doi:10.1006/geno.2002.6854, available online at http://www.idealibrary.com on IDEAL **Article**

Characterization of the Genomic and Transcriptional Structure of the *CRX* Gene: Substantial Differences between Human and Mouse

Matthew D. Hodges,[1,*] Helena Vieira,[1] Kevin Gregory-Evans,[2] and Cheryl Y. Gregory-Evans[1,†]

[1]*Section of Cell and Molecular Biology, Faculty of Medicine, Imperial College, London SW7 2AZ, UK*
[2]*Department of Ophthalmology, The Western Eye Hospital, Marylebone Road, London NW1 5YE, UK*

**Present address: Department of Cancer Medicine, Faculty of Medicine, Imperial College, London W6 8RF, UK*

[†]To whom correspondence and reprint requests should be addressed. Fax: +44 (0)20 7594 3015. E-mail: c.gregory-evans@ic.ac.uk.

We have previously shown that there is a temporal difference in human *CRX* gene expression compared with that of mouse *Crx*. We have now characterized these genes at the genomic and transcriptional levels and here we expand on this earlier report. Human *CRX* spans 25 kb and has six exons, and mouse *Crx* spans 15 kb and has four exons. We isolated seven human and two mouse mRNAs generated by alternative splicing of a variable 5′ untranslated region. The human and mouse genes share an evolutionarily conserved promoter, which contains OTX/CRX type and SP1/AP2 binding sites and drives expression of two conserved transcripts in both species. Additionally, the human gene has a second human-specific promoter, which has OTX/CRX type binding sites and drives expression of five other transcripts. Band shift assays have shown that six of the seven candidate OTX/CRX elements bind CRX *in vitro*, possibly implying that the gene can regulate its own expression. These data may account for the differences in temporal expression *in vivo* we have previously reported between these two species.

Key Words: CRX, transcription factor, retina, alternative splicing, promoter

INTRODUCTION

Cone-rod homeobox (CRX) belongs to the OTD/OTX family of transcription factors [1–3]. It is expressed in an as yet undetermined cell type in the inner nuclear layer (INL) and in the photoreceptors of the outer nuclear layer (ONL), where it is essential for photoreceptor integrity [3–5]. It is also expressed in the pinealocytes of the pineal gland [3,4,6]. So far, 18 different mutations causing either autosomal dominant cone-rod dystrophy (adCRD), Leber congenital amaurosis (LCA), or retinitis pigmentosa (RP) have been identified in *CRX* at the cone rod dystrophy-2 (*CORD2*) locus on chromosome 19q13 [reviewed in 7]. Patients exhibit an abnormal electroretinogram consistent with degeneration of the cones and rods [1,2,8]. Both human and mouse genes encode a 299-residue protein that is 97% conserved [1–4]. The gene is evolutionarily conserved with an ortholog isolated or inferred in all gnathostomata species investigated: human, rat, mouse, oxen, chicken, and zebrafish [1–3,9–11].

Mice carrying a targeted disruption of *Crx* fail to form mature cone and rod photoreceptor outer segments and exhibit a progressive degeneration of the ONL, consistent with a loss of cones and rods [12]. Furthermore, overexpression of *Crx* was shown to disrupt the differentiation of multipotent progenitor cells, highlighting the importance of this gene in morphogenesis [4]. *Crx*-deficient mice also had a reduced expression of the genes necessary for phototransduction and absence of expression of genes necessary for melatonin synthesis [12,13]. Thus, a role for *Crx* in both the phototransduction cascade and the melatonin synthesis pathway is evident. *Crx*-deficient mice also exhibit abnormal circadian entrainment, a feature not yet described in patients with *CRX* mutations.

The OTD/OTX family of transcription factors are all structurally similar to the *Drosophila melanogaster* bicoid protein, having a lysine residue at position 50 of the homeodomain (position 9 of the recognition helix). The closest family members to CRX are the OTX and OTD proteins, which have a consensus-binding site of CCTAATCCT [14]. A putative CRX/Crx binding site with the consensus sequence (C/TTAATC/T) was

0888-7543/02 $35.00

doi:10.1006/geno.2002.6854, available online at http://www.idealibrary.com on IDEAL

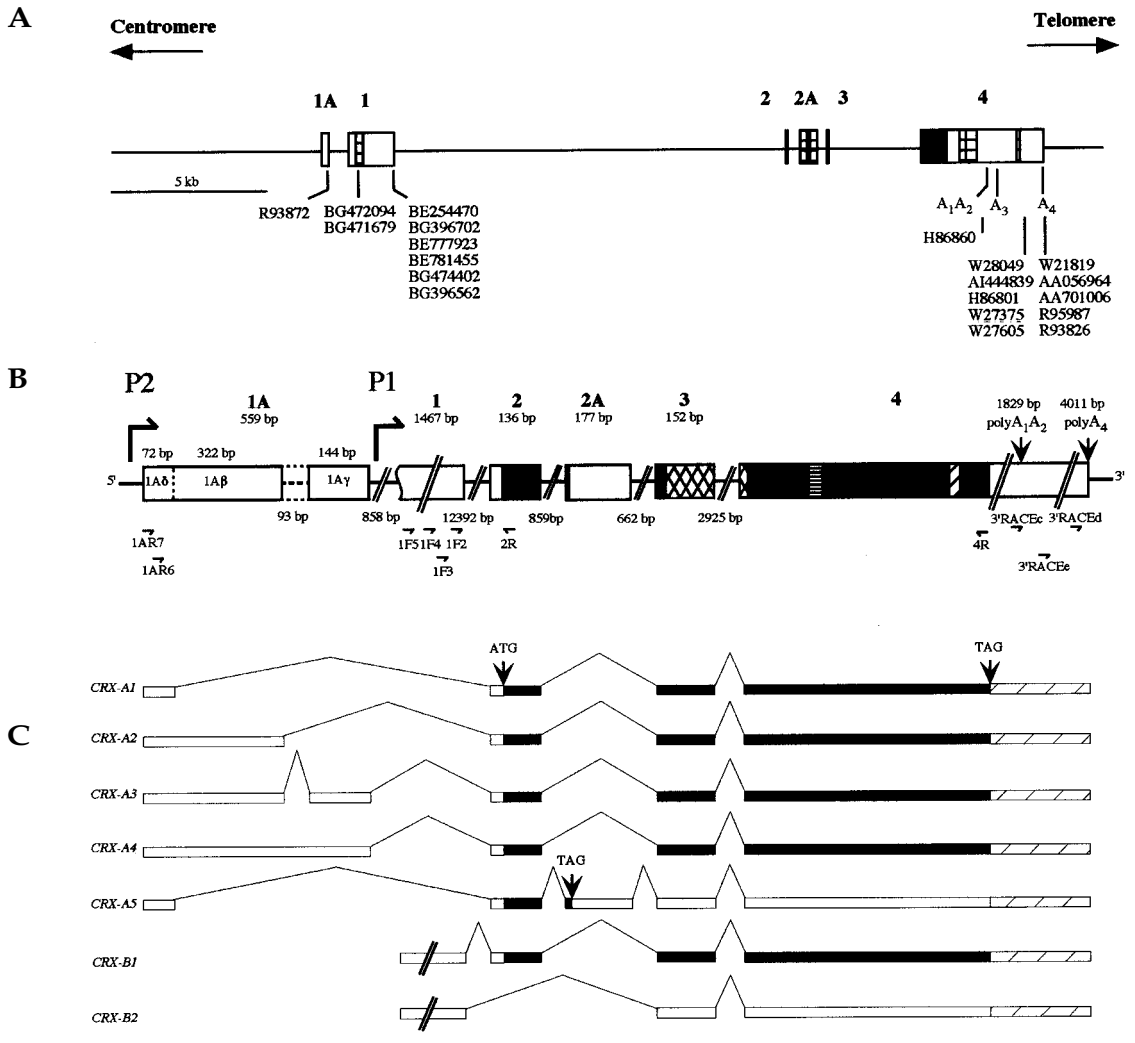

FIG. 1. Genomic structure and organization at the human *CRX* locus. (A) Genomic structure of the *CRX* gene. The gene has six exons: 1A, 1, 2, 2A, 3, and 4, spanning 25 kb of DNA. Open boxes indicate noncoding exons and filled boxes indicate coding exons; the crosshatched region is *Alu-Sq* repeat match surrounding exon 2A and in 3′-UTR. A1–A4 are four consensus polyadenylation signals. Approximate positions of ESTs within the 5′ and 3′ regions are indicated. Intron sizes were derived from the genomic sequence AC008745. (B) Exon–intron boundaries and protein motifs. Lines indicate introns with size below. Open boxes and filled boxes indicate noncoding and coding exons, respectively, with size of exon above; crosshatched, horizontal, and diagonal filled boxes indicate the homeodomain, WSP, and OTX motifs, respectively. The 5′ end of exon 1 has not been precisely determined. The 93-bp intron between variant exons 1Aβ and 1Aγ may be omitted through alternative splicing of exon 1A. Positions of putative promoters *CRX*P1 and *CRX*P2 are shown. Two polyadenylation signals used by *CRX* are indicated. Primer positions annotated by half arrows. (C) *CRX* retinal transcripts. Seven different transcripts have been isolated from retinal cDNA at the *CRX* locus. Filled and open boxes indicate translated and non-translated mRNA, respectively. Thin lines indicate introns and hashed open boxes indicate 3′ untranslated sequence not investigated in this study. The initiation and termination sites are indicated with arrows.

identified through *in vitro* binding studies of several photoreceptor and pineal-specific genes that were thought to be targets for these genes, including those encoding the interstitial retinol-binding protein (*IRBP*), rhodopsin, and the pineal night-specific ATPase (*PINA*) [3,4,6]. Previous work from our laboratory [5] investigated the temporal and spatial expression patterns of *CRX/Crx* and their putative target genes. One of these was the gene encoding the β-subunit of phosphodiesterase (*PDEB*), which is expressed before *CRX*, even though it is driven *in vitro* by CRX [3]. Additionally, OTX2 was identified in preference to CRX in a yeast one-hybrid assay using *IRBP* as bait [15] and its expression remained unaltered in the *Crx*-null mouse [12]. More recent studies have shown that several *Crx*-type binding sites upstream of the chicken hydroxyindole-O-methyltransferase (*Hiomt*) gene will not bind chicken *Crx*, even though these sites, which are completely conserved in humans, are CRX binding sites in the human gene *HIOMT* and vice versa [10]. Thus, it seems likely that not

