PROTEINS: Structure, Function, and Genetics 1:312–325 (1986)

Primary Structure of the Reaction Center From *Rhodopseudomonas sphaeroides*

J. C. Williams,[1,2*] L. A. Steiner,[3] and G. Feher[2]
Departments of [1]Biology and [2]Physics, University of California, San Diego, La Jolla, California 92093;
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

ABSTRACT **The reaction center is a pigment-protein complex that mediates the initial photochemical steps of photosynthesis. The amino-terminal sequences of the L, M, and H subunits and the nucleotide and derived amino acid sequences of the L and M structural genes from *Rhodopseudomonas sphaeroides* have previously been determined. We report here the sequence of the H subunit, completing the primary structure determination of the reaction center from *R. sphaeroides*. The nucleotide sequence of the gene encoding the H subunit was determined by the dideoxy method after subcloning fragments into single-stranded M13 phage vectors. This information was used to derive the amino acid sequence of the corresponding polypeptide. The termini of the primary structure of the H subunit were established by means of the amino and carboxy terminal sequences of the polypeptide. The data showed that the H subunit is composed of 260 residues, corresponding to a molecular weight of 28,003. A molecular weight of 100,858 for the reaction center was calculated from the primary structures of the subunits and the cofactors. Examination of the genes encoding the reaction center shows that the codon usage is strongly biased towards codons ending in G and C. Hydropathy analysis of the H subunit sequence reveals one stretch of hydrophobic residues near the amino terminus; the L and M subunits contain five such stretches. From a comparison of the sequences of homologous proteins found in bacterial reaction centers and photosystem II of plants, an evolutionary tree was contructed. The analysis of evolutionary relationships showed that the L and M subunits of reaction centers and the D1 and D2 proteins of photosystem II are descended from a common ancestor, and that the rate of change in these proteins was much higher in the first billion years after the divergence of the reaction center and photosystem II than in the subsequent billion years represented by the divergence of the species containing these proteins.**

Key words: membrane proteins, nucleotide sequence, evolution, photosynthesis

INTRODUCTION

The conversion of light into chemical energy by cells occurs in a pigment-protein complex called the reaction center (RC). RCs from the bacterium *Rhodopseudomonas sphaeroides* (recently renamed *Rhodobacter sphaeroides*) have been extensively characterized (reviewed in ref. 1). They are isolated in detergent from the cytoplasmic membrane and are composed of three subunits (L, M, and H), four bacteriochlorophylls, two bacteriopheophytins, two ubiquinones, and one iron. The roles of the cofactors in the primary photochemistry and subsequent electron transfer steps have been elucidated in part, but until recently, complementary information on the structure was lacking. In particular, the primary structures of the polypeptides proved difficult to obtain by protein chemical methods, owing to the intractability of the hydrophobic peptides[2]. This problem was circumvented by utilizing recombinant DNA techniques to determine the nucleotide sequences of the structural genes encoding the subunits.

Progress in the crystallization of membrane proteins made the determination of the three dimensional structure of RCs by x-ray diffraction possible. The structure of the RC from *R. viridis* has been determined to a resolution of ~3 Å[3] and of the RC from *R. sphaeroides* to a resolution of ~3.3 Å.[4] The major structural features of the RCs are conserved between the two species. Advances in molecular genetics and crystallography have greatly increased the knowledge of the RC structure and provided the basis for detailed structure-function analyses. In addition, the application of gene sequencing has expanded the number of available sequences of proteins involved in photosynthesis in both bacteria and plants, allowing comparisons of homologous sequences.

The primary structures of the L and M subunits of the RC from *R. sphaeroides* have been determined.[5,6] We report here the primary structure of the H subunit. This gene was isolated on a BamH1 fragment by Donahue et al.[7] We determined the nucleotide sequence of this gene by the dideoxy method[8] and derived the amino acid sequence of the corresponding polypeptide. To corrobate the sequence obtained from the DNA, the sequences of several peptides and of the carboxy terminus were determined. The complete se-

Received October 9, 1986; accepted November 25, 1986.

Address reprint requests to G. Feher, Department of Physics, B-019, University of California, San Diego, La Jolla, CA 92093.

*Work performed in partial fulfillment of the requirement for the Ph.D. degree. Present address: Medical Research Council, Hills Road, Cambridge CB2 2QH, England.

quences of the three subunits of the bacterial RC have been determined for two other species, *R. capsulata*[9] and *R. viridis*[10,11] These sequences were compared to establish the conserved features.

Photosystem II (PSII) of green plants, algae, and cyanobacteria shares many characteristics with RCs from purple bacteria (for review, see ref. 12). Five proteins are generally found in PSII preparations (for review, see ref. 13). One of these,the D1 protein, also called the Q_B or 32-kilodalton (kD) protein, has been studied extensively (for review, see ref. 14). The gene encoding this protein has been sequenced in a number of species. The D1 sequences from plants are highly conserved, differing by only a few amino acid residues. The D2 or 34-kD polypeptide of PSII is less well characterized, but its gene has been sequenced in several organisms.[15–18] The D1 and D2 sequences have been found to be homologous to each other as well as to the L and M subunits of the RC. A quantitative analysis of the evolutionary relationships between PSII and bacterial RC proteins is presented.

MATERIALS AND METHODS

Materials

Sources for enzymes and vectors have been described previously.[5,6] A plasmid (pRHB2) containing a 1.45-kilobase (kb) BamH1 fragment carrying the H subunit gene was obtained from T. Donahue.[7] The fragment was isolated from *R. sphaeroides* 2.4.1. The following mixture of 16 oligonucleotides each 20 nucleotides long was synthesized by the phosphodiamidite method on an Applied Biosystems 380A DNA synthesizer:

$$\mathrm{TTCGG}\begin{matrix}\mathrm{A}\\\mathrm{G}\\\mathrm{T}\\\mathrm{C}\end{matrix}\mathrm{AACTTCGA}\begin{matrix}\mathrm{T}\\\mathrm{C}\end{matrix}\mathrm{CT}\begin{matrix}\mathrm{G}\\\mathrm{C}\end{matrix}\mathrm{GC} \qquad (1)$$

The sequences of these oligonucleotides are based on the codons for residues 7–13 (Phe-Gly-Asn-Phe-Asp-Leu-Ala) of the previously determined amino-terminal sequence of the H subunit.[2] The codon usage in the L and M subunit genes[5,6] was used to eliminate some possible codons for phenylalanine, asparagine, and leucine (see table I). The mixture of oligonucleotides was originally used to screen clones for those containing the H subunit gene, and was later used as a primer for M13 subclones in sequencing.

Preparation and Sequence Analysis of DNA

The M13 subclones used as templates for dideoxy sequencing reactions were derived from the plasmid pRHB2.[7] M13 subclones of the Sph I-Xho I fragments of pRHB2 were obtained from T. Donahue. Bgl II-BamH1 and Sal I-BamH1 fragments were subcloned in M13 mp18.[19] DNA sequencing was performed as described previously, with the following modifications. The mixed-sequence oligonucleotide described above was used as a primer for one of the M13 subclones. Formamide (20% v/v) was added to the gel mixture to minimize the effects of secondary structure in the DNA. The gels were wedge-shaped, with the thickness increasing from 0.1 mm at the top to 0.3 mm at the bottom.

