FEBS Letters 473 (2000) 316–322 FEBS 23682

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 opsin families 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.

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 parapinopsin (PP) shows a yet further difference in its genomic structure lacking the second intron of the Rh, M1, M2 and S opsins [7]. The significance of these differences in the genomic structure is unknown. They may represent changes that were introduced when P, L and PP opsins were duplicated from an ancestral gene. Alternatively, they may represent post-duplication events specific to P, L and PP opsin families. Either way, these differences in genomic structure appear to serve as additional features to differentiate the opsin families of the vertebrates.

We have described previously the primary structure of a novel ocular cDNA isolated from the eye of Atlantic salmon (*Salmo salar*) which was the single member of a new opsin family named the vertebrate ancient (VA) opsin family [8]. Salmon VA (sVA) opsin forms a functional photopigment with a maximal absorbance of 451 nm and is expressed in a subset of retinal cells with the location and morphology characteristic of horizontal and amacrine cells [9]. Additionally sVA opsin is also expressed within cells of the pineal gland and the sub-habenular region of the brain [10]. These diverse sites of VA opsin expression, taken together with several unique features present in its amino acid sequence, suggested that VA opsin is a novel photopigment specialised for non-image forming light detection [8].

To date, the only available VA opsin sequence is that described for ocular sVA opsin and no information currently exists on the genomic structure of this gene. Furthermore, preliminary data suggests that VA opsin may exist in two variants with either a short or a long carboxyl terminus [11,12]. We have therefore undertaken a detailed study of another member of the VA opsin family from the common carp (*Cyprinus carpio* L.), a species that is widely used in teleost retinal electrophysiology. In this paper we report and consider: (i) the isolation and characterisation of carp VA opsin (cVA opsin) from both retinal and pineal tissues; (ii) the finding that VA opsin is present in two carboxyl terminus variants in teleosts; (iii) the genomic structure of VA opsin in both carp and salmon and, finally, (iv) we provide evidence that the previously described genomic sequence isolated from lamprey (*Petromyzon marinus*) and termed Lamprey P opsin is, in fact, a member of the VA opsin family.

2. Materials and methods

2.1. Isolation of cVA opsin cDNA

Poly(A)+mRNA was isolated from retinae and pineal glands of the common carp (*C. carpio* L.) using the QuickPrep Micro mRNA Kit (Pharmacia). Retinal and pineal cDNA was synthesised using the 3′ RACE System (Life Technologies).

*Corresponding author.
E-mail: r.foster@ic.ac.uk

PII: S0014-5793(00)01550-7

A 363 bp VA opsin cDNA fragment was amplified using degenerate primers VM1, 5′-TAYTTYGTIATHTGYCGICCICTIGG-3′ (corresponding to the following amino acid sequence of sVA opsin, Tyr-138 to Gly-146), and VM2, 5′-IGCCATRTACATSACSACSACCAT-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 NH_4 buffer at 3.0 mM $MgCl_2$. 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 3′ region of the cDNA was recovered using the following gene specific primers in conjunction with the 3′ RACE System (Life Technologies): VM3, 5′-CTCGCTGTGATCATCGTCT-3′ (primary PCR), and VM4, 5′-TCTAACACGCACGGCAGGCT-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′-GTCCAGACGAAGGAGAA-3′ (cDNA synthesis), VM6, 5′-CAGACCCAGAGCTGCATGCTT-3′ (primary PCR), and VM7 5′-CTTTCCTCGCAGACGGATGT-3′ (secondary PCR).

All PCR products were ligated into pGEM-T-Easy Vector (Promega) and transformed into DH5α 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: VMVA1, 5′-GATGCACGCAGACTTCTTCCA-3′ and VMVA2, 5′-CGCTGGATGAAGGTCATTTAG-3′

