ELSEVIER

FEBS Letters 342 (1994) 7–11

FEBS 13825

FEBS LETTERS

Hypothesis

Two hypotheses – one answer

Sequence comparison does not support an evolutionary link between halobacterial retinal proteins including bacteriorhodopsin and eukaryotic G-protein-coupled receptors

Jörg Soppa*

Max-Planck-Institut für Biochemie, Am Klopferspitz, 82152 Martinsried, Germany

Received 21 December 1993; revised version received 8 February 1994

Abstract

The structure of bacteriorhodopsin (BR) of *Halobacterium halobium* is known. Despite the lack of sequence similarities it is often taken as a model for eukaryotic G-protein-coupled receptors (GPCRs). Recently two hypotheses were used to support the homology of BR and GPCRs, namly evolution by exon shuffling and evolution by gene duplication. BR is a member of a family of halobacterial retinal proteins. The sequences of eight members of this family were used to test the two hypotheses. Based on sequence comparison, no indication for an evolutionary linkage between the two protein families could be found.

Key words: G-protein-coupled receptor; Bacteriorhodopsin; Halorhodopsin; Sensory rhodopsin; Retinal protein

1. Introduction

Bacteriorhodopsin (BR) is one of the best studied membrane proteins (for a review see [1]). Furthermore, it is one of only a few membrane proteins for which the structure could be determined [2]. Therefore, it has often been taken as a model protein for the family of G-protein-coupled receptors (GPCRs) [3]. The proteins of this family, like BR, are believed to contain seven membrane-spanning α-helices, but the approach of taking BR as a model is severely hampered by the fact that there are no sequence similarities between BR and GPCRs and thus it is not at all clear whether or not these proteins are homologous.

It is sometimes taken as an argument in favor of a common ancestor that a subset of GPCRs, the rhodopsins, have the chromophore, retinal, covalently bound to a lysine residue, as is also found in BR. However, the retinal is bound in different conformations in both cases, i.e. *all-trans* retinal is bound in BR and 11-*cis* retinal in the rhodopsins, and also the relative conformation of the retinal ring to the polyene chain is different. Thus the fact that retinal is bound can be explained by convergent evolution as well as homology.

Recently two different hypotheses have been put forward which try to explain the low sequence similarities found thus far and which both argue in favor of an evolutionary linkage of BR and GPCRs [4,5]. One hypothesis infers that exon shuffling events have occurred during the evolution of BR and GPCRs from a common ancestor and that therefore the homologous helices of the existing proteins are not collinear, e.g. that helix 7 of BR is homologous to helix 3 of the GPCRs [4]. The second hypothesis includes a gene duplication event in the evolution of the common ancestor of BR and GPCRs [5]. Thus the helices 1–3 were once identical to the helices 5–7 and the two lineages evolved by chance in a way that the highest degree of similarity in the recent proteins is found between e.g. helix 7 of BR and helix 3 of the GPCRs. In both hypotheses, BR and GPCRs are homologous and thus they justify the use of the BR structure as a template for modeling GPCRs, but it has to be done in different ways.

Bacteriorhodopsin is not the only retinal protein found in Halobacteria. They also possess a light-driven

*Corresponding author. Fax: (49) (89) 8578 3557.
E-mail: MX%"Soppa@vms.biochem.mpg.de"

SSDI 0014-5793(94)00210-M

chloride pump, halorhodopsin (HR), and two proteins involved in phototaxis, sensory rhodopsins I and II (SR I and SR II).

The polypeptide moieties of these proteins are named bacterio-opsin (Bop), halo-opsin (Hop) and sensory opsin (Sop), the chromoproteins including the chromophore retinal BR, HR and SR. This terminology will be followed throughout the rest of this paper. During the last years, the primary structures of Bop, Hop and Sop I of different halobacterial species have been determined, so that today 10 non-identical sequences and a number of partial sequences are known [6–13]. These proteins are clearly homologous based on the similarities of the primary sequences as well as biochemical and biophysical properties [13,14]. Therefore, a hypothesis about the relationship of BR to the GPCRs is in fact a hypothesis about the relationship of two protein families. Here the two above-mentioned hypotheses are tested using not BR alone but the whole retinal protein family.

2. Evolution of the common ancestor of BR and GPCRs involving gene duplication

Taylor and Agarwal propose that a primordial gene coding for a protein with three or four helices is duplicated partially or in total, giving rise to a translational fusion of the tandem genes with a gene product of seven helices. Their hypothesis is based mainly on two facts: (i) that helices 1 and 5 of Bop of *Halobacterium halobium* and helices 1 and 5, 2 and 6, and 3 and 7 of the GPCRs show some similarity, and (ii) that, if the first half of Bop is aligned to the second half, then helices 1–3 are in register with the helices 5–7.

