Zebrafish Pax9 Encodes Two Proteins with Distinct C-terminal Transactivating Domains of Different Potency Negatively Regulated by Adjacent N-terminal Sequences*§

(Received for publication, November 28, 1995, and in revised form, July 3, 1996)

Svanhild Nornes‡§, Ingvild Mikkola‡§, Stefan Krauss†**, Madjid Delghandi‡, Maria Perander‡, and Terje Johansen‡‡

From the ‡Departments of Biochemistry and §Molecular Genetics, Institute of Medical Biology, University of Tromsø, 9037 Tromsø, Norway

We describe the isolation of cDNA clones for zebrafish Pax9. Pax9 expression was initiated at the end of the segmentation period in mesenchymal sclerotome cells on both sides of the notochord similarly to the corresponding mouse and chick genes. Two transcripts, Pax9a and -b, are generated by alternative splicing. The gene contains 4 exons with exon 3 being included in the Pax9a transcript and spliced out in the Pax9b transcript. The Pax9a and -b proteins are identical for 212 amino acids from the N terminus but contain distinct C-terminal regions of 131 and 58 amino acids, respectively. The paired domain of Pax9 displayed a binding-site specificity distinct from Pax6 but similar to Pax1 and -2. Both Pax9a and -b activated a promoter containing a paired domain binding site. However, this activation was observed when low amounts of Pax9 expression vectors were used. Higher amounts led to a sharp decrease in the activation and even turned into repression. Both the distinct C-terminal regions of Pax9a and -b harbored transcriptional activating domains of different potency not revealed in the context of the full-length proteins due to a negative influence of the N-terminal region including the paired domain.

Several of the multigene families acting as developmental control genes encode transcription factors containing conserved DNA-binding domains such as the homeodomain and the paired domain (1, 2). The paired domain consisting of 125–128 amino acids, encoded by the paired box, was named after the paired domain (1, 2). The paired domain consisting of 125–128 DNA-binding domains such as the homeodomain and the control genes encode transcription factors containing conserved gene families, which is organized in large gene clusters, the nine Pax genes are dispersed on eight chromosomes in humans and six chromosomes in mice (reviewed in Ref. 13). In addition to the paired domain prd, gsb, gsb, ey, and Pax3, -4, -6, and -7 also contain a paired-type homeodomain (2, 3, 13). Pax2, Pax5, and Pax8 encode an amino acid stretch related to the first predicted α-helix of the paired type homeodomain (8, 14). All the Pax genes, except for Pax6 (9, 15), encode a conserved octapeptide of unknown function (12). This sequence is also absent in the Prd protein and the Drosophila Pax6 homologue Ey (3, 12).

The Pax genes show distinct spatiotemporal expression patterns during embryogenesis. Several Pax genes, including Pax2, Pax3, Pax5, Pax6, Pax7, and Pax8 are expressed in dorsoventrally restricted domains along the hindbrain and spinal cord (i.e. see Refs. 16 and 17). The expression of Pax1 is unique as it is confined exclusively to mesodermal tissue that gives rise to the axial skeleton (18). An evolutionary relationship between Pax1 and Drosophila "pox meso" is apparent both in terms of paired domain sequence homology and mesodermal expression (2, 19).

Strikingly, mutations in the three murine Pax genes Pax1, Pax3, and Pax6 have been shown to cause the developmental mutants undulated, Splotch, and Small eye, respectively. Furthermore, mutations in the PAX3 and PAX6 genes are associated with the human inherited disorders Wardenburg’s syndrome and aniridia (reviewed in Ref. 13). Studies with transgenic mice lacking Pax5 show this gene to be important for normal development of the midbrain and absolutely essential for B cell development (20), and deregulated expression of Pax2 leads to severe kidney abnormalities (21). In zebrafish, Pax2 is required for proper formation of the midbrain-hindbrain boundary (22). Convincing evidence for Pax6 as a master control gene for eye morphogenesis was recently provided (23). Overexpression of Pax proteins in rodent fibroblasts is oncogenic (24), and in the pediatric solid tumor alveolar rhabdomyosarcoma, chromosomal translocations result in fusions of the paired and homeo domains of PAX3 or PAX7 to the transactivating domain of a forkhead family transcription factor (25, 26). Thus, the association with severe developmental defects and oncogenesis both in mouse and man suggest that the members of the relatively small Pax gene family seem to exert crucial functions in the regulation of development. Consistent with this notion, evidence showing that Pax proteins function as transcription factors has recently been provided. Pax proteins show specific binding to isolated sequences and to pro-

* This work was funded by grants from the Norwegian Cancer Society, the Aakre Foundation, and the Norwegian Research Council (to T. J.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Contributed equally to this work.

‡ Contributed equally to this work. Fellow of the Norwegian Research Council.

¶ Recipient of a short term fellowship from the European Science Foundation.

** To whom correspondence should be addressed. Tel.: 47 776 44720; Fax: 47 776 45350.

26914
Transactivation by Alternatively Spliced Forms of Pax9

Poly(A)+ RNA was purified using Dynabeads oligo(dT)25 (Dynal). Northern blot analysis was carried out following standard procedures (40). Expression of Isolated Paired Domains in Escherichia coli and Gel Mobility Shift Assays of DNA Binding—The paired domain coding sequences of zebrafish Pax2, Pax6, and Pax9 were amplified from their respective cDNA clones by PCR. The specific primers were constructed so that the 5’ primer contained a NcoI site, whereas the 3’ primer contained a BamHI site. The paired domain PCR products were cut with NcoI and BamHI and ligated into the correct E. coli expression vector pET-15b (Novagen). The following primers were used: Pax2, GAATTCCATGACCATATGATGGCAGCGGCGACCGG and CCGG- GATCTTGGAAATGTTTCTGAATG; Pax6, GAATTCCATGGAGAAACCTGACATGGTGGT and CGGCGATCTTTAGCAGTACCTGCGT- TTATCG; and for Pax9, primers 6 and 7 described above were used. E. coli BL21 (DE3) was used as host strain for protein expression, which was performed according to the Novagen Instruction Manual for pET vectors. Dilutions of crude extracts were run on a 20% (21:1) polyacrylamide gel (47), and the percentage of the overexpressed paired domain relative to total bacterial proteins was determined by densitometric scanning following Coomassie Blue staining using the UVP System 5000 gel documentation system with the GelBase software package.

