THE RHODOPSIN GENE OF THE CUTTLEFISH SEPIA OFFICINALIS: SEQUENCE AND SPECTRAL TUNING

JAMES BELLINGHAM, ALEX G. MORRIS AND DAVID M. HUNT*

Department of Molecular Genetics, Institute of Ophthalmology, University College London, Bath Street, London EC1V 9EL, UK

*Author for correspondence (e-mail: d.hunt@ucl.ac.uk)

Accepted 12 May; published on WWW 14 July 1998

Summary

The cephalopod molluscs are a group of invertebrates that occupy a wide range of oceanic photic environments. They are an ideal group of animals, therefore, in which to study the evolution of rhodopsin. The cDNA sequence of the rhodopsin gene of the cuttlefish Sepia officinalis (L.) (Sub-class Coleoidea, Order Sepiida) is presented, together with an analysis of the structure of the gene. A proline-rich C terminus is present; this structure is characteristic of cephalopod rhodopsins. In common with all invertebrate opsins studied so far, the equivalent site to the counterion in vertebrate opsins is occupied by an aromatic amino acid. An intron is present that splits codon 107, in contrast to the intronless rhodopsin gene in two species of myopsid squid. A spectral tuning model involving substitutions at only three amino acid sites is proposed for the spectral shifts between the rhodopsins of Sepia officinalis, three species of squid and Paroctopus defleini.

Key words: rhodopsin, spectral tuning, cephalopod, intron, cuttlefish, Sepia officinalis.

Introduction

The cephalopods are a class of marine mollusc that emerged approximately 500 million years ago at the end of the Cambrian era (Clarkson, 1993) with two extant subclasses: the Nautiloidea represented by a single genus, Nautilus; and the Coleoidea, which is further divided into four orders: Sepiida (cuttlefishes), Teuthida (squid), Octopoda (octopuses) and Vampyromorpha (Nesis, 1987). Members of the coleoids inhabit a wide range of photic environments from the surface waters of the continental shelf to the bathypelagic depths (Nesis, 1987).

In the species studied so far, only a single type of rhodopsin would appear to be expressed in the eyes of cephalopods (the different pigments described in certain species probably result from incorporation of different chromophores), and the coding sequence of the corresponding gene has been determined for an octopus, Paroctopus defleini (Ovchinnikov et al. 1988), and for three species of squid, Loligo forbesi (Hall et al. 1991), Alloteuthis subulata (Morris et al. 1993) and Todarodes pacificus (Haranishimura et al. 1993). Of the four species, gene structure has been determined for only two, the myopsid squids Alloteuthis subulata and Loligo forbesi; in both cases, the gene is intronless (Morris et al. 1993), a feature in common with the opsin gene expressed in rod photoreceptors of teleost fish (Fitzgibbon et al. 1995; Hope et al. 1997) and certain opsins of Drosophila spp. (Carulli et al. 1994).

The rhodopsins of cephalopods uniquely possess a proline-rich C terminus in the form of repeat units, typically of the imperfect repeat sequence tyrosine–proline–proline–glutamine–glycine interspersed with alanine and further proline residues. This poly-proline tail has recently been shown to be important in maintaining the mobility of rhodopsin in the microvillar membrane (Venien-Bryan et al. 1995).

To investigate further the structure, evolution and adaptation of cephalopod rhodopsins, we have examined the rhodopsin gene of the common cuttlefish Sepia officinalis (L.), a member of a third order of coleoids (the Sepiida). This species is found over the continental shelf mainly in the upper sublittoral zone at depths down to 100 m (Nesis, 1987).

