BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **240,** 196–202 (1997)
ARTICLE NO. RC977623

Alternative Splicing of *Pax6* in Bovine Eye and Evolutionary Conservation of Intron Sequences

Cynthia Jaworski, Sally Sperbeck, Caroline Graham, and Graeme Wistow
Section on Molecular Structure and Function, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892

Received October 1, 1997

***Pax-6* is essential for eye development and for tissue-specific gene expression within the eye. Splicing of an alternative exon (5a) leads to variant DNA binding specificity in Pax6. Here we show that in the mature bovine eye the relative abundance of the alternative Pax6-5a mRNA varies markedly, particularly between the lens and iris. Additional alternative splicing products were identified which may lead to "domain swapping" or truncation of Pax6 proteins. Comparison of amphibian (*Xenopus*) and mammalian (bovine) *Pax-6* genes revealed highly conserved intronic sequences flanking exon 5a, including potential lariat sites which may have a role in the mechanism of alternative splicing.**

Pax6 has a key role early in the cascade of events leading to differentiation of the eye in both vertebrates and invertebrates (1-4). It also has specific effects in different eye tissues. Haplo-insufficiency of *Pax6* in human *aniridia* generally produces more serious effects in iris than in other eye tissues (5). Pax6 is also implicated in expression of several crystallin genes in lens (6, 7) and is essential for lens-specific expression of guinea pig ζ-crystallin (8). However, *Pax6* gene expression itself is not lens or even eye specific (9-11).

Pax6 genes in vertebrates contain a characteristic alternative exon (called 5a in mammals) which, when spliced into mRNA, produces a variant paired domain (PD) with altered DNA recognition (12). Previously we observed that the relative abundance of canonical Pax6 mRNA and the alternative Pax6-5a form differs markedly between mature mouse brain and lens (8) such that lens shows a marked preference for canonical Pax6 while brain shows similar levels of Pax6 and Pax6-5a (8). It has also been reported that ratios of Pax6 splice forms vary with development in embryonic chick retina (13). Here we have examined alternative splice forms of Pax6 mRNA in mature bovine eye and used phylogenetic footprinting (14) to identify evolutionarily conserved *Pax6* intronic sequences with potential for involvement in control of alternative splicing in this "master gene" for eye development.

MATERIALS AND METHODS

RNA from eye tissues. Bovine eyes were obtained from a local slaughterhouse. Lens, retina, iris (plus ciliary body) and cornea were dissected. Pigmented material was removed from lens and cornea.

Reverse transcription-polymerase chain reaction (RT-PCR). For RT-PCR (15), 200ng total RNA was reverse transcribed using Superscript 1 (Life Sciences, Gaithersburg, MD) and amplified in a reaction using 0.2mM dNTP, 1μM primers and 5U Taq DNA polymerase (Boehringer Mannheim, Indianapolis, IN) in 10mM Tris-HCl, 50mM KCl, 2mM $MgCl_2$, pH 8.3. Following preliminary 5min incubation at 95° to denature DNA, amplification used 30 cycles of 1min 94°, 1min 55°, 1min 72°. Initially, primers flanking the PD, were designed from human and mouse cDNA sequences (11, 16): (5′) ATGCAGAACAGTCACAGCGGAGTGAA; (3′) GCTAGCCAGGTTGCGAAGAACTCTGTT and used to amplify bovine lens RNA. For comparison of different eye tissues, a second pair of primers were designed from the determined bovine sequence: (5′) CAGCTCGGTGGTGTCTTTGTC; (3′) AACTCTGTTTATTGATGACAG.

Products were visualized on agarose gels. Nucleotide sequences of PCR products were determined by direct sequencing (17) and by subcloning (pCRII system, Invitrogen, San Diego CA) followed by sequencing (Sequenase, USB, Cleveland OH).

Sequence comparison and molecular graphics. Sequences were analyzed using the GCG package (18) and the ALIGN program (Scientific and Educational Software, Durham, NC). Protein structures were examined using Quanta (Molecular Simulations Inc) with coordinates from the crystallographic structure of the *paired* PD (19), communicated by Dr. Carl Pabo.

Pax6 PD coding sequences from the present study and from GenBank were aligned using PILEUP (18) and a cladogram was drawn using Neighbor-joining in MEGA (Molecular Evolutionary Genetics Analysis) version 1.01 (20). Distance calculations used simple nucleotide differences with 1000 bootstrap replications.

