties of the Pax6(5a) and Pax8(S) isoforms by EMSA analysis. Both isoforms were equally unable to interact with the H2A-2.1 site (Figure 1C) and with a large panel of paired domain-binding sequences (Z.Kozmik, unpublished data) which was previously described by Czerny *et al.* (1993). Hence, both the Pax6(5a) and Pax8(S) isoforms fail to interact with bipartite paired domain recognition sequences, suggesting that the DNA-binding function of the paired domain is similarly affected by the amino acid insertions in both proteins.

Constitutive splicing of Pax8(S) and Pax6(5a) mRNA during development and in adult tissues

We next determined the relative abundance of the different Pax6 and Pax8 mRNA splice forms during ontogeny by using total RNA from mouse embryos, adult tissues and human cell lines for RNase protection analysis. A riboprobe containing paired domain sequences of the murine Pax8(S) cDNA was expected to give rise to a diagnostic fragment of 159 nt for the Pax8(S) mRNA, while protection by the Pax8 mRNA should result in two

Fig. 1. Structure and DNA-binding properties of the alternative splice forms Pax8(S) and Pax6(5a). **(A)** Generation of the Pax8(S) isoform by alternative splicing. Intron 2 of the human and mouse *Pax8* genes is shown together with the sequences of the exon–intron junctions, the cloned cDNAs and the corresponding proteins. Filled arrowheads point to the cleavage sites within the consensus 5' and 3' splice sequences, and an open symbol indicates cleavage at the alternative 3' splice site used for the generation of the Pax8(S) variant. Amino acid residues are numbered according to the published Pax8 sequence (Plachov *et al.*, 1990; Kozmik *et al.*, 1993). **(B)** Schematic diagram of the splice variant Pax6(5a). The alternatively spliced exon 5a of the *Pax-6* locus codes for 14 amino acids which are shown for the mouse Pax6(5a) isoform (Walther and Gruss, 1991). **(C)** Protein–DNA binding assay. Pax8a and Pax8b isoforms lacking (–S) or containing (+S) the serine insertion as well as the Pax6 splice variants lacking (–5a) or containing (+5a) exon 5a were expressed in transiently transfected COP-8 cells. Nuclear extracts containing equivalent protein amounts were subjected to EMSA with the paired domain-binding site of the sea urchin H2A-2.1 gene (upper panel). The same extracts were analysed by SDS–PAGE (10% gel) followed by immunoblotting with anti-paired domain antibodies (Adams *et al.*, 1992) (lower panel).

fragments of 59 and 97 nt due to cleavage of the probe at the inserted serine codon (Figure 2A). As shown by the abundance of these three protected fragments, the Pax8 and Pax8(S) mRNAs were detected at a constant ratio of 5:1 throughout embryogenesis as well as in adult kidney and ovary, thus suggesting that this alternative splicing event is not regulated during development. Likewise, analysis of total RNA isolated from human kidney cell lines confirmed the presence of the two Pax8 mRNA variants at the same molar ratio (Figure 2B). A similar strategy was used to map the relative abundance of the two Pax6 mRNA isoforms. As shown in Figure 2C, the Pax6 and Pax6(5a) transcripts are also expressed at a constant ratio (8:1) throughout embryogenesis and in the adult brain. In conclusion, the alternative splicing in the paired domain region of Pax6 and Pax8 is apparently not developmentally regulated, in contrast to the splicing events which are responsible for assembling the C-terminal sequences of the Pax8 protein (Kozmik *et al.*, 1993).

Amino acid insertions restrict the DNA sequence specificity of the paired domain

The inability of the Pax8(S) and Pax6(5a) proteins to bind to bipartite paired domain-binding sites could arise in two different ways. The amino acid insertions may result either in a loss of DNA-binding activity or in a change in sequence specificity of the paired domain. To distinguish between these two possibilities, we have bacterially expressed and subsequently purified paired domain (pd) polypeptides of the different Pax6 and Pax8 isoforms (see

Materials and methods). Equivalent amounts of each protein were analysed for binding to a mixture of DNA sequences consisting of 25 random nucleotides in the centre of the probe. As shown by the EMSA experiment in Figure 3A, the Pax6-pd and Pax8-pd peptides could form complexes with the degenerate DNA probe in a concentration-dependent manner and were thus capable of selecting binding sites from this mixture of random sequences. In contrast, the Pax8(S)-pd and Pax6(5a)-pd proteins were unable to bind to the degenerate DNA probe, indicating that both insertion versions could interact only with very few sequences, if any at all, of the random oligonucleotide pool. It appears therefore that amino acid insertions into the N-terminal part of the paired domain result in a dramatic reduction or loss of the DNA-binding potential of the paired domain.

Epstein *et al.* (1994a) have recently demonstrated that the paired domain of the Pax6(5a) protein is still able to bind to DNA. Using a PCR-based binding site selection assay, they have identified a 22-bp consensus sequence (termed 5aCON) as a high-affinity binding site for the Pax6(5a) protein (Epstein *et al.*, 1994a). As the Pax6(5a) and Pax8(S) isoforms exhibit similar DNA-binding properties, we tested the possibility that the Pax8 paired domain peptides may also interact with the 5aCON sequences (Figure 3B). In the absence of unspecific competitor DNA, both Pax8 paired domain proteins, either lacking (–S) or containing (+S) the serine insertion, bound as efficiently to the 5aCON site (Figure 3B, lanes 2 and 4) as the two Pax6 paired domain polypeptides (lanes 6 and 8).

