Review Letter

THE STRUCTURAL BASIS OF THE FUNCTIONING OF BACTERIORHODOPSIN: AN OVERVIEW

Yu. A. OVCHINNIKOV, N. G. ABDULAEV, M. Yu. FEIGINA, A. V. KISELEV and N. A. LOBANOV
Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, Moscow 117312, USSR

Received 12 February 1979

1. Introduction

Mass transport across biological membranes is a pressing problem of modern physicochemical biology, determining to a considerable extent our understanding of the regulatory and bioenergetic processes in the living cell. Latterly, as well as continued interest in the general aspects and theories of transport and in the study of model systems, one may clearly discern a tendency in this field towards detailed investigation of the proteins directly associated with translocation of ions and molecules, independently of whether in the membrane such proteins serve as carriers, channels or receptors. Such a membrane protein, one of the most interesting and extensively explored, undoubtedly is bacteriorhodopsin.

Bacteriorhodopsin, the only protein of the purple membrane of halophilic bacteria which thrive under extremal conditions and effectively utilize light energy in the performance of their vital activity [1], has been found to function as a light-driven proton pump, establishing a considerable pH gradient in the membrane, utilized in the synthesis of ATP [2].

Despite available source and relatively small size, bacteriorhodopsin has long foiled attempts to elucidate its structure by numerous protein chemists and physicists, and only very recently has a way been found to decode its amino acid sequence and elucidate the chemical nature of its functionally important groups [3,17]. These recent studies have also given the first indications of how the molecule is packed in the membrane.

The present paper is a summing up of our studies, begun in 1976, that have resulted in the complete structural elucidation of bacteriorhodopsin (fig.1), the first of the true membrane proteins for which this has been successfully accomplished. The results have opened wide perspectives for further work on the mode of action of this membrane-located proton pump and the approach can be of help in the studies of other membrane proteins.

In setting up the strategy for the structural analysis of bacteriorhodopsin one had to take into account the unusual properties of this protein. As is well known [4] bacteriorhodopsin repeatedly spans the purple membrane and is therefore embedded within the lipid phase. Hence the high affinity of this protein for lipids and its relative 'hydrophobicity', characteristic of the majority of integral membrane proteins. We, therefore, had reason to expect that the isolation of bacteriorhodopsin and its structural study could hardly be based on the standard arsenal of methods of protein chemistry. This circumstance as well as the microbial nature of bacteriorhodopsin first gave us the idea of using, for the structural study of the protein, a now popular procedure of rapid sequencing of the appropriate DNA fragment especially as we had available the methodology applicable to *Halobacterium halobium*.

However, we accepted an alternative approach, namely, direct study of the protein since, in addition to the primary structure, it also provides valuable information on the reactivity and other properties of the whole molecule and its fragments, on the location of the active center, polypeptide folding, etc., all these data being directly relevant to the structure–function relationship of the protein. In other words we believe that 'reading-out' of the protein sequence

```
                                                                    20
GLU-ALA-GLN-ILE-THR-GLY-ARG-PRO-GLU-TRP-ILE-TRP-LEU-ALA-LEU-GLY-THR-ALA-LEU-MET-
                                                                    40
GLY-LEU-GLY-THR-LEU-TYR-PHE-LEU-VAL-LYS-GLY-MET-GLY-VAL-SER-ASP-PRO-ASP-ALA-LYS-
                                                                    60
LYS-PHE-TYR-ALA-ILE-THR-THR-LEU-VAL-PRO-ALA-ILE-ALA-PHE-THR-MET-TYR-LEU-SER-MET-
                                                                    80
LEU-LEU-GLY-TYR-GLY-LEU-THR-MET-VAL-PRO-PHE-GLY-GLY-GLU-GLN-ASN-PRO-ILE-TYR-TRP-
                                                                   100
ALA-ARG-TYR-ALA-ASP-TRP-LEU-PHE-THR-THR-PRO-LEU-LEU-LEU-LEU-ASP-LEU-ALA-LEU-LEU-
                                                                   120
VAL-ASP-ALA-ASP-GLU-GLY-THR-ILE-LEU-ALA-ILE-VAL-GLY-ALA-ASP-GLY-LEU-MET-ILE-GLY-
                                                                   140
THR-GLY-LEU-VAL-GLY-ALA-LEU-THR-LYS-VAL-TYR-SER-TYR-ARG-PHE-VAL-TRP-ALA-ILE-SER-
                                                                   160
THR-ALA-ALA-MET-SER-TYR-ILE-LEU-TYR-VAL-LEU-PHE-PHE-GLY-PHE-THR-SER-LYS-ALA-GLU-
                                                                   180
SER-MET-ARG-PRO-GLU-VAL-ALA-SER-THR-PHE-LYS-VAL-LEU-ARG-ASN-VAL-THR-VAL-VAL-LEU-
                                                                   200
TRP-SER-ALA-TYR-PRO-VAL-VAL-TRP-LEU-ILE-GLY-SER-GLU-GLY-ALA-GLY-ILE-VAL-PRO-LEU-
                                                                   220
ASN-ILE-GLU-THR-ALA-LEU-PHE-MET-VAL-LEU-ASP-VAL-SER-ALA-LYS-VAL-GLY-PHE-GLY-LEU-
                                                                   240
ILE-LEU-LEU-ARG-SER-ARG-ALA-ILE-PHE-GLY-GLU-ALA-GLU-ALA-PRO-GLU-PRO-SER-ALA-GLY-
ASP-GLY-ALA-ALA-ALA-THR-SER
```

