Sequence, genomic structure and tissue expression of carp (Cyprinus carpio L.) vertebrate ancient (VA) opsin

Paraskevi Moutsaki, James Bellingham, Bobby G. Soni, Zoë K. David-Gray, Russell G. Foster*

Department of Integrative and Molecular Neuroscience, Division of Neuroscience and Psychological Medicine, Imperial College School of Medicine, Charing Cross Hospital, Fulham Palace Road, London W6 8RF, UK

Received 28 February 2000; received in revised form 12 April 2000

Edited by Takashi Gojobori

Abstract We report the isolation and characterisation of a novel opsin cDNA from the retina and pineal of the common carp (Cyprinus carpio L.). When a comparison of the amino acid sequences of salmon vertebrate ancient opsin (sVA) and the novel carp opsin are made, and the carboxyl terminus is omitted, the level of identity between these two opsins is 81% and represents the second example of the VA opsin family. We have therefore termed this C. carpio opsin as carp VA opsin (cVA opsin). We show that members of the VA opsin family may exist in two variants or isoforms based upon the length of the carboxyl terminus and propose that the mechanism of production of the short VA opsin isoform is alternative splicing of intron 4 of the VA opsin gene. The VA opsin gene consists of five exons, with intron 2 significantly shifted in a 3′ direction relative to the corresponding intron in rod and cone opsins. The position (or lack) of intron 2 appears to be a diagnostic feature which separates the image forming rod and cone opsins from the more recently discovered non-visual opsin families (pin-opsins (P), vertebrate ancient (VA), parapinopsin (PP)). Finally, we suggest that lamprey P opsin should be reassigned to the VA opsin family based upon its level of amino acid identity, genomic structure with respect to the position of intron 2 and nucleotide phylogeny.

© 2000 Federation of European Biochemical Societies.

Key words: Retina; Pineal; Vertebrate ancient opsin; Photopigment; Phylogeny; Teleost fish

1. Introduction

Based upon cDNA sequences, the well-characterised opsin families of the vertebrates appear to have emerged in the following phylogenetic sequence: long wavelength cone opsins (L), pin-opsins (P), violet cone opsins (S), blue cone opsins (M1), green cone opsins (M2) and finally rod opsins (Rh) [1]. The genomic structure of S [2], M1 [3], M2 [4] and Rh [5] opsins are identical; they all contain four introns at equivalent coding positions (designated in this paper as 1, 2, 3 and 4). The genomic structure of P and L opsins, however, are slightly different. In the P opsins the second intron is displaced 14 nucleotides in the 3′ direction [6], whilst the L opsins contain an additional intron site [2], designated as 0 in this paper. In addition, the recently discovered parapinop-
A 363 bp VA opsin cDNA fragment was amplified using degenerate primers VM1, 5'-TAYTTYTGAHTHTGYGICCTTGGT-3' (corresponding to the following amino acid sequence of sVA opsin, Tyr-138 to Gly-146), and VM2, 5'-IGCTCATRTATGCTACCACTTACCAGCTATT-3' (complementary to amino acid sequence of sVA opsin, Ala-258 to Met-251). Polymerase chain reaction (PCR) amplification was performed using BioTaq polymerase (Bioline) in the manufacturers NH4 buffer at 3.0 mM MgCl2. The following thermal profile was used: an initial denaturation step of 3 min at 94°C, followed by 35 cycles of 1 min at 94°C, 1.5 min at 53°C and 1.5 min at 72°C.

The 5' region of the cDNA was recovered using the following gene specific primers in conjunction with the 5' RACE System (Life Technologies): VM3, 5'-CTCGATGATGCATCTGG-3' (primary RCR), and VM4, 5'-CTACACCCAGCAGCGAGCTG-3' (secondary PCR). The 5' end of the cDNA was obtained using the following gene specific primers in conjunction with the 5' RACE System (Life Technologies): VM5, 5'-GTCCAGAGGAGAGGAA-3' (cDNA synthesis), VM6, 5'-CAGACCCAGAGCTGACTGCTT-3' (primary PCR), and VM7-CTTCTGCAGCCAGCAGGT-3' (secondary PCR).

All PCR products were ligated into pGEM-T Easy Vector (Promega) and transformed into DH5a Sub-Cloning Efficiency Competent Cells (Life Technologies). Nucleotide sequence determination was carried out on an ABI PRISM 377 DNA Sequencer (Perkin Elmer) using the ABI PRISM BigDye Terminator Cycle Sequencing Kit (Perkin Elmer). The cloned products from three independent PCR amplifications were sequenced on both DNA strands.

