Ersch. in: The EMBO Journal ; 8 (1989), 13. - S. 3963-3971

Primary structure of sensory rhodopsin I, a prokaryotic photoreceptor

A.Blanck, D.Oesterhelt, E.Ferrando, E.S.Schegk and F.Lottspeich

Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG

Communicated by D.Oesterhelt

The gene coding for sensory rhodopsin I (SR-I) has been identified in a restriction fragment of genomic DNA from the *Halobacterium halobium* strain L33. Of the 1014 nucleotides whose sequence was determined, 720 belong to the structural gene of SR-I. In the 5′ non-coding region two putative promoter elements and a ribosomal binding site have been identified. The 3′ flanking region bears a potential terminator structure. The SR-I protein moiety carries no signal peptide and is not processed at its N terminus. The C terminus, however, lacks the last aspartic acid residue encoded by the gene. Analysis of the primary structure of SR-I reveals no consistent homology with the eukaryotic photoreceptor rhodopsin, but 14% homology with the halobacterial ion pumps, bacteriorhodopsin (BR) and halorhodopsin (HR). Residues conserved in all three proteins are discussed with respect to their contribution to secondary structure, retinal binding and ion translocation. The aspartic acid residue which mediates in BR the reprotonation of the Schiff base (D_{96}) is replaced in SR-I by a tyrosine (Y_{87}). This amino acid replacement is proposed to be of crucial importance in the evolution of the slow-cycling photosensing pigment SR-I.

Key words: bacteriorhodopsin/*Halobacterium halobium*/halorhodopsin/sensory rhodopsin I

Introduction

The sequence of molecular events occurring in living cells upon stimulation is referred to as signal transduction. Chemical and electrical stimuli as well as light signals are transduced along a molecular interaction chain to their intracellular target, which might be the gene expression system or an enzyme complex or the flagellar motor of a motile cell. The signals follow linear, branched or network pathways. Very thoroughly studied examples are the systems of hormone action for growth regulation (Beato, 1989), the visual transduction cascade (Chabre and Deterre, 1989) and chemotaxis in bacteria (Steward and Dahlquist, 1987). Amplification and adaptation are essential features of these signalling systems. They allow the swimming cell to optimize its sensitivity and to detect, for example, spatial stimulus gradients. One model system which has been exploited for sensory mechanistic studies of micro-organisms is the chemotaxis of enteric bacteria. Another example comes from halobacterial phototaxis.

Some halobacteria are specialized for photosynthesis mediated by two retinal proteins, bacteriorhodopsin (BR) and halorhodopsin (HR), which act as light-driven ion pumps (Lanyi, 1984, 1986; Oesterhelt and Tittor, 1989). One of the best studied species in this respect is *Halobacterium halobium*. These bacteria also show light-induced motility responses which allow them to migrate toward optimal light conditions for ion pump activity (Hildebrand and Dencher, 1975). Light quality and intensity are sensed by two other retinal proteins, sensory rhodopsin I (SR-I) and sensory rhodopsin II (SR-II, also called P_{480} or phoborhodopsin) (Takahashi *et al.*, 1985; Wolff *et al.*, 1986; Spudich *et al.*, 1986; Marwan and Oesterhelt, 1987; Spudich and Bogomolni, 1988). [We follow the suggestion of Lanyi *et al.* (1989) that as new bacterial rhodopsins are discovered, the original names which refer to their function be retained, i.e. bacteriorhodopsin for proton pumps, halorhodopsin for chloride pumps, and sensory rhodopsin for sensory pigments. These names should be either preceded by the species name when appropriate, or followed by roman numerals when the species are not identified, or more than one pigment with the same function is found in the same species. The genes for these proteins should be called *bop, hop* and *sop* respectively, together with the species designations to indicate their origin.]

The motility of bacterial cells is based on a motor-driven flagellar bundle. Clockwise rotation of the flagellar bundle causes forward swimming. Direction reversals are brought about by spontaneous switches to the counter-clockwise mode (Alam and Oesterhelt, 1984). As there is no bias for clockwise or counter-clockwise rotation, the cells perform a random walk unless a light stimulus hits the photoreceptor molecules. Increasing green light and decreasing blue light, both attractant stimuli, lead to a prolongation of a single run before adaptation takes place. Repellent light (green/blue), on the other hand, shortens the period between two reversals (Hildebrand and Schimz, 1986; Schimz and Hildebrand, 1988). This stimulus-controlled behaviour enables halobacterial cells to approach and to avoid specific light sources. Although the light response in halobacterial phototaxis has been described quantitatively (Marwan and Oesterhelt, 1987), little is known about the molecular components of the signal transduction chain so far. The photosensing system in vertebrates, however, has been elucidated in detail (Applebury *et al.*, 1986). The photoreceptor molecule in the retina, rhodopsin, is an integral membrane protein with an M_r of 40 000. It consists of seven transmembrane helices and carries retinal as a covalently bound chromophoric group. This structural principle is shared by the smaller sized halobacterial ion pumps, whose M_r ranges around 25 000. In contrast to these proteins rhodopsin contains large interhelical loops which carry the sites for interaction with several components of the visual cascade. Most important is the