Partial genomic cloning of the Xenopus and bovine genes for Pax6. *Xenopus* genomic DNA (gift of Dr. T. Sargent) was from the "homozygous" HD1 line derived by gynogenesis (21). Following a PCR protocol similar to that used for mRNA except for an annealing temperature of 60°C, 1μg DNA was amplified using degenerate primers designed from *Drosophila* and human *Pax6* PD sequences (2, 16): (5′) GTCACAGTGGAGTAAATCAGCTGGGTGG(CT)GT(CT)T; (3′) AGTAATCT(AG)TC(CT)CG(AG)ATTTCCCAAGCAAA. Five cloned PCR products were fully sequenced. Primers from the bovine *Pax6* cDNA

0006-291X/97 $25.00

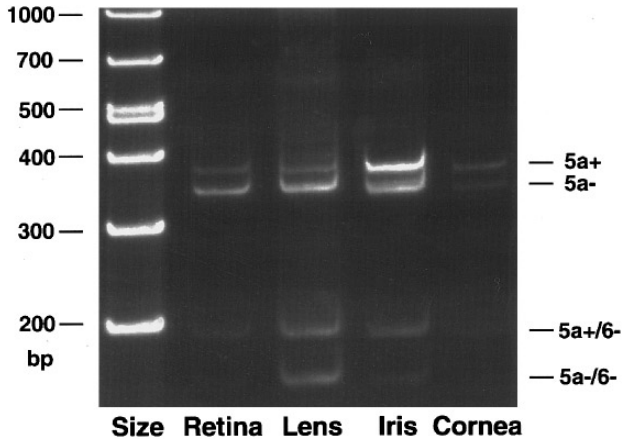

FIG. 1. Alternative splicing of *Pax6* mRNA in adult bovine eye. Representative RT-PCR amplification of RNA from eye tissues. Presence or absence of exonic sequences in the PD region are indicated by 5a+, 5a−, 6+, and 6−, where "6−" indicates partial absence of exon 6.

sequence: (5′) GATCGTAGAGCTAGCTCACA ; (3′) GGGCACTCCCGCTTATACTG, were similarly used to amplify the equivalent exon 5-exon 6 region of the bovine *Pax6* gene. A single product was obtained and was completely sequenced on both strands by direct sequencing (22).

RESULTS

Alternative splicing of Pax6 mRNA. PCR primers for the *Pax6* PD coding sequence were used to amplify RNA from bovine lens, retina, iris and cornea (fig 1). Although this is not a quantitative method between tissues, it allows comparison of relative abundance of mRNA splice forms within each tissue. In different preparations, consistently high yields of PD RT-PCR products were obtained from iris and lens with moderate levels in retina and low levels in trimmed cornea. Four major product sizes observed in lens were isolated and sequenced, revealing the presence of five distinct products of alternative splicing (fig 2a). Other eye tissues contained similar PCR products, but with distinctly different relative abundance. Most notably, lens showed a clear preference for the canonical form of Pax6 mRNA over Pax6-5a, just as was previously observed for mouse lens and lens-derived cells (8) while, in contrast, iris yielded much higher levels of the Pax6-5a form (fig 1). Retina was generally similar to lens, while cornea yielded low but detectable levels of both Pax6 and Pax6-5a.

Additional minor splice forms. In addition to the major Pax6 and Pax6-5a mRNA species, other shorter RT-PCR products corresponding to the PD coding region were also detected in eye tissues, most noticeably in lens (fig 1). Cloning and sequencing revealed that these bands contained three additional alternative splice forms of Pax6 mRNA (fig 2a).

One of these (type 4) corresponds to a form previously found in quail retina (23) in which most of exon 6 is deleted ("6−") through use of an alternative splice donor site (6′) in exon 6 which splices to the acceptor site of exon 7. Type 3 (fig 2a) contains the same 6′ donor site, but also includes the alternative exon 5a, giving a "5a+/6−" pattern. A fifth type cloned from the same size band as type 4 uses a second novel splice donor site (6″), 20 bases further downstream, also spliced to the common exon 7 acceptor site.