Fig.1. Complete amino acid sequence of bacteriorhodopsin.

from the structure of the corresponding gene deprives us of important information, perhaps even more important than the primary structure as such. Our structural work with bacteriorhodopsin provided a good example for the above statement, since it yielded valuable knowledge of its topography and mode of action.

2. Results and discussion

In the course of the structural analysis of bacteriorhodopsin use was made of splitting the polypeptide chain by chemical means in media effectively solubilizing the delipidized protein. When bacteriorhodopsin was subjected to cyanogen bromide treatment, 10 fragments were identified, 6 of which were obtained in the individual state after gel filtration on Sephadex. Determination of the amino acid sequence of these peptides showed that they comprised the following parts of the polypeptide chain: the N-terminal fragment 1–20 with a pyroglutamic acid residue [5], the fragments 21–32 and the retinal (on Lys 41 [6]) carrying fragment 33–56, the fragments 57–60 and 61–68 and finally the C-terminus (209–247). Attempts to isolate the large fragments 69–118 and 163–208 by all available procedures met with no success. The structure of these peptides was determined by direct analysis of their mixtures on a sequencer in combination with analysis of short peptides obtained by cleavage of large fragments.

A small amount (~100 nM) of peptide 119–144

could be isolated in the individual state from the cyanogen bromide cleavage products of bacteriorhodopsin when the reagent was in large excess (500 mol BrCN/mol Met) and its structure was established by tryptic hydrolysis, and also by automatic analysis in a mixture with the N-terminal fragment. Isolation of peptide 145–162 was achieved by exhaustive cleavage of the protein but, owing to the resistant Met–Ser bond, the amount obtained was very small and, moreover, it was contaminated with the peptide 1–20; however, we were able to determine its sequence by means of the procedure described in [7]. The structure and location of fragment 145–162 were substantiated indirectly by data on the sequence of the fragment 159–171, obtained by tryptic degradation of bacteriorhodopsin in the form of a fine suspension of the delipidized protein in a saturated solution of guanidinium chloride from which the latter had been removed by dialysis.

In order to determine the sequence in the chain of the cyanogen bromide fragments the protein was cleaved at the tryptophan residues by means of BNPS-skatole. The results made possible reassembly of the entire polypeptide chain. The BNPS-skatole [8] cleavage yielded a mixture of peptides from which only the fragment 13–80 could be isolated in the individual state. But analysis of its structure made possible not only establishment of the sequence of the 6 cyanogen bromide fragments but also reassembly of the protein molecule from the N-terminus to the 119th residue.

Separation of the other large BNPS-skatole fragments of bacteriorhodopsin was carried out chromatographically on Sephadex in 8 M guanidinium chloride solution. Determination of the N-terminal sequence of the fragment 189–247 by automatic degradation on a sequencer revealed the location of the cyanogen bromide fragments 163–208 and 209–247.