Preparation and Sequence Analysis of Peptides

The H subunit was isolated as described[2] and digested with trypsin.[5,6] The digest was lyophilized and the peptides that were soluble in 0.1% trifluoroacetic acid (~40% of the total material) were fractionated by reversed-phase high pressure liquid chromatography.[5] Selected fractions were subjected to amino acid analysis and to sequence analysis by either the manual dansyl-Edman procedure or by automated Edman degradation.[20] In the dansyl-Edman procedure, glutamic acid and aspartic acid are not distinguished from the amides.

Digestion With Carboxypeptidase A

The H subunit was dialyzed against H_2O that had been adjusted to pH ~9 with NH_4OH; aliquots of ~90 μg were digested with 9 μg of carboxypeptidase A in 0.05M NH_4HCO_3 for 30 min or for 24 hr at room temperature. Substrate and enzyme blanks were subjected to similar treatment. The digests were lyophilized and applied directly to the amino acid analyzer.

Alignment of Homologous Sequences

The NEWAT protein sequence data base[21] was searched for sequences homologous to each subunit. Homologous sequences were aligned to show the maximum amount of similarity, by means of the computer programs described by Feng et al.,[22] which are based on the method of Needleman and Wunsch.[23] Different pairs of sequences were compared by the percentage identity, defined as the number of identical residues observed after alignment of two sequences divided by the number of residues in the smaller sequence. Sequences were also compared by the scoring system of Dayhoff et al.,[24] modified so that all values are positive (see Table 1 in Feng et al.)[22] A penalty score of 1.0 was assigned per gap. The scoring system of Dayhoff rates the probability of any amino acid exchange, and is based on the observed changes in a large group of closely related sequences. The scores obtained by the Dayhoff method are referred to as similarity scores.

Multiple alignments were used for comparing more than two sequences, and were created by the stepwise addition of sequences to an initial pair defined as the two most closely related sequences as determined by binary alignments of all sequences under consideration.[25] These two sequences were aligned by the insertion of gaps to maximize their homology. Of the remaining sequences, that closest to either member of the original pair was added to form a new group. The sequence added was aligned with the sequence in the group to which it was the most similar. Any gaps added to a sequence in the group during the alignment with a newly added sequence were also added at the same position in all members of that

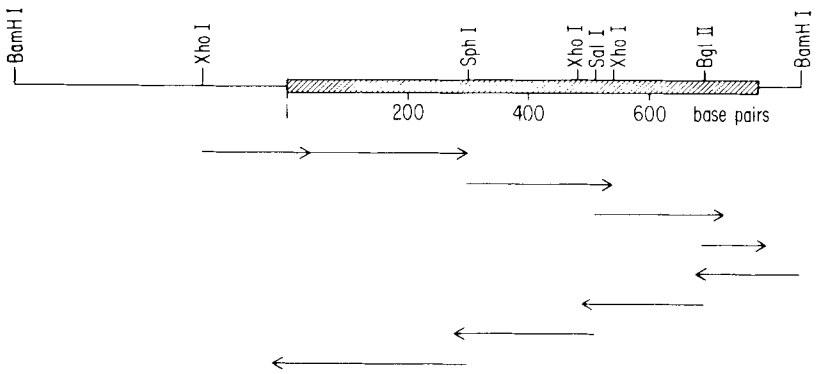

Fig. 1. Sequencing strategy for the H subunit gene. Only restriction sites used for subcloning are shown. Arrows indicate the extent and direction of each sequence determination. Box indicates the position of the H subunit gene.

group. If another pair of sequences was more similar to each other than either was to a member of the initial group, then that aligned pair defined a separate group. Two groups were joined by aligning the two most similar sequences from each group.

Construction of a Phylogenetic Tree

The evolutionary distance between a number of sequences can be represented graphically by the branches of a phylogenetic tree.[26] The branch lengths are a measure of the number of mutations that have taken place. It is assumed that the mutation rate is constant over time. The number of mutations is greater than the observed number of sequence changes in distantly related sequences due to multiple mutations at the same site; the two being related exponentially. Furthermore, any two sequences will exhibit a background random similarity because there are only 20 different amino acids. A score that accounts for the exponential relation between the number of mutations and the observed number of sequence changes, as well as the background similarity, has been defined as follows:[22]

$$D = -\ln\left(\frac{S_{real} - S_{rand}}{S_{iden} - S_{rand}}\right) \times 100 \qquad (2)$$

where D is the difference score and is a measure of the evolutionary distance between two aligned sequences, S_{real} is the similarity score of the sequences in the multiple alignment, S_{iden} is the average similarity score of the sequences aligned with themselves and is used to normalize the scores, and S_{rand} represents the background similarity. The S_{rand} score was approximated by a score of 60 per 100 residues; this number was arrived at empirically from the computation of a large number of trees.[25]

For a given tree, there is a unique path along branches between one sequence and every other sequence. The difference score between two sequences is equated to the sum of the branch lengths connecting those two sequences. If the number of sequences is larger than three, the number of difference scores (equal to the number of equations generated) is greater than the number of branches whose length is to be determined (equal to the number of variables). Consequently, the problem is overdetermined and there is no unique solution for the branch lengths. The optimum solution was obtained by a linear least-squares fit. The branch lengths determined by the least-squares fit were used to calculate a new set of difference scores. The percent standard deviation was determined by comparing the calculated difference scores to the true difference scores.

RESULTS

Nucleotide Sequence

The nucleotide sequence of the gene encoding the H subunit was determined from eight recombinant phage (Fig. 1). Sequence data were obtained on both strands over the entire length of the gene. Amino acid sequence data confirmed 20% of the H subunit sequence. The nucleotide and derived amino acid sequences are shown in Figure 2. The codons used in the L, M, and H subunit genes in *R. sphaeroides* are compiled in Table I.

Sequences of Tryptic Peptides

Four peptides were isolated in 30–40% yield. The seven amino-terminal residues of one peptide were determined by automated Edman degradation and these were found to correspond to positions 147 to 153, as determined by nucleotide sequencing. The sequences of the three other peptides were determined by the manual dansyl-Edman method; the carboxy-terminal residue in each of these peptides was deduced from the amino acid composition. These pep-