TABLE 1: Exon–intron splice boundaries of human *CRX*

Exon	Size (bp)	Acceptor splice site	Donor splice site	Intron size (bp)
1Aα	559	–	TCAGCCAG**gc**ctgtagc	14717[a]
1Aβ	322	–	TCCCCAGG**gc**aggtgat	14954[a] or 93[b]
1Aγ	144	cctgctgc**ag**GGCCAGGG	TCAGCCAG**gc**ctgtagc	14717[a]
1Aδ	77	–	GATTGGAT**gt**aagtggt	15276[a]
1	> 1473	–	TTCTGAAG**gt**gagcgtc	12392[a] or 14226[c]
2	136	tctcttgc**ag**GCCCCCTG	CTACCCAA**gt**gagtaca	1698[d] or 858[e]
2A	177	cgagacac**ag**TTTAGACG	TAAGAGAA**gt**gagtact	663
3	152	cccaccc**ag**GCGCCCCC	GGGTTCAG**gt**ggggtgg	2925
4[f]	1829	tatcccc**ag**GTTTGGTT	–	–
4[f]	4011	tatcccc**ag**GTTTGGTT	–	–

[a]Exons spliced to exon 2.
[b]If exon 1Aβ is spliced to exon 1Aγ there is a 93-bp intron.
[c]If exon 1 is spliced to exon 3 there is a 14,226-bp intron.
[d]Exon spliced to exon 3.
[e]If exon 2A is spliced to exon 2 there is a 859-bp intron.
[f]Size of exon 4 depends on whether polyA_1A_2 (1829 bp) or polyA_4 (4011 bp) is used.

all the putative CRX binding sites identified by *in vitro* studies are actual *in vivo* targets for CRX. There are several ways of interpreting this: other proteins drive expression of these genes and not CRX, for example the retinal homeobox (Rx) and the empty spiracles-related retinal homeobox (Erx) have been shown to bind to the Ret-1 site of rhodopsin at a greater affinity than CRX [16]; CRX expression at the transcriptional and translational levels is modulated in either *cis* or *trans*; the neural retina leucine zipper (NRL) [17], p300/CBP [18], RX [16], phosducin [19], and QRX [20] have all been reported as interacting and altering *CRX* activity in *trans*; CRX function is redundant with another similar protein, such as OTX2, RX, and NRL, for example CRX will bind to the PCE-1/Ret-1 element [3,16] and a *Nrl* null mouse has reduced expression of several genes also seen in a *Crx* null mouse [12,21].

The published transcript sizes of human and mouse *CRX* have been reported as 1.4 kb and 1.6 kb, respectively. These do not correspond to the predicted sizes of transcripts observed in a northern hybridization study, which were 4.5 kb and 3 kb for human *CRX* [1] and 3 kb for mouse *Crx* [4]. Here we investigate *CRX* at the transcriptional level to sequence the remainder of the gene and identify the *CRX* promoter in human and mouse.

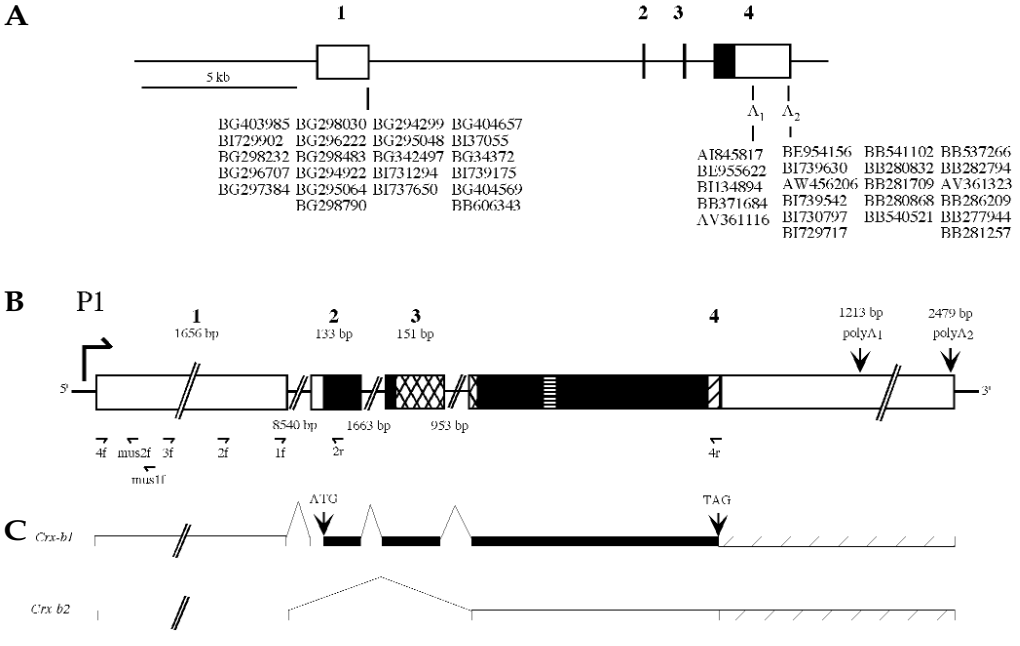

FIG. 2. Genomic structure and organization at the mouse *Crx* locus. (A) Genomic structure of *Crx*. The gene has four exons spanning 15 kb of DNA. Open boxes indicate noncoding exons and filled boxes indicate coding exons. A_1 and A_2 are consensus polyadenylation signals. Approximate positions of ESTs within the 5′ and 3′ regions are indicated. Intron sizes were derived from genomic sequence of AC08167 and AC073748. (B) Exon–intron boundaries and protein motifs. Lines indicate introns, with size indicated below. Open boxes and filled boxes indicate noncoding and coding exons, respectively, with size of exon indicated above; crosshatched, horizontal, and diagonal filled boxes indicate the homeodomain, WSP and OTX motifs, respectively. The position of putative promoter P1 is indicated. Two polyadenylation signals used by *Crx* are indicated. Primer positions are annotated by half arrows.(C) *Crx* transcripts. Two different transcripts have been isolated from retinal cDNA. Filled boxes and open boxes indicate translated and non-translated mRNA, respectively. Thin lines indicate introns and hashed open boxes indicate 3′ untranslated sequence not investigated in this study. The initiation and termination sites are indicated with arrows.

doi:10.1006/geno.2002.6854, available online at http://www.idealibrary.com on IDEAL

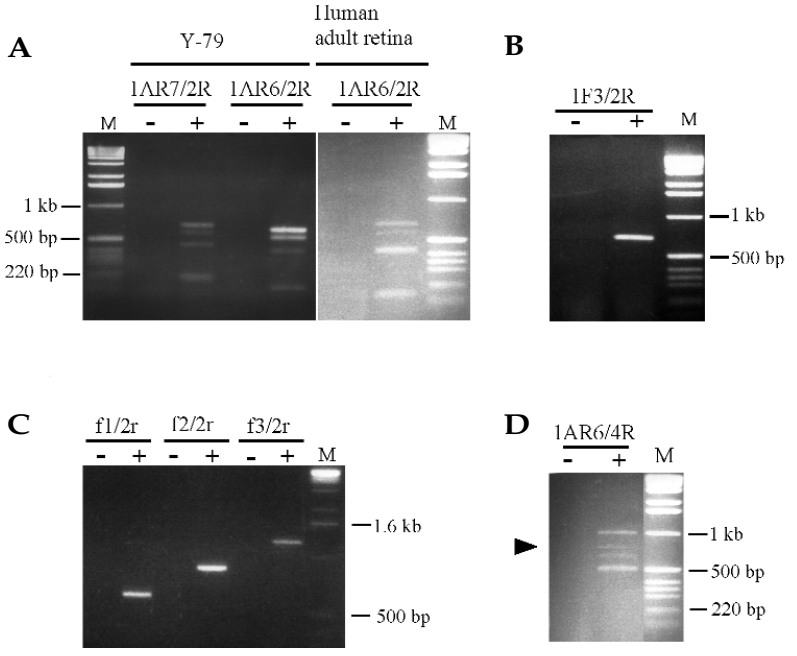

FIG. 3. RT-PCR analysis of human and mouse *CRX/Crx* transcripts. (A) Human RT-PCR amplifying exon: 1Aα (largest band), 1Aβγ (second band), 1Aβ (third band), and 1Aδ (smallest band); +/- denotes presence or absence of RT in PCR mix. M, kb ladder (Invitrogen). (B) RT-PCR with human retinal cDNA, amplifying partial sequence of *CRX* exon 1. (C) RT-PCR amplifying mouse *Crx* exon 1 from C57BL/6 cDNA. (D) RT-PCR from human retinal cDNA amplifying a splice variant containing exon 2A (850 bp; arrowhead). Other products correspond to exons 1Aα, 1Aβ, and 1Aβγ.

RESULTS

Human *CRX* 3′-UTR Identification

A BLAST search with AF024711 (the *CRX* cDNA sequence) revealed that the entire *CRX* sequence was within the genomic clone AC008745. NIX analysis [22] isolated human retinal and pineal ESTs that overlapped with AF024711 but extended in a 3′ direction. Furthermore, four consensus poly(A) sequences were within the extended region: A_1A_2 (ATTAAAAATAAA), A_3 (AATAAA), and A_4 (AATTAAA) (Fig. 1A). This implied that there was additional *CRX* sequence at the 3′ end of AF024711. We used 3′-RACE to amplify two retinal 3′-UTRs from *CRX* of 1.2 kb and 3.4 kb (data not shown). Cloning and sequencing of these products showed that they terminated after the polyA_1A_2 and polyA_4 sites, respectively. Several EST clones that overlapped the 3′-RACE sequence were sequenced and similarly terminated adjacent to the polyA_1A_2 or polyA_4 sites. No *CRX* mRNAs terminated at the polyA_3 site from retinal cDNA as determined by 3′-RACE or from retinal or pineal gland cDNA as determined from ESTs in the database. BLAST analysis of the entire 3′-UTR sequence revealed two regions with high identity to the *Alu-sq* repeat sequence (Fig. 1A).