Fig. 2. Constitutive splicing of *Pax6* and *Pax8* transcripts during development. **(A)** Expression of the two *Pax8* mRNA isoforms during mouse embryogenesis and in adult tissues. Total RNA was prepared at day 9.5 and 10.5 post-coitum from the whole embryo (we, including extra-embryonic membranes and placenta) and at day 11.5 from the embryo proper (e). Later embryos were dissected into head (h) and trunk (t) prior to RNA preparation. Total RNA (20 µg) was analysed by RNase protection assay with a *Pax8*(S) riboprobe mapping paired domain sequences from positions 250 (*Hind*III) to 405 (*Bal*I) of the mouse *Pax8* gene (Plachov *et al.*, 1990). RNase-protected fragments were separated on an 8% polyacrylamide sequencing gel together with an end-labelled DNA size marker (lane M; pUC19 digested with *Msp*I; sizes given in nt). The expected pattern of protection (given in nt) is indicated below for the two *Pax8* mRNA isoforms. A black box denotes the inserted serine codon. **(B)** Expression of the alternatively spliced *PAX8* transcripts in human kidney cells. Total RNA (10 µg) of human kidney cell lines was analysed by RNase protection with a *Pax8*(S) riboprobe spanning the sequences from positions 98 (*Ava*II) to 260 (*Bal*I) of the human *PAX8* gene (Kozmik *et al.*, 1993). **(C)** Expression of the two *Pax6* mRNA variants in embryogenesis and adult tissues of the mouse. The same RNA preparations described in (A) were analysed with a *Pax6*(5a)-specific riboprobe containing mouse *Pax6* sequences from positions 134 to 598 (Walther and Gruss, 1991).

Moreover, binding to the 5aCON site was cooperative, as it could not be observed at a 10-fold lower protein concentration (Figure 3B, lanes 1, 3, 5 and 7). Interestingly, complex formation between the 5aCON site and the paired domain polypeptides lacking amino acid insertions was preferentially lost, if bacterial plasmid sequences was included as competitor DNA in the binding reaction (Figure 3C, lanes 9 and 10; also unpublished data). However, binding of *Pax8*(S) was even observed at a 10-fold higher competitor DNA concentration (Figure 3C, compare lanes 12 and 13). This finding strongly suggests that only the insertion versions *Pax6*(5a)-pd and *Pax8*(S)-pd bind to the 5aCON site with high affinity. The peptides *Pax8*-pd and *Pax6*-pd are instead titrated away from the 5aCON site by the competitor DNA, probably due to the high DNA-binding potential of the intact paired domain. This finding is consistent with the binding data obtained with the degenerate DNA probe which revealed a relatively broad DNA sequence specificity of the unmodified paired domain (Figure 3A). Taken together, these results demonstrate therefore that N-terminal amino acid insertions dramatically reduce the DNA-binding potential of the

paired domain by restricting the sequence specificity to a few sites.

The 5aCON element consists of four interdigitated 5' half-sites of the Pax consensus recognition sequence

The low electrophoretic mobility of the complexes formed between the 5aCON site and the *Pax6*(5a) or *Pax8*(S) proteins suggested to us that more than one protein molecule was bound to the DNA (Figure 3C). Binding of multiple molecules was particularly evident in the presence of competitor DNA which gave rise to four distinct complexes, each differing by one additional protein molecule. In their analysis of the 5aCON sequence, Epstein *et al.* (1994a) reached the conclusion that the 5aCON site consists of two direct repeats which lack any extended homology with the bipartite consensus recognition sequence of the paired domain. Since the amino acid insertions of *Pax6*(5a) or *Pax8*(S) appear to inactivate the N-terminal DNA-binding region of the paired domain (see Figure 6A), we focused our attention to the sequence motif recognized by the intact C-terminal subdomain.

Fig. 3. Amino acid insertions restrict the DNA-binding specificity of paired domain. **(A)** DNA-binding activity of purified paired domain polypeptides. Histidine-tagged proteins were generated by the addition of six histidine residues at the N-terminus of the paired domain (pd) of Pax6 (amino acids 1–131) and Pax8 (amino acids 1–195). These paired domain polypeptides were expressed in *E. coli* and purified to homogeneity by chromatography on Ni^{2+} -NTA agarose (see Materials and methods). Increasing amounts (20, 100 or 500 ng) of the purified proteins were incubated with 1 μg of poly[d(I-C)] and ~0.1 ng of a mixture of ^{32}P -labelled oligonucleotides containing 25 random nucleotides (N₂₅) in the centre of the probe (see Materials and methods). Protein-DNA complexes were analysed by EMSA. **(B and C)** Binding of Pax6 and Pax8 isoforms to the 5aCON sequence. Increasing amounts of the indicated proteins (5 ng for odd-numbered lanes; 50 ng for even-numbered lanes and lane 13) were incubated with ~0.1 ng of the end-labelled 5aCON probe (Epstein *et al.*, 1994a) in the absence (B) or presence (C) of non-specific competitor DNA prior to EMSA analysis. The plasmid pKW10 (Adams *et al.*, 1992) was used as competitor DNA at either 10 ng (lanes 9–12) or 100 ng (lane 13) per binding reaction. Poly[d(I-C)] was omitted in all cases. Numbers to the right indicate the distinct multimeric complexes formed by the Pax8(S)-pd polypeptide.

