Structure and DNA-binding Properties of *Pax-QNR*, a Paired Box- and Homeobox-containing Gene[1]

Christine Dozier, Catherine Carrière, Delphine Grévin, Patrick Martin, Brigitte Quatannens, Dominique Stéhelin, and Simon Saule[2]

Centre National de la Recherche Scientifique, URA 1160, Institut Pasteur de Lille, 1 Rue Calmette, 59019 Lille cedex, France

Abstract

After differential screening of a complementary DNA library constructed from quail neuroretina cells (QNR) infected with the v-*myc*-containing avian retrovirus MC29, we have isolated a complementary DNA clone which identifies a mRNA essentially expressed in the neuronal layer of the retina. This complementary DNA encodes a protein containing paired box and homeobox domains. This gene, called *Pax-QNR*, is homologous to the murine *Pax*-6 and human *AN* genes, which are mutated in the autosomal dominant mutation *small eye* (*Sey*) of the mouse and *aniridia* in humans. Here, we report the genomic exon-intron organization, as well as the existence of alternative splicing events taking place at both the 5′ end and the middle part of the gene. A *Pax-QNR* clone translated in reticulocyte lysate directed the synthesis of a 46 kilodalton protein able to bind specifically to the e5 sequence present upstream of the *Drosophila even-skipped* gene and target of the *Drosophila* paired protein. The Pax-QNR paired and homeobox domains individually expressed in bacteria are both able to recognize the e5 sequence.

Introduction

The vertebrate embryonic eye offers exceptional advantages for investigations of the molecular basis of differentiation, since it exhibits a wide repertoire of possible tissue type interconversions *in vivo* and *in vitro*. Cells of the NR[3] have the potential to transdifferentiate into lens cells or pigmented cells (1). Cell differentiation is the product of differential gene expression, and transcription factors may regulate such events. Homeobox-containing genes encoding developmentally regulated transcription factors defined domains in the developing eye (2–4), and previous studies of *Drosophila* homeobox-containing genes have shown that these genes are involved in the process of pattern formation (for reviews, see Refs. 5 and 6). Data from vertebrate systems are consistent with this role (7–9).

The homeobox encodes a conserved DNA-binding domain (10), and several vertebrate homeo proteins have been identified as transcription factors required for expression of lineage-specific genes (11–13). The paired motif found in *Drosophila* developmental genes, such as *paired* and *gooseberry*, has also been shown to be conserved in vertebrates and found to encode another DNA-binding motif (14, 15). This motif is present alone (in the murine genes *Pax*-1, *Pax*-2, and *Pax*-8) or together with the homeobox domain (*Pax*-3, *Pax*-4, *Pax*-6, and *Pax*-7) (14). In contrast to homeobox-containing genes whose expression is region specific, the *Pax* genes are expressed along the entire anteroposterior axis in the neural tube (16). Several developmental mutations in the mouse have been associated with mutations in the *Pax* genes (17). Mutations in the *Pax*-6 gene have been associated with the mouse mutation *small eye* (*Sey*) (18). The same gene has been found to be deleted in some cases of the human congenital disorder *aniridia* (*AN*), which is a similar disorder to *small eye* (19), strengthening the role of this gene in eye development.

We have previously isolated from MC29 (an avian retrovirus bearing the v-*myc* oncogene) infected quail NR cells (QNR-MC29), a cDNA clone expressed in the neuroretina and in the cerebellum of chicken and quail embryos, but not in the pigmented retina. *In situ* hybridization performed in developing neuroretina reveals strong expression of *Pax-QNR* in the ganglionic cell layer and in the lower part of the inner nuclear layer. The derived amino acid sequence of this cDNA shows paired box and homeobox sequences, and the corresponding gene, which has been named *Pax-QNR*, appears to be the quail homologue of the recently described *Pax*(*zf-a*) gene from zebrafish (20) and *Pax*-6 from the mouse (3). To understand the evolution of the *Pax* family and to facilitate the analysis of their transcriptional regulation, the chromosomal genes must be cloned. To date, no complete genomic organization of a *Pax* gene has been reported.

In this paper, we show that the quail *Pax-QNR* gene is composed of a transcriptional unit spanning more than 17 kbp of genomic DNA and comprising at least 15 exons. Both the paired box and the homeobox domains are encoded by three exons. Alternative splicing takes place both at the 5′ end and in the middle of the gene. The expected products each have a distinct amino-terminus and, due to the use of an alternative splice acceptor, differ by 6 amino acids in the first exon following the homeobox domain. The *Pax-QNR* gene encodes at least a 46 kd protein that specifically binds to the e5 sequence present upstream of the *Drosophila even-skipped* gene. This e5 sequence is recognized *in vitro* by the *Drosophila* paired and the murine Pax-2, Pax-3, and Pax-5 proteins (15, 21–23) but not by Pax-1 and Pax-8 proteins (24, 25). When expressed in bacteria, both the Pax-QNR paired

Received 11/12/92; revised 1/6/93; accepted 1/11/93.

[1] This work was supported by grants from the Centre National de la Recherche Scientifique, Institut Pasteur de Lille, and the Association Française Retinitis Pigmentosa.

[2] To whom requests for reprints should be addressed.

[3] The abbreviations used are: NR, neuroretina; QNR, quail NR; cDNA, complementary DNA; bp, base pair(s); kbp, kilobase pair(s); kd, kilodalton(s); ORF, open reading frame; 5′-UTR, 5′-untranslated region; nt, nucleotides; SDS, sodium dodecyl sulfate; EMSA, electrophoretic mobility shift assay; PCR, polymerase chain reaction.

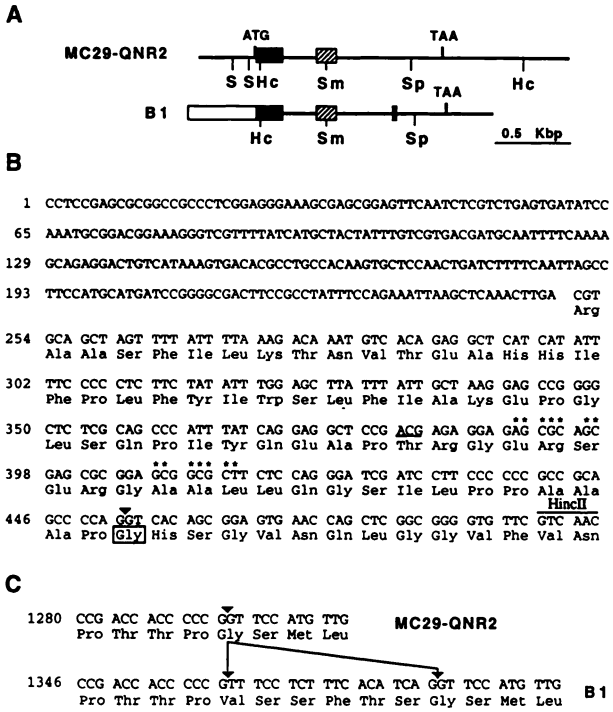


Fig. 1. *A*, structure of the two *Pax-QNR* cDNAs isolated. *White box*, the B1-specific 5′-UTR; *gray box*, paired domain; *hatched box*, homeodomain; *black box*, the 6 amino acids found in clone B1 but absent in clone MC29-QNR2. Initiation and stop codons are indicated. Restriction enzymes used: *S*, *Sst*II; *Hc*, *Hinc*II; *Sm*, *Sma*I; *Sp*, *Sph*I. *B*, nucleotide and deduced amino acid sequence of the B1-specific 5′-end. *Arrowhead*, position of intron found in the *Pax-QNR* gene. The first paired domain amino acid is *boxed*. *Asterisks*, the potential stem-loop structure arising from the inverted repeats. The potential translation codon ACG is *underlined*. *C*, sequence comparison between MC29-QNR2 and B1 cDNAs showing the 6-amino acid insertion (*arrows*) in the B1 clone due to an alternative splice acceptor site in exon 10.

domain and homeodomain peptides recognize the e5 sequence. Thus, despite the fact that *Pax-QNR* bears the most divergent paired box and homeobox sequences in the *Pax* family, its product shows the same *in vitro* DNA-binding specificity as the other *Pax* gene products already characterized.