Materials and methods

Preparation of genomic DNA and retinal cDNA

Genomic DNA was isolated from approximately 2 g of mantle tissue from a male Sepia officinalis, as described by Morris et al. (1993). Poly(A+) mRNA was extracted from 2 g of retinal tissue using a QuickPrep micro mRNA purification kit (Pharmacia).

cDNA cloning and sequencing

Single-stranded cDNA was synthesised using the Gibco BRL 3’ rapid amplification of cDNA ends (RACE) technique. Polymerase chain reactions (PCRs) were performed in a total volume of 25 μl with 12.5 pmol of each primer, 0.5 nmol each of dATP, dCTP, dGTP and dTTP, 1.5 mmol l⁻¹ MgCl₂ and 0.25 units of Taq polymerase (BioTaq) in the manufacturer’s
NH₄ buffer. A consensus primer (5'-CAATGAGACATGG-TGGTATAAYCC-3') designed to a conserved region of the rhodopsin sequences of *Paroctopus deffeinii* (Ovchinnikov et al. 1988), *Loligo forbesi* (Hall et al. 1991), *Alloteuthis subulata* (Morrison et al. 1993) and *Todarodes pacificus* (Hara-Nishimura et al. 1993), was used in conjunction with the oligo(dT) reverse primers supplied with the 3' RACE system at an annealing temperature of 60 °C to amplify a fragment of approximately 1800 base pairs (1.8 kb), which encompassed the coding region for all seven α-helices and the 3' untranslated region (UTR). The amplified products from three separate PCRs were cloned into pGEM-T (Promega) and sequenced using a T7 sequencing kit (Pharmacia) with [35S]dATP and visualised by autoradiography.

The remaining 5' coding sequence was obtained by amplifying across a circularised/concatamerised double-stranded cDNA generated from poly(A') mRNA with a reverse primer designed to the *Sepia officinalis* 5' UTR (5'-GCATCTACAAATAAGTCTACAG-3') using the GibcoBRL cDNA synthesis system. The double-stranded product was isolated in deionised water using a Centricron-100 (Amicon) column. Ligation reactions were set up in a total volume of 20 µl with 15 µg of double-stranded cDNA and 6 units of T4 DNA ligase (Promega) in the manufacturer's protocol (Domínguez and López-Larrea, 1994). Species-specific nested primers were designed for use in gene walking. A 571 bp fragment of approximately 650 bp was generated for the 3' UTR and a specific primer (primer D, reverse 5'-GGAGTCCCTATGCTGTCGTG-3') was designed to amplify across the ligation between the 3' and 5' UTRs of the circularised/concatamerised cDNA. PCRs containing 3 µl of ligated cDNA and the above pair of primers at an annealing temperature of 58 °C yielded a fragment of approximately 500 bp which was cloned and sequenced as above.

**Amplification of genomic DNA**

The *Sepia officinalis* rhodopsin gene was amplified by a combination of conventional PCR and gene-walking. A 571 bp product was initially produced by PCR amplification with primers designed from the rhodopsin gene of *Loligo forbesi* (Hall et al. 1991) (5'-GGAGTCCCTATGCTGTCGTG-3'; 5'-GCCTGATAAGGCTGTTATC-3'). This sequence was then extended using an unpredictably primed PCR (UP-PCR) protocol (Domínguez and López-Larrea, 1994). Species-specific nested primers were designed for use in gene walking into both the 5' end of the gene (A1, reverse 5'-GCATAAGGTTTACCCACTCG-3'; A2, reverse 5'-CCGAACCTG-GGCGAGCAGACG-3') and the 3' flanking region (B1, forward 5'-TCAATGCCTCAATTCGACGAG-3'; B2, forward 5'-AAGGAGCAACTTAAAGTGC-3') with two 'universal' primers (C69, 5'-TTTTTTTTTTTTTTTTTTTTTGT- TGGGTTGGTTGGT-3'; C47, 5'-TTTTTTTTTTTTTTTTTTTTTGTGGG-3'). The first-round UP-PCR mix contained 10 ng of DNA, 12.5 pmol of primer C69, 0.5 nmol each of dATP, dCTP, dGTP and dTTP, 3 mMol1-1 MgCl₂, and NH₄ PCR buffer in a volume of 23 µl. This mix was denatured at 94 °C for 60 s, during which 0.5 units of *Taq* polymerase was added in a volume of 1 µl. The sample temperature was reduced from 80 °C for 30 s, to 15 °C for 1–2 min and to 25 °C for 10 min, followed by heating to 72 °C for 60 s prior to denaturation at 94 °C for 60 s during which 12.5 pmol of the outer species-specific primer (A1 or B1) was added in a volume of 1 µl. The sample was then cycled at 94 °C for 1 s, 62 °C for 30 s and 72 °C for 30 s over 35 cycles, followed by a final extension of 30 s at 72 °C.