TABLE I. Codon Usage in L, M, and H Subunit Genes

Amino acid	Codon	L and M subunit genes					H subunit genes				
		*R.s.**	*R.c.*†	*R.v.*‡	Total	%§	*R.s.**	*R.c.*†	*R.v.*‡	Total	%§
Arg	CGC	14	9	13	36	59	9	9	9	27	61
	CGU	3	6	4	13	21	1	1	9	11	25
	CGG	2	3	4	9	15	1	1	4	6	14
	CGA	1	1	0	2	3	0	0	0	0	0
	AGA	0	0	1	1	2	0	0	0	0	0
	AGG	0	0	0	0	0	0	0	0	0	0
Leu	CUG	40	24	32	96	51	13	13	18	44	56
	CUC	29	23	17	69	37	11	11	19	32	41
	CUU	2	7	8	17	9	0	0	1	1	1
	UUG	0	4	2	6	3	0	0	1	1	1
	UUA	0	0	0	0	0	0	0	1	1	1
	CUA	0	0	0	0	0	0	0	0	0	0
Ser	UCG	17	18	17	52	57	8	8	7	23	66
	AGC	4	5	8	17	18	2	2	2	6	17
	UCC	7	2	4	13	14	2	2	1	5	14
	AGU	1	2	1	4	4	0	0	0	1	3
	UCU	1	0	4	5	5	0	0	0	0	0
	UCA	0	0	0	0	0	0	0	0	0	0
Thr	ACC	13	22	8	43	56	6	6	7	19	48
	ACG	15	11	8	26	34	5	5	8	18	45
	ACU	1	0	6	7	9	1	1	0	2	5
	ACA	0	0	1	1	1	0	0	1	1	3
Pro	CCG	16	20	25	61	70	16	16	16	48	75
	CCC	8	6	6	20	23	7	7	2	16	25
	CCU	2	1	3	6	7	0	0	0	0	0
	CCA	0	0	0	0	0	0	0	0	0	0
Ala	GCC	30	30	14	74	43	11	11	10	32	40
	GCG	20	18	22	60	34	14	14	9	37	46
	GCU	4	9	14	27	6	2	2	6	10	13
	GCA	5	2	6	13	7	0	0	1	1	1
Gly	GGC	43	40	43	126	61	22	22	16	60	87
	GGU	9	10	14	33	16	1	1	3	5	7
	GGG	12	12	9	33	16	1	1	0	2	3
	GGA	4	3	7	14	7	1	1	0	2	3
Val	GUC	22	21	16	59	56	13	13	15	41	59
	GUG	9	13	8	30	28	5	5	7	17	25
	GUU	2	5	9	16	15	2	2	6	10	14
	GUA	1	0	0	1	1	0	0	1	1	1
Lys	AAG	6	5	9	20	77	14	14	9	37	100
	AAA	2	3	1	6	23	0	0	0	0	0
Asn	AAC	21	17	13	51	91	8	8	1	17	94
	AAU	1	2	2	5	9	0	0	1	1	6
Gln	CAG	11	12	11	34	85	5	5	7	17	89
	CAA	2	2	2	6	15	0	0	2	2	11
His	CAC	14	14	11	39	85	4	4	3	11	69
	CAU	0	0	7	7	15	2	2	1	5	31
Glu	GAA	9	12	5	26	54	7	7	10	24	47
	GAG	9	4	9	22	46	9	9	9	27	53
Asp	GAC	9	15	11	35	67	11	11	16	38	88
	GAU	4	3	10	17	33	2	2	1	5	12
Tyr	UCA	13	9	13	34	50	4	4	8	16	62
	UAU	8	14	12	34	50	3	3	4	10	38
Cys	UGC	3	5	6	14	88	2	2	1	5	100
	UGU	0	0	2	2	12	0	0	0	0	0
Phe	UUC	49	43	46	138	93	9	9	6	24	92
	UUU	1	4	5	10	7	1	1	0	2	8
Ile	AUC	34	33	32	99	85	13	13	8	34	100
	AUU	3	3	12	18	15	0	0	0	0	0
	AUA	0	0	0	0	0	0	0	0	0	0
Met	AUG	16	27	10	53	100	9	9	1	19	100
Trp	UGG	36	35	33	104	100	3	3	6	12	100

**R. sphaeroides* [5,6] and this work.
†*R. capsulata*.[9]
‡*R. viridis*.[10,11]
§Calculated as the % of the total codons for each amino acid.

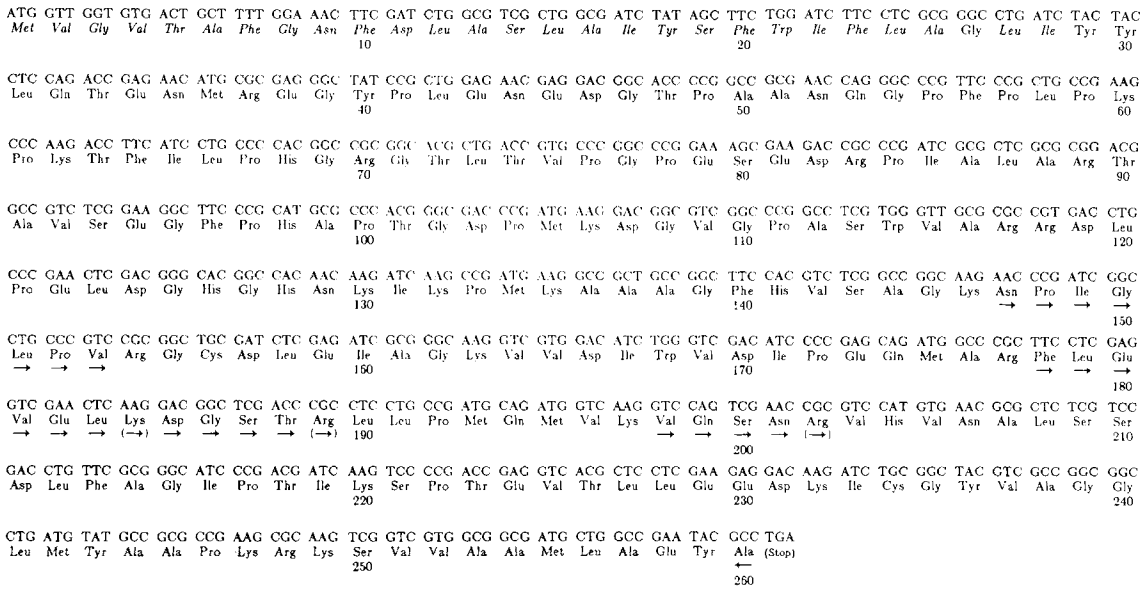

Fig. 2. Nucleotide sequence of the H subunit gene. Residues that confirm previously determined amino-terminal sequences[2] are in italics. Residues identified by Edman degradation of isolated tryptic peptides are indicated by →; residues deduced from the amino acid compositon of isolated tryptic peptides are indicated by (→); the residue identified by carboxypeptidase A digestion is marked by ←.

tides correspond to positions 178 to 184, 185 to 189, and 198 to 202, as determined by nucleotide sequencing.

Carboxy-terminal Analysis

After 30 min of digestion of H with carboxypeptidase A, alanine was the amino acid released in highest (~60%) yield; after 24 hr of digestion, the yield of alanine was ~80%. From these results we conclude that the carboxy-terminal residue is alanine.

Homologies

The L, M, and H sequences of *R. sphaeroides, R. capsulata*, and *R. viridis* were compared. An alignment of the three H sequences is shown in Fig. 3. Alignments of the L and M sequences have been made previously.[11] The percent identity between the same subunits in the different species is summarized in Table II.

Three D1 sequences (*Spinacia oleracia, Chlamydomonas reinhardtii*, and *Anacystis nidulans*); two D2 sequences (*Spinacia oleracia* and *Chlamydomonas reinhardtii*); and the L and M sequences from *R. sphaeroides, R. capsulata*, and *R. viridis* were used to determine the evolutionary relatedness among these proteins. A multiple alignment of these eleven sequences is shown in Figure 4. The extreme amino- and carboxy-terminal sequences of the proteins were judged to be insufficiently similar for a multiple alignment and were not included. The difference scores calculated from this alignment are shown in Table III.