The previously described H2A-2.2, H2B-2.2, CD19-1 (48), CD19-2 (49), Thy-1 (50), H2A-2.2, and Tyr-1 (50) cDNA libraries, using a 313-bp HindIII-XbaI fragment from pBlueScriptSK- (+) (Promega) by ligating AvaI-NcoI cut cDNA fragments into the EcoRI and NcoI sites of the vector. Rabbit polyclonal antibodies were raised against the Pax9 paired domain and the unique C-terminal domain (amino acids 207–270) of Pax9b using glutathione S-transferase fusion proteins. Protein extracts for Western blots were made by crumbling and boiling zebrafish embryos (24–26 h) frozen in liquid nitrogen in SDS sample buffer. Following electrophoresis on SDS-polyacrylamide gel electrophoresis, 10% gels, the proteins were blotted to a Immobilon-P membrane. The membrane was blocked overnight at 4 °C with 1% bovine serum albumin, 5% non-fat dry milk, 0.1% Tween-20 in TBS (Tris-buffered saline) (pH 7.4). Anti-Pax9 sera and secondary antibody (goat anti-rabbit, alkaline phosphatase conjugate, Santa Cruz Biotechnology, SC2007) were diluted 1:5000 and 1:2500 in blocking buffer, respectively, and incubated for 1 h at room temperature. Following each incubation the membrane was washed five to six times 1 h with 0.1% Tween 20, TBS. A chemiluminescence substrate, CDpstar (Boehringer Mannheim), was used to develop the signals.

Total RNA for Northern blot was isolated from zebrafish embryos (26–36 h) using the TRIsolv™ reagent (Life Technologies, Inc.), and

Materials and Methods

Isolation of cDNA Clones and DNA Sequencing—A 5’-truncated Pax9 cDNA clone obtained in a low stringency screen of a 33-h zebrafish embryonic gt11 cDNA library, using a 313-bp HindIII-XbaI fragment from the paired box of murine Pax as a probe, was used to screen 7.8 × 10⁸ plaques of a 20–28-h embryonic IZAP-II library at high stringency. This way four new clones were isolated. Two of these clones were sequenced completely on both strands and correspond to the Pax9a and -b sequences shown in Fig. 1. Screening of cDNA libraries, subcloning, and DNA sequencing were done according to standard procedures (40).

Computer-assisted Analyses of DNA and Protein Sequences—The GCG software package (version 8.0) was used for phylogenetic analyses (42, 43). For secondary structure predictions the Alexis program of the Seqsee program of the GCG software package (version 8.0) was used for general sequence alignments. This way four new clones were isolated. Two of these clones were sequenced completely on both strands and correspond to the Pax9a and -b sequences shown in Fig. 1. Screening of cDNA libraries, subcloning, and DNA sequencing were done according to standard procedures (40).

Long PCR on Genomic DNA—A Takara LA PCR Kit (Takara Shuzo Co. Ltd.) was used to analyze the genomic structure of the Pax9 gene with genomic DNA isolated from adult zebrafish as the template. Pax9 genomic DNA fragments were amplified using different combinations of 8 oligonucleotide primers: 1, CGGGATCCATGGAAATGCATAC; 2, GCTGATGACAACTGTCTACG; 3, TTAATGGCAGATCCTGCTTCC; 4, CGGCGATCTTTAGCAGTGACG; 5, CACACATCTGCCTGGACG; 6, CGGCGATCTTTAGCAGTGACG; and 7, ATGCTCCGTCTATATCGTGC. The PCR products were gel-purified and sequenced directly from their ends using the appropriate primers essentially as described by Khrona et al. (46).

In Vitro Translation and Western and Northern Blot Analysis—The two proteins Pax9a and Pax9b were expressed in vitro in the presence of [35S]methionine using the Promega TNT in vitro transcription and translation kit. Pax9a and Pax9b cDNAs were cloned in the sense orientation downstream of the SP6 promoter in pGEM-Zf(+) (Promega) by ligating AvaI-NcoI cut cDNA fragments into the EcoRV and NcoI sites of the vector.

The previously described H2A-2.2, H2B-2.2, CD19-1 (48), CD19-2, Thy-1 (50), H2A-2.2, and Tyr-1 (50) cDNA libraries, using a 313-bp HindIII-XbaI fragment from pBlueScriptSK- (+) (Promega) by ligating AvaI-NcoI cut cDNA fragments into the EcoRV and NcoI sites of the vector. Rabbit polyclonal antibodies were raised against the Pax9 paired domain and the unique C-terminal domain (amino acids 207–270) of Pax9b using glutathione S-transferase fusion proteins. Protein extracts for Western blots were made by crumbling and boiling zebrafish embryos (24–26 h) frozen in liquid nitrogen in SDS sample buffer. Following electrophoresis on SDS-polyacrylamide gel electrophoresis, 10% gels, the proteins were blotted to a Immobilon-P membrane. The membrane was blocked overnight at 4 °C with 1% bovine serum albumin, 5% non-fat dry milk, 0.1% Tween-20 in TBS (Tris-buffered saline) (pH 7.4). Anti-Pax9 sera and secondary antibody (goat anti-rabbit, alkaline phosphatase conjugate, Santa Cruz Biotechnology, SC2007) were diluted 1:5000 and 1:2500 in blocking buffer, respectively, and incubated for 1 h at room temperature. Following each incubation the membrane was washed five to six times 1 h with 0.1% Tween 20, TBS. A chemiluminescence substrate, CDpstar (Boehringer Mannheim), was used to develop the signals.