Alternatively Spliced Upstream Human *CRX* Exons

We performed 5′-RACE with primers designed from the 5′ end of AF024711. A 300-bp product was generated, cloned, and sequenced (data not shown). Novel sequence from this clone had an identical match to a region ~ 14.7 kb upstream of exon 2 within the sequence of AC008745. RT-PCR was carried out using Y-79 and human adult retinal cDNA with primers designed from this sequence and amplified a PCR product, confirming this as a *CRX* spliced exon, which we have termed exon 1A. Additionally, several ESTs were present in the database that had exon 1A spliced to exon 2 (Fig. 1A) and the acceptor splice site of exon 2 and donor splice of exon 1A fitted a consensus splice sequence (Table 1). Further RT-PCR primers were designed from AC008745 upstream of exon 1A sequence and primed with a reverse primer in exon 2. The most upstream primer that would amplify a product was 1AR7. In the course of characterizing exon 1A it became apparent that it was alternatively spliced, as we obtained four RT-PCR products with primers 1AR7/2R and 1AR6/2R (Fig. 3A). Cloning and sequencing of these products identified four variants, which we have termed exon 1Aα (559 nt; full-length exon 1A), 1Aβ (322 nt), exon 1Aγ (144 bp), and exon 1Aδ (72 nt). Exon 1Aγ has only been isolated in conjunction with exon 1Aβ, such that a 93-bp intron is generated in exon 1A (Fig. 1B). We termed the exon 1Aδ splice variant transcript *CRX-A1*, the exon 1Aβ variant, *CRX-A2*, the exon 1Aβγ variant, *CRX-A3*, and the exon 1Aα variant, *CRX-A4* (Fig. 1C).

Examination of ESTs in the database revealed at least six that had a sequence unrelated to exon 1A (Fig. 1A), but were spliced to exon 2. A BLAST search of AC008745 with this unrelated sequence identified a region ~ 12.4 kb upstream of exon 2. We carried out RT-PCR with the 1F2 primer designed from this sequence and the 2R reverse primer in exon 2, and amplified a product from adult retinal and Y-79 cDNA (data not shown). Additionally, the donor splice site of exon 1 matched a consensus sequence (Table 1), confirming this to be a *CRX* exon, which we termed exon 1 (Fig. 1A). We were able to extend exon 1 in a 5′ direction through RT-PCR, with primers 1F3 and 1F4 (Fig. 3B). Exon 1 is presumably not alternatively spliced in adult retina as determined by the single amplification product generated with each primer pair, neither is exon 1 spliced to exon 1A, as RT-PCR primers designed to span the two exons did not amplify a product (data not shown). Two human retinal ESTs, BG472094 and BG471679, overlapped with the sequence of exon 1 and extended the sequence in a 5′ direction (Fig. 1A). An RT-PCR product was generated from adult retina cDNA with a forward primer (1F5) from within the EST overlap and primed with the exon

TABLE 2: Oligonucleotides used in band shift assay

Name	Description of sequence	Oligonucleotide sequence (5′-3′)
CRX-1	Putative CRX binding-site in *CRX* promoter 1	CACAAGGAG**GGATTA**CCGGGCGTG
CRX-2	Putative CRX binding-site in *CRX* promoter 1	CTGTAGCC**TTAATC**TCTCCTAGCA
CRX-3	Putative CRX binding-site in *CRX* promoter 2	AAATCTAAA**GATTAG**TTTGAGAAG
CRX-4	Putative CRX binding-site in *CRX* promoter 2	GAGTCTCCT**TTAATT**TCTTCCAGC
CRX-5	Putative CRX binding-site in *CRX* promoter 2	ACATTTTC**TTAATCC**TTAACCCCA
Crx-6	Putative Crx binding-site in *Crx* promoter 1	CAGAGGCC**CTAATC**TCTCCTAGCA
Crx-7	Putative Crx binding-site in *Crx* promoter 1	GGTGAGCAG**GGATTAG**AGTGTGCG
TRE F2	Putative TRE(F2) site in *Crx* promoter 1	GACAAGCAG**TGACC**CAGTGGTGAG
BAT-1	CRX binding site in bovine rhodopsin promoter	GTGA**GGATTAA**TA**TGATTAA**TAAC

For each oligonucleotide there is a reverse complementary sequence not shown. Bold sequence highlights the putative functional sequence.

2 reverse primer, 2R, confirming the validity of these ESTs (data not shown). In total we were able to determine exon 1 to be more than 1467 bp. BLAST analysis of this exon shows it to have a high identity match to the *Alu-sq* repeat sequence over a short region (Fig. 1A). We have termed the exon 1 splice variant transcript *CRX-B1* (Fig. 1C). Additional primers designed upstream of this exon did not generate RT-PCR products. Similarly, 5′-RACE experiments with primers designed near the 5′ end of the exon were unable to amplify a product.

Mouse *Crx* 3′-UTR Identification

A BLAST search with the mouse *Crx* sequence (U77615) identified two genomic sequences, AC073748 and AC080167, that together spanned the entire *Crx* cDNA sequence. More than 20 mouse retinal EST clones (Fig. 2A) overlapped with the cDNA sequence and extended downstream of the 3′ end of U77615. Within this region two putative poly(A) signals, A_1 and A_2, were also present. PolyA_1 correspondes to the 3′ end of U77615 and polyA_2 is presumably from a longer *Crx* transcript. The two 3′-UTRs are predicted to be 568 bp and 1.8 kb, respectively (data not shown). An alignment of the human polyA_4 3′-UTR and mouse polyA_2 3′-UTR identified a 32-bp sequence that was 86% conserved. This region of homology was adjacent to but downstream of the first poly(A) signal in both genes and probably corresponds to a G/U-rich element [23] (Fig. 4A). A smaller core motif was 100% conserved between human, mouse, canine *Crx* (AF454668), and a rat *Crx* genomic clone (AC099348) in all cases after a putative poly(A) signal (Fig. 4A; unpublished data). No other OTX/OTD family homology was observed in the 3′-UTR.

Mouse *Crx* 5′-UTR Identification

A BLAST search with human *CRX* exon 1 identified a highly homologous mouse sequence present in AC080167 and AC073748 that was ~ 8.5 kb upstream of *Crx* exon 2 (previously exon 1). The oligonucleotide primer 1f was designed from within this region and RT-PCR was carried out with the reverse primer 2r from *Crx* exon 2 on C57BL/6 adult retinal cDNA (Fig. 3C). In addition there were more than 22 mouse retinal ESTs (Fig. 2A) that matched this sequence and the splice site between exon 1 and 2 conformed to the GT/AG splice consensus (-CCAAGgtagtc...intron...Py7gtagGTGCC-), confirming this as a *Crx* exon, which we termed exon 1. We were able to extend the size of exon 1 in a 5′ direction through RT-PCR, with forward primers 2f and 3f and reverse primer 2r (Fig. 3C), and 5′-RACE, using primer pairs mus1F and mus2f (data not shown). The most 5′ RT-PCR primer (4f) that would amplify a product from adult C57BL/6 retinal cDNA was 4 bp away from the most 5′-RACE clone sequenced (Fig. 5B). We propose this as the start of mouse *Crx* transcription and determine *Crx* exon 1 to be

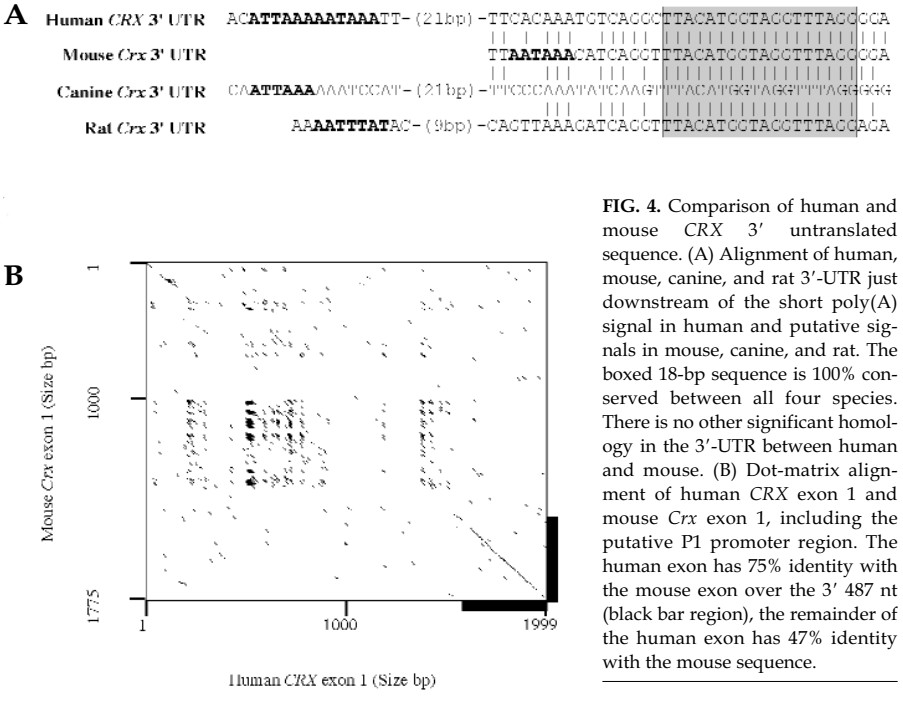

FIG. 4. Comparison of human and mouse *CRX* 3′ untranslated sequence. (A) Alignment of human, mouse, canine, and rat 3′-UTR just downstream of the short poly(A) signal in human and putative signals in mouse, canine, and rat. The boxed 18-bp sequence is 100% conserved between all four species. There is no other significant homology in the 3′-UTR between human and mouse. (B) Dot-matrix alignment of human *CRX* exon 1 and mouse *Crx* exon 1, including the putative P1 promoter region. The human exon has 75% identity with the mouse exon over the 3′ 487 nt (black bar region), the remainder of the human exon has 47% identity with the mouse sequence.