We have previously demonstrated that the C-terminal region of the paired domain recognizes the 5' half-site (A/G)-GCA-T(C/G)A of the bipartite consensus sequence (Czerny *et al.*, 1993; Czerny and Busslinger, 1995). As shown in Figure 4A, each repeat of the 5aCON sequence is indeed composed of two 5' half-sites which are interdigitated in opposite orientation. In this arrangement, several base pairs are part of two overlapping recognition sequences and must therefore be contacted by two paired

Fig. 4. The 5aCON sequence is a tetrameric paired domain-binding site. **(A)** Analysis of the 5aCON sequence. The up-dated bipartite consensus recognition sequence for the paired domain of Pax6 and Pax8 (Czerny and Busslinger, 1995) is shown together with the 5aCON sequence (Epstein *et al.*, 1994a). The 5' half-site as defined by Czerny *et al.* (1993) is indicated by a long arrow. Nucleotides of the 5aCON sequence which are homologous to the 5' half-site are highlighted by black overlay. Base mutations introduced into the 5aCON sequence are denoted by asterisks, and the corresponding C- and G-residues in the complementary strand, which still conform to the consensus sequence of the 5' half-site, are indicated by grey shading. The tetrameric consensus sequence is obtained by alignment of the two complementary DNA strands of the 5aCON sequence. **(B)** Verification of the tetrameric nature of the 5aCON site. The purified Pax8(S)-pd protein (50 ng) was incubated with labelled wild-type or mutated 5aCON oligonucleotides in the presence of the competitor DNA pKW10 (100 ng). The distinct multimeric complexes revealed by EMSA analysis are indicated by numbers to the left. The Pax8-pd protein lacking the serine insertion (-S) was incubated with the bipartite recognition sequence of the *I ϵ* promoter (Czerny *et al.*, 1993) to indicate the position of the monomeric complex.

domain molecules from opposite sides of the DNA helix. This is only possible, if these base pairs are simultaneously recognized by one protein through the major groove and by the second protein through the minor groove (see Discussion and Figure 6C). Moreover, the two repeats of the 5aCON sequence are separated by 11 bp, i.e. by one turn of the DNA helix, thus indicating that 5' half-sites with the same orientation are accessible from the same side of the DNA helix. Four 5' half-sites can be organized in this manner only in one possible arrangement, thus resulting in the definition of a unique 'tetrameric' consensus recognition sequence (Figure 4A). The tetrameric nature of the 5aCON site also offers an explanation why four paired domain polypeptides can interact with this DNA sequence. To verify our model experimentally, we have introduced mutations that selectively destroyed individual half-sites of the 5aCON sequence, thus generating 'trimeric' and 'dimeric' binding sites (Figure 4A). As predicted, the 'tetrameric', 'trimeric' and 'dimeric' sites could maximally bind four, three or two Pax8(S)-pd molecules, respectively (Figure 4B). The presence of one Pax8(S)-pd molecule in complex #1 was directly demonstrated by co-migration with the complex formed between the *I ϵ* site and Pax8-pd (Figure 4B), as the intact paired domain binds only as a monomer to bipartite recognition sequences such as the *I ϵ* site (Adams *et al.*, 1992; Czerny *et al.*, 1993). In summary, we conclude

Fig. 5. Differential binding of Pax6 and Pax8 isoforms to bipartite and tetrameric recognition sequences *in vivo*. The transactivation potential of Pax8-VP16 fusion proteins and full-length Pax6 isoforms was measured by transient transfection assay using the luciferase reporter genes Iε-luc and 5aCON-luc which contain the bipartite Iε site or the tetrameric 5aCON sequence upstream of the TATA box (see diagram below). The Pax8-VP16 proteins were generated by fusion of the mPax8 paired domain at codon 153 (Plachov *et al.*, 1990) to the transactivation domain of VP16 (amino acids 413–490; Triezenberg *et al.*, 1988). Increasing amounts (0.01–1 μg) of expression plasmids directing the synthesis of the indicated isoforms were transfected together with the Iε-luc (A and C) or 5aCON-luc reporter gene (B and D) into RAC65 cells. The amount of expression plasmid was equalized by the addition of the empty expression vector pKW10 to normalize for promoter interference effects. The luciferase activities of representative experiments are shown as values relative to the basal activity observed with the vector pKW10. All data were normalized for differences in transfection efficiencies by co-transfection of a reference CMV-CAT gene and by standardizing luciferase activities to CAT values.

therefore that the 5aCON sequence represents a unique assembly of four interdigitated recognition sequences for the C-terminal DNA-binding motif of the paired domain. These data furthermore confirm that only highly specialized binding sites can be recognized by Pax protein containing N-terminal amino acid insertions in the paired domain.

Selective binding of Pax6(5a) and Pax8(S) to the tetrameric 5aCON site *in vivo*

The ability of the two isoforms of Pax6 and Pax8 to mediate transcriptional activation from bipartite and tetrameric recognition sequences was next investigated by transient transfection assay *in vivo* in the embryonal carcinoma cell line RAC65 (Pratt *et al.*, 1990). To this end, four copies of a bipartite binding site originating from the Iε promoter (Czerny *et al.*, 1993) were cloned upstream of the TATA box in the luciferase reporter construct Iε-luc (Figure 5). The Iε site was chosen because it conferred only low background activity to the luciferase gene in RAC65 cells. On the other hand, a single tetrameric 5aCON oligonucleotide (also containing four Pax-binding sites) was inserted in the luciferase gene 5aCON-luc (Figure 5). Initial co-transfection experiments with full-length Pax8 expression vectors resulted only in low transcrip-

tional stimulation of these reporter genes quite in contrast to Pax6 (T.Czerny, unpublished data), which may be caused by the presence of a C-terminal inhibitory domain in Pax8 (Dörfler and Busslinger, 1996). Hence, a hybrid transcription factor was created by directly linking the paired domain of Pax8 to the transactivation domain of VP16. Since the transactivation potential of a Pax protein can critically depend on the protein concentration (Czerny and Busslinger, 1995), we performed a series of transfection experiments using increasing amounts of each transactivator plasmid.

