Mouse Opsin

GENE STRUCTURE AND MOLECULAR BASIS OF MULTIPLE TRANSCRIPTS*

(Received for publication, March 16, 1990)

Muayyad R. Al-Ubaidi, Steven J. Pittler, Mary S. Champagne, John T. Triantafyllos, James F. McGinnis, and Wolfgang Baehr

From the Department of Ophthalmology, Cullen Eye Institute, Baylor College of Medicine, Houston, Texas 77030 and the Department of Anatomy and Cell Biology, Brain Research Institute and The Mental Retardation Research Center, UCLA School of Medicine, Los Angeles, California 90024

The single copy mouse opsin gene produces five major transcripts, varying in size from 1.7 to 5.1 kilobases. The mRNAs are present at levels that vary over 2 orders of magnitude and can be detected as early as postnatal day 1. Each of the transcripts is polyadenylated and can be identified in polysome-bound RNA, suggesting that each is translated in vivo. To elucidate the molecular basis of this complex transcription pattern, we have characterized genomic fragments covering the entire mouse opsin gene, including several kilobases of 5'- and 3'-flanking regions. Transcription initiates at a single site 97 base pairs upstream of the translation start codon. Northern hybridization with exon- and intron-specific probes demonstrated that the various transcripts are not generated by partial or alternative splicing. Sequence analysis of the 3' end of the gene showed the presence of multiple polyadenylation signals. Analysis by polymerase chain reaction of the 3' end of opsin cDNA demonstrated that the complex transcription pattern originated from the selective use of these polyadenylation sites, generating transcripts that differ only in the length of the 3'-untranslated region. Transcript heterogeneity similar to that observed in mouse was also found in rat and, to a lesser degree, in human and frog opsin mRNAs.

Phototransduction in mammalian rod photoreceptors operates via a complex enzyme cascade initiated by the absorption of light by rhodopsin (1, 2). Rod rhodopsins are comprised of a chromophore, 11-cis retinal, and a 49-kDa apoprotein, opsin (3, 4). The amino acid sequences of bovine (5, 6) and ovine opsins (7) consisting of 348 residues have been determined by peptide sequencing. The mammalian opsin genes that have been identified thus far are present in single copy in their respective genomes (8). The human gene was mapped to the long arm of chromosome 3 (9, 10), and the mouse gene, to chromosome 6 (11). Genes and cDNAs encoding opsins have been described for various vertebrate and invertebrate species, including bovine (12), human (13), Drosophila (14, 15), mouse (16), octopus (17), and chicken (18). The characteristic organization of the mammalian opsin gene is a 5-exon, 4-intron arrangement with highly homologous exon sequences and with precise conservation of intron positions. Mutations in the opsin gene may disrupt phototransduction and therefore lead to blindness. Recently, a point mutation in exon 1 (a C to A conversion in codon 23 resulting in a Pro to His substitution) of the human opsin gene has been implicated as causative of one form of autosomal dominant retinitis pigmentosa (19). In an earlier, unrelated study, an autosomal dominant retinitis pigmentosa gene in a large Irish pedigree was mapped to the long arm of chromosome 3 (20). In order to facilitate studies aimed at the establishment of a mouse model for this blinding disorder we have extended our preliminary characterization of the mouse opsin gene (16). It was shown previously that transcription of the gene produces at least five major transcripts (16, 21, 22). The sizes of the RNAs range from 1.7 to 5.1 kb, and the various species are present in the rod cytoplasm at levels that vary over 2 orders of magnitude (16). The major transcript of the bovine opsin gene, in contrast, is a 3-kb species which was shown to contain a large 3'-untranslated region (12). The molecular basis leading to the heterogeneity observed in mouse is not understood.

In this paper, we present the complete sequence of the mouse opsin gene, including introns and flanking sequences (9.5 kb). We show that the transcript heterogeneity is not generated by alternative splicing or by the use of multiple initiation sites. The gene sequence contains several putative polyadenylation sites resembling the AATAAA consensus site (23) in the 3'-flanking region. We provide evidence, by PCR amplification of 3'-flanking sequences of opsin cDNA, that the transcripts observed are generated by the selective use of these polyadenylation sites. Moreover, we show that the pattern of multiple transcripts is not unique to the mouse, but is also present in the rat and human (three transcripts), and frog retina (two transcripts).