An important contribution to the complete amino acid sequence of bacteriorhodopsin was made by study of the action of papain on the native purple membrane. Sequencing of the resultant large peptides 4–65 and 73–230 independently confirmed the correctness of reassembly of the polypeptide chain based of the earlier isolated short fragments. The results showed that the bacteriorhodopsin molecule consisted of 247 amino acid residues (mol. wt 26 534) of the following composition: Ala-30, Arg-7, Asn-3, Asp-9, Gln-2, Glu-10, Gly-25, Ile-15, Leu-34, Lys-7, Met-9, Phe-13, Pro-11, <Glu-1, Ser-14, Thr-18, Trp-7, Tyr-11, Val-21.

The complete primary structure of the protein is represented in fig.1. Noteworthy is the high content of 'hydrophobic' amino acids (~66% of the total number and their 'clustering' in definite regions of the chain (e.g., Leu–Val–Pro–Ala–Ile–Ala–Phe–, Pro–Leu–Leu–Leu–Leu–, Leu–Ala–Leu–Leu–Val–, Ile–Leu–Ala–Ile–Val–Gly–Ala–, Val–Leu–Phe–Phe–Gly–Phe–, Val–Gly–Phe–Gly–Leu–Ile–Leu–Leu–).

Of considerable interest also is the anomalously high number of tryptophan residues (7) in such a small protein, and that bacteriorhodopsin lacks a free N-terminal α-amino group, the N-terminal residue being a pyroglutamic acid. In addition, judging from its amino acid composition, bacteriorhodopsin should be classified among the 'acidic proteins'. Finally the Lys-41 residue whose ε-amino group is Schiff-bonded to the retinal chromophore, lies very close to the N-terminus of the protein.

In the course of the complete amino acid sequence determination of bacteriorhodopsin, a number of highly important findings were made as to the topography of the molecule in the membrane. As already indicated, a decisive part was played here by experiments on the limited proteolysis of the protein by 'non-penetrating' reagents such as papain and α-chymotrypsin. As is well known, the purple membrane is a quite dense semi-crystalline formation with more-or-less rigid, well-defined mutual disposition of its protein and lipid components [9]. One could thus expect that relatively mild treatment of the purple membrane by proteolytic enzymes should lead to splitting only in the regions exposed to the exterior (if such are present) whereas the inaccessible regions of the protein deeply embedded in the membrane would withstand attack. In fact, short time (2 h) treatment of the purple membrane with papain (at an enzyme/bacteriorhodopsin ratio of 1:200) led to the cleavage of 17 amino acid residues from the C-terminus of the protein ([3,10] cf. [11]), showing that the fragment consisting of the last 17 residues should be in or very near the outer space. More drastic conditions (enzyme/bacteriorhodopsin ratio 1:20 and increase in incubation time to 24 h) led to splitting

of the Gly–Ile (3–4) and Gly–Gly (72–73) bonds, as well. A more thorough analysis of the degradation products revealed that not only the Gly–Gly (72–73) bond was affected, but the entire short region of the chain between the residues Gly-65 and Gly-73, so that as a result of the papain treatment, the short peptides Leu–Thr and Val–Pro–Phe–Gly [12] pass into solution. From this one may conclude that there must be some kind of peptide 'loop' on the membrane surface. The ready accessibility of this section to proteolysis is confirmed also by the fact that when the purple membrane is treated with α-chymotrypsin under a variety of conditions splitting of the Phe–Gly (71–72) bond always takes place.

Prolonged (7 days with an enzyme : substrate ratio of 1:20) papain digestion of the purple membrane leads to cleavage of yet another peptide bond; namely Ser–Met (161–162). It can thus be seen that the exposed segments of bacteriorhodopsin in the native purple membrane are 1–4, 65–73, 161–162 and 230–247, in all 13% of the molecule.