As shown in Figure 5A, only the Pax8-VP16 protein lacking the serine insertion (–S) was able to stimulate transcription by binding to the bipartite recognition sequence of the Iε-luc reporter gene. An inverse picture was seen for transcriptional activation of the reporter gene 5aCON-luc (Figure 5B), as only the Pax8-VP16 protein with the serine insertion (+S) could induce transcription of this gene. Hence, the paired domains of Pax8 and Pax8(S) bind *in vivo* in a mutually exclusive manner to their preferred target sequences, i.e. to ‘conventional’ bipartite binding sites (Iε) or to the tetrameric 5aCON sequence. Moreover, transactivation by the Pax8(S)-VP16 protein (Figure 5B) was higher and strictly concentration-dependent compared with Pax8-VP16 (Figure 5A),

indicating that the C-terminal region of the paired domain binds also *in vivo* cooperatively to the 5aCON site. Taken together, these *in vivo* data therefore complement and confirm our *in vitro* binding studies.

Similar, though not identical, transactivation data were obtained with the two full-length Pax6 isoforms. Pax6 activated transcription of the $\text{I}\epsilon$ -luc reporter gene, in comparison with the insertion version Pax6(5a), already at a ~ 10 -fold lower concentration of the expression plasmid (Figure 5C). Moreover, the Pax6(5a) isoform stimulated transcription of the 5aCON-luc gene to significantly higher levels than the Pax6 protein (Figure 5D). Although *in vivo* binding of the two Pax6 isoforms was not mutually exclusive as in the case of Pax8, each isoform preferentially activated transcription from its high-affinity binding site in agreement with the *in vitro* data. The discrepancy between the Pax6 and Pax8 transactivation results could be explained by the different location of the inserted amino acids in the paired domain (see Figure 6A) and/or the additional presence of a homeodomain in Pax6, which may help to stabilize weak DNA interactions of the paired domain. In summary, we conclude that inactivation of the N-terminal DNA-binding motif of the paired domain results in preferential targeting of Pax6(5a) and Pax8(S) to the highly specialized tetrameric 5aCON sequence *in vivo*.

Discussion

Alternative splicing of transcripts from a single gene is often used as a mechanism for generating protein variants with diverse functions (reviewed by McKeown, 1992). In the case of transcription factor genes, alternative splicing frequently gives rise to protein isoforms with distinct or even opposing transcriptional activities (reviewed by Foulkes and Sassone-Corsi, 1992). However, few cases are known where the DNA sequence specificity of a transcription factor is altered by alternative splicing. Pax6 belongs to this class of genes which normally code for transcription factors with modular DNA-binding domains such as the mammalian WT-1, the *Drosophila* Tramtrack and CF2 zinc finger proteins (Bickmore *et al.*, 1992; Gogos *et al.*, 1992; Read and Manley, 1992). Here we have demonstrated that a second Pax gene, Pax8, also codes for alternative splice products with drastically different DNA-binding specificities.

The paired domain is a bipartite DNA-binding region consisting of an N-terminal and a C-terminal subdomain (Czerny *et al.*, 1993; Xu *et al.*, 1995). Each subdomain contains a HTH motif which can potentially bind to DNA (Xu *et al.*, 1995). The novel splice variant Pax8(S) contains an extra serine residue inserted in α -helix 3 of the paired domain (Figures 1A and 6A). This amino acid insertion is expected to inactivate the DNA-binding function of the N-terminal HTH motif by disrupting helix 3 which is responsible for all major groove DNA contacts of the N-terminal subdomain (Xu *et al.*, 1995). Consistent with this notion, Pax8(S) fails to interact with bipartite paired domain recognition sequences similar to the previously characterized isoform Pax6(5a) which contains a 14-amino-acid insertion immediately N-terminal of α -helix 3 (Epstein *et al.*, 1994a; Figure 6A). Both isoforms could, however, bind to the highly specialized recognition sequence 5aCON (Epstein *et al.*, 1994a; this study). In

this context it is interesting to note that purified paired domain peptides of Pax6(5a) and Pax8(S) were unable to interact with a pool of random DNA sequences, while the unmodified paired domains of both Pax proteins could quite efficiently select binding sites from the same DNA mixture (Figure 3A). Collectively, these data therefore demonstrate that the DNA sequence recognition of the paired domain is severely restricted by amino acid insertions in the N-terminal subdomain. This conclusion is supported further by the fact that the 5aCON sequence is the only binding site identified to date for the Pax6(5a) and Pax8(S) isoforms. The 5aCON sequence consists of a unique assembly of four copies of the 5' half-site which we previously characterized as the recognition sequence for the C-terminal region of the paired domain (Czerny *et al.*, 1993). In agreement with this finding, four molecules of each insertion isoform bind to the tetrameric 5aCON sequence (Figures 4 and 6B). Four C-terminal domains can indeed be arranged on the 5aCON sequence without spatial constraints (Figure 6C), as is demonstrated by molecular modelling of the paired domain-DNA interaction (Xu *et al.*, 1995; see Materials and methods). In this model, each polypeptide contacts the 5' half-site in both the major and minor groove of the DNA via the C-terminal recognition helix $\alpha 6$ and the linker region connecting the two subdomains, respectively (Figure 6C). Interestingly, the N-terminal region of the paired domain is capable of binding to a single optimized 3' half-site in the absence of a functional C-terminal region (Czerny *et al.*, 1993; Jun and Desplan, 1996), whereas the C-terminal subdomain depends on cooperative binding to multiple 5' half-sites. Hence, the two DNA-binding regions of the paired domain differ in their interaction with DNA, although both subdomains contain similar HTH motifs (Xu *et al.*, 1995).