2.2. Determination of exon/intron structure

High molecular weight genomic DNA was isolated from *C. carpio* brain tissue using the soft tissue protocol of the Nucleon DNA Extraction Kit (Nucleon Biosciences, UK). PCR amplification was performed on 50 ng of genomic DNA as described previously using 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′-ATGGAGTCGTTGGCTGCG-3′ and exon 2, VM11R, 5′-GATGACGAAGAACCGCTCGAA-3′; exon 2, VM12F, 5′-ATCGTGGCGCTGTGG-3′ and exon 3, VM13R, 5′-TGAGCTTGCGGAGCAGTTTGC-3′; exon 3, VM14F, 5′-TTCATCATCACGTTCTTCACC-3′ and exon 4, VM15R, 5′-TGTAAACCGCTGCAGTTTTGG-3′; exon 4, VM16F, 5′-ACTCGTATGGTGGTTGTGATGATC-3′ and exon 5, VM19R, 5′-CATCGGACATACTTTGTTCTC-3′.

2.3. Phylogenetic analysis

Nucleotide and amino acid alignments were carried out using the ClustalX 1.8 suite of programs using default values ([13]; ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/). For phylogenetic purposes, the nucleotide alignments were adjusted using the SeaView program to maintain codon integrity prior to analysis in the Phylo_win package (both programs- [14]; http://pbil.univ-lyon1.fr/). Maximum parsimony and neighbour-joining trees were constructed with bootstrap confidence values based on 1000 replicates.

```
-179 bp
 aaaacccgtccagattcaagtagctaccgtcaaacaacgctcaacaaagtccacggacc
gcggcgcgcccgggttatgaagcgagaacaagcacgaaacatcacagaatagcctcgtct
ttgcgatgcacgcagacttcttccagatgaagcagccactgctcggtttggtttgaagcg

ATGGAGTCGTTGGCTGCGGCGGTGAACGGCGTGTCTCACACCGAGGACCCGTTCTCCGGA    60
 M  E  S  L  A  A  A  V  N  G  V  S  H  T  E  D  P  F  S  G   20
CCGCTCTCCTCCATCGCACCCTGGAACTACCGAGTCCTGGCGGCGCTGATGTTCGTGGTG   120
 P  L  S  S  I  A  P  W  N  Y  R  V  L  A  A  L  M  F  V  V   40
ACCTCGGTGTCTCTGTGCGAGAACTTCACCGTGATGCTCGTCACCTTCAGGTTCAAGCAG   180
 T  S  V  S  L  C  E  N  F  T  V  M  L  V  T  F  R  F  K  Q   60
CTCCGACAGCCGCTCAATTATATCATCGTCAACCTGTCTCTGGCGGACTTCCTCGTGTCT   240
 L  R  Q  P  L  N  Y  I  I  V  N  L  S  L  A  D  F  L  V  S   80
CTGAGCGGAGGAACCATCAGTTTTCTCACCAACTTCCACGGCTATTTCTTCTTGGGCAGA   300
 L  S  G  G  T  I  S  F  L  T  N  F  H  G  Y  F  F  L  G  R  100
TGGGCTTGTGTTCTGGAGGGATTCGCAGTCACCTTTTTTGGCATCGTGGCGCTGTGGTCT   360
 W  A  C  V  L  E  G  F  A  V  T  F  F  G  I  V  A  L  W  S  120