Taylor and Agarwal state that 'one of the most striking matches is in BR itself, between H1 and H5, which have an impressive 43% identity and 70% similarity, with one deletion'. To test whether this surprising similarity is a solid basis for proposing a gene duplication, the same alignment of helices 1 and 5 was done for the other members of the halobacterial retinal protein family. The result is summarized in Table 1. As can be seen, Bop of *H. halobium* is an extreme case, the other extreme being Sop I of *H. halobium* with 0% of identical amino acids. The average degree of identity is 16% which does not seem to be significant for a comparison of hydrophobic helices. That the striking similarity of 42% for the case of *H. halobium* Bop is incidental seems quite feasible if it is realized that (i) a gap has to be introduced to reach this alignment, (ii) the '42%' are equal to only 10 amino acids and (iii) that 5 of the 10 matches are matches of alanine or leucine which together comprise 28% of the retinal proteins [13].

The second argument in favor of a gene duplication was that in an alignment of the first to the second half of *H. halobium* Bop the helices are in register. The alignment shown by Taylor and Agarwal includes 9 gaps in a length of 130 amino acids. They also state that, if the alignment parameters are varied, different results can be obtained. By changing the 'gap weight' and the 'gap length weight' systematically over a wide range it was found that the alignment, which is most insensitive against parameter variation, includes 4 gaps and a 7 amino acid offset of helices 3 and 7. Therefore I feel that the alignment of the two protein halves, which both contain 3 hydrophobic helices and rather small loops, might not be a suitable method to investigate a possible gene duplication event.

As an alternative, helix 7 of all members of the retinal protein family was aligned to the rest of the respective proteins to see whether a systematic similarity to a specific part can be found which would argue in favor of a gene duplication. The same was repeated for helices 1 and 2 (in this case to the rest of the proteins excluding helix 1). The most similar regions and the overlapping helices are tabulated in Table 2. As can be seen, helix 1 is most similar to helices 3, 4, 5 or 7, depending on the retinal protein used. Furthermore, in all cases the quality of the alignment of the actual helix is not better than of the helix after random shuffling of the sequence (data not shown). There is no systematic similarity of one of the three helices investigated to another part of the proteins. Thus it can be concluded that using the information content of the halobacterial retinal protein family, no indication for a gene duplication event could be detected.

3. Evolution of BR and GPCRs from a common ancestor involving exon shuffling

Pardo et al. aligned any helix of Bop to any helix of GPCRs, not allowing gaps. They found similarities between Bop helix 7 and GPCR helix 3, Bop helix 3 and GPCR helix 5, Bop helix 1 and GPCR helix 7. To test whether or not these similarities were significant they aligned each of the three helices of Bop with the respective helices of 17 GPCRs using the program, Gap [15].

Table 1
Similarities of helices 1 and 5 of halobacterial retinal proteins

	Bop	Hop	Sop I
H. h.	42%	8%	0%
H. SG1	25%	13%	8%
H. aus2	13%		
N. p.		17%	

The alignments of helices 1 and 5 were done as from Taylor and Agarwal for the respective helices of bacterio-opsin of *H. halobium* [5]. The fractions of identical amino acids are tabulated. Bop, bacterio-opsin; Hop, halo-opsin; Sop I, sensory opsin I; *H. h.*, *Halobacterium halobium*; *H. SG1*, *Halobacterium* spp. *SG1*; *H. aus2*, *Halobacterium* spp. *aus2*; *N. p.*, *Natronobacter pharaonis*. The sequences (top-to-bottom, left-to-right) were taken from [6], [13], [11], [7], [13], [10], [8], [13].

For all alignments the 'quality of the alignment' (Q) and the 'average quality of 100 alignments in which one of the sequences was randomly shuffled' (Q_{random}) were tabulated. In each case they found that Q was higher than Q_{random}. Thus they concluded that the similarity was significant and that Bop and GPCRs are homologous, but not in a collinear fashion. Instead, exon shuffling events had taken place in the evolution of Bop and GPCRs from a common ancestor leading to a different sequence of homologous membrane-spanning helices.