Alternative splicing has recently been shown to affect also the paired domains of Pax3 and Pax7 (Vogan *et al.*, 1996). In both cases, a glutamine residue is inserted next to α -helix 4 in the C-terminal subdomain. This glutamine insertion has only a mild effect on the DNA-binding potential of Pax3 and Pax7. In contrast, a glutamine residue introduced at the equivalent position in Pax6 drastically affects DNA-binding of the paired domain by weakening the DNA contacts of the C-terminal subdomain (Vogan *et al.*, 1996). These observations lead to two predictions. First, the C-terminal region of the paired domain appears to make only a minor contribution to the DNA binding of Pax3 and Pax7 in analogy to the corresponding *Drosophila* homologue Paired (Cai *et al.*, 1994; Bertuccioli *et al.*, 1996; Jun and Desplan, 1996). Second, the DNA-binding specificity of Pax proteins can be modified through alternative splicing in two different ways by inserting extra amino acids either into the N-terminal (Pax6 and Pax8) or C-terminal (Pax3 and Pax7) region of the paired domain.

The insertion of a serine codon into the Pax8 mRNA is caused by the presence of an AG dinucleotide at a 3-bp distance from the normal intron 2-exon 3 junction (Figure 1A). This AG dinucleotide constitutes an invariant feature of 3' splice sites (Stephens and Schneider, 1992) and has been conserved in both the human and mouse Pax8 genes consistent with the synthesis of Pax8(S) isoforms in both species. However, the closely related genes Pax2 and

Fig. 6. Model for binding of Pax proteins to bipartite and tetrameric recognition sequences. **(A)** Location of the amino acid insertions within the paired domains of Pax6 and Pax8. The secondary structure of the paired domain as determined by X-ray crystallographic analysis (Xu *et al.*, 1995) is schematically shown together with the position of the inserted amino acids (filled arrowheads). Numbers refer to the respective amino acid residue of the paired domain. Antiparallel β -sheets and α -helices are represented by arrows and boxes, respectively. The three amino acid residues, which are responsible for differential sequence recognition by the Pax6 and Pax8 paired domains (Czerny and Busslinger, 1995), are indicated by grey shading. **(B)** Model explaining the DNA sequence recognition of the paired domain. N- and C-terminal subregions of the paired domain are referred to as N and C, and the 5' and 3' half-sites of the bipartite binding site are indicated by solid and broken lines, respectively. The N-terminal subregion, whose DNA-binding is inactivated in the alternative splice products Pax6(5a) and Pax8(S), is diagrammed as an unstructured domain. **(C)** Molecular model showing the spatial arrangements of four C-terminal subdomains of the paired domain on the 5aCON sequence. The molecular modelling of the protein-DNA interaction was based on the sequence and structural homology of the C-terminal domain (Xu *et al.*, 1995) with the Hin recombinase (Feng *et al.*, 1994). For detailed explanations, see Materials and methods. Only the backbones of both the protein (red) and DNA (green) are shown. As it is best illustrated for the second polypeptide from the left, the C-terminal recognition helix 6 (Xu *et al.*, 1995) contacts the DNA in the major groove (M), while the N-terminal tail (linker) of the same subdomain interacts with the minor groove (m).

Pax5 do not have the potential to code for a similar isoform, as they lack an AG dinucleotide at the equivalent position in intron 2 (Stapleton *et al.*, 1993; Busslinger *et al.*, 1996). The generation of Pax8(S) and Pax6(5a) mRNA by alternative splicing appears to be constitutive rather than regulated during mouse development. While the inclusion of the alternative exon 5a into the Pax6(5a) mRNA is a more complex process involving the splicing of two introns, the generation of the Pax8(S) transcripts depends solely on the differential use of adjacent 3' splice sites. The *Pax8* sequences surrounding the normal intron 2-exon 3 junction (TAG/GT) conform well to the optimal consensus for 3' splice sites [(Py)₁₂₋₁₈NPYAG/GT; Stephens and Schneider, 1992] in contrast to the sequences at the alternative upstream 3' splice site (GAG/TA; Figure 1A). Hence, the spliceosomes may recognize the two 3' splice site with different affinities, thus explaining the higher abundance and constitutive splicing of the Pax8 mRNA relative to the Pax8(S) transcript. Similarly, the synthesis of alternatively spliced transcripts from the *Pax3*, *Pax7* and *I-POU* genes is also constitutive and relies on the differential use of two closely spaced 3' splice sites (Treacy *et al.*, 1992; Vogan *et al.*, 1996). In contrast, the alternative splicing affecting the C-terminal sequences of Pax8 is spatially and temporally regulated during development (Kozmik *et al.*, 1993) and is thus uncoupled from the constitutive splicing events occurring in the paired domain region.

The exon 5a of *Pax6* has been conserved from zebrafish to man, thus suggesting an important function for alternatively spliced insertions in the paired domain (Glaser *et al.*, 1992; Püschel *et al.*, 1992). As these insertions severely restrict the sequence recognition of the paired domain, they may almost entirely switch the DNA-binding mode of certain Pax proteins [such as Pax6(5a)] from the paired domain to the homeodomain. Alternatively, Pax proteins lacking a homeodomain [such as Pax8(S)] may be more readily available for interaction with other proteins in analogy to the mutually exclusive use of either a protein interaction or DNA-binding function by the two alternative splice products of the *Drosophila* I-POU gene (Treacy *et al.*, 1992). Last but not least, insertions in the paired domain may preferentially target Pax proteins to gene control regions containing a 5aCON-like sequence which is suggested by the highly selective binding of Pax8(S) and Pax6(5a) to the 5aCON site *in vivo* in transiently transfected cells. In support of this latter possibility, we have recently identified a Pax6-binding site in the lens-specific enhancer of the mouse γ F-crystallin gene (Lok *et al.*, 1989; Goring *et al.*, 1993) by using the tetrameric consensus recognition sequence (Figure 4A) to search the genome database for 5aCON-like sequences (Z.Kozmik, unpublished data). Interestingly, mutation of this Pax6-binding site interfered with the lens-specific activity of the γ F-crystallin enhancer (Z.Kozmik, unpublished data), indicating that 5aCON-like sequences can indeed mediate transcriptional regulation of endogenous genes by Pax proteins.

