ELSEVIER

Gene 215 (1998) 111–122

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Organization of the *ABCR* gene: analysis of promoter and splice junction sequences

Rando Allikmets [a], Wyeth W. Wasserman [b], Amy Hutchinson [c], Philip Smallwood [d,e], Jeremy Nathans [d,e,f], Peter K. Rogan [g], Thomas D. Schneider [h], Michael Dean [c,*]

[a] *Intramural Research Support Program, SAIC-Frederick, Frederick, MD 21702, USA*
[b] *Bioinformatics, SmithKline Beecham Pharmaceuticals, King of Prussia, PA 19406, USA*
[c] *Laboratory of Genomic Diversity, NCI-FCRDC, Frederick, MD 21702-1201, USA*
[d] *Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA*
[e] *Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA*
[f] *Departments of Neuroscience and Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA*
[g] *Department of Human Genetics, Allegheny University of the Health Sciences, Pittsburgh, PA 15212, USA*
[h] *Laboratory of Experimental and Computational Biology, NCI-FCRDC, Frederick, MD 21702, USA*

Received 16 January 1998; received in revised form 1 May 1998; accepted 1 May 1998; Received by T. Sekiya

Abstract

Mutations in the human *ABCR* gene have been associated with the autosomal recessive Stargardt disease (STGD), retinitis pigmentosa (RP19), and cone-rod dystrophy (CRD) and have also been found in a fraction of age-related macular degeneration (AMD) patients. The *ABCR* gene is a member of the ATP-binding cassette (ABC) transporter superfamily and encodes a rod photoreceptor-specific membrane protein. The cytogenetic location of the *ABCR* gene was refined to 1p22.3–1p22.2. The intron/exon structure was determined for the *ABCR* gene from overlapping genomic clones. *ABCR* spans over 100 kb and comprises 50 exons. Intron/exon splice site sequences are presented for all exons and analyzed for information content (R_i). Nine splice site sequence variants found in STGD and AMD patients are evaluated as potential mutations. The localization of splice sites reveals a high degree of conservation between other members of the ABC1 subfamily, e.g. the mouse *Abc1* gene. Analysis of the 870-bp 5′ upstream of the transcription start sequence reveals multiple putative photoreceptor-specific regulatory elements including a novel retina-specific transcription factor binding site. These results will be useful in further mutational screening of the *ABCR* gene in various retinopathies and for determining the substrate and/or function of this photoreceptor-specific ABC transporter.

Keywords: ABC genes; Splice sites; Information theory

1. Introduction

The ATP-binding cassette (ABC) transporter superfamily is one of the largest gene families and encodes a functionally diverse group of membrane proteins involved in energy-dependent transport of a wide variety of substrates across membranes (Dean and Allikmets, 1995). The ATP-binding domains of ABC transporters contain characteristic motifs conserved throughout evolution and distinguish ABC proteins from other ATP-binding proteins. In humans, ABC genes typically encode four domains consisting of two ATP-binding segments and two transmembrane (TM) domains (Dean and Allikmets, 1995).

More than 35 human ABC genes have been characterized to date to various extents (Allikmets et al., 1996). Several ABC proteins have been implicated in different inherited human diseases, including cystic fibrosis transmembrane conductance regulator (CFTR), and those involved in peroxisome biogenesis (ALD, PMP70). ABCR, a rod photoreceptor-specific ABC

* Corresponding author. Tel: +1 301 846 5931; Fax: +1 301 846 1909; e-mail: dean@fcrfv1.ncifcrf.gov

Abbreviations: A, adenosine; aa, amino acid(s); ABCR, retina-specific ABC transporter; bp, base pair(s); C, cytidine; cDNA, DNA complementary to RNA; EST, expressed sequence tag; G, guanosine; kb, kilobase(s) or 1000 bp; ORF, open reading frame; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; T, thymidine; UTR, untranslated region.

0378-1119/98/$19.00
PII: S0378-1119(98)00269-8

protein, is involved in at least four eye disorders—in autosomal recessive Stargardt disease (STGD), in autosomal recessive retinitis pigmentosa (RP19), in cone-rod dystrophy (CRD), and in age-related macular degeneration (AMD) (Allikmets et al., 1997a,b; Martinez-Mir et al., 1998; Cremers et al., 1998). STGD and AMD share a similar phenotype including progressive loss of central vision and geographic atrophy of the central macular region (Allikmets et al., 1997a,b). AMD is the most common cause of vision loss in the elderly and is estimated to affect in the US about 35% of population over the age of 75 (Allikmets et al., 1997b). Some *ABCR* sequence variants have been shown to be associated with a fraction of the 'dry' form of AMD (Allikmets et al., 1997b), the extent of the *ABCR* gene involvement in AMD is currently under investigation. Clinical manifestations of RP19 are radically different, starting from night blindness and developing into peripheral scattered pigmentation and choriocapillaris atrophy (Martinez-Mir et al., 1998).

The *ABCR* gene is expressed at high levels in only one cell type, rod photoreceptors (Allikmets et al., 1997a), and encodes a protein that has been previously identified as the rod outer segment rim protein (RmP, Papermaster et al., 1976). ABCR is a full-length, four-domain protein, and belongs to the ABC1 subfamily within the ABC superfamily (Allikmets et al., 1996). Other members of this subfamily include the *ABC1*, *ABC2* and *ABC3* (ABC-C) genes, characterized in both humans and mice (Luciani et al., 1994; Allikmets et al., 1995; Klugbauer and Hofmann, 1996). It seems that members of this subfamily have evolved to perform specialized functions in multicellular organisms since no yeast ortholog has been described (Michaelis and Berkower, 1995). To date, however, no endogenous function or substrate has been described for these proteins with the exception of the mouse Abc1 protein. Mouse Abc1 is highly expressed in macrophages and is proposed to be involved in the scavenging of apoptotic cells (Luciani and Chimini, 1996).

Involvement of the *ABCR* gene in at least four distinct eye disorders prompted us to fully characterize the genomic structure and sequences of intron/exon boundaries of this gene to enable exhaustive screening for mutations in the eye diseases mentioned above, and in others. At the same time, the cell specificity of *ABCR* expression raised intriguing possibilities concerning the promoter elements that may regulate this expression pattern as compared with other photoreceptor-specific genes, such as rhodopsin (Morabito et al., 1991).

2. Materials and methods

2.1. Sequence analysis

The sequence analysis was performed as described earlier (Allikmets et al., 1996). Briefly, searches of the dbEST database were performed with BLAST on the NCBI file server. Amino acid alignments were generated with PILEUP. Sequences were analyzed with programs of the Genetics Computer Group package on a VAX computer. Phylogenetic trees were generated from the amino acid alignments using PHYLIP (Phylogeny Inference Package) Version 3.5c. Two programs were utilized: NEIGHBOR, implementing the Neighbor-Joining distance matrix method, and PROTPARS, Protein Sequence Parsimony Method. Bootstrap resampling of 100 iterations was performed to test the reliability of the associations in both methods. In these analyses, resampling of the original data set was performed to create 100 new data sets. A consensus of the resulting 100 trees measures the consistency of the phylogenetic signal within the original data. Bootstrap proportions greater than 70% were considered as strong support for the adjacent node.

2.2. Computational identification of regulatory sequences

The *ABCR* promoter was screened for transcription factor binding sites represented in the Transfac database (Quandt et al., 1995) with the tools available on the database internet page (http://transfac.gbf-braunschweig.de/welcome.html). Screening for sites resembling known photoreceptor regulatory elements was performed with the FindPatterns program of the GCG package.

2.3. Genomic and cDNA cloning

cDNA clones containing *ABCR* sequences were obtained from a human retina cDNA library and sequenced fully. Primers were designed from the sequences of cDNA clones from 5′ and 3′ regions of the gene and used to link the identified cDNA clones by RT-PCR with retina QUICK-Clone cDNA (Clontech) as a template. PCR products were cloned into pGEM-T vector (Promega). cDNA clones from various regions of the *ABCR* gene were used as probes to screen a human genomic library in Lambda FIX II (#946203, Stratagene).