Results

Isolation and Nucleotide Sequence Comparison of *Pax-QNR* cDNAs. We previously described the isolation of a cDNA clone (MC29-QNR2) corresponding to a *Pax* gene (*Pax-QNR*) expressed in the neuronal layers of the retina. The sequence of MC29-QNR2 (2509 bp) has been already published along with the derived sequence of the longest open reading frame (4). The deduced protein (416 amino acids) revealed a paired domain (Fig. 1*A*, *gray box*) and a paired-type homeodomain (Fig. 1*A*, *hatched box*). The paired domain resides close to the amino-terminus of the protein, as do all the paired box-containing genes isolated to date.

Subsequently, by screening a quail spinal cord cDNA library using MC29-QNR2 as a probe, we isolated another cDNA clone, named B1 (Fig. 1*A*). Nucleotide sequence analysis of cDNA B1, as compared with the MC29-QNR2 clone, shows two differences. First, the 5′ ends of the two cDNAs diverge, and no AUG can be found in the ORF used to translate the B1 deduced protein (Figs. 1*B* and 2). Second, in the middle of the gene, a stretch of 18 nucleotides, also found in zebrafish (20), murine (3), and human species (19), is present in clone B1. This insert encodes 6 amino acids (Val–Ser–Ser–Phe–Thr–Ser) inserted at amino acid 306 of the deduced MC29-QNR2 protein (Fig. 1*C*). No AUG was found to translate the B1 ORF, but an ACG in a nonoptimal in-frame sequence (GCTCCG**ACG**A, *underlined* in Fig. 1*B*) could initiate the protein. This ACG is followed by a potential stem-loop present 7 nucleotides downstream. If the stem-loop hinders ribosome scanning appropriately to promote initiation at the in-frame ACG site (26), the polypeptide encoded by B1 cDNA, and starting with a threonine, would contain 24 amino acids upstream from the paired domain, whereas in the MC29-QNR2 cDNA, only 3 amino acids precede the paired domain. However, as observed for the other species (murine, fish, and human), the MC29-QNR2 ORF can be extended 5′ of the putative translational start site (ATG). When compared to the murine Pax-6, the zebrafish Pax(zf-a), or the human AN deduced protein, the extended amino-terminus of Pax-QNR appears divergent (Fig. 2). Stretches of homology can be found with murine and human species, although no homology can be found with the reported zebrafish sequence in this extended ORF. The observed differences can be due either to distinct alternative splicing in the 5′-UTR of these genes or to the divergence accumulated during species evolution.

Nevertheless, differences observed between the two *Pax-QNR* cDNAs strongly suggested that cDNA B1 could represent an alternative splice product of the *Pax-QNR* gene. In order to clarify this point, we characterized the genomic *Pax-QNR* gene.

Isolation of Genomic Clones and Characterization of the *Pax-QNR* Locus. Three phages, named λPax-QNR1, λPax-QNR2, and λPax-QNR3, were isolated from two different quail genomic libraries using parts of the *Pax-QNR* cDNA or genomic fragments as probes (Fig. 3*A*). Inserts containing the different exons of *Pax-QNR* were isolated, subcloned, and sequenced. Results obtained show that the *Pax-QNR* gene is organized in at least 15

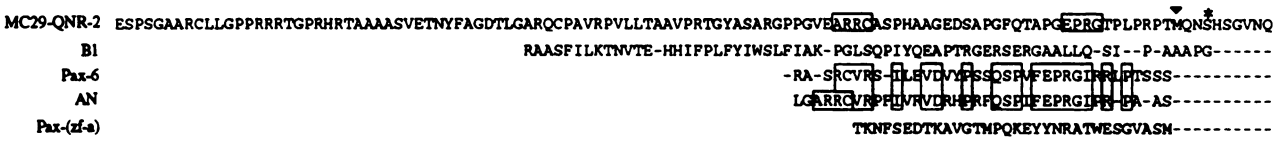


Fig. 2. Comparison of the Pax-QNR protein sequence extended in the 5-UTR with those from other species. Sequences were from mouse Pax-6 (3), human AN (19), and zebrafish Pax(zf-a) (20). *Boxes*, conserved amino acids; *arrowhead*, the putative initiation Met; *asterisk*, the first paired domain amino acid differing in the two quail cDNA clones as the result of the alternative splicing.