Predicted protein structures. The sequence of the PD is highly conserved throughout vertebrates and invertebrates (24), allowing the crystallographic structure of the PD from *paired*, the family archetype from *Drosophila* (19), to serve as a model for bovine Pax6. The PD consists of two subdomains (19), here called N-domain and C-domain. Both have similar arrangements of three α-helices (N1-3 and C1-3) (fig 2b). In the crystallographic structure only the N-domain binds the DNA fragment used for co-crystallization, with N3 playing a major part in DNA recognition. No structure is presently available for the product of the alternative 5a exon.

The result of the 6′ alternative splice is a deletion, removing most of N3, the connecting peptide between N- and C-domains and helices C1 and C2. Most of helix C3 would be retained in frame, as illustrated in fig 3a. It is likely that the N-domain would fold normally, with the substitution of part of helix C3 for N3. This would resemble "domain swapping" (25) in which a structural element normally folding in one context swaps for similar interactions with another domain or subunit. Indeed, comparison of the "swapped" regions of N3 and C3 show several conserved amino acid positions (fig 3c), corresponding to specific packing contacts in the N-domain. The 6′ splice would thus reduce the PD to a single DNA binding domain structurally similar to the N-domain but altered in sequence in the DNA-binding N3 helix. The homeodomain (HD) and the activator PST region of Pax6 (26) would remain intact.

The 6″ splice causes a deletion after 3 turns of helix N3, but since it splices to exon 7 out of frame, three new residues, CHQ, would be included before reaching a premature stop codon. This would be enough to complete a modified version of N3 (fig 3b), but the protein would resemble an isolated N-domain lacking all downstream sequences. While it could retain DNA binding activity, just as in the crystal structure (19), it would probably lack transactivation activity and might even be inhibitory. Future experiments to express this recombinant domain should allow the determination of any binding activity.

Phylogenetic footprinting of intron sequences. Sequences involved in alternative splicing events in the PD regions were examined by phylogenetic footprinting (14, 27, 28). Part of the *Pax6* gene from exons 5 to 6