The full length coding region of VA opsin was PCR amplified from retinal cDNA using primers directed against the 5' and 3' untranslated regions of the gene: VM1, 5'-GATGCGAGCTCTTTCC-3' (primary PCR), VM2, 5'-CTTCCTCGCAGACGGATGT-3' (secondary PCR). The primer pairs designed to encompass the entire coding sequence. The following primer pairs were designed to encompass the putative intron insertion sites based upon the genomic structure of Rh opsin [5]: exon 1, VM8F, 5'-AGGGCTGGTTGGAGTGG-3' and VM8R, 5'-GATGACGAAGAACCGCTCGAA-3'; exon 2, VM11F, 5'-ACTCGTATGGTGGTTGTGAT-3' and VM11R, 5'-AT-CTCGGACATACTTTGTT-3'; exon 3, VM13F, 5'-TTCATCATCACGTT-3' and VM13R, 5'-TGAGCTTGCG-3'; exon 4, VM15F, 5'-TGTAGGGCCGATGTGGG-3' and VM15R, 5'-CATCGGACATACTTTGTT-3'; exon 5, VM16F, 5'-CTTCATCATCACGTT-3' and VM16R, 5'-ACTCGTATGGTGGTTGTGAT-3'; exon 6, VM17F, 5'-ATGGAGTCGTTGGCTGCG-3' and VM17R, 5'-GATGCACGCAGACTTGGAGG-3'; exon 7, VM18F, 5'-CTC-3' and VM18R, 5'-GATGCACGCAGACTTGGAGG-3'.

3. Results and discussion
3.1. Characterisation of VA opsin from the common carp (C. carpio)

Using a combination of degenerate primer PCR and 3' RACE PCR techniques, we isolated the complete coding sequence of an opsin-like cDNA from the retina of the common carp, C. carpio. The cDNA possesses a 1149 bp open reading frame predicting a 382 amino acid protein (Fig. 1).

The complete nucleotide sequence and the predicted reading frame predicting a 382 amino acid protein (Fig. 1).

Table 1: Comparison of amino acid positions in Rh opsin implicated in the binding or activation of Gt, with those of cVA opsin

<table>
<thead>
<tr>
<th>Opsin domain</th>
<th>Amino acid position in bovine Rh opsin</th>
<th>Substitution effect</th>
<th>Equivalent position in cVA opsin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second cytoplasmic loop</td>
<td>E134, R135</td>
<td>no Gt binding</td>
<td>E127, R128</td>
</tr>
<tr>
<td>Third cytoplasmic loop</td>
<td>L226, T243</td>
<td>no Gt activation</td>
<td>L217, A234</td>
</tr>
<tr>
<td>Transmembrane domain VII and cytoplasmic tail</td>
<td>Y306</td>
<td>reduced Gt activation</td>
<td>K235</td>
</tr>
</tbody>
</table>

The biochemical effect observed in Rh opsin when substitutions are made at these positions is noted.

Fig. 1. Nucleotide sequence and conceptual translation of cVA opsin cDNA (GenBank: AF233520). The nucleotide coding sequence is presented in upper case and the 5' and 3' untranslated regions in lower case, with the predicted polyadenylate signal shown as the single letter code below the nucleotide sequence. The translation stop-codon is designated by the asterisk and the polyadenylation signal is underlined. The seven transmembrane α-helices as predicted by the model of Baldwin et al. [32] are shaded.
amino acid translation opsin share 60 and 66% identity, respectively, with the VA opsin isolated from the Atlantic salmon, S. salar (sVA) [8]. Significantly, the carboxyl terminus (the cytoplasmic tail region after the seventh transmembrane domain) of the C. carpio opsin is 79 amino acid residues in length, whereas the corresponding region in sVA opsin is 13 residues long (see Fig. 2). When a comparison of the C. carpio opsin and sVA opsin amino acid sequences are made, and the carboxyl terminus is omitted, the level of identity between these two opsins increases to 81%. We have therefore termed this C. carpio opsin as cVA opsin. It seems, therefore, that the members of the VA opsin family exist in two variants, a long and a short carboxyl terminus, cVA opsin being an example of the long form and sVA opsin of the short form.