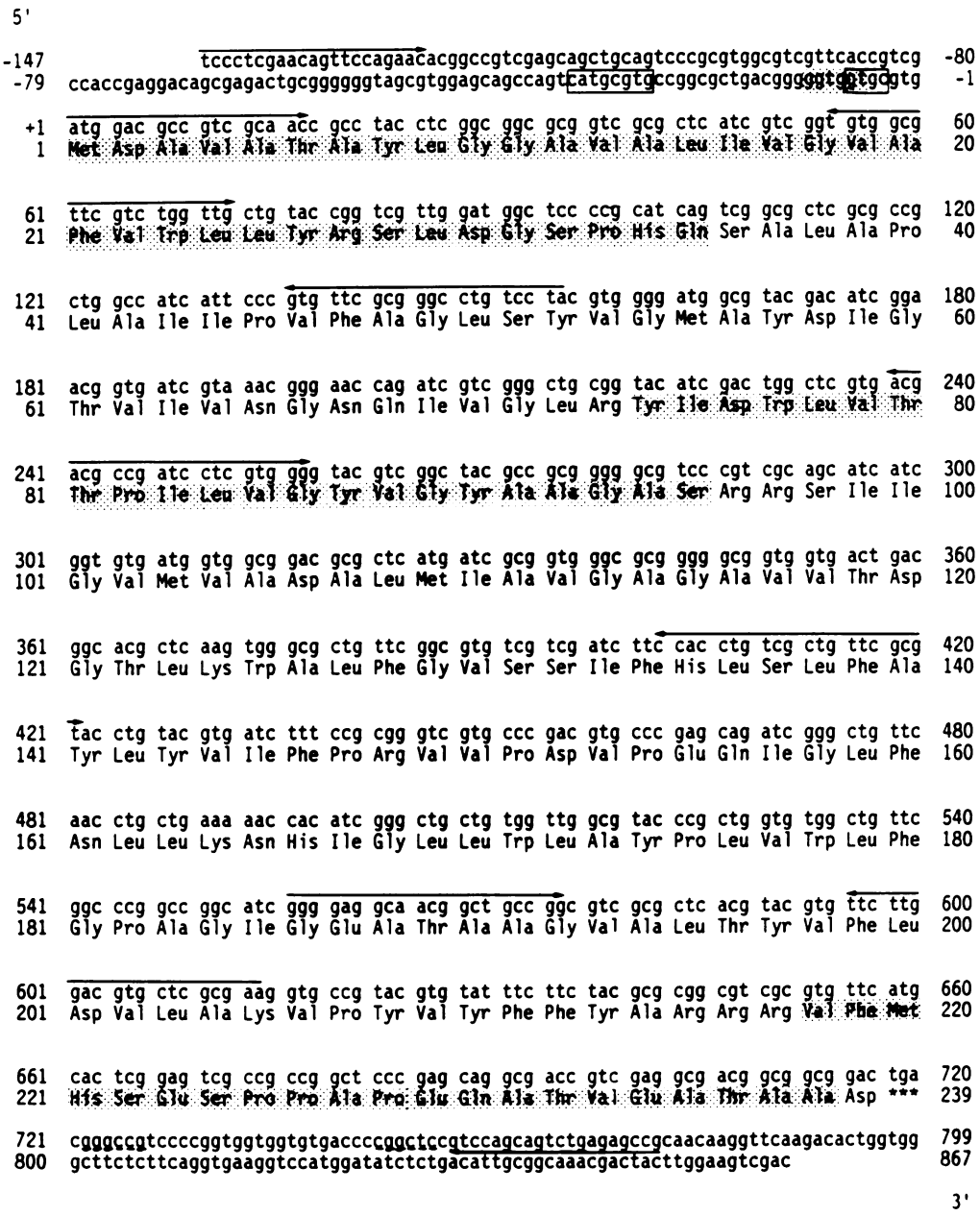

Fig. 1. Nucleotide sequence of the *sop*-I gene with its 5′ and 3′ flanking regions. Arrows indicate the positions of synthetic oligonucleotides used as sequencing primers. An ORF begins at position +1 and ends at position +720. Upstream of the ATG two putative promoter elements, box A and box B, are shown in boxed letters. The ribosomal binding site is located at −11 to −6 and is shown in dotted letters. The dotted regions in the protein sequence have been confirmed by amino acid sequencing.

interaction with a G protein. The activation of this protein is the next step in signal transduction upon light stimulation of rhodopsin.

In a first effort to investigate the signal chain underlying halobacterial phototaxis we have isolated the photoreceptor molecule, sensory rhodopsin I. The photointermediates arising upon light stimulation of the purified protein match the action spectrum of phototaxis. The apparent mol. wt of the protein was determined to be 24 kd (Schegk and Oesterhelt, 1988). Assuming a seven helices transmembrane structure, the size of SR-I does not allow for large cytoplasmic domains, thus assigning SR-I to the family of halobacterial retinal proteins rather than to that of visual pigments. On the other hand, both photoreceptors fulfil similar functions, suggesting that the study of the halobacterial system will contribute to an improved understanding of signal transduction mechanisms. A further step in this direction, the determination and analysis of the primary structure of sensory rhodopsin I, will be presented in this paper.

Results

Protein sequence

The amino acid sequence of SR-I was determined to 33% by automated Edman degradation. These protein data were later confirmed by DNA sequencing (Figure 1). Thirty five amino acids of the N terminus were sequenced. A methionine residue occurs in the first position indicating that SR-I is not N-terminally processed. Further protein sequence data were

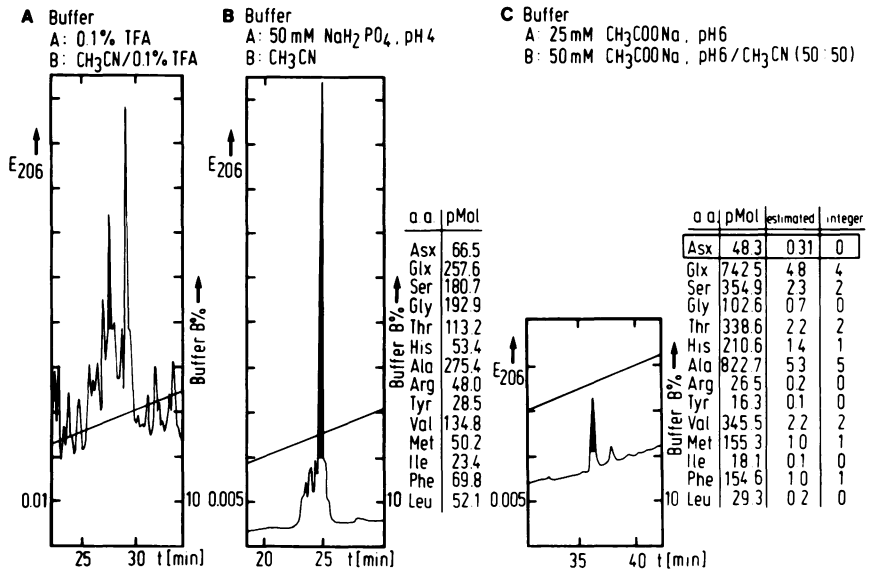

Fig. 2. Purification of the C-terminal peptide of SR-I by reverse-phase HPLC on a LiChrospher 100 RP-18 column. The flow rate was 1 ml/min and the buffers are indicated on top of the chromatograms. (**A**) Separation of a tryptic digest of SR-I. Only the region containing the C-terminal peptide (black coloured peak) is shown. (**B**) First rechromatography of the fraction containing the C-terminal peptide. The amino acid composition of the peak of interest is shown on the right. (**C**) Second rechromatography of the C-terminal peptide. Comparison of the two subsequent amino acid analyses in (B) and in (C) shows that the desired peptide is now obtained in pure form. The experimentally determined amino acid composition of the pure peptide indicates that no aspartic acid is present.

obtained by tryptic digestion. Two tryptic fragments from the SR-I polypeptide were isolated. By comparison with the DNA sequence one of them was identified as the C terminus of the protein.

As is shown in Figure 2 the HPLC fraction containing this peptide had to be further purified in order to determine its exact amino acid composition. This C-terminal peptide lacks the last aspartic acid residue coded by the open reading frame (ORF) (Figure 1), thus indicating C-terminal processing of the gene product. Although this peptide is the only C-terminal one which could be identified in tryptic hydrolysates, we cannot exclude the presence of small quantities of unprocessed or further processed photoreceptor molecules *in vivo*. The kind of C-terminal processing determined here for the SR-I polypeptide is found also in BR and HR and suggests the existence of an aspartate-specific exopeptidase activity in halobacteria.

