Volume 232, number 1, 69–72 FEB 05828 May 1988

Octopus rhodopsin

Amino acid sequence deduced from cDNA

Yu.A. Ovchinnikov†, N.G. Abdulaev, A.S. Zolotarev, I.D. Artamonov, I.A. Bespalov, A.E. Dergachev and M. Tsuda*

Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, 117871 Moscow V-437, USSR and **Department of Physics, Sapporo Medical College, Sapporo, Japan*

Received 9 March 1988

The primary structure of rhodopsin from the octopus *Paroctopus defleini* has been determined by parallel analysis of the protein and corresponding cDNA. The amino acid sequence is most similar to the recently cloned *Drosophila* opsins. Similarities to bovine and human opsins are also evident. The transmembrane topology of octopus rhodopsin is discussed.

Rhodopsin; Primary structure; Nucleotide sequence; Transmembrane topology; (*Paroctopus defleini*)

1. INTRODUCTION

The determination of the complete amino acid sequence of bovine rhodopsin was the culmination of over a decade of intensive work by many investigators [1,2]. Recent progress in recombinant DNA techniques provided cDNA sequence information for a number of visual pigments of both vertebrate and invertebrate origins. These studies demonstrated the gross structural and topological similarities in the opsin family of proteins [3]. However, some striking differences such as greater thermal stability, and photoregeneration of the isomerized chromophore [4,5] are characteristic of invertebrate rhodopsins. Investigation of the structural features of octopus rhodopsin might account for these and other differences. Here, we report on the nucleotide sequence of cDNA encoding octopus opsin and on the principal structural and topological peculiarities of this protein.

† Deceased

Correspondence address: N.G. Abdulaev, Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, 117871 Moscow V-437, USSR

2. MATERIALS AND METHODS

Photoreceptor membranes from octopus retina were prepared as in [6]. The preparation of membranes containing mainly rhodopsin was solubilized in 10 mM Tris-HCl (pH 8.0), 2% SDS, 5% β-mercaptoethanol, 1 mM EDTA overnight in the dark at room temperature. Proteins were fractionated on Ultragel AcA44 (LKB, Sweden) equilibrated in 50 mM NH_4HCO_3 (pH 8.0), 0.1% SDS, 0.1% β-mercaptoethanol and 1 mM EDTA. Fractions with denatured rhodopsin were pooled according to PAGE in the presence of SDS using a highly purified sample of octopus rhodopsin as a standard.

A sample of 0.5 μM octopus opsin solubilized in 0.15 M Tris-HCl (pH 8.5), 3 mM EDTA and a 50-fold excess of DTT per mol cysteine was subjected to gel filtration on Bio-Gel P2 equilibrated in 0.1 M Na acetate (pH 4.0), 1% SDS and 3 mM EDTA. The protein was coupled to insoluble carrier (CPG-thiol, Pierce, USA), cleaved with CNBr and the peptides separated as in [7].

A cDNA library in vector pBR322 was produced by a poly(A)$^+$ RNA fraction from whole octopus eyes essentially as described [8]. To initiate synthesis of the first cDNA chain use was made of: (i) oligo(dT)$_{12-18}$, (ii) a statistical mixture of hexanucleotide primers. The nucleotide sequence of DNA fragments was determined according to Sanger et al. [9]. The strategy for nucleotide sequencing was that of Henikoff [10].

3. RESULTS AND DISCUSSION

CNBr peptides of puridied rhodopsin were

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

isolated by gel filtration on Bio-Gel and reverse-phase HPLC. The complete or partial amino acid sequences of 13 peptides comprising 25% of the polypeptide chain were determined on a gas-phase sequencer. Oligodeoxyribonucleotides

5' AAANCC$^{A}_{G}$CA$^{A}_{G}$AA$^{A}_{G}$TACAT 3'
CAT
GAANCC$^{A}_{G}$CA$^{A}_{G}$AA$^{A}_{G}$TACAT

corresponding to peptide Met-Tyr-Phe-Cys-Gly-Phe-Met were synthesized and used as hybridization probes to screen the cDNA library. Three hybridization-positive clones were isolated from 1 × 105 transformants. One of them, pORh104 (fig.1) encodes the central part of the polypeptide chain and two other clones pORh124 and pORh126 – the C-terminal part of the protein; pORh126 also contained the 3'-untranslated region.