The nested UP-PCR contained 1 µl of a 1:1000 dilution of the first-round UP-PCR, 12.5 pmol each of primer C47 and either species-specific primer A2 or B2 (depending on the first-round species-specific primer), 0.5 mmol each of dATP, dCTP, dGTP and dTTP, 1.5 mMol1-1 MgCl₂, and NH₄ PCR buffer in a volume of 24 µl. This mixture was denatured at 94 °C for 60 s, during which 0.25 units of *Taq* polymerase was added in a volume of 1 µl. The sample was then cycled at 94 °C for 1 s, 56 °C for 1 s and 72 °C for 30 s over 35 cycles, followed by a final extension of 30 s at 72 °C.

A PCR fragment of approximately 1.9 kb was generated for the 5' end of the gene and partially sequenced in order to design a specific primer (C, forward 5'-CATATTGGTGTTACAGTTAGCGG-3') for use against primer A2. Similarly, a PCR fragment of approximately 650 bp was generated for the 3' UTR and a specific primer (primer D, reverse 5'-CATATTGGTGTTACAGTTAGCGG-3') designed for use against primer B2. For the 5' region, primers C and A2, when cycled at an annealing temperature of 58 °C, produced a fragment of approximately 1.8 kb. A 3' UTR fragment of approximately 600 bp was generated by primers D and B2 cycled at an annealing temperature of 56 °C.

**Sequence analysis**

Sequence alignments were performed with ClustalV (Higgins et al. 1992) using default fixed-gap and floating penalties, and unweighted for transitions. Pairwise divergence values of amino acid sequences were determined by the MEGA computer package (Kumar et al. 1993). GeneWorks was used to generate a Kyte–Doolittle hydrophobicity plot using a window of 11 residues.

**Results**

The *rhodopsin cDNA* of *Sepia officinalis*

The complete coding sequence of the rhodopsin cDNA together with the 5' and 3' UTRs is shown in Fig. 1A. A putative poly(A) tail and a poly(A) addition signal (AATAAA) 345 bp downstream from the stop codon would appear to be present in the 3' UTR. However, the sequence of a 545 bp genomic DNA fragment generated from the 3' flanking region by gene-walking (Fig. 1B) indicates that this may not be the case, since the string of 17 A residues at the 3' end of the cloned cDNA sequence is part of a stretch of 27 A nucleotides 346 bp downstream from the stop codon. The poly(A) region at the 3' end of the cDNA would appear, therefore, to be a direct product of transcription with the
A cDNA sequence

```
agaattcaagcgccgtaatgtttctagttattgttgctgaggaatagtaaatagtggtgccc
```

Functional poly(A) + addition signal further downstream from the sequenced region.

By alignment with other cephalopod rhodopsins, a translation product of 464 amino acid residues was deduced (Fig. 2). A hydrophobicity plot (not shown) of this sequence shows the presence of seven potential transmembrane regions. Each of these regions consists of an α-helix of approximately 26 residues, with the middle 18 residues embedded in the membrane (Baldwin, 1993). A lysine residue for the attachment of retinal is present at position 305 in helix VII, without tyrosine rather than glutamate occupying site 111 just outside the transmembrane region of helix III, the position of the counterion to the Schiff base (site 113 in bovine rod opsin numbering) in vertebrate opsins (Zhukovsky and Oprian, 1989). This site is invariably occupied by an aromatic amino acid in invertebrate pigments, but recent work (Hashimoto et al. 1996) indicates that it probably does not serve this role in invertebrate pigments. Other typical opsin features include two cysteine residues (Cys-108 and Cys-186) which form a disulphide bridge that is essential for opsin stability (Karnik et al. 1988), a conserved Asp–Arg–Tyr motif at the junction of helix III with the second intracellular loop that is required for G-protein binding (Franke et al. 1990), a doublet of cysteine residues (Cys-336 and Cys-337), which may be palmitoylated (Ovchinnikov et al. 1988; Nakagawa et al. 1997), and one glycosylation site in the N-terminal part of the protein that may be important for anchoring the nascent polypeptide in the membrane. The third intracellular loop is highly conserved across all cephalopod rhodopsins. This region is thought to be involved in the binding of the Gq class of G-proteins and in the end-to-end contact between molecules in cephalopod rhodopsin crystals (Davies et al. 1996).