Gene isolation

Attempts to identify the gene of SR-I (*sop*I gene) in a cosmid library of genomic DNA from *H.halobium* strain L33 (Blanck and Oesterhelt, 1987) with gene fragments coding for bacteriorhodopsin (*bop*) and halorhodopsin (*hop*) as heterologous probes failed. Further trials with oligonucleotides coding for the conserved sequence around the retinal binding lysine residue or for the first six amino acids of SR-I were not successful. In these experiments partially degenerated short oligonucleotides were hybridized against the target DNA under stringent conditions in order to allow base pairing only between fully complementary DNA sequences. Nevertheless cross hybridization always occurred, impairing the identification of a *sop*I-specific signal.

Following a new strategy we synthesized a fully degenerated oligonucleotide mixture containing long oligonucleotides (59mers) which coded for the first 20 amino acid residues of SR-I. This mixture consisted of 10^{11} individual oligonucleotides. Further, we applied hybridization conditions which permitted hybridization between the *sop*I gene and oligonucleotides not fully complementary over the whole sequence. This resulted in an improved signal-to-noise ratio and led to the identification of a specific signal in *Bam*HI digests of genomic DNA. The detected 9.5 kb fragment was subcloned and subjected to double-strand sequencing. Eight different (16-fold) degenerated oligonucleotide mixtures corresponding to the first six amino acid residues of SR-I were first used as sequencing primers. One of them allowed for sequencing the 5′ coding region of the *sop*I gene, whose complete nucleotide sequence is shown in Figure 1.

Comparison of the DNA sequence with protein sequence data confirmed that the 720 bp long ORF encodes SR-I. Furthermore, the M_r of 25 501 calculated for the gene product is in good agreement with the mol. wt value determined by sedimentation analysis (Schegk and Oesterhelt, 1988).

Since the ATG start codon of the ORF coincides with the N terminus of the protein as determined by amino acid sequencing we conclude that SR-I is not processed at its N terminus. Thus, insertion of the SR-I polypeptide is achieved without signal sequence, in contrast to BR. The 5′ non-coding region of the *sop*I gene contains a putative ribosomal binding site at position −11 to −6 complementary to the 3′ end of the 16S rRNA (Kagramanova *et al.*, 1982) and two putative promoter elements. These polymerase binding sites were defined according to the archaebacterial consensus sequence (Reiter *et al.*, 1988; Thomm and Wich, 1988) and are located at the conserved positions −33 to −26 and −7 to −4. Furthermore, the gene bears in its 3′ flanking region a potential terminator structure with a 6 bp G-C stem sequence and a 21 bp loop between the two palindromes. The GC content of the *sop*I gene (67%), as well as the codon usage, are as expected from average values of the halobacterial genome.

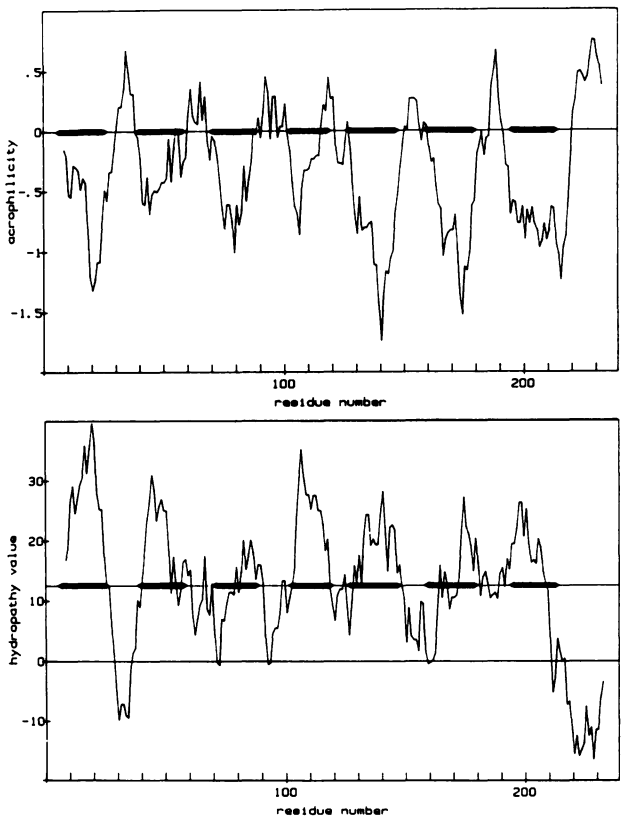

Fig. 3. Acrophilicity and hydropathy plot of SR-I. The acrophilicity plot in the upper part of the figure clearly shows seven transmembrane helices (black bars) which are centred around the minimum values. The hydropathy plot in the lower part separates the hydrophilic (minima correspond to connective loops) and hydrophobic (helices 1–7) protein areas less well, especially around helix 3. The line additional to the zero line represents the average hydrophilic value of all 238 amino acids which gives the best separation between the hydrophilic and hydrophobic areas. In both plots a window size of 14 was used.

Secondary structure

For a secondary structure prediction the amino acid sequence of SR-I has been scanned using the algorithms for hydrophobicity and acrophilicity calculations (Kyte and Doolittle, 1982; Hopp, 1985). The results are shown in Figure 3. Seven hydrophobic domains were determined by this analysis and can be very well arranged in the secondary structure model previously developed for the halobacterial ion pumps, BR and HR (Engelman *et al.*, 1986; Blanck and Oesterhelt, 1987).

In the Kyte–Doolittle plot in Figure 3 the hydrophobicity of the third domain is not very evident. In the acrophilicity plot, however, the hydrophobic domains are defined more clearly. Figure 3 shows seven transmembrane helices together with six interconnecting hydrophilic loops. The discrepancy between the two plots in the third domain may arise from the negatively charged residues which seem to have a stronger effect on the hydropathy rather than on the acrophilicity calculation. Generally, in these structure predictions basic and acidic amino acid residues are assumed to occur in their ionic forms, a fact which is not always granted. For example, it was demonstrated that two out of four intramembrane aspartates in BR exist in their protonated form (Engelhard *et al.*, 1984; Eisenstein *et al.*, 1987; Gerwert *et al.*, 1989).

Figure 4 shows an alignment of the primary structures of SR-I, BR and HR. Gaps were introduced in the individual sequences only if this permitted a significant increase in the number of sequence identities. The three sequences were arranged in seven blocks representing the seven transmembrane helices. Sequence identities between all three proteins are indicated by dotted areas. The model is identical to those already published for BR and HR, except for the SR-I specific helix borders which were drawn according to predictions based on Figure 3. The few slight differences to the model of BR and HR are listed in the legend to Figure 4.

Comparison of the primary structures shows that all three proteins share significant homology. The percentages of identical amino acid positions in the processed gene products were determined by a computer calculation (Needleman and Wunsch, 1979; Devereux *et al.*, 1984) and are as follows: 26% (four gaps) between SR-I and BR, 23% (five gaps) between SR-I and HR and 32% (five gaps) between HR and BR. Finally, 14% of all residues are at identical positions in all three proteins.