TABLE II. Percent Identities* Among RC Subunits†

	R. sphaeroides R. capsulata	*R. sphaeroides R. viridis*	*R. capsulata R. viridis*
L:L	78	59	59
M:M	76	50	50
H:H	64	39	38

*The percent is the number of identical residues divided by the number of residues in the smaller sequence × 100.
†Gaps were introduced to optimize the similarities between the sequences (see Fig. 3).

DISCUSSION

Molecular Weights

The data show that the H subunit is composed of 260 residues, corresponding to a molecular weight of 28,003. The molecular weights of the other subunits and of the entire RC complex are shown in Table IV. The molecular weights of the subunits had previously been estimated using the technique of $NaDodSO_4$ polyacrylamide gel electrophoresis. The estimated molecular weights of the L and M subunits are ~30% less than the calculated values, while the estimate for the H subunit agrees very well. The estimated molecular weights were based upon comparison with the migration of water-soluble proteins. The H subunit is much less hydrophobic than the L and M subunits, and appears to be comparable to the proteins used as standards, leading to a much better estimated value. An alternative method of estimating molecular weights is based on the relative migration in $NaDodSO_4$ gels of different concentrations of acry-

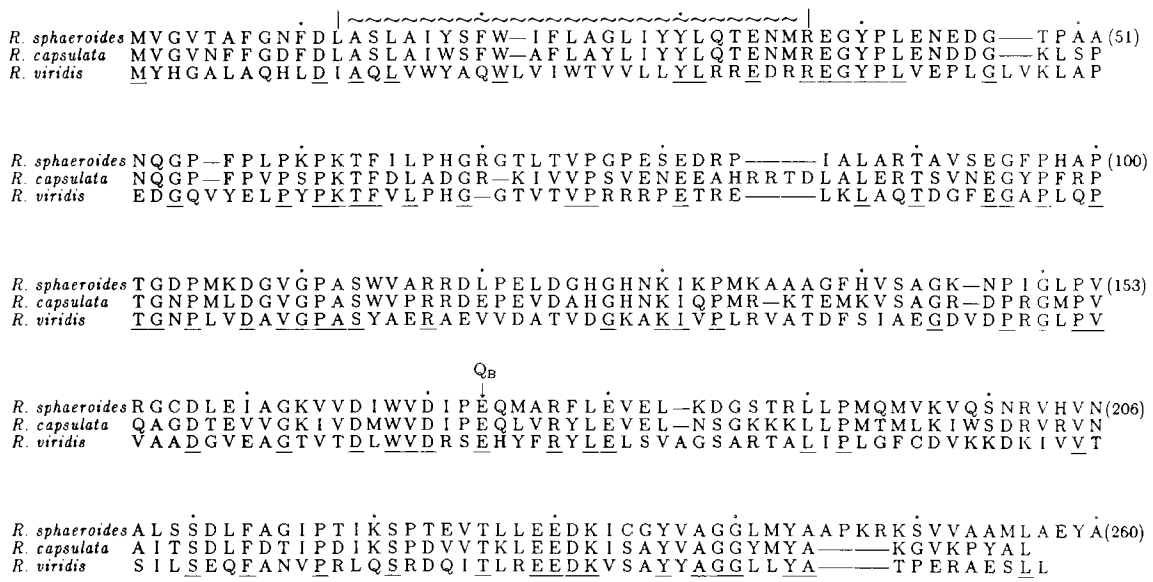

Fig. 3. Comparison of sequences of H subunits from *R. sphaeroides*, *R. capsulata*,[9] and *R. viridis*.[10] Gaps were introduced to maximize the homology. Residues identical in all three sequences are underlined. Numbers indicate the sequence number of the *R. sphaeroides* sequence; every tenth residue is marked. Positions of a putative transmembrane helix (| ~ |) and a conserved glutamic acid (↓) located near the secondary quinone (Q_B) are indicated.[3]

lamide.[30] This technique appears to be much less sensitive to hydrophobicity of the protein.[31] The molecular weights of the RC subunits determined by this technique were within 10% of those calculated from the sequences for all three subunits (Gardlund and Steiner, unpublished results, cited in ref. 32). The minimum molecular weights had also been calculated based on the experimentally determined amino acid compositions of the RC subunits.[32] In this calculation the molecular weight that best gives integral values for the numbers of amino acids is determined. This determination is very sensitive to errors in the experimentally determined amino acid composition (e.g., if a subunit contains 25 residues of a given amino acid, the error with which this residue has to be determined has to be smaller than 2%). Consequently, the method failed to yield the correct molecular weights.

Stoichiometry

The stoichiometry of the RC subunits was estimated in a previous study to be 1:1:1. The stoichiometry was evaluated by two methods: comparison of

TABLE III. Difference Scores* Among RC and PSII Proteins

			1	2	3	4	5	6	7	8	9	10	11
1	PSII D2	*Chlamydomonas*	0	2	78	77	79	147	143	146	156	156	170
2		*Spinacia*	2	0	80	78	80	149	146	145	158	156	168
3	PSII D1	*Chlamydomonas*	78	80	0	4	6	144	140	146	157	153	161
4		*Spinacia*	77	78	4	0	6	144	141	148	154	153	159
5		*Anacystis*	79	80	6	6	0	147	143	150	156	155	160
6	RC L	*R. sphaeroides*	147	149	144	144	147	0	12	30	81	85	82
7		*R. capsulata*	143	146	140	141	143	12	0	35	84	88	84
8		*R. viridis*	146	145	146	148	150	30	35	0	78	83	85
9	RC M	*R. sphaeroides*	156	158	157	154	156	81	84	78	0	13	44
10		*R. capsulata*	156	156	153	153	155	85	88	83	13	0	41
11		*R. viridis*	170	168	161	159	160	82	84	85	44	41	0

*Calculated for the multiple alignment in Figure 4 by the method of Feng et al[22] rounded to the nearest unit. Difference score is a measure of evolutionary distance and is approximately equal to the actual number of mutations separating two sequences, calculated from the observed number of changes corrected for the superposition of mutations and random background similarities (see Eq. 2).