Total RNA for Northern blot was isolated from zebrafish embryos (26–36 h) using the TRIsolv™ reagent (Life Technologies, Inc.), and
Transactivation by Alternatively Spliced Forms of Pax9

Fig. 1. Zebrafish Pax9 encodes two proteins differing in their C termini. A, nucleotide and derived amino acid sequences of the Pax9a and -b cDNAs. The paired domain is indicated by double underlining while the octapeptide, the noncoding sequence shown to mediate mRNA instability (58), and the polyadenylation signal are marked by single underlining. The vertical arrows demarcate the 140-bp exon present in Pax9a relative to Pax9b and the location of the intron in the second codon of the gene. The sequences deleted in Pax9b, indicated by dashes, and a single A to C transversion are shown below the Pax9a sequence as well as the C-terminal amino acid sequence of Pax9b encoded by a different reading frame. Putative proline-directed protein kinase phosphorylation sites are shown in bold lettering. Asterisks denote in frame stop codons in the 5′-untranslated region. B, schematic drawing of the Pax9a and -b cDNAs with the coding regions boxed to emphasize the location of the paired domain (striped boxes), the 84-amino acid common region (open boxes) with the octapeptide (black boxes), and the distinct C-terminal domains. The amino acids which are overrepresented in the different regions downstream of the paired domain are indicated by single letters: (P) proline, (Q) glutamine, (S) serine, and (H) histidine. The black bars denote the locations of the helix-turn-helix motifs in the N- and C-terminal halves of the paired domain. The numbering refers to amino acid positions. The GenBank™ accession nos. for the Pax9a and Pax9b sequences are U40931 and U40932, respectively.

assigned since the 5′-UTR is devoid of other potential start codons and contains two in frame stop codons (see Fig. 1A). The start codon sequence context compares favorably with the Kozak consensus for initiator methionine codons (54). The putative proteins (denoted Pax9a and Pax9b) are identical for 212 amino acids from the N terminus, including the 125-amino acid paired domain and the conserved octapeptide sequence, but differ in their C-terminal sequences downstream of the octapeptide. Due to the presence of the 140-bp exon, the Pax9a cDNA encodes a 131-amino acid C-terminal region distinct from the 58-amino acid C-terminal region encoded by Pax9b, since the two C termini are encoded by different reading frames (see Fig. 1). The common 84-residue long downstream of the paired domain is very hydrophilic with a preponderance of proline (14.3%), serine (10.7%), and glutamine (9.5%) residues. The unique C-terminal region of Pax9a is rich in serine (17%) and proline (13%) residues. The 58-amino acid C-terminal region of Pax9b contains histidine and serine as the most abundant amino acids (27.5%). These features are characteristic for many transcription factors (55). Interestingly, eight putative phosphorylation sites for proline-directed protein kinases (56) are present in Pax9a (three in Pax9b, Fig. 1A). Pax9a is a 343 amino acids long protein of 37.3 kDa, whereas Pax9b is only 270 amino acids long (29.7 kDa), making it the smallest Pax protein known to date. The predicted sizes of the proteins were confirmed by coupled in vitro transcription and translation of the two cDNAs. As seen in Fig. 2A, the apparent molecular masses of 38 and 30 kDa determined by SDS-polyacrylamide gel electrophoresis followed by in vitro translation correspond very well to the molecular masses predicted from the cDNA sequences. To confirm the presence of two Pax9 proteins, we performed immunoblot analyses of whole zebrafish embryo extracts using antisera raised against the paired domain of Pax9 and against the unique C-terminal domain of Pax9b. As seen from Fig. 2B, immunoreactive bands with molecular masses corresponding to both Pax9a and -b were seen with the paired domain antisemur (lane 2), while only the band corresponding to the size of Pax9b was detected with the antisemur raised against the C-terminal domain of Pax9b (lane 4).

The 3′-UTRs of Pax9a and -b do not contain the consensus polyadenylation signal AATAAA. However, the sequence AT-TAA located 12 nt upstream of the poly(A) tail is also known as a functional polyadenylation signal (57). The Pax9a and -b cDNA clones are 1939 and 1782 nt long, respectively. Northern blot analysis of zebrafish mRNA isolated at 26–36 h of development displayed two broad bands of 2.1 ± 0.1 kb and 1.3 ± 0.1 kb, respectively (see Fig. 2C). The Pax9a and -b transcripts differing only by the presence/absence of the 140-bp exon will not be resolved on this gel. Assuming poly(A) tracts of ~200 nt, this suggests that the cDNA sequences shown in Fig. 1A are close to full-length. The 1.3-kb transcript is most probably due to the use of an alternative ATTTA polyadenylation signal located about 470 nt upstream of the ATTTAA sequence directing polyadenylation of the cDNA sequences shown in Fig. 1A. The 3′-UTR of Pax9a is (A + T)-rich (64%), contains an uninterrupted stretch of 27 A-T residues, and one copy of the nonamer motif TTATTTATT recently shown to be the crucial AU-rich sequence motif that mediates degradation of unstable mRNAs (58).