Again, only Bop of *H. halobium* was taken for the comparison and the method used might not be suitable to solve the question. To make the latter clear, I would like to introduce some alignments of model peptides (compare Table 3). The peptides of each alignment contained only two amino acids, having the same composition (50% of each amino acid) but a different sequence. Peptides consisting of alanine and leucine (AL-alignment), valine and leucine (VL-alignment), alanine and tryptophane (AW-alignment) and alanine and aspartic acid (AD-alignment) were used. Each alignment yielded 70% 'matches' and 30% 'mismatches'. The 'significance' of the alignments was also checked by comparing the quality of the alignment with the quality after randomizing the second sequence 100 times. The parameter 'quality' is calculated by alignment programs by adding up values of 'similarity' or 'exchangeability' for each pair of amino acids in the alignment. Different approaches have been undertaken to quantitate how 'similar' or 'exchangeable' any possible pair of amino acids is and the results are tabulated as so-called 'scoring matrices' [16]. The matrix which is most often used and is supplied with the GCG software package as the default matrix is the 'mutation data matrix' of Dayhoff and colleges ([17]; for discussion of the matrix see [16]). It was calculated on the basis of 1,600 mutations which were found in 71 groups of proteins containing proteins of less than 15% difference. Basically, it represents mutation frequencies of families of soluble proteins known in 1979 which have not evolved far. Therefore it may be misleading if it is applied to the comparison of membrane proteins instead of soluble proteins or of proteins which are related only distantly.

A second scoring matrix supplied with the GCG program package is the 'structure data matrix' based on amino acid replacements observed in 32 proteins of known structure [18], but also here only soluble proteins could be taken. These two scoring matrices as well as a self-constructed 'hydrophobicity matrix' were used with the program, Gap, to investigate the model alignments. The 'qualities' and 'random qualities' are tabulated in Table 3. As can be seen, the results differ substantially. The AL-alignment is clearly significant if the mutational data matrix is used, slightly significant on the basis of the structure data matrix and not significant on the basis of the hydrophobicity matrix. In contrast, the VL-alignment is slightly significant according to the mutational data matrix and not significant on the basis of the other two matrices. The AW-alignment, again, can be interpreted very differently depending on which matrix is used, and only the AD-alignment is significant regardless of the matrix. For membrane-spanning helices with most of the amino acids in contact with the lipid phase, alanine–leucine mismatches and valine–leucine mismatches seem to be equivalent and even alanine–tryptophan mismatches are energetically not too unfavored, but the

Table 2
Similarities between helices 1, 2 and 7 and the remaining part of the respective halobacterial retinal proteins

		Used for the search					
		Helix 1		Helix 2		Helix 7	
		Most similar		Most similar		Most similar	
		Region,	Helix	Region,	Helix	Region,	Helix
H. h.	Bop	148–170	5	126–150	4	61–85	2
H. SG1	Bop	217–239	7	196–220	6 (7)	20–44	1
H. aus2	Bop	146–168	5	177–201	6	18–42	1
H. h.	Hop	111–133	3	115–139	3 (4)	29–53	1
H. SG1	Hop	165–187	5	121–145	3 (4)	144–166	4
N. p.	Hop	175–197	5	146–170	4	31–55	1
H. h.	Sop I	181–203	7	189–213	7	5–29	1
H. SG1	Sop I	99–121	4	153–177	4	6–30	1

The alignments were done with the program, Gap [15]. The helix boundaries were taken from the structural model of *H. halobium* BR [2] and the alignment of the retinal proteins [13]. The helices were aligned to the remaining C-terminal (helices 1 and 2) or N-terminal (helix 7) sequence of the helix used. No gaps were allowed by choosing a high gap weight. The 'most similar regions' as well as the helices overlapping these regions are tabulated. If a helix is most similar to a region including parts of 2 helices and the connecting loop, the helix with the shorter overlap is set in brackets. The alignments need not be reciprocal, e.g. helix 7 can align to a region overlapping helix 1 and at the same time helix 1 to a region overlapping helix 3. 100 'random runs' (alignments after randomly shuffling the sequence of the helix) were included in the alignments. Sequences and abbreviations are the same as in Table 1.

Table 3
Significance of model alignments

Scoring matrix	Mutational data		Structure		Hydrophobicity	
	Q	Q_{random}	Q	Q_{random}	Q	Q_{random}
AAAAAAAAAALLLLLLLLL ::::::: ::::::: AAAAAAALLLLLLLLLAAA	20.4	14.7 ± 3.6	38.6	34.9 ± 2.0	194	189.6 ± 2.2
VVVVVVVVVVLLLLLLLLL ::::::: ::::::: VVVVVVVLLLLLLLLLVVV	25.8	23.1 ± 1.7	42.8	42.0 ± 0.5	194	190.1 ± 2.3
AAAAAAAAAAWWWWWWWWW ::::::: ::::::: AAAAAAAWWWWWWWWWAAA	16.2	6.6 ± 5.1	25.4	12.9 ± 7.4	194	190.4 ± 2.4
AAAAAAAAAADDDDDDDDD ::::::: ::::::: AAAAAAADDDDDDDDDAAA	22.8	17.9 ± 2.5	32.0	24.1 ± 4.5	158	132.5 ± 14.4