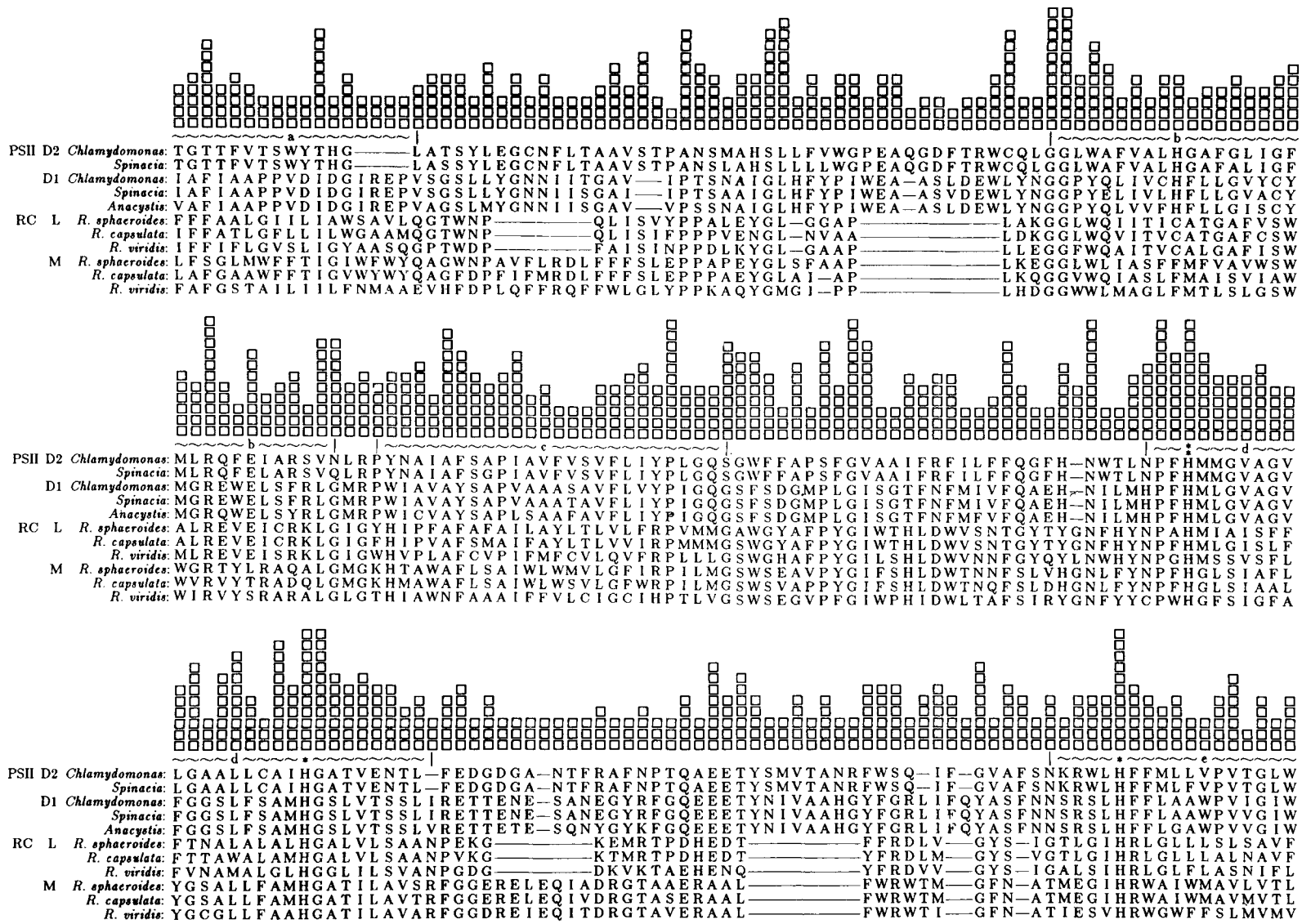

Fig. 4. Comparison of sequences of RC and PSII proteins. Gaps were introduced to maximize the homology. The amino- and carboxy-terminal segments of the proteins have been deleted. Shown are residues 50–280 (68% of the sequence) of the D2 protein from *Chlamydomonas reinhardtii*,[18] residues 51–281 (65% of the sequence) of the D2 protein from *Spinacia oleracia*,[16] residues 50–284 (67% of the sequence) of the D1 protein from *Chlamydomonas reinhardtii*,[27] residues 50–284 (67% of the sequence) of the D1 protein from *Spinacia oleracia*,[28] residues 50–284 (65% of the sequence) of the D1 protein from *Anacystis nidulans*,[29] residues 39–242 (73% of the sequence) of the L subunit from *R. sphaeroides*,[6] residues 40–243 (72% of the sequence) of the L subunit from *R. capsulata*,[9] residues 39–242 (75% of the sequence) of the L subunit form *R. viridis*,[11] residues 60–278 (71% of the sequence) of the M subunit from *R. sphaeroides*,[5] residues 59–276 (71% of the sequence) of the M subunit from *R. capsulata*,[9] and residues 59–276 (67% of the sequence) of the M subunit from *R. viridis*.[11] The boxes indicate the largest number of identical residues at each position. Positions of putative transmembrane helices a–e (| ~ |) and ligands to the donor bacteriorchlorophyll (:) and iron (*) in the L and M subunits are indicated.[3]

TABLE IV. Subunit Compositions

Species	No. of residues			Molecular weight				
	L	M	H	L	M	H	Total	+Cofactors
R. sphaeroides	281	307	260	31,319	34,388	28,003	93,660	100,858*
R. capsulata†	282	306	254	31,565	34,440	25,534	91,539	98,737*
R. viridis‡	273	323	258	30,571	35,902	28,345	94,818	101,102§

*Cofactors: 4 bacteriochlorophyll a (911), 2 bacteriopheophytin a (886), 2 ubiquinone-10 (863), 1 iron (56).
†Youvan et al.[9]
‡Michel et al.[10,11]
§Cofactors: 4 bacteriochlorophyll b (909), 2 bacteriopheophytin b (884), 1 menaquinone-9 (785), 1 ubiquinone-9 (795), 1 iron (56).

radiolabeled LM complex to H after separation in a sucrose gradient,[33] and comparison of the amino acids compositions of the individual subunits and the entire RC.[34] Both of these methods relied on molecular weight determinations, some of which (for the L and M subunits) we now know to be incorrect. Recalculation with the molecular weights determined from the sequence yields the same answer for the latter but not for the former method. Apparently, the amount of the H subunit was overrepresented in the sample, or LM was underestimated. This may be due to an insufficient correction for overlapping of the LM and H bands in the sucrose gradient fractions (see Fig. 6 in Okamura et al.[33]) The number of subunits observed in the structure of RCs determined by x-ray diffraction of crystals has confirmed the 1:1:1 stoichiometry[3,4].

Extinction Coefficient

The molecular weights can also be used to determine the extinction coefficient of the RC if both the amount of protein and the optical density of a sample are determined. The amount of protein in a one ml sample of RCs with an optical absorbance $A_{802}^{1cm} = 1.00$ (ie. $ODV_{802} = 1.00$) as determined from the molar amounts of the individual amino acids, was reported to be 0.319 mg by Steiner et al.[34] The precision of this determination was approximately 1%. The average difference in the mole percent of each amino acid determined experimentally and from the sequence is 4%, when weighted by the contribution of each amino acid relative to the total protein. This reflects the error in the accuracy of the determination of the amino acid composition; we assume that a similar error is involved in determining the amount of protein. Calculation of the ϵ_{802} using the experimental data and the molecular weights of the subunits determined from the sequences yields a value of $2.94 \pm 0.12 \times 10^5$ $M^{-1}cm^{-1}$. This is in agreement with the value calculated using the ratio of the pigments in the RC, which is independent of the molecular weight [$\epsilon_{802} = 2.88 \pm 0.14 \times 10^5$ $M^{-1}cm^{-1}$].[35]

Gene Location

The structural gene encoding the H subunit from *R. sphaeroides* is contained on a 1.45-kb BamH1 fragment.[7] In *R. sphaeroides*, the genes encoding the L and M subunits are adjacent, and the carboxy terminus of the L subunit gene overlaps by 8 bp the amino terminus of the M subunit gene. The genes encoding the α and β subunits of the B870 antenna complex, which is closely associated with the RC, are near the L and M subunit genes on a 12.5-kb BamH1 fragment.[5,6] It is not known how far away the fragment containing the L and M subunits and the fragment containing the H subunit are on the chromosome of *R. sphaeroides*. In *R. capsulata*, a number of genes involved in the synthesis of the photosynthetic apparatus are clustered.[36] These include the genes encoding the L, M, and H subunits of the RC as well as genes encoding proteins involved in bacteriochlorophyll and carotenoid biosynthesis. The L, M, and B870 antenna genes have the same arrangment in both species, and in *R. capsulata*, the gene encoding the H subunit is ~40 kb away from the genes encoding the L and M subunits.[9] It is likely that a similar arrangement of genes associated with photosynthesis occurs in *R. sphaeroides*.