CTGGCCGTCCTGGCGTTCGAGCGGTTCTTCGTCATCTGTCGGCCGCTGGGAAACATCCGT   420
 L  A  V  L  A  F  E  R  F  F  V  I  C  R  P  L  G  N  I  R  140
CTGCGAGGAAAGCATGCAGCTCTGGGTCTGCTGTTCGTCTGGAGCTTCTCCTTCGTCTGG   480
 L  R  G  K  H  A  A  L  G  L  L  F  V  W  S  F  S  F  V  W  160
ACCATTCCTCCCGTCCTGGGCTGGAGCAGCTACACCGTCAGCAAGATCGGCACCACCTGC   540
 T  I  P  P  V  L  G  W  S  S  Y  T  V  S  K  I  G  T  T  C  180
GAACCCAACTGGTATTCGGGAAACTTCCAGGATCACACCTTCATCATCACGTTCTTCACC   600
 E  P  N  W  Y  S  G  N  F  Q  D  H  T  F  I  I  T  F  F  T  200
ACCTGCTTCATCTTTCCTCTCGCTGTGATCATCGTCTGCTACTGCAAACTGCTCCGCAAG   660
 T  C  F  I  F  P  L  A  V  I  I  V  C  Y  C  K  L  L  R  K  220
CTCAGAAAGGTGTCTAACACGCACGGCAGGCTGGGTAATGCCAGGAAGCCGGAGCGTCAG   720
 L  R  K  V  S  N  T  H  G  R  L  G  N  A  R  K  P  E  R  Q  240
GTGACTCGTATGGTGGTTGTGATGATCGTGGCATTCATGGTGGCCTGGACGCCGTACGCT   780
 V  T  R  M  V  V  V  M  I  V  A  F  M  V  A  W  T  P  Y  A  260
GCGTTCTCCATCGTGGTCACGGCCCTTCCCAGCATCCACCTGGACCCCCGGCTGGCAGCC   840
 A  F  S  I  V  V  T  A  L  P  S  I  H  L  D  P  R  L  A  A  280
GCTCCGGCGTTCTTCTCCAAAACTGCAGCGGTTTACAACCCCATCATCTACGTGTTCATG   900
 A  P  A  F  F  S  K  T  A  A  V  Y  N  P  I  I  Y  V  F  M  300
AACAAACAGTTCAGGAAGTGTGTGGTTCAGCTCTTGAGCTGCAGAGACGTCACCGTGGTG   960
 N  K  Q  F  R  K  C  V  V  Q  L  L  S  C  R  D  V  T  V  V  320
GAGGGGAACATCAACCAGACCACCGACCGAGCAGGACTGACCAATGAGAGCAACACAGGA  1020
 E  G  N  I  N  Q  T  T  D  R  A  G  L  T  N  E  S  N  T  G  340
GAGATGTCCGCCATCGCTGCACGCATCCCAGCGGCTGGAACCGTCCCACCCAAAACAGAA  1080
 E  M  S  A  I  A  A  R  I  P  A  A  G  T  V  P  P  K  T  E  360
GAGCCTCCCAACGAACGCAGCTCCTTCACACAAATTCCCATTCCTGAGAACAAAGTATGT  1140
 E  P  P  N  E  R  S  S  F  T  Q  I  P  I  P  E  N  K  V  C  380
CCGATGtgaacacaccaacatctctgtttgtcactcgaactttagtctaaatgaccttca  1200
 P  M  *                                                      382
tccagcgttaaattctgcagcttcgtgcacgacagacagacaaagttaaaatcacaggat  1260
aaacacacttgagacgtggtgttggggacaaaacaaggctaaggggtgatatcttttgtt  1320
tattttaaaagcaacaagttttgtttggcagcagtttacaaatatggatgcaaaacgtta  1380
aaacttggtaaatcttgcaatgcaattatttaatttggttacagtttaaatggtaacact  1440
ttcatttagaagcctttatgtgtaatgcatttatacattttcattaaataaagttaagta  1500
agttc                                                          1505
```