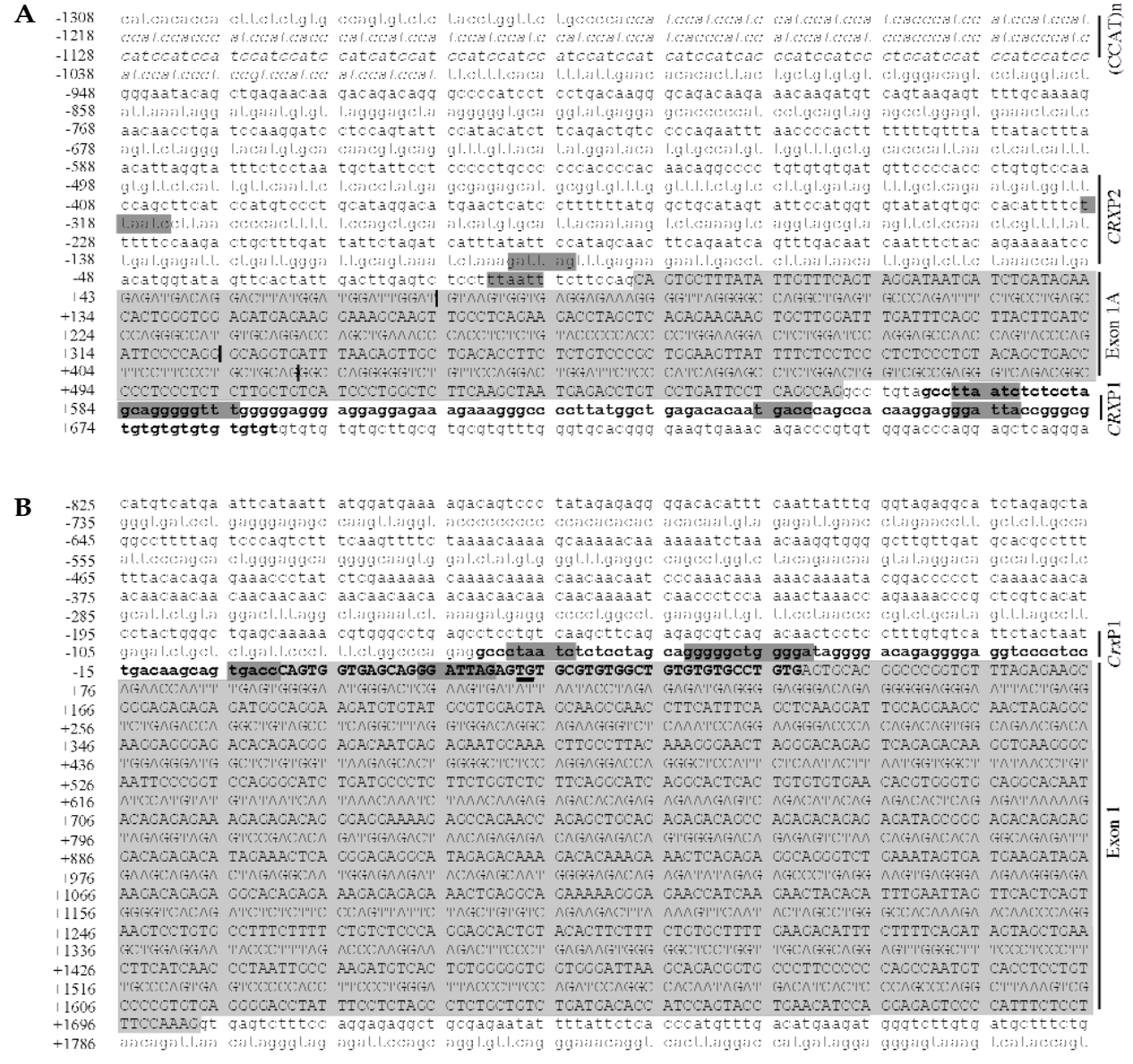

FIG. 5. Genomic upstream regions of human and mouse *CRX*. (A) Human genomic region encompassing *CRX* exon 1A and putative 5′ end of exon 1. Light gray uppercase boxed region is exon 1A, vertical bars indicate splice sites used in exons 1Aβ, 1Aγ, and 1Aδ. Putative human-specific promoter *CRX*P2 with three CRX/OTX2 type binding sites are boxed and positioned relative to the putative start of transcription of CRX exon 1A: (CRX-4) TTAATT –14 to –9, (CRX-3) GATTAG –112 to –107, and (CRX-5) TTAATC –319 to –314. Bold sequence (+568 to +688) is 70% conserved in mouse and is the putative 5′ end of human exon 1 (not shown) and promoter *CRX*P1. The position of two conserved CRX/OTX2 type binding sites (CRX-2) TTAATC +472 to +477 and (CRX-1) GGATTA +662 to +667, a TRE(F2) site TGACC +643 to +647, and a SP1/AP2 site, GCAGGGGGTT +584 to +594, are boxed and positioned relative to the start of transcription of CRX exon 1A, within this highlighted sequence. A tetranucleotide $(ccat)_n$ repeat is in italics. (B) Mouse genomic region encompassing *Crx* exon 1. Sequence of exon 1 is boxed in light gray. A 120-bp region in bold (–72 to +48, numbered relative to start of transcription of *Crx* exon 1) is 70% conserved with human sequence and the position of two conserved CRX/OTX2 type binding sites (Crx-6) CTAATC –69 to –64 and (Crx-7) GGATTAG +14 to +20, TRE(F2) TGACC –5 to –1, and a SP1 site GGGGGCTGGGGA –53 to –42 are boxed in this putative *Crx*P1 promoter. Double underlined TG (+23 to +24) indicates the 5′ end of two independent 5′-RACE sequences.

1656 bp in size. A *Crx* mouse sequence related to human *CRX* exon 1A was not identified through BLAST analysis and there is no evidence that human *CRX* exon 1A is evolutionarily conserved in mouse.

Novel Transcripts from *CRX/Crx*

In the progress of this study, a search for additional *CRX/Crx* ESTs identified the mouse ESTs BI735039 and BG296222, which consisted of *Crx* exon 1 spliced directly to exon 4. To verify the existence of this mRNA and search for additional

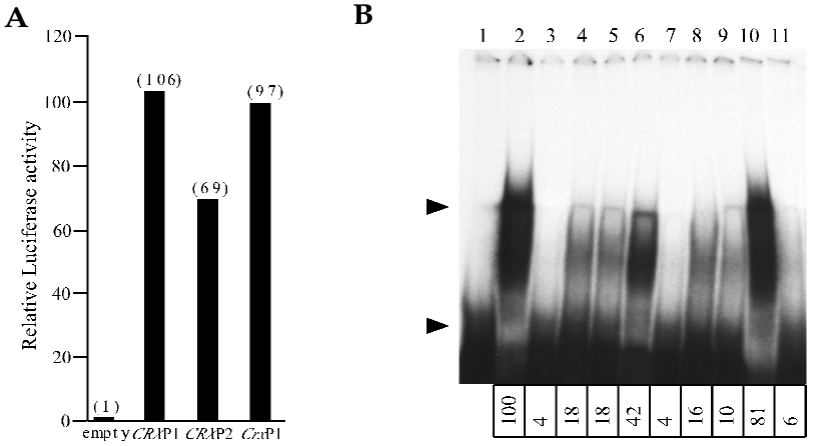

FIG. 6. Promoter analysis. (A) *CRX/Crx* promoter assays. Lane 1, empty pGL3 basic vector; lane 2, *CRX*P1; lane 3, *CRX*P2; lane 4, *Crx*P1. Luciferase values were measured in confluent populations of cells and expressed as relative light units per microgram of protein (B). Competition binding assay of CRX-HD on the BAT-1 element and OTX/OTD type binding sites from the human and mouse *CRX/Crx* promoter. CRX-HD-GST was generated in BL21 cells and the GST portion removed. CRX-HD protein (10 ng; lanes 2–11) was pre-incubated at 4°C with 200 pM of cold competitor DNA (see Table 2 for oligonucleotide sequences). Competitor DNA was as follows: CRX-1 to CRX-5 (lanes 3–7) and Crx-6 and Crx-7 (lanes 8, 9), TRE(F2) (lane 10), and BAT-1 (lane 11). Radiolabeled BAT-1 probe (1 pm) was added to all lanes (and lane 1) after the pre-incubation period and left to incubate for a further 30 minutes at 4°C. Samples were resolved on a non-denaturing 4.5% polyacrylamide gel and exposed on a phosphoimager. Values below lanes 2–11 are quantified intensities of retarded band as a percentage of BAT-1 CRX-HD complex with no competitor DNA (lane 2). Upper arrow indicates protein-DNA complex and lower arrow unbound BAT-1 probe. The experiment was repeated three times with similar results.

CRX/Crx splice variants, we used RT-PCR to amplify between exon 1 of *CRX/Crx* (and exon 1A of *CRX*) and exon 4 of *CRX/Crx*, using mouse primers 1f/4r and human primers 1AR6/4R and 1F2/4R.

From human retinal cDNA we amplified an additional *CRX* mRNA (850 bp) with the 1AR6 and 4R oligonucleotide primers that could not be accounted for by the alternative splicing previously identified in exon 1A (Fig. 3D). Cloning and sequencing of this mRNA revealed that it contained a novel exon that was between exons 2 and 3. We called this exon 2A and the transcript it was in *CRX-A5* (Fig. 1C). Exon 2A has 88% sequence identity with the *Alu-sq* repeat [24]. The 5′ and 3′ ends of exon 2A fitted the consensus splice sequence (Table 1). The *Alu* identity extends outside of exon 2A into the intronic sequence between exon 2 and 3 and may be the result of an ancestral insertion of a full-length *Alu* sequence into this intron.

We were also able to RT-PCR amplify two RT-PCR products with the 1F2/4R primer pair: a product comprising exons 1, 2, 3, and 4 (which was amplified at high intensity), which we had previously identified as *CRX-B1*; and a smaller product (amplified at ~ 10% the intensity of *CRX-B1*) consisting of exon 1 spliced to exon 3, which we called *CRX-B2* (Fig. 1C).

In total we were able to detect seven human adult retinal transcripts: transcript *CRX-A1* to *CRX-A4*, comprising variant forms of exon 1A spliced at the 5′ end of a common coding structure (Fig. 1C); transcript *CRX-A5*, which has the novel exon 2A spliced after exon 2; transcript *CRX-B1*, which had exon 1 at the 5′ end, but otherwise the same coding exons as transcripts *CRX-A1* to *CRX-A4*; and transcript *CRX-B2*, which had exon 1 at the 5′ spliced directly to exons 3 and 4.

We were able to amplify two mRNA transcripts from mouse: a major transcript corresponding to exons 1, 2, 3, and 4 spliced together, which we termed transcript *Crx-b1*; and a minor transcript at < 1% the intensity of *Crx-b1*, which consisted of exon 1 spliced directly to exon 4 (data not shown), which we termed *Crx-b2* (Fig. 2C).