A C-terminal poly-proline tail, characteristic of squid (Hall et al. 1991; Morris et al. 1993; Hara-Nishimura et al. 1993) and octopus (Ovchinnikov et al. 1988) rhodopsins, is also found in Sepia officinalis rhodopsin. Twelve imperfect Tyr–Pro–Pro–Gly–Gly repeats are present compared with the 10 repeats in the three squid species and 11 repeats in the octopus. A perfectly conserved region of 16 amino acid residues precedes the poly-proline tail in each species (Fig. 3).

Identification of an intron in the Sepia officinalis rhodopsin gene

Previous analysis of the rhodopsin genes in two species of myopsid squid, Alloteuthis subulata and Loligo forbesi (Morris et al. 1993), showed that introns are totally lacking from the coding region. In order to determine whether the Sepia officinalis gene also lacks introns, a PCR was set up with Sepia officinalis genomic DNA using primers designed to the 5′ and 3′ ends of the Sepia officinalis cDNA sequence. No amplified product resulted, indicating that the gene may be interrupted by one or more introns. Three overlapping genomic fragments covering approximately 2.8 kb of the coding region and 545 bp of the 3′ flanking regions were amplified by a combination of conventional PCR and a gene-walking protocol (Domínguez and López-Larrea, 1994). The sequence of this region revealed a typical intron acceptor site boundary that splits codon 107 (Fig. 4). The Sepia officinalis rhodopsin gene is interrupted, therefore, by an intron at this position that is in excess of 1 kb in length.
Fig. 2. Alignment of the five cephalopod rhodopsin amino acid sequences. The positions of the seven α-helical transmembrane regions are indicated by horizontal lines. Amino acid identity across species is indicated by a dot (.) and alignment insertions are indicated by a dash (–).

Fig. 3. The C terminus of cephalopod rhodopsin based on the imperfect repeat Tyr–Pro–Pro–Gln–Gly. Note the conserved region of 16 amino acid residues (denoted by a horizontal line) immediately before the poly-proline tail region.
The relative position of the intron in the *Sepia officinalis* rhodopsin gene in relation to those in other opsin genes was determined by adding the *Sepia officinalis* sequence to the alignment in Bellingham et al. (1997) of the Rh1, Rh2 and Rh4 rhodopsin genes of *Drosophila melanogaster* (O’Tousa et al. 1985; Cowman et al. 1986; Montell et al. 1987), the ultraviolet opsin gene of the honeybee *Apis mellifera* (Bellingham et al. 1997) and the human blue opsin gene (Nathans et al. 1986). The position of the intron in *Sepia officinalis* does not correspond to the position of introns in any of these genes (data not shown).

**Sequence identity of cephalopod rhodopsins**

The per cent amino acid identity of the different cephalopod rhodopsins shows a high level of sequence conservation over all five species (Table 1). The identity between the two myopsid squids is particularly high at 95 %, but even when comparisons of *Paroctopus defleini* with the other four species are made this only falls to approximately 76 %. This contrasts with values of 40–50 % when the different classes of vertebrate cone opsins are compared.