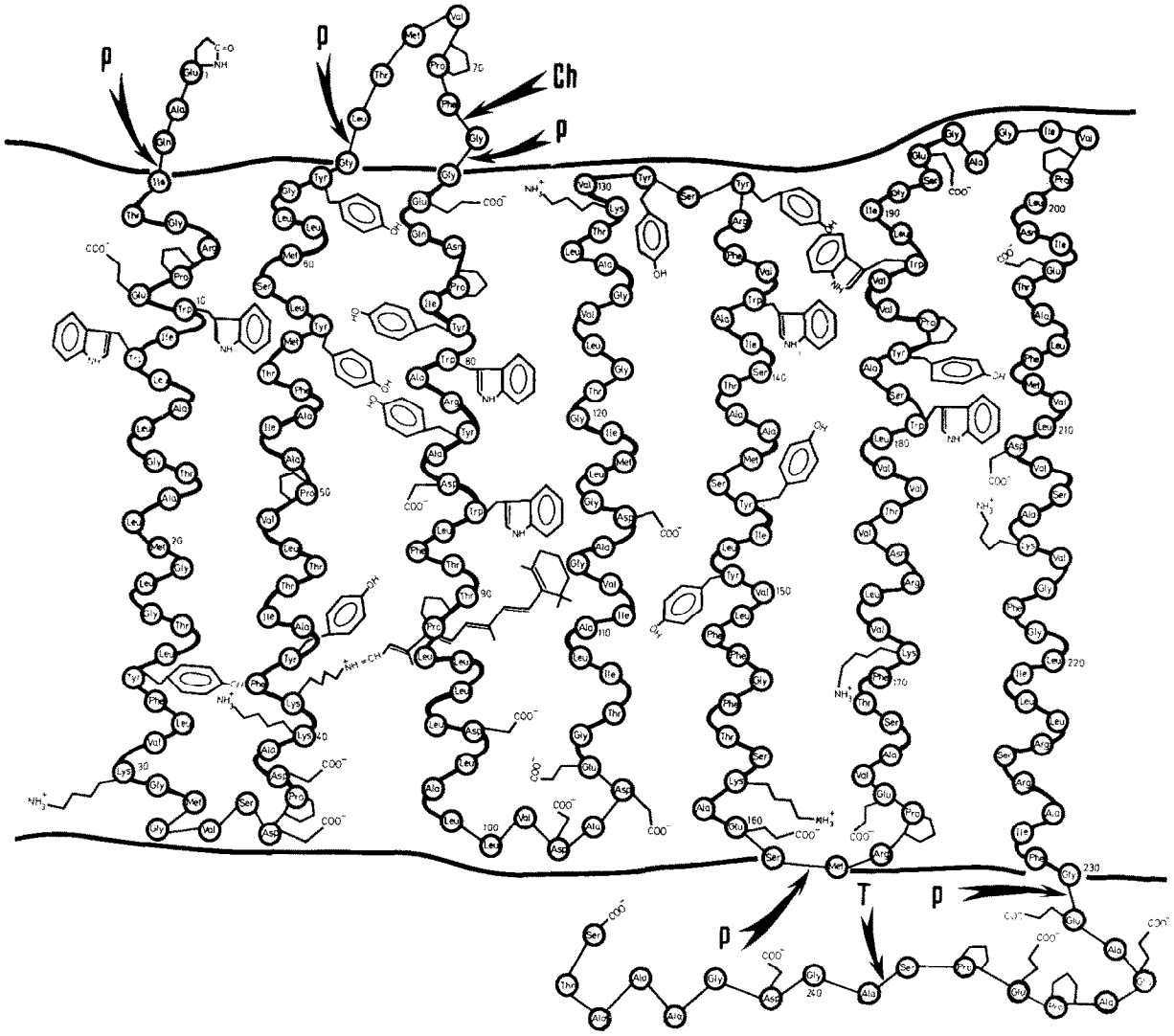
It is noteworthy that the purple membrane preparations digested by the proteolytic enzymes, although devoid of the N- and C-termini of bacteriorhodopsin (1–3 and 231–247, respectively) and split into a number of fragments at various bonds, yet retain almost completely their biological activity. Thus, embedded in the proteoliposome and incorporated into a planar bilayer [12] they maintain their ability to effectively 'pump' protons on a level with the control purple membrane specimens containing intact bacteriorhodopsin. This means that apparently even partially 'cut up', bacteriorhodopsin must retain the folding in the membrane which determines its functioning as a proton pump, and that a number of peptide fragments and bonds in this molecule are not essential for manifestation of its function. At present we are investigating the question of the ability of the retinal carrying fragment 4–65, the 'heart' of the molecule, to function independently, without the rest of the molecule, when embedded in the membrane. What is the role of the other segments of the polypeptide chain, within the membrane and on its surface? The answer to these questions will undoubtedly give us a deeper understanding of the structure–function relation in bacteriorhodopsin and thereby will shed more light on its mode of action.