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 polypeptide 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.

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′ and 5′ 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 coding sequence and the predicted

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

Opsin domain	Amino acid position in bovine Rh opsin	Substitution effect	Equivalent position in cVA opsin	Reference
Second cytoplasmic loop	E134, R135	no Gt binding	E127, R128	[34–37]
Third cytoplasmic loop	L226, T243	no Gt activation	L217, A234	[34,35,38]
	T229, V230, A233, A234, T242, Q244	reduced Gt activation	K220, L221, V224, S225, N233, R235	
Transmembrane domain VII and cytoplasmic tail	Y306	no Gt activation	Y297	[17,34,39]
	I307, M309, N310, K311, Q312, F313, M317	reduced Gt activation	V298, M300, N301, K302, Q303, F304, V308	

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

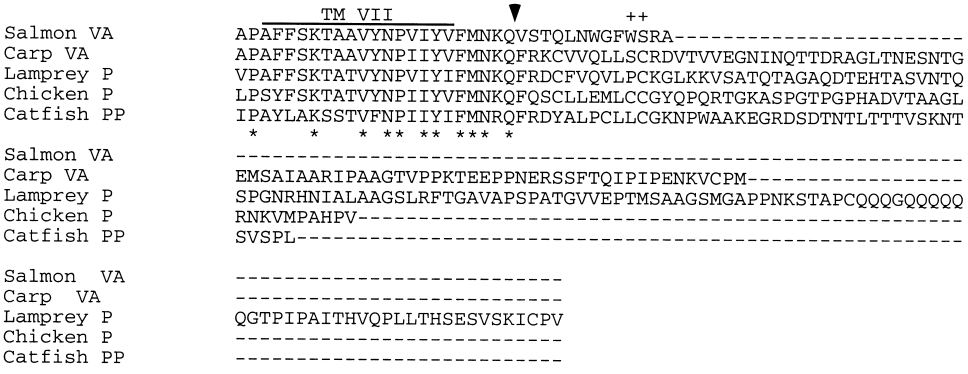

Fig. 2. Partial amino acid alignment of sVA opsin, cVA opsin, lamprey P opsin, chicken P opsin and catfish PP opsin. The alignment shown is restricted to the seventh transmembrane domain (indicated by bar) and the carboxyl terminus (whose start is indicated by an arrow). Amino acid identity is indicated by an asterisk. The plus sign (+) indicates the presence of possible palmitoylation cysteines, absent in sVA opsin.

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 the 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 phototransduction [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 2
Exon/intron splice junctions for cVA opsin gene

Intron	5' junction	3' junction	Size
	340	341	
1	CAGTCACCTTTTTTGgtgagtcaaaagttt---//---aatctgaatctgcagGAATCGTGGCTCTG		1.2 Kb
	551	552	
2	CTGCGAACCCAACTGgtgaggagcacaaac---//---tggtctactctgcagGTATTCGGGAAACT		3.5 Kb
	669	670	
3	CGGAAGCTCAGAAAGgtcagtcgctccaca---//---tgtgtatgtttgcagGTGTCTAACTCGCA		208 bp
	909	910	
4	TTCATGAACAAACAGgtgacaattgcattt---//---ctgtgatctcttcagTTCAGGAAGTGTGT		2.1 Kb

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.

	TM IV	Extra-cellular Loop 2	TM V
		**	
Carp VA opsin	TIPP	VLGWSSYTVSKIGTTCEPNWYSG--NFQDHTFIIT	FFTTCFIFP
Salmon VA opsin	TIPP	VFGWCSYTVSKIGTTCEPNWYSN--NIWNHTYIIT	FFVTCFIMP
Lamprey P-opsin	TLPP	VLGWSSYRPSMIGTTCEPNWYSG--ELHDHTFILM	FFSTCFIFP
Chicken P-opsin	STPP	LLGWSSYVPEGLRTSCGPNWYTG--GSNNNSYILS	LFVTCFVLP
Anolis P-opsin	TLPP	LFGWSSYIPEGLRTSCGPNWYTG--GNDNNSYIMT	LFVTCFITP
Toad P-opsin	TSPP	LIGWCSYVPEGLGTSCGPNWYTG--GTNNNSYILA	LFTTCFMMP
Catfish Parapinopsin	NTPP	LFGWGSYQLEGVMTSCAPNWYRR--DPVNVSYILC	YFMLCFALP
Bovine Rod-opsin	AAPP	LVGWSRYIPEGMQCSCGIDYYTPHEETNNESFVIY	MFVVHFIIP

Fig. 3. Partial amino acid alignment of some of the non-visual photopigments that are found to date with that of the visual bovine Rh opsin. The alignment shown is restricted between the end of the transmembrane IV and the beginning of the transmembrane domain V. The asterisks denote the lack of a doublet of amino acids in all the non-visual opsins.