**Spectral tuning of cephalopod rhodopsins**

Previous work on vertebrate visual pigments has shown that most of the amino acid changes that affect the wavelength of maximum absorbance ($\lambda_{max}$) of a pigment are located in the transmembrane helices and involve either a change of charge or the gain/loss of a hydroxyl group (Nathans, 1990; Nakayama and Khorana, 1990; Merbs and Nathans, 1993; Asenjo et al. 1994). Forty-four transmembrane sites show substitution in one or more of the five cephalopod species. To determine which of the substituted sites are in a position to interact with the chromophore, these sites were mapped on to the three-dimensional G-protein-linked receptor model of Baldwin (1993). This model is based on a comparison of the amino acid sequences of 204 G-protein receptors, including 32 visual pigments, and the two-dimensional crystal structure of bovine rhodopsin (Schertler et al. 1993). The applicability of this model to cephalopod rhodopsins is supported by the recent demonstration of the similarity between the two-dimensional crystal structures of squid and bovine rhodopsin (Davies et al. 1996). In the absence of a high-resolution crystal structure for rhodopsin, this model is generally considered to provide a satisfactory basis for determining the relative positions of residues around the seven $\alpha$-helical transmembrane regions (Hunt et al. 1996; Bellingham et al. 1997; Hope et al. 1997). Where the role of spectral tuning substitutions has been confirmed by site-directed mutagenesis in opsins expressed in vitro, as in the case of the human red and green visual pigments (Merbs and Nathans, 1993; Asenjo et al. 1994), such sites are located by the Baldwin model in the central hydrophilic chromophore-binding pocket.

Using this model, each helix of the five cephalopod rhodopsins was orientated in relation to the exterior lipid membrane and this central pocket. Only four substitutions, at

Table 1. Amino acid sequence identity

<table>
<thead>
<tr>
<th>Species</th>
<th>Per cent identity</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sepia officinalis</em></td>
<td>89.9</td>
</tr>
<tr>
<td><em>Alloteuthis subulata</em></td>
<td>95.0</td>
</tr>
<tr>
<td><em>Loligo forbesi</em></td>
<td>88.1</td>
</tr>
<tr>
<td><em>Todarodes pacificus</em></td>
<td>76.3</td>
</tr>
<tr>
<td><em>Paroctopus defleini</em></td>
<td>76.3</td>
</tr>
</tbody>
</table>

Table 2. Amino acid substitutions and proposed spectral shifts at the candidate spectral tuning sites

<table>
<thead>
<tr>
<th>Species</th>
<th>$\lambda_{max}$ (nm)</th>
<th>Amino acid site</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Alloteuthis subulata</em></td>
<td>499</td>
<td>Thr, Ala, Tyr, Ser</td>
</tr>
<tr>
<td><em>Loligo forbesi</em></td>
<td>494</td>
<td>Thr, Ala, Tyr, Phe, -5 nm</td>
</tr>
<tr>
<td><em>Sepia officinalis</em></td>
<td>492</td>
<td>Ser, Ser, Tyr, Phe, -5 nm</td>
</tr>
<tr>
<td><em>Todarodes pacificus</em></td>
<td>482</td>
<td>Ala, Ala, Phe, -5 nm</td>
</tr>
<tr>
<td><em>Paroctopus defleini</em></td>
<td>480</td>
<td>Ala, Ser, Tyr, Phe, -5 nm</td>
</tr>
</tbody>
</table>
sites 127, 167, 205 and 270 (numbering from the Sepia officinalis sequence) in transmembrane regions III, IV, V and VI, respectively, fulfill the above criteria for identifying potential spectral tuning sites (Fig. 5). The replacement of Ser-270 in Alloteuthis subulata by Phe in Loligo forbesi has previously been proposed (Morris et al. 1993) to account for the blue shift of 5 nm in the latter species, and this is supported by the presence of Phe at this site in the other cephalopod opsins. The additional blue shifts to 492 nm in Sepia officinalis (Messenger, 1981), to 482 nm in Todarodes pacificus and to 480 nm in Paroctopus defleini (Seidou et al. 1990) can be explained by different combinations of substitution at sites 127 and 167 (Table 2). In this model, site 205, which is substituted only in Todarodes pacificus, would have little or no effect on spectral tuning.