MATERIALS AND METHODS

RNA—Retinas of various vertebrate species (adult normal C57BL/6j mice, rats (Long Evans), Irish setter dogs, humans (a generous

*This investigation was supported in part by National Eye Institute Grants EY08123 (to W. B.) and EY06085 and EY06680 (to J. F. M.); Individual National Research Service Award F32 EY06172-01 (to S. J. P.); grants (to W. B.) from the National Retinitis Pigmentosa Foundation/Fighting Blindness, Texas Advanced Technology Program, the Retina Research Foundation (Houston); a National Institutes of Health Biomedical Research Support Grant (to W. B.); and a grant from the Knights Templar Foundation (to M. R. A.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

†Recipient of a Jules and Doris Stein Research to Prevent Blindness professorship. To whom correspondence should be addressed: Dept. of Ophthalmology, Baylor College of Medicine, 6501 Fannin St., Houston, TX 77030. Tel.: 713-798-5965.
Mouse Opsin Gene

Gift of the Lions' Eye Bank, Houston, TX), and frog (Rana pipiens) were excised from enucleated eyes and immediately dropped into liquid nitrogen. Total RNA was isolated according to Chirgwin et al. (24), and poly(A) mRNA was isolated by two sequential passages through oligo(dT) columns (25). In some cases, poly(A) mRNA was also isolated according to the Fastrack procedure (Invitrogen (26)). No apparent differences in mRNAs from these procedures were observed. For Northern blotting, RNA was denatured in 0.5 M NaOH at 65°C for 5 min and neutralized with an equal volume of 1:1 M ammonium acetate. After precipitation with ethanol, RNA samples were dissolved in 300 μl of water, quantitated by spectrophotometry at 260 nm, and diluted to 100 ng/μl. End labeling of oligonucleotides was performed using [γ-32P]ATP (model 391A, Applied Biosystems, Inc.) and deprotected in 14 M HCl, 20% hydroxylamine at 60°C for 10 min. The labeled oligonucleotides were purified by agarose gel electrophoresis and stored at -20°C.

**Complementary DNA**—The three cDNA clones M2-M4 (Fig. 1) were isolated as described previously (16, 29). Exon-specific fragments were isolated by digesting the 1.1-kb M2 insert with various restriction endonucleases according to the following strategy. M2 insert digested with XhoI yielded a 270-bp 5' fragment which is exon 1-specific (Fig. 1B, x1). The 848-bp 3' fragment was digested with SauI to yield a 509-bp fragment. The 339-bp XhoI/PvuII fragment was digested with RsaI. The resulting 424-bp fragment is exon 3-specific (x3); the larger 197-bp fragment was then digested with HaeIII to yield a 102-bp exon 2 probe (x2). The 509-bp PvuII/XhoI fragment was digested with SacI, yielding two fragments that were digested with MnlI. The resulting 194-bp fragment is exon 4-specific (x4). The 140-bp fragment is exon 5-specific (Fig. 1B, x5). All exon-specific fragments were shown not to cross-hybridize to each other by a standard dot blot procedure and to hybridize to the original M2 cDNA (results not shown).

**Genomic DNA**—A 5-kb genomic EcoRI fragment (Fig. 1, MOPS) was isolated as described previously (16). Genomic clones with approximately 15-kb inserts (M01 and M02) were isolated from a bacteriophage library which was prepared by cloning partial M01 digests into the BamHI site of EMBL3 (Clontech). PsI and XhoI subclones of MOPS1 and EcoRI/SalI, XhoI, and BamHI subclones representing flanking sequences of the opsin gene were generated using standard procedures (30) with pUC or Bluescript vectors. Subclones were sequenced using the double-stranded plasmid sequencing technique (31), M13 universal and sequence-specific primers, and Sequenase (U. S. Biochemical Corp.). Sequencing reactions followed the Sequenase protocol except that 100 ng of primer were used for each set, and incubations for annealing were omitted. The coding region and introns of opsin were sequenced on both strands (the untranslated regions in only one direction), using dITP, dCTP, dGTP, and dATP.