Let us see just what is the folding of the peptide chain of bacteriorhodopsin in the light of the evidence now available. In conformity with Henderson's model [9] the predominant pattern should comprise 7 α-helical segments or rods spanning the membrane normally to its surface. If one subtracts the 20 amino acid residues situated at the N- and C-termini outside the membrane, each rod should have on an average 32 residues. However, we now know that 7 more amino acid residues should be on the outside forming the loop 65 73, and, moreover 3–4 residues should go to form each of the 6 connections between the α-helices. As a result each rod should then contain on an average 29–30 amino acid residues and the general scheme of the bacteriorhodopsin packing in the membrane should be that shown in scheme 1.

The scheme is the result not only of the above general considerations but, as can readily be seen, is based also on the limited proteolysis experiments. The results of these experiments have led to correction of the former proposal of the bacteriorhodopsin packing in the membrane [3] such that the loop 65–73 is now on the same side of the membrane as the N-terminus. Here, of particular value was the information obtained from chymotryptic and papain action on the vesicles obtained from *H. halobium* according to Blaurock et al. [13] with the correct, i.e., natural orientation of the membrane. Electron microscopic examination showed that in this case ⩾90% of the vesicles are closed and that their morphology is not affected by the enzymic treatment. Comparison of the rate of chymotryptic splitting of the 71–72 bond for the vesicle preparations and the native membranes under identical conditions together with the results of analysis of the cleavage of 17 C-terminal amino acid residues for the same specimens led to the conclusion that the 71–72 bond and the C-terminus of bacteriorhodopsin are situated at opposite sides of the membrane: the C-terminus being within the vesicle (see also Gerber et al. [11]) the bond 71–72 on its surface. As a result we have proposed a model of the membrane shown in scheme 1.

What are the principal features of the model? First, all its 7 rods are in the α-helical configuration in conformity with the data of Henderson [9] and Long [14]. The residue Lys-41 which carries the chromophore is in the second rod and closer to the cytoplasmic side of the membrane. The residue seems to be quite randomly oriented, although one should bear

Scheme 1
Disposition of bacteriorhodopsin polypeptide chain in the membrane



External surface above, cytoplasmic surface below. Arrows labelled Ch, P and T show points of attack by chymotrypsin, papain and trypsin, respectively, on intact purple membranes

in mind recent indications [15] of its possible interaction with a tryptophan (Trp-86) residue.

Noteworthy is the marked asymmetry of the positions of the charged functional groups of bacteriorhodopsin. Most basic amino acid residues (Lys and Arg) are close the surface of the membrane and can interact with negatively charged phosphate and sulfate groups of lipids, thereby stabilizing the lipoprotein structure. Another stabilizing factor can be the formation of ion pairs between the oppositely charged groups well in the membrane such as Arg-82, Asp-85, Asp-115, Lys-171, Arg-174, Asp-211, Lys-215. A still greater factor is in the location of the negatively charged carboxylate groups of bacteriorhodopsin

(Asp and Glu residues, see scheme 1), most of them being close to the cytoplasmic surface of the membrane. At the same time practically all the tryptophan residues are, contrariwise, much closer to the external surface and can stabilize the protein structure not only by hydrophobic interaction, but also can facilitate its functioning by forming a π-electron-saturated 'shield'. In general, the structural asymmetry of bacteriorhodopsin in the membrane is an important factor in establishing the required 'vectorality' of the active proton transport.