3.2. Key features in the amino acid sequence of VA opsin

The predicted cVA opsin amino acid sequence retains, like that of sVA opsin, the conserved features required for opsin function—retinal attachment site (Lys287) located in transmembrane domain VII; the Schiff base counterion (Glu106) in transmembrane domain III; cysteine residues at positions 103 and 180 required for conformational stability [8]. However, cVA opsin contains one probable palmitoylation site (Cys-314) in its carboxyl terminus whereas sVA opsin has no palmitocysteines (Fig. 2). The palmitocysteines in the rod and cone opsins are believed to be used as the site where the C-terminal tail anchors to the plasma membrane creating a fourth cytoplasmic loop [15,16]. The absence of palmitoylation sites in sVA opsin might relate to the greatly reduced length of the carboxyl terminus, which does not need ‘anchoring’ and presumably floats freely in the cytoplasm. Removal of palmitoylation sites in rod and cone opsins appears not to disrupt the activation of the G-protein, transducin (Gt) [16,17]. However, differences in the length of the carboxyl terminus of sVA and cVA may have an effect on the dynamics of the phototransduction cascade. Phosphorylation of serines and threonines in the carboxyl terminal of rhodopsin (by rhodopsin kinase) is required for the termination of the G-protein mediated cascade [18,19]. The long form of VA opsin in carp has an increased number of serines and threonines (13 in total), whereas only three are found in the short sVA opsin (Fig. 2). Thus, the increased number of phosphorylation sites in the long form (cVA) may allow a rapid deactivation of the G-protein mediated cascade, whereas the reduced number of serines and threonines in the short form (sVA) may allow an extended activation of the cascade.

Site directed mutagenesis studies have identified three specific domains of the cytoplasmic surface of Rh opsin that are known to be important in the binding and activation of Gt. These domains and the amino acids associated with binding

<table>
<thead>
<tr>
<th>Exon/intron splice junctions for cVA opsin gene</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intron</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

Upper case letters represent coding sequence, lower case letters represent intron sequence. Numbers above sequence indicate the coding nucleotides that surround the intron insertion site.
and/or activation are reviewed in Table 1. Although some of these key amino acid positions are conserved in cVA opsin, several are not. This suggests that VA opsin may interact with a unique form of G-protein. In common with other non-visual opsins cVA opsin lacks a doublet of amino acids in the second extra-cellular loop (Fig. 3) that are present in the image forming (rod and cone) opsins [20].

3.3. Tissue expression of cVA opsin

To determine whether cVA opsin is expressed in the pineal gland of *C. carpio*, we used gene specific primers located at the start and the end of the coding sequence of cVA opsin (VM8F and VM19R) to amplify the entire coding sequence from pineal cDNA. The cDNA that we isolated from the pineal is 99.8% identical to that isolated from the retina. The nucleotide substitutions that are present are silent in nature and may occur due to the tetraploid nature of the *C. carpio* genome [21]. A similar phenomenon has been observed in *C. carpio*, where two Rh opsin alleles have been isolated from the retina which are 97.2% identical [22,23]. The expression of cVA opsin in both the retina and pineal parallels our finding with sVA opsin [10].

3.4. Genomic organisation of cVA and sVA opsin

Using PCR primers designed to amplify introns based on the genomic structure of other opsin classes [2,5–7,20], we determined that the cVA opsin gene is approximately 8.2 kb in length. The gene consists of five exons and all the introns conform to the splice site consensus [24] (see Table 2). The first, third and fourth intron sites are at equivalent positions to all of the other opsin class. The second intron, however, is shifted 42 nucleotides in the 3' direction in comparison with the Rh-like and L opsins, and 14 nucleotides in comparison to P opsins. Although the second intron is absent in PP opsin [7], it is apparent that the position (or absence) of intron 2 clearly separates the M1, M2, Rh, S and L opsins (traditionally associated with the rod and cone photoreception) from P, PP and VA opsins (which are not expressed in rod or cone photoreceptors) (see Fig. 4). Whether intron 2 occurs at these different sites as a result of intron slippage of a single ancestral intron, or independent intron gain/loss in each of the lineages [25] remains unclear.

