The EMBO Journal vol.5 no.6 pp.1149–1158, 1986

The 'light' and 'medium' subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: isolation of the genes, nucleotide and amino acid sequence

H.Michel, K.A.Weyer, H.Gruenberg, I.Dunger, D.Oesterhelt and F.Lottspeich

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

Communicated by D.Oesterhelt

The 'light' (L) and the 'medium' (M) subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis* were isolated and their amino-terminal sequences, as well as the sequences of several chymotryptic peptides, determined. *Rps. viridis* DNA was cloned in the *Escherichia coli* plasmid pBR322. Mixed oligonucleotide probes derived from the amino acid sequences were synthesised and utilised to isolate one clone which contained the genes for the L and M subunits of the reaction centre as well as the α and β subunits of the light-harvesting complex and part of the gene for the reaction centre cytochrome. The nucleotide sequences of the L and M subunit genes and the derived amino acid sequences are presented. The L subunit consists of 273 amino acids and has a mol. wt of 30 571. The M subunit consists of 323 amino acids and has a mol. wt of 35 902. The primary structure is discussed in the light of the recently published secondary and tertiary structure which has shown that both subunits contain five membrane-spanning helices.

Key words: hydropathy plot/membrane protein structure/photosynthesis/reaction centre/sequence

Introduction

The present knowledge of the three-dimensional structure of membrane proteins is limited since X-ray crystallography is the only method of determining the spatial structure of biological macromolecules at high resolution but well-ordered three-dimensional crystals of membrane proteins could not be obtained for a long time. Recently this situation has changed completely and membrane protein crystals diffracting X-rays to high resolution have now been grown both from bacterial photosynthetic reaction centres (Michel, 1982; Allen and Feher, 1984; Chang *et al.*, 1985) and from porin, an outer membrane protein of *Escherichia coli* (Garavito *et al.*, 1983).

Photosynthetic reaction centres are complexes comprised of integral membrane proteins and pigments, which catalyse light-driven electron transport across the photosynthetic membranes. Knowledge of the structure of photosynthetic reaction centres will help understanding of the mechanism of light-driven charge separation, biological electron transport and the principles governing membrane protein structure. Most of the photosynthetic reaction centres from purple photosynthetic bacteria (for review, see Feher and Okamura, 1978; Hoff, 1982; Parson, 1982) consist of three protein subunits which are called H (heavy), M (medium) and L (light) subunits according to their apparent mol. wts as determined by SDS–polyacrylamide gel electrophoresis (SDS–PAGE). The apparent mol. wts are misleading since at least in *Rhodopseudomonas capsulata* (Youvan *et al.*, 1984), which is now called *Rhodobacter (Rb.) capsulatus* (Imhoff *et al.*, 1984) and in *Rhodopseudomonas (Rps.) viridis* the H subunit is the smallest one. In *Rps. viridis* these subunits have apparent mol. wts of 35 000 (H); 28 000 (M) and 24 000 (L) (Thornber *et al.*, 1980). The L and M subunits are known to bind the photosynthetic pigments, since in *Rhodopseudomonas sphaeroides*, which is now called *Rhodobacter sphaeroides* (Imhoff *et al.*, 1984), and *Rhodospirillum (Rs.) rubrum* (Wiemken and Bachofen, 1984) the H subunit can be removed, and the residual complex remains photochemically active (Feher and Okamura, 1978; Wiemken and Bachofen, 1984). The reaction centres from several purple photosynthetic bacteria, e.g. that from *Rps. viridis*, contain a tightly bound cytochrome molecule, which re-reduces the photo-oxidised primary electron donor.

Recently, an electron density map at 3 Å resolution of the reaction centre from *Rps. viridis* has been calculated and the chromophore arrangement determined (Deisenhofer *et al.*, 1984). The sequences of the four protein subunits were then needed to trace the polypeptide chains unambiguously. The amino acid sequence of the H subunit has since been determined (Michel *et al.*, 1985). Here we report the isolation of the genes coding for the α and β subunits of the light-harvesting complex (LHC) and for the L, M and (part of) the cytochrome subunits of the reaction centre. The nucleotide sequences coding for the L and M subunits are presented along with the derived amino acid sequences. These sequences have been incorporated into the atomic model of the reaction centre (Deisenhofer *et al.*, 1985). The H, L and the M subunits of the photosynthetic reaction centre of *Rps. viridis* are thus the first membrane proteins whose structures are known at nearly atomic resolution. A detailed discussion of pigment–protein interaction, the distribution of amino acids within membrane proteins and an evaluation of the various prediction schemes for membrane-spanning helices now becomes possible.

Results

The protein sequence work was started with the hope of finding sequences which would allow the synthesis of a mixed oligonucleotide probe of low degeneracy. The following amino acid sequences were obtained (see also Figure 1): (i) amino-terminal residues 1–46 of the L subunit, (ii) the sequences of two chymotryptic peptides corresponding to residues 52–62 and 63–73 of L, (iii) the amino-terminal residues 1–30 of the M subunit and (iv) the sequence of one chymotryptic peptide corresponding to residues 170–175 of M. The amino acid sequence most suited for the synthesis of an oligonucleotide probe was L22–L26 (-F-D-F-W-V-) which allowed the synthesis of a tetradecamer with 8-fold degeneracy. Assuming that the gene arrangement may be the same as in the genome of *Rb. capsulatus*, (where the genes of the β and α subunits of the LHC are located just upstream of the L and M subunit genes, Youvan *et al.*, 1984), we also synthesised an octadecamer with 16-fold degeneracy corresponding to amino acids 46–51 (-N-W-W-E-F-Q-) of the α subunit of the LHC from *Rps. viridis* (Brunisholz *et al.*, 1985). This oligonucleotide probe was used for colony hybridisation. Among