2.4. Exon/intron structure of the human ABCR gene

Primers for the cDNA sequences of the *ABCR* were designed with the PRIMER program (Lincoln et al., 1991). Both *ABCR* cDNA clones and genomic clones were used as templates for sequencing. Sequencing was performed with the Taq dye-deoxy terminator cycle sequencing kit (Applied Biosystems), according to the manufacturer's instructions. Sequencing reactions were resolved on an ABI 373A automated sequencer. The positions of the introns were determined by comparison between genomic and cDNA sequences. Primers for amplification of individual exons were designed from

adjacent intron sequences 20–50 bp from the splice site. Amplification of exons was performed with AmpliTaq Gold polymerase in a 25-ml volume in 1× PCR buffer supplied by the manufacturer (Perkin Elmer). Samples were heated to 95°C for 10 min and amplified for 35–40 cycles of 96°C, 20 s; 60°C, 30 s; 72°C, 30 s. Amplification of the introns was carried out in the same way, except that the extension time at 72°C was usually 4 min. PCR products were analyzed on 1–1.5% agarose gels and in some cases digested with an appropriate restriction enzymes to verify their sequence. Primer sequences and specific reaction conditions have been deposited with GDB.

2.5. Splice site sequence analysis

Sequence walkers are a graphical representation of the individual information content at specific binding sites (Schneider, 1997a,b). Characters representing the sequence are either oriented normally and placed above a line indicating favorable contact, or upside-down and placed below the line indicating unfavorable contact. Functional sites therefore have most letters pointing upwards, whereas nonfunctional sites have many letters pointing downwards.

The weight matrix used to model the splice junctions is computed from:

$$R_{\mathrm{iW}}(b,l) = 2 - \{-\log_2 f(b,l) + e[n(l)]\} \text{ (bits per base)},$$

where $f(b,l)$ is the frequency of each base b at position l in the aligned binding site sequences, and $e[n(l)]$ is a sample size correction factor for the n sequences at position l used to create $f(b, l)$. The matrix, $R_{\mathrm{iW}}(b, l)$, can be used to rank-order the sites and search for new sites, compare the sites with each other and with other quantitative data, and detect any errors in splice junction sequences.

To evaluate a DNA sequence, the bases of the sequence are aligned with the matrix entries, and the values corresponding to each base are added together to produce the total value. This measure has several advantages over other methods. First, the scale is in bits, which allows direct comparison with many other systems. Second, by adding the weights together for various positions in a particular binding site, we obtain the total 'individual information' (R_i) for that site. Third, the average for all of the binding sites used to create the matrix is the average information content. This is the same as the area under the sequence logo (Schneider and Stephens, 1990; Schneider, 1997b). Fourth, unlike a neural network that needs to be cyclically trained and requires both sites and non-sites, the matrix can be created immediately using only proven sites as examples, thereby avoiding the danger of training against unknown functional sites. Fifth, functional binding sites have positive values, within the error of the method, allowing predictions to be made. Finally, unlike consensus sequences that destroy the available sequence data by arbitrarily rounding the frequencies up or down, the individual information method uses the base frequencies directly and so it preserves subtleties in the data.

The effect of nucleotide substitutions was evaluated by comparing the individual information of the wild-type and variant alleles. To assess the effects of different substitutions on a specific splice site, R_i was computed for the common and variant sites with the program Scan and displayed with MakeWalker and Lister. See http://www-lecb.ncifcrf.gov/~toms/walker for additional information.

3. Results

3.1. ABCR genomic structure and intron/exon boundaries

The *ABCR* cDNA sequence was assembled combining dbEST database screening, RT-PCR, cDNA library screening and 5′ Marathon RACE methods (Allikmets et al., 1997a). A total of 8184 bp of the *ABCR* sequence was assembled, including an open reading frame of 6819 bp (2273 amino acids). A bacteriophage l human genomic library was screened with cDNA probes. A contig of 10 overlapping phage clones containing the entire *ABCR* coding region and spanning more than 100 kb was obtained (data not shown). The exon/intron structure of *ABCR* was determined by direct sequencing of phage and cDNA clones. A total of 50 exons were identified (Table 1) and primers designed for each individual exon to screen for mutations in STGD and AMD patients (Allikmets et al., 1997a,b). Exon sizes range from 33 to 266 bp (Table 1) and are relatively small on average. Primers used to amplify individual exons have been submitted to GDB (GDB: 9300763–9300827). Intron sizes were estimated from the sizes of PCR products generated using primers from adjacent exons with phage clones and/or genomic DNA as templates (Table 1). Introns larger than 10 kb were not amplified due to the limitations of the method. Our phage contig covered all but one intron; therefore, despite a large transcript and 50 exons, the *ABCR* gene spans a relatively short genomic distance on chromosome 1p22. Earlier reports have placed the STGD1 locus on 1p13–p21 (Kaplan et al., 1993). The latest version of the chromosome 1 Cytogenetic Map (HUGO_CC1) localizes two markers that flank the ABCR gene, D1S3361 and D1S236, to 1p22.3–1p22.2, therefore repositioning the gene to 1p22.

During PCR amplification and cloning of the *ABCR* gene using human retina cDNA (Clontech) as a template, we observed the in-frame absence of 38 amino acids in exon 30 in many subclones. The poly(A)$^+$