Codon Usage

Examination of the codon usage in the L, M, and H subunit genes (Table I) shows that it is distinctly nonrandom. The most prominent feature is the preferential usage of G or C in the third position of the codon (89% G or C in the third position, 46% in the second, and 59% in the first). The L, M, and H subunit genes contain an overall GC content of 62%, 66%, and 67%, respectively, which is in accord with the high GC content in the total DNA (69% in the total DNA, Mandel et al.[37]). The relation between the GC content in the third position of the codons and in the total DNA has been observed in other organisms.[38] A similar high GC content is found in the L, M, and H genes of *R. capsulata*[9] and *R. viridis*.[10,11]

In *R. sphaeroides*, as in *R. capsulata* and *R.viridis*, the H gene contains a higher proportion of codons ending in G and C than do the L and M genes (Table I). It has been suggested that difference in codon usage among the genes encoding the subunits of the RC may be a regulatory phenomenon.[11] For example, the infrequently used codons (those ending in A and

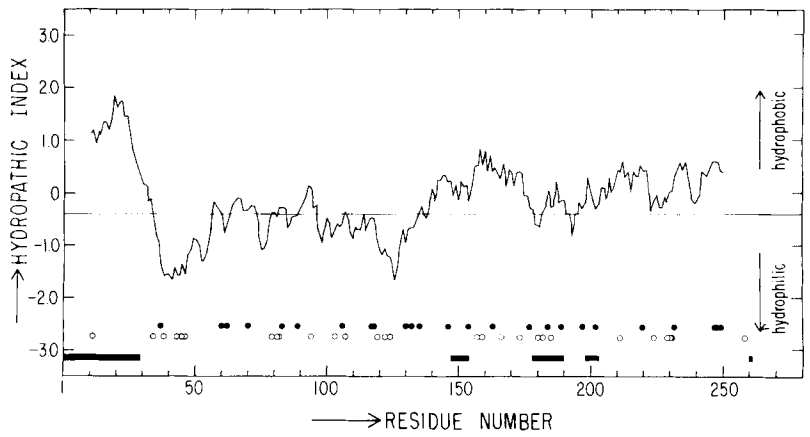

Fig. 5. Hydropathy profile of the H subunit. The average hydropathy value[46] of a moving window of 19 residues is plotted at the midpoint of the window. Positions of basic (Arg, Lys) and acidic (Glu, Asp) residues are marked by ● and ○, respectively; segments determined by sequence analysis of the protein are marked by ■.

T) may be read by less abundant tRNAs. In that case the near absence of these codons in the H subunit gene helps assure that the H subunit will be synthesized even after synthesis of the L and M subunits has ceased because the pool of minor tRNAs has been depleted.

Membrane Orientation

The RC is an integral membrane protein and has been shown by proteolysis, iodination and antibody labeling techniques to span the plasma membrane.[39–43] A thermodynamically favorable structure for spanning a lipid bilayer membrane is an α helix, since the hydrogen bonding requirements of the backbone atoms will be met internally rather than by water. The hydrocarbon portion of a membrane is ~30 Å wide for phospholipids with alkyl chains in the C_{16} to C_{18} range.[44] The major fatty acid in membranes from *R. sphaeroides* is an octadecenoic acid.[45] It is expected that the residues in the membrane-spanning portion of a protein be hydrophobic. The hydropathy scale of Kyte and Doolittle[46] was used to generate a profile by averaging the hydropathy value over a window of 19 residues over the entire sequence (Fig. 5). Clusters of hydrophobic residues appear as peaks in this analysis. A peak in a window of 19 that has an average hydropathic score of 1.6 or more is indicative of a potential membrane-spanning α helix. There is one such peak at the amino-terminal end of the H subunit; five peaks in both the L and the M subunit sequences were found previously.[5,6] In each case, there are no charged residues in a stretch of hydrophobic amino acids at least 19 residues long. From the x-ray diffraction studies of RCs from *R. sphaeroides* and *R. viridis*, these stretches of hydrophobic residues have been shown to correspond to α helices, and are thought to be the membrane spanning portions of the RC.[4,47] Consistent with this analysis are the results of circular dichroism and polarized infrared spectroscopy. These experiments indicate that the RC contains a large amount of α-helical structure that is approximately perpendicular to the plane of the membrane and that the H subunit contains less helical structure than do the L and M subunits.[48]

From the positions of the putative membrane-spanning helices, one can predict the topology of the RC in the membrane if the polypeptides can be oriented with respect to the cytoplasmic and periplasmic sides of the membrane. Labeling experiments indicated that the bulk of the H subunit lies on the cytoplasmic side.[39–41,43] Proteolytic digestion of chromatophores from *R. rubrum* showed that the amino terminus of the L subunit and an internal segment (residues 246–249) of the M subunit are on the cytoplasmic side of the membrane.[49] The position of the internal segment of the M subunit is consistent with placing the amino terminus of the M subunit also on the cytoplasmic side.

The pathways of the backbones of the three subunits, as determined from topological and structural studies, are represented in Figure 6. The charged residues in the L and M sequences are found in the putative interhelical stretches outside of the core of the membrane. The distribution of charges is similar to that observed in *R. viridis* and *R. capsulata*. More charged residues are found on the cytoplasmic side of the membrane than on the periplasmic side. In both the L and M subunits, a net negative charge exists on the periplasmic side and a slight net positive charge exists on the cytoplasmic side. This is of the opposite polarity of the potential of the membrane created by the extrusion of protons during electron