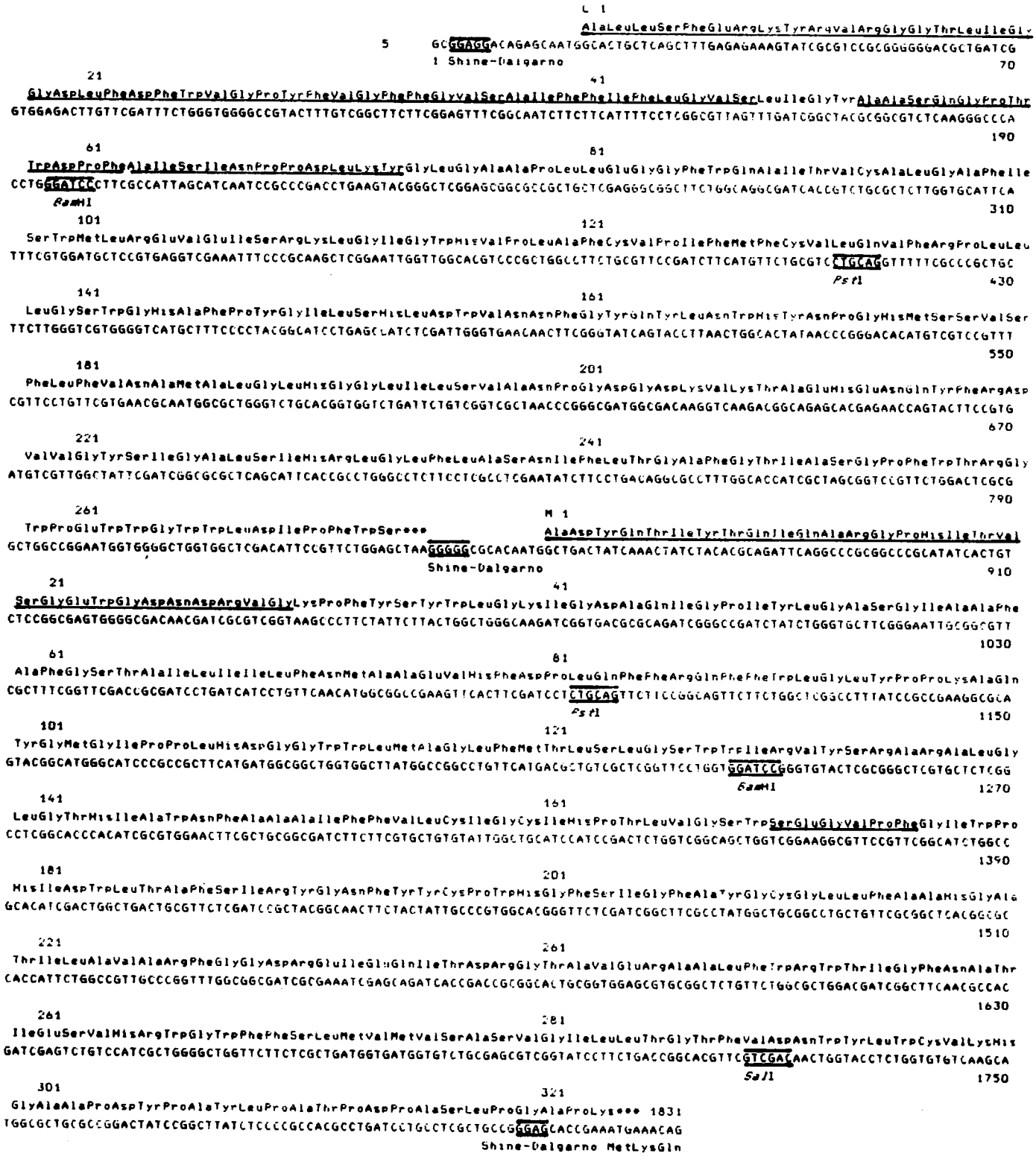

Fig. 1. Nucleotide sequence of the genes coding for the L and M subunits, the start of the signal sequence of the cytochrome subunit and the derived amino acid sequences. The nucleotide sequence of the non-coding strand is shown. Both strands were completely sequenced. Underlined amino acid sequences were confirmed by protein sequencing. The *Bam*HI, *Pst*I and *Sal*I retriction sites used in sequence work, as well as the Shine−Dalgarno sequences, are indicated.

5000 clones containing *Sau*3A restriction fragments of 5 − 10 kb which resulted from partial digestion, five were detected which hybridised with the octadecamer oligonucleotide probe. However, after isolation of plasmid DNA only one of the clones, LHC2, yielded restriction fragments upon digestion with endonucleases (*Bam*HI, *Sau*I and *Pst*) which hybridised on a Southern blot with the same oligonucleotide probe. The same restriction fragments also hybridised with the tetradecamer oligonucleotide probe derived from the amino acid sequence of the L subunit of the reaction centre, indicating that the genes of the L subunit of the

reaction centre and of the α subunit of the LHC are indeed close together. The *Apa*I fragment (2778 bp) containing the genes of the LHC I and the reaction centre L and M subunits of *Rb. capsulatus* became available later as a probe and also hybridised to the same restriction fragments as well as to a 2.3-kb *Eco*RI restriction fragment of pLHC2.

The α and β subunits of the LHC in *Rps. viridis* contain the sequence -Glu-Phe- each (see Brunisholz *et al.*, 1985). Bearing in mind the codon usage for the H subunit (see Michel *et al.*, 1985), a 50% chance existed that one Glu-Phe sequence corresponded to an *Eco*RI restriction site. It was therefore likely that the 2.3-kb *Eco*RI restriction fragment starts at the end of the LHC and would just encode the L and M subunits, if they were arranged in tandem as in *Rb. sphaeroides* and *Rb. capsulatus* (Williams *et al.*, 1983, 1984; Youvan *et al.*, 1984). The sequence of the 2.3-kb *Eco*RI fragment confirmed this assumption and contained, in addition to L and M, the first 77 amino acids of the cytochrome subunit of the reaction centre. Unfortunately, the rest of the cytochrome gene is not part of the insert in pLHC2 because the 2.3-kb *Eco*RI fragment is an artificial fragment which must have been created by ligating two *Sau*3A fragments prior to their insertion into the compatible *Bam*HI site of pBR322.

The genes of the L and M subunit initiate with ATG [coding for (N-formyl)methionine] and are preceded by Shine–Dalgarno sequences (Shine and Dalgarno, 1974) within the correct distance from the start. In the mature L and M subunits only the presumed N-formylmethionines are removed. This is in contrast to the cytochrome gene which starts with a typical bacterial signal peptide of 20 amino acids that is not found in the mature cytochrome. The carboxy terminus of the L subunit was confirmed by digestion with carboxypeptidase Y. In addition, the compatibility of electron density and derived amino acid sequence (including the carboxy termini) proves the position of the stop codons. Figure 1 shows the nucleotide sequences coding for the L and M subunits and the derived amino acid sequences. As expected, both the L and M subunits are very hydrophobic proteins. The L subunit consists of 273 amino acids, corresponding to a mol. wt of 30 531, while the M subunit has 323 amino acids and a mol. wt of 35 902. Both mol. wts are significantly higher than those determined by SDS–PAGE (24 000 and 28 000, respectively). In contrast to *Rb. sphaeroides* (Williams *et al.*, 1983, 1984) and *Rb. capsulatus* (Youvan *et al.*, 1984), the genes of the L and M subunits do not overlap. However, the last codon of the M subunit and the start codon of the cytochrome gene overlap by a single base.

Quite interestingly the codon usage for the L and M subunits seems to be significantly different from that of the H subunit as shown in Table I. With the exception of CGC/CGT coding for arginine, the use of T in the third position is significantly higher in the L and M subunit genes than in the H subunit gene. This is particularly striking for the codons of aspartic acid, threonine, isoleucine and leucine. Codons containing A in the third position are rarely used in all three subunits.

Table I. Codon usage for the L and M subunits from *Rps. viridis*

5	(0) UUU Phe	4	(0) UCU Ser	12	(4) UAU Tyr	2	(0) UGU Cys
46	(6) UUC Phe	4	(1) UCC Ser	13	(8) UAC Tyr	6	(1) UGC Cys
0	(1) UUA Leu	0	(0) UCA Ser	1	(0) UAA ***	1	(1) UGA ***
2	(1) UUG Leu	17	(7) UCG Ser	0	(0) UGA ***	33	(6) UGG Trp
8	(1) CUU Leu	3	(0) CCU Pro	7	(1) CAU His	4	(9) CGU Arg
17	(10) CUC Leu	6	(2) CCC Pro	11	(3) CAC His	13	(9) CGC Arg
0	(0) CUA Leu	0	(0) CCA Pro	2	(2) CAA Gln	0	(0) CGA Arg
32	(18) CUG Leu	25	(16) CCG Pro	11	(7) CAG Gln	4	(4) CGG Arg
12	(0) AUU Ile	6	(0) ACU Thr	2	(1) AAU Asn	1	(1) AGU Ser
32	(8) AUC Ile	8	(7) ACC Thr	13	(1) AAC Asn	8	(2) AGC Ser
0	(0) AUA Ile	1	(1) ACA Thr	1	(0) AAA Lys	1	(0) AGA Arg
10	(1) AUG Met	8	(8) ACG Thr	9	(9) AAG Lys	0	(0) AGG Arg
9	(6) GUU Val	14	(6) GCU Ala	10	(1) GAU Asp	14	(3) GGU Gly
16	(15) GUC Val	14	(10) GCC Ala	11	(16) GAC Asp	43	(16) GGC Gly
0	(1) GUA Val	6	(1) GCA Ala	5	(10) GAA Glu	7	(0) GGA Gly
8	(7) GUG Val	22	(9) GCG Ala	9	(9) GAG Glu	9	(0) GGG Gly

The codon usage for the H subunit (Michel *et al.*, 1985) is given in brackets.