CRX/Crx Promoter Identification

Upstream of mouse *Crx* exon 1 we identified a transcription factor binding site for an Otd/Otx paired-like protein with the sequence CTAATCT (–69 to –64). A second site was found on the reverse strand, GGATTAG (+14 to +20), and a SP1/AP2 type sequence (GGGGGCTGGGGA; –53 to –42) was identified close to the start of transcription and a TRE(F2) partial sequence TGACC (–5 to –1). This region was highly conserved in the human *CRX* gene; in total, a 120-bp region of mouse *Crx* sequence, spanning 48 bp of exon 1 and 72 bp upstream of exon 1, had 70% sequence identity with a 121-bp region upstream of human *CRX* exon 1 (Fig. 5B). We have termed this common putative promoter P1 (Figs. 1B and 2B). Human *CRX*P1, conserves the two Otd/Otx *paired*-like binding sites, the Sp1/AP2 site and TRE(F2) partial sequence (Fig. 5A). A comparison of mouse *Crx* exon 1 with the proposed *CRX* full-length exon 1 (assuming that the 5′ end of human *CRX* exon 1 is the same as mouse *Crx* exon 1) shows 47% homology over the entire sequence. Over the most 3′ region there is a 487-bp region that has 75% homology (Fig. 4B).

The most 5′ RT-PCR primer from human exon 1A that would amplify a product, when primed with a reverse primer in either exon 1 or 4, was 1AR7. The 5′ end of this primer therefore defines the 5′ end of the exon. This sequence is identical to the consensus acceptor splice site (Fig. 5A) and led us to believe that further upstream exons may be present. We searched ~ 80 kb of genomic sequence upstream of exon 1A within the sequence of AC008745 for additional exons and promoter-like sequence without success. We are therefore unable to discount the possibility that there are additional upstream exons of *CRX*. Analysis of the sequence immediately upstream of human *CRX* exon 1A, the human specific promoter P2 (*CRX*P2), identified three putative binding sites for OTX/OTD paired-like transcription factors: TTAATT (–14 to –9), GATTAG (–112 to –107) on the reverse strand, and TTAATC (–319 to –314). We did not find TATA or CAAT-like boxes in close proximity to the 5′ end of exon 1A.

The putative *Crx* mouse promoter P1 (*Crx*P1), putative *CRX* human promoter P1 (*CRX*P1), and putative *CRX* human promoter P2 (*CRX*P2) were cloned upstream of a luciferase reporter gene. Initial experiments in COS-7 cells suggested that these elements do act as promoters (Fig. 6A). Luciferase activity from human *CRX*P1 and mouse *Crx*P1 resulted in similar levels of transactivation of the reporter, whereas the human *CRX*P2 fragment resulted in one-third less activity compared with the other two promoters.

TABLE 3: *CRX*/*Crx* retinal transcript sizes

Transcript	Short[a] poly(A) (bp)	Long[b] poly(A) (bp)	Observed transcript size (bp)	Predicted poly(A)
CRX-A1	2189	**4371**	**~4500** and ~3000	$PolyA_4$
CRX-A2	2439	**4621**	**~4500** and ~3000	$PolyA_4$
CRX-A3	2583	**4765**	**~4500** and ~3000	$PolyA_4$
CRX-A4	2676	**4858**	**~4500** and ~3000	$PolyA_4$
CRX-A5	2366	**4548**	**~4500** and ~3000	$PolyA_4$
CRX-B1	**3590**	5772	~4500 and **~3000**	$PolyA_1A_2$
CRX-B2	**3454**	5636	~4500 and **~3000**	$PolyA_1A_2$
Crx-b1	**3153**	4419	**~3000**	$PolyA_1$
Crx-b2	2869	4135	~3000	rare transcript

Bold numbers indicate best matches between observed and predicted sizes.
[a]Short poly(A) refers to $polyA_1A_2$ for *CRX* and $polyA_1$ for *Crx*.
[b]Long poly(A) refers to $polyA_4$ for *CRX* and $polyA_2$ for *Crx*.

CRX Binds to Upstream OTX/CRX Elements in Human and Mouse *CRX* Promoters

The binding site of CRX has been well documented [3,4,6,10,13]. To initiate an analysis of the human and mouse *CRX* promoters, we wanted to see if CRX could bind upstream in the putative promoter region to the OTX/OTD elements we had identified in human and mouse (Table 2), and thus determine if CRX could auto-regulate. To see which of the sequences would bind to CRX, we tested their ability to compete for binding of CRX-HD, the homeodomain portion of CRX, on a BAT-1 probe (a nucleotide sequence that was known to bind to CRX with high affinity) [3], and the candidate sequences. The addition of labeled BAT-1 probe was able to efficiently displace the protein–DNA complexes generated by CRX-4, and an unrelated control sequence TRE(F2), implying that these were not targets for CRX. All other candidate sequences generated a high degree of complex formation, particularly CRX-1 and CRX-5, as determined by the inability of the BAT-1 probe to displace these sequences from the CRX-HD protein. Moreover, CRX-1 and CRX-5 bound CRX-HD as efficiently as BAT-1 (Fig. 6B).

Size of *CRX/Crx* Retinal Transcripts

We had successfully PCR-amplified human *CRX* specific products from Y-79 and human retinal cDNA with oligonucleotide primers designed from the noncoding exons 1A and 1 with reverse primers designed from exon 2 and exon 4 (Figs. 3A, 3B, and 3D). Similarly, we were able to PCR amplify *CRX*-specific products with oligonucleotide primers designed in the 3′-UTR for 3′-RACE with human retinal cDNA (data not shown). Primer pairs between the 5′ end and 3′ region of *CRX* would not, however, amplify a *CRX*-specific product from human retinal or Y-79 cDNA. As a consequence it was impossible to determine which 3′-UTR and promoter was associated with each of the seven retinal transcripts isolated. Nevertheless, based on the size of the *CRX* 3′-UTR sequence (AF335247S2), the exon structure of each transcript (Fig. 1C), and the size of predicted full-length *CRX* transcript (~ 3 kb and 4.5 kb) [1], it was possible to predict the sizes of the retinal transcripts and determine which 3′-UTR was associated with each transcript. In an identical manner we were able to extend this technique for the two mouse *Crx* transcripts isolated (Table 3).

Amplification of *CRX* transcripts *CRX-A1* to *-A5* generated a RT-PCR product of similar intensity when visualized on an agarose gel (Figs. 3A and 3D). It is reasonable to surmise that these transcripts account for the ~ 4.5-kb band seen on a northern blot (Table 3) with the long 3′-UTR ($polyA_4$). Further evidence of this is available from the EST database, where the 5′ end of a *CRX* retinal EST (R93872) is from exon 1A and the 3′ end of this clone (R93826) is from the $polyA_4$ region. *CRX-B1* and *CRX-B2* together probably account for the band seen at ~ 3 kb with the short 3′-UTR ($polyA_1A_2$). Mouse *Crx* transcript *Crx-b1* probably accounts for the transcript of ~ 3 kb seen on a northern blot [4] with the short 3′-UTR ($polyA_1$). Transcript *Crx-b2* may also correspond to this ~ 3 kb transcript, but this is not certain, as we would expect this transcript to be expressed at a low level, judging from our RT-PCR results, and to be undetectable on a northern blot. We have predicted a second, longer *Crx* 3′-UTR that is supported by retinal ESTs in this region (Fig. 2A) and therefore *Crx-b2* may have this longer 3′-UTR ($polyA_2$). Alternatively, both *Crx-b1* and *Crx-b2* have the short 3′-UTR and there are other *Crx* retinal mRNA transcripts that we failed to isolate.

Mutational Screening of *CORD2* Family

The original adCRD family [25] defining the *CORD2* locus containing *CRX* did not have a detectable mutation in the coding region. We therefore investigated whether a mutation in the noncoding *CRX* regions may be attributable to the dystrophy observed in the family. Oligonucleotide primers to screen exon 1A, exon 1, exon 2A (including splice sites), and putative poly(A) signals in the 3′ UTR region were designed from the genomic sequence AC008745. Exon 1A variants 1Aα, 1Aβ and 1Aβγ were amplified with a primer pair (5′-CTGATAGAA-GAGATGACAGGA-3′ and 5′-CTTCCAGCGGGACAGA-GAA-3′), generating a 335-bp product. Within this sequence we identified a C to A single-nucleotide polymorphism (SNP; A216C). The SNP also creates or abolishes a *Mse*I site that was found not to segregate with adCRD in the *CORD2* family (data

not shown). No pathological mutations were identified from the new *CRX* sequences in the *CORD2* family.

Functional Analysis of *CRX/Crx* Upstream Exons

Human transcripts *CRX-A1* to *CRX-A4* and *CRX-B1* (Fig. 1C) have a common sequence comprising exons 2, 3, and 4. Exon 2 contains the initiation codon, implying that these transcripts encode a full-length CRX protein of 299 residues [1]; however, these transcripts all have an additional, different exon spliced to the 5′ end. Transcript *CRX-A1* has exon 1Aδ spliced to the 5′ end, which has three in-frame stop codons when in the same open reading frame (ORF) as CRX. Therefore, this exon is deemed to be noncoding. Transcripts *CRX-A2* to *CRX-A4* contain the A216C SNP that introduces or removes a termination codon within the additional sequence encoded by variant exons 1Aα, 1Aβ and 1Aβγ. These exons are therefore judged to be noncoding, as this SNP has been shown not to segregate with adCRD. Transcript *CRX-B1* has exon 1 spliced to the 5′ end of the common coding sequence. Exon 1 has several weak Kozak consensus sequences, such as RNNATGG and YNNATGG [26], each of which may generate small ORFs, all of which terminate before the previously reported *CRX* initiation complex. Therefore, this exon is deemed to be noncoding when in-frame with the major CRX ORF.

Transcript *CRX-A5* comprises exons 1Aδ, 2, 2A, 3, and 4. Splicing of exon 2 to exon 2A instead of exon 3 generates an AGC to ATT change at codon 34 (AF024711 nucleotides 101–102). The next codon introduced from exon 2A is a termination codon (TAG). Exon 1Aδ, as previously shown in transcript *CRX-A1*, does not offer any protein-coding potential. Thus, a CRX-A5 peptide would be predicted to have the first 34 residues of CRX with isoleucine replacing serine at the carboxy terminus. An investigation into the domain structure of CRX did specifically look at a truncated CRX protein comprising the first 54 residues activity [27]. This construct did not bind the BAT-1 element, but may transactivate the rhodopsin proximal promoter region at very low levels. All the *CRX* mutations identified in LCA, adCRD, or RP patients, except the Pro9 (1-bp ins) mutation that was identified in an LCA patient [28], affect the primary *CRX* coding sequence after codon 35. The *CRX-A5* transcript and CRX-A5 peptide would therefore remain unaffected in these patients.