Discussion

We have isolated and sequenced the gene corresponding to a third halobacterial retinal protein, the photoreceptor molecule SR-I. The sequence bears stop codons 5′ flanking of the ATG at positions −2, −18 and −51. Therefore the question arose as to whether this gene is transcribed as a polycistronic mRNA. Size determination of the *sop*I mRNA by Northern blot analysis, however, showed that the *sop*I gene occurs as a single transcriptional unit (our unpublished results). Later, we compared the 5′ flanking region of the *sop*I gene with the published consensus sequence for halobacterial promoters and identified two putative promoter elements, box A and box B. Box B is very close to the start codon (−7 to −4) and consists of the sequence GTGC; compared with the consensus sequence there is a purine exchange in the first position from A to G. Box A, as it was determined for several archaebacterial genes is shown in Figure 5. The position of this box relative to that of the *sop*I gene is as expected and its sequence agrees with the consensus except for the fourth base, which is a G instead of A or T. These two discrepancies with the archaebacterial consensus sequence might have an influence on the affinity of the transcription machinery to the *sop*I promoter.

As already mentioned (see 'Results') the SR-I polypeptide lacks the aspartic acid residue encoded by the gene, as is also the case for the two other sequenced retinal proteins from *H.halobium*, BR and HR. With respect to N-terminal processing, however, the proteins show remarkable differences. While HR exists in several different N-terminal heterogeneous species (Kehl *et al.*, 1989) and BR carries a precursor sequence (Dunn *et al.*, 1981), SR-I is not processed at all at its N terminus. Since all three proteins are membrane located there seems to be no need for a leader peptide for membrane insertion, as has been assumed up to now on the basis of BR data. Futhermore, the methionine residue at position 1 in the SR-I sequence is not formylated, which is a common feature between archaebacterial and eukaryotic proteins (König, 1986).

Processing of the N terminus in prokaryotes may also depend on the location of the protein in the membrane, as

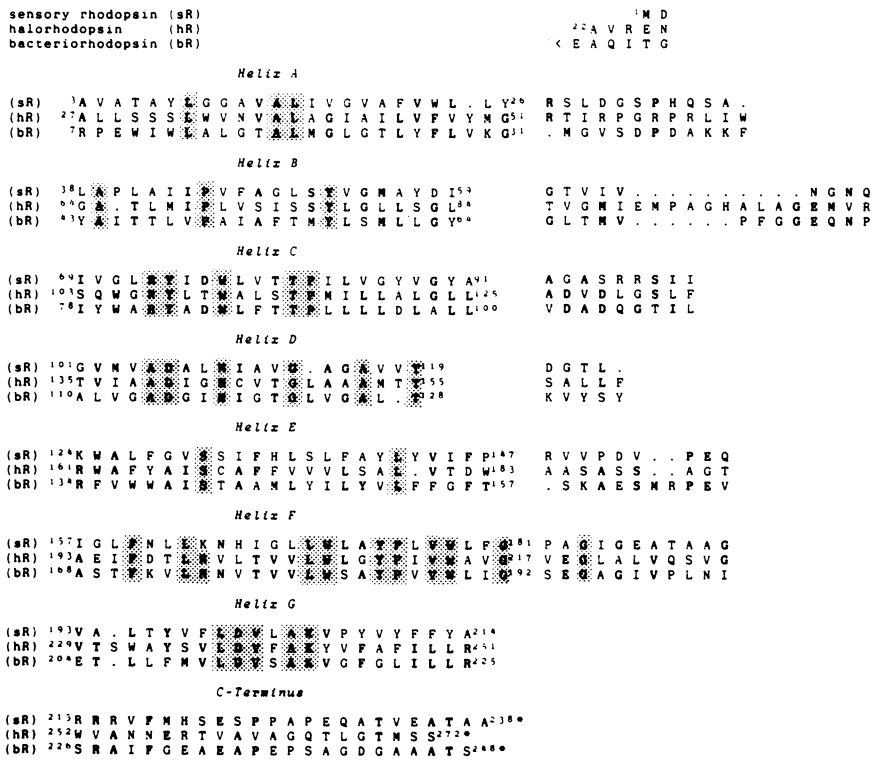

Fig. 4. Linear alignment of the protein sequences of SR-I, BR and HR. Residues which occur at identical positions are shown in bold letters, those at identical positions in all three proteins are dotted (with the exception of K_{164}). For SR-I and HR the numbering starts with 1 at the ATG codon of the ORF. The first amino acid in the BR sequence is the pyroglutamic acid encoded at position 14 of the gene for easier evaluation of literature data. The transmembrane helical regions and the C terminus are listed one below the other. The breaks between the helices and the interconnecting loops were drawn according to the SR-I sequence. In helices D, F and G the breaks occur at a different position from those defined in the secondary structure model of BR and HR. Helical borders differing from those in SR-I are given below.
helix D: BR: D_{102}, V_{130}
HR: L_{130}, S_{156}
helix F: BR: R_{164}, G_{192}
helix G: BR: L_{206}, F_{230}

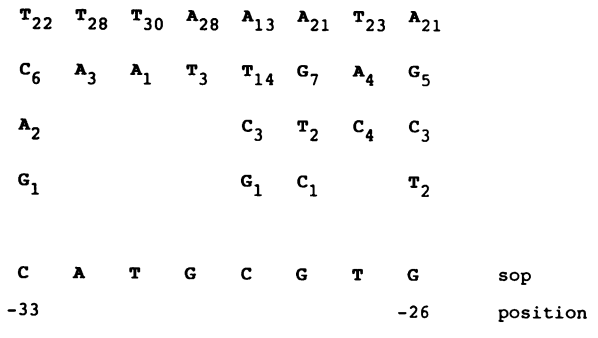

Fig. 5. Comparison of a putative SR-I gene promoter element with the consensus sequence for archaebacterial promoters. The 5′ flanking regions of several archaebacterial genes were aligned with respect to the conserved sequence TTTAT/AATA. Subscripts indicate the base frequency at each position (from Thomm and Wich, 1988).

it has been proposed for the subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis*. Only the H-subunit, whose N terminus faces the outside of the cell and thus escapes the action of cytoplasmic peptidases, retains its N-terminal fMet residue (Michel *et al.*,1986).

The coding region of the *sop*I gene was also compared with the *bop* and *hop* sequences with respect to silent mutations in the conserved amino acids. Silent base replacements amount to 13.4% between *sop*I and *bop*, 16.0% between *bop* and *hop* and 16.6% between *sop*I and *hop*. Therefore we presume that all three genes evolved from a common ancestor of retinal proteins in a relatively short period of evolution.

A further comparison concerns the primary structure of the prokaryotic photoreceptor SR-I with that of its functional counterpart in vertebrates, eye rhodopsin. As already mentioned, these two proteins have different molecular sizes (SR-I, 25 kd; rhodopsin, 40 kd) and complex two different isomers of retinal (SR-I, all-*trans*; rhodopsin, 11-*cis*). Further differences lie in their biochemical environment with respect to lipids and ionic strength. For these reasons it is not amazing that the primary stuctures of the two light sensors do not share any consistent homology. Besides the AK motif of the retinal binding site only one further conserved sequence motif could be identified, which is W_{171}xAYP$_{175}$ in helix F of SR-I.; The corresponding motif in eye rhodopsin is WxPYA, where the A and the P residue are exchanged against each other but the distance between the tryptophan and the tyrosine residues is kept constant. One more invariant tryptophan residue occurs in helix C at position 77 of SR-I. On the other hand, the eye rhodopsin polypeptide chain is folded in seven amphipathic transmembrane helices thus showing the same secondary structure as SR-I and as the halobacterial ion pumps, BR and HR.