Table 1
ABCR splice junction sequences and their information content

Exon	Size (bp)	Splice acceptor	R_i (bits)	Splice donor	R_i (bits)	Intron (kb)
1	5′			AAGGCAAAAG**gt**aacagttactgtct	7.4	2
2	94	tgttttgtttttcc**ag**ATTCGCTTTG	13.2	CATCATGAAT**gt**aagcatagcagggt	7.8	1.5
3	142	attttctgttttaa**ag**GCCATTTCCC	10.4	ACAACTCCAT**gt**aagtgttgagatcc	9.8	2
4	139	ggctttgtccttac**ag**CTTGGCAAGG	8.0	AGAATTGCAG**gt**aagcatgactgcag	11.3	>10
5	128	tcttattcatatgt**ag**GAAGAGGAAT	10.0	TCCAGAGCAG**gt**agggggatgtcact	9.3	4.5
6	198	cctcttctccctgc**ag**TTCGCTCATG	13.2	CTTCCGTGTG**gt**aagggaggggtttg	8.6	>10
7	88	ttagtgtcaattac**ag**GTTCCCACAC	6.4	AATTCAAGAG**gt**gagtatcctgatgg	11.2	3
8	238	ctttgccttgccct**ag**TTTATCCATC	3.8	AGAAGAACAA**gt**aagttttctgagtc	9.1	1
9	139	ccttgtctcctggc**ag**CATCCTTTTG	9.1	ACTGAAGAAT**gt**aagatcccacctgg	7.1	1
10	117	ctttgtctggttttt**ag**GCCAACTCAA	7.2	CATGATCAGA**gt**aaggggggggttgga	6.4	0.8
11	198	atttcttcttccca**ag**GATACCCTGG	8.3	ATACCTGGAG**gt**aaggggctgcaacc	10.8	>10
12	206	ttgtgactctctgc**ag**TGCTTGGTCC	8.7	TTAAAGACAG**gt**gatgtttcaggaag	6.7	0.4
13	177	ctctgtgtcttctc**ag**GTATTGGGAT	14.0	TGGACGATTC**gt**gagtctgaagttcg	4.4	2
14	223	tctccttttgctttt**ag**TTTCATGATC	11.7	ATTCATCATG**gt**aagccagatggaga	8.8	3.5
15	222	ggcctcttggtttc**ag**CATGGAAGAA	4.9	GAAGGCTGTG**gt**gaggcccttgggct	7.6	1
16	205	ctccgcctcactgc**ag**AGCTTACTGT	9.9	GTGTTTCCAG**gt**aagcatcctcctct	11.3	3.5
17	65	ttgtctctatttttt**ag**GAGACTATGG	11.6	GGCGGTGAAG**gt**gagtcctttaaaac	12.0	2.5
18	90	cttctgattggtgc**ag**GGTGTTCAAC	4.0	GGAATACACG**gt**aaaacccgataaag	5.2	1.5
19	175	ttttttttcctccc**ag**ACTCCTTCTT	14.5	CCACCACCTT**gt**gagtctgccagcag	6.7	2
20	132	cgcctgtgttctgc**ag**GTCCATCCTG	8.0	TGTTCCACCA**gt**aagcgacacaggaa	6.6	1
21	140	tgtcccatcctcct**ag**CCTCACGGTG	8.5	GACCTATCAG**gt**gctcanagctggat	3.9	0.8
22	138	cttcttgtgcctcc**ag**GTGGCATGCA	14.7	TATCGCTCAG**gt**aacagctgctgctc	7.6	1
23	194	ggcccctaccttgc**ag**GCAGAACCAT	9.6	AGGCAGTGAG**gt**aggtgtctgcccac	9.4	0.5
24	85	catccatctgttgc**ag**GGGACCTGCA	9.4	GTCCTGGATG**gt**aaggactggacggg	9.2	2.5
25	206	ttttttattgtcat**ag**GGGATGTAAA	7.9	CCTGGAAGAG**gt**aaagtagagattcc	6.3	0.5
26	49	ttcttcttcttgac**ag**ATTTTTCTGA	15.4	CTGTTTGCGG**gt**atggtgctggagcc	5.6	>10
27	266	caatctctcaaaac**ag**GTGGCGCTCA	6.6	CCTGGCGCAG**gt**actattgtcggtcg	3.6	1
28	125	gtctattctcccac**ag**ATCGTGCTCC	12.2	CCTTCTTCAG**gt**gcgcggactcgggg	8.2	1
29	99	gtctcactctgtct**ag**CATGGATGAA	4.9	GGTGGCTTCC**gt**aagtgcctacgcgc	5.5	0.8
30A	187	tcttgtcttccacg**ag**GGAGTACCCC	5.0	GCCCCCCCAG**gt**acctgacctccaaa	6.1	0
30B	73	cttcaccatcctgc**ag**GTGCAGCACC	9.4	GCCCCCCCAG**gt**acctgacctccaaa	6.1	3.5
31	95	tgtctcattgcctc**ag**AGAACACAGC	5.2	TAAGAAGCAG**gt**aagaagaaatcctt	11.1	1.5
32	33	ttattttggctttc**ag**CTTAAAGAGC	10.6	ATGAACAGAG**gt**aagaaactattttt	10.0	1.5
33	106	cctgtttccttgtc**ag**GTATGGAGGA	10.6	TGTGAGCGGG**gt**atgtaaacagactg	6.2	0.13
34	75	tttacatggtttttt**ag**GGCCCTATCA	5.9	CAACATTAAG**gt**acttgacctatgta	6.0	0.23
35	170	cctctccaccctct**ag**GTGTGGTTTA	11.3	AGATTACAGT**gt**aagccaccacagcc	5.3	1.5
36	178	tctggccctgctgc**ag**GCTGACCACT	9.6	CTGGGACATC**gt**aagtgtcagtttac	6.8	3.5
37	116	ctcttctctttccg**ag**ATGAATTATT	9.7	TGCTGTATGG**gt**aagccgtttgggcc	7.2	1.2
38	158	gtctcatttcacac**ag**ATGGGCGGTC	8.5	TAATAACCGG**gt**gagcataactttct	8.6	3.0
39	125	tacttctctgtttc**ag**ACGCTGCTCA	10.3	GCCCGGTTTG**gt**gggtggtacccgag	6.4	0.8
40	130	tcctgttgatgccc**ag**GTGAGGAGCA	5.9	TCTCCCAATG**gt**acgtccatgccaca	7.2	1.5
41	121	ccttttcttcatct**ag**GATTGCCGAG	12.8	ACTAACCAAG**gt**aagggaatgggtat	11.8	0.4
42	63	tcttcatatcttgc**ag**ATTTATCTGG	10.6	CCCTGGAGAG**gt**gggtactctgcaga	8.6	0.495
43	107	ctgtgttttctcct**ag**TGCTTTGGCC	7.7	CAGGCAAGAG**gt**gagtatccngctcc	11.2	2
44	142	actttttttcttgc**ag**TATTTTAACC	12.9	AATCGAAAAG**gt**gaaaaatgttttgt	7.1	3.5
45	135	tctttctcttcccct**ag**GTTGCAAACT	12.5	GGTGCTGCTG**gt**aactgcgggcttgg	7.1	1
46	104	tctntccacccccc**ag**GATGAGCCCA	10.4	CATCCCACAG**gc**aagagattcccagg	4.1	0.07
47	93	ctctcctgccccac**ag**CATGGAAGAA	12.9	TCAAGTCCAA**gt**aagcagatggtggg	8.3	2.5
48	250	tgtgtgcatcccct**ag**ATTTGGAGAT	5.6	ACTGGACCAG**gc**aagttggccctggg	4.7	1.5
49	87	cttctcttacctct**ag**GTGTTTGTAA	11.4	ACAAGCCCAG**gt**acccctgctgctta	4.1	2.5
50	3′	tgatctccttccac**ag**GACTGATCTT	9.2			

RNA, used for that analysis, had been obtained from 76 pooled adult human retinas, making it impossible to assess the extent of this observation in individual cases. At the same time, analysis of the exon 30 sequence revealed the presence of a strong cryptic acceptor site (R_i 9.4 bits; Table 1) 114 bp downstream (position 4467) of the original exon 30 splice acceptor (R_i 5.0 bits; Table 1). This cryptic splice acceptor site is apparently used in some cases. Other researchers have noticed the same phenomenon in independent studies (Cremers et al., 1998) and proposed to designate the two variants exon 30A (187 bp) and exon 30B (73 bp; Table 1).

3.2. 5′ and 3′ UTRs of the ABCR gene

To determine the 5′ end of the *ABCR* mRNA, 5′ RACE was performed using Marathon Ready retina cDNA (Clontech) as a template. The resulting sequence was compared with those obtained from ESTs and with the sequence of the genomic phage clone. The putative transcription initiation site was identified 90 bp upstream of the ATG (Fig. 3). The CpG content of the promoter region was about 50%, showing no obvious CpG island in the immediate 5′ region of the gene. This is consistent with the observation that genes with tissue-specific expression do not harbor CpG islands in their 5′ ends in at least 50% of all cases. The 3′UTR was contained in exon 50 and spanned only 402 bp (data not shown). A polyadenylation signal (AATAAA) was identified 372 bases distal from the stop codon.

3.3. Information analysis of the ABCR splice acceptor and donor sites sequences

Splice junction sequences of all 50 exons were analyzed for their individual information content (R_i, bits; Schneider, 1997a,b; Rogan et al., 1998). The information theory-based model allows one to estimate which nucleotides are permissible at both conserved and variable positions of splice donor and acceptor sites (Stephens and Schneider, 1992). The information content (R_i, in bits) of a member of any sequence family describes the degree to which that member contributes to the conservation of the entire family (Schneider, 1997a,b). The effects of all, even seemingly insignificant, nucleotide changes are detectable, given that the R_i value is cumulative over all positions in a splice site.

The mean of the distribution of individual information contents (R_i values, $R_{sequence}$) for splice acceptor and donor sites has been estimated to be 9.35 ± 0.12 bits for the 28 nucleotide long splice acceptors and 7.92 ± 0.09 bits for the 10-nucleotide-long splice donors (Stephens and Schneider, 1992). Splice sites with R_i values significantly greater or lower than $R_{sequence}$ were considered strong and weak sites, respectively. Sites with R_i values ≤ 0 are considered definitely non-functional (Schneider, 1997b). The minimal information ($R_{i,min}$), required for splicing, has been estimated to be approximately 2.4 bits, but the confidence interval around $R_{i,min}$ is not known (Rogan et al., 1998).