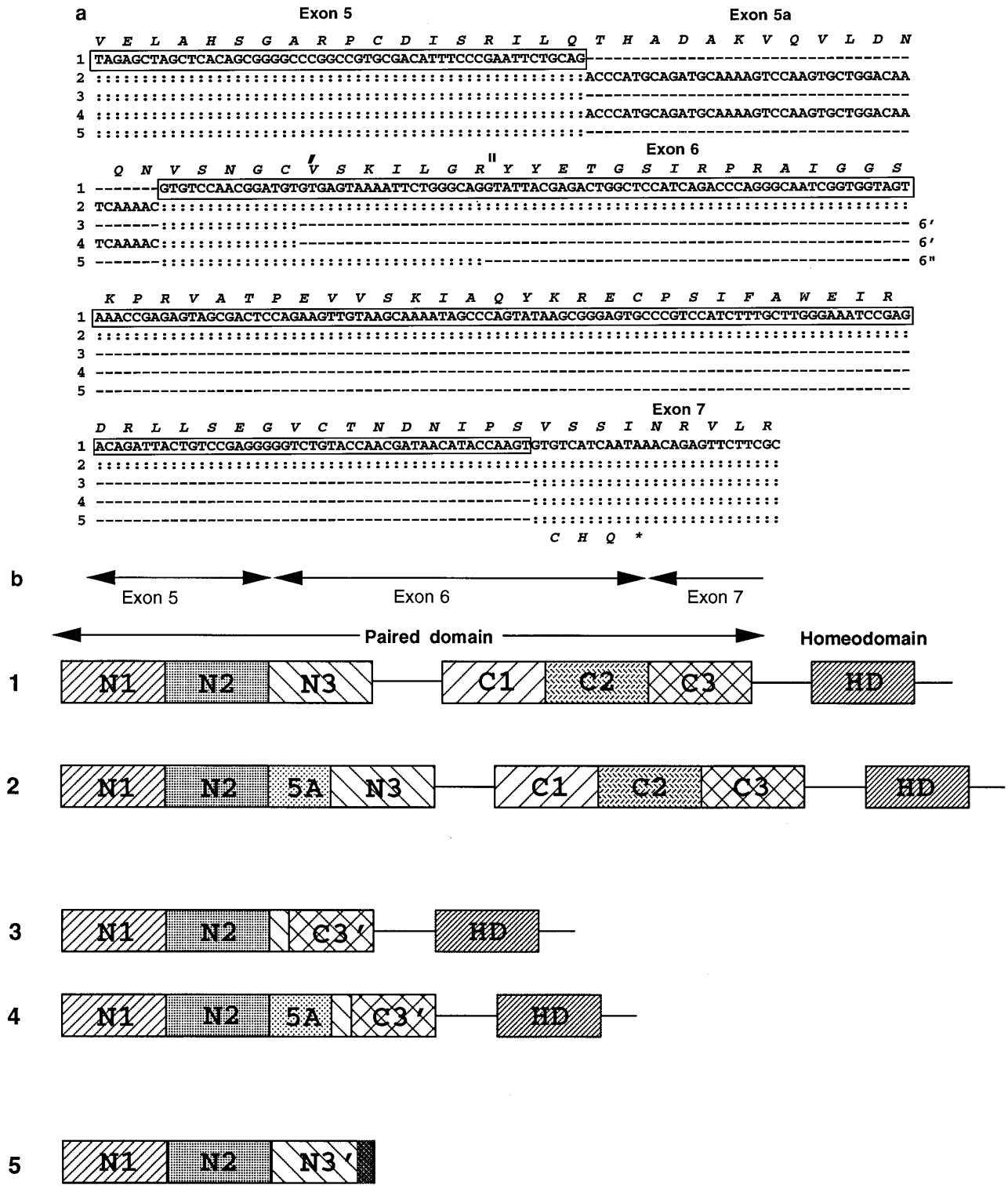

FIG. 2. (a) Partial sequences of five alternative splice forms of bovine *Pax6* mRNA. Exons 5 and 6 are boxed, and 5a and 7 are unboxed. Conceptual translation is shown in italics. The frame-shifted protein sequence for type 5 is shown below. Alternative splice donor sites in exon 6 are shown by ′ and ″ and sequences using these splices are labelled 6′ and 6″. Identities are shown by colons (:) and deletions by hyphens (-). GenBank accession No. U73855. (b) Schematic of protein products of the five variant Pax6 mRNA sequences. Boxes illustrate α-helices N1-3 and C1-3. Approximate correspondence with exons is shown by arrows.