The Genomic Structure of the Pax9 Gene Reveals That the Pax9a and -b Transcripts Are Generated by Alternative Splicing—The fact that the human PAX9 gene contains an intron at exactly the same position (4) as the 140-bp insertion in Pax9a suggested that the Pax9a and -b transcripts could be generated
by alternative splicing. Furthermore, the absence of consensus splice donor and acceptor sequences at the junctions of the 140-bp insertion made it unlikely that Pax9a represented an incompletely spliced transcript. Also, PCR on two different cDNA libraries and reverse transcription-PCR analyses of RNA from different developmental stages with primers flanking the putative 140-bp exon in Pax9a consistently confirmed the presence of both transcripts (data not shown). PCR analyses and sequencing on zebrafish genomic DNA revealed that the gene contains four exons and three introns with exon 3 being the alternatively spliced exon included in the Pax9a transcript and spliced out in the Pax9b transcript (Fig. 3). The first long intron is located in the second codon of the gene. Similar to other class I paired-box genes (4, 12, 18, 19) the paired domain is not interrupted by any introns. Comparison of the splice donor and acceptor sequences for the three introns reveal that the splice donor site for the third intron deviates from the consensus sequence by having an A at position +5 (Table I). The other two splice donor sites have the consensus G at this position. Exon 3 is more AT-rich (57%) than the other exons (46%) also noted as a feature of alternatively spliced exons (59). These features may contribute to the exon skipping occurring during generation of the Pax9b transcript.

Comparison of the cDNA-derived Amino Acid Sequences of Zebrafish Pax9a and -b to Other Class I Paired Domain Proteins—Based on sequence homology, the presence of class-specific amino acids at certain positions, and conservation of exon-intron locations, the known paired domains can be divided into six different classes (2, 6). Using three different phylogenetic tree construction methods including 23 representative vertebrate and Drosophila paired domain amino acid sequences we were able to confirm the division of paired domains into six evolutionary classes by showing that the class sequences actually form distinct phylogenetic assemblages (see Fig. 4A). Pax1 and Pax9 constitute together with Drosophila pax meso class I of paired domain-containing proteins (4, 5). While this work was in progress a partial genomic sequence for chicken Pax9 (60) and a cDNA sequence for murine Pax9 (61) were reported. As shown in Fig. 4B, the paired domain of zebrafish Pax9 contains only three amino acid substitutions relative to human, murine, and chicken Pax9; seven compared to human Pax1 and 16 relative to pax meso. Comparisons of the C-terminal regions of Pax9a and -b to that of murine Pax1 and Pax9 clearly show the latter sequence to be the homologue to zebrafish Pax9a. The overall sequence identity is 73%, but if changes to chemically similar amino acids are considered the sequence similarity increases to 90%. However, the C-terminal region found here to harbor the transcriptional activation domain (see below) is remarkably less conserved compared to the region encoded by the paired domain exon (51.5% versus 87% identity).

The paired domain of Pax9 contains putative helix-turn-helix motifs both in the N- and C-terminal subdomains—Recently developed secondary structure prediction algorithms (44, 45) were used to predict the location of a-helices in the paired domain. As seen from Fig. 4B, six helices are predicted with helix-turn-helix (HTH) motifs both in the N-terminal (helices II and III) and C-terminal subdomains (helices IV and V). The C-terminal HTH has been postulated earlier along with the presence of helix I (12, 19). However, the N-terminal HTH has not been predicted before. Interestingly, this HTH is also predicted by the algorithm of Dodd and Egan (62) specifically developed to detect such motifs. All the helices except helix II are jointly predicted by the neural network method (45) and the Alexis program of the Seqsee program suite (44). Helix II is predicted by Alexis on the basis of homology to solved structures and the likelihood of this prediction is further strengthened by the fact that the Dodd and Egan algorithm suggests that this helix together with helix III forms a HTH structure. The presence of putative HTH motifs both in the N- and C-terminal subdomains is completely consistent with recent functional analyses of specific DNA sequence recognition by the paired domain which show that sequence specific recognition modules reside in both subdomains (48, 63). A corectral structure of the paired domain of Drosophila Prd in complex with an in vitro selected optimal binding site was recently reported (64). As shown in Fig. 4B, our helix predictions show a striking correlation with the Prd paired domain crystal structure model.

DNA-binding properties of the Pax9 paired domain—in order to study its DNA-binding specificity we overexpressed the paired domain of Pax9 in E. coli and performed gel mobility shift assays with several previously established Pax protein-binding motifs as probes (see the legend to Fig. 5). To allow
direct comparisons of binding specificities and affinities of the Pax9 paired domain relative to other paired domains, the zebrafish Pax2 and Pax6 paired domains were expressed in a similar way and included in the binding assays. As shown in Fig. 5, Pax9 bound the same probes as Pax2 but showed a lower affinity than Pax2 for the sea urchin histone H2A-2.2 and H2B-2.2 sites. Pax9 showed highest affinity for the mutated CD19-2(A-ins) site and almost equal affinity for CD19-1 and PRS4. H2B-2.2 and H2A-2.2 were bound to a lesser extent. These results from equilibrium binding experiments of the Pax9 paired domain were confirmed by competition assays except that H2B-2.2 and PRS4 were equally effective in competing the CD19-2(A-ins) site and slightly more effective than CD19-1 (see Fig. 5). As recently demonstrated by Czerny and Busslinger (65), the Pax6 paired domain displayed a distinct specificity binding only to the H2B-2.2, CD19-1, and CD19-2(A-ins) probes. None of the paired domains bound the TgC probe under the conditions used here although this site was previously shown to bind Pax8 (35), which contains a paired domain highly related to Pax2 (8). Comparison of the results shown in Fig. 5 with those reported for murine Pax1 with some of the same probes (48) clearly shows that Pax1 and Pax9 display very similar binding preferences.