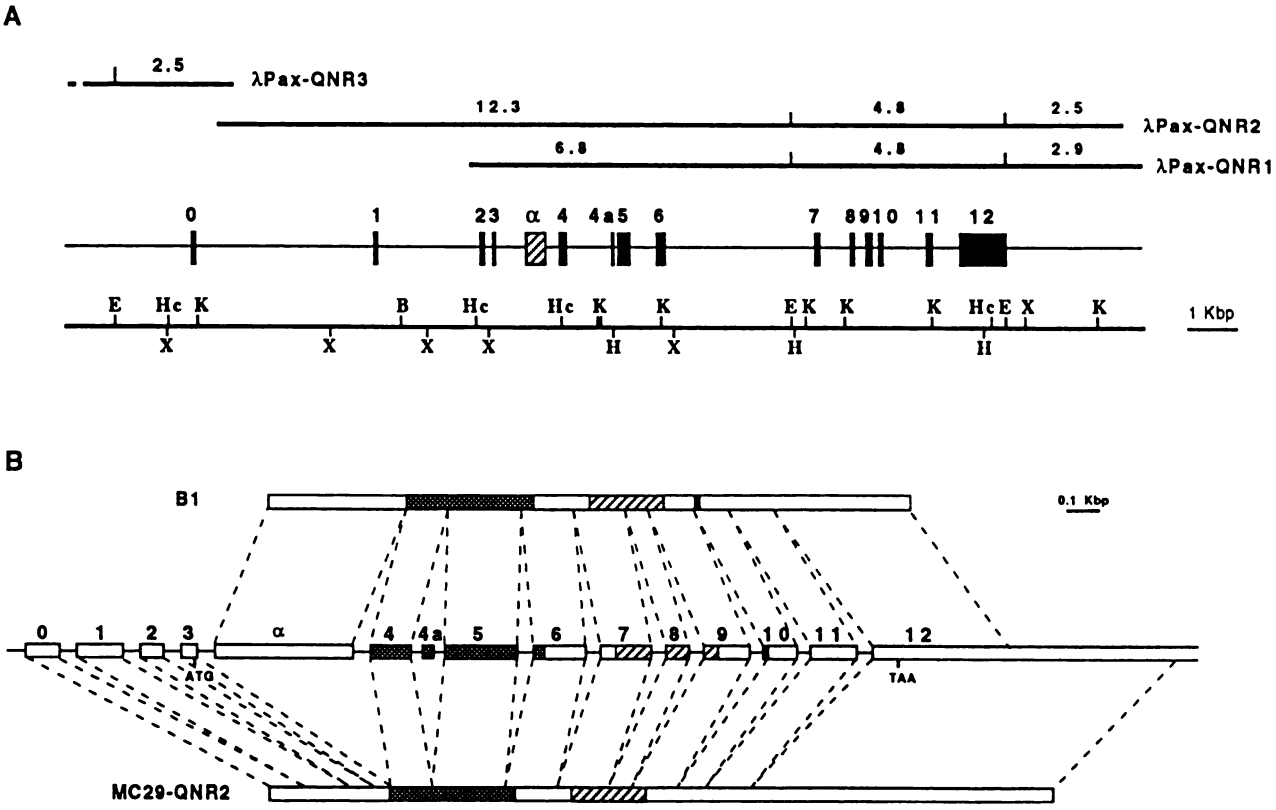


Fig. 3. Schematic diagram of the structural organization of the *Pax-QNR* locus. *A*, genomic organization of *Pax-QNR*. At *top*, the three overlapping genomic phages are aligned with respect to *Pax-QNR* locus. Only part of the λPax-QNR3 is presented. The exon/intron distribution is shown *below*. *Black boxes*, exons numbered from *0* to *12*; *exon 4a* represents an alternative exon identified in murine cDNA (3); *hatched box*, B1-specific exon denoted *α*. Restriction enzymes used: *X*, *Xba*I; *B*, *Bam*HI; *Hc*, *Hinc*II; *K*, *Kpn*I; *H*, *Hind*III; *E*, *Eco*RI. *B*, alternative splicing of the *Pax-QNR* gene. The assumed initiation codon of *Pax-QNR* is in exon 3, and the TAA stop codon is in exon 12. *Gray box*, paired domain; *hatched box*, homeodomain; *small black box*, the 6-amino acid insertion in the B1 clone, due to an alternative splice acceptor in exon 10.

exons extending over 17 kbp of cellular DNA (Fig. 3*A*). The MC29-QNR2 cDNA contains 13 exons, and the B1 cDNA contains 10 exons (Fig. 3*B*). We have mapped three noncoding exons (named exons 0, 1, and 2) in the 5′ part of MC29-QNR2 cDNA, assuming that the Pax-QNR protein synthesized from the MC29-QNR2 mRNA starts with the first ATG in exon 3. Thus, exon 3 is expected to be the first coding exon of the *Pax-QNR* gene. The first B1 exon (exon *α*), devoid of an ATG to initiate the Pax-QNR ORF, ends with a classical splice donor consensus sequence (see Table 1 for exon/intron boundaries) and is fused with exon 4, the first paired domain exon. As a result of the splicing process, the first amino acid of the paired domain encoded by B1 cDNA is a glycine (Fig. 1*B*) instead of a serine, as in all cDNAs isolated to date (see Fig. 2).

The sequence of the genomic *Pax-QNR* gene shows that the paired domain is encoded by three exons (exons 4, 5, and 6) (Fig. 3*B*). The first exon (exon 4) encodes the first helical structure of the paired domain (amino acids 23–31), which plays a major role in the DNA binding activity of the paired protein (15). The two helix-turn-helix motifs found in the carboxy-terminal part of the paired domain (amino acids 80–89 and 94–106) (14) are encoded by the second paired exon (exon 5). The conserved 115–120 amino acids encoded by the third exon (exon 6) have no defined function but probably are important because of their complete conservation among the different paired domains. Thus, each exon may encode a distinct functional motif in the paired domain. An insertion of 42 bp encoding 14 amino acids has been found in the *Pax-6* paired domain of some murine cDNAs (3). This insertion is located at the end of the first paired exon, suggesting that these 42 bp arise from an alternative splicing event. Therefore, we examined the nucleotide sequence of the intron between *Pax-QNR* exons 4 and 5 (paired exons 1 and 2) for the presence of this alternative exon. A sequence homologous to these 42 bp (ACCCATGCAGATGCAAAAGTCCAAGTGCTGGACAATCAAAAC) was found 150 bp upstream of the fifth exon (named exon 4a; Fig. 3*B*). Two nucleotide differences with the murine sequence were found in this sequence (*underlined letters*). At the protein level, these resulted only in the change of a Gln residue into an Asp residue.

As for the paired domain, the homeodomain is also encoded by three exons (exons 7, 8, and 9) (Fig. 3*B*). The homeodomain contains a helix-turn-helix sequence specifying a DNA-binding motif (27). Of the three helix motifs, helix 1 and helix 2 are the less conserved. Helix 1 is encoded by the end of exon 7 and the beginning of exon 8. Exon 8 also encodes the second helix and the

Table 1 Nucleotide sequences of exon-intron boundaries

	Intron		Exon		Intron
			Exon 0	.....CGCCGGACAG	gtaacg
	ctgtctccttcccag	GACCTCGGCA.....	Exon 1	.....CCTCTGCGAG	gtgagt
	tctgctttgtcctag	AGGGCCGCCG.....	Exon 2	.....CGGCTTCCAG	gcaagt
	gtttcgttgccgcag	ACCGCTCCAG.....	Exon 3	.....ATGCAGAACA	gtaagt
			Exon α(B1)	.....GCAGCCCCAG	gtcggc
	atttttgtgttatag	ACCCATGCAG.....	Exon 4a	.....CAATCAAAAC	gtaagc
	ctgtgcttcccgcag	GTCACAGCGG.....	Exon 4	.....AATCCTGCAG	gtgaag
	cgcttccttatccag	GTGTCGAATG.....	Exon 5	.....CATACCCAGT	gtaagt
	tcctttgccttgcag	GTGTCGTCGA.....	Exon 6	.....CCCGCACAAG	gtgaga
	ctctccttcccacag	ACGGCTGCCC.....	Exon 7	.....CTTGAGAAAG	gtgagc
	tacttctccttgcag	AGTTTGAGAG.....	Exon 8	.....AAGGATACAG	gtacgg
	acgctgtgcttccag	GTGTGGTTTT.....	Exon 9	.....ACCACCCCCG	gtaatg
	tctattctattatag	TTTCCTCTTT.....	Exon 10(B1)	.....GCCTATGCAA	gtaagc
	cctctttcacatcag	GTTCCATGTT.....	Exon 10 (MC29-QNR2)	.....GCCTATGCAA	gtaagc
	gctctatctttgcag	CCCCCGGTAC.....	Exon 11	.....ACTTCCACAG	gtgagc
	ttttcgg	GTCTCATTTC.....	Exon 12	.....	
Consensus sequences	$SA\begin{pmatrix}t\\c\end{pmatrix}_{15\text{-}30}{}^{t}_{c}n^{t}_{c}ag$	G^{T}_{G}			$SD\,{}^{A}_{C}AG$ $gt^{a}_{g}agt$

beginning of the third helix. The third helix of the paired-type homeodomain is highly conserved among all *Pax* genes, and the serine at position nine, which confers sequence-specific recognition to the homeodomain (10), is encoded by exon 9.