Materials and methods

Cell lines

The human kidney cell lines were obtained from American Type Culture Collection (Rockville, MD) and grown as previously described (Kozmik

et al., 1993). The mouse embryonal carcinoma cell line RAC65 (Pratt *et al.*, 1990) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum.

Oligonucleotides

The following oligonucleotides were used in this study:

- 1 5'-GTGGGATCCGGAGGCTGCCAACCC-3'
- 2 5'-CGCAAGCTTTACTGTAAATCGAGGCCAGT-3'
- 3 5'-GCGGGATCCAAAGCTGCGAGTGTCCCTCAG-3'
- 4 5'-GCGAAGCTTGCTGGTTCCTCCCACTGT-3'
- 5 5'-GCGAGCTCGAGAGACTTGGTGGCCACACA-3'
- 6 5'-GCGGGATCCGCGCTCGAGTTCGCCCCCGACCGATGTC-3'
- 7 5'-CGCGAATTTACCCACCGTACTCGTCAAT-3'
- 8 5'-GCGGGATCCCGTGTGAGCCATGGCTGTGAAGCAAGATC-3'
- 9 5'-GCGGTTCGACGATCTCCCAAGCAACATGGTAGGGTCTG-3'
- 10 5'-AATTGGGCACTGAGGAGAGCG-3'
- 11 5'-AATTCGCTCTGCCTCAGTGCCC-3'
- 12 5'-TCGAATCTGAACATGCTCAGTGAATGTTTACTTCTCTCC-3'
- 13 5'-TCGAGGAGAGAGTCAATGAACATTCACTGAGCATGTTTCAAT-3'
- 14 5'-GCGGGATCCACCATGCTCACAACCTCGATCAGAT-3'
- 15 5'-GCGAAGCTTAGTTGTCATTGCCAGGCTGAGCA-3'
- 16 5'-GCGGGATCCAGCTCCAGCATGCAGAACAG-3'
- 17 5'-GAGAAGCTTCTAAGCCAGGTTGCGAAGAAGTCTG-3'
- 18 5'-AGACGGATCCTCTGAA(N)₂₅TCCAGTGGAAATTCGGA-3'
- 19 5'-TCCGAATTCCTACTGGA-3'

Expression plasmids

The coding sequences of the murine Pax6 and Pax6(5a) proteins were amplified from RNA of 12.5-day-old embryos by RT-PCR with primers 1 and 2, and the coding sequences of murine Pax8 and Pax8(S) by RT-PCR using adult kidney RNA and the primer pair 3/4. All cDNAs were cloned into the polylinker of the eukaryotic expression vector pKW10 (Adams *et al.*, 1992) and verified by DNA sequencing. The Pax8-VP16 and Pax8(S)-VP16 fusion constructs were generated by introduction of the appropriate Pax8 sequences (PCR with oligos 3 and 5) and VP16 sequences (PCR with oligos 6 and 7) into the pKW10 vector. The intron 2 sequences of the murine *Pax8* gene were isolated by PCR with the primer pair 8/9, cloned and sequenced on an automated sequencer (Applied Biosystems, model 373A) by primer walking.

Generation of recombinant *Escherichia coli* proteins

All recombinant proteins were expressed in the *E. coli* strain BL21(DE3, pLysS) (Studier *et al.*, 1990) and purified on Ni²⁺-nitriloacetic acid (NTA) agarose (Hochuli *et al.*, 1988). The paired domain sequences of the Pax8 isoforms were amplified by PCR with oligonucleotides 14 and 15 using murine Pax8 or Pax8(S) cDNA templates and were then inserted into the *Bam*HI and *Hind*III sites of pETH-2a (Adams *et al.*, 1992) for expression of the Pax8-pd and Pax8(S)-pd peptides, respectively. The expression vectors for Pax6-pd and Pax6(5a)-pd were generated by cloning PCR fragments amplified from the respective cDNA with the primer pair 16/17.

EMSA analysis

cDNA expression plasmids directing the synthesis of various Pax8 and Pax6 isoforms were transiently transfected into COP-8 cells by the DEAE-dextran method, and cell extracts were subsequently prepared as described (Adams *et al.*, 1992). These extracts were used for electrophoretic mobility shift assay (EMSA) as described in detail by Barberis *et al.* (1989) except that 10 μ g of BSA was additionally included into the binding reaction. The degenerate DNA probe N₂₅ was synthesized by annealing oligonucleotides 18 and 19 followed by second strand synthesis with DNA polymerase I (Klenow fragment) in the presence of 30 μ Ci of [α -³²P]dCTP. The reaction was subsequently chased with an excess of unlabelled dNTPs to ensure complete synthesis of the second strand.

Western blot analysis

10 μ g of nuclear extract proteins prepared from COP-8 cells, which were transiently transfected with Pax8 or Pax6 expression plasmids, were separated by SDS-PAGE and subsequently processed for immunoblotting with affinity-purified anti-BSAP paired domain antibodies (Adams *et al.*, 1992).

RNAse protection analysis

Pax8 and *Pax6* gene transcripts were detected in total RNA (10 μ g) from human kidney cell lines, mouse embryos and adult tissues by RNase protection analysis as described (Vitelli *et al.*, 1988). The mouse Pax8 paired domain probe (Figure 2A) was obtained by cloning a 159-bp

HindIII–*Ba*II fragment of the murine Pax8a(S) cDNA into the blunted *Sal*I and *Bam*HI sites of pSP64. The human Pax8 probe (Figure 2B) was generated by cloning a blunt-ended 166-bp *Ava*II–*Ba*II fragment of the human Pax8a(S) cDNA into the blunted *Hind*III site of pSP64. A 467-bp Pax6 probe (Figure 2C) was amplified from murine Pax6(5a) cDNA by PCR using primer pair 1/17 followed by subcloning into the *Bam*HI and *Hind*III sites of pSP64.