Figure 6 shows a helical wheel model (Schiffer and Edmundson, 1967) of SR-I with seven transmembrane helices arranged around the retinal binding pocket. The bold

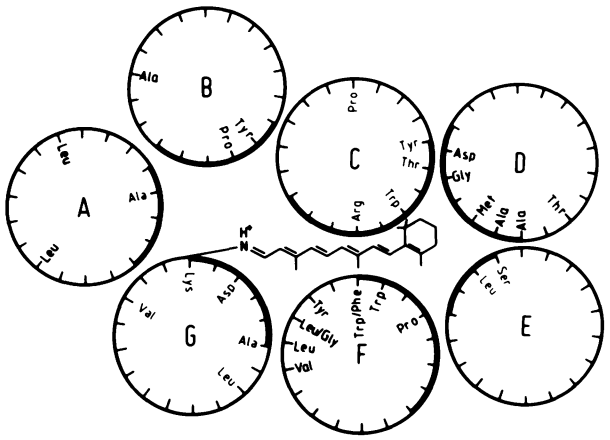

Fig. 6. A helical wheel model of SR-I. The helices were positioned in the contours of the electron diffraction map of BR. The hydrophilic side of each helix is marked by a bold segment and was calculated using the values of Kyte and Doolittle (1982). Conserved residues are indicated (see text). The rotational orientation of the helices was chosen in order to position most of the putative functional residues towards the retinal binding cleft. This orientation is also in agreement with the hydrophobic moment of the helices (Eisenberg *et al.*, 1982).

segment in each circle represents the hydrophilic side of the amphipathic helix which faces the inner part of the molecule. The hydrophobicity of the transmembrane regions of SR-I is more pronounced compared with BR and HR and the amphipathic character of its helices is weaker than in the ion pumps. These findings are consistent with the function of SR-I as a photoreceptor, since no polar intramembrane regions are required in SR-I for ion translocation.

In spite of the different functions performed by the halobacterial retinal proteins, there are common features in their mechanism of action:

(i) all three proteins contain a covalently bound all-*trans* retinal molecule, which shows the characteristic colour shift upon binding to the protein;
(ii) in the dark, the all-*trans* form of the chromophore equilibrates specifically with the 13-*cis* isomer;
(iii) all three proteins share the fundamental process of thermoreversible isomerization of the retinal molecule upon light absorption, called the photocycle; and
(iv) the stereochemistry of this photoreaction and the kinetics of its reversion are determined by the protein moiety.

The crucial difference between SR-I and the ion pumps consists, however, in the rates of photocycling. The cycling time of the ion pumps' photocycle amounts to 10 ms while SR-I needs ~1 s to return to the initial state. The photocycles of BR and HR are always strictly coupled to ion translocation even in mutants which have a photocycle speed comparable with SR-I (Butt *et al.*, 1989). In contrast, no charge displacement or other electrical effect could be associated with SR-I function up to now (Ehrlich *et al.*, 1984; Oesterhelt and Marwan, 1987). Therefore it is assumed that the phototactical response requires direct molecular interactions between photoproducts of SR-I and the components of the signal transduction chain. The slow cycling rate of SR-I meets exactly this need for longer-lived photointermediates.

The mutation of a fast-cycling, ion pumping transmembrane retinal protein to a homologous, but slow-cycling light sensor molecule is achieved by changing the protein environment of the retinal chromophore. Comparison of the primary structures of BR and HR, on the one hand, and SR-I on the other hand should therefore allow us to distinguish between residues responsible for common structural elements, especially for the binding of retinal and those chiefly involved in ion translocation which do not necessarily have to be conserved in SR-I.

The sequence alignment in Figure 4 demonstrates that most of the residues conserved between SR-I, BR and HR are located in the membrane-spanning segments of the proteins. A, V and L residues are often retained because they are energetically favoured in hydrophobic helices. Glycine residues which induce helix breaks are concentrated at the membrane borders. All three intermembrane P residues in helices B, C and F of BR and HR are conserved in SR-I and three more occur at positions 40, 147 and 207. *Cis/trans* isomerization of these conserved prolines might play a role in conformational changes occurring during the catalytic cycle (Dunker, 1982). Further conserved amino acids are one F, several S and T residues and one G residue in the middle of helix D. Lacking atomic stuctural resolution of the retinal proteins no functional interpretation can be given for these residues at present.

The helices C, D, F and G which are in the closest proximity of the chromophore as deduced for BR from electron density maps and neutron diffraction studies (Baldwin *et al.*, 1988; Heyn *et al.*, 1988) share the highest degree of homology. The SR-I sequence, in concert with the BR and HR data, has confirmed the assumption that the following conserved residues (numbered according to the SR-I sequence) may directly interact with the retinal molecule.

(i) The retinal binding lysine residue (K_{205}) in helix G.
(ii) The motif L_{200} DVxA_{204} is assumed to contribute to the negative charge balancing the protonated Schiff base. This residue might also be involved in the wavelength regulation of the chromophore's absorption maximum.
(iii) Three tryptophan residues (W_{77}, W_{171}, W_{178}) in helix C and F; they are good candidates to interact closely with the retinal molecule and could be necessary to its correct positioning in the binding cleft. Evidence for a close spatial relationship between tryptophan side chains and retinal in BR and HR was obtained by demonstration of exciton coupling between these moieties (Polland *et al.*, 1986). The contribution of a further tryptophan residue in helix C (W_{80} and W_{105} in BR and HR) to this interaction seems neglectable since this residue is mutated to a glycine residue in SR-I (position 71).
(iv) P_{175} in helix F. This residue is embedded in the conserved sequence motif L_{170} W x x Y P x V W_{178}. In BR mutation of the homologous residue P_{186} → Leu causes a blue-shift of the absorption maximum of 70 nm and destroys proton pump activity (Hackett *et al.*, 1987). This proline residue probably determines the helical conformation needed for best interaction between retinal and the neighbouring tryptophan residues.
(v) Two tyrosine residues, Y_{52} in helix B and Y_{174} in helix F are conserved between all three proteins. In BR there is evidence that the tyrosine residue at position 57 corresponding to Y_{52} influences the photocycle and light–dark adaptation (Soppa *et al.*, 1989). The residue homologous to Y_{174} occurring in BR at position 185 is

expected to contribute negative charge density as a tyrosinate to balance the protonated Schiff base (de Groot *et al.*, 1989) and is reversibly protonated during the BR photocycle (Braiman *et al.*, 1988).