The alignments shown on the left were calculated using the program, Gap [15]. No gaps were allowed due to a high gap weight and the 'end-weighting' of gaps. The qualities (Q) and the average qualities of 100 alignments after random shuffling of the second sequence (Q_{random}) were tabulated using three different scoring matrices as local data files, as indicated: the mutational data matrix [16,17] and the structure matrix [18] provided by the GCG program package, and a self-constructed hydrophobicity matrix based on the different amino acid hydrophobicities as tabulated by Engelman et al. [22].

three are treated totally differently by the mutational data matrix. These examples show that the calculated 'significance' of peptide alignments depends on the amino acid composition of these peptides and the scoring matrix used and is not a true measure of the evolutionary relatedness of these peptides, especially for helices of membrane proteins.

Table 4
Similarities of GPCR helix 5 to helices of halobacterial retinal proteins

		5HT1α	α2	β1	D1	M1
H. h.	Bop	5	6	5	5	5
H. SG1	Bop	5	6	6	6	6
H. Aus2	Bop	5	6	5	6	1
H. h.	Hop	7	1	1	5	7
H. SG1	Hop	5	1	1	7	4
N. p.	Hop	7	1	5	5	1
H. h.	Sop I	1	5	1	1	1
H. SG1	Sop I	1	5	6	1	1

The retinal protein helices which are most similar to helix 5 of five different GPCRs (see below) are tabulated. The GPCR helices were aligned to the halobacterial retinal proteins with the program, Gap [15]. To allow only helix-to-helix alignments, the loops of the retinal proteins were replaced by stretches of ten times the symbol, 'B', which was given a high negative value in the scoring matrix. The introduction of gaps was prohibited due to a high gap weight. Because the parameter which is optimized by the program, the alignment quality, depends on alignment length, a uniform length for the different helices were chosen. To lower the stringency of the helix alignment and thereby compensate for a possible non-exact helix prediction in the GPCRs, the helices of the retinal proteins were defined to consist of 27 amino acids and the GPCR helices to consist of 23 amino acids. The following GPCR sequences were chosen: 5HT1α, rat 5-hydroxytryptamine-1A receptor [23]; α2, human platelet α-2-adrenergic receptor [24]; β1, human β-1-adrenergic receptor [25]; D1, human D_1 dopamine receptor [26]; M1, human m1 (Hm1) muscarinic acetylcholine receptor [27]. The other abbreviations are as in Table 1.

Therefore a different approach was used to investigate whether a non-collinear relationship between helices of halobacterial retinal proteins and GPCRs exist which could be explained by exon shuffling. Using a GPCR helix as a 'probe' the most similar helix in halobacterial retinal proteins was determined with the program, Gap, allowing only helix-to-helix alignments on no gaps. Five different GPCRs and 8 halobacterial retinal proteins were used for the alignments. The tabulated results for the GPCR helix 5 is shown in Table 4. As can be seen, GPCR helix 5 is most similar to helices 1, 4, 5, 6 or 7, depending on the specific GPCR and retinal protein used for the alignment. Equivalent results were obtained using GPCR helices 3 and 7. Using only one protein of a family can be misleading, e.g. helix 3 of the muscarinic acetylcholine receptor is most similar to helix 1 of the halobacterial retinal proteins in 6 out of 8 cases, a result which is not obtained with helix 3 of the other receptors. Thus it can be concluded that there is no systematic similarity between helices of halobacterial retinal proteins and G-protein-coupled receptors.

4. Discussion

Using the sequences of 8 halobacterial retinal proteins no indication for an evolutionary relationship between this protein family and the family of eukaryotic G-pro-

tein coupled receptors could be detected. In previous studies bacterio-opsin had been compared to some members of the GPCR family without finding similarities [19] and the retinal protein family was used to search for related sequences in protein sequence databases without success [13]. Of course this lack of sequence similarties does not exclude the development of the two protein families from a common ancestor. A highly divergent evolution leading to the loss of sequence similarity between homologous proteins cannot be distinguished from a convergent development resulting in two unrelated protein families having the same structural principle of seven membrane-spanning helices and, in the case of the rhodopsins, retinal binding. Nevertheless, the consequences for practical purposes are the same. The structure of bacteriorhodopsin cannot be taken as a model for GPCR structure. Moreover, structure predictions using the sequences of about 200 GPCRs [20], as well as the determination of the rhodopsin structure to a limited resolution [21], indicate that the structures of the halobacterial retinal proteins and the GPCRs are different. Researchers working with halobacterial retinal proteins might like the idea that their work has an impact on another field with great medical importance, but they will have to live without this comfort. The question of whether or not the two protein families are related in evolution has to wait until high-resolution three-dimensional structural models of several members of both families are available.