Pax9 Is Expressed in Sclerotomal Tissue—In situ hybridizations on tissue sections and whole mount embryos with Pax9a and Pax9b probes showed similar staining patterns confined to sclerotome tissue. Since the two splicing variants Pax9a and Pax9b differ only by the presence/absence of the 140-bp exon, it was not possible to dissect variations in the expression of both. Pax9 expression was first detected at the end of the segmentation period. In situ hybridizations using 35S-labeled Pax9a probes on 24-h embryos show a strong signal in cells of the mesenchymal sclerotome on both sides of the notochord (Fig. 6, E–H). Similar staining was seen on caudal transverse sections of embryos that were hybridized whole mount with Pax9. Interestingly, in addition to the Pax9-positive cells in the perichordal tube, also a lateral group of cells express Pax9. From their position between the myotomes and the neural keel, we assume that these cells are sclerotome cells that give rise to the neural arch. This notion is further substantiated by lateral views of the trunk and tail demonstrating a homogenous ventral Pax9 staining and a metamerical dorsomedial extension at the caudal margin of the myotomes (Figs. 6, B and D, and 7E). In addition to the expression in the trunk and tail, at 24 h, Pax9 expression could also be detected in the lateral head mesoderm (Fig. 6, A and C). At 48 h of development, Pax9 expression remains basically unaltered. The ventral peripheral Pax9 staining appears strong and homogenous, and the dorsal Pax9 expressing cells have increased in number and comprise now a major component of Pax9 expression in the trunk and tail (Fig. 7, C and D).

The Distinct C-terminal Domains of Pax9a and -b Display Different Transactivating Activities Negatively Regulated by N-terminal Sequences—All the Pax proteins that have been tested by transient transfection assays in cell culture have been reported to act as transcriptional activators (see Introduction). However, Pax proteins have also been reported to repress certain promoters/enhancers (66–68). To test the transactivation/repression potential of intact Pax9 proteins we transfected NIH 3T3 cells with increasing amounts of Pax9a and -b expression vectors together with a reporter vector containing a single Pax-binding site (CD19-2(A-ins)) upstream of the tk promoter in pBLCAT2. A seen from Fig. 8, both Pax9a and -b were capable of activating this reporter. No activation was observed when pBLCAT2 lacking the Pax-binding site was transfected together with Pax9 expression vectors (data not shown). Interestingly, the transactivation by Pax9 proteins was critically
dependent on the concentration of the Pax9 expression vectors. As seen for Pax9a in Fig. 8, 2 μg led to a sharp decrease of transcriptional activity. Such a behavior was recently reported for Pax6 and explained as due to self-squelching of the Pax6 transactivation domain (65). In several separate experiments using different cell lines we found that higher concentrations of Pax9 expression vectors even led to a repression below the basal level of the reporters inherent promoter activity (data not shown).

To characterize further the transactivational or transrepressional potential of the Pax9 proteins, we fused different parts of Pax9a and Pax9b to the DNA-binding domain (DBD) of yeast GAL4 and carried out cotransfection assays in human HeLa cells using three different reporters: pG5E1bTATA-CAT, containing 5 GAL4-binding sites upstream of the TATA box of the herpes simplex virus tk promoter, facilitated measurements of transcriptional activity. Such behavior was recently reported when acting from an enhancer position. Surprisingly, nearly full-length sequences of murine Pax1 and -9 (28, 61), amino acid sequences to the partial sequences of human PAX9 and PAX1 (4, 12), chicken Pax9 (60) and Drosophila pox meso (19), as well as the full-length sequences of murine Pax1 and -9 (28, 61). Dots indicate identity, while gaps are represented by dashes. Black boxes above the sequences indicate α-helices jointly predicted by the neural network method and the Alexis program (see “Materials and Methods”), while open boxes denote residues predicted by the Alexis program only to be part of α-helices. For the paired domain (helices I–VI) the secondary structure predictions are based on 20 different paired domain sequences present in the Swissprot data base (Release 30). The extent of α-helices determined from the x-ray structure model of the paired domain of Drosophila Prd (64) is indicated by double underlining. The vertical arrows denote the approximate locations of introns in both the Pax9 and Pax1 genes.