(vi) The Met_{109} residue in helix D. This amino acid faces the β-ionone ring in the binding cavity (see also helical wheel model). Mild oxidation of methionine residues by H_2O_2 to methionine sulphoxide (our unpublished results) has been shown to influence planarization of the β-ionone ring using retinol as a probe (Schreckenbach *et al.*, 1978).

In addition to these conserved amino acids in the retinal binding pocket we examined the distribution of basic and acidic residues. Two conserved motifs involving basic amino acid residues are $R_{73}Y_{74}$ in helix C and $L_{163}K_{164}$ in helix F. In BR and HR K_{164} is replaced by an R while it is conserved in the chloride pump from the alkalophlic archaebacterium *Natronobacterium pharaonis* (Lanyi *et al.*, 1989). These two mentioned basic sequence motifs, which are the only ones deeply buried in the membrane, are good candidates for binding of chloride ions in halorhodopsin. Their conservation in BR and SR-I, on the other hand, might indicate contribution to the overall structure.

Five positively charged amino acids are found in SR-I in the cytoplasmic interhelical segments, while the extracellular loops bear three negative charges. This charge distribution corresponds to the pattern expected for prokaryotic membrane proteins (Michel *et al.*, 1986; von Heijne, 1986) and supports the hypothesis that membrane insertion of these proteins could be sustained by the membrane potential.

No unpaired charges occur in the membrane-spanning regions of SR-I under the premise that D_{106} exists in its protonated form. However this has only been shown for BR (Renthal *et al.*, 1988). Otherwise D_{106} in SR-I would provide a negative point charge of unknown function.

Some more charged residues are found in BR (seven) and HR (four) when compared with SR-I. They are all arranged toward the membrane surface. This increased surface charge density might be required for ion pump activity.

Recently, the role of two particular acidic residues in the mechanism of BR function could be determined. Experiments were carried out with BR mutants gained by random and by site-specific mutagenesis. Both the random and the specific mutated proteins have identical wild-type sequences (Marinetti *et al.*, 1989; Soppa and Oesterhelt, 1989). Strains which carry a mutation in either D_{85} or D_{96} are defective in phototrophic growth. Time-resolved studies on the proton pump activity of these BR mutants showed that these two aspartic acid residues are involved in the proton displacements occurring at the Schiff base during the photocycle (Butt *et al.*, 1989). After light stimulation protons are released from the positively charged Schiff base possibly via D_{85} with subsequent formation of the M-intermediate (50 μs). This species returns to the initial state in 5–10 ms after reprotonation of the Schiff base via D_{96}. These two residues are therefore the structural prerequisites for both, the fast-cycling rate and the proton translocation activity of BR.

In the BR mutant $Asp_{96} \rightarrow Asn$ (Soppa *et al.*, 1989), for example, the decay of the M-intermediate is slowed down by a factor of ~50 at physiological pH. The Schiff base is reprotonated so slowly (in 500 ms at pH 7) that no or little proton pumping activity may be detected. The wild-type phenotype, however, could be restored by addition of azide. It was suggested that hydrogen azide acts as an internal proton donor instead of Asp_{96} (Tittor *et al.*, 1989).

In HR both D_{85} and D_{96} from BR, are exchanged for neutral amino acid residues and no changes in the protonation state of the Schiff base have been detected for the active ion translocation cycle (Oesterhelt *et al.*, 1985). Indeed, binding and transport of chloride ions should follow a different mechanism than for protons in BR.

In SR-I only one of the two above mentioned aspartate residues is conserved, namely D_{85}, while D_{96} is substituted by a tyrosine residue. This should, in principle, allow fast release of protons to form a long-lived intermediate. In fact, the photointermediate arising upon light stimulation within 270 μs has its absorption maximum at 380 nm and decays in 750 ms (Spudich and Bogomolni, 1988). The SR-I photocycle resembles therefore the one of the BR mutants $Asp_{96} \rightarrow Asn$, which also lacks the aspartate residue responsible for reprotonating the Schiff base. We assume that during formation of SR-I the Schiff base is either deprotonated or hydrolysed. Since this process does not elicit any potential difference there has to be an internal proton acceptor for the case that the Schiff base is deprotonated. Hydrolysis of the Schiff base, on the other hand, would end up in the protonated free ϵ-amino group of K_{205}. In fact, D_{76} may accept a proton from the Schiff base as it might occur in Br (D_{85}), or it may act as a catalyst of the hydrolysis of the Schiff base. Preliminary results based on the pH-dependence of the decay of SR-I indicate that in this species the retinal moiety occurs as a free aldehyde.

Since there is no internal proton donor comparable to D_{96} the charge displacement at the Schiff base in SR-I cannot be quickly reverted; this leads to the slow-cycling rate typical of SR-I. Azide shows no significant effects on the time constants of the SR-I photocycle. A possible reason could lie in a different arrangement of amino acid residues between the chromophore and the cytoplasmic surface of the molecule. This might impair the diffusion of the azide ion and its interaction with the Schiff base. Nevertheless, the comparison of SR-I with BR and its mutant $Asp_{96} \rightarrow Asn$ allows for a very attractive hypothesis about the evolution of this sensory protein. The crucial step has probably been the mutation of the amino acid residue responsible for the fast decay of one intermediate in the photocycle of the cell's light energy transducing molecule. This resulted in a new protein which presumably gained the ability to interact with other molecules in the cell in a light-dependent manner, i.e. to act as a photoreceptor.

The translation of the light signal into the physiological answer of phototaxis is the subject of current research. Aggregation of SR-I molecules might play a role since oligomerization seems to influence the stability of the solubilized protein (our unpublished results). Although no SR-I oligomers have been detected in membranes by spectroscopic analysis (Hasselbacher *et al.*, 1988) transient aggregation may occur and trigger signal transduction. Interactions, both with cytoplasmic or with other membrane proteins, are possible. Recently, the involvement of cyclic nucleotides in halobacterial phototaxis has been demonstrated and evidence for the occurrence of a G protein-like compound presented (Schimz and Hildebrand, 1987; Schimz *et al.*, 1989). Furthermore, a 94 kd methyl-accepting membrane protein has been identified which eventually could

be implicated in the adaptation to phototactical stimuli (Spudich *et al.*, 1988). The connections between these compounds of the signal transduction chain and SR-I remain to be elucidated.

Materials and methods

Materials

[γ-^{32}P]dATP (5000 Ci/mmol) was obtained from Amersham-Buchler (Braunschweig, FRG); caesium chloride was from J.T.Baker Chemical Co. (Grossgerau, FRG), molecular biological enzymes, chloramphenicol, DNA-marker and ethidium bromide were obtained from Boehringer Mannheim (Mannheim, FRG); Bacto-Agar, yeast extract and tryptone were from Difco Laboratories (Detroit, MI), nitrocellulose (0.45 μm) and Elutip-D-columns were purchased from Schleicher & Schüll (Dassel, FRG), agarose ME was obtained from Seakem (Rockland, ME) and sodium dodecyl sulphate (SDS) from Serva (Heidelberg, FRG).Ampicillin, bromophenol blue, bovine serum albumin, Ficoll, calf thymus DNA, polyvinyl pyrrolidon and Triton X-100 were obtained from Sigma (München, FRG). The pGem-vector system was from Promega Biotec/Atlanta (Heidelberg, FRG). All chemicals needed for SR-I isolation are listed in Schegk and Oesterhelt (1988). Trifluoroacetic acid (TFA), sequencing grade, was from Pierce (Rockford, IL). Trypsin (affinity purified) and chemicals were purchased from Merck (Darmstadt, FRG).