Exon 10 differs in MC29-QNR2 and B1 cDNAs. A sequence of 18 bp coding for 6 amino acids (Val–Ser–Ser–Phe–Thr–Ser) is present in clone B1 precisely at the junction between exons 9 and 10 of cDNA MC29-QNR2 (Table 1; Figs. 1C and 3*B*). All other sequences reported to date contained the B1-like exon 10. Analyses of the genomic sequences reveal that this difference between the two cDNAs comes from the use of an alternative splice acceptor in the *Pax-QNR* exon 10 (Table 1). To examine whether Pax-QNR mRNAs expressed in neuroretina cells contained one, the other, or both types of exon 10, total RNA extracted from QNR was retrotranscribed using oligo(dT) as a primer. The single-strand DNA molecules obtained were used as templates for PCR experiments using oligonucleotide primers corresponding to sequences present in exons 9 and 10. The resulting bands were subjected to *Hpa*II restriction enzyme hydrolysis. This enzyme was chosen because, due to the splicing event, an additional *Hpa*II site is present within the fragment resulting from the PCR amplification of MC29-QNR2 mRNA. After enzymatic digestion, we found both B1-like (*Hpa*II uncut) and MC29-QNR2-like (*Hpa*II cut) fragments, suggesting that neuroretina cells expressed both types of exon 10 (data not shown). Part of the protein encoded by exons 9, 10, 11, and 12 is particularly rich in proline, serine, and threonine amino acids. The last identified exon of the *Pax-QNR* gene (exon 12) is more than 1 kbp long. Since the *Eco*RI site present at the 3′ end of the genomic DNA is also present at the end of the longest cDNA isolated, the exact size of the noncoding sequence in exon 12 is unknown.

Alternative Splicing Takes Place in the 5′ Part of the *Pax-QNR* Gene *in Vivo*. Since B1 cDNA was found without any AUG to initiate the Pax-QNR open reading frame, we performed RNase protection experiments in order to look for the existence of exon α-containing mRNA *in vivo*. We subcloned a *Hin*cII fragment encompassing the 5′ part of the B1 cDNA (exon α and part of exon 4) into the plasmid pGEM4 (Fig. 4*A*). Similarly, we subcloned the 5′ *Eco*RI-*Hin*cII fragment of the MC29-QNR2 cDNA (containing the exons 0, 1, 2, 3, and part of 4) (Fig. 4*B*). Antisense RNA probes (637 nt long for B1 cDNA and 455 nt long for MC29-QNR2 cDNA) were synthesized and hybridized to total RNA extracted from various quail tissues or from QNR cells. After RNase digestion of the hybridized probes, the mRNAs corresponding to B1 and MC29-QNR2 cDNAs could be detected by full-length protected fragments of 490 and 424 nt, respectively. Results obtained show that a mRNA containing exon α (Fig. 4*A*, *arrow*) is detected in neuroretinas (MC29 infected or not) and to a lesser extent in the cerebellum, but not in heart or brain (Fig. 4*A*). Although a similar tissue distribution is observed with the MC29-QNR2 mRNA, surprisingly, very little of this mRNA is detected in neuroretinas (Fig. 4*B*). However, with this probe, we detect several partially protected fragments (Fig. 4*B*, *arrows*). These fragments correspond to alternatively spliced Pax-QNR mRNAs that differ in their 5′-UTR (see "Discussion").

***Pax-QNR* Encodes a DNA-binding Protein.** In order to analyze the size of the protein encoded by MC29-QNR2 cDNA, *in vitro* synthesized MC29-QNR2 RNA was translated in rabbit reticulocyte lysate in the presence of [^{35}S] methionine. Results revealed a major labeled protein present in a doublet band of 46/47 kd specific for the MC29-QNR2 cDNA (Fig. 5*A*, *Lane b*). The apparent molecular weight of 46,000 for Pax-QNR protein in a SDS-polyacrylamide gel is consistent with the use of the first ATG in the cDNA to translate a protein of 416 amino acids with a predicted molecular weight of 45,800.

Drosophila paired protein has been shown to bind and footprint *in vitro* a region upstream of the *Drosophila even-skipped* gene, known as the e5 sequence (15). The 5′ part of this sequence, containing the ATTA motif, binds the homeodomain of the protein, whereas the 3′ part of the sequence binds the paired domain (15). The e5 sequence is also recognized by the murine Pax-2 (22), Pax-5 (23) (each containing only a paired domain), and Pax-3 gene products (21). To determine whether Pax-QNR (which differs in many positions from the other paired domain and homeodomain proteins) was able to specifically recognize the e5 sequence, EMSAs were performed using Pax-QNR protein synthesized *in vitro*