The second cDNA library was produced as mentioned above [8] except that the statistical mixture of hexanucleotides was used to prime the first chain synthesis. Screening of this library with nick-translated pORh104 insert made identification of five clones possible. Clone pORh462, carrying the largest cDNA insert, was sequenced to overlap the entire coding region of octopus rhodopsin.

Fig.2 shows the 1365-nucleotide sequence of cDNA encoding octopus rhodopsin. All the complete and partial amino acid sequences of CNBr peptides determined by peptide analysis were found to be encoded by the cDNA sequence in the same reading frame. Triplet ATG(75–77) is the initiation codon of octopus rhodopsin, since it is the first ATG codon that appears downstream of a nonsense codon (TAA at positions 33–31) found in the frame and since the second ATG codon codes for amino acid residue 72 preceded by one of the peptide sequences determined. The 455th codon specifying alanine is followed by the termination codon. The octopus rhodopsin amino acid sequence derived from the cDNA nucleotide sequence contains 455 residues (fig.2) of molecular mass 50324 Da.

Octopus rhodopsin, like mammalian and *Drosophila* pigments, has lysine (306) corresponding to position 296 of bovine rhodopsin, the site of covalent attachment of 11-*cis*-retinal. The

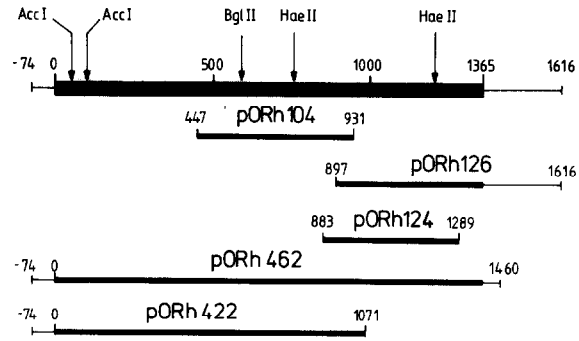

Fig.1. Location of fragments of the isolated clones in the restriction map of octopus rhodopsin cDNA.

polypeptide chain is divided into several alternating hydrophobic and hydrophilic sequences, the latter being much more extended than those in bovine and human opsins. A characteristic feature of the octopus opsin polypeptide chain is the unexpectedly long C-terminal tail adjacent to the last hydrophobic stretch. This tail is very rich in prolines and glutamines. Except for the fact that this tail is easily proteolysed in both octopus and squid rhodopsin, data are not available on its functional role. Finally, there are two potential glycosylation sites closer to the N-terminus of the protein.

The delineation of the transmembrane sections of octopus rhodopsin does not differ markedly from that of vertebrate and *Drosophila* counterparts [1,11]. The model depicts seven transmembrane regions connected by polypeptide loops on both sides of the membrane (fig.3). These loops are of the same size as in *Drosophila* pigments except for a slight difference in amino acid sequences.