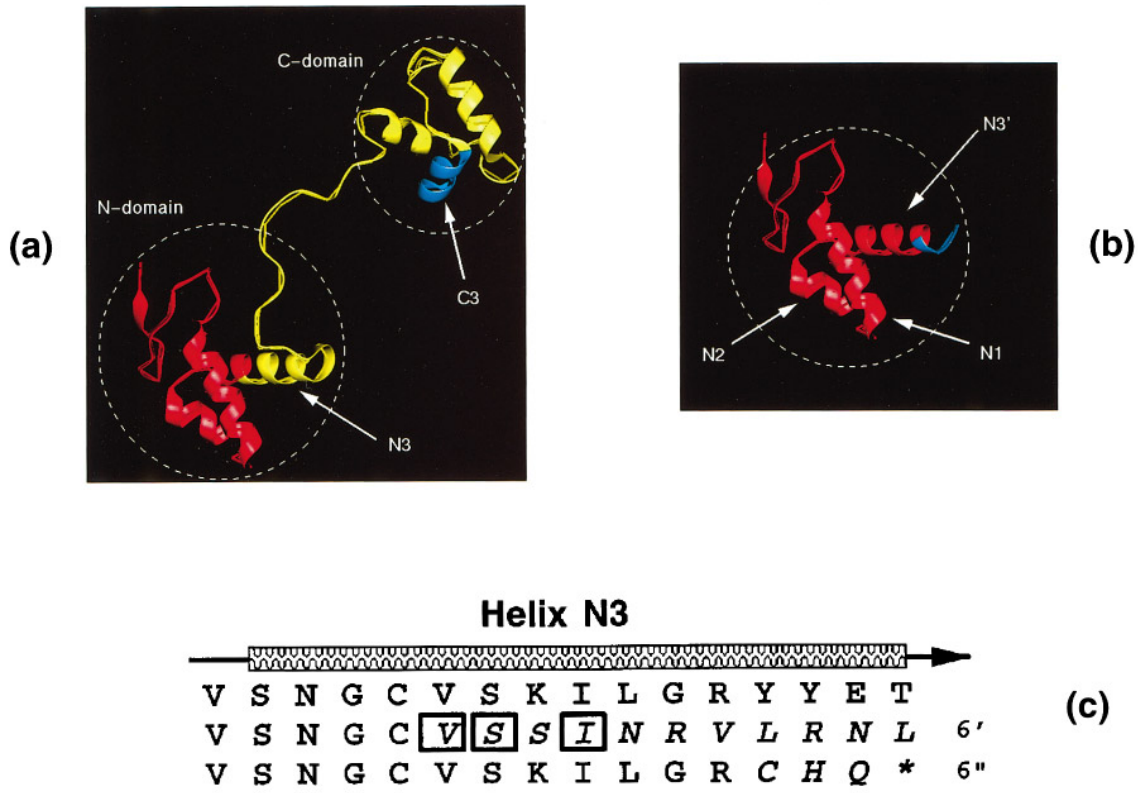

FIG. 3. The structural consequences of alternative splicing in the Pax6 PD. A ribbon trace superimposed on Cα positions of the PD is shown. (a) The 6′ splice deletes most of N3, C1, and C2. Regions deleted through this splice are shown in yellow. Part of helix C3 (blue) could substitute for most of helix N3 (yellow). (b) The 6″ splice causes a deletion of sequences from exon 6 and a frame shift of helix C3 (blue). (c) Alignment of sequences predicted for the N3 helix region. Top line shows the sequence of N3 compared with the variants produced by alternative splices 6′ and 6″. Residues from helix C3 or from a shifted reading frame are shown in italics. Boxed residues are identical between N3 and C3.

was cloned from a mammal (*Bos taurus*) and from an amphibian (*Xenopus laevis*) and compared.

The *Xenopus* genomic sequence was derived from HD1, a gynogenetic line identical at all loci (21). A single consensus sequence was assembled from multiple PCR clones and exon sequences were found to be highly conserved with those of other species. Within intron 5, the cloned region contained the alternative exon 5a, diagnostic for *Pax6* of vertebrates. After this sequence was determined, two *Pax6* cDNA sequences from *Xenopus* were deposited in Genbank (accession numbers U64513 and U67887). One of these (U67887) matches the genomic sequence exactly for the regions they have in common. Phylogenetic comparisons for PD nucleotide sequences show that both *Xenopus* sequences cluster together and are similarly related to *Pax6* sequences of other species (fig 4), suggesting that they represent pseudo-alleles. Interestingly, although mouse (11) and fish (29) 5a protein sequences are identical (THADAKVQVLDNEN), that predicted for *Xenopus* 5a differs at 4 out of 14 positions (sHADAKVpVLDsqN). All changes result from single base substitu-

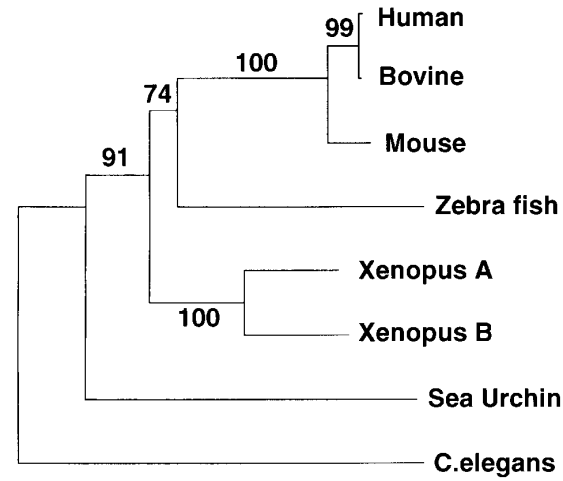

FIG. 4. Phylogenetic comparison of Pax6 sequences. Pax6 PD nucleotide sequences for several species were compared with those for bovine and *Xenopus* genes using MEGA 1.01 (20). Xenopus A is from the present sequence (GenBank No. U73621), and Xenopus B is from GenBank entry No. U64513. Bootstrap values are indicated.