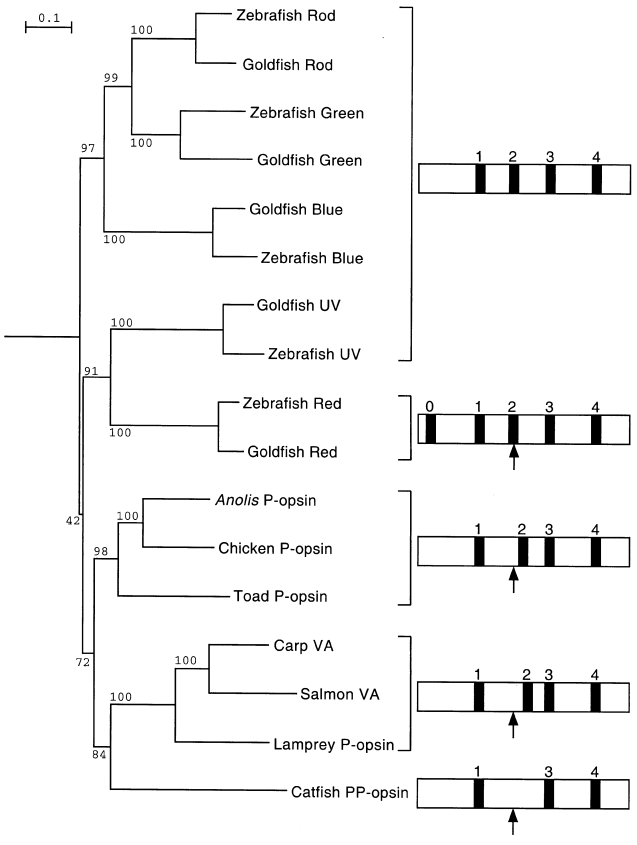

Fig. 4. A neighbour-joining tree of the different classes of opsins (Rh, rod opsin; M1, blue cone opsin; M2, green cone opsin; S, ultraviolet cone opsin; L, red cone opsin; P, pin-opsin; PP, parapinopsin; VA, vertebrate ancient opsin), constructed from nucleotide data. Schematic representations of the intron positions in the different opsin classes are indicated. Intron 0 is only present in the L class opsins. Intron 2 varies in position depending on opsin class, and may be absent, see text for details. The relative position of intron 2 in the Rh/L class opsins is indicated by an arrow. The nucleotide neighbour-joining tree is corrected by the Kimura method [33] and the scale bar is calibrated in substitutions per site. Bootstrap confidence values are based on 1000 replicates. Cuttlefish rhodopsin (*Sepia officinalis*, AF000947) and *Calliphora erythrocephala* Rh1 opsin (M58334) were used as an outgroup. Nucleotide accession numbers: zebrafish: AF109372, AF109369, AF104904, AF109373, AF109368; goldfish: L11864, L11865, L11867, D85863, L11863; *Anolis*: AF134767, AF134768, AF134769, AF134770, AF134771; chicken: U15762; toad: AF200433; carp: AF233520; salmon: AF001499; lamprey: U90667, U90668, U90669, U90670, U90671; catfish: AF028014.

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 families. 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

(A)

```
TACAACCCAGTAATCTACGTCTTCATGAACAAACAGGTCTCAACCCAACTGAACTGGGGA
 Y  N  P  V  I  Y  V  F  M  N  K  Q  V  S  T  Q  L  N  W  G
TTCTGGAGCCGCAGCTGAgacagtgttttccatcaacaaaacaccacattatggaatttA
 F  W  S  R  A  *
ttgtggaagaatggtgtcacacagcactgcactgattgtactgtctgtaatataattcat
gagtaaaacaaattcctgtcaagcattgttgataggcaattaagttgtctgctctattta
ataagtaaatgaataacggaagataaactacacaataaaatgagaatct
```