The retinal-binding site (Lys-306) in the seventh transmembrane segment is strongly conserved. It should, however, be noted that the tyrosine residue nearest to this lysine which is conserved in all known sequences is replaced by histidine in octopus rhodopsin. The second transmembrane segment of octopus rhodopsin contains an obvious candidate (Asp-81) for a counterion to protonate the retinal-lysine Schiff base. If this is the case, there are no additional negatively charged amino acids which could contribute to the regulation of chromophore absorption [12]. This may well be accomplished by an appropriate disposition of a large number of aromatic amino acids in the transmembrane segments [13].

```
                             -74 ATTGGGTTGTACTCTAGAGGGGTAGAATACCTAGTATTCCCTAAAAAGCACAAGCGTTAACCCAAGCATTAAAA   -1

ATG GTG GAA TCA ACA ACG TTA GTA AAC CAG ACA TGG TGG TAT AAT CCA ACC GTA GAC ATC CAT CCT CAT TGG GCC AAG TTC GAT CCC ATC   90
Met-Val-Glu-Ser-Thr-Thr-Leu-Val-Asn-Gln-Thr-Trp-Trp-Tyr-Asn-Pro-Thr-Val-Asp-Ile-His-Pro-His-Trp-Ala-Lys-Phe-Asp-Pro-Ile   30

CCA GAT GCA GTC TAC TAT TCT GTA GGT ATC TTC ATC GGT GTT GTT GGA ATT ATC GGA ATC CTA GGC AAT GGT GTC GTC ATC TAC CTT TTC  180
Pro-Asp-Ala-Val-Tyr-Tyr-Ser-Val-Gly-Ile-Phe-Ile-Gly-Val-Val-Gly-Ile-Ile-Gly-Ile-Leu-Gly-Asn-Gly-Val-Val-Ile-Tyr-Leu-Phe   60

TCC AAA ACG AAA TCT CTA CAG ACC CCG GCT AAC ATG TTT ATC ATC AAT CTC GCT ATG TCT GAC TTG AGT TTC TCA GCT ATT AAT GGA TTT  270
Ser-Lys-Thr-Lys-Ser-Leu-Gln-Thr-Pro-Ala-Asn-Met-Phe-Ile-Ile-Asn-Leu-Ala-Met-Ser-Asp-Leu-Ser-Phe-Ser-Ala-Ile-Asn-Gly-Phe   90

CCG CTT AAA ACA ATA TCA GCG TTT ATG AAA AAG TGG ATT TTC GGT AAA GTT GCT TGT CAA CTT TAT GGT TTG CTG GGC GGT ATC TTC GGA  360
Pro-Leu-Lys-Thr-Ile-Ser-Ala-Phe-Met-Lys-Lys-Trp-Ile-Phe-Gly-Lys-Val-Ala-Cys-Gln-Leu-Tyr-Gly-Leu-Leu-Gly-Gly-Ile-Phe-Gly  120

TTC ATG TCA ATC AAC ACC ATG GCC ATG ATC TCC ATC GAT CGT TAT AAC GTC ATT GGA AGA CCT ATG GCA GCG TCC AAA AAA ATG TCC CAT  450
Phe-Met-Ser-Ile-Asn-Thr-Met-Ala-Met-Ile-Ser-Ile-Asp-Arg-Tyr-Asn-Val-Ile-Gly-Arg-Pro-Met-Ala-Ala-Ser-Lys-Lys-Met-Ser-His  150
AGA AGA GCT TTC CTC ATG ATT ATC TTT GTG TGG ATG TGG TCC ATT GTT TGG TCA GTC GGA CCC GTC TTC AAC TGG GGA GCA TAC GTC CCC  540
Arg-Arg-Ala-Phe-Leu-Met-Ile-Ile-Phe-Val-Trp-Met-Trp-Ser-Ile-Val-Trp-Ser-Val-Gly-Pro-Val-Phe-Asn-Trp-Gly-Ala-Tyr-Val-Pro  180

GAA GGT ATT CTC ACA TCC TGC TCT TTC GAT TAC CTC TCC ACT GAT CCT AGT ACC AGA TCT TTC ATC TTG TGC ATG TAC TTC TGT GGT TTC  630