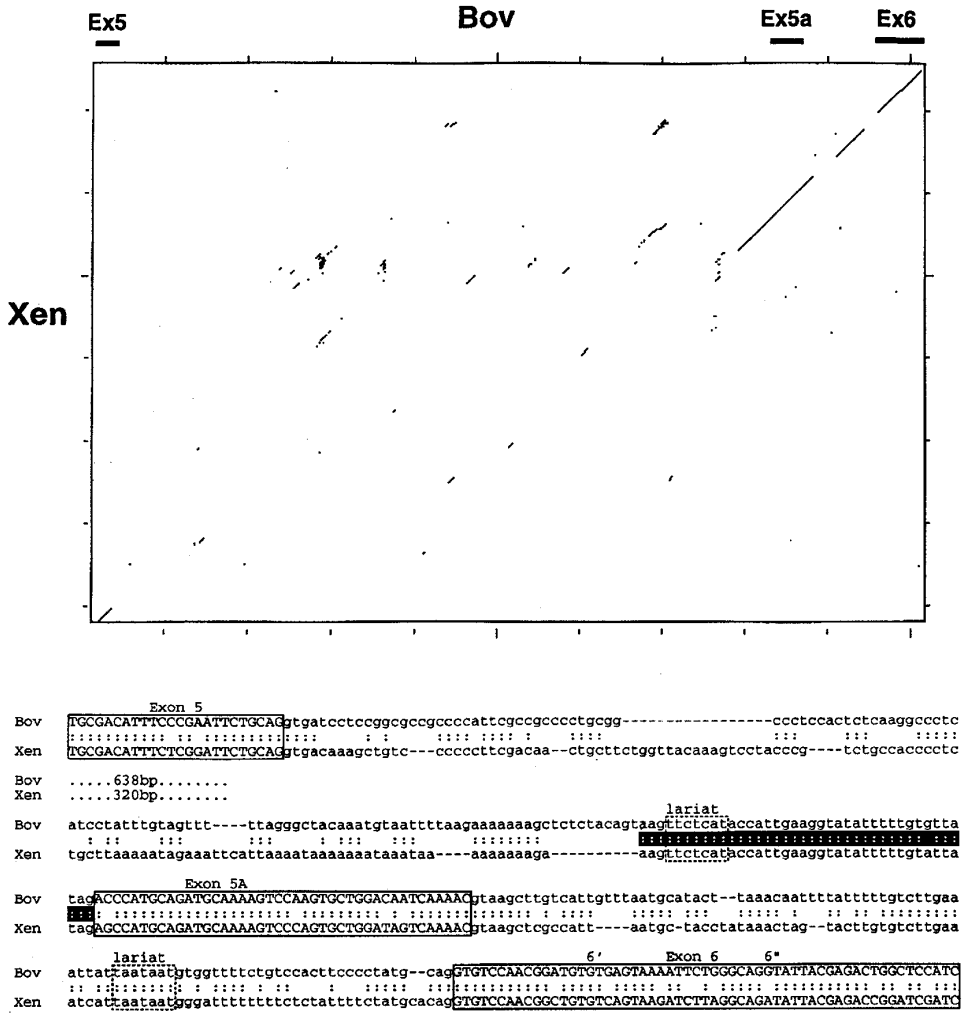

FIG. 5. Sequence comparison of *Pax6* genes from bovine and *Xenopus* genomes in the exon 5 -exon 6 region. (Top) Dot matrix comparison of bovine (Bov) and *Xenopus* (Xen) intron 5 sequences. Diagonal line shows conserved sequences. Positions of exons in the bovine sequence are indicated. Plot was generated using DOTPLOT (18) (window size 21, stringency 14). (Bottom) Alignment of conserved regions. Exons and potential consensus lariat sites are indicated. The highly conserved intron sequence adjacent to exon 5a is highlighted. Positions of alternative splices 6′ and 6″ are shown. GenBank Accession Nos. U73622 and U73621.

tions while other silent positions are identical, giving the appearance of selected evolution for changes in amino acid sequence. This might be the result of specialization in pseudo-alleles of this tetraploid species.

The partial bovine sequence is very similar to both rodent and human Pax6 sequences, but actually lies closer to human in phylogenetic analysis (fig 4). Unlike other mammals and fish, the predicted bovine protein sequence contains a single amino acid change (THADAKVQVLDqN) in the 5a insert.

In both bovine and *Xenopus* genes (fig 5), there is a longer distance between exons 5 and 5a (475bp in *Xenopus*, 794bp in bovine) than between exons 5a and 6 (90bp in *Xenopus*, 93bp in bovine). Exonic sequences are well conserved while, as expected, there is essentially no similarity throughout most of the intronic sequence (confirmed by dot matrix analysis, not shown). In the *Xenopus* gene, the 6′ alternative splice site is conserved, but at the 6″ site, GT is changed to AT (fig 5), suggesting that this alternative splice may not occur in frog.

Remarkably, the sequence immediately upstream of 5a and including the lariat site for the alternative splice is identical between frog and mammal at 38/39 positions; indeed this non-coding sequence is more highly conserved than the exons. The short intron 5a/6 is also almost 70% identical between the two species and again shows identity at a potential lariat site.