The transactivating domain of the herpes simplex virus VP16 protein, the Pax9a transactivation domain was only 5–7-fold less potent. As seen in Fig. 9D the transactivating domains of both Pax9a and -b also activated the tk promoter from an enhancer position. In fact, the activation domain of Pax9a was just as efficient in activating the tk promoter from distal downstream binding sites as from proximal upstream sites. For Pax9b, however, the activation was 2-fold lower from the distal sites relative to the proximal sites. Thus, Pax9a was 6-fold more potent than Pax9b when acting from an enhancer position. Surprisingly, nearly full-length sequences of Pax9a (amino acids 17–343) or Pax9b (amino acids 17–270), while the weakest effect was mediated by a construct containing the very full-length sequences of Pax9a (amino acids 17–343) and -b (amino acids 207–270) both displayed a strong transactivating activity on both promoters. The transactivating domain of Pax9a was about 2.5 times more potent than that of Pax9b. Compared to the very strong acidic transactivating domain of the herpes simplex virus VP16 protein, the Pax9a transactivation domain was only 5–7-fold less potent. As seen in Fig. 9D the transactivating domains of both Pax9a and -b also activated the tk promoter from an enhancer position. In fact, the activation domain of Pax9a was just as efficient in activating the tk promoter from distal downstream binding sites as from proximal upstream sites. For Pax9b, however, the activation was 2-fold lower from the distal sites relative to the proximal sites. Thus, Pax9a was 6-fold more potent than Pax9b when acting from an enhancer position. Surprisingly, nearly full-length sequences of Pax9a (amino acids 17–343) or Pax9b (amino acids 17–270), while the weakest effect was mediated by a construct containing the very full-length sequences of Pax9a (amino acids 17–343) and -b (amino acids 207–270) both displayed a strong transactivating activity on both promoters. The transactivating domain of Pax9a was about 2.5 times more potent than that of Pax9b. Compared to the very strong acidic transactivating domain of the herpes simplex virus VP16 protein, the Pax9a transactivation domain was only 5–7-fold less potent. As seen in Fig. 9D the transactivating domains of both Pax9a and -b also activated the tk promoter from an enhancer position. In fact, the activation domain of Pax9a was just as efficient in activating the tk promoter from distal downstream binding sites as from proximal upstream sites. For Pax9b, however, the activation was 2-fold lower from the distal sites relative to the proximal sites. Thus, Pax9a was 6-fold more potent than Pax9b when acting from an enhancer position. Surprisingly, nearly full-length sequences of Pax9a (amino acids 17–343) or Pax9b (amino acids 17–270), while the weakest effect was mediated by a construct containing the very full-length sequences of Pax9a (amino acids 17–343) and -b (amino acids 207–270) both displayed a strong transactivating activity on both promoters. The transactivating domain of Pax9a was about 2.5 times more potent than that of Pax9b. Compared to the very strong acidic transactivating domain of the herpes simplex virus VP16 protein, the Pax9a transactivation domain was only 5–7-fold less potent. As seen in Fig. 9D the transactivating domains of both Pax9a and -b also activated the tk promoter from an enhancer position. In fact, the activation domain of Pax9a was just as efficient in activating the tk promoter from distal downstream binding sites as from proximal upstream sites. For Pax9b, however, the activation was 2-fold lower from the distal sites relative to the proximal sites. Thus, Pax9a was 6-fold more potent than Pax9b when acting from an enhancer position. Surprisingly, nearly full-length sequences of Pax9a (amino acids 17–343) or Pax9b (amino acids 17–270), while the weakest effect was mediated by a construct containing the very full-length sequences of Pax9a (amino acids 17–343) and -b (amino acids 207–270) both displayed a strong transactivating activity on both promoters. The transactivating domain of Pax9a was about 2.5 times more potent than that of Pax9b. Compared to the very strong acidic transactivating domain of the herpes simplex virus VP16 protein, the Pax9a transactivation domain was only 5–7-fold less potent. As seen in Fig. 9D the transactivating domains of both Pax9a and -b also activated the tk promoter from an enhancer position. In fact, the activation domain of Pax9a was just as efficient in activating the tk promoter from distal downstream binding sites as from proximal upstream sites. For Pax9b, however, the activation was 2-fold lower from the distal sites relative to the proximal sites. Thus, Pax9a was 6-fold more potent than Pax9b when acting from an enhancer position. Surprisingly, nearly full-length sequences of Pax9a (amino acids 17–343) or Pax9b (amino acids 17–270), while the weakest effect was mediated by a construct containing the very full-length sequences of Pax9a (amino acids 17–343) and -b (amino acids 207–270) both displayed a strong transactivating activity on both promoters. The transactivating domain of Pax9a was about 2.5 times more potent than that of Pax9b. Compared to the very strong acidic transactivating domain of the herpes simplex virus VP16 protein, the Pax9a transactivation domain was only 5–7-fold less potent. As seen in Fig. 9D the transactivating domains of both Pax9a and -b also activated the tk promoter from an enhancer position. In fact, the activation domain of Pax9a was just as efficient in activating the tk promoter from distal downstream binding sites as from proximal upstream sites. For Pax9b, however, the activation was 2-fold lower from the distal sites relative to the proximal sites. Thus, Pax9a was 6-fold more potent than Pax9b when acting from an enhancer position. Surprisingly, nearly full-length sequences of Pax9a (amino acids 17–343) or Pax9b (amino acids 17–270), while the weakest effect was mediated by a construct containing the very full-length sequences of Pax9a (amino acids 17–343) and -b (amino acids 207–270) both displayed a strong transactivating activity on both promoters. The transactivating domain of Pax9a was about 2.5 times more potent than that of Pax9b. Compared to the very strong acidic transactivating domain of the herpes simplex virus VP16 protein, the Pax9a transactivation domain was only 5–7-fold less potent. As seen in Fig. 9D the transactivating domains of both Pax9a and -b also activated the tk promoter from an enhancer position. In fact, the activation domain of Pax9a was just as efficient in activating the tk promoter from distal downstream binding sites as from proximal upstream sites. For Pax9b, however, the activation was 2-fold lower from the distal sites relative to the proximal sites. Thus, Pax9a was 6-fold more potent than Pax9b when acting from an enhancer position. Surprisingly, nearly full-length sequences of Pax9a (amino acids 17–343) or Pax9b (amino acids 17–270), while the weakest effect was mediated by a construct containing the very full-length sequences of Pax9a (amino acids 17–343) and -b (amino acids 207–270) both displayed a strong transactivating activity on both promoters. The transactivating domain of Pax9a was about 2.5 times more potent than that of Pax9b. Compared to the very strong acidic transactivating domain of the herpes simplex virus VP16 protein, the Pax9a transactivation domain was only 5–7-fold less potent. As seen in Fig. 9D the transactivating domains of both Pax9a and -b also activated the tk promoter from an enhancer position. In fact, the activation domain of Pax9a was just as efficient in activating the tk promoter from distal downstream binding sites as from proximal upstream sites. For Pax9b, however, the activation was 2-fold lower from the distal sites relative to the proximal sites. Thus, Pax9a was 6-fold more potent than Pax9b when acting from an enhancer position. Surprisingly, nearly full-length sequences of Pax9a (amino acids 17–343) or Pax9b (amino acids 17–270), while the weakest effect was mediated by a construct containing the very full-length sequences of Pax9a (amino acids 17–343) and -b (amino acids 207–270) both displayed a strong transactivating activity on both promoters. The transactivating domain of Pax9a was about 2.5 times more potent than that of Pax9b. Compared to the very strong acidic transactivating domain of the herpes simplex virus VP16 protein, the Pax9a transactivation domain was only 5–7-fold less potent. As seen in Fig. 9D the transactivating domains of both Pax9a and -b also activated the tk promoter from an enhancer position. In fact, the activation domain of Pax9a was just as efficient in activating the tk promoter from distal downstream binding sites as from proximal upstream sites. For Pax9b, however, the activation was 2-fold lower from the distal sites relative to the proximal sites. Thus, Pax9a was 6-fold more potent than Pax9b when acting from an enhancer position. Surprisingly, nearly full-length sequences of Pax9a (amino acids 17–343) or Pax9b (amino acids 17–270), while the weakest effect was mediated by a construct containing the very full-length sequences of Pax9a (amino acids 17–343) and -b (amino acids 207–270) both displayed a strong transactivating activity on both promoters. The transactivating domain of Pax9a was about 2.5 times more potent than that of Pax9b. Compared to the very strong acidic transactivating
sequence, but due to the exclusion of exon 3 the C-terminal of Pax9b is translated in a different reading frame compared to that of Pax9a. The resulting C-terminal regions differ both in length and sequence. Comparisons of the two zebrafish protein sequences to the recently published murine Pax9 cDNA sequence (61) clearly show Pax9a to represent the zebrafish homologue to this sequence. Alternative splicing has also been reported for other Pax genes. Similar to zebrafish Pax9, human Pax2 and Pax8 give rise to proteins differing in their C-terminal sequences (69, 70). For human Pax6, alternative splicing leads to the generation of two proteins differing by a 14-amino acid insertion in the paired domain that alters the DNA binding-specificity (63). This insertion has also been found in Pax6 cDNA sequences from zebrafish, quail, and mouse, but not in the sea urchin (15, 65, 71, 72). Furthermore, the quail Pax6 gene gives rise to several alternatively spliced products, including a protein truncated in the paired domain (71). Finally, alternative splicing has also been demonstrated for the human Pax3 gene (73). Although we do not know the mechanism allowing skipping of exon 3 in some of the zebrafish Pax9 transcripts, we note that the splice donor site of intron 3 of the Pax9 gene deviates from the consensus sequence by a substitution of G to A in position +5. Interestingly, this is also seen in one of the splice donor sites of the human Pax8 gene and, as in Pax9, the exon upstream of this splice donor site may be excluded (69). In addition, exon 3 of the Pax9 gene is more A-T-rich than the other exons, another recognized feature that may be characteristic of exon skipping (59). It has been suggested that tissue-specific factors, such as U2AF, SF2/ASF, and other members of the SR protein family, may be responsible for regulating alternative splicing events (74). Alternative splicing products of Pax8 are spatially and temporally regulated in distinct parts of the embryo (69). In this work we were not able to discriminate between the expression patterns of the two Pax9 transcripts in the zebrafish embryo due to the small size of the exon 3 insertion in Pax9a. Thus, it remains to be seen