For the 50 splice acceptor and donor sites of the *ABCR* gene, the average R_i values were 9.49 ± 0.37 and 7.45 ± 0.32, respectively, in good correlation with those predicted. The R_i values ranged between 3.8 and 15.4 for splice acceptors and 3.6 to 12.0 for splice donors (Table 1). In general, splice sites contained well-known conserved AG and GT sequences in acceptors and donors, respectively. In two cases, in splice donors of exons 46 and 48, a GC sequence was present instead of GT (Table 1, Fig. 1a). The individual information R_i values for these two donor sites were 4.1 and 4.7 bits, respectively, which is below average but still well above the assumed minimum information required for splicing ($R_{i,min} \approx 2.4$ bits; Rogan et al., 1998). The site conservation is the sum of its parts. Although the GC sequence is rare in splice donors, it does not violate the theory since the remainder of the site compensates for the T to C change. Stephens and Schneider (1992) found a total of 10 cases out of 1799 splice donors with this change. Two splice sites with this change have not been described in a single gene before. In addition, these sites are not recognized as splice donors by programs used for exon detection, e.g. GRAIL, GeneFinder, etc. A sequence analysis of these splice junctions on at least 10 different individual DNAs confirmed that they represent the actual splice donor sequences.

3.4. Analysis of potential mutations in ABCR splice sites

Usually, nucleotide alterations in splice junction recognition sequences can be divided into three categories:

(1) Primary mutations, resulting in inactivation of a particular splice site without creating a cryptic site.
(2) Those that create (activate) a cryptic site near the primary splice site, with or without inactivation of the primary site.
(3) Variants not altering splicing (Krawczak et al., 1992; Rogan et al., 1998).

A total of nine nucleotide variations, occurring near splice acceptor and donor sites of 150 STGD and 167 AMD patients, were analyzed for being potential splice altering mutations. Three of these variants were in splice acceptors (G863A, 5585+1G/A, V2050L), and six in donors (4253+5G/T, 5196+1G/A, 5196+2T/C, 5714+5G/A, 5898+1G/T, 6005+1G/T). Two out of three splice acceptor alterations were at the first base of an exon (+1 position), resulting also in a change of amino acid. Splice donor variants were all in intron sequences.

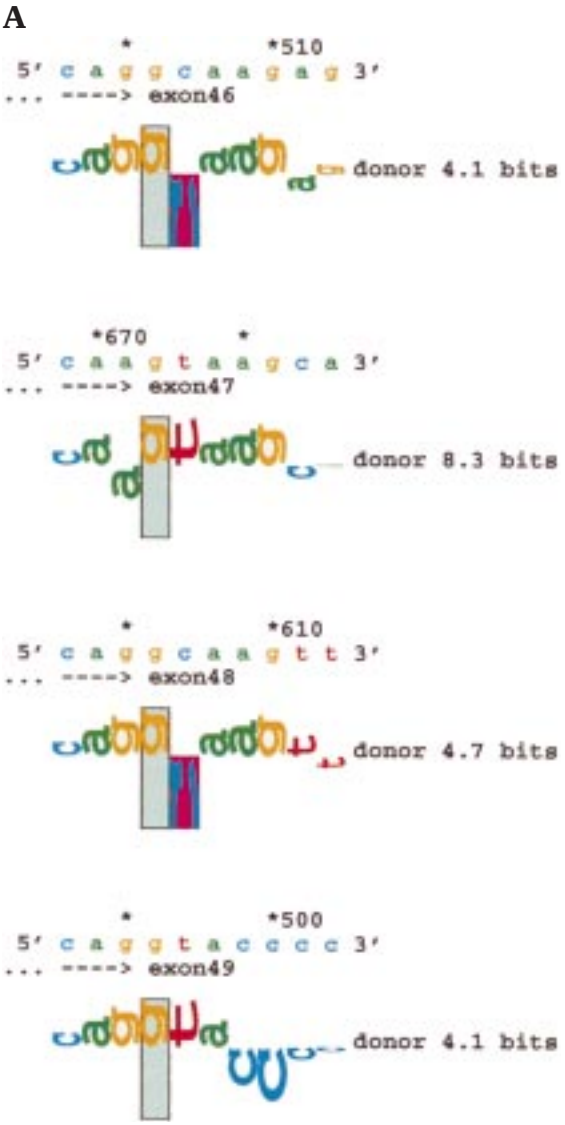
Sequence changes in the first two intronic nucleotides (5585+1G/A, 5196+1G/A, 5196+2T/C, 5898+1G/T and 6005+1G/T) constituted primary mutations, directly altering splicing (Table 2). R_i values for three out of five splice junctions harboring these changes dropped below zero (Table 2), presumably totally inactivating the corresponding splice site. The V2050L mutation lowered the information content of the splice acceptor site in exon 45 insignificantly, suggesting that this variant did not affect splicing but represented a missense mutation (data not shown).

Another sequence change in the coding region of the exon 17 splice acceptor site, G863A, had a dual effect (Fig. 1b). Besides being a putative missense mutation, this change lowered the R_i value of the original splice acceptor at position 365 by two bits and created an

Table 2
Analysis of ABCR splice site variants

Exon	Mutation	Disease	Number of patients	wt R_i	mt R_i	Predicted effect on the ABCR protein
17A	G863A	STGD/AMD	11/1	11.6	9.7	Change of aa or deletion of one aa[a]
28D	4253+5G/T	STGD	1	8.2	4.3	Partially functioning splice site
36D	5196+1G/A	AMD	1	6.8	−6.0	Non-functional protein
36D	5196+2T/C	STGD	1	6.8	−0.7	Non-functional protein
40A	55851G/A	STGD	1	5.9	−1.6	Non-functional protein
40D	5714+5G/A	STGD	8	7.2	3.7	Partially functioning splice site
42D	5898+1G/T	STGD	3	8.6	0.8	Non-functional protein
43D	6005+1G/T	STGD	1	11.2	3.4	Partially functioning splice site
45A	V2050L	STGD	2	12.5	10.6	Change of aa, no effect on splicing

'A' or 'D' after the exon number indicates splice acceptor or donor sequences, respectively. The numbering of the nucleotides is according to the *ABCR* cDNA sequence.
STGD, Stargardt disease; AMD, age-related macular degeneration. wt, wild type; mt, mutant; R_i, individual information content of the splice junction in bits.
[a]This change created an equally strong, in-frame, cryptic splice site.