Transcript *CRX-B2* consists of exons 1, 3, and 4. It does not have exon 2 and consequently loses the recognized *CRX* initiation codon [1]. A search for other possible initiation sites in all ORFs identified poor Kozak consensus sequences. For this reason it is thought that transcript *CRX-B2* is noncoding. Similar to transcript *CRX-A5*, the *CRX* mutation Pro9 (1-bp ins) [28] is within exon 2 of the gene whose transcript *CRX-B2* does not contain, so would therefore be unaffected in the LCA patient.

Mouse transcript *Crx-b1* is similar to human transcript *CRX-B1*, having several small ORFs encoded by exon 1 from weak Kozak initiation complexes. However, these ORFs are not conserved between the species. Mouse transcript *Crx-b2* consists of exon 1 spliced to exon 4. Similar to human transcript *CRX-B2*, it does not contain exon 2 and therefore does not have the initiation start codon, and may also be a noncoding transcript.

DISCUSSION

We have identified seven adult retinal transcripts of human *CRX* and two retinal transcripts of mouse *Crx*. The human and mouse genes share two highly conserved transcripts, *CRX-B1* and *Crx-b1*, respectively. These transcripts are predicted to be driven by the *CRX*P1 promoter, they both have a large noncoding 5′ exon 1 (which has a degree of sequence similarity), and they are both predicted to have the shorter of the two possible 3′-UTRs. Both transcripts also have the potential to encode the full-length, 299-residue protein. The human and mouse genes also share two lesser-conserved transcripts, *CRX-B2* and *Crx-b2*, respectively. As before, they share the same P1 promoter and noncoding 5′ exon 1. Initial data suggest that these transcripts are expressed to a lesser degree than the other transcripts driven by the P1 promoter. Additionally, both transcripts skip exon 2 and therefore remove the recognized initiation codon. In the absence of other strong Kozak initiation sequences, they are thought to be noncoding. Crucially, however, the mouse transcript also skips exon 3. The function of *CRX-B2* and *Crx-b2* transcripts has yet to be determined.

Human *CRX* has five additional transcripts, including *CRX-A1* to *CRX-A4*, each of which has an alternative splice variant (1Aα, 1Aβ, 1Aβγ, and 1Aδ) of the human-specific exon 1A at its 5′ end and are predicted to be driven by the *CRX*P2 promoter. Furthermore, these transcripts are predicted to have a much longer 3′-UTR than the transcripts under the control of the *CRX*P1 promoter. Human *CRX* also contains the additional transcript *CRX-A5*, which is related to the *CRX-A1* transcript, except it has the inclusion of exon 2A, an *Alu*-like repeat sequence that alters the CRX ORF and is predicted to generate a novel, truncated CRX-A5 protein.

Recently a report describing the 5′ region of mouse *Crx* has been published [29]. The authors place a mouse promoter approximately 1500 bp 3′ to the position we have for *Crx*P1. We found no evidence of a short *Crx* transcript from mouse retinal cDNA with our 5′-RACE or RT-PCR analysis (Fig. 3) and have shown that the region upstream of exon 1 can function as a promoter *in vitro*. However, this disagreement does not affect the authors' conclusions as their Pcrx2k-*lac*Z transgenic line spans the genomic region we have detailed here as the mouse *Crx*P1 promoter. Moreover, the combined data delineate the genomic region necessary for endogenous *Crx* expression to within an approximately 500-bp region. We have also noticed (unpublished data) that a rat *Crx* genomic clone (AC099348) shares a high degree of homology with *Crx*P1 and *CRX*P1, further identifying the P1 region as the major promoter.

It is apparent that human and mouse *CRX* have a high degree of divergence in transcriptional structure, yet the putative 299-residue protein products are 97% conserved [1–4]

and have a generally conserved function [5]. It is likely that the transcripts *CRX-B1* and *Crx-b1* (and possibility *CRX-B2* and *Crx-b2*) account for the similarity in activation of target genes and that the human-specific transcripts *CRX-A1* to *CRX-A5* have a human-specific function yet to be assigned.

Human exon 1A and its alternative splice variants have no significant homology with any other sequence in the databases, and although human exon 1 has some weak homology to *Alu* repeat elements, there is no other database match apart from the 3′ portion of mouse exon 1. The function of the 5′ noncoding exons in both human and mouse genes awaits elucidation and will be a key issue in understanding the regulation of the gene. Possible clues are available from other structurally complex or highly related genes, in particular mouse *Otx2*, the most closely related gene to *Crx*. In an investigation to see if *Otd* was functionally equivalent with *Otx2* [30], it was apparent that both the 5′- and 3′-UTR of *Otx2* were necessary for nucleo-cytoplasmic export and essential for the polyribosome complex, and therefore translation of the protein. Genes with upstream noncoding exons or genes with several or more promoters have been associated with complex regulatory functions [31] and may give clues to the function. The inward rectifier potassium channel gene *Kir7.1* has upstream noncoding exons and is also associated with alternative promoters that may modulate cell and developmental regulation [32]. In the case of *CRX/Crx*, alternative promoters may be related to the INL and ONL cell-specific expression [5]. The CC chemokine receptor-3 gene has three alternatively spliced upstream noncoding exons. The absence of the most upstream exon has been shown to reduce promoter activity [33]. Again, *CRX/Crx* expression in the INL and ONL may be modulated at developmental stages by a similar mechanism. Little is known of *CRX/Crx* expression in the pineal gland; this gene may be under strict developmental and temporal control here as well. Homeobox genes that undergo extensive alternative splicing are not uncommon in the mammalian genome either. For example, there are six isoforms of mouse *Pax8*, which display differential *trans*-activating potentials based on their C-terminal structure [34]. The paired-like transcription factor gene *RINX/Rinx* (*VSX1/Vsx1*) has been identified [35,36]. This gene has a similar spatial expression profile to *CRX/Crx* and has six alternative transcripts, some of which seem to be expressed to different degrees in different retinal layers [37]. Possibly, this implies that the six splice variants are used differently in distinct retinal cell layers and this may extend to a difference between embryonic and adult expression as well.

The human and mouse *CRX/Crx* genes generate long and short mRNA transcripts in adult retina through the selection of one of two alternative tandem poly(A) sites. From a review of genes containing alternative, tandem poly(A) of this type [38], it is clear that these genes regulate either the stability of the mRNA, the level of protein translation from the mRNA, or the tissue specificity of the mRNA. Multiple genes within this class have been shown or suggested to have a regulatory element between the poly(A) sites that influences the stability or translatability of the longer mRNA transcript. A conserved element has recently been identified among vertebrate *OTX2* genes that is important for translational control [39]. We could not find any homology with this region in *CRX/Crx*, but we did find a highly conserved G/U-rich element in the 3′-UTR between human, mouse, rat, and canine *CRX* that may have a similar or related function.

Six of the seven candidate OTX/CRX binding elements present in the promoter regions of *CRX/Crx* bound to the CRX-HD protein. Two of the six elements, CRX-1 and CRX-5, bound as efficiently as the BAT-1 control sequence. The CRX-5 sequence contains the reverse complement of the first binding site present in the BAT-1 probe, followed by a second imperfect site of TTAACC. CRX-1 also contains an almost exact match to the first BAT-1 CRX binding element. This confirms an earlier report that the BAT-1 element has high affinity for CRX [3]. Crx-7 and Crx-6, which we had predicted to be the mouse equivalents of CRX-1 and CRX-2, respectively, bound CRX but at a lower affinity compared with the human counterparts, suggesting that *Crx* regulates its own expression. However, the CRX homeodomain is identical in human, mouse, and bovine [3], and should therefore bind equally well to the CRX-HD protein we have used. An explanation for this discrepancy may be that other vertebrate *CRX* genes are under different regulation controls, as shown for mouse and chicken [5,10]. Furthermore, these data imply that CRX may auto-regulate or modulate its own expression, a phenomena that has also been identified in the *PAX-6* gene [40].

The data we have presented here will have significant relevance to the human *CRX* mutations previously documented at the *CORD2* locus. We were unable to identify a *CRX* mutation in the original *CORD2* family when searching the novel sequence we report here, although it remains a possibility that the *CORD2* mutation lies in an intron of *CRX* or in enhancer sequences either 5′ or 3′ to the gene, as has been shown to be the case for the retinal transcription factors *PAX6* [41–43] and *Pitx3* [44]. One of the most curious *CRX* mutations described is Pro 9 (1-bp ins) [28]. The proband had a range of clinical phenotypes that have not been reported for other *CRX* mutations, including depigmented skin lesions and amenorrhea [45]. This is the only documented *CRX* mutation that affects the *CRX-A5* transcript and may suggest an unrelated function for this splice variant. All reported *CRX* mutations are thought to inhibit binding of *CRX* to either DNA, in the case of homeodomain missense mutations such as Glu80Ala [1] and Arg41Trp [2], or other proteins, such as the *CRX* truncated proteins Glu168 (1-bp del) [1] and Ala196-Pro197 (4-bp del) [2]. With the insight into *CRX* transcriptional and translational structure presented here, it may be the case that a specific *CRX* mutation, for example the Glu80Ala mutation, has a unique affect on each of the *CRX* transcripts it is present in. In this instance, the mutation is thought to affect five (*CRX-A1* to *CRX-A4* and *CRX-B1*) of the seven possible *CRX* transcripts (*CRX-A5* terminates before residue 80 and *CRX-B2* does not seem to encode a protein). This hypothesis does not take into account the relative abundance of each *CRX* transcript, which may

themselves be regulated at key developmental times, further complicating the interpretation. This may account for the critical heterogeneity seen in LCA, adCRD, and RP patients that are associated with *CRX* mutations.

MATERIALS AND METHODS

Tissue and cell lines. Human Y-79 retinoblastoma cells (American Type Culture Collection HTB18) were cultured in RPMI 1640 medium (Sigma) supplemented with 15% fetal bovine serum (Invitrogen) and 2 mM glutamine at 37°C with 5% CO_2. COS-7 cells were cultured in Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% fetal bovine serum, 5% CO_2, and 100 U of penicillin and streptomycin (Invitrogen). Tissue was obtained from human adult retinas with informed consent, and whole eyes were obtained from adult C57BL/6 mice.

RNA extraction. Total cellular RNA was extracted from human Y-79 cells, human retina tissue, or C57BL/6 mouse eyes using the TRIzol reagent (Invitrogen) following the manufacturer's instructions. Total RNA (1–2 μg) was treated with RNase-free DNaseI (Invitrogen) and cDNA synthesized using the SuperScript Pre-amplification System for first strand cDNA synthesis using oligo (dT) primer (Invitrogen).