**Polymerase Chain Reaction**—To map the polyadenylation sites in the 3' end of the opsin gene, the polymerase chain reaction (PCR) was performed using the primer XhoI (MOPS, 5' TACCTAGCTAGACCCGCCGATCTAGA-3', 26) and primer M2 (MOPS, 5' GCTGTTGATGATGTTTGG-3'), which hybridizes outside the coding region, including the four introns but only short flanking sequences. PCR amplification. The cDNA regions to be amplified were delimited by a primer (5'-CCGATCGCGGCCGATCTAGA, mcs23) which contained the first 23 nucleotides of Tp1-mcs, and by opsin-specific primers (W9-W69, Figs. 2 and 4). Three cycles of amplification were performed in a thermal cycler (Perkin-Elmer Cetus Instruments) at 94/72/55°C (denaturation, 1 min; extension, 2 min; annealing, 2 min; final extension, 10 min) with 2-5 μl of Taq polymerase (Perkin-Elmer Cetus Instruments) and 50 μM each of dATP, dCTP, dGTP, and dTTP. PCR reactions were centrifuged in a thermal cycler (Ericomp) at 94/72/55°C (denaturation, extension, and annealing) for 30 sec, 1 min, and 1 min, respectively, for each set of primers. The length of the amplified DNA fragments was determined after electrophoresis on 1% agarose gels, Southern blotting, and hybridization with opsin-specific probes. In some cases, fragments of interest were purified by agarose gel electrophoresis and sequenced directly.

**RESULTS**

Characterization of Mouse Opsin Clones—We previously described the isolation of a 5-kb genomic clone MOPS1 (16), which was shown to contain the complete opsin-coding sequence, including the four introns but only short flanking regions. In order to characterize larger genomic fragments containing several kilobases of 5'- and 3'-flanking sequences, we used MOPS1 to isolate two overlapping genomic EMBL3 clones (λM01 and λM02). λM01 is the parent clone for all genomic subclones (except MOPS1) described in Fig. 1A. It contains approximately 6 kb of sequence upstream from the translation start codon (ATG) and 3.5 kb of the 3'-untranslated region. The 1.3-kb XhoI fragment XX1, whose 3' XhoI site is located just 10 bp upstream from ATG, contains putative CCAAT and TATAA boxes (Fig. 2) and, presumably, other upstream transcription regulatory sequences. The 3'-flanking fragment ES4 contains several polyadenylation sites (see below). The cDNA clones M2-M4 (Fig. 1B) were described previously (16).

Structure and Sequence of the Mouse Opsin Gene—A complete composite sequence (9.5 kb) of the mouse opsin gene, starting with the 5' XhoI site of XX1 and ending at the 3' end of the opsin gene and its mRNA. A restriction map and extent of the mouse opsin genomic fragment λ M01 (15 kb). The vector is indicated by horizontal bars, Single letters symbolize restriction sites that were used for sequencing and subcloning (S, Sall; A, EcoRI; X, XhoI; A, XhoI; P, PstI). Lines underneath the restriction map represent the extent of genomic clones between two sites. Except for MOPS1, genomic clones are named using the single letter symbols of their flanking restriction sites, followed by a number to designate the approximate length of the fragment. MOPS1 is a schematic representation of the opsin gene. Boxes indicate the extent of exons and lines connecting the exons demarcate the length of introns. B, schematic representation of opsin cDNA. The filled box represents the coding sequence, and open boxes are the 5'- and 3'-untranslated regions. The actual 3' end of the structural gene is upstream (not shown). The extent of cDNA clones M3, M4, and M2 is indicated below. A restriction map of M2 with sites used for excision of exon-specific fragments is shown at the bottom (M, MnlI; X, XhoI; t, PvuII; H, HaeIII; R, RsaI; C, SacI). The positions of introns in M2 are marked with arrowheads. x1-25, exon-specific probes.