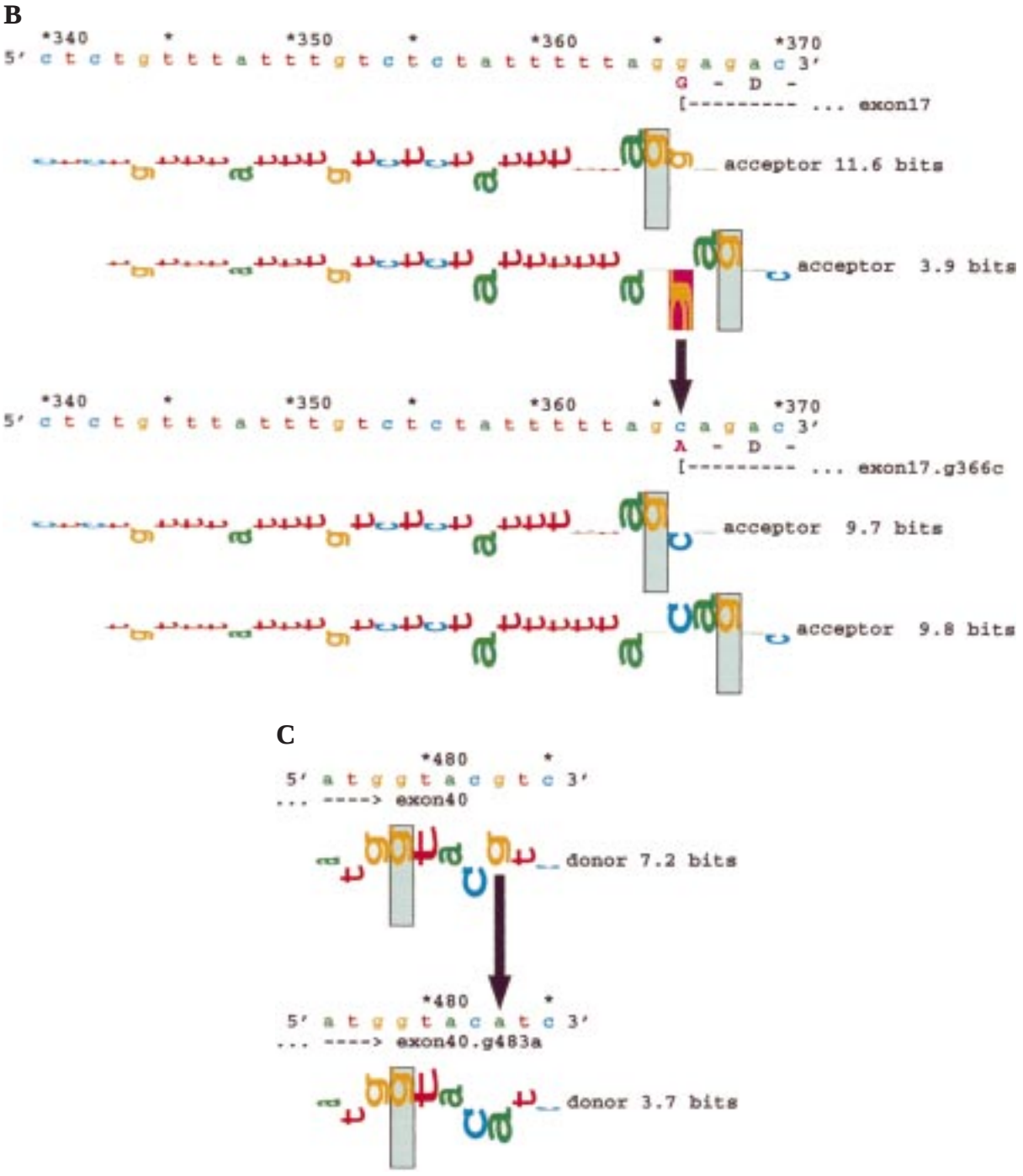


Fig. 1. Splice site walkers for six *ABCR* exon/intron junctions. Each walker is shown by a set of consecutive letters. The height of every letter is in bits of sequence conservation, with the scale given by the vertical bar at the zero coordinate running from −3 to +2 bits. The total conservation given for each site is the sum of the individual letter heights (Schneider, 1997a). (**A**) Last four splice donor sites of the *ABCR* gene. Two of the sites, exons 46 and 48, contain a GC sequence instead of the usual GT at positions +1 and +2. Nevertheless, the total information content (R_i) of these two splice donors is the same as in exon 49, where the usual GT dinucleotide is present. Exon 47 represents an example of an average strength splice donor. (**B**) Cryptic site creation concurrent with the weakening of the natural site. A G→C (G863A) mutation in the first nucleotide of exon 17. In addition to being a putative missense mutation, this change decreases the information content of the natural acceptor at position 365 and creates an equally strong in-frame cryptic site at position 368. (**C**) Mild splice junction mutation. A G→A mutation five nucleotides downstream of exon 40 donor site. This variant was observed in eight out of 148 (5.4%) STGD patients but was not detected in over 400 control individuals. The reduction in information content is significant even though the R_i value is still greater than $R_{i,min}$.

equally strong in-frame cryptic site at position 368. At this point, we cannot evaluate which one of these splice acceptors is actually used in vivo because the *ABCR* gene is expressed only in rod photoreceptors, therefore making it difficult to obtain patient RNA.

Two nucleotide changes occurred in exons 28 and 40, at the IVS+5 position of the splice donor sites. These variants were detected in one out of 148 and eight out of 148 STGD patients, respectively, and were not found in over 400 general population controls. Comparison of the information content of the mutant sites with the wt sites revealed that R_i values of both mutated sites dropped by approximately 4 bits (Fig. 1c). Although the R_i values of the mutant splice donors are still above $R_{i,min}$, so that the total inactivation of the splice site is unlikely, these changes could frequently alter the splicing of corresponding exons, yielding a non-functional protein product.

3.5. *Evolutionary conservation of the* ABCR *gene*

The *ABCR* gene is a typical representative of the ABC superfamily that consists of two TM domains, each consisting of six membrane spanning segments, and two highly conserved ATP domains (Allikmets et al., 1997a). This gene is a member of the ABC1 subfamily that includes only transporters of multicellular organisms (Michaelis and Berkower, 1995) and is most closely related to the mouse *Abc1* gene (Luciani et al., 1994; Allikmets et al., 1997a). A comparison of *ABCR* intron locations with those of mouse *Abc1* revealed an unexpectedly high degree of identity, even in regions significantly different in amino acid sequence (data not shown). The overall identity of ABCR and Abc1 amino acid sequences was estimated to range from 50 to 85%, depending on the gene region (Allikmets et al., 1997a). At the same time, 45 out of 50 (90%) splice site locations in *Abc1* were identical to those in *ABCR* (data not shown). The mouse *Abc1* gene has been shown to be highly expressed in macrophages and associated with the consumption of apoptotic cells (Luciani and Chimini, 1996). The human *ABC1* gene is located on chromosome 9q22–q31 and is not yet fully characterized.

Previously, we have identified six subfamilies of genes within the ABC superfamily (Allikmets et al., 1996). Here, we studied the ABC1 subfamily in greater detail. Since our last report, full-length sequences of five genes from different organisms belonging to this subfamily (*ABCR*, *ABC-C*, *Abc1*, *Abc2* and *Ceabc*) have become available. In addition, we identified and partially sequenced a new gene, EST1133530, another member of the subfamily.

N-terminal and C-terminal conserved ATP-binding domains of human *ABCR* and *ABC-C*, mouse *Abc1* and *Abc2*, and *C. elegans abcc* genes were analyzed (Luciani et al., 1994; Allikmets et al., 1995, 1996, 1997a). In addition, C-terminal domains of three partially characterized human genes, designated as EST155051, EST90625 and EST1133530, were included in the analysis, as well as the human oligoA binding protein gene as a representative of another subfamily (Allikmets et al., 1996). The amino acid sequences were aligned with PILEUP and trimmed to a common overlapping region of 220 residues. Two distinct methods of phylogenetic reconstruction, maximum parsimony and distance-based neighbor-joining, yielded similar results shown as the maximum parsimony phylogram on Fig. 2. In addition, a bootstrap resampling of 100 iterations in both methods was performed to gauge the reliability of our data to derive the same tree consistently (see Section 2.1). Analysis resulted in three distinct groups statistically supported by bootstrap values greater than 70 (Fig. 3). As in the case of the CFTR/MRP subfamily (Allikmets et al., 1996), the N- and C-terminal domains of the different proteins clustered together and diverged significantly from one another. The C-terminal