(B)

```
                880                 900       909
                 |                  |         |
Salmon   TACAACCCAGTAATCTACGTCTTCATGAACAAACAGGTCTCAACCCAACTGAACTGGGGA
Carp     TACAACCCCATCATCTACGTGTTCATGAACAAACAGgtgacaattgcatttttttatcca
         ********  * ******** *****************  ***    * *    *    *

Salmon   TTCTGGAGCCGCAGTTGAgacagtgttt-----tccatcaacaaaacaccacattatgga
Carp     aa--aaagttcaacttcatgcattgcctggggatcaagctgcatgtcttctctcgactga
               **    * ** *  ** **  *     ** * *  **   *  * *   *  **

Salmon   att-tattgtggaagaatggtgtcacacagcactgcactgattgt-----actgtctgta
Carp     ttgatagtttagagttttgtatccagtcaaaactcttttaacttttacaaaacgtccagc
          *  ** * * **    **    **  **  ***    * * * *     *  ***

Salmon   atataattcatgagtaaaacaaattcctgtcaagcattgttgataggca-attaagttgt
Carp     ttgtgcttctggtgtcagatctttcctgattgtgatcaagttacggtcagaataag--gt
          * *  ***  * ** * *    * *   *   *      * *  * ** * ****  **

Salmon   ctgctctatttaataagta--aatgaataacggaagataaac---tacacaaataaaatg
Carp     cagtgatatc-gataggcatcatttaataaagcagttcaagcatctgaagaattaaacct
         * *   ***   *** * *  * * ***** * *    ** *   *  * ** ****

Salmon   agaatct
Carp     gatat--
            **
```

(C)

```
TACAACCCCATCATCTACGTGTTCATGAACAAACAGgtgacaattgcatttttttatcca
 Y  N  P  I  I  Y  V  F  M  N  K  Q  V  T  I  A  F  F  Y  P
292                      300      303
aaaaagttcaacttcatgcattgcctggggatcaagctgcatgtcttctctcgactgatt
 K  K  F  N  F  M  H  C  L  G  I  K  L  H  V  F  S  R  L  I
gatagtttagagttttgtatccagtcaaaactcttttaacttttacaaaacgtccagctt
 D  S  L  E  F  C  I  Q  S  K  L  F  *
gtgcttctggtgtcagatctttcctgattgtgatcaagttacggtcagaataaggtcagt
gatatcgataggcatcatttaataaagcagttcaagcatctgaagaattaaacctgatat
```

Fig. 5. Illustration of the possible mechanism for the production of the short and long VA opsin isoforms. A: Partial cDNA sequence and translation of the sVA opsin carboxyl terminus. Coding sequence is in upper case and 3′ UTR is in lower case. The predicted intron donor site is shown in bold. The polyadenylation signal is underlined. B: Alignment of the sVA opsin cDNA sequence from (A) and a partial genomic sequence of exon IV and intron 4 of cVA opsin (with numeration relative to the cVA opsin cDNA, see Fig. 1). The intron sequence of cVA opsin shares 44% identity with the carboxyl terminus and 3′ UTR cDNA sequence of sVA opsin (shaded). Asterisks indicate identity. C: Partial genomic sequence and theoretical translation of exon IV and intron 4 of cVA opsin from (B) (with numeration relating to the amino acid translation in Fig. 1). The cDNA coding sequence is shown in upper case, intron sequence in lower case and the hypothetical translation in italic. Two potential polyadenylation signals that lie approximately 100 and 130 bp 3′ of the potential stop codon are underlined.

predicted insertion site of intron 4 occurs in the motif ....MNKQ-intron-FR.... 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 are present 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 opsin [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 agnatha. 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*ong* 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