Luciferase reporter constructs

The luciferase construct *luc*–*luc* was generated by insertion of four copies of the annealed oligonucleotides 10/11 upstream of a β -globin TATA box and initiator region linked to a luciferase gene. The reporter gene 5aCON–*luc* was obtained by inserting one copy of the double-stranded oligonucleotide 12/13 upstream of the same luciferase gene.

Transient transfection assays

The luciferase reporter gene (1 μ g), the reference CMV–chloramphenicol acetyltransferase (CAT) gene (1 μ g) and the transactivator plasmids (0.01–1 μ g as indicated in the legend to Figure 5) were transiently transfected into RAC65 cells by the calcium phosphate co-precipitation method. Two days later, the cells were washed twice with phosphate-buffered saline (PBS), and resuspended in 250 mM Tris, pH 7.0. Following three freeze–thaw cycles and a final centrifugation step, the supernatant was directly used for measuring the luciferase and CAT activities.

Molecular modelling

The interaction of the C-terminal subdomain of the paired domain with the 5aCON site was analysed by molecular modelling in a similar manner as described by Xu *et al.* (1995). The modelling (Figure 6C) was based on the sequence and structural homology of the C-terminal domain with the *Hin* recombinase. First, the structure of the entire C-terminal domain (Xu *et al.*, 1995) was superimposed with that of the *Hin* recombinase (Feng *et al.*, 1994). For this purpose, the coordinates of the recognition helix of the *Hin* recombinase (residues 173–179; Feng *et al.*, 1994) were used for positioning of the homologous α -helix 6 of the paired domain (residues 117–123). The flexible linker region of the paired domain (residues 70–77) was subsequently replaced by the structure of the homologous region of the *Hin* recombinase (residues 139–145), resulting in a proper docking arrangement of this region with the minor groove. The coordinates of the DNA in the *Hin* recombinase structure were then used for correct positioning of the 5aCON sequence (in B-DNA conformation) by superimposing the 5' half-site of the paired domain with the recognition sequence of the *Hin* recombinase via their shared CA dinucleotide. This alignment resulted in the major and minor groove contacts predicted for the 5' half-site of the recognition sequence (Czerny *et al.*, 1993). Finally, the same procedure was repeated for the remaining three C-terminal domains which are bound to the 5aCON site. All four domains could be arranged on the 5aCON sequence without spatial constraints (Figure 6C).

Acknowledgements

We are grateful to A.Beyer for help with the molecular modelling, to G.Schaffner for oligonucleotide synthesis, to I.Botto for DNA sequencing, to H.Tkadletz for graphical work and to L.Huber and P.Pfeffer for critical reading of the manuscript. This work was supported by the I.M.P. and, in part, by a grant from the Austrian Industrial Research Promotion Fund.

References

- Adams,B., Dörfler,P., Aguzzi,A., Kozmik,Z., Urbánek,P., Maurer-Fogy,I. and Busslinger,M. (1992) *Pax-5* encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev.*, **6**, 1589–1607.
- Barberis,A., Superti-Furga,G., Vitelli,L., Kemler,I. and Busslinger,M. (1989) Developmental and tissue-specific regulation of a novel transcription factor of the sea urchin. *Genes Dev.*, **3**, 663–675.
- 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.
- Bickmore,W.A., Oghene,K., Little,M.H., Seawright,A., van Heyningen,V. and Hastie,N.D. (1992) Modulation of DNA binding specificity by alternative splicing of the Wilms tumor *wil* gene transcript. *Science*, **257**, 235–237.
- 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 of *Drosophila*. *Cell*, **47**, 1033–1040.
- Busslinger,M., Klix,N., Pfeffer,P., Graninger,P.G. and Kozmik,Z. (1996) Deregulation of *PAX-5* by translocation of the Em enhancer of the *IgH* locus adjacent to two alternative *PAX-5* promoters in a diffuse large-cell lymphoma. *Proc. Natl Acad. Sci. USA*, **93**, 6129–6134.
- Cai,J., Lan,Y., Appel,L.F. and Weir,M. (1994) Dissection of the *Drosophila* *Paired* proteins: 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.
- 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*). *Mol. 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.
- Dörfler,P. and Busslinger,M. (1996) C-terminal activating and inhibitory domains determine the transactivation potential of BSAP (*Pax-5*), Pax-2 and Pax-8. *EMBO J.*, **15**, 1971–1982.
- Dozier,C., Carrière,C., Grévin,D., Martin,P., Quatannens,B., Stéhelin,D. and Saule,S. (1993) Structure and DNA-binding properties of *Pax-QNR*, a paired box- and homeobox-containing gene. *Cell Growth Differ.*, **4**, 281–289.
- Epstein,J.A., Glaser,T., Cai,J., Jepeal,L., Walton,D.S. and Maas,R.L. (1994a) Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. *Genes Dev.*, **8**, 2022–2034.
- Epstein,J., Cai,J., Glaser,T., Jepeal,L. and Maas,R. (1994b) Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J. Biol. Chem.*, **269**, 8355–8361.
- Feng,J.-A., Johnson,R.C. and Dickerson,R.E. (1994) *Hin* recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science*, **263**, 348–355.
- Foulkes,N.S. and Sassone-Corsi,P. (1992) More is better: activators and repressors from the same gene. *Cell*, **68**, 411–414.
- Glaser,T., Walton,D.S. and Maas,R.L. (1992) Genomic structure, evolutionary conservation and aniridia mutations in the human *PAX6* gene. *Nature Genet.*, **2**, 232–239.
- Gogos,J.A., Hsu,T., Bolton,J. and Kafatos,F.C. (1992) Sequence discrimination by alternatively spliced isoforms of a DNA binding zinc finger domain. *Science*, **257**, 1951–1955.
- Goring,D.R., Bryce,D.M., Tsui,L.-C., Breitman,M.L. and Liu,Q. (1993) Developmental regulation and cell type-specific expression of the murine γ F-crystallin gene is mediated through a lens-specific element containing the γ F-1 binding site. *Dev. Dyn.*, **196**, 143–152.
- Grindley,J.C., Davidson,D.R. and Hill,R.E. (1995) The role of *Pax-6* in eye and nasal development. *Development*, **121**, 1433–1442.
- Halder,G., Callaerts,P. and Gehring,W. (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,V. (1994) Mutations at the *PAX6* locus are found in heterogeneous anterior segment malformations including Peters' anomaly. *Nature Genet.*, **6**, 168–173.
- Hill,R.E. *et al.* (1991) Mouse *Small eye* results from mutations in a paired-like homeobox-containing gene. *Nature*, **354**, 522–525.
- Hochuli,E., Bannwarth,W., Döbeli,H., Gentz,R. and Stüber,D. (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate absorbent. *BioTechnology*, **6**, 1321–1325.
- Jun,S. and Desplan,C. (1996) Cooperative interactions between paired domain and homeodomain. *Development*, **122**, 2639–2650.
- Kozmik,Z., Kurzbauer,R., Dörfler,P. and Busslinger,M. (1993) Alternative splicing of *Pax-8* gene transcripts is developmentally regulated and generates isoforms with different transactivation properties. *Mol. Cell. Biol.*, **13**, 6024–6035.
- Lok,S., Stevens,W., Breitman,M.L. and Tsui,L.-C. (1989) Multiple regulatory elements of the murine γ 2-crystallin promoter. *Nucleic Acids Res.*, **17**, 3563–3582.
- McKeown,M. (1992) Alternative mRNA splicing. *Annu. Rev. Cell Biol.*, **8**, 133–155.
- Plachov,D., Chowdhury,K., Walther,C., Simon,D., Guenet,J.L. and Gruss,P. (1990) *Pax8*, a murine paired box gene expressed in the developing excretory system and thyroid gland. *Development*, **110**, 643–651.