We also determined the genomic structure of sVA opsin and introns 1, 2 and 3 are inserted at similar locations to their cVA opsin counterparts (data not shown), but intron 4 appears to be absent in the sVA opsin gene. Comparison of the sVA opsin sequence with that of cVA opsin indicates that the
predicted insertion site of intron 4 occurs in the motif \(\ldots MNKQ\)-intron-FR\(\ldots\). This motif is highly conserved and present in all opsin classes at the junction of the seventh transmembrane domain and the carboxyl terminus, but is disrupted in sVA opsin occurring as MNKQVS. Inspection of the cDNA sequence of sVA opsin reveals that the disruption of the MNKQFR motif occurs at the predicted intron insertion site present in cVA opsin (see Fig. 5A). Interestingly, the nucleotide sequence of the sVA opsin carboxyl terminus and 3' UTR shares approximately 44% identity with the initial sequence of cVA opsin intron 4 (see Fig. 5B). Analysis of the sequence at the 5' region of cVA opsin intron 4 reveals two potential polyadenylation signals that may lead to an alternatively spliced transcript [26], forcing a read-through at translation that will produce an open reading frame that contains an MNKQVT motif followed by a potential carboxyl terminus of 38 amino acids (see Fig. 5C). Significantly, the NKQ tripeptide remains intact in sVA opsin, and this is likely...
to be important, as it has recently been shown that disruption of this tripeptide diminishes Gt activation in bovine Rh opsins [17].

3.5. The relationship of VA opsin and lamprey P opsin

We compared the deduced amino acid sequences and the genomic structure of both cVA opsin and an opsin isolated from the marine lamprey, P. marinus, designated lamprey P opsin [20]. This designation, however, may be misleading as lamprey P opsin shows greater similarity to the VA opsins than to the P opsins. Both the VA opsins and lamprey P opsin retain all four introns and share a 42 bp shift in their second intron site. This feature alone distinguished these opsins from all other opsin families. Furthermore, a comparison of the deduced amino acid sequences between chicken (Gallus gallus) [27], toad (Bufo japonicus) [28] and American chameleon (Anolis carolinensis) [29] P opsins shows an amino acid identity of >80%. Lamprey P opsin, however, shares only 46-48% identity with these P opsins. By contrast, lamprey P opsin shares 61 and 65% identity with sVA and cVA opsin, respectively. Finally, the size of lamprey P opsin carboxyl terminus (124 amino acids) is more comparable to that of the cVA opsin (79 amino acids) than the rest of the P opsins (43 amino acids in pigeon P opsin [30] and 56 amino acids in chameleon P opsin [29]). One parsimonious explanation for these data, especially in light of the evolutionary position of lampreys, is that lamprey P opsin is the evolutionary precursor of the VA opsin family. The hypothesis is further supported by the phylogenetic placement (based upon nucleotide coding sequences) of lamprey P opsin, sVA opsin and sVA opsin within the same clade (Fig. 4).

VA opsin has been isolated from both a marine fish (Atlantic salmon) and a fresh water fish (carp). Furthermore, Southern blot analysis was performed on the DNA from a deep-sea fish (bristlemouth) and zebrafish, and potential VA opsin orthologues were found in both species (data not shown). It also appears that lampreys have an orthologue of VA opsin (see discussion above). These findings suggest that VA opsin is widely distributed amongst the teleosts and agnath. By contrast, our preliminary high stringency Southern analysis failed to find potential VA orthologues in birds and several species of mammal (data not shown). We are currently extending this analysis to include amphibian and reptilian species. It is interesting that P opsin orthologues have been reported in birds [27], reptiles [29] and amphibia [28], but have not been found in teleost fish [29]. If VA opsins and P opsins were performing some common photosensory task then their non-overlapping taxonomic distribution might be expected. Further physiological analysis will be required to clarify this possibility.

4. Note added in proof

Since this work was accepted for publication we have learnt that similar work carried out on the zebrafish (Danio rerio) has produced two VA opsin isoforms that have homologies to the short carboxyl terminus sVA opsin (termed zVA opsin) and the long carboxyl terminus cVA opsin (termed zVAL opsin) [31]. The carboxyl terminus of zVAL opsin is 74 amino acids in length and shares 73.5% identity with the 79 amino acid carboxyl terminus of cVA opsin, whereas the carboxyl terminus of zVA opsin is 7 amino acids long and is comparable to the 13 amino acid carboxyl terminus of sVA opsin. The authors of this paper suggest a similar alternative splicing mechanism to the one proposed in this paper for the production of the two zebrafish VA opsin isoforms, and we anticipate that the 5’ region of intron 4 of the zVAL gene will contain the nucleotide sequence that encodes the short carboxyl terminus of zVA opsin and its associated 3’ UTR and polyadenylation mechanism.

Acknowledgements: This publication is the result of research sponsored by the Biotechnology and Biological Sciences Research Council (BBSRC), UK. Our appreciation also goes to the anonymous reviewer who generously supplied us with a pre-print of Ref. [31].

References