Discussion

Genetic aspects

The sequence work described here shows that the arrangement of the genes encoding the L and M subunits is the same as in *Rb. sphaeroides* (Williams *et al.*, 1983, 1984) and *Rb. capsulatus* (Youvan *et al.*, 1985). However, due to a deletion of five bases at the end of the gene coding for L in *Rps. viridis*, the genes coding for L and M do not overlap. Therefore, overlap of the L and M genes is not necessary to assure a 1:1 stoichiometry of the L and M subunits as was previously discussed (Williams *et al.*, 1984). The gene encoding the reaction centre cytochrome is located at the end of the gene encoding the M subunit and overlaps the M gene by 1 bp. This differs from *Rb. capsulatus* and *Rb. sphaeroides* where the reaction centres do not contain this type of cytochrome molecule. Our preliminary sequence work also shows that the genes encoding the β and α subunits of the LHC are located upstream from the genes encoding the L and M subunits in the same order as in *Rb. capsulatus* (Youvan *et al.*, 1984). We expect that all five genes form one polycistronic operon.

The L and M subunits are integral membrane proteins and contain five membrane-spanning helices each (Deisenhofer *et al.*, 1985). They have their amino termini on the cytoplasmic side of the membrane, as has also been shown for the L subunit of the reaction centre from *Rs. rubrum* (Brunisholz *et al.*, 1984). The N-formylmethionine residues must have been removed by post-translational modification. In contrast, the H subunit which has its amino terminus on the periplasmic side of the membrane (Deisenhofer *et al.*, 1985) still contains the N-formylmethionine (Michel *et al.*, 1985). The reason for this difference may be that the peptidase responsible for the removal of the N-formylmethionine is located in the cytoplasm and the amino terminus of the H subunit moves across the membrane immediately after, or during, translation so that its amino terminus is inaccessible to the peptidase. The cytochrome subunit, on the other hand, which is located in the periplasm, has to cross the membrane entirely. Our preliminary gene sequence shows that it possesses a typical bacterial signal, or leader, sequence of 20 amino acids which is removed in connection with the transport of the protein across the membrane. One might therefore be able to infer the location of the amino terminus of membrane proteins in *Rps. viridis* by comparing gene and protein sequence in fortunate cases. However, the presence of N-formylmethionine at the amino terminus may also be due to reasons other than its location in the periplasmic space.

A surprising observation is the difference in codon usage of the genes coding for the L and M subunits when compared with the H subunit gene (Table I). A similar observation has been made in the case of the D1 and D2 proteins of photosystem II in spinach chloroplasts (Alt *et al.*, 1984). The difference could be related

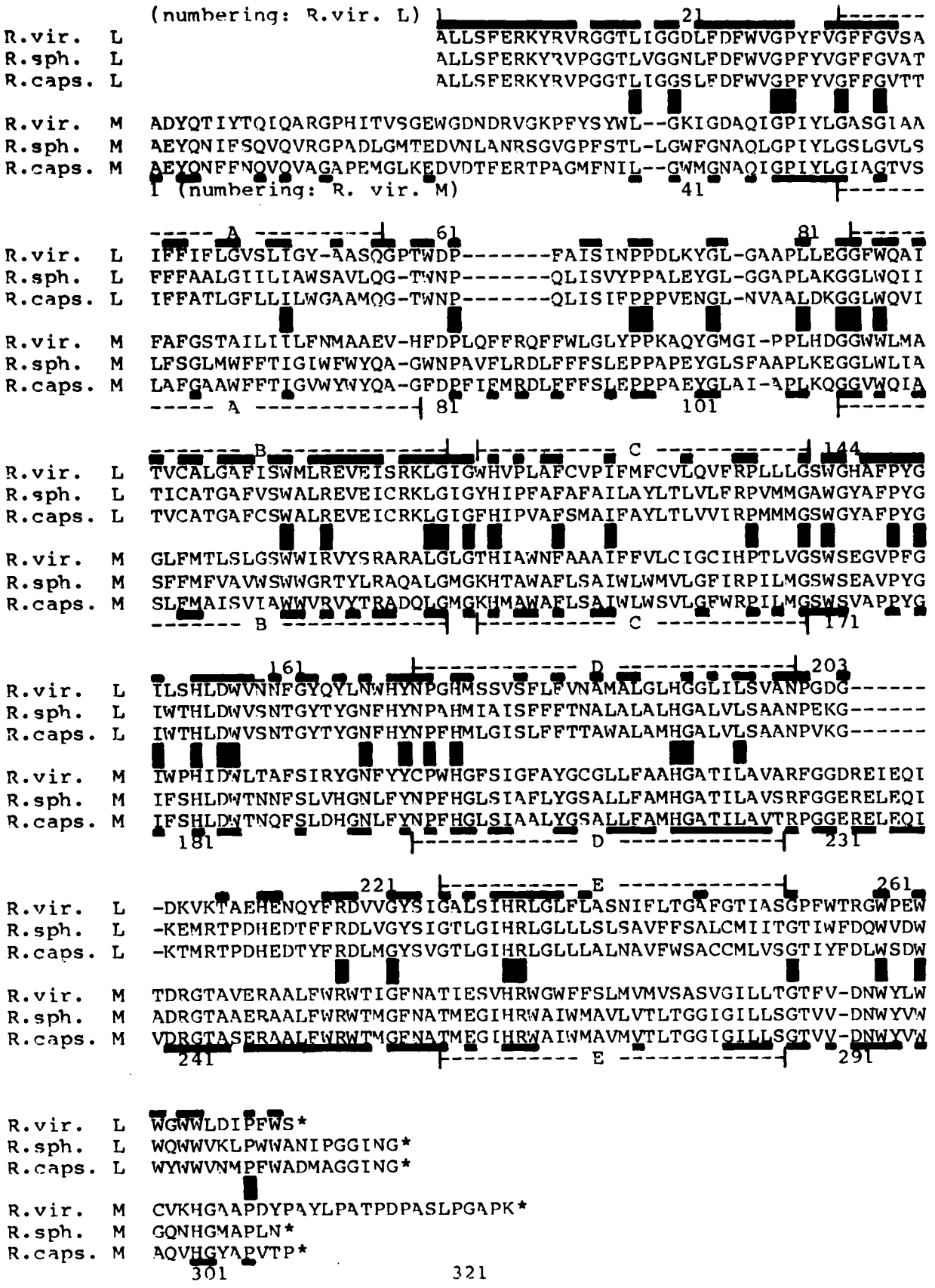

Fig. 2. Alignment of amino acid sequences of the L and M subunits from *Rps. viridis*, *Rb. sphaeroides* (Williams *et al.*, 1983, 1984) and *Rb. capsulatus* (Youvan *et al.*, 1984). The position of the transmembrane helices in *Rps. viridis* (A–E, Deisenhofer *et al.*, 1985) is indicated by dotted lines with a letter indicating the corresponding helix. Residues conserved among all six subunits are indicated by black boxes between the L and M subunits. Residues conserved between the L subunits are indicated by black bars above the sequence of the L subunit from *Rps. viridis*. Residues conserved between the M subunits are indicated by black bars below the M subunit from *Rb. capsulatus*.