- Pratt,M.A.C., Kralova,J. and McBurney,M.W. (1990) A dominant negative mutation of the alpha retinoic acid receptor gene in a retinoic acid-nonresponsive embryonal carcinoma cell. *Mol. Cell. Biol.*, **10**, 6445–6453.
- Püschel,A.W., Gruss,P. and Westerfield,M. (1992) Sequence and expression pattern of *pax-6* are highly conserved between zebrafish and mice. *Development*, **114**, 643–651.
- 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.
- Read,D. and Manley,J.L. (1992) Alternatively spliced transcripts of the *Drosophila tramtrack* gene encode zinc finger proteins with distinct DNA binding specificities. *EMBO J.*, **11**, 1035–1044.
- Stapleton,P., Weith,A., Urbánek,P., Kozmik,Z. and Busslinger,M. (1993) Chromosomal localization of seven *PAX* genes and cloning of a novel family member, *PAX-9*. *Nature Genet.*, **3**, 292–298.
- Stephens,R.M. and Schneider,T.D. (1992) Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J. Mol. Biol.*, **228**, 1124–1136.
- St-Onge,L., Sosa-Pineda,B., Chowdhury,K., Mansouri,A. and Gruss,P. (1997) *Pax6* is required for differentiation of glucagon-producing α -cells in mouse pancreas. *Nature*, **387**, 406–409.
- Strachan,T. and Read,A.P. (1994) *PAX* genes. *Curr. Opin. Genet. Dev.*, **4**, 427–438.
- Stuart,E.T., Kiousi,C. and Gruss,P. (1994) Mammalian *PAX* genes. *Annu. Rev. Genet.*, **28**, 219–236.
- Studier,W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.
- Ton,C.C.T. *et al.* (1991) Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell*, **67**, 1059–1074.
- Treacy,M.N., Neilson,L.I., Turner,E.E., He,X. and Rosenfeld,M.G. (1992) Twin of I-POU: a two amino acid difference in the I-POU homeodomain distinguishes an activator from an inhibitor of transcription. *Cell*, **68**, 491–505.
- Treisman,J., Harris,E. and Desplan,C. (1991) The paired box encodes a second DNA-binding domain in the Paired homeo domain protein. *Genes Dev.*, **5**, 594–604.
- Triezenberg,S.J., Kingsbury,R.C. and McKnight,S.L. (1988) Functional dissection of VP16, the *trans*-activator of herpes simplex virus immediate early gene expression. *Genes Dev.*, **2**, 718–729.
- Vitelli,L., Kemler,I., Lauber,B., Birnstiel,M.L. and Busslinger,M. (1988) Developmental regulation of micro-injected histone genes in sea urchin embryos. *Dev. Biol.*, **127**, 54–63.
- Vogan,K.J., Underhill,D.A. and Gros,P. (1996) An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Mol. Cell. Biol.*, **16**, 6677–6686.
- Walther,C. and Gruss,P. (1991) *Pax-6*, a murine paired box gene, is expressed in the developing CNS. *Development*, **113**, 1435–1449.
- 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.
- Xu,W., Rould,M.A., Jun,S., Desplan,C. and Pabo,C.O. (1995) Crystal structure of the paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell*, **80**, 639–650.

Received on May 27, 1997; revised on August 28, 1997