**Discussion**

The presence of an intron in the Sepia officinalis rhodopsin gene was an unexpected observation since the coding regions of two other cephalopod rhodopsin genes, those of the myopsid squids Alloteuthis subulata and Loligo forbesi, are both intron-free (Morris et al. 1993). This raises the possibility that intron(s) may be present in the rhodopsin gene of the oegopsid squid Todarodes pacificus and the octopus Paroctopus defleini. At present, only the cDNA sequences are known for these two species (Ovchinnikov et al. 1988; Hara-Nishimura et al. 1993), although an intron in the rhodopsin gene of another octopus, Octopus bimaculoides, has been reported (Nishita et al. 1995) but its precise position was not determined. From our data, therefore, it would seem probable that introns were present in the rhodopsin gene of the ancestral cephalopods and that the single intron present in the Sepia officinalis gene may represent a step in the total loss of introns as found in the rhodopsin genes of the two myopsid squids. The opsin gene expressed in rod photoreceptors of teleost fish also lacks introns, whereas at least four introns are present in the orthologous gene of all other vertebrates sequenced to date (Nathans and Hogness, 1984; Nathans et al. 1986; Baehr et al. 1988; Takao et al. 1988; Yokoyama and Yokoyama, 1993), indicating that introns have also been lost from the rod opsin gene during the evolution of fish (Fitzgibbon et al. 1995).

The position of the intron in Sepia officinalis is at a site not seen in any other rhodopsin gene, whether invertebrate or vertebrate (Nathans and Hogness, 1984; O’Tousa et al. 1985;
Bellingham et al. 1997; Cowman et al. 1986). At least 12 unique locations for introns have now been described in invertebrate (mostly insect) opsin genes (Montell et al. 1987) in what is a relatively short coding sequence of just over 1000 bp (if the poly-proline tail region that is unique to cephalopods is excluded); the addition of one or more of these subsequent to the evolution of the ancestral opsin gene must be considered a real possibility.

A characteristic feature of cephalopod rhodopsins is the presence of imperfect repeats of the Tyr–Pro–Pro–Gly–Gly motif. Alignments of this region show that the number of repeats varies from 10 in the three squid species to 12 in Sepia officinalis. Immediately prior to the poly-proline tail is a region of 16 amino acid residues that is perfectly conserved over the five cephalopod species. The proline-rich C terminus has been shown to be important in the stabilisation of rhodopsins in the cell membrane (Venien-Bryan et al. 1995). The role of the conserved 16-residue motif remains unclear.

An interesting feature of the cephalopod rhodopsins sequenced to date is that their \( \lambda_{\text{max}} \) values range from 499 nm in Alloteuthis sublata to 480 nm in Paroctopus deflei. By comparing the deduced amino acid sequence from the five cephalopod rhodopsins now reported, it is possible to identify candidate amino acid substitutions for these spectral shifts. Previous work with vertebrate opsin genes has demonstrated that those substitutions that alter the spectral location of the \( \lambda_{\text{max}} \) of a visual pigment generally involve either a charge change or the gain/loss of a hydroxyl group (Nathans, 1990; Nakayama and Khorana, 1990; Merbs and Nathans, 1993; Asenjo et al. 1994) and are positioned in close proximity to the chromophore on the inner face of the transmembrane \( \alpha \)-helices (Montell et al. 1987; Yokoyama and Yokoyama, 1993). Only four substitutions meet these criteria across the five cephalopod rhodopsin sequences, and a fully additive model for the spectral tuning of cephalopod rhodopsins can be constructed from only three of these. In this model, the replacement of Ser by Ala at site 127 results in a blue shift of 12 nm and accounts for the major part of the spectral difference between the Alloteuthis sublata (499 nm), Loligo forbesi (494 nm) and Sepia officinalis (492 nm) on the one hand and Todarodes pacificus (482 nm) and Paroctopus deflei (480 nm) on the other. The 5 nm shift from 499 nm in Alloteuthis sublata to 494 nm in Loligo forbesi can be explained by a Ser to Phe substitution at site 270. This site is equivalent to site 277 of primate red and green opsinps, which has been shown by site-directed mutagenesis (Merbs and Nathans, 1993; Asenjo et al. 1994) to cause a blue shift when the hydroxyl-bearing amino acid Tyr is replaced by Phe. Finally, an Ala to Ser substitution at site 167 would account for the blue shift of 2 nm between Loligo forbesi and Sepia officinalis and between Todarodes pacificus and Paroctopus deflei.

This work was supported by a grant from the BBSRC.

References