Oligonucleotides. The oligonucleotides which were used for hybridization experiments and for sequencing were synthesized on an automatic DNA synthesizer (Applied Biosystems, 381A).

Strains. Halobacterium halobium strain L33 (Bop$^-$, Hop$^+$, Rub$^-$, Ret$^+$) is a bacterioopsin-deficient mutant of S9 (Wagner *et al.*, 1981). *Escherichia coli* DH1 F$^-$ gyrA96 recA1 relA1 endA1 thi-1 hsdR17 (r_k^-,m_k^+) *supE44* λ^- (Hanahan, 1983).

Methods

Purification of sensory rhodopsin I. Sensory rhodopsin I was isolated from *H.halobium* strain flx3 as desribed by Schegk and Oesterhelt (1988). Some printing errors occurred in 'Methods' in this publication and are corrected here: (i) buffer D is prepared by adding 8 ml of the detergent–retinal solution to buffer C; (ii) buffers E and F contained the same amount of the detergent–retinal solution; (iii) the tween-washed membranes were washed once in buffer B and not in buffer D.

After native gel filtration, SR-I was subjected to chromatography in organic solvent also as described in order to obtain very pure samples for N-terminal sequencing.

Enzymatic digestion of sensory rhodopsin I. 3–6 nmol of SR-I purified by hydroxylapatite chromatography or by native gel filtration were dialysed against 100 mM $(NH_4)_2CO_3$ and digested with 2% (w/w) trypsin at 37°C for 6–8 h.

Separation of tryptic fragments by reverse-phase HPLC. SR-I samples digested with trypsin were lyophilized, dissolved in 10% formic acid and subjected to reverse-phase HPLC on a LiChrospher 100 RP-18 column. The column was equilibrated with 0.1% TFA and the material eluted with a linear gradient of acetonitril in 0.1% TFA. Effluent was monitored by absorbance at 206 nm (for details see legend to Figure 2).

Amino acid analysis. For amino acid analysis 0.5–2 μg peptide material was hydrolysed by a gas phase hydrolysis method using a TFA:6 M hydrochloric acid mixture (1:10, v/v) at 159°C for 20 min (Tsugita *et al.*, 1987). The hydrolysate was derivatized by *O*-phthalaldehyde/mercapto-propionic acid and subjected to reverse-phase HPLC as described (Godel *et al.*, 1984).

Amino acid sequencing. The amino acid sequence of the N terminus of SR-I and of its tryptic fragments were determined by automated Edman degradation. The samples were dissolved in 70% formic acid and applied to a gas phase sequencer (Applied Biosystems, 470A). The phenyl-thiohydantoin derivatives were analysed by an on-line isocratic HPLC system (Lottspeich, 1985).

Isolation of the sopI gene. Standard molecular biological methods were performed, if not otherwise indicated, according to Maniatis *et al.*(1982). Chromosomal DNA from *H.halobium* strain L33 was isolated and purified according to Vogelsang *et al.* (1983). As a *sopI* gene specific probe, a fully degenerated 59meric oligonucleotide mixture corresponding to the N terminus of sensory rhodopsin I was synthesized. Southern blot analysis of genomic L33 DNA digested with *Bam*HI was performed with the radioactively labelled oligonucleotide mixture in 50% formamide at 35°C. This resulted in a single band at ~9.5 kb.

For cloning, a *Bam*HI digest was separated on an agarose gel and the DNA from the region around 9.5 kb was recovered by electroelution. Hybridization to slot blots of this material revealed a fraction that was ligated into a *Bam*HI cut pGem3 vector and transformed into the *E.coli* strain DH1. Around 1000 clones containing recombinant DNA were screened with the above mentioned oligonucleotide mixture and three positive clones AB2H11, AB3B5 and AB8D3 were used for further analysis. The plasmids were amplified overnight (Clewell, 1972), isolated from the cells (Katz *et al.*, 1973) and purified on caesium chloride gradients. As all three plasmid DNAs showed the same restriction pattern, pAB2H11 was used for sequencing analysis according to the USB sequenase protocol.

Eight different 16-fold degenerated 17meric oligonucleotide mixtures corresponding to the first N-terminal amino acids of SR-I were synthesized as sequencing primers. One oligonucleotide mixture enabled us to sequence the 5′ end of the *sop*I gene. Based on the obtained sequence data, further primers were synthesized and the sequence of 1014 nucleotides was determined for both strands.

Acknowledgements

We wish to thank Dr J.Tittor for his help in drawing Figure 4 and the 'Tabinger' students, S.Klostermann, V.Döring, B.Sam and M.Krah for their support in the isolation of the gene.

References

Alam,M. and Oesterhelt,D. (1984) *J. Mol. Biol.*, **176**, 459–475.
Applebury,M.L., Cavaggione,A., Chabre,M., Findlay,J.B.C., Goldberg, N.D., Kaupp,U.B., Kühn,H., Liebmann,P.A., Prinz,H., Saibil,H.R., Schleicher,A. and Stryer,L. (1986) In Stieve,H. (ed.), *The Molecular Mechanism of Photoreception*. Dahlem Konferenzen. Springer-Verlag, Berlin, pp. 401–429.
Baldwin,J.M., Henderson,R., Beckman,E. and Zemlin.F. (1988) *J. Mol. Biol.*, **202**, 585–591.
Beato,M. (1989) *Cell*, **56**, 335–344.
Blanck,A. and Oesterhelt,D. (1987) *EMBO J.*, **6**, 265–273.
Braiman,M.S., Mogi,T., Stern,L.J., Hackett,N.R., Chao,B.H., Khorana, G.B. and Rothshild,K.J. (1988) *Proteins: Structure, Function and Genetics*, **3**, 219–229.
Butt,H.-J., Fendler,K., Bamberg,E., Tittor,J. and Oesterhelt,D. (1989) *EMBO J.*, **8**, 1657–1663.
Chabre,M. and Deterre,P. (1989) *Eur. J. Biochem.*, **179**, 255–266.
Clewell,D.B. (1972) *J. Bacteriol.*, **110**, 667–676.
de Groot,H.J.M., Harbison,G.S., Herzfeld,J. and Griffin,R.G. (1989) *Biochemistry*, **28**, 3346–3353.
Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
Dunker,A.K. (1982) *J. Theor. Biol.*, **97**, 95–127.
Dunn,R., McCoy,J., Simsek,M., Majumdar,A., Chang,S.H., RajBhandary,U.L. and Khorana,H.G. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6744–6748.
Ehrlich,B.E., Schen,C.R. and Spudich,J.L. (1984) *J. Membr. Biol.*, **82**, 89–94.
Eisenberg,D., Weiss,R.M. and Terwilliger,T.C. (1982) *Nature*, **299**, 371–374.
Eisenstein,L., Lin,S.-L., Dollinger,G., Odashima,K., Termini,J., Kuuno,K., Ding,W.-D. and Nakanishi,K. (1987) *J. Am. Chem. Soc.*,**109**, 6860–6862.
Engelhard,M., Gerwert,K., Hess,B., Kreutz,W. and Siebert,F. (1984) *Biochemistry*, **24**, 400–407.
Engelman,D.M., Steitz,T.A. and Goldman,A. (1986) *Annu. Rev. Biophys. Chem.*, **15**, 321–353.
Gerwert,K., Hess,B., Soppa,J. and Oesterhelt,D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4843–4947.
Godel,H., Graser,Th., Földi,P., Pfaender,P. and Fürst,P. (1984) *J. Chromatogr.*, **297**, 49–61.
Hackett,N.R., Stern,L.J., Chao,B.H., Kronis,K.A. and Khorana,H.G. (1987) *J. Biol. Chem.*, **262**, 9277–9274.
Hanahan,D. (1983) *J. Mol. Biol.*, **166**, 557–580.
Hasselbacher,C.A., Spudich,J.L. and Dewey,T.G. (1988) *Biochemistry*, **27**, 2540–2546.
Heyn,M.P., Westerhausen,J., Wallat,I. and Seiff,F. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2146–2150.