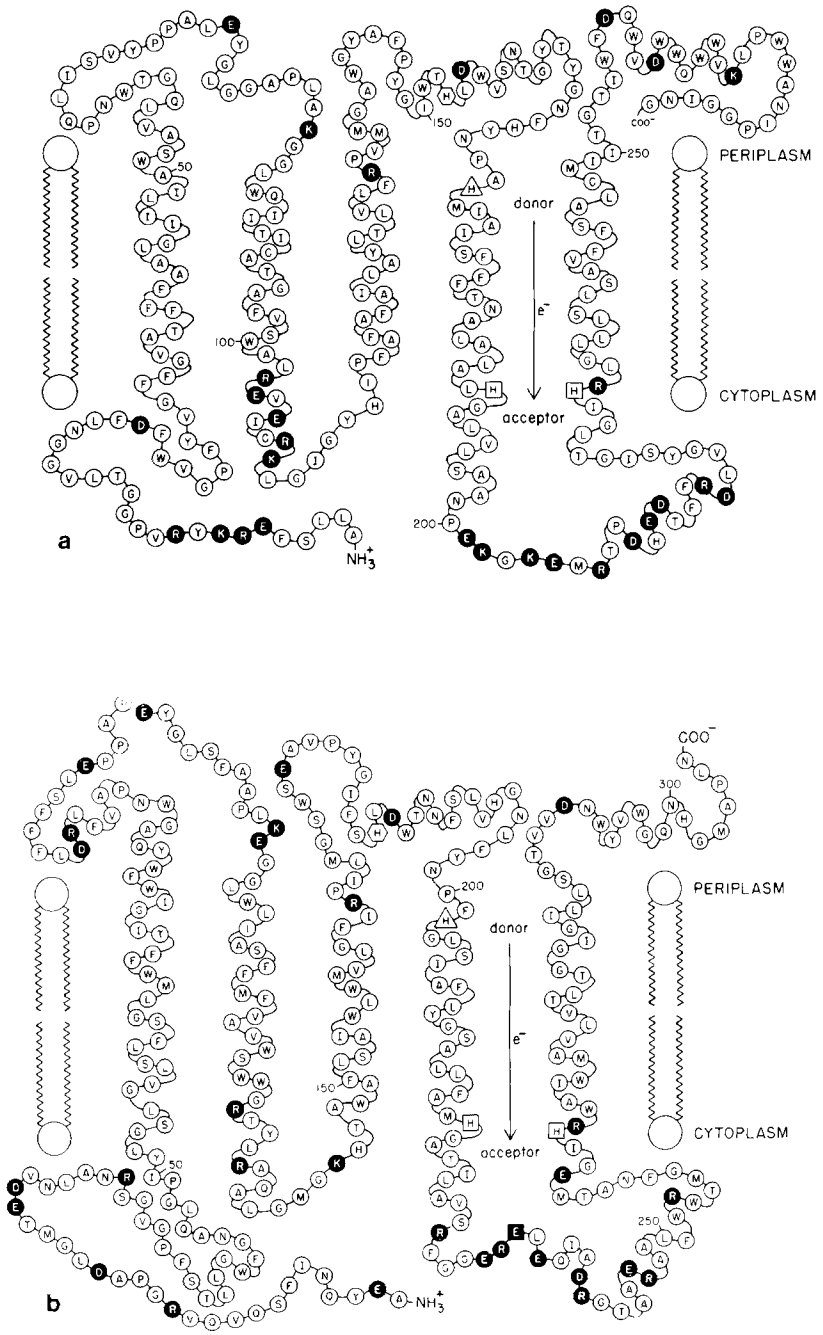

Fig. 6. Continued on the following page.

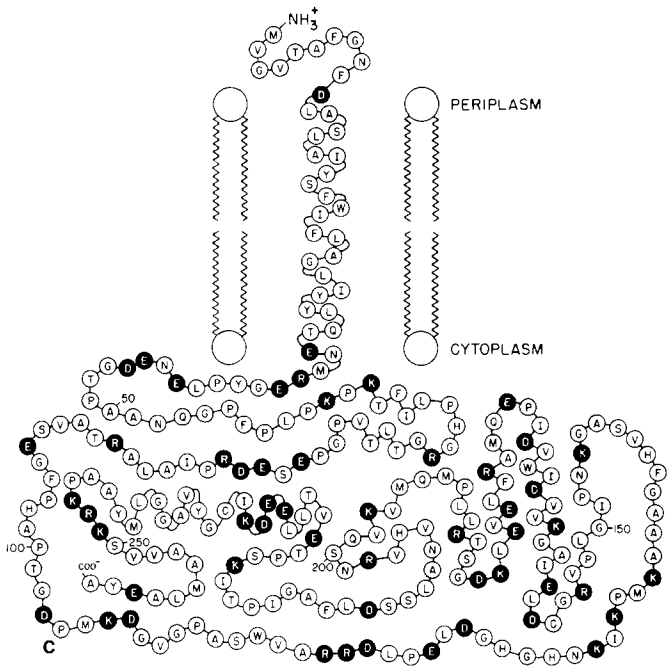

Fig. 6. Schematic representation of the orientation with respect to the membrane of the (**a**) L, (**b**) M, and (**c**) H subunits. Charged residues (Glu, Asp, Lys, Arg) are shown as filled circles. Ligands to the accessory bacteriochlorophyll (hexagon), donor bacteriochlorophyll (triangle), and iron (square) are marked.[3]

transport. As discussed by Michel et al.,[11] this charge distribution may facilitate the orientation of the protein in the membrane during assembly of the RC, and may also be involved in electron transfer from the periplasmic donor to the cytoplasmic acceptor. The negative charges on the periplasmic side of the L and M subunits include residues containing carboxyl groups involved in binding cytochrome c_2, the secondary electron donor to the RC.[50]

Sequence Comparisons

For each of the three subunits, the subunits of *R. sphaeroides* and *R. capsulata* are more closely related to each other than they are to *R. viridis* (Table II). This is consistent with other taxonomic criteria such as 16S ribosomal RNA and cytochrome sequences.[51] Among the conserved residues in each subunit are those that participate in binding ligands and those at structurally strategic positions such as turns and ends of helices.[11] The H subunit sequences are less well conserved than the L and M sequences, indicating that there are less constraints on its structure. This is probably because it does not bind cofactors and consequently plays a less critical role in the functioning of the RC.

Purple bacteria perform photosynthesis only under anaerobic conditions. This type of photosynthesis is thought to be older (having originated ~3 billion years ago) than oxygenic photosynthesis which is found in plants, green algae, and cyanobacteria.[52] If photosystems I and II of green plant evolved from bacterial RCs, their proteins should be related, especially the parts used to form the initial charge separation and electron transfer reactions. Comparison of the protein sequences shows that the L, M, D1, and D2 sequences are all related, although comparison of the RC and PSII sequences indicates a very distant relationship. There is variation in the extent of the homology in different segments of the sequences. The middle portion is the most homologous, in particular the region corresponding to the d helix of the L and M subunits, which contains the binding sites for the donor bacteriochlorophyll and for the iron (see Fig. 4).

Difference scores were obtained for the sequences shown in figure 4 and an evolutionary tree was derived from them (Fig. 7). For each of the four types of proteins studied, the closest relationships were those of the same protein from different species, indicating that the genes for the proteins were established before the species diverged. The next closest relationships were between D1 and D2 and between L and M and the least close between the PSII and bacterial RC proteins. Plants (e.g., *Spinacia*) diverged from green algae (e.g., *Chlamydomonas*) ~0.4 billion years ago, and green algae diverged from cyanobacteria (e.g., *Anacystis*) ~1 billion years ago.[52] Referring to the phylogenetic tree (Fig. 7), these time scales correspond to 1–2 and 2–4 units, respectively. Oxygenic photosynthesis utilizing PSII is thought to have emerged ~2 billion years ago, which is represented