**Discussion**

In this study we present the full-length cDNA sequences derived from two different transcripts of the zebrafish Pax9 gene. Elucidation of the exon-intron structure of the zebrafish Pax9 gene revealed that the two transcripts, denoted Pax9a and -b, arise by alternative splicing. This results in inclusion of exon 3 in the Pax9a transcript and exclusion of this exon in the Pax9b transcript. Both of the putative Pax9 proteins contain the conserved N-terminal paired domain and a common 84-amino acid region, including the conserved octapeptide sequence.
whether the two Pax9 proteins differ in their temporal and spatial expression patterns. The recent reports on chicken (60) and murine (61) Pax9 sequences give no indications of alternative splicing occurring in these species. However, by analyzing the murine Pax9 cDNA sequence, which encodes the orthologous protein to Pax9a, we were able to locate a region that would encode a 58-amino acid C-terminal region homologous to Pax9b (40% identity, 59% similarity) following skipping of a putative exon 3 of the same size as the one found in zebrafish (140 bp). This suggests that a Pax9b protein may be expressed in vertebrates in general.

Direct amino acid sequence comparisons and our phylogenetic analyses of the known paired domain sequences suggest the presence of four different classes of vertebrate paired domains if the putative pseudogene Pax4 is excluded. Thus, the four classes may differ in their DNA-binding specificities and/or affinities with Pax2, -5, and -8 in one, Pax6 in a second, Pax3 and -7 in a third, and Pax1 and -9 in a fourth class. In fact, the results of Czerny et al. (48) for the paired domains of Pax1, -3, -5, and 6 confirm this notion. As expected, our data show a DNA-binding specificity and affinity of Pax9 for different sites, which is similar to that previously reported for Pax1, and confirm that Pax6 has a distinct specificity (31, 48, 65), while Pax2, like Pax5 and Pax8, displays the most promiscuous DNA-binding pattern.

Using a combination of several novel secondary structure prediction methods, we predicted that the paired domain contains an N-terminal and a C-terminal subdomain consisting of three $\alpha$-helices with the second and third helices in each subdomain forming HTH motifs. Comparison of our predictions to the recently published crystal structure of the paired domain of Drosophila Prd in complex with an in vitro selected optimal binding site (64) show a striking correlation, suggesting that our approach may be valid for other protein families containing conserved domains for which there is presently no structural information. The structural model of the paired domain is consistent with functional analyses showing the paired domain to act as a bipartite DNA-binding domain with the N- and C-terminal subdomains contacting 3' and 5'-half sites in consecutive major grooves of the DNA (48).