Isolation of CRX/Crx exons by RT-PCR. *CRX* oligonucleotide primers used for amplification of human *CRX* exon 1A (1Aα, 1Aβ, 1Aγ and 1Aδ) were 1AR6 (5′-GAT CTGATAGAAGAGATGACAG-3′) or 1AR7 (5′-CAGTGCTT-TATATTGTTTCAGTA-3′) with 2R (5′-CCTGGTGCATCAGATCCACAC-3′). For *CRX* exon 1, we used primers 1F2 (5′-GAGGGGATTAAGCAGACGG-3′), 1F3 (5′-CTGAGGAGTAGGGGCTTAT-3′), 1F4 (5′-GACCATCTGACAGTCTT-TATATTGTTTCAG-3′), and 1F5 (5′-CCAGGAGAGGGATCCCGAA-3′) with 2R. Forward primers used to amplify *CRX* internal splice variants were 1AR6 and 1F2 with reverse primer 4R (5′-GGGACTGTAGGAATCTGAGAT-3′). Mouse *Crx* oligonucleotide primers used for the amplification of exon 1 were 1f (5′-CTAGCCTCTGCTGTCTGAT-3′), 2f (5′-CAGAGAGGCAGGGTCT-GAAA-3′), 3f (5′-AGAGTCAGAGACAAGGTGAAGGG-3′), 4f (5′-CAGTG-GTGAGCAGGGATTAG-3′), and 2r (5′-CCTGGTGCATCAGGTCCACAT-3′). *Crx* internal splice variants were isolated with 1f and 4r (5′-GGGCTGTAA-GAATCTGAGAT-3′). RT-PCR was carried out using Y-79 cDNA, human adult retinal cDNA, or C57BL/6 eye cDNA under the following conditions: one cycle of 94°C for 3 minutes; 32–40 cycles of 55°C for 1 minute, 72°C for 1–3 minutes, and 94°C for 1 minute; followed by one cycle of 72°C for 10 minutes. RT-PCR fragments were cloned into the pGEM-T Easy vector (Promega) and sequenced using ABI Prism DNA sequencing and ABI 377 genetic analyzer (Perkin-Elmer).

Isolation of Crx exon 1 by 5′-RACE. Full-length *Crx* exon 1 was isolated using the 5′-RACE system for RACE version 2 (Invitrogen) and *Crx*-specific primers mus 1f and mus 2f. Approximately 3 μg of DNaseI treated total eye RNA from C57BL/6 adult mice was tailed with dCTP for 1 hour at 37°C using TdT. This was purified as recommended by the manufacturer and first round 5′-RACE-PCR performed using the abridged anchor primer (5′-GGCCACGCGTCGAC-TAGTACGGGIIGGGIIGGGIIG-3′) and Mus 1f (5′-CTAATCCCTGCTCAC-CACTG-3′) with *Taq* polymerase (Bioline). The cycling conditions were as follows: one cycle of 94°C for 3 minutes; 32 cycles of 60°C for 1 minute, 72°C for 2 minutes, and 94°C for 1 minute; followed by one cycle of 72°C for 10 minutes. Nested 5′-RACE-PCR used primers AUAP (5′-GGCCACGCGTCGAC-TAGTAC-3′) and mus 2f (5′-CCCTTTGTAAGGCAAGTT TGCAT-3′) using a 1/10 dilution of the primary PCR product as template with cycling conditions as before. Final PCR product was gel-purified (Qiagen) and cloned into the pGEM-T Easy vector (Promega) and sequenced using M13 forward and reverse primers as before.

Isolation of CRX 3′-UTR by 3′-RACE. Full-length *CRX* 3′-UTR was isolated using the 3′-RACE system (Invitrogen) and *CRX*-specific primers: 3′ RaceC (5′-GGTCAGAATCACCGTGCCCTTGAA-3′), 3′ RaceD (5′-GCA-CAATATATGCTTACGAGTTGGTA-3′), and 3′ RaceE (5′-CGTGGACACTTC TTTACATATGGTTA-3′). Approximately 3 μg of DNaseI treated total RNA from Y-79 human retinoblastoma cells was used for first strand synthesis using AP primer (5′-GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTT-3′). This was amplified with either 3′ RaceC, 3′ RaceD, or 3′ RaceE oligonucleotides and AUAP primer. Products were purified, cloned, and sequenced as before.

CRX mutation screening of the CORD2 family. Oligonucleotide primers to screen *CRX* for mutations within the CORD2 family members were designed from the genomic sequence AC008745 and spanned all previously isolated *CRX* exons [1]. Primer pairs were also designed to span exon 1A: 5′-CCATAG-CAACTTCAGAATCAGT-3′ and 5′-GGGTTTCAGCTGGTCCTGC-3′ (441 bp); 5′-CTGATAGAAGAGATGACAGGA-3′, and 5′-CTTCCAGCGGGACAGA-GAA-3′ (335 bp); 5′-CCCACCCCTGGAAGGACT-3′ and 5′-GGCTCCT-GATGGGAGAAT-3′ (197 bp); 5′-CCCGCTGGAAGTTATTTTCTC-3′ and 5′-CCCCTGCTAGGAGAGATTAA-3′ (232 bp); exon 1, 5′-GGGATCCC-GAAAGCTCCGT-3′ and 5′-GGCCGGGAGTGATGTCATTT-3′ (365 bp), 5′-GGGATTAAGCAGACGGGT-3′, and 5′-CCCCTCCGCCTCTCCTG-3′ (267 bp); exon 2A, 5′-GGCCTGCAAAGCCTTATTT-3′ and 5′-CCTGGTAACAGAA-CAAGGCT-3′ (314 bp); and putative poly(A) signals in the 3′-UTR: 5′-GGA-CACTTCTTTAGCATA-3′ and 5′-CCCAATGTGCTAATGAGCGA-3′ (284 bp) spanning A_1A_2, 5′-CGTTCTACGTCCTAAAGCCA-3′, and 5′-CCCAGCCC-CTGATCCTTCA-3′ (355 bp) spanning A_4.

Sequence analysis. Genomic and mRNA sequences were analyzed using the NIX program available from the United Kingdom MRC Human Genome Project Resource Centre [22]. This included BLAST searches against Expressed Sequence-Tag, Swissprot, EMBL, TREMBL, mRNA, and High Throughput Genome Sequences databases. BLAST analysis of two sequences was performed with the Pairwise BLAST program available at the National Center for Information at the National Institute of Health (Bethesda, MD). Dot-matrix alignment of DNA was with Dotmatcher (http://bioweb.pasteur.fr/seqanal/interfaces/dotmatcher.html).

CRX homeodomain construct. A 378-bp fragment of *CRX* cDNA encompassing the homeodomain region (residues 10–134) was amplified with *Bam*HI and *Eco*RI oligonucleotide linkers with *Pfu* Turbo polymerase (Stratagene). The fragment was digested with *Bam*HI and *Eco*RI, purified (Qiagen), and cloned into a *Bam*HI and *Eco*RI digested and purified pGex4T-1 vector (Amersham Pharmacia Biotech) in frame with GST. The resulting pGST-CRX-HD construct was sequenced to insure integrity.

Purification of GST fusion protein. GST-CRX-HD was generated in *Escherichia coli* strain BL21 and purified by passage through glutathione-sepharose beads (Amersham Pharmacia Biotech) as recommended by the manufacturer. GST was cleaved with 200 U of thrombin (Sigma) overnight in PBS and CRX-HD Protein eluted by centrifugation. Thrombin was removed with benzamidine sepharose 6b (Amersham Pharmacia Biotech). CRX-HD protein integrity was confirmed on a SDS-PAGE gel and protein concentration estimated using a Bradford assay.

Band shift assay. Aliquots of CRX-HD protein were pre-incubated with 200 pM of unlabeled competitor oligonucleotide DNA (Table 2) for 30 minutes at 4°C. To this mixture 1 pM ^{32}P end labeled BAT-1 ds oligonucleotide was added and incubated for a further 30 minutes at 4°C in a final concentration of 20 mM Hepes, pH 7.4, 5% glycerol, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, and 0.2 μg polyDI.dc. Binding reactions were electrophoresed on a 4.5% polyacrylamide gel in 0.25× TBE, vacuum-dried onto 3MM paper and autoradiographed or exposed on a phosphoimager system (Cyclone) and bands quantified with Optiquant software. The band shift assay was repeated three times with similar results.

Luciferase reporter constructs. Oligonucleotide primers were designed to amplify the two human promoters and single mouse promoter regions. An *Mlu*I site was introduced into the forward primer and a *Bgl*II site on the reverse primer using human PAC clone RPCI1-236C18 (M.D.H. *et al*., manuscript in preparation) and mouse PAC clone RPCI23-464O10 (identified from the RPCI23 library available from the HGMP, http://www.hgmp.mrc.ac.uk/Biology/descriptions/mouse_pac.html with a human *CRX* probe following standard techniques.) as templates: *CRX*P1F (5′-CGCGCACGCGTGCCTGTAGCCT-TAATCTCTCCTA-3′) and *CRX*P1R (5′-CCCGCAGATCTCGCCCGGTAATC-CCTCCTT-3′), 113 bp; *CRX*P2F (5′-CGGCGACGCGTGCATAGGACAT-GAACTCATCCT-3′) and *CRX*P2R (5′-CGGCCAGATCTCTGGAAGAAAT-TAAAGGAGAC-3′), 388 bp; *Crx*P1F (5′-CGGGCACGCGTGCCCAGAGGCC-CTAATCTCTCCTA-3′) and *Crx*P1R (5′-CGGCCAGATCTGCCACACGCA-CACTCTAATC-3′), 114 bp. Purified PCR products were cloned into the pGEM-T Easy vector (Promega), and inserts were excised with *Mlu*I/*Bgl*II, purified, and cloned into the *Mlu*I/*Bgl*II site of pGL3-Basic (Promega). Clones were

sequenced to confirm integrity and plasmid DNA for transfection was purified with midiprep columns (Promega).

Luciferase assays. Liposome-mediated transfections were performed in six-well tissue culture plates using 12 μg of LipofectAMINE (Invitrogen) and 1μg of each construct, essentially as described [46]. Thirty-hours post-transfection, the medium was removed and the cells were rinsed twice in PBS buffer. Luciferase Assay Systems (Promega) was used to determine promoter activity in cell lysates using a luminometer (Nichols Institute Diagnostics). Each construct was transfected six times and mean values taken in at least two independent transfections. Relative light values were normalized per microgram of total protein in each sample.