![Fig. 1. Map of the mouse opsin gene and its mRNA.](http://www.jbc.org/content/263/9/20564)
FIG. 2. Complete sequence of the mouse opsin gene. The composite sequence of 9466 nucleotides is numbered in 300-bp intervals on the right, position 1 is the transcription start point (tsp). Sequences upstream of this point and introns are in lowercase letters, and exon sequences are in capital letters. TATAA and CCAATT boxes preceding the transcription start point and polyadenylation signals downstream of the translation stop codon are boldface. EcoRI, XbaI, and XhoI sites used for subcloning are italic and boldface. Abbreviations for oligonucleotides (W9-W60) used for the polymerase chain reaction, in Northern blots, and for DNA sequencing are shown in the left margin, and their sequences are underlined. Arrows to the left or right indicate antisense or sense direction. Also shown in the left margin are positions of polyadenylation sites (Al-A9), exon designations framed by rectangles, and the CA repeat in intron 1. The predicted amino acid sequence for mouse opsin is shown in single letter symbols underneath the exon sequences.
end of ES4, is shown in Fig. 2. Exon/intron junctions were
determined by identifying acceptor/donor consensus se-
cquences (35) within the gene sequence and by comparison
with the cDNA sequences of M2–M4 (16). The results show
that the mouse opsin gene is split into five exons and four
introns, an organization characteristic of other mammalian
opsin genes, and that the position of introns has been precisely
conserved. Corresponding exons in mouse, human, and bovine
opsin genes are highly homologous and, except for exon 5
containing 3‘-untranslated regions, are of identical length.

The lengths of the introns in base pairs are 1487 (1782), 1072
(1205), 117 (116), and 971 (833) (the lengths of human introns
(13) are shown in parentheses). A dot matrix comparison (not
shown) reveals that there is no sequence similarity between
corresponding human and mouse introns immediately before
and after intron/exon junctions. Islands of similar sequences,
containing 3’untranslated regions, are of identical length.

The mouse opsin gene is split into five exons and four
introns, an organization characteristic of other mammalian
species of opsin mRNA is unique to the mouse, we analyzed
whether each opsin mRNA species contains each of the five
exons, a set of exon-specific probes, x1-x5, was generated by
extension of a synthetic oligonucleotide (W40, exon 2) using mouse retina poly(A) mRNA as a template. A
single extension product was identified (Fig. 4), indicating
absence of differential splicing suggests that the heterogene-
ity in mouse opsin RNA may originate either from multiple
transcription start points or from utilization of multiple
polyadenylation signals. The transcription start point was
determined by extension of a synthetic oligonucleotide (W40,
FIG. 4. Analysis of the transcription start point by primer extension. An end-labeled synthetic primer (W40, see Fig. 2) was extended with reverse transcriptase using mouse retina poly(A) mRNA as a template (lanes 1 and 2; lane 1 contains 3 times more extension product than lane 2). For size verification, a sequencing ladder generated with W40 and denatured MOPS1 DNA (Fig. 1) as template was used (GATC). Only the sequence flanking the transcription start site (arrow) is shown.

Gene Utilizing PCR—To map the location of polyadenylation signals in the 3' end of the gene, we applied a procedure based on amplification of cDNA copies of specific polyadenylated mRNAs by the polymerase chain reaction. The principle of this procedure is to first produce a library of retina cDNAs with an oligo(dT) primer (Fig. 5A, dT15-mcs). This reverse transcription primer introduces a nucleotide sequence (mcs) that can be used in a second amplification step to specifically amplify DNA between a site near the C-terminal end of opsin and this site (Fig. 5A, mcs23 primer). The size of the amplified DNA allows identification of polyadenylation signals used in the gene sequence (Fig. 5, B and C). The point of poly(A) addition may then be precisely determined by direct sequencing of the purified amplified DNA. Since the exon-specific primer W9 and the last putative polyadenylation site (A9) are too far apart (approximately 3.4 kb) to be amplified by the polymerase chain reaction (39), amplifications were carried out in several steps (Fig. 5). In each step, a distinct opsin-specific primer (W9-W60) was designed to yield short fragments whose sizes can be precisely determined after electrophoresis and blotting. All amplified fragments visible after ethidium bromide staining were shown by Southern blotting to be derived from the opsin cDNA template (Fig. 5B).