[1] Okano, T., Kojima, D., Fukada, Y., Shichida, Y. and Yoshizawa, T. (1992) Proc. Natl. Acad. Sci. USA 89, 5932–5936.
[2] Nathans, J., Darcy, T. and Hogness, D.S. (1986) Science 232, 193–202.
[3] Yokoyama, R. and Yokoyama, S. (1993) FEBS Lett. 334, 27–31.
[4] Wang, S.-Z., Adler, R. and Nathans, J. (1992) Biochemistry 31, 3309–3315.
[5] Nathans, J. and Hogness, D.S. (1984) Cell 34, 807–814.
[6] Max, M., McKinnon, P.J., Seidenman, K.J., Barrett, R.K., Applebury, M.L., Takahashi, J.S. and Margolskee, R.F. (1995) Science 267, 1502–1506.
[7] Blackshaw, S. and Snyder, S. (1997) J. Neurosci. 17, 8083–8092.
[8] Soni, B.G. and Foster, R.G. (1997) FEBS Lett. 406, 279–283.
[9] Soni, B.G., Philp, A., Knox, B.E. and Foster, R.G. (1998) Nature 394, 27–28.
[10] Philp, A.R., Garcia-Fernandez, J.-M., Soni, B.G., Lucas, R.J., Bellingham, J. and Foster, R.G. (2000) J. Exp. Biol., in press.
[11] Foster, R.G. (1999) in: Rhodopsins and Phototransduction, Vol. Novartis Foundation Symposium 224, pp. 36–43. Wiley, Chichester.
[12] Kojima, D. and Fukada, Y. (1999) in: Rhodopsins and Phototransduction, Vol. Novartis Foundation Symposium 224, pp. 265–282. Wiley, Chichester.
[13] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) Nucleic Acids Res. 25, 4876–4882.
[14] Galtier, N., Gouy, M. and Gautier, C. (1996) Comput. Appl. Biosci. 12, 543–548.
[15] Karnik, S.S., Ridge, K.D., Bhattacharya, S. and Khorana, H.G. (1993) Proc. Natl. Acad. Sci. USA 90, 40–44.
[16] Ostrer, H., Pullarkat, R.K. and Kazmi, M.A. (1998) Mol. Vis. 4, 28–32.
[17] Marin, E.P., Krishna, A.G., Zvyaga, T.A., Isele, J., Siebert, F. and Sakmar, T.P. (2000) J. Biol. Chem. 275, 1930–1936.
[18] Chen, J., Makino, C.L., Peachey, N.S., Baylor, D.A. and Simon, M.I. (1995) Science 267, 374–377.
[19] Brannock, M.T., Weng, K. and Robinson, P.R. (1999) Biochemistry 38, 3770–3777.
[20] Yokoyama, S. and Zhang, H. (1997) Gene 202, 89–93.
[21] Larhammar, D. and Risinger, C. (1994) Trends Genet. 10, 418–419.
[22] Tsai, H.J., Shih, S.R., Kuo, C.M. and Li, L.K. (1994) Comp. Biochem. Physiol. B Biochem. Mol. Biol. 109, 81–88.
[23] Lim, J., Chang, J.L. and Tsai, H.J. (1997) Biochim. Biophys. Acta 1352, 8–12.
[24] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.
[25] Mattick, J.S. (1994) Curr. Opin. Genet. Dev. 4, 823–831.
[26] Zhao, J., Hyman, L. and Moore, C. (1999) Microbiol. Mol. Biol. Rev. 63, 405–445.
[27] Okano, T., Yoshizawa, T. and Fukada, Y. (1994) Nature 372, 94–97.
[28] Yoshikawa, T., Okano, T., Oishi, T. and Fukada, Y. (1998) FEBS Lett. 424, 69–72.
[29] Kawamura, S. and Yokoyama, S. (1997) Vis. Res. 37, 1867–1871.
[30] Kawamura, S. and Yokoyama, S. (1996) Gene 182, 213–214.
[31] Kojima, D., Mano, H. and Fukada, Y. (2000) J. Neurosci. 20, 2845–2851.

[32] Baldwin, J.M., Schertler, G.F. and Unger, V.M. (1997) J. Mol. Biol. 272, 144–164.
[33] Kimura, M. (1980) J. Mol. Evol. 16, 111–120.
[34] Konig, B., Arendt, A., McDowell, J.H., Kahlert, M., Hargrave, P.A. and Hofmann, K.P. (1989) Proc. Natl. Acad. Sci. USA 86, 6878–6882.
[35] Franke, R.R., Konig, B., Sakmar, T.P., Khorana, H.G. and Hofmann, K.P. (1990) Science 250, 123–125.
[36] Acharya, S., Saad, Y. and Karnik, S.S. (1997) J. Biol. Chem. 272, 6519–6524.
[37] Hargrave, P.A., Hamm, H.E. and Hofmann, K.P. (1993) BioEssays 15, 43–50.
[38] Yang, K., Farrens, D.L., Hubbell, W.L. and Khorana, H.G. (1996) Biochemistry 35, 12464–12469.
[39] Cai, K., Klein-Seetharaman, J., Farrens, D., Zhang, C., Altenbach, C., Hubbell, W.L. and Khorana, H.G. (1999) Biochemistry 38, 7925–7930.