ACKNOWLEDGMENTS

This work was supported by the British Retinitis Pigmentosa Society, Grant GR518, and Fundação para a Ciência e Tecnologia, Portugal, Grant BD/837/2000.

RECEIVED FOR PUBLICATION JANUARY 22;
ACCEPTED AUGUST 30, 2002.

REFERENCES

1. Freund, C. L., *et al.* (1997). Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell* **91:** 543–553.
2. Swain, P. K., *et al.* (1997). Mutations in the cone-rod homeobox gene are associated with the cone-rod dystrophy photoreceptor degeneration. *Neuron* **19:** 1329–1336.
3. Chen, S., *et al.* (1997). Crx, a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes. *Neuron* **19:** 1017–1030.
4. Furukawa, T., Morrow, E. M., and Cepko, C. (1997). Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell* **91:** 531–541.
5. Bibb, L. C., *et al.* (2001). Temporal and spatial expression patterns of the CRX transcription factor and its downstream targets. Critical differences during human and mouse eye development. *Hum. Mol. Genet.* **10:** 1571–1579.
6. Li, X., *et al.* (1998). A pineal regulatory element (PIRE) mediates transactivation by the pineal/retina-specific transcription factor CRX. *Proc. Natl. Acad. Sci. USA* **95:** 1876–1881.
7. Rivolta, C., Berson, E. L., and Dryja, T. P. (2001). Dominant Leber congenital amaurosis, cone-rod degeneration, and retinitis pigmentosa caused by mutant versions of the transcription factor CRX. *Hum. Mutat.* **18:** 488–498.
8. Freund, C. L., *et al.* (1998). De novo mutations in the CRX homeobox gene associated with Leber congenital amaurosis. *Nat. Genet.* **18:** 311–312.
9. Sakamoto, K., *et al.* (1999). Molecular cloning of the cone-rod homeobox gene (Crx) from the rat and its temporal expression pattern in the retina under a daily light-dark cycle. *Neurosci. Lett.* **261:** 101–104.
10. Bernard, M., Dinet, V., and Voisin, P. (2001). Transcriptional regulation of the chicken hydroxyindole-O-methyltransferase gene by the cone-rod homeobox-containing protein. *J. Neurochem.* **79:** 248–257.
11. Liu, Y., *et al.* (2001). Isolation and characterization of a zebrafish homologue of the cone rod homeobox gene. *Invest. Ophthalmol. Vis. Sci.* **42:** 481–487.
12. Furukawa, T., Morrow, E. M., Li, T., Davis, F. C., and Cepko, C. L. (1999). Retinopathy and attenuated circadian entrainment in Crx-deficient mice. *Nat. Genet.* **23:** 466–470.
13. Livesey, F. J., Furukawa, T., Steffen, M. A., Church, G. M., and Cepko, C. L. (2000). Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx. *Curr. Biol.* **10:** 301–310.
14. Briata, P., Ilengo, C., Bobola, N., and Corte, G. (1999). Binding properties of the human homeodomain protein OTX2 to a DNA target sequence. *FEBS Lett.* **445:** 160–164.
15. Fong, S. L., and Fong, W. B. (1999). Elements regulating the transcription of human interstitial retinoid-binding protein (IRBP) gene in cultured retinoblastoma cells. *Curr. Eye Res.* **18:** 283–291.
16. Kimura, A., *et al.* (2000). Both PCE-1/RX and OTX/CRX interactions are necessary for photoreceptor-specific gene expression. *J. Biol. Chem.* **275:** 1152–1160.
17. Mitton, K. P., *et al.* (2000). The leucine zipper of NRL interacts with the Crx homeodomain. a possible mechanism of transcriptional synergy in rhodopsin. *J. Biol. Chem.* **275:** 29794–29799.
18. Yanagi, Y., Masuhiro, Y., Mori, M., Yanagisawa, J., and Kato, S. (2000). p300/CBP acts as a coactivator of the cone-rod homeobox transcription factor. *Biochem. Biophys. Res. Commun.* **269:** 410–414.
19. Zhu, X., and Craft, C. M. (2000). Modulation of Crx transactivation activity by phosducin isoforms. *Mol. Cell. Biol.* **20:** 5216–5226.
20. Wang, Q., *et al.* (2001). Mutations of QRX, a novel retinal homeobox gene, are associated with retinal degeneration. *Am. J. Hum. Genet.* **69:** 366.
21. Mears, A. J., *et al.* (2001). Nrl is required for rod photoreceptor development. *Nat. Genet.* **29:** 447–452.
22. Williams, G.W., Woollard, P. M., and Hingamp, P. (1998) NIX: a nucleotide identification system at the HGMP-RC (http://www.hgmp.mrc.ac.uk/NIX/).
23. Beyer, K., Dandekar, T., and Keller, W. (1997). RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3′-end processing of pre-mRNA. *J. Biol. Chem.* **272:** 26769–26779.
24. Jurka, J., and Milosavljevic, A. (1991). Reconstruction and analysis of human Alu genes. *J. Mol. Evol.* **32:** 105–121.
25. Evans, K., *et al.* (1994). Genetic linkage of cone-rod retinal dystrophy to chromosome 19q and evidence for segregation distortion. *Nat. Genet.* **6:** 210–213.
26. Kozak, M. (1996). Interpreting cDNA sequences: some insights from studies on translation. *Mamm. Genome* **7:** 563–574.
27. Chau, K.-Y., Chen, S., Zack, D. A., and Ono, S. J. (2000). Functional domains of the cone-rod homeobox (CRX) transcription factor. *J. Biol. Chem.* **275:** 37264–37270.
28. Silva, E., *et al.* (2000). A CRX null mutation is associated with both Leber congenital amaurosis and a normal ocular phenotype. *Invest. Ophthalmol. Vis. Sci.* **41:** 2076–2079.
29. Furukawa, A., Koike, C., Lippincott, P., Cepko, C. L., and Furukawa, T. (2002). The mouse *Crx* 5′-upstream transgene sequence directs cell-specific and developmentally regulated expression in retinal photoreceptor cells. *J. Neurosci.* **22:** 1640–1647.
30. Acampora, D., *et al.* (2001). OTD/OTX2 functional equivalence depends on 5′ and 3′ UTR-mediated control of Otx2 mRNA for nucleo-cytoplasmic export and epiblast-restricted translation. *Development* **128:** 4801–4813.
31. Ayoubi, T. A., and Van De Ven, W. J. (1996). Regulation of gene expression by alternative promoters. *FASEB J.* **10:** 453–460.
32. Nakamura, N., Suzuki, Y., Ikeda, Y., Notoya, M., and Hirose, S. (2000). Complex structure and regulation of expression of the rat gene for inward rectifier potassium channel Kir7.1. *J. Biol. Chem.* **275:** 28276–28284.
33. Zimmermann, N., *et al.* (2000). Analysis of the CC chemokine receptor 3 gene reveals a complex 5′exon organization, a functional role for untranslated exon 1, and a broadly active promoter with eosinophil-selective elements. *Blood* **96:** 2346–2354.
34. Kozmik, Z., Kurzbauer, R., Dorfler, P., and Busslinger, M. (1993). Alternative splicing of Pax-8 gene transcripts is developmentally regulated and generates isoforms with different transactivation properties. *Mol. Cell. Biol.* **13:** 6024–6035.
35. Semina, E. V., Mintz-Hittner, H. A., and Murray, J. C. (2000). Isolation and characterization of a novel human paired-like homeodomain-containing transcription factor gene, VSX1, expressed in ocular tissue. *Genomics* **63:** 289–293.
36. Hayashi, T., Huang, J., and Deeb, S. S. (2000). RINX(VSX1), a novel homeobox gene expressed in the inner nuclear layer of the adult retina. *Genomics* **67:** 128–139.
37. Hayashi, T., Huang, J., and Deeb, S. S. (2001). Expression of the homeobox protein Rinx/Vsx1 during postnatal eye development. *Invest. Ophthalmol. Vis. Sci.* **42:** S657.
38. Edwards-Gilbert, G., Veraldi, K. L., and Milcarek, C. (1997). Alternate poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res.* **25:** 2547–2561.
39. Pilo Boyl, P., *et al.* (2001). Forebrain and midbrain development requires epiblast-restricted OTX2 translational control mediated by its 3′ UTR. *Development* **128:** 2989–3000.
40. Okladnova, O., Syagailo, Y. V., Mossner, R., Riederer, P., and Lesch, K. P. (1998). Regulation of *PAX-6* gene transcription: alternate promoter usage in human brain. *Brain Res. Mol. Brain Res.* **60:** 177–192.
41. Williams, S. C., Altmann, C. R., Chow, R. L., Hemmati-Brivanlou, A., and Lang, R. A. (1998). A highly conserved lens transcriptional control element from the *Pax-6* gene. *Mech. Dev.* **73:** 225–229.
42. Ton, C. C., *et al.* (1991). Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell* **67:** 1059–1074.
43. Lauderdale, J. D., Wilensky, J. S., Oliver, E. R., Walton, D. S., and Glaser, T. (2000). 3′ deletions cause aniridia by preventing PAX6 gene expression. *Proc. Natl. Acad. Sci. USA* **97:** 13755–13759.
44. Rieger, D. K., Reichenberger, E., McLean, W., Sidow, A., and Olsen, B. (2001). A double-deletion mutation in the Pitx3 gene causes arrested lens development in Aphakia mice. *Genomics* **72:** 61–72.
45. Dharmaraj, S. R., *et al.* (2000). Mutational analysis and clinical correlation in Leber congenital amaurosis. *Ophthalmic Genet.* **21:** 135–150.
46. Hellqvist, M., Mahlapuu, M., Blixt, A., Enerbäck, S., and Carlsson, P. (1998). The human forkhead protein FREAC-2 contains two functionally redundant activation domains and interacts with TBP and TFIIB. *J. Biol. Chem.* **273:** 23335–23343.

Sequence data from this article have been deposited with the DDBJ/EMBL/GenBank Data Libraries under accession numbers AF335246 (genomic *CRX* exon 1A and proximal promoter), AF335247S1 (partial *CRX* exon 1 cDNA), AF335248 (genomic *CRX* exon 1 and proximal promoter), AF335247S2 (genomic *CRX* 3′-UTR), AF335593 (*CRX* exon 1Aβ cDNA), AF335594 (*CRX* exon 1Aγ cDNA), AF335595 (*CRX* exon 1Aδ cDNA), and AF442496 (*CRX* exon 2A cDNA).