Such high sequence identity suggests that an important secondary structure or sequence recognition site is under selective protection. In terms of RNA secondary structure, there is a potential RNA stem of 11bp (ATTTTTGT(AG)TT/GATGCAAAAGT) between the conserved intron and part of 5a itself. A combination

of sequence and structure recognition is a possibility, as is binding by an antisense RNA. Whatever the recognition mechanism, masking or enhancement of the 5a lariat site would be a straightforward way of influencing alternative splice site selection. However, at present, the functional significance of these conserved sequences is not known.

DISCUSSION

Pax6, a member of a family of paired box genes (24), has been described as a "master gene" in eye development since mutations in *Pax6* abolish or disrupt eye development in mammals and flies (1, 2) while the highly conserved *Pax6* of mouse can induce ectopic eye formation in *Drosophila* (4). In addition to this high-level role, Pax6 is a transcription factor involved in regulation of lens-specific genes in vertebrates (6, 7). In spite of these eye or lens-specific effects, *Pax6* expression itself occurs in several different tissues in eye, in CNS and even in pancreas (9, 10).

Previously we showed that Pax6 is essential for expression of ζ-crystallin in transgenic mouse lens and lens-derived cells (8) and that *Pax6* is expressed appropriately for such a role in mature mouse and guinea pig tissues (8). However, we also found that although mouse brain RNA yielded similar levels of RT-PCR products corresponding to both canonical Pax6 and Pax6-5a mRNA, lens and lens-derived cells gave a great predominance of canonical Pax6 and very little Pax6-5a. The protein products of these splicing events differ in PD structure and DNA binding (12, 30). To determine whether this preference for one form of *Pax6* mRNA in mature lens extends to other species and to look for differences among eye tissues, tissues of the adult bovine eye were examined. The same bias against Pax6-5a was again observed in lens while iris showed a much greater abundance of Pax6-5a.

At the same time three other forms of *Pax6* mRNA were identified in eye. Two of these make use of an alternative donor splice site in exon 6 which was previously observed in one alternative transcript from quail retina (23). At the protein level this probably results in "domain swapping" (25) in the PD, with replacement of most of the DNA-binding α-helix of the N-domain with the equivalent helix of the C-domain. Another splice form makes use of a second alternative donor site and is predicted to result in a truncation of the PD after the N-domain, with consequent loss of the HD and activating regions. All of the variant Pax6 proteins predicted could have distinct DNA binding activities and different target genes.

In *aniridia*, a human haplo-insufficiency of *Pax6*, the iris is generally more severely effected than other eye tissues (5), suggesting that iris may require higher levels of *Pax6* mRNA for normal development and function. However, one interesting case of *aniridia* has been described in which splicing of the Pax6-5a form is greatly enhanced (12). This mutation is associated with a relative sparing of iris and more noticeable effects in lens. This is consistent with the preference for Pax6-5a seen in adult bovine iris. The retina is also relatively spared in this form of *aniridia*. The reason for this is unknown, but it might be explained if retinal targets for Pax6 are less dependent on activation by PD binding. Interestingly, it has been suggested that the conserved "P3" site common to the promoters of many retina genes is a binding site for dimers of Pax6 HD (rather than PD) (7) and a truncated form of Pax6 (33kDa) lacking the PD region has been observed in quail retina (23).

Functionally important elements in gene promoters have been identified by comparing sequences of distantly related species, a method known as "phylogenetic footprinting"(14, 27, 28). The same approach was used to compare intron 5 sequences of frog and bovine *Pax6* genes. This showed that sequences immediately flanking the alternative exon 5a are remarkably well conserved. Although we do not yet know whether frog eye tissues display similar patterns of alternative splicing to those of mammals, the striking conservation of intron sequences through more than 300Myr of evolution suggests an important function. Indeed, specific recognition of the 5′ conserved sequences by protein or nucleic acid could mask the lariat site and thereby inhibit splicing of 5a. Clearly, control of alternative splicing, perhaps varying during and after embryogenesis, could contribute to the multiple functions of a single "master" gene like *Pax6* with distinct roles at different levels in the molecular hierarchy of eye and CNS development.