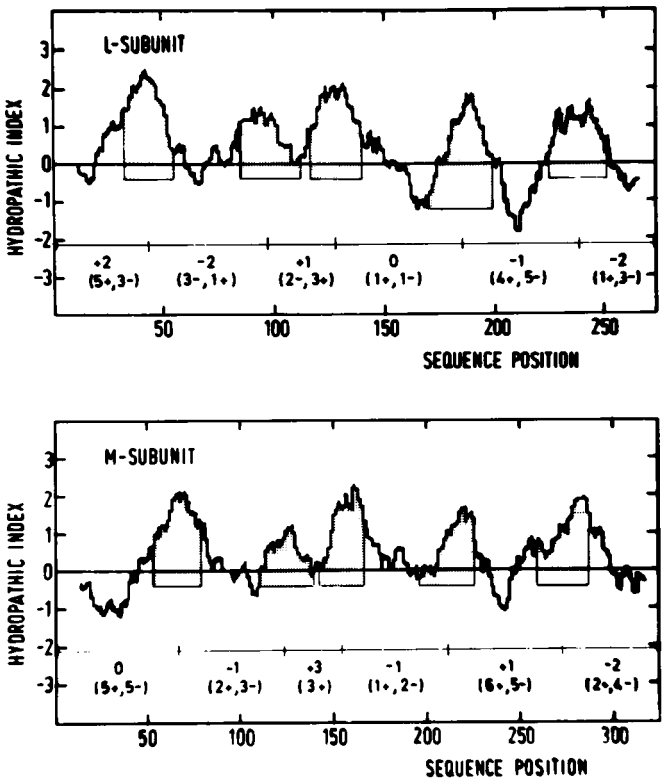

Fig. 3. Hydropathy plot of the L and M subunits according to Kyte and Doolittle (1982). A window of 19 amino acids was used. The hydropathic index is plotted against the 10th amino acid of the window. The position of the transmembrane helices as found by X-ray structural analysis is indicated by the dotted area. The numbers given at the bottom of the plots indicate the net charges of the connections starting with the cytoplasmic N terminus. In brackets the respective numbers of positively and negatively charged amino acids are given.

to regulatory phenomena. If, for example, one of the protein subunits is to be synthesised faster than other components after induction of the photosynthetic apparatus, or this protein has a more rapid turnover, it would be of advantage to use those codons whose tRNAs occur in high concentrations in the cells. At present, the levels of the various tRNA species are not known in *Rps. viridis*. If the tRNAs for the codons which have T in the third position are present in low levels the rate of synthesis of the L and M subunits would be decreased compared with the H subunit. In this context it is interesting to note that the H subunit of the bacterial reaction centre has been proposed to initiate the assembly of the reaction centre complex (Chory *et al.*, 1984). This would require a faster synthesis of the H subunit and would be possible on the basis of different codon usage compared with L and M subunits.

General aspects of membrane protein structure

The aim of this work has been to provide the sequence information needed for a reliable interpretation of the electron density map (Deisenhofer *et al.*, 1984). The sequences presented here have been incorporated into the atomic model of the reaction centre (Deisenhofer *et al.*, 1985) and, after correcting a few minor errors in chain tracing, were found to fit very well into the electron density map. The electron density map is incomplete for only a small fraction of the side chains, probably due to disorder in the crystal. With the availability of the amino acid sequences and the three-dimensional arrangement of the amino acids, a detailed discussion of pigment protein interaction and membrane protein structure becomes possible. The reaction centre from *Rps. viridis* will also be a valuable test object for the various methods of predicting membrane-spanning helices in membrane proteins with known sequence.

The X-ray structure analysis has shown that residues L32–L55 (helix LA), L84–L112 (LB), L115–L140 (LC), L170–L199 (LD), L225–L251 (LE) as well as M52–M78 (MA), M110–M139 (MB), M142–M167 (MC), M197–M225 (MD) and M259–M285 (ME) form membrane-spanning helices. The positions of these helices in the sequences are indicated in Figures 2 and 3. Most of them possess glycine residues, sometimes prolines, at, or very close to, both ends. This is probably because these amino acids allow the formation of turns in the peptide chains more easily. Several of these helices contain up to five charged residues at one of their ends, but there is always a stretch of 19 amino acid residues which does not contain any basic or acidic side chain. Different helices carry the acidic or basic side chains close to their cytoplasmic or periplasmic end, but never at both. Neither the exact position of the membrane nor the thickness of the hydrophobic part are known. A reliable determination of the thickness of lipid bilayers has been made using bilayer membranes prepared from 1,2-dipalmitoyl-glycero-3-phosphocholin (Büldt *et al.*, 1978). Under biologically relevant conditions (above phase transition, fully hydrated) a thickness of 30–31 Å was found for the hydrocarbon core of the bilayer. The predominant fatty acids in the photosynthetic membranes from *Rps. viridis* possess 18 carbon atoms (Welte and Kreutz, 1982). Therefore a thickness of ~36 Å must be expected for the hydrophobic core of the bilayer in photosynthetic membranes from *Rps. viridis*. Since the membrane-spanning helices consist of 24–30 amino acids, and are tilted up to 38°, they cannot extend much from the hydrophobic core of the bilayer.

If the helices contain more than one basic or acidic residue then these residues are found on the cytoplasmic side of the membrane. These side chains, therefore, do not have to be transferred across the membrane during biosynthesis and incorporation of the protein into the membrane. They could well serve as a stop-transfer or stop-insertion signal. The membrane-spanning helix of the H subunit also contains five charged residues on the cytoplasmic side of the membrane. If one includes the basic and acidic amino acids of the connecting loops and the termini, considerably less acidic and basic residues are found on the periplasmic side of the membrane. The L subunit has 17 basic or acidic residues on the cytoplasmic side of the membrane and only 10 on the periplasmic side, similarly M has 24 acidic or basic residues on the cytoplasmic side and 14 on the periplasmic side.

In Table II the amino acid compositions of the transmembrane helices of the L and M subunit are compared with that of the helix connections, and a distribution factor indicating the enrichment (or depletion) of each amino acid in the transmembrane helices is given. To broaden the basis for this comparison the respective compositions of the presumed transmembrane helices and connecting segments of the *Rb. sphaeroides* and the *Rb. capsulatus* reaction centres (see below) have been included. The sulphur-containing amino acids cysteine and methionine are clearly over-represented in the transmembrane helices. To our surprise alanine and serine are also over-represented in the membrane-spanning helices, whereas the high frequency of histidine occurrence is explained by the fact that most of the histidines are involved in binding of the non-heme iron and the bacteriochlorophylls. Glycine is nearly equally distributed between the membrane-spanning helices and their connections. Tryptophan is under-represented in the membrane-spanning helices, but we

Table II. Amino acid compositions of transmembrane helices (H) and the connections (C) of the L and M subunits from *Rps. viridis*, *Rb. sphaeroides* and *Rb. capsulatus*