Hildebrand,E. and Dencher,N. (1975) *Nature*, **257**, 46–48.
Hildebrand,E. and Schmiz,A. (1986) *J. Bacteriol.*, **166**, 689–692.
Hopp,P.T. (1985) In Alitalo,K., Partanen,P. and Vahen,A. (eds), *Synthetic Peptides in Biology and Medicine*. Elsevier Science Publishers B.V., Amsterdam, pp. 3–12.
Kagramanova,V.K., Mankin,A.S., Baratova,L.A. and Bogdanov,A.A. (1982) *FEBS Lett.*, **144**, 177–180.
Katz,L., Kingsbury,D.T. and Helsinki,D.R. (1973) *J. Bacteriol.*, **114**, 577–591.
Kehl,M., Lottspeich,F. and Oesterhelt,D. (1989) In Voelter,W. and König,W. (eds), *Chemistry of Peptides and Proteins*, Vol. **4**. Attempto Verlag, Tübingen.
König,H. (1986) *BIUZ*, **16**, 71–82.
Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.
Lanyi,J.K. (1984) In Ernster,L. (ed.), *Bioenergetics*. Elsevier Science Publishers B.V., pp. 315–350.
Lanyi,J.K. (1986) *Annu. Rev. Biophys. Biophys. Chem.*, **15**, 11–28.
Lanyi,J.K., Duschl,A., Hatfield,G.W., May,K. and Oesterhelt,D. (1989) *J. Biol. Chem.*, in press.
Lottspeich,F. (1985) *J. Chromatogr.*, **326**, 321–327.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Marinetti,T., Subramaniam,S., Mogi,T., Marti,Th. and Khorana,H.G. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 529–533.
Marwan,W. and Oesterhelt,D. (1987) *J. Mol. Biol.*, **195**, 333–342.
Michel,H., Weyer,K.A., Gruenberg,H., Dunger,I., Oesterhelt,D. and Lottspeich,F. (1986) *EMBO J.*, **5**, 1149–1158.
Needleman,S.B. and Wunsch,C.D. (1970) *J. Mol. Biol.*, **48**, 443–453.
Oesterhelt,D. and Marwan,W. (1987) *J. Bacteriol.*, **169**, 3515–3520.
Oesterhelt,D. and Tittor,J. (1989) *Trends Biochem. Sci.*, **14**, 57–61.
Oesterhelt,D., Hegemann,P. and Tittor,J. (1985) *EMBO J.*, **4**, 2351–2356.
Polland,H.J., Franz,M.A., Zinth,W., Kaiser,W. and Oesterhelt,D. (1986) *Biochim. Biophys. Acta*, **851**, 407–415.
Reiter,W.-D., Palm,P. and Zillig,W. (1988) *Nucleic Acids Res.*, **16**, 1–19.
Renthal,R., Brogley,L. and Vila,J. (1988) *Biochim. Biophys. Acta*, **935**, 109–114.
Schegk,E.S. and Oesterhelt,D. (1988) *EMBO J.*, **7**, 2925–2933.
Schiffer,M. and Edmundson,A. (1967) *Biophys. J.*, **7**, 121–135.
Schimz,A. and Hildebrand,E. (1987) *Biochim. Biophys. Acta*, **923**, 222–232.
Schimz,A. and Hildebrand,E. (1988) *Bot. Acta*, **101**, 111–117.
Schimz,A., Hinsch,K.-D. and Hildebrand,E. (1989) *FEBS Lett.*, **249**, 59–61.
Schreckenbach,T., Walckhoff,B. and Oesterhelt,D. (1978) *Biochemistry*, **17**, 5353–5359.
Soppa,J. and Oesterhelt,D. (1989) *J. Biol. Chem.*, **264**, 13043–13044.
Soppa,J., Otomo,J., Straub,J., Tittor,J., Meeßen,S. and Oesterhelt,D. (1989) *J. Biol. Chem.*, **264**, 13049–13056.
Spudich,J.L. and Bogomolni,R.A. (1988) *Annu. Rev. Biophys. Chem.*, **17**, 193–215.
Spudich,E.N., Sundberg,S.A., Manor,D. and Spudich,J.L. (1986) *Proteins: Structure, Function and Genetics*, **1**, 239–246.
Spudich,E.N., Hasselbacher,C.A. and Spudich,J.L. (1988) *J. Bacteriol.*, **170**, 4280–4285.
Steward,R.C. and Dahlquist,F.W. (1987) *Chem. Rev.*, **87**, 997–1025.
Takahashi,T., Tamioka,H., Kamo,N. and Kobatake,Y. (1985) *FEMS Microbiol. Lett.*, **28**, 161–164.
Thomm,M. and Wich,G. (1988) *Nucleic Acids Res.*, **16**, 151–163.
Tittor,J., Soell,C., Oesterhelt,D., Butt,H.-J. and Bamberg,E. (1989) *EMBO J.*, **8**, 3477–3482.
Tsugita,A., Uchida,T., Mewes,H.W. and Ataka,T. (1987) *J. Biochem.*, **102**, 1593–1597.
Vogelsang,H., Oertel,W. and Oesterhelt,D. (1983) *Methods Enzymol.*, **97**, 226–241.
von Heijne,G. (1986) *EMBO J.*, **5**, 3021–3027.
Wagner,G., Oesterhelt,D., Krippahl,G. and Lanyi,J.K. (1981) *FEBS Lett.*, **131**, 341–345.
Wolff,E.K., Bogomolni,R.A., Scherrer,P., Hess,B. and Stockenius,W. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7272–7276.