Glu-Gly-Ile-Leu-Thr-Ser-Cys-Ser-Phe-Asp-Tyr-Leu-Ser-Thr-Asp-Pro-Ser-Thr-Arg-Ser-Phe-Ile-Leu-Cys-Met-Tyr-Phe-Cys-Gly-Phe  210

ATG CTG CCC ATA ATT ATC ATC GCT TTC TGT TAT TTC AAC ATT GTC ATG TCT GTA TCC AAC CAC GAA AAG GAA ATG GCT GCC ATG GCA AAG  720
Met-Leu-Pro-Ile-Ile-Ile-Ile-Ala-Phe-Cys-Tyr-Phe-Asn-Ile-Val-Met-Ser-Val-Ser-Asn-His-Glu-Lys-Glu-Met-Ala-Ala-Met-Ala-Lys  240

AGG TTG AAT GCC AAA GAA TTG CGT AAA GCA CAG GCT GGT GCG AGC GCT GAA ATG AAA CTT GCC AAA ATT TCA ATG GTA ATT ATT ACC CAA  810
Arg-Leu-Asn-Ala-Lys-Glu-Leu-Arg-Lys-Ala-Gln-Ala-Gly-Ala-Ser-Ala-Glu-Met-Lys-Leu-Ala-Lys-Ile-Ser-Met-Val-Ile-Ile-Thr-Gln  270

TTC ATG CTT TCC TGG TCT CCA TAC GCC ATC ATC GCT CTT CTT GCA CAG TTT GGG CCA GCT GAA TGG GTT ACT CCA TAC GCA GCC GAA TTG  900
Phe-Met-Leu-Ser-Trp-Ser-Pro-Tyr-Ala-Ile-Ile-Ala-Leu-Leu-Ala-Gln-Phe-Gly-Pro-Ala-Glu-Trp-Val-Thr-Pro-Tyr-Ala-Ala-Glu-Leu  300

CCT GTA CTG TTT GCT AAA GCT TCA GCT ATC CAC AAC CCA ATT GTC TAC TCT GTT TCC CAT CCA AAG TTC AGA GAG GCC ATC CAA ACC ACA  990
Pro-Val-Leu-Phe-Ala-Lys-Ala-Ser-Ala-Ile-His-Asn-Pro-Ile-Val-Tyr-Ser-Val-Ser-His-Pro-Lys-Phe-Arg-Glu-Ala-Ile-Gln-Thr-Thr  330

TTC CCA TGG TTG CTG ACA TGT TGT CAA TTC GAT GAG AAA GAA TGC GAA GAT GCT AAT GAT GCC GAA GAA GAA GTC GTA GCT TCC GAA CGC 1080
Phe-Pro-Trp-Leu-Leu-Thr-Cys-Cys-Gln-Phe-Asp-Glu-Lys-Glu-Cys-Glu-Asp-Ala-Asn-Asp-Ala-Glu-Glu-Glu-Val-Val-Ala-Ser-Glu-Arg  360

GGC GGT GAA TCC CGT GAT GCC GCA CAA ATG AAA GAA ATG ATG GCA ATG ATG CAG AAA ATG CAA GCA CAA CAA GCT GCC TAC CAA CCA CCA 1170
Gly-Gly-Glu-Ser-Arg-Asp-Ala-Ala-Gln-Met-Lys-Glu-Met-Met-Ala-Met-Met-Gln-Lys-Met-Gln-Ala-Gln-Gln-Ala-Ala-Tyr-Gln-Pro-Pro  390

CCA CCA CCT CAG GGC TAC CCA CCA CCA GGC TAC CCA CCC CAA GGC GCC TAT CCA CCA CCT CAG GGC TAC CCA CCA CAA GGC TAC CCA CCA 1260
Pro-Pro-Pro-Gln-Gly-Tyr-Pro-Pro-Pro-Gly-Tyr-Pro-Pro-Gln-Gly-Ala-Tyr-Pro-Pro-Pro-Gln-Gly-Tyr-Pro-Pro-Gln-Gly-Tyr-Pro-Pro  420

CAA GGC TAC CCA CCT CAA GGC TAC CCA CCC CAG GGA GCA CCA CCC CAA GTA GAG GCA CCC CAA GGA GCA CCA CCC CAA GGA GTC GAC AAC 1350
Gln-Gly-Tyr-Pro-Pro-Gln-Gly-Tyr-Pro-Pro-Gln-Gly-Ala-Pro-Pro-Gln-Val-Glu-Ala-Pro-Gln-Gly-Ala-Pro-Pro-Gln-Gly-Val-Asp-Asn  450

CAG GCC TAT CAA GCT TGA GAAGCAGGTCTTTTAAGAATTACTTAGAATTCTGTCGTAGAAACTGCAAGAAAGTGTTATCACTGGAAAAGACTCTTGAACAAGGAAAAACAAAA 1463
Gln-Ala-Tyr-Gln-Ala ***                                                                                                  455

AATAACATGTTCAAATTTTTTTGTGCTCTTTTATGAATTTTTTTTCTTCAAATTTTTATTTTAAATATTGAGGCAAAATGGTTTGTCGGAATAGAATAAAAGTATTTTCTATTTGGTTG 1582

TTTATTTTCGAAAGAGATGAAAAAAAAAAAAAAAA
```