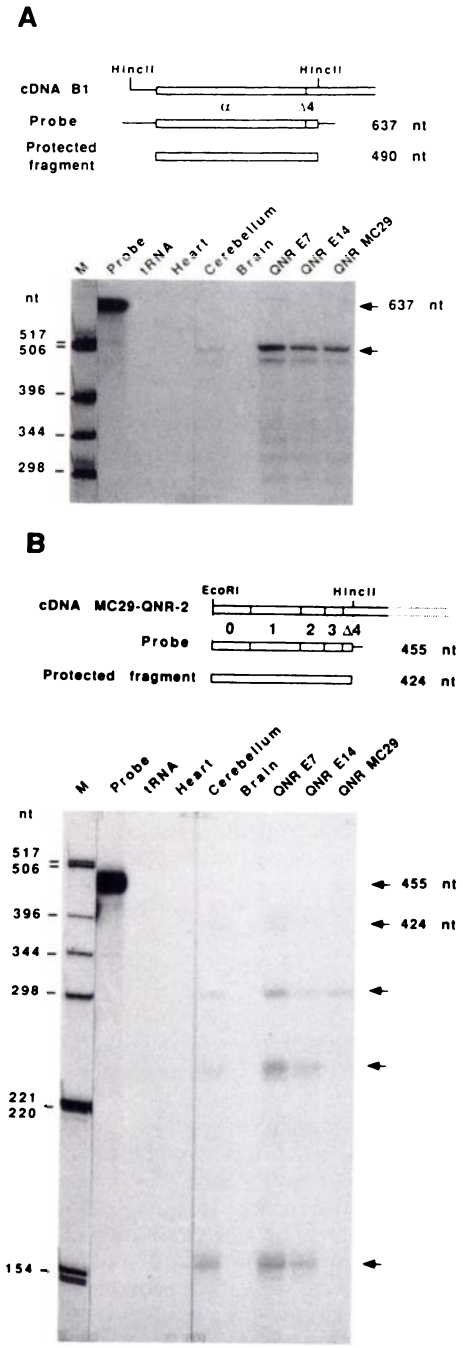


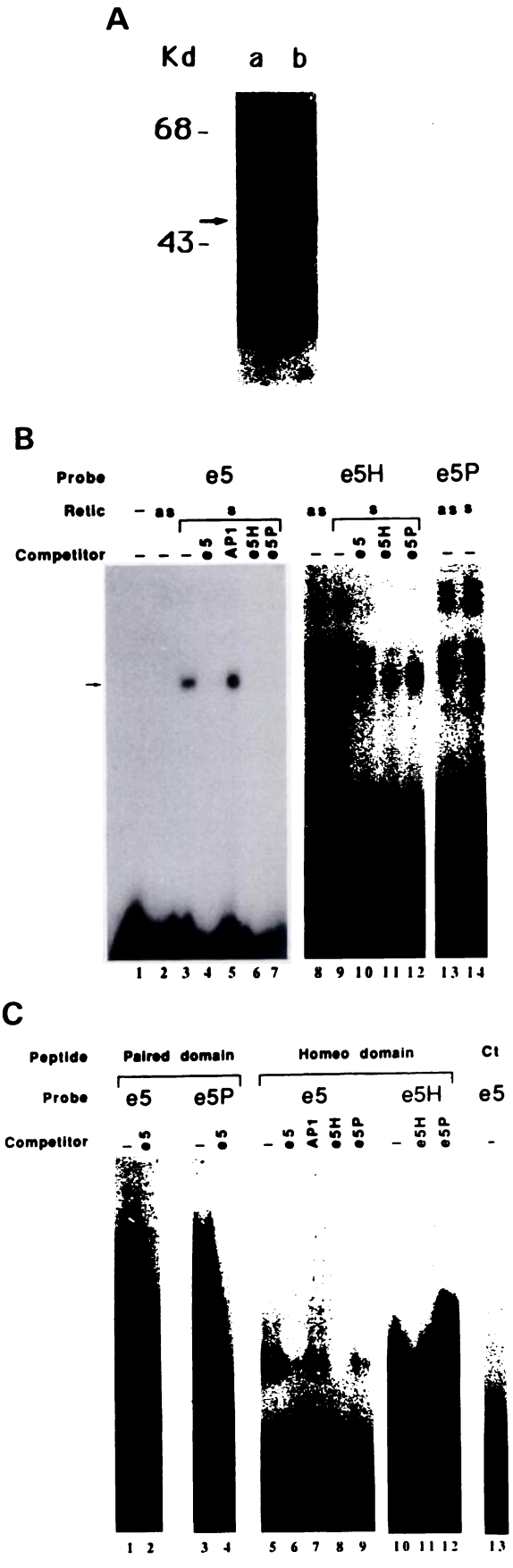
Fig. 4. RNase protection experiments. *A, top: open box,* B1 fragment subcloned in pGEM4 plasmid; *bold line,* linker sequence. The size of the expected RNase-protected fragment is indicated. *Bottom:* the probe was hybridized to 10 μg of total RNA. Yeast tRNA was used as a control of RNase digestion; *heart,* 3-week-old quails; *cerebellum* and *brain,* 14-day-old quail embryos; *QNR E7* and *E14,* neuroretinas from 7- and 14-day-old quail embryos, respectively; *QNR MC29,* quail neuroretina cells infected by MC29 virus; *M,* end-labeled *HinfI*-digested pBR322. *B,* same experiment as in *A,* but with a RNA probe derived from an MC29-QNR2 cDNA fragment. *Arrows,* the protected fragments.

and ^{32}P-labeled e5 DNA (Fig. 5*B*). Very little endogenous e5 binding activity was found in reticulocyte lysates containing the antisense Pax-QNR RNA (*Lane 2*). Expression of Pax-QNR protein resulted in an e5 DNA-binding complex (*Lane 3*). This complex was specific for the e5 sequence, since it disappeared in the presence of a 100-fold excess of unlabeled e5 oligonucleotide (*Lane 4*). In contrast, a 100-fold excess of an oligonucleotide containing the AP-1 consensus recognition sequence failed to compete with the e5 sequence for Pax-QNR binding (*Lane 5*). To determine the nature of the Pax-QNR protein interaction with the e5 sequence, we performed mutational analysis of the homeo (e5P) or the paired (e5H) recognition sites. When each of the mutated oligonucleotides, e5P and e5H, was used as competitor, both competed for Pax-QNR binding on the e5 sequence (*Lanes 6* and *7*). These mutated oligonucleotides were then tested for binding to Pax-QNR protein. The oligonucleotide mutated in the paired binding site (e5H) produced a complex with Pax-QNR as efficiently as the e5 oligonucleotide (*Lane 9*). e5H binding was efficiently competed with a 100-fold excess of unlabeled e5 (*Lane 10*) or e5H (*Lane 11*) oligonucleotides. The oligonucleotide mutated in the homeodomain binding site (e5P) exhibited only negligible binding with Pax-QNR (*Lane 14*) and weakly competed with the e5H sequence for Pax-QNR binding (*Lane 12*).

Both the Paired Domain and the Homeodomain of Pax-QNR Are Individually Able to Bind DNA. The results obtained with the mutated e5 oligonucleotides suggested an influence of both the paired domain and the homeodomain on the DNA binding ability of the Pax-QNR protein. To ascertain the DNA binding ability of each domain, we expressed separately the paired domain and the homeodomain in bacteria and tested the DNA binding properties of each gel-purified peptide in EMSA (Fig. 5*C*). Results show that the paired peptide binds strongly to e5 (*Lane 1*), and this binding is competed with a 100-fold excess of unlabeled e5 oligonucleotide (*Lane 2*). However, in contrast to the full Pax-QNR protein, the paired peptide binds efficiently to e5P (e5 oligonucleotide mutated in the homeodomain binding site; *Lane 3*) and very weakly to e5H (e5 oligonucleotide mutated in the paired domain binding site; data not shown). The homeodomain also binds to e5 (*Lane 5*), and this binding is competed with a 100-fold excess of unlabeled e5 (*Lane 6*) or e5H (*Lane 8*) oligonucleotides but is not competed with a 100-fold excess of unlabeled e5P (*Lane 9*) or AP-1 (*Lane 7*) oligonucleotides. As expected, the homeodomain peptide binds efficiently to e5H (*Lane 10*), and this binding is competed with e5H (*Lane 11*) but not with e5P (*Lane 12*) oligonucleotides. As a control, the carboxy-terminal part of Pax-QNR protein was unable to bind the e5 oligonucleotide (*Lane 13*), demonstrating that the 99 MS2 amino acids fused with each of the peptide domains are devoid of DNA binding properties. Thus, the paired domain and the homeodomain are individually able to bind specifically to the e5 oligonucleotide.