	Rps. viridis				*Rb. sphaeroides*				*Rb. capsulatus*				H		C		Distribution factor		
	L		M		L		M		L		M								
	H	C	H	C	H	C	H	C	H	C	H	C	sum	%	sum	%	ϕ	$1/\phi$	
Ala	14	6	21	15	22	7	15	16	18	6	20	15	110	13.33	65	6.87	1.94	0.52	Ala
Arg	4	5	4	9	4	4	4	8	4	4	4	7	24	2.91	37	3.91	0.74	1.34	Arg
Asn	4	5	2	4	3	6	1	12	3	7	1	8	14	1.70	42	4.44	0.38	2.61	Asn
Asp	0	9	0	12	0	7	0	6	0	9	1	8	1	0.12	51	5.39	0.022	44.92	Asp
Cys	3	0	4	1	3	0	0	0	5	0	0	0	15	1.82	1	0.11	16.55	0.06	Cys
Gln	3	2	0	8	1	4	2	6	1	2	3	8	10	1.21	30	3.17	0.38	2.62	Gln
Glu	2	5	2	5	2	5	1	10	2	3	1	10	10	1.21	38	4.02	0.30	3.32	Glu
Gly	16	19	17	21	11	21	18	18	13	19	17	16	92	11.15	114	12.05	0.93	1.08	Gly
His	4	4	6	4	4	3	4	3	4	3	4	3	26	3.15	20	2.11	1.49	0.67	His
Ile	11	7	13	13	14	7	12	4	9	7	13	7	72	8.73	45	4.76	1.83	0.55	Ile
Leu	20	12	15	12	22	11	21	17	21	10	15	12	114	13.82	74	7.82	1.77	0.57	Leu
Lys	1	4	0	5	1	5	1	1	1	4	1	2	5	0.61	21	2.21	0.28	3.62	Lys
Met	4	0	5	1	4	1	7	4	8	4	7	6	35	4.24	16	1.69	2.51	0.40	Met
Phe	16	10	12	13	15	9	12	14	13	10	8	15	76	9.21	71	7.51	1.21	0.83	Phe
Pro	4	11	2	17	3	11	2	10	3	10	2	13	16	1.94	72	7.61	0.25	3.92	Pro
Ser	12	6	10	6	7	4	10	9	6	7	8	6	53	6.42	38	4.02	1.60	0.63	Ser
Thr	3	4	7	9	7	8	7	6	9	9	8	7	41	4.97	43	4.54	1.09	0.91	Thr
Trp	3	12	8	10	3	13	12	8	5	11	12	7	43	5.21	61	6.44	0.81	1.24	Trp
Tyr	1	9	2	13	2	10	3	6	1	11	4	7	13	1.58	56	5.92	0.27	3.75	Tyr
Val	11	7	9	6	8	9	7	10	10	9	10	10	55	6.67	51	5.39	1.24	0.81	Val
sum	136	137	139	184	136	145	139	168	136	145	139	167	825	100	946	100			

To calculate the percentage of each amino acid in the helices and connections the sum of all amino acids in the helices or connections was set as 100%. The distribution factor (ϕ) was calculated as ratio of the percentage of the respective amino acid in the helices and in the connections.

have to note the occurrence of a conserved cluster of five tryptophan residues at the ends of the L subunits from all three bacterial species. Arginine is only enriched in the connections by a factor of 1.35, which is not much considering its strong basicity. The amino acids which are most under-represented in the transmembrane helices are tyrosine, proline and, in particular, aspartic acid.

The distribution factors in Table II may be used as the basis for an improved hydropathy scale to predict membrane-spanning helices. One outstanding feature is that the amino acids preferring the membrane-spanning helices are less enriched in the helices than the amino acids preferring the connections are enriched in the connections: only two amino acids, cysteine and methionine, are enriched >2-fold in transmembrane helices, whereas six amino acids, asparagine, glutamine, glutamic acid, lysine, tyrosine, proline and aspartic acid, are enriched >2-fold in the helix connections (compare the ϕ and $1/\phi$ values of the respective amino acid in Table II). This observation should result in an 'asymmetric' hydropathy scale, where the positive values assigned to the hydrophobic amino acids are smaller than the negative values assigned to the polar amino acids, compared with a symmetric one like the Kyte–Doolittle scale (Kyte and Doolittle, 1982).

Another weakness of the currently used scales (von Hejne, 1981a,b; Kyte and Doolittle, 1982; Engelman and Steitz, 1981) is that they are mainly based on thermodynamic parameters like solubilities of amino acid chains in organic solvents and water, or vapour pressures of side chain analogs. They do not consider the fact that charges may be needed for the function of a particular membrane protein. If, for example, a positive charge would be needed, it is of advantage to use arginine instead of lysine, which has a lower p*K*. This may explain why arginine is more frequently found in the membrane-spanning helices of the reaction centres than lysine.

A problem not overcome by any general predictive scheme for transmembrane helices are the different functions of the proteins, e.g. electron transfer in the reaction centre, proton translocation in bacteriorhodopsin and chloride translocation in halorhodopsin. In the reaction centres, tyrosines are enriched nearly 4-fold in the helix connections. In bacteriorhodopsin, where the transmembrane helices are well defined by various biochemical experiments (see Ovchinnikov *et al.*, 1985, and references therein), tyrosines are nearly exclusively found within membrane-spanning helices. This result could mean that either tyrosines in the membrane-spanning helices of the reaction centres would negatively interfere with the light-driven electron transport, or that in bacteriorhodopsin these tyrosines are involved in the function. A similar situation is found with respect to aspartic acid where only the M subunit of *Rb. capsulatus* carries one residue in a membrane-spanning helix, whereas in the helices of bacteriorhodopsin four aspartic acids occur. Thus it will be very difficult, or impossible, to define and use general hydropathy scales or membrane propensities for amino acids and to predict precisely membrane-spanning helices in membrane proteins of different function.

Figure 3 compares the result of the most popular scheme to predict membrane-spanning helices, the 'Kyte–Doolittle plot' (Kyte and Doolittle, 1982) and the actual position of the membrane-spanning helices. It is clear that (when a large window is used) five maxima of hydrophobicity appear which correspond roughly to the hydrophobic parts of the membrane-spanning helices. It is also clear that the ends of the helices are predicted rather imprecisely. There must be some doubts whether a prediction method can be developed which predicts that the last 10 residues of the membrane-spanning helix B of the L subunit, containing a total of five basic and acidic residues, are in fact part of a transmembrane helix. Application of other methods which predict

turns of the peptide chain (e.g. Paul and Rosenbusch, 1985) also fail to trace most of the ends of the transmembrane helices.

If one counts (assuming that glutamic and aspartic acids are always negatively charged and lysine and arginine are always positively charged) the net charges of the peptide chains on the periplasmic side of the membrane and compares them with the net charges on the cytoplasmic side of the membrane, one finds that the polar ends of the transmembrane helices and their respective helix connections on the cytoplasmic side are nearly always less negatively charged than the polar ends of the transmembrane helices and their respective helix connections on the periplasmic side of the membrane. This finding is documented in Figure 3. Within the M subunit the amino-terminal peptide chain located on the cytoplasmic side and preceding the first transmembrane helix, carries no net charges. The (periplasmic) connection of the first and second transmembrane helix including the polar ends of the helices carries one negative charge, the (cytoplasmic) connection of the second and third transmembrane helices including the polar ends of the helices carries three positive charges, the next interhelix linkage on the periplasmic side one negative charge, the next (cytoplasmic) connection one positive charge, and the carboxy terminus on the periplasmic side two negative charges. As a result the cytoplasmic side of the M subunit carries four positive net charges and the periplasmic side four negative net charges. This asymmetric charge distribution across the membrane leads to an electrical dipole moment perpendicular to the plane of the membrane. The same consideration for the L subunit reveals two positive net charges on the cytoplasmic side and four negative net charges on the periplasmic side. The electric polarity of the L and M subunits becomes even more pronounced if one considers that the reaction centres from the purple bacteria contain a firmly bound ferrous non-heme iron close to the cytoplasmic side of the membrane carrying two positive charges. The connection between the fourth and fifth transmembrane helix of the L subunit showing one negative net charge is therefore overcompensated by being involved in the iron binding and in binding the secondary quinone (Deisenhofer *et al.*, 1985).