Fig.2. Nucleotide sequence of cDNA encoding octopus rhodopsin and corresponding amino acid sequence. Amino acid sequences determined by peptide analysis are underlined.

In the invertebrate photoreceptors some components of the amplification cascade were characterized. Octopus rhodopsin was shown to interact with mammalian transducin [6]. This implies that a binding site for transducin should be conserved in both invertebrate and vertebrate opsins. Close homology of cytoplasmic loops in octopus and *Drosophila* opsins is indicative of transducin interacting with one or more of these loops.

There is no obvious homology between the C-terminal sequences of mammalian, *Drosophila* and octopus rhodopsins. However, octopus rhodopsin, like other counterparts, contains several threonine and serine residues in this region and thus meets the important requirement of being phosphorylated in a light-dependent manner [14]. Another interesting observation which distinguishes octopus rhodopsin from other pigments is the presence in the C-terminal region of a large number of prolines and glutamines clustered into eight blocks of repetitive sequences (Pro-Pro-Gln-Gly) interrupted by dipeptides or single amino acids. Whether this unusual sequence together with extended cytoplasmic loops contributes to the thermal stability of octopus rhodopsin remains to be seen.

Recently it was shown that two adjacent cysteines 322–323 in bovine rhodopsin are modified with palmitate [15]. The two cysteines 337–338 are also conserved in the C-terminal tail of octopus