Discussion

Genomic Organization of *Pax-QNR*. We have previously described the isolation of *Pax-QNR* (4), the quail homologue of the gene recently described as *Pax(zf-a)* in zebrafish, *Pax-6* in mouse, and *AN* in humans. We have determined the exon-intron organization of this gene. The 5′-UTR of *Pax-QNR* appears variable, since we have isolated two cDNAs showing distinct 5′-UTRs. One of these cDNAs, clone B1, does not contain any AUG to



initiate the Pax-QNR protein. However, we believe that this cDNA is not an artifact because the nucleotide sequence of the first B1 exon (exon α) is identical in the cDNA and the genomic DNA, and because RNase protection experiments using a probe containing this exon demonstrate the presence of this latter sequence in mRNAs of neuroretinas. Moreover, experiments show that the MC29-QNR2 and B1 mRNAs are transcribed from two distinct promoters located, respectively, upstream from exons 0 and α. Recent reports suggest that non-AUG translation initiation occurs much more frequently in eukaryotic cells than was previously anticipated (26). *In vitro* translation in a reticulocyte lysate of B1 RNA indicates that this RNA is able to direct the synthesis of a Pax-QNR product (data not shown). However, the codon used to initiate this protein has not yet been defined. Further studies will be necessary to address this issue. We do not know whether other species also contain exon α or whether this particular transcript is avian specific. The previously isolated cDNA, MC29-QNR2, shows one in-frame AUG translation start site, which places the paired domain very close to the amino-terminus of the protein. Analysis of the sequence revealed that the open reading frame extended upstream from the first AUG. The conservation of this extension in all of the species studied (fish, quail, mouse, and human) suggests a possible role for the upstream part of the open reading frame, but only mouse and human cDNAs showed clear homology in the protein deduced from the extended reading frame. This finding implies that this part does not play a critical role in the biological function, or that the function played by this part of *Pax-QNR* is species specific. However there is no AUG codon that could be used to translate this extended open reading frame, a situation already found for other transcription factors (28). The 5′-UTR of MC29-QNR2 cDNA is organized in 4 exons (exons 0, 1, 2, and 3, the latter containing the initiation AUG). RNase protection experiments performed on neuroretina RNA using this 5′-UTR as a probe reveal the expected fully protected band, as well as several partially protected bands, the sizes of which could correspond to mRNAs generated from alternative splicing events arising in the 5′ part of the *Pax-QNR* gene. This assumption is confirmed by the isolation of two other cDNAs, one showing an exon 0-exon 2 junction, and the other containing a 5′ extended exon 1. The mouse and the human cDNAs also exhibited multiple 5′-UTRs (3, 19), suggesting a complex translational regulation of this gene in these species.

The paired box domain of *Pax-QNR* is encoded by three exons, like all other known murine paired boxes

Fig. 5. A, in vitro translated Pax-QNR protein. The MC29-QNR2 cDNA was transcribed *in vitro* in a sense and antisense orientation and translated in a rabbit reticulocyte lysate. Translated proteins were separated on a 10% acrylamide gel. *Lane a*, antisense RNA; *Lane b*, MC29-QNR2 sense RNA. *B*, Pax-QNR binding to the *even-skipped* e5 sequence. Reticulocyte lysate expressing the Pax-QNR protein (*Retic*) and ^{32}P-labeled double-stranded oligonucleotide were used for each EMSA reaction. *s*, sense; *as*, antisense. *e5H*, e5 oligonucleotide mutated in the paired binding site; *e5P*, e5 oligonucleotide mutated in the homeodomain binding site. Unlabeled competitor DNA was used in a 100-fold molar excess. *C*, binding of bacterially expressed paired domain and homeodomain peptides to e5 and mutated oligonucleotides. The peptides encoded by the paired, homeo-, and carboxy-terminal part (*Ct*) of the Pax-QNR protein were purified after electrophoresis of the bacterial proteins on a 15% SDS-polyacrylamide gel and used as in *B*.

with the exception of *Pax*-1. (14). The intron-exon boundaries are also conserved in the paired box domain between quail and mouse. The murine *Pax*-6 cDNA previously described (3) shows an insertion of 14 amino acids between the first and the second exon of the paired box domain with respect to the *Pax-QNR* sequence. We found a highly homologous sequence in the first paired box intron of *Pax-QNR* locus (between exons 4 and 5), and this sequence exhibits a classical exon-intron boundary. Therefore, we called this sequence exon 4a. However, we have not isolated any quail cDNA containing this exon. The genomic structure of the human *Pax*-6 gene exhibited similar organization (29).

The homeodomain of *Pax-QNR* is encoded by three exons. Thus, the *Pax-QNR* homeobox is distinct from *Pax*-3 and *Pax*-7 genes at the structural level, since no intron in the homeodomain has been reported for these genes (21, 30). Nucleotide sequence comparison of the two cDNAs isolated reveals that an alternative splicing event takes place in the exon following the homeodomain (exon 10). Indeed, a stretch of six amino acids (Val-Ser-Ser-Phe-Thr-Ser) was added in the protein encoded by cDNA B1. Serine and threonine amino acids are potential targets for cellular kinases, and phosphorylation events may modulate the activity of the two Pax-QNR products. Further studies will be necessary to understand the biological relevance of this particular splicing event in exon 10.

DNA-binding Properties of Pax-QNR Product. The Pax-QNR protein is able to bind to the *Drosophila* e5 sequence like Pax-2, Pax-3, and Pax-5 proteins (21-23), and this binding is most probably mediated by the paired domain and the homeodomain, as it as been demonstrated for the Pax-3 protein (21). From their degree of homology in conjunction with class-specific amino acids at certain positions, the paired domains of the Pax proteins fall into six different classes (14). Three distinct binding specificities of paired domains can be defined by the following DNA sequences: e5 (bound by Prd, Pax-3, Pax-2, and Pax-5), PRS-1 (bound by Pax-1), and CT (bound by Pax-8) (25). Prd and Pax-3 fall into class II, but Pax-2, Pax-5, and Pax-8 fall into class III, although showing a distinct DNA-binding recognition sequence. Pax-QNR/Pax-6 is the first member of the class IV-type paired domain, and we found that an efficient binding on e5 is obtained with Pax-QNR protein expressed in reticulocyte lysates. These data show that Pax proteins of different subclasses exhibit different sequence-binding specificities, due to amino acid changes in their paired domains, and that distinct sequence-binding specificities can be found in the same class as well as identical sequence-binding specificities in distinct classes. Therefore, the degenerated sequence recognition emphasizes the limited value of consensus sequences and core motifs for Pax proteins.

The fact that an efficient binding on e5 is obtained with Pax-QNR protein expressed in reticulocyte lysates suggests that Pax-QNR binding does not require the presence of additional proteins; similar results have been obtained for the Pax-5 protein (25). However, we have been unable to obtain DNA binding of the protein encoded by cDNA B1, translated in reticulocyte lysates, using e5 as probe (data not shown). This suggests that some modifications affecting the paired domain may be critical for the binding of this particular Pax-QNR protein isoform or that accessory proteins are needed for the DNA binding activity of this B1 Pax-QNR.

Modification of the homeodomain binding site in e5 results in a mutated oligonucleotide (e5P) unable to bind Pax-QNR, but able to compete with wild-type e5 for Pax-QNR binding. This indicates that Pax-QNR requires both homeodomains and paired domains for high-affinity binding, and that binding by the paired domain alone results in the formation of a transient complex, unstable under EMSA conditions. Modification of the paired domain binding site in e5 results in a mutated oligonucleotide (e5H), able to bind Pax-QNR. Thus, paired domains and homeodomains of Pax-QNR proteins exhibit distinct affinity for the e5 sequence. Such a situation also exists for another paired domain- and homeodomain-containing *Pax* gene, namely *Pax*-3. However, for *Pax*-3, oligonucleotides containing the paired domain recognition sequence alone exhibited weak binding, whereas the homeodomain recognition sequence alone gave no specific complexes with the Pax-3 protein (21).