With the exception of those bacteria living in an acidic environment, bacteria are negatively charged inside the cells due to the action of electrogenic ion pumps. Thus the L and M subunits are oriented in the cells with their dipoles in the energetically more favourable direction. *Vice versa*, a membrane potential may be needed for the correct incorporation of the L and M subunits into the membrane and correct protein folding. If a long hydrophobic helix, or a pair of helices (see Engelman and Steitz, 1981) has been formed during protein translation it may become associated with the membrane. Then the membrane potential and the net charges at the ends of the helix (or helix pair) may decide which end will be transferred to the periplasmic side of the membrane. The question whether the electric dipole moments of the L and M subunits influence the direction and speed of the light-driven electron transfer will be discussed elsewhere since one has to consider also the possible formation of salt bridges with the H subunit and the cytochrome.

Sequence comparisons

Reaction centres from other purple photosynthetic bacteria. Figure 2 shows the aligned sequences of the L and M subunits of *Rps. viridis*, *Rb. sphaeroides* and *Rb. capsulatus*. As has been previously noted (Williams, *et al.*, 1983, 1985; Youvan *et al.*, 1984), there exists considerable sequence homology between the L and M subunits of all these species (black boxes between the L and M subunits in Figure 2) indicating that their gene arose by gene duplication of one single gene. The architecture of the reaction centre of current day's purple bacteria shows that the reaction centres of their predecessors must have been a symmetric dimer of two identical protein subunits containing four bacteriochlorophylls, two bacteriopheophytins and two quinones. The reaction centre contained two identical pathways of electron transport across the membrane (for discussion see Michel and Deisenhofer, 1986). The L and M subunits from *Rps. viridis* still have 72 amino acid residues in common, corresponding to a sequence homology of 22% [common residues as percent of total residues in M (longer sequence)] or 26.4% [common residues as percent of total residues in L (shorter sequence)]. The major differences between the L and M subunits are longer amino termini in all three M subunits and a longer carboxy terminus in the M subunit of the *Rps. viridis* reaction centre. All three M subunits also contain two inserts of seven amino acids. The first of the insertions is located in the connection of the first and second membrane-spanning helices and creates a short helix parallel to the membrane (see Deisenhofer *et al.*, 1985). The second insertion is found in the connection of the fourth and fifth membrane-spanning helix and creates an additional loop which provides glutamic acid-M232 as a fifth ligand to the ferrous non-heme iron atom. Residues conserved between the L and M subunits of the three bacteria are mainly the glycine and proline residues at, or close to, the ends of helices and other turns of the peptide chains. The ligands to the bacteriochlorophyll magnesium atoms and the non-heme irons are also conserved. The overall sequence homologies between the L subunits are 57.8 (59.5)% for *Rps. viridis* and *Rb. sphaeroides*, and 57.4 (59.1)% for *Rps. viridis* and *Rb. capsulatus*, and between the M subunits 46.2 (48.9)% for *Rps. viridis* and *Rb. sphaeroides*, and 47.4 (50.0)% for *Rps. viridis* and *Rb. capsulatus*. (The lower and higher percentage values are used as defined above.) These values are unexpectedly low. On the other hand, apart from the deletion at the end of the L subunit (see Results) there is only a deletion of one amino acid at the end of the first transmembrane helix in L of the *Rps. viridis* reaction centre and this deletion is cancelled by an insertion of one amino acid at a position six amino acids later in the sequence. The M subunit from *Rps. viridis* has a carboxy-terminal extension of 17 (18) amino acids compared with that of *Rb. capsulatus* (*Rb. sphaeroides*). This extension is in contact with the cytochrome subunit (see Deisenhofer *et al.*, 1985). Most of the amino acid changes are conservative, thus the structure of the cores of all three reaction centres must be very similar. *Rb. capsulatus* and *Rb. sphaeroides* are more closely related to each other (sequence homologies of 77 and 78%, see Williams *et al.*, 1984) than either *Rb. capsulatus* and *Rps. viridis* or *Rps. viridis* and *Rb. sphaeroides*. This relationship is the same as that deduced from sequence comparison of both cytochrome c_2 and 16S RNA (Gibson *et al.*, 1979).

The amino acid sequences of the transmembrane helices are not better conserved than those in the helix connections: among the L subunits the sequences of the second transmembrane helix are well conserved whereas the sequences of the third and fifth transmembrane helices are less conserved. Likewise the degree of conservation among the connections of the transmembrane helices also varies as shown by the lower degree of conservation in the helix connections AB than in CD or DE.

Proteins of photosystem II reaction centre. A weak, but significant, sequence homology exists between the L and M subunits and the protein from photosystem II reaction centre, known as D1, Q_B, herbicide-binding or 32-kd protein (Zurawski *et al.*,

Fig. 4. Comparison of the amino acid sequences of the L and M subunits from *Rps. viridis* and the D1 and D2 proteins from spinach chloroplasts (Zurawski *et al.*, 1982; Alt *et al.*, 1984). Residues common to all four subunits are marked by dots above the L subunit sequence, residues conserved between L and D1, and between M and D2 are boxed. Two possible alignments **(a)** and **(b)** are given for the carboxy-terminal sequences. The histidine ligands to the magnesium atoms of the bacteriochlorophylls of the special pair in *Rps. viridis* (sp.p.) and to the non-heme iron (Fe) are indicated. Tryptophans forming the major part of the binding site for the primary quinone (see Deisenhofer *et al.*, 1985) are also labelled (Q).

1982; Williams *et al.*, 1983, 1984; Youvan *et al.*, 1984). Later on another photosystem II reaction centre protein, the D2 protein, was also found to possess sequence homologies to the D1 protein and the L and M subunits (Rochaix *et al.*, 1984; Alt *et al.*, 1984; Holschuh *et al.*, 1984). Both the L subunits and the D1 proteins bind the photoaffinity-labelled herbicide azidoatrazine (Pfister *et al.*, 1981; DeVitry and Diner, 1984). In addition, the X-ray structure analysis has shown that both the L and M subunits are needed to establish the primary electron donor, which is made up of two nearly parallel bacteriochlorophyll molecules ('special pair'), and the electron accepting quinone iron complex. These and other findings have led to the proposal (Michel and Deisenhofer, 1986) that the D1 and D2 proteins form the core of the photosystem II reaction centre, in a similar way to the L and M subunits forming the reaction centre core from purple photosynthetic bacteria. The L subunit is assumed to be the equivalent of D1 and M to that of D2. Figure 4 shows the aligned sequences of the four proteins. Significant sequence homology starts at the Gly-Gly pair (M110,111; L83,84) at the beginning of the transmembrane helix B. Other amino acids of structural importance are also conserved, such as glycine M176 (L149) at the start of the short helix parallel to the membrane in the C−D connection, as well as the histidine ligands to the magnesium atoms of the special pair chlorophylls (M200, L173) and to the ferrous non-heme iron close to the end of transmembrane helix D. Between the start of transmembrane helix B and the end of transmembrane helix D only one single insertion in the bacterial subunits exists. The transmembrane helix connections D−E, which form the binding sites for the primary quinone (M subunit) and the secondary quinone (L subunit) are considerably longer in D1 and D2. However, tryptophan M250, which forms part of the binding site of the primary quinone, is conserved between M and D2. A phenylalanine (L216) is found in the equivalent position in the binding site of the secondary quinone and is conserved between L and D1. When this phenylalanine (255 in D1 of *Chlamydomonas*) mutates to tyrosine, then photosystem II reaction centre becomes resistant to the herbicide atrazine (Erickson *et al.*, 1985) which is thought to be a competitive inhibitor of the secondary quinone (Pfister *et al.*, 1981). The structural and genetic data are in good agreement, thus confirming the proposal that the L subunit corresponds to D1 and the M subunit to D2. More similarities between L and D1 and M and D2 are indicated in Figure 4.