References

[1] Tittor, J. (1991) Curr. Opin. Struct. Biol. 1, 534–538.
[2] Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E. and Downing, K.H. (1990) J. Mol. Biol. 213, 899–929.
[3] Trumppkallmeyer, S., Hoflack, J., Bruinvels, A. and Hibert, M. (1992) J. Med. Chem. 35, 3448–3462.
[4] Pardo, L., Ballesteros, J.A., Osman, R. and Weinstein, H. (1992) Proc. Natl. Acad. Sci. USA 89, 4009–4012.
[5] Taylor, E.W. and Agarwal, A. (1993) FEBS Lett. 325, 161–166.
[6] Dunn, R., McCoy, J., Simsek, M., Majumdar, A., Chang, S., RajBhandary, U. and Khorana, G. (1981) Proc. Natl. Acad. Sci. USA 78, 6744–6748.
[7] Blanck, A. and Oesterhelt, D. (1987) EMBO J. 6, 265–273.
[8] Blanck, A., Oesterhelt, D., Ferrando, E., Schegk, E. and Lottspeich, F. (1989) EMBO J. 8, 3963–3971.
[9] Sugiyama, Y., Maeda, M., Futai, M. and Mukohata, Y. (1989) J. Biol. Chem. 264, 20859–20862.
[10] Lanyi, J., Duschl, A., Hatfield, G. and May, K. Oesterhelt, D. (1990) J. Biol. Chem. 265, 1253–1260.
[11] Uegaki, K., Sugiyama, Y. and Mukohata, Y. (1991) Arch. Biochem. Biophys. 286, 107–110.
[12] Otomo, J., Urabe, Y., Tomioka, H. and Sasabe, H. (1992) J. Gen. Microbiol. 138, 2389–2396.
[13] Soppa, J., Duschl, J. and Oesterhelt, D. (1993) J. Bacteriol. 175, 2720–2726.
[14] Oesterhelt, D. and Tittor, J. (1989) Trends Biochem. Sci. 14, 57–61.
[15] Genetics Computer Group (April 1991) 575 Science Drive, Madison, Wisconsin, USA 53711.
[16] George, D., Barker, W. and Hunt, L. (1990) Methods Enzymol. 183, 333–351.
[17] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1979) Atlas of Protein Sequence and Structure, Vol. 5 (M.O. Dayhoff, ed.) pp. 345, National Biomedical Research Foundation, Washington, DC.
[18] Risler, J.L., Delorme, M.O., Delacroix, H. and Henaut, A. (1988) J. Mol. Biol. 204, 1019–1029.
[19] Henderson, R. and Schertler, G.F.X. (1990) Phil. Trans. Royal Soc. London B Biol. Sci. 326, 379–389.
[20] Baldwin, J. (1993) EMBO J. 12, 1693–1703.
[21] Schertler, G.F.X., Villa, C. and Henderson, R. (1993) Nature 362, 770–772.
[22] Engelman, D., Steitz, T. and Goldman, A. (1986) Annu. Rev. Biophys. Biophys. Chem. 15, 321–353.
[23] Albert, P.R., Zhou, Q.Y., Van Tol, H.H., Bunzow, J.R. and Civelli, O. (1990) J. Biol. Chem. 265, 5825–5832.
[24] Kobilka, B.K., Matsui, H., Kobilka, T.S., Yang-Feng, T.L., Francke, U., Caron, M.G., Lefkowitz, R.J. and Regan, J.W. (1987) Science 238, 650–656.
[25] Frielle, T., Collins, S., Daniel, K.W., Caron, M.G., Lefkowitz, R.J. and Kobilka, B.K. (1987) Proc. Natl. Acad. Sci. USA 84, 7920– 7924.
[26] Zhou, Q.Y., Grandy, D.K., Thambi, L., Kushner, J.A., van Tol, H.H., Cone, R., Pribnow, D., Salon, J., Bunzow, J.R. and Civelli, O. (1990) Nature 347, 76–80.
[27] Chapman, C.G. and Browne, M.J. (1990) Nucleic Acids Res. 18, 2129.