That both domains are involved in Pax-QNR DNA binding is further indicated by the fact that paired domains and homeodomains individually expressed in bacteria are able to bind to e5. With bacterially expressed peptides, the e5P mutated oligonucleotide, unable to form stable complexes with Pax-QNR, forms stable complexes with the paired domain peptide, suggesting that conformational events involving other part of the Pax-QNR protein are involved in the DNA binding activity of the protein. Such interaction was also demonstrated for the *Drosophila* Prd protein (10, 15), for which the homeodomain binding appears to be masked by the COOH-terminal part of the protein. In addition, the inhibitory effect of the COOH-terminal part of the protein appears to be sequence specific (15). This COOH-terminal modulation is not a general regulation of *Pax* gene DNA binding, since Pax-5 DNA binding is not modulated by this part of the protein (25). However, Pax-5 is devoid of homeodomain, and this particular type of DNA binding regulation may be specific to paired domain and homeodomain-containing proteins.

The core sequence of the Pax-1, Pax-2, and Prd paired domain binding site, TTCC, is also the core sequence of the binding site of the *ets* gene family products (31). We observed that certain types of *ets* target sequences, devoid of any ATTA homeobox binding sites, can be recognized by the MC29-QNR2 Pax-QNR protein, and that a stable complex can be formed under conditions in which the e5P oligonucleotide (also devoid of ATTA) does not allow the formation of a stable complex. Thus, the Pax-QNR binding on e5 demonstrates a cooperation between the two DNA binding domains in the protein, but the type of the sequence recognized is also involved in the stabilization of the DNA-protein complex.

Presently, it is unclear whether the e5 sequence is a target for Pax-QNR binding *in vivo* and whether this sequence will be instrumental in defining Pax-QNR as a transcription factor. Experiments performed with Pax-3 indicate that efficient binding on e5 is not correlated with any activation of transcription (21).

Materials and Methods

Cells and Viruses. NRs were dissected from 7-day-old quail embryos. QNR cultures from 7-day-old embryonic

tissue were maintained and passaged in Eagle's basal medium supplemented with 10% fetal calf serum. MC29 RAV-1 virus was obtained by cotransfection of DNA from the molecular clone pMC38 containing the myelocytomatosis provirus and Rous associated virus type 1 DNA into quail embryo cells (4).

Isolation of Quail Clones Containing *Pax-QNR* Homologous Sequences. Using MC29-QNR2 cDNA as probe, clone B1 was isolated from a cDNA library constructed in the pCDM8 vector from 15-day-old quail spinal cords (kindly supplied by C. Dulac). Clone λPax-QNR1 was isolated from a genomic library constructed from peripheral blood DNA of quail embryos in the EMBL4 vector, screened with the MC29-QNR1 cDNA probe (4). Clones λPax-QNR2 and λPax-QNR3 were isolated from a genomic library constructed from 14-day-old quail embryonic tissue DNA in the λDASH vector, using as probes a 1.5-kbp *Bam*HI genomic fragment containing exon 1 and a 94-bp *Eco*RI-*Hae*III fragment representing the 5′ part of MC29-QNR2 cDNA, respectively.

DNA Sequencing. Appropriate subclones of genomic DNA as well as B1 cDNA were ligated into M13mp18 and M13mp19. Nucleotide sequence was determined in both orientations by the dideoxynucleotide chain-termination method using standard techniques and an Applied Biosystem 370A automatic sequencer. All sequence homology searches were done with software from Bisance (32).

RNase Protection Assays. To obtain a B1-specific RNA probe, a 572-bp *Hin*cII fragment containing 490 bp of cDNA B1 and 82 bp of linker sequences was cloned into the *Hin*cII site of the pGEM4 plasmid. The antisense RNA probe, 637 nt long, was transcribed *in vitro* from the *Hin*dIII-linearized plasmid using SP6 polymerase in the presence of [α-^{32}P]CTP under conditions provided by the riboprobe system (Promega). To obtain an MC29-QNR2-specific RNA probe, a 424-bp *Eco*RI-*Hin*cII 5′ cDNA fragment was cloned into the *Eco*RI-*Hin*cII sites of the pGEM4 plasmid. The antisense RNA probe, 455 nt long, was transcribed *in vitro* from the *Eco*RI-linearized plasmid using T7 polymerase under the same conditions as described above. RNase protection assays were performed with 10 μg of total quail RNA extracted from MC29-transformed neuroretina cells, 14-day-old embryonic tissues (cerebellum, brain, and QNR), QNR from 7-day-old embryonic tissue, and heart from 3-week-old quails, as described previously (33). As a negative control, RNase protection assays were performed on yeast tRNA.

Expression of Paired, Homeo-, and Carboxy-terminal Domains of Pax-QNR in Bacteria. The appropriate fragments were obtained from the MC29-QNR2 plasmid used as template in PCR experiments using oligonucleotides corresponding to sequences present in paired, homeo-, and carboxy domains of Pax-QNR (the sequences of the oligonucleotides used are available upon request). The expected bands were subjected to *Bam*HI-*Hin*dIII restriction enzyme digestion and inserted into the *Bam*HI-*Hin*dIII sites of the pPLc24 vector. The subcloning resulted in the in-frame fusion of the desired sequence to the first 99 amino acids of the polymerase of phage MS2. For expression, plasmids were transferred into an *Escherichia coli* host which has a temperature-sensitive repressor of the PL promoter. Cultures of exponentially growing bacteria that carried the desired vector were induced at 42°C for 3 h. Bacterial pellets were washed twice, boiled in sample buffer, and electrophoresed in a 15% SDS-polyacrylamide gel. Proteins were visualized by staining with Coomassie brilliant blue and cut from the gel for purification.

***In Vitro* Transcription, Translation, and EMSA.** Plasmids pSP64 and pSP65 MC29-QNR2 were linearized using the *Pvu*II restriction site, and the sense (pSP65) and antisense (pSP64) RNA strands were synthesized with SP6 RNA polymerase. Half (5 μl) of the RNA template was translated in a rabbit reticulocyte lysate. All reactions were performed according the manufacturer's specifications (BRL).

To prepare the DNA probes for gel shift analysis, an oligonucleotide containing the e5 binding site CTCAGCACCGCACGATTAGC**ACCGTTCCGCTTC** and its complementary strand were synthesized, annealed, and used as DNA probe after 5′-end labeling. The recognition sequence bound by the homeodomain is underlined, and the paired domain recognition sequence is shown in boldface. Two other sets of oligonucleotides, e5P with mutations (bold letters) in the homeodomain binding site CTCAGCACCGCACG**CGCT**GCACCGTTCCGCTTC, and e5H with mutations in the paired binding site CTCAGCACCGCACGATTAGCACCGTT**TT**GCTTC and their complementary strands, were synthesized, annealed, and used for the competition experiments.