Two possibilities exist for the alignment of the sequences after the quinone binding sites, both of which are indicated in Figure 4. The upper alignment (a in Figure 4) leads to a slightly higher sequence homology. In this alignment histidines M264 and L230, which are ligands of the ferrous non-heme iron in the reaction centre from *Rps. viridis*, would not be conserved in D1 and D2. There are, however, histidine residues in the D1 and D2 proteins three residues later in the sequences. If in this sequence region L and D1 and M and D2 would have exactly the same protein folding, then the histidine residues in the D1 and D2 proteins could not be iron ligands. The other alignment (b in Figure 4) would allow these histidine residues to be ligands to the iron atom, but the sequence homology is slightly poorer. It may well be that the alignment (a) better reflects the evolutionary homology, whereas the alignment (b) shows more clearly the structural similarities. This would mean that the transmembrane helix E could have been shifted perpendicular to the membrane to allow the D1 and D2 histidine residues to interact with the ferrous non-heme iron atom. The most important difference between the cores of the bacterial and the photosystem II reaction centres may be that the histidine ligands to the magnesium atoms of the accessory bacteriochlorophylls are not conserved, which could mean that in the photosystem II reaction centre the histidines do not act as ligands of the accessory chlorophylls, if such accessory chlorophylls exist at all.

Materials and methods

Materials

The sources of the materials used corresponded to those described in a previous paper (Michel *et al.*, 1985).

Growth of Rps. viridis and isolation of reaction centres

The procedures used to growth *Rps. viridis* and to isolate the reaction centres were described recently (Michel *et al.*, 1985).

Isolation of L and M subunits

SDS and dithiothreitol (DTT) were added to a concentrated reaction centre solution (10 mg/ml) to give a final concentration of 8% and 0.5%, respectively. After 1 h at room temperature the L and M subunits were separated from the other subunits by molecular sieve chromatography in 4% SDS, 2 mM DTT, 0.02% NaN_3 (column length: 120 cm) using an AcA 44 column (LKB). L and M subunits eluted together in one peak. Separation of L and M subunits was achieved by two methods:

Preparative electrophoresis. The fractions containing L and M subunits were concentrated by ultrafiltration (Amicon PM 10 membrane) followed by the re-addition of a solution containing 2 mM DTT and 0.02% NaN_3. This procedure was repeated until the SDS concentration of the concentrated sample was <3%. It was then subjected to preparative electrophoresis using an LKB 7900 Uniphor apparatus. The method of Waehnelt (1971) was used with the following modifications: the length of the stacking gel was increased to 1.9 cm and the sample volume was reduced to 1.5 ml.

Molecular sieve chromatography. The fractions containing the L and M subunits were concentrated by ultrafiltration and re-chromatographed on an AcA 44 column (column length 120 cm) in a buffer containing 4% SDS, 400 mM Tris-HCl, pH 8.8, 2 mM DTT, 0.02% NaN_3. The increase in buffer concentration and/or pH led to a separation of the L and M subunits on the column that otherwise left these subunits undissociated but removed the bound cytochrome and the H subunit. Fractions enriched with one subunit were re-chromatographed once.

Determination of N-terminal sequences. The subunits were precipitated by the addition of 4 volumes ice-cold ethanol, washed with 80% ethanol four times, dried and dissolved in concentrated formic acid and subjected to Edman degradation as already described (Michel *et al.*, 1985).

Preparation and isolation of chymotryptic peptides. Protein subunits, isolated by molecular sieve chromatography, were suspended in 50 mM NH_4HCO_3 (containing 0.1% SDS in the case of the L subunit) to a final concentration of 2 mg/ml in a total volume of 5 ml. Chymotrypsin (0.75 mg) was added and, after incubation at 37°C for 24 h, the samples were lyophilised and resuspended in 4 ml 0.16% trifluoroacetic acid (TFA). Unsolubilised material was removed by centrifugation (2 min, 8000 *g*) and 1 ml aliquots were applied to a reversed phase column (Lichrosorb RP 18, Merck). A gradient from 0.16% TFA in water to 0.12% TFA in acetonitrile was used to elute the peptides. Isolated peptides were re-chromatographed on a second reversed phase column (ODS, Shandon) using a gradient from 25 mM triethylammonium acetate (TEAA) pH 6.0 in water to 60% acetonitrile.

DNA isolation

Genomic DNA was isolated from *Rps. viridis* by resuspending 20 g of cells (wet weight) in 1 l of 0.1 M Tris-HCl, 50 mM EDTA, pH 7.5 followed by the addition of 300 mg lysozyme (egg white) and incubation at room temperature for 5 min. Pronase (200 mg, pre-incubated) and sodium dodecyl sarcosinate (final concentration 3%) were added and the suspension was incubated for 4 h at 60°C. After centrifugation (28 000 *g*; 1 h) the DNA was found in the viscous upper part of the pellet. It was removed with a spatula, solubilised in 200 ml of 50 mM Tris-HCl, 1 mM EDTA, pH 7.5 and incubated after addition of 60 mg pronase at 60°C for another hour. The DNA was precipitated with 2.5 volumes of ethanol in the presence of 1 M LiCl, coiled around a glass rod, and thoroughly washed with ethanol. It was resolubilised in 200 ml of 50 mM Tris-HCl, 1 mM EDTA, pH 7.5, and phenolised three times. The DNA was precipitated and washed again, solubilised in 75 ml of 50 mM Tris-HCl, 1 mM EDTA, pH 7.5, and incubated in the presence of 3.7 mg RNase A (pre-incubated at 90°C) at 60°C for 1 h. It was further purified on an isopycnic CsCl gradient. Plasmid DNA from *E. coli* was isolated using the Triton X-100 lysis procedure (Kahn *et al.*, 1979) and purified on a CsCl gradient in the presence of ethidium bromide.

Cloning

Rps. viridis DNA was partially digested using the restriction endonuclease *Sau*3A, size fractionated by sucrose density gradient centrifugation, and 5–10 kb fragments were ligated into the *Bam*HI site of pBR322 as described (Vogelsang *et al.*, 1983). The 2.3-kb *Eco*RI fragment of pLHC2 was isolated by agarose electrophoresis, and digested with endonuclease *Sau*3A.

Oligonucleotide probes

Two mixed oligonucleotide probes, a tetradecamer of 8-fold degeneracy and an octadecamer of 16-fold degeneracy, were synthesised by the phosphoamidite method using an Applied Biosystems 380A oligonucleotide synthesiser. The protecting groups were removed with concentrated NH_3 overnight and the oligonucleotides purified on a reversed phase RP 18 column in a 0.1 M TEAA buffer, pH 7.2, with a gradient of 5–100% acetonitrile. Detritylation was achieved by incubation of the oligonucleotide in 80% acetic acid for 45 min at room temperature. The acetic acid was removed in a speed vac centrifuge. Traces of acetic acid were removed by resolubilising the oligonucleotides in ethanol and drying the samples in the speed vac.