The structural elucidation of bacteriorhodopsin does not, of course, permit an unequivocal hypothesis as to its mode of action, but it provides a basis for the proposal of working hypotheses. Thus, of significance in this respect is that the path of the proton from the protonated Schiff base to the outside and from the cytoplasm to the azomethine grouping takes place along the peptide chain which forms a sort of well, the walls of which are strewn with HO groups. This shows that the tyrosine, serine and threonine side chains play a significant part in proton transport forming a channel of a continuous chain of hydrogen bonds as postulated by Morowitz [16]. In any case such a mechanism could provide the required rate of proton transport and also could explain the effect of the overall spatial structure of its charges on the transport process. In this light the role of the protein's surface COO^- groups could be (might be) that of possible regulators of the second stage of proton transport through changes in their pK values. The lysine-retinal residue in conformity with this hypothesis could fulfill the role of an energy trigger participating in the proton transport chain through light-induced changes in its stereochemistry and donor–acceptor properties.

But, whatever the mode of action the structural elucidation of bacteriorhodopsin has provided a reliable probe for the functional study and a guideline as well.

What perspectives have been opened up by elucidation of the bacteriorhodopsin structure? The model we have proposed of the packing of the bacteriorhodopsin molecule in the membrane shows the way towards the complete determination of its tertiary structure. We are at present carrying out the selective modification of the functionally important residues of the protein in order to ascertain the concrete role of each in the transport process and at the same time to derive possible clues as to actual tertiary structure. The model thus serves as a working hypothesis which permits analysis and prediction of results under conditions as close as possible to those occurring in nature.

The progress in structural elucidation of bacteriorhodopsin makes us feel that it will not be very long before we shall be in a position to comprehend within the limits of our present day science the way whereby this magic molecule utilizes light for its functioning. However, perhaps of still more far reaching implications are the prospects of a clearer approach to the other proton translocating systems that are widespread in the various membranes of a cell.

References

[1] Oesterhelt, D. (1976) Angew. Chem. 15, 17–24.
[2] Racker, E. and Stoeckenius, W. (1974) J. Biol. Chem. 249, 662–663.
[3] Ovchinnikov, Yu, A., Abdulaev, N. G., Feigina, M. Yu., Kiselev, A. V. and Lobanov, N. A. (1977) FEBS Lett. 84, 1–4.
[4] Blaurock, A. E. and King, G. I. (1977) Science 196, 1101–1104.
[5] Ovchinnikov, Yu. A., Abdulaev, N. G., Feigina, M. Yu., Kiselev, A. V., Lobanov, N. A. and Nasimov, I. V. (1978) Bioorg. Khim. 4, 979–981.
[6] Bridgen, J. and Walker, I. (1976) Biochemistry 15, 792–796.
[7] Lipkin, V. M., Aldanova, N. A., Feigina, M. Yu., Zhikulina, E. B. and Vinogradova, E. I. (1972) Biokhimiya V.37, N2, 410–413.
[8] Fontana, A. (1972) Methods Enzymol. 25, 419–423.
[9] Henderson, R. and Unwin, P. N. T. (1975) Nature 257, 28–31.
[10] Abdulaev, N. G., Feigina, M. Yu., Kiselev, A. V. and Ovchinnikov, Yu. A. (1977) Bioorg. Khim. 3, 709–711.
[11] Gerber, G., Gray, C., Wildenauer, D. and Khorana, H. G. (1977) Proc. Natl. Acad. Sci. USA 74, 5426–5430.
[12] Abdulaev, N. G., Feigina, M. Yu., Kiselev, A. V., Ovchinnikov, Yu. A., Drachev, L. A., Kaulen, A. D., Khitrina, L. V. and Skulachev, V. P. (1978) FEBS Lett. 90, 190–194.
[13] Blaurock, A., Stoeckenius, W., Oesterhelt, D. and Scherphef, G. (1976) J. Cell. Biol. 71, 1–22.
[14] Long, M., Urry, D. and Stoeckenius, W. (1977) Biochem. Biophys. Res. Commun. 75, 725–731.
[15] Konishi, T. and Packer, L. (1977) FEBS Lett. 79, 369–373.
[16] Nagle, J. and Morowitz, M. (1978) Proc. Natl. Acad. Sci. USA 75, 298–302.
[17] Ovchinnikov, Yu. A., Abdulaev, N. G., Feigina, M. Yu., Kiselev, A. V., Lobanov, N. A. and Nasimov, I. V. (1978) Bioorg. Khim. 4, 1573–1574.