One-fiftieth of the reticulocyte lysate (1 μl) or 2 μg of the gel-purified, bacterially expressed proteins were mixed with 1 ng of the DNA probe and ^{32}P labeled with T4 polynucleotide kinase in binding buffer [10% glycerol, 10 mM 4-(2-hydroxy-ethyl)-1-piperazineethanesulfonic acid (pH 7.9), 30 mM KCl, 0.5 mM dithiothreitol, 4 mM spermidine, 1.5 μg polydeoxyinosinic-deoxycytidylic acid, 1 μg salmon sperm DNA, protease inhibitors (0.5 mM phenylmethylsulfonyl fluoride, 1 μg/μl each of leupeptin, antipain, and pepstatin, and 2 mM benzamidine] and phosphatase inhibitors (10 mM β-glycerophosphate, 2 mM levamisole, and 10 μM orthovanadate). Reactions were performed in 16 μl final volume for 10 min on ice. DNA-protein complexes were analyzed by electrophoresis on 6% polyacrylamide gels (Tris-borate-EDTA buffer) at room temperature for 2 h (10 V/cm). The gels were dried and exposed to film.

Nucleotide Sequence Accession Number. The nucleotide sequence data reported in this paper will appear in the EMBL nucleotide sequence databases under the accession numbers X68168 and X68169.

Acknowledgments

We thank Kathryn Nason-Burchenal for critical reading of the manuscript.

References

1. Okada, T. S. *In:* J. D. Ebert, and M. Marois (eds.), Tests of Teratogenicity *in Vitro*, pp. 91-105. Amsterdam: Elsevier, 1976.

2. Monaghan, P., Davidson, D. R., Sime, C., Graham, E., Baldock, R., Bhattacharya, S., and Hill, R. The Msh-like homeobox genes define domains in the developing vertebrate eye. Development, *112:* 1053-1061, 1991.

3. Walther, C., and Gruss, P. Pax-6, a murine paired box gene, is expressed in the developing CNS. Development, *113:* 1435-1449, 1991.

4. Martin, P., Carrière, C., Dozier, C., Quatannens, B., Mirabel, M-A., Vandenbunder, B., Stéhelin, D., and Saule, S. Characterization of a paired box- and homeobox-containing quail gene (Pax-QNR) expressed in the neuroretina. Oncogene, *7:* 1721-1728, 1992.

5. Akam, M. The molecular basis for metameric pattern in the *Drosophila* embryo. Development, *101:* 1-22, 1987.

6. Ingham, P. W. The molecular genetics of embryonic pattern formation in *Drosophila*. Nature (Lond.), *335:* 25-34, 1988.

7. Izpisua-Belmonte, J-C., Tickle, C., Dolle, P., Wolpert, L., and Duboule, D. Expression of the homeobox *Hox-4* genes and the specification of position in chick wing development. Nature (Lond.), *350:* 585-589, 1991.

8. Murphy, P., Davidson, D. R., and Hill, R. E. Segment-specific expression of a homeobox-containing gene in the mouse hindbrain. Nature (Lond.), *341:* 156-159, 1989.

9. Tabin, C. J. Retinoids, homeoboxes, and growth factors; toward molecular models for limb development. Cell, *66:* 199-217, 1991.

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

11. Ingraham, H. A., Chen, R., Mangaiam, H., Elsholtz, S., Flynn, C. R., Lin, D. M., Simmons, L., Swanson, L., and Rosenfeld, M. G. A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell, *55:* 519-529, 1988.

12. Frain, M., Swart, G., Monaci, P., Nicosia, A., Stämpfli, S., Frank, R., and Cortese, R. The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. Cell, *59:* 145-157, 1989.

13. Karlson, O., Thor, S., Norberg, T., Ohlsson, H., and Edlund, T. Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. Nature (Lond.), *344:* 879-882, 1990.

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

15. Treisman, J., Harris, E., and Desplan, C. The paired box encodes a second DNA-binding domain in the paired homeo domain protein. Genes & Dev., *5:* 594-604, 1991.

16. Kessel, M., and Gruss, P. Murine developmental control genes. Science (Washington DC), *249:* 374-379, 1990.

17. Hastie, N. *Pax* in our time. Dev. Genet., *1:* 342-344, 1991.

18. Hill, R. E., Favor, J., Hogan, B., Ton, C. C., Saunders, G., Hanson, I., Prosser, J., Jordan, T., Hastie, N., and Heyningen, V. V. Mouse *small eye* results from mutations in a paired-like homeobox-containing gene. Nature (Lond.), *354:* 522-525, 1991.

19. Ton, C. C., Hirvonen, H., Miwa, H., Weil, M., Monaghan, P., Jordan, T., Van Heyningen, V., Hastie, N., Meijers-Heijboer, H., Drechsler, M., Royer-Pokora, B., Collins, F., Swaroop, A., Strong, L. C., and Saunders, G. F. Positional cloning and characterization of a paired-box- and homeobox-containing gene from the aniridia region. Cell, *67:* 1059-1074, 1991.

20. Krauss, S., Johansen, T., Korzh, V., Moens, U., Ericson, J. U., and Fjose, A. Zebrafish *pax[zf-a]*: a paired box-containing gene expressed in the neural tube. EMBO J., *10:* 3609-3619, 1991.

21. Goulding, M. D., Chalepakis, G., Deutch, U., Erselius, J. R., and Gruss, P. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. EMBO J., *10:* 1135-1147, 1991.

22. Dressler, G. R., and Douglass, E. Pax-2 is a DNA-binding protein expressed in embryonic kidney and Wilms tumor. Proc. Natl. Acad. Sci. USA, *89:* 1179-1183, 1992.

23. Adams, B., Dörfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurer-Fogy, I., and Busslinger, M. Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. Genes & Dev., *6:* 1589-1607, 1992.

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

25. Zannini, M., Francis-Lang, H., Plachov, D., and Di Lauro, R. Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters. Mol. Cell. Biol., *12:* 4230-4241, 1992.

26. Kozack, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA, *87:* 8301-8305, 1990.

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

28. Lemaire, P., Vesque, C., Schmitt, J., Stunnenberg, H., Frank, R., and Charnay, P. The serum-inducible mouse gene Krox-24 encodes a sequence-specific transcriptional activator. Mol. Cell. Biol., *10:* 3456-3467, 1990.

29. Glaser, T., Welton, D. S., and Maas, R. Genomic structure, evolutionary conservation and aniridia mutations in the human Pax-6 gene. Nat. Genet., *2:* 232-239, 1992.

30. Jostes, B., Walther, C., and Gruss, P. The murine paired box gene, Pax 7, is expressed specifically during the development of the nervous and muscular system. Mech. Dev., *33:* 27-38, 1991.

31. Macleod, K., Leprince, D., and Stéhelin, D. The ets gene family. Trends Biochem. Sci., *17:* 251-256, 1992.

32. Dessen, P., Fondrat, C., Valencien, C., and Mugnier, C. Comput. Appl. Biosci., *6:* 355-356, 1990.

33. Dozier, C., Ansieau, S., Ferreira, E., Coll, J., and Stéhelin, D. An alternatively spliced *c-mil/raf* mRNA is predominantly expressed in chicken muscular tissues and conserved among vertebrate species. Oncogene, *6:* 1307-1311, 1991.