Colony hybridisation

The oligonucleotides were labelled at their 5′ ends using terminal deoxyribonucleotide transferase as described (Vogelsang *et al.*, 1983). The protocol used for colony hybridisation was as in the same reference with the modification that 0.1 mg of poly(C) were added per ml, and the hybridisation temperatures were 32°C for the tetradecamer nucleotide and 40°C for the octadecamer nucleotide. Alternatively, the 2778-bp *Apa*I restriction fragment coding for the L and M subunits of the reaction centre and the α and β subunits of the LHC I from *Rb. capsulatus* (Youvan *et al.*, 1984) was isolated, labelled with [^{32}P]dCTP by nick translation following the protocol of Maniatis *et al.* (1982) and used for hybridisation at 40°C.

DNA sequence analysis

The M13mp8/mp9 system (Messing and Vieira, 1982) was used for sequence analysis by the dideoxy method (Sanger *et al.*, 1977) using ^{35}S-labelled α-thio-dATP and the buffer gradient system (Biggin *et al.*, 1983). The size of the glass plates was 20 × 50 cm, the Teflon spacers had a thickness of 0.25 mm. The gels (8% polyacrylamide) were run at a power of 55 W. The heptadecamer primers were used, and dITP routinely replaced dGTP (Mills and Kramer, 1979). It was found that under these conditions urea could be omitted from the sequencing gels.

The complete sequence of the 2.3-kb *Eco*RI restriction fragment (see above) was obtained from 32 randomly selected subclones containing *Sau*3A restriction fragments, 10 subclones containing pieces of sonicated DNA (prepared as described by Bankier and Barrel, 1983), and from the subclones which could be obtained making use of the *Pst*I, *Sal*I and *Bam*HI restriction sites. To confirm the sequences on the complementary strand, several mixed *Taq*I/*Sau*3A and *Taq*I/*Bam*HI restriction fragments were also cloned and sequenced.

Acknowledgements

We thank Drs D.Youvan, G.Drews and G.Klug for the generous gift of plasmids carrying the genes for the L and M subunit from *Rb. capsulatus*, Dr J.Tittor for help in preparing Figure 3 and Dr J.Deisenhofer for valuable discussions. Thanks are extended to Drs J.Farchaus and F.Jay for reading the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 143).

References

Allen,J. and Feher,G. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4795–4799.
Alt,J., Morris,J., Westhoff,P. and Herrmann,R.G. (1984) *Curr. Genet.*, **8**, 597–606.
Bankier,A.T. and Barrel,B.G. (1983) In Flavell,R.A. (ed.), *Techniques in Nucleic Acid Biochemistry*. Elsevier, County Clare, Ireland, pp. B508/1–B508/34.
Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.
Brunisholz,R.A., Wiemken,V., Suter,F., Bachofen,R. and Zuber,H. (1984) *Hoppe-Seyler's Z. Physiol. Chem.*, **365**, 689–701.
Brunisholz,R.A., Jay,F., Suter,F. and Zuber,H. (1985) *Biol. Chem. Hoppe-Seyler*, **366**, 87–98.
Büldt,G., Gally,H.U., Seelig,A. and Seelig,J. (1978) *Nature*, **271**, 182–184.
Chang,C.H., Schiffer,M., Tiede,D., Smith,U. and Norris,J. (1985) *J. Mol. Biol.*, **186**, 201–203.
Chory,J., Donohue,T.J., Varga,A.R., Staehelin,L.A. and Kaplan,S. (1984) *J. Bacteriol.*, **159**, 540–554.
Deisenhofer,J., Epp,O., Miki,K., Huber,R. and Michel,H. (1984) *J. Mol. Biol.*, **180**, 385–398.
Deisenhofer,J., Epp,O., Miki,K., Huber,R. and Michel,H. (1985) *Nature*, **318**, 618–624.
DeVitry,C. and Diner,B. (1984) *FEBS Lett.*, **167**, 327–331.
Engelman,D.M. and Steitz,T.A. (1981) *Cell*, **23**, 411–422.
Erickson,J.M., Rahire,M., Rochaix,J.D. and Mets,L. (1985) *Science*, **228**, 204–207.
Feher,G. and Okamura,M.Y. (1978) In Clayton,R.K. and Sistrom,W.R. (eds), *The Photosynthetic Bacteria*. Plenum Press, NY, pp. 349–386.
Garavito,R.M., Jenkins,J., Jansonius,J.N., Karlsson,R. and Rosenbusch,J.P. (1983) *J. Mol. Biol.*, **164**, 313–327.
Gibson,J., Stackebrandt,E., Zablen,L.B., Gupta,R. and Woese,C.R. (1979) *Curr. Microbiol.*, **3**, 59–64.
Holschuh,K., Bottomley,W. and Whitfield,P.R. (1984) *Nucleic Acids Res.*, **12**, 8819–8834.
Hoff,A.J. (1982) In Fong,F.K. (ed.), *Molecular Biology, Biochemistry, and Biophysics*. Springer, Berlin, Vol. **35**, pp. 80–151, 322–326.
Imhoff,J.F., Trüper,H.G. and Pfenning,N. (1984) *Int. J. Syst. Bacteriol.*, **34**, 340–343.
Kahn,M., Kolter,R., Thomas,Ch., Figurski,D., Meyer,R., Remaut,E. and Helinski,D.R. (1979) *Methods Enzymol.*, **68**, 268–279.
Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269–276.
Michel,H. (1982) *J. Mol. Biol.*, **158**, 567–572.
Michel,H. and Deisenhofer,J. (1986) In Staehelin,A.C. and Arntzen,C.J. (eds), *Encyclopedia of Plant Physiology: Photosynthesis III*. Springer, Berlin, Vol. **19**, pp. 371–381.
Michel,H., Weyer,K.A., Gruenberg,H. and Lottspeich,F. (1985) *EMBO J.*, **4**, 1667–1672.
Mills,D.R. and Kramer,F.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 2232–2235.
Ovchinnikov,Yu.A., Abdulaev,N.G., Vasilov,R.G., Vturina,I.Yu., Kuryatov, A.B. and Kiselev,A.V. (1985) *FEBS Lett.*, **179**, 343–350.
Paul,C. and Rosenbusch,J.P. (1985) *EMBO J.*, **4**, 1593–1597.
Parson,W.W. (1982) *Annu. Rev. Biophys. Bioeng.*, **11**, 57–80.
Pfister,K., Steinback,K.E., Gardner,G. and Arntzen,C.J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 981–985.

Rochaix,J.D., Dron,M., Rahire,M. and Malnoe,P. (1984) *Plant Mol. Biol.*, **3**, 363–370.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342–1346.
Thornber,J.P., Cogdell,R.J., Seftor,R.E.B. and Webster,G.D. (1980) *Biochim. Biophys. Acta*, **593**, 60–75.
Vogelsang,H., Oertel,W. and Oesterhelt,D. (1983) *Methods Enzymol.*, **97**, 226–241.
von Heijne,G. (1981a) *Eur. J. Biochem.*, **116**, 419–422.
von Heijne,G. (1981b) *Eur. J. Biochem.*, **120**, 275–278.
Waehneldt,T.V. (1971) *Anal. Biochem.*, **43**, 306–312.
Welte,W. and Kreutz,W. (1982) *Biochim. Biophys. Acta*, **692**, 479–488.
Wiemken,V. and Bachofen,R. (1984) *FEBS Lett.*, **166**, 155–159.
Williams,J.C., Steiner,L.A., Ogden,R.C., Simon,M.I. and Feher,G. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6505–6509.
Williams,J.C., Steiner,L.A., Feher,G. and Simon,M.I. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7303–7307.
Youvan,D.C., Bylina,E.J., Alberti,M., Begusch,H. and Hearst,J.E. (1984) *Cell*, **37**, 949–957.
Zurawski,G., Bohnert,H.J., Whitfield,P.R. and Bottomley,W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7699–7703.

Received on 21 February 1986; revised on 27 March 1986