The results (Fig. 5, B and C) show that the polymerase chain reaction performed with mcs23 and opsin gene-specific primers on a cDNA template yields a series of fragments, consistent with the presence of multiple polyadenylation signals in the 3' end of the opsin gene. In a control experiment, opsin-specific primers were replaced by primers specific for the rod cGMP phosphodiesterase α subunit whose mRNA occurs as a single predominant species in mouse (40). Only one major amplified fragment was detected (Fig. 5B, lanes 7 and 8). The various extension products shown in Fig. 5B are summarized in Fig. 5C. The 660-, 980-, 1800-, and 2300-bp fragments generated with W9/mcs23 (Fig. 5B, lane 1) indicate that the polyadenylation signals A1, A2, A4, and A5 are used to generate the 1.75-, 2.05-, 2.8-, and 3.5-kb mRNA species shown on the Northern blot (Fig. 3). The A1 signal, but no poly(A) tail, was present in two cDNA clones (M1 and M3) described previously (16), indicating that mRNAs exist that extend beyond the A1 signal. The A2 polyadenylation site and the precise base at which poly(A) was attached were determined by direct sequencing of a 280-bp fragment (marked by an asterisk in Fig. 5B) generated by amplification with W34 and

FIG. 5. Mapping the 3' end of the opsin gene. A, schematic diagram of first-strand cDNA synthesis using T7-rnas primer (see “Materials and Methods”) and PCR amplification of cDNA with opsin-specific primers (W9-W60) and the multiple cloning site primer mcs23 (see “Materials and Methods”). B, Southern blot of PCR-amplified DNA. Lanes 1–7 were hybridized with nick-translated ES4 (Fig. 1); lanes 8–9 were hybridized with a mouse cGMP phosphodiesterase α subunit probe (W. Baehe, unpublished data). Primer pairs used for amplification were as follows: lane 1, W9/mcs23; lane 2, W34/mcs23; lane 3, W36/mcs23; lane 4, W41/mcs23; lane 5, W49/mcs23; lane 6, W60/mcs23; lane 7, W9/W39 (positive control with two opsin-specific primers); lane 8, SP66/mcs23 (SP66 is a primer specific for the α subunit of rod cGMP phosphodiesterase, positions 2185–2205 in the bovine sequence in Ref. 52); and lane 9, PCR40/mcs23 (PCR40 is a phosphodiesterase primer corresponding to positions 2030–2070 in the bovine sequence in Ref. 52). SP66 and PCR40 correctly amplify both mouse and bovine phosphodiesterase cDNA. All major fragments shown in lanes 1–9 are also visible on ethidium bromide-stained gels. The 280-bp fragment marked with an asterisk (lane 2, also marked in C) was directly sequenced with the amplification primers (see text). C, schematic representation of opsin mRNA with poly(A) signals. Shaded box, the opsin-coding sequence. A1–A9, putative consensus polyadenylation signals identified on the gene sequence (Fig. 2). Signals used to generate the polyadenylated mRNA in Fig. 3 are boxed. The vertical line on the right of A9 identifies the 3' end of the sequence shown in Fig. 2. Underneath the mRNA are schematic representations of amplified DNA using the primer sets outlined in B. Filled rectangles represent opsin-specific primers; open rectangles denote mcs23. The control experiment W9/W39 (B, lane 9) is shown above the W9 amplification set. The shaded rectangle represents primer W39.
mcs23. The sequence obtained (not shown) indicated that the poly(A) attachment point is an A (Fig. 2, position 5549) which is preceded by an AT-rich area (TAATATAATTAATTAA) presumably serving as the polyadenylation signal (Fig. 2, boldface). Polyadenylation signals A5 and A9 are predicted to be used by four amplified fragments originating from four different primer sets (Fig. 5C). Polyadenylation signals A2 and A4 are predicted by at least two primer sets. Using the 3'-most primer W59 (Fig. 2, position 7815–7838) and mcs23, amplification products could still be identified that are longer than predicted from the last polyadenylation signal (Fig. 2, position 8070). This would indicate the presence of additional poly(A) sites downstream from the sequence shown in Fig. 2. No major transcripts of the required length, however, can be detected on Northern blots (Fig. 3).