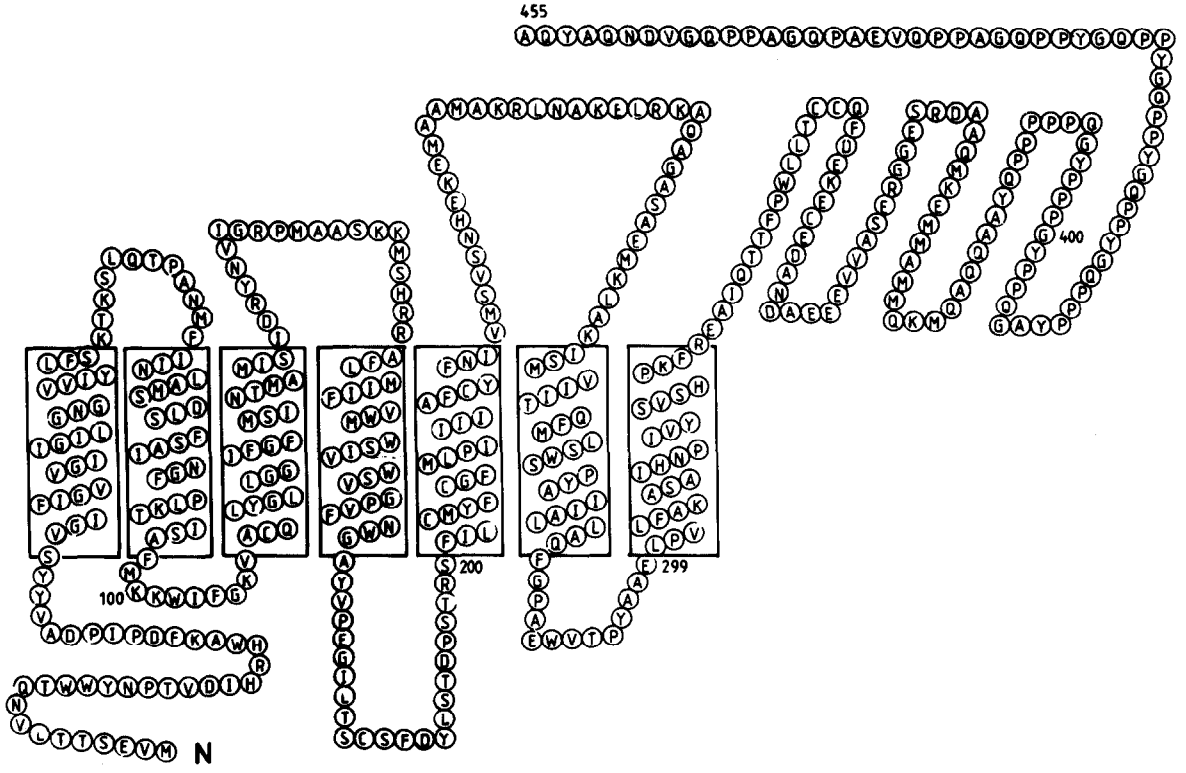

Fig.3. The transmembrane topology of octopus rhodopsin.

rhodopsin. The conservation of from one to three cysteines in this region of all known visual pigments poses an interesting question about their involvement in posttranslational fatty acid acylation and the functional implications of this modification. Being available in large quantities octopus rhodopsin provides additional advantages for such investigations.

Rhodopsin is not known to interact with extracellular factors. However, what is intriguing is the remarkable homology of connectivity between the IV-V transmembrane segments in all the visual pigments. Whether this high degree of conservation is necessary simply for the structural stability of rhodopsin or is of some still unknown functional importance remains to be explained.

REFERENCES

[1] Ovchinnikov, Yu.A. (1982) FEBS Lett. 148, 179–191.
[2] Hargrave, P.A., McDawell, J.H., Curtis, D.R., Wang, J.K., Jusczak, E., Fong, S.L., Mohanna Rao, J.K. and Argos, P. (1983) Biophys. Struct. Mech. 9, 235–244.
[3] Applebury, M.L. and Hargrave, P.A. (1986) Vision Res. 26, 1881–1895.
[4] Tsuda, M. (1979) Biochim. Biophys. Acta 545, 537–546.
[5] Hamdorf, K. (1979) in: Handbook of Sensory Physiology, vol.VII/6A, pp.145–224, Springer, Berlin.
[6] Tsuda, M., Tsuda, T., Terayama, Y., Fukada, Y., Akino, T., Yamanaka, G., Stryer, L., Katada, T., Ui, M. and Ebrey, T. (1986) FEBS Lett. 198, 5–10.
[7] Ovchinnikov, Yu.A., Abdulaev, N.G., Bogachuk, A.S. and Morris, C.A. (1987) in: Methods in Protein Sequence Analysis (Walsh, K. ed.) pp.189–209, Human Press, Clifton, NJ.
[8] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[9] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[10] Henikoff, S. (1984) Gene 28, 351–359.
[11] Zuker, C.S., Cowman, A.F. and Rubin, G.M. (1986) Cell 40, 851–858.
[12] Honig, B., Dinur, U., Nakanishi, K., Bologh-Nair, V., Gawinowicz, M.A., Annaboldi, M. and Motto, M.G. (1979) J. Am. Chem. Soc. 101, 7084–7086.
[13] Kakitani, H., Kakitani, T., Rodman, H. and Honig, B. (1985) Photochem. Photobiol. 41, 471–479.
[14] Vandenberg, C. and Montal, M. (1984) Biochemistry 23, 2347–2352.
[15] Ovchinnikov, Yu.A., Abdulaev, N.G. and Bogachuk, A.S. (1988) FEBS Lett. 230, 1–5.