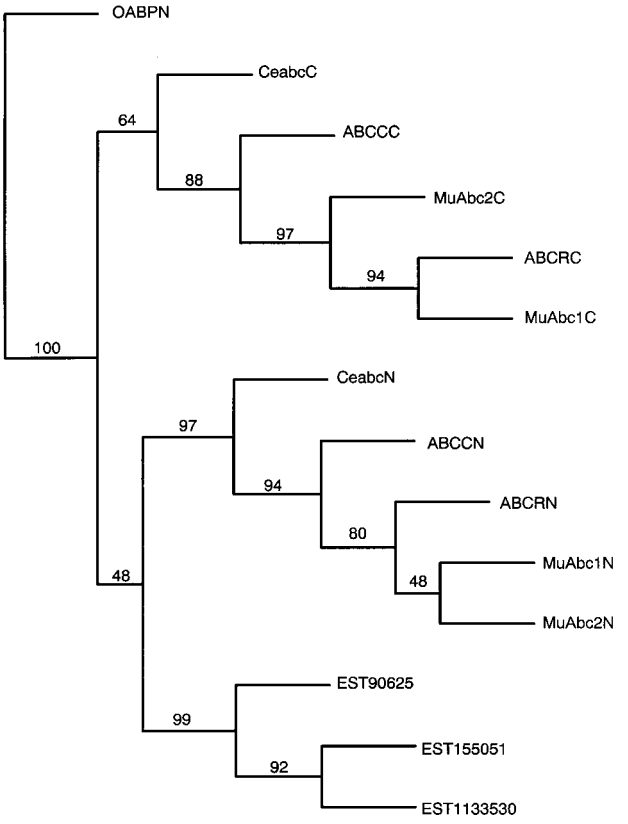


Fig. 2. Evolutionary relationship of ABC transporters within the ABC1 subfamily. An unrooted maximum parsimony phylogram of members of the subfamily is presented. EST numbers indicate NCBI ID numbers of the partially characterized genes. N and C at the end of the gene name indicate the N-terminal or C-terminal ATP-binding domain. OABP, human oligoA binding protein; ABCR, human rod photoreceptor-specific transporter; ABCC, human ABC-C gene; MuAbc1, mouse Abc1 gene; MuAbc2, mouse Abc2 gene; Ceabc; *C. elegans* Abc gene. Numbers on the internal nodes represent bootstrap values out of 100.

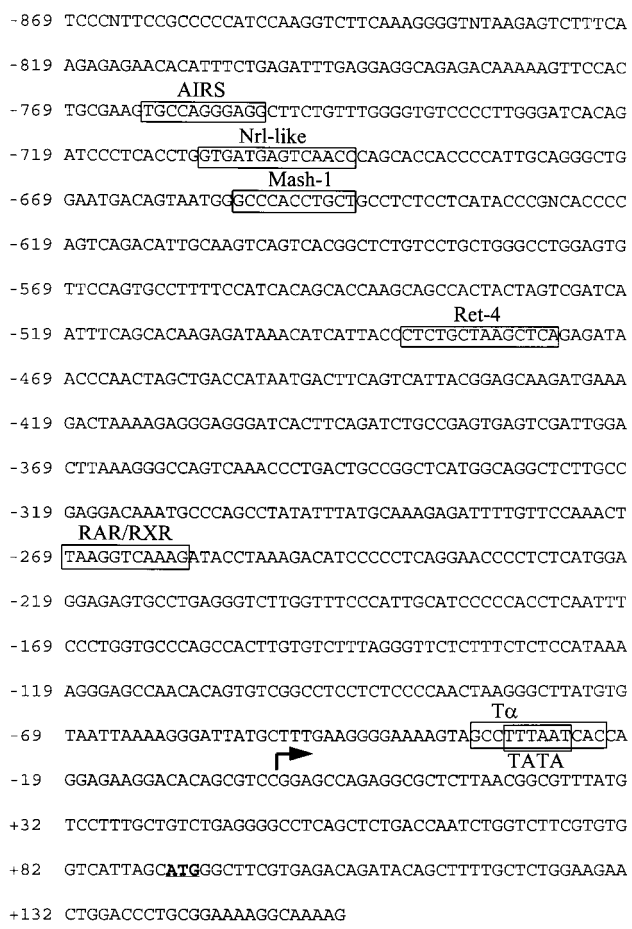


Fig. 3. *ABCR* promoter sequence with predicted transcription factor binding sites. Potential binding sites for transcription factors associated with gene expression in photoreceptor cells are boxed. A potential TATA box sequence is boxed as well. The putative transcription start site is labelled with an arrow, and the first ATG in the *ABCR* open reading frame is underlined.

sequences of the partially characterized genes formed a separate, statistically supported subgroup, indicating an earlier divergence of these genes from the rest of the proteins of this subfamily.

3.6. *Analysis of the* ABCR *promoter*

Tissue-specific gene expression is primarily controlled at the transcriptional level through regulatory sequences in gene promoters. The promoters of a variety of genes with photoreceptor-specific expression patterns have been partially analyzed to identify some of the enhancers conferring tissue specific expression. Promoters have been studied from the following genes: rhodopsin, S-antigen/arrestin, rod cyclic monophosphate phosphodiesterase-b, rod a-transducin, QR1, interphotoreceptor retinoid binding protein (IRBP), red/green opsin, c-transducin, blue opsin, cone a-transducin/GNAT2, and metabotropic glutamate receptor 6 (Table 3). While a diverse set of transcription factors have been linked to defects in eye development (Freund et al., 1996), the experimentally determined regulatory enhancers can be primarily classified into three groups: Nrl, Ta/CRX, and PCE-I/Ret-1 sites. Observations of other elements have been reported, including binding sites in the well-studied rhodopsin gene for Mash-1 (34) and Ret-4 (Ahmad, 1995), and a binding site in the IRBP gene (Bobola et al., 1995) for an unknown protein. Computational analysis of the *ABCR* promoter was attempted to determine whether any similar sites are present.

Analysis of the *ABCR* promoter with the on-line programs of the transcription factor database Transfac (Quandt et al., 1995) failed to identify any photoreceptor-specific elements. Within the set of 219 matches to Transfac sites, the most interesting sites were for members of the AP-1 (position −708) and Retinoic Acid/Retinoid Receptor (RAR/RXR) (position −269) families. The RXR/RAR factors have been linked to eye development (Freund et al., 1996), although no binding site for these factors has been proven to have a role in photoreceptor-specific gene expression. The AP-1 site is interesting due to a past case in which an AP-1-like site in a photoreceptor promoter was subsequently identified as a functional site for the Nrl transcription factor (Rehemtulla et al., 1996).

Directed computer searches for *ABCR* promoter sequences similar to the known photoreceptor gene regulatory sites identified several potential sites (Fig. 3). The putative AP-1-like sequence at −708 is similar to the known Nrl sites (Table 2). Two apparently dissimilar types of binding sites, PCE-I (Kikuchi et al., 1993) and Ta (Ahmad et al., 1994), were initially linked to the same transcription factor, Ret-1. Recently, the photoreceptor-specific factor CRX was identified as the activator of expression through the Ta elements (Chen et al., 1997; Furukawa et al., 1997). No matches to the PCE-I sites are present in the *ABCR* promoter, but a CRX/Ta-like sequence was found at position −33 (Table 2). This site overlaps a potential TATA box (Fig. 3), which is similar to the positioning of the Ta sequences in the IRBP (Ahmad et al., 1994) and PDE-b (observation) genes. A site resembling the recently described Ret-4 element was identified at position −489, and a Mash-1-like sequence was found at position −655 (Table 2). A subset of the putative sites are in close proximity (positions −770 to −620 in Fig. 3).