Zebrafish Pax9 expression in the sclerotomes is clearly similar to that seen in chick and mice with expression further dorsally relative to Pax1 and includes tissues involved in the formation of the neural arch (60, 61). It remains to be seen whether a mutation in the Pax9 gene may lead to a defect that includes sclerotomal tissue of more dorsal origin like the neural arch, and whether Pax9 deficiency in more ventral tissue may be rescued by the overlapping Pax1 expression. Recently, it has been shown that the induction of Pax1 expression in the

**Fig. 7. Distribution of Pax9 transcripts as detected by in situ hybridizations on whole mount embryos.** Anterior is to the left. A, sagittal view of Pax9 expression in the tail of a 24 h embryo. B, transverse section through the tail of the same embryo. Lines demarcate dorsolateral sclerotome expression of Pax9 in presumptively migrating chondrocytes that will contribute to the neural arch. Note the strong expression in the ventral sclerotome. Arrowheads denote the notochord. C, sagittal view of Pax9 expression in the tail of a 48-h embryo. D, transverse section through the tail of the same embryo. The Pax9-expressing cells have continued to wrap the neural tube and show a dorsal concentration (compare the positions of the lines that demarcate dorsolateral scleroderm with those in B). E, sagittal view of Pax9 expression in the trunk of a 24-h embryo. Note the metamer distribution of the Pax9 positive sclerotomal cells. Abbreviations: nt, neural tube; nc, notochord; and s, sclerotome.

**Fig. 8.** Pax9a and -b activate a promoter containing a Pax-binding site. NIH 3T3 fibroblasts were cotransfected with the reporter plasmid together with increasing amounts of Pax9a and -b expression plasmids. The CAT activity obtained with 2 $\mu$g of empty pRc-CMV expression vector (control) was set to 1. The amounts of expression plasmids were equalized by the addition of the empty expression vector to a total of 2 $\mu$g to normalize for possible promoter interference effects. The reporter plasmid contained a single CD19-2(A-ins)-binding site in front of the tk promoter. The data from one experiment with three independent transfections are shown.
somites depends on antagonizing signals between the notochord and the dorsal surface ectoderm (75, 76), with sonic hedgehog (SHH) (53, 77, 78) being the ventral inducting signal. In response to an induction by SHH, Pax1 expression is initiated in the ventral, but not in the dorsal portion of the sclerotome. It is there essential for further vertebrate differentiation (28, 79). Whether Pax9 exhibits a similar crucial role in vertebrate development as Pax1 remains to be seen. We detect Pax9 expression first at a time when shh has ceased to be expressed in the notochord and only weak shh expression is maintained in a single row of floor plate cells (53). Thus, if Pax9 expression is, similar to the expression of Pax1, induced by SHH, a sufficient concentration of active SHH peptide must be present several hours after the shh gene has been switched off in the notochord.

We found both Pax9a and -b to be able to activate transcription of a reporter containing a single Pax-binding site upstream of the tk promoter. This transactivation was concentration-dependent, and high concentrations of the Pax9 expression vectors caused repression compared to the same amounts of empty expression vector. When fused to the GAL4 DNA-binding domain, the C-terminal domains of both Pax9a and Pax9b were shown to be potent activators of transcription from both a minimal and a more complex promoter, as well as being able to activate the tk promoter from an enhancer position. However, this activation was abolished and even turned into repression when the common N-terminal part, including the paired domain, was present in the fusions. Interestingly, similar observations have been made for other transcription factors such as the serum response factor (SRF) (80), ATF-2 (see Ref. 81 and references therein) the visna virus Tat protein (82), and the ETS family proteins ER81 and Sap-1a (83). In these cases the full-length proteins fused to GAL4 did not show any activation, while smaller fragments of these proteins fused to GAL4 did. For both SRF, ER81, Sap-1a, and visna virus Tat, inhibitory domains were shown to repress the activity of the activation domain, while the activation domain of ATF-2 is masked in the intact protein and must be phosphorylated by JNK/SAPK kinases to display activity. Our findings indicate that the paired domain is able to directly repress the tk promoter, when fused alone to GAL4, as well as inhibiting the function of the C-terminal activation domains. This points to multiple functions of this domain besides its role in specific DNA binding. That the DNA-binding domain in itself can work as an inhibitory domain, independent of its DNA-binding capacity, has also been found for SRF (80) and ATF-2 (81). The inhibitory domains could work through an intramolecular interaction that regulates the activation domain, through an intermolecular interaction with another protein or by a combination of both. For ATF-2, a direct interaction between the N-terminal region containing the activation domain and the DNA-binding domain prevents the activation domain from being able to stimulate transcription unless the activation domain is phosphorylated (81). This model may also apply to Pax9. However, our results show that the paired domain by itself represses the tk promoter, indicating that an intermolecular interaction may also be involved.

Given the fact that the two different C-terminal activation domains of Pax9a and Pax9b display different potency one can speculate as to whether they play different or redundant roles during embryogenesis. That alternative splicing may generate distinct C-terminal domains with different transactivating potential was also recently demonstrated for human PAX8 (69).

Thus, in future studies it will be of interest both to elucidate the mechanisms involved in the regulation of the transactivating domains of Pax9 proteins, and to understand why two proteins with transactivating domains of different potency are produced, as well as determining what genes they regulate.

Acknowledgments—We thank Fang Liu and Michael Green for GAL4 plasmids, Thomas Czerny and Meinrad Busslinger for oligonucleotides, and Kai Zinn and David Grunwald for cDNA libraries. The skillful technical assistance of Turid Holm is greatly appreciated.
Zebrfish Pax9 Encodes Two Proteins with Distinct C-terminal Transactivating Domains of Different Potency Negatively Regulated by Adjacent N-terminal Sequences

Svanhild Nornes, Ingvild Mikkola, Stefan Krauss, Madjid Delghandi, Maria Perander and Terje Johansen

J. Biol. Chem. 1996, 271:26914-26923.
doi: 10.1074/jbc.271.43.26914

Access the most updated version of this article at http://www.jbc.org/content/271/43/26914

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 83 references, 32 of which can be accessed free at http://www.jbc.org/content/271/43/26914.full.html#ref-list-1