Expression of the Opsin Gene during Development of the Retina—We next asked if there was a differential expression of any of the five opsin mRNA species during the development of the retina in normal C57BL/6J mice that would identify a possible function for the multiple mRNAs. Animals of selected ages (Fig. 6, postnatal day (PN) 1–50) were killed and the RNA was isolated from whole eyes (lenses removed) and hybridized on a Northern blot with labeled M2 cDNA (Fig. 1). Lanes showing PN 1–7 were exposed for 15 h; lanes showing PN 3–50 were exposed for 5 h with an intensifier screen.

DISCUSSION

Whereas the gene encoding opsin is abundantly expressed in mammalian rod photoreceptors, opsin or opsin-like anti-
at the poly(A) tail are amplified using the polymerase chain reaction. The results show that, of nine putative polyadenylation sites in the 3′-untranslated region of the mouse opsin gene, five are utilized to produce the major species seen on the Northern blot. Use of the most proximal site thus produces the shortest transcript, whereas the most distal yields the longest. Many genes contain internal AATAAA sequences that apparently do not function in 3′ processing, suggesting the requirement for additional signal sequences (23). A transcript heterogeneity similar to that in mouse is also observed in rat and, to a lesser extent, in human and frog (Fig. 4). No sequence information is available for the rat and frog opsin genes, and the 3′ end of the bovine gene has been only partially sequenced (12). The human sequence (13) has at least two polyadenylation sites that may lead to differential polyadenylation. Apart from multiple poly(A) sites, the organization of the mouse opsin gene is relatively uncomplicated. With a single transcription starting point, no alternative splicing, and multiple polyadenylation sites, the mouse opsin gene is similar to a group of genes (46) encoding very diverse protein products, among them murine dihydrofolate reductase (48), chicken vimentin (49), and human N-ras (50). The human N-ras gene has two and the dihydrofolate reductase gene 11 functional polyadenylation sites. While 3′-un-translational heterogeneity similar to that in mouse is also observed in rat and, to a lesser extent, in human and frog (Fig. 4). No sequence information is available for the rat and frog opsin genes, and the 3′ end of the bovine gene has been only partially sequenced (12). The human sequence (13) has at least two polyadenylation sites that may lead to differential polyadenylation. Apart from multiple poly(A) sites, the organization of the mouse opsin gene is relatively uncomplicated.

The recent discovery that a Pro to His mutation at position 23 near the N terminus of human opsin may be causative for one form of autosomal dominant retinitis pigmentosa (19), an inherited photoreceptor degeneration in the human retina, has renewed interest in the mouse opsin gene as a candidate gene to study retinal degeneration. Compared with the predicted human sequence (13), only 19 residues differ in mouse opsin (16). Pro is present in the mouse sequence (Fig. 2) and is conserved in all sequenced vertebrate and invertebrate opsins and in related G protein receptors (51). Apart from asparagine-linked glycosylation, a precise function for the N-terminal end of opsin or for Pro has not been determined. The detailed characterization of the mouse opsin gene and availability of cloned gene fragments as described in this study will allow the introduction of specific point mutations, the following of the consequences of expression of the mutated opsin gene in a transgenic mouse, and, possibly, the facilitation of the establishment of an animal model which mimics one autosomal dominant form of human retinitis pigmentosa.

Acknowledgments—We thank Drs. Gregory I. Liou and Richard Hurwitz for critically reading this paper. We are grateful to Dr. Kuyas for generating the exon-specific cDNA probes and to Nancy Baumann for help with the manuscript.

REFERENCES
Mouse opsin. Gene structure and molecular basis of multiple transcripts.
M R al-Ubaidi, S J Pittler, M S Champagne, J T Triantafyllos, J F McGinnis and W Baehr


Access the most updated version of this article at http://www.jbc.org/content/265/33/20563

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

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

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/33/20563.full.html#ref-list-1