In addition to the above putative sites, an interesting sequence was present at position −762. This sequence is similar to the IRBP gene 'B' site (Bobola et al., 1995). To determine whether this sequence is common in regulatory regions of photoreceptor genes, the available photoreceptor promoter sequences were analyzed. In addition to *ABCR* and *IRBP*, similar sequences were present at two locations in both the rhodopsin and S-antigen/arrestin genes (Table 2). To reflect the gene distribution of the sequences, we have collected the first

Table 3
Regulatory sequences linked to gene expression in photoreceptor cells

Site	Gene	Species	Sequence	Proof	Reference
RefNRL	Rhodopsin	Human	TGCTGATTCAGCCA	E	Rehemtulla et al. (1996)
		Mouse	TGCTGAATCAGCCT		
	PDE-b	Human	GAGTGAGTCAGCTG	E	Di Polo et al. (1996)
		Mouse	GTATGAGTCAGCTG		
	S-Ag/Arr	Human	GGTTGACATTTCTC	C	
		Mouse	GGCTGACCTTTCTC		
	QR1	Quail	AGCTGACAGGACAG	E	Pouponnot et al. (1995)
	ABCR	Human	GGTTGAGTCATCAC	P	
PCE/Ret-1	Rhodopsin	Human	AGCCAATTAGGCCC	E	Morabito et al. (1991); Kumar et al. (1996)
		Mouse	AGCCAATTAGGCCC		
	S-Ag/Arr	Human	TTTCATTTAGCTGT	C	
		Mouse	TTTCAATTAGCTAT		
	IRBP	Human	GGTCAATTAGCTAA	C	
		Mouse	GGTCAATTAGCTAA		
	Rd/Gr Op	Human	CATCAATTAGCAGA	C	
		Mouse	CATCAATTAGCACT		
	Rd/Gr Op	Human	GCCCAATTAAGAGA	C	
		Mouse	GCCCAATTAAGAGA		
	ABCR	Human	No site detected		
CRX/Ta	Rhodopsin	Human	GTGATTATGCCC	E	Nie et al. (1996)
		Mouse	GTGATTAAGACC		
	Rhodopsin	Human	GGGATTAATATG	C	
		Mouse	GGGATTAGCGTT		
	PDE-b	Human	GAGATTAGGAAC	C	
		Mouse	ATGATTAGGGAG		
	S-Ag/Arr	Human	TTGATTAAACTC	C	
		Mouse	CTGATTAAGCTC		
	IRBP	Human	AGGATTAAAGGC	E	Bobola et al. (1995)
	Transduc-a	Mouse	CAGATTAAAATG	E	Ahmad et al. (1994)
	ABCR	Human	GTGATTAAAGGC	P	
Ret-4	Rhodopsin	Human	GAGCTTAGGGAGGG	E	Chen and Zack (1996)
		Mouse	TGGCTTAGGAGGGG		
	ABCR	Human	TGAGCTTAGCAGAG	P	
Mash-1	Rhodopsin	Rat	TTCCACCTGAT	E	Ahmad (1995)
		Human	TTCCTCCTAGT		
	PDE-b	Human	CACCACCTTCC	C	
		Mouse	CACCACCTTCC		
	ABCR	Human	GCCCACCTGCT	P	
AIRS	IRBP	Human	CCACCCTGGCC	E	Bobola et al. (1995)
	Rhodopsin	Human	TCACCTTAACC	E	Nie et al. (1996)
		Mouse	CCACCTTGACC		
	Rhodopsin	Human	TCACCTTGGCC	C	
		Mouse	TCACCTTGGCC		
	S-Ag/Arr	Human	CCTCTTTGGAT	C	
		Mouse	CCTCTTTAGGT		
	S-Ag/Arr	Human	TCTCCTTGACC	C	
		Mouse	TCTCCTTGACC		
	ABCR	Human	CCTCCCTGGCA	P	

Experimentally verified elements (E) and similar promoter sequences conserved through evolution (C) are presented. Putative (P) ABCR sites are listed at the end of each section.

letter of each gene name to generate the term 'AIRS' element.

4. Discussion

Here, we present the structural characterization of the *ABCR* gene, a member of the ABC transporter superfamily involved in at least three different inherited eye diseases. This study included characterization of all splice junctions of the *ABCR* gene and variants occurring in their close vicinity, promoter sequence analysis and determination of phylogenetic relationship of this gene to other known members of the ABC1 subfamily. Analysis of the genomic organization of the *ABCR* gene showed that it is a typical representative of the ABC

transporter superfamily, comprised of two transmembrane domains and two conserved ATP-binding segments. The coding region of the gene is divided into 51 exons, spanning about 100 kb on the short arm of the human chromosome 1p22.2–1p22.3 region. Evolutionary analysis revealed that this gene is most closely related to mouse and human *Abc1* and *Abc2* genes, two other members of the ABC1 subfamily. Genes from this subfamily have been found only in multicellular organisms, as opposed to genes from other ABC subfamilies, which are present in all characterized genomes. Recent advances in sequencing of complete genomes of bacteria and yeast have allowed the identification of all putative ABC genes in these organisms. Since no orthologs of ABC1 subfamily genes have been found in bacteria or yeast, it means that genes from this subfamily diverged in multicellular organisms and evolved to perform highly specialized functions. The phylogenetic analysis showed that genes from the ABC1 subfamily resemble those from the CFTR/MRP subfamily (Allikmets et al., 1996) in that their N-terminal and C-terminal ATP-binding domains are significantly diverged from each other.

An estimated 15% of all genetic disease-causing point mutations affect mRNA splicing (Krawczak et al., 1992). Essential sequences in donor and acceptor splice junctions have been defined by different methods: consensus sequences (Mount, 1982), nucleotide frequency analysis (Senapathy et al., 1990) and neural network predictions (Brunak et al., 1990). The information theory-based model of splice junctions takes into account all information contained in both highly conserved and variable positions (Schneider, 1997a). Substitutions at highly conserved positions of a splice site have always been considered deleterious; however, the total information content (R_i) must be computed and compared with that of other sites before making a conclusion of a possible splice site mutation (Rogan et al., 1998). The variants that result in negligible changes in the information content are usually considered polymorphisms, whereas mutant sites have significantly less information than the corresponding wt sites. Utilizing this methodology, we were able to characterize all known sequence variants of the ABCR splice junctions within three categories. Four out of nine variants were predicted to abolish splicing completely and therefore constituted primary mutations (Table 2). Three changes created cryptic splice sites or weakened the original splice site sufficiently to predict reduced splicing fidelity. One remaining variant did not change the information content of the splice acceptor and was classified as a putative missense mutation (Table 2). The information theory-based method of splice site analysis was determined to be a powerful tool for assessing individual information in splice junction sequences, allowing us to draw conclusions about potential functional impact of any given sequence variant.

Computational analysis of the *ABCR* promoter region identified putative sites at positions −762 (AIRS), −708 (AP-1/Nrl-like), −655 (Mash), −489 (Ret-4), −269 (RAR/RXR), and −33 (CRX/Ta). Of course, the functional roles for these putative sites must be assessed by experimental measures, but there are two interesting features to note. First, the AIRS element appears to be a common sequence in the promoters of several photoreceptor genes. It has been found to be specifically bound by a protein in gel mobility shift assays with the IRBP site (Bobola et al., 1995) and the rhodopsin site is protected in a footprint assay (Nie et al., 1996). Second, the cluster of putative sites at positions −762, −708, and −655 is suggestive of a tissue-specific regulatory module. Studies of a variety of genes, for instance b-globin, myosin light chain 1/3, and the T-cell receptor, have shown that regulatory regions conferring tissue-specific expression feature multiple functional transcription factor binding sites within short regions (Arnone and Davidson, 1997). It will be interesting to determine whether the computationally predicted features are functional. Study of the AIRS element and its binding protein could be useful in deciphering how genes are specifically expressed in photoreceptor cells.

In conclusion, a detailed characterization of the *ABCR* gene sequence will enable identification of all sequence variants found in the gene as well as potential function-altering mutations and enhances the functional characterization of this gene.

Acknowledgement

We thank Stan Cevario for oligonucleotide synthesis and assistance with DNA sequencing and James Fickett for support and advice. Computing resources were in part provided by the Frederick Biomedical Supercomputing Center. P.K.R. acknowledges support from the American Cancer Society (DHP-132) and the National Cancer Institute (CA74683-02). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

References

Ahmad, I., Yu, X., Barnstable, C.J., 1994. A *cis*-acting element, T alpha-1, in the upstream region of rod alpha-transducin gene that binds a developmentally regulated retina-specific nuclear factor. J. Neurochem. 62, 396–399.

Ahmad, I., 1995. Mash-1 is expressed during ROD photoreceptor

differentiation and binds an E-box, E(opsin)-1 in the rat opsin gene. Brain Res. 90, 184–189.

Allikmets, R., Gerrard, B., Glavač, D., Ravnik-Glavač, M., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., Modi, W., Dean, M., 1995. Characterization and mapping of three new mammalian ATP-binding transporter genes from an EST database. Mammal. Genome 6, 114–117.