ACKNOWLEDGMENTS

We thank Tom Sargent for *Xenopus* DNA, Carl Pabo for structure coordinates, and the Frederick Cancer Research and Development Center and the Scientific Computing Resource and Molecular Modelling Centers at NIH for computing support.

REFERENCES

1. Hill, R. E., Favor, J., Hogan, B. L., Ton, C. C., Saunders, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. D., and van Heyningen, V. (1991) *Nature* **354,** 522–525.
2. Quiring, R., Walldorf, U., Kloter, U., and Gehring, W. J. (1994) *Science* **265,** 785–789.
3. Li, H. S., Yang, J. M., Jacobson, R. D., Pasko, D., and Sundin, O. (1994) *Dev. Biol.* **162,** 181–194.
4. Halder, G., Callaerts, P., and Gehring, W. J. (1995) *Science* **267,** 1788–1792.
5. Nelson, L. B., Spaeth, G. L., Nowinski, T. S., Margo, C. E., and Jackson, L. (1984) *Surv. Ophthalmol.* **28,** 621–642.
6. Wistow, G. (1995) Molecular Biology and Evolution of Crystallins: Gene Recruitment and Multifunctional Proteins in the Eye Lens, Landes, Austin, TX, and Springer, New York, NY.
7. Cvekl, A., and Piatigorsky, J. (1996) *Bioessays.* **18,** 621–630.

8. Richardson, J., Cvekl, A., and Wistow, G. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 4676–4680.
9. Hanson, I., and van Heyningen, V. (1995) *Trends Genet.* **11,** 268–272.
10. Turque, N., Plaza, S., Radvanyi, F., Carriere, C., and Saule, S. (1994) *Mol. Endocrinol.* **8,** 929–938.
11. Walther, C., and Gruss, P. (1991) *Development* **113,** 1435–1449.
12. Epstein, J. A., Glaser, T., Cai, J., Jepeal, L., Walton, D. S., and Maas, R. L. (1994) *Genes Dev.* **8,** 2022–2035.
13. Koroma, B. M., Yang, J. M., and Sundin, O. H. (1997) *Invest. Ophthalmol. Vis. Sci.* **38,** 108–120.
14. Gumucio, D. L., Shelton, D. A., Bailey, W. J., Slightom, J. L., and Goodman, M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 6018–6022.
15. Spiewak Rinaudo, J. A., and Zelenka, P. S. (1992) *Exp. Cell Res.* **199,** 147–153.
16. Glaser, T., Walton, D. S., and Maas, R. L. (1992) *Nat. Genet.* **2,** 232–239.
17. Palumbi, S., Martin, A., Romano, S., McMillan, W. O., Stice, L., and Grabowski, G. (1991) The Simple Fool's Guide to PCR. Version 2, Department of Zoology and Kewalo Marine Laboratory, University of Hawaii, Honolulu, HI.
18. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
19. Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995) *Cell* **80,** 639–650.
20. Kumar, S., Tamura, K., and Nei, M. (1994) *Comput. Appl. Biosci.* **10,** 189–191.
21. Tompkins, R., and Reinschmidt, D. (1991) *Methods. Cell. Biol.* **36,** 35–44.
22. Rao, V. B. (1994) *Anal. Biochem.* **216,** 1–14.
23. Carriere, C., Plaza, S., Martin, P., Quatannens, B., Bailly, M., Stehelin, D., and Saule, S. (1993) *Mol. Cell. Biol.* **13,** 7257–7266.
24. Walther, C., Guenet, J. L., Simon, D., Deutsch, U., Jostes, B., Goulding, M. D., Plachov, D., Balling, R., and Gruss, P. (1991) *Genomics* **11,** 424–434.
25. Bennett, M. J., Schlunegger, M. P., and Eisenberg, D. (1995) *Protein Sci.* **4,** 2455–2468.
26. Hill, R. E., and Hanson, I. M. (1992) *Curr. Opinion Cell Biol.* **4,** 967–972.
27. Jaworski, C. J., Chepelinsky, A. B., and Piatigorsky, J. (1991) *J. Mol. Evol.* **33,** 495–505.
28. Wistow, G., and Graham, C. (1995) *Biochim. Biophys. Acta* **1263,** 105–113.
29. Puschel, A. W., Gruss, P., and Westerfield, M. (1992) *Development* **114,** 643–651.
30. Epstein, J., Cai, J., Glaser, T., Jepeal, L., and Maas, R. (1994) *J. Biol. Chem.* **269,** 8355–8361.