Allikmets, R., Gerrard, B., Hutchinson, A., Dean, M., 1996. Characterization of the human ABC superfamily: Isolation and mapping of 21 new genes using the EST database. Hum. Mol. Genet. 5, 1649–1655.

Allikmets, R., Singh, N., Sun, H., Shroyer, N.F., Hutchinson, A., Chidambaram, A., Gerrard, B., Baird, L., Stauffer, D., Peiffer, A., Rattner, A., Smallwood, P., Li, Y., Anderson, K.L., Lewis, R.A., Nathans, J., Leppert, M., Dean, M., Lupski, J.R., 1997a. A photoreceptor cell-specific ATP-binding transporter gene *(ABCR)* is mutated in recessive Stargardt macular dystrophy. Nature Genet. 15, 236–246.

Allikmets, R., Shroyer, N.F., Singh, N., Seddon, J.M., Lewis, R.A., Bernstein, P., Peiffer, A., Zabriskie, N., Li, Y., Hutchinson, A., Dean, M., Lupski, J.R., Leppert, M., 1997b. Mutation of the Stargardt disease gene *(ABCR)* in age-related macular degeneration. Science 277, 1805–1807.

Arnone, M.I., Davidson, E.H., 1997. The hardwiring of development: organization and function of genomic regulatory systems. Development 124, 1851–1864.

Bobola, N., Hirsch, E., Albini, A., Altruda, F., Noonan, D., Ravazzolo, R., 1995. A single *cis*-acting element in a short promoter segment of the gene encoding the interphotoreceptor retinoid-binding protein confers tissue-specific expression. J. Biol. Chem. 270, 1289–12943.

Brunak, S., Engelbrecht, J., Knudsen, S., 1990. Neural network detects errors in the assignment of mRNA splice sites. Nucleic Acids Res. 18, 4797–4801.

Chen, S., Zack, D.J., 1996. Ret 4, a positive acting rhodopsin regulatory element identified using a bovine retina in vitro transcription system. J. Biol. Chem. 271, 28549–28557.

Chen, S., Wang, Q.L., Nie, Z., Sun, H., Lennon, G., Copeland, N.G., Gilbert, D., Jenkins, N.A., Zack, D.J., 1997. Crx, a novel otx-like paired-homeodomain protein binds to and transctivates photoreceptor cell-specific genes. Neuron 19, 1017–1030.

Cremers, F.P., van de Pol, D.J., van Driel, M., den Hollander, A.I., van Haren, F.J., Knoers, N.V., Tijmes, N., Bergen, A.A., Rohrschneider, K., Blankenagel, A., Pinckers, A.J., Deutman, A.F., Hoyng, C.B., 1998. Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. Hum. Mol. Genet. 7, 355–362.

Dean, M., Allikmets, R., 1995. Evolution of ATP-binding cassette transporter genes. Curr. Opin. Gen. Dev. 5, 79–785.

Di Polo, A., Rickman, C.B., Farber, D.B., 1996. Isolation and initial characterization of the 5′ flanking region of the human and murine cyclic guanosine monophosphate-phosphodiesterase beta-subunit genes. Invest. Ophthalmol. Vis. Sci. 37, 551–560.

Freund, C., Horsford, D.J., McInnes, R.R., 1996. Transcription factor genes and the developing eye: a genetic perspective. Hum. Mol. Genet. 5, 1471–1488.

Furukawa, T., Morrow, E.M., Cepko, C.L., 1997. *Crx*, a novel *otx*-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. Cell 91, 531–541.

Kaplan, J., Gerber, S., Larget-Piet, D., Rozet, J.M., Dollfus, H., Dufier, J.L., Odent, S., Postel-Vinay, A., Janin, N., Briard, M.L., Frezal, J., Munnich, A., 1993. A gene for Stargardt's disease (*fundus flavimaculatus*) maps to the short arm of chromosome 1. Nature Genet. 5, 308–311.

Kikuchi, T., Raju, K., Breitman, M.L., Shinohara, T., 1993. The proximal promoter of the mouse arrestin gene directs gene expression in photoreceptor cells and contains an evolutionarily conserved retinal factor-binding site. Mol. Cell Biol. 13, 4400–4408.

Klugbauer, N., Hofmann, F., 1996. Primary structure of a novel ABC transporter with a chromosomal localization on the band encoding the multidrug resistance-associated protein. FEBS Lett. 391, 61–65.

Krawczak, M., Reiss, J., Cooper, D.N., 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum. Genet. 90, 41–54.

Kumar, R., Chen, S., Scheurer, D., Wang, Q.L., Duh, E., Sung, C.H., Rehemtulla, A., Swaroop, A., Adler, R., Zack, D.J., 1996. The bZIP transcription factor Nrl stimulates rhodopsin promoter activity in primary retinal cell cultures. J. Biol. Chem. 271, 29612–29618.

Lincoln, A.L., Daly, M., Lander, E., 1991. PRIMER: a computer program for automatically selecting PCR primers. Whitehead Institute Technical Report.

Luciani, M.F., Denizot, F., Savary, S., Mattei, M.G., Chimini, G., 1994. Cloning of two novel ABC transporters mapping on human chromosome 9. Genomics 21, 150–159.

Luciani, M.-F., Chimini, G., 1996. The ATP binding cassette transporter ABC1, is required for the engulfment of corpses generated by apoptotic cell death. EMBO J. 15, 226–235.

Martinez-Mir, A., Paloma, E., Allikmets, R., Ayuso, C., del Rio, T., Dean, M., Vilageliu, L., Gonzales-Duarte, R., Balcells, S., 1998. Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene *ABCR*. Nature Genet. 18, 11–12.

Michaelis, S., Berkower, C., 1995. Sequence comparison of yeast ATP binding casette (ABC) proteins. In: Cold Spring Harbor Symposium on Quantitative Biology, vol. LX: Protein Kinesis—The Dynamics of Protein Trafficking and Stability. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Morabito, M.A., Yu, X., Barnstable, C.J., 1991. Characterization of developmentally regulated and retina-specific nuclear protein binding to a site in the upstream region of the rat opsin gene. J. Biol. Chem. 266, 9667–9672.

Mount, S.M., 1982. A catalogue of splice junction sequences. Nucleic Acids Res. 10, 459–472.

Nie, Z., Chen, S., Kumar, R., Zack, D.J., 1996. RER, an evolutionarily conserved sequence upstream of the rhodopsin gene, has enhancer activity. J. Biol. Chem. 271, 2667–2675.

Papermaster, D.S., Converse, C.A., Zorn, M., 1976. Biosynthetic and immunochemical characterization of large protein in frog and cattle rod outer segment membranes. Exp. Eye Res. 23, 105–115.

Pouponnot, C., Nishizawa, M., Calothy, G., Pierani, A., 1995. Transcriptional stimulation of the retina-specific QR1 gene upon growth arrest involves a Maf-related protein. Mol. Cell Biol. 15, 5563–5575.

Quandt, K., Frech, K., Karas, H., Wingender, E., Werner, T., 1995. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. 23, 4878–4884.

Rehemtulla, A., Warwar, R., Kumar, R., Ji, X., Zack, D.J., Swaroop, A., 1996. The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. Proc. Natl. Acad. Sci. USA 93, 191–195.

Rogan, P.K., Faux, B.M., Schneider, T.D., 1998. Information analysis of human splice site mutations. Hum Mutat. 12, 153–171.

Schneider, T.D., Stephens, R.M., 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100.

Schneider, T.D., 1997a. Sequence Walkers: a graphical method to display how binding proteins interact with DNA or RNA sequences. Nucleic Acids Res. 25, 4408–4415.

Schneider, T.D., 1997b. Information content of individual genetic sequences. J. Theor. Biol. 189, 427–441.

Senapathy, P., Shapiro, M.B., Harris, N.L., 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Meth. Enzymol. 183, 252–278.

Stephens, R.M., Schneider, T.D., 1992. Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. J. Mol. Biol. 228, 1124–1136.