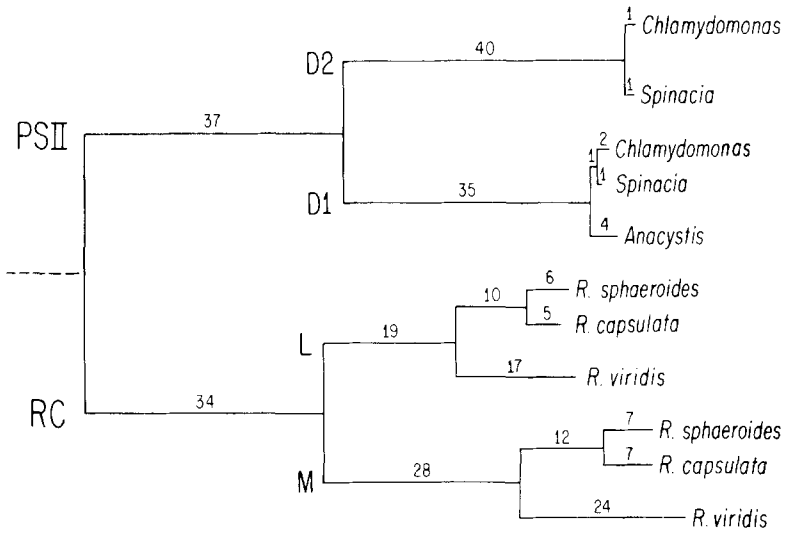

Fig. 7. Evolutionary tree of RC and PSII proteins. Vertical lines indicate divergencies of species or genes. The lengths of the horizontal branches (indicated by numbers, rounded to the nearest unit) show the relative evolutionary distance between proteins as determined from a least square fit of difference scores (see text). The standard deviation between the difference scores calculated from the tree and the difference scores calculated from the original alignments (Table III) is 6%. The common ancestor represented by the root of the tree was arbitrarily placed so that it is approximately equidistant to each descendant.

by ~70 units on the phylogenetic tree (Fig. 7). This branch length (70 units) is not consistent with the branch lengths corresponding to the 1.0- and 0.4-billion year time scales. This discrepancy can be reconciled by assuming that the changes during the first billion years after the divergence of RCs and PSII occurred at a rate that is an order of magnitude greater than the more recently diverged sequences.

The sequence homology in the L, M, D1, and D2 sequences suggests that a functional homology may exist as well, leading to the hypothesis that D1 and D2 make up the core of the PSII reaction center as the L and M subunits do in the bacterial RC.[12,53] Strong experimental evidence in support of this hypothesis has come from the recent observation of a spin polarized triplet (indicative of a charge separation)[54] in the D1D2 complex of PSII particles from spinach.[55] The L subunit is thought to correspond to D1 and the M subunit to D2. This identification is based on the conservation of specific residues involved in binding cofactors and on experimental data such as photoaffinity labeling with herbicides that associate the secondary quinone (Q_B) with L in the RC and D1 in PS II[56–58] and the location of mutations leading to herbicide resistance in both the D1 and L subunits (reviewed in refs. 53, 59–61). In the evolutionary tree, however, there is no obvious line of descent from one of the RC sequences to one of the PS II sequences, as might be expected if only one gene duplication event had given rise to both pairs of polypeptides. Thus the data indicate that two gene duplications took place, leading in both systems to a specialization of the functions of the subunits.

No homologous protein for the H subunit of bacterial RCs has so far been found in the available (NEWAT) data base. It may be that the role of the H subunit is assumed by an entirely different protein in plants or that the gene encoding a homologous protein has not yet been sequenced.

NOTE ADDED IN PROOF

Recently, the complete nucleotide sequences of the chloroplast genome of tobacco (*Nicotiana tabacum*)[62] and of liverwort (*Marchantia polymorpha*)[63] have been published. We have searched the translated DNA sequence of the tobacco chloroplast[61] (155 kilobases) with the overlapping segment search employing the Dayhoff Minimum Mutation Matrix[64]. No significant matches with the H-subunit sequence were found. In addition, we searched for sequences that are similar to four short peptide segments that are strictly conserved in the H subunits of the three bacterial species (REGYPL-searched for 4 identities; PKTF-3ident.; VGPAS-4ident. EEDK-4ident.) (see Fig. 3). Within segments of the size of the H gene no more than two matches were found. We conclude that either eukary-

otes do not have a protein homologous to the H-subunit of prokaryotes or that such a protein, if it exists, is encoded in the nuclear DNA. We thank Drs. M. Sugiura for sending us the DNA sequence of the tobacco chloroplast and R.F. Doolittle for his help in the search.

ACKNOWLEDGMENTS

We thank E.C. Abresch for preparation of the RCs, M.R. Sutton for carrying out digestions with carboxypeptidase A, R.A. Steiner and L.-P. Li for preparation and sequence analysis of peptides, T.J.Donahue and S. Kaplan for the clones containing the H subunit gene, D.F. Feng and R.F. Doolittle for access to computing facilities and unpublished procedures, and M.I. Simon for his help in the early stages of this project. This work was supported by grants from the National Institutes of Health (GM-13191), the National Science Foundation (DMB 82-02811), and the U.S. Department of Agriculture (84-CRCR-1-1475).

REFERENCES

1. Feher, G., Okamura, M.Y. Chemical composition and properties of reaction centers. In: "The Photosynthetic Bacteria." Clayton, R.K., Sistrom, W.R., eds. New York: Plenum Press. 1978:349–386.
2. Sutton, M.R., Rosen, D., Feher, G., Steiner, L.A. Amino-terminal sequence of the L, M, and H subunits of RCs from the photosynthetic bacterium, *Rhodopseudomonas sphaeroides* R-26. Biochemistry 21:3842–3849, 1982.
3. Deisenhofer, J., Epp, O., Miki, K., Huber, R., Michel, H. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3Å resolution. Nature 318:618–624, 1985.
4. Allen, J.P., Feher, G., Yeates, T.O., Rees, D.C., Deisenhofer, J., Michel, H., Huber, R. Structural homology of reaction centers from *R. sphaeroides* and *R. viridis* as determined by X-ray diffraction. Proc. Natl. Acad. Sci. USA, 83:8589–8593, 1986.
5. Williams, J.C., Steiner, L.A., Ogden, R.C., Simon, M.I., Feher, G. Primary structure of the M subunit of the reaction center from *Rhodopseudomonas sphaeroides*. Proc. Natl. Acad. Sci. USA 80:6505–6509, 1983.
6. Williams, J.C., Steiner, L.A., Feher, G., Simon, M.I. Primary structure of the L subunit of the reaction center from *Rhodopseudomonas sphaeroides*. Proc. Natl. Acad. Sci. USA 81:7303–7307, 1984.
7. Donahue, T.J., Hoger, J.H., Kaplan, S. Cloning and expression of the *Rhodobacter sphaeroides* reaction center H gene. J. Bacteriol., 168:953–961, 1986.
8. Sanger, F., Nicklen, S., Coulson, A.R. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467, 1977.
9. Youvan, D.C., Bylina, E.J., Alberti, M., Begusch, H., Hearst, J.E. Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-center, B870 antenna, and flanking polypeptides from R. capsulata. Cell 37:949–957, 1984.
10. Michel, H., Weyer, K.A., Gruenberg, H., Lottspeich, F. The 'heavy' subunit of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: Isolation of the gene, nucleotide and amino acid sequence. EMBO J. 4:1667–1672, 1985.
11. Michel, H., Weyer, K.A., Gruenberg, H., Dunger, I., Oesterhelt, D., Lottspeich, F. The 'light' and 'medium' subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: Isolation of the genes, nucleotide and amino acid sequence. EMBO J. 5:1149–1158, 1986.
12. Michel, H., Deisenhofer, J. X-ray diffraction studies on a crystalline bacterial photosynthetic reaction center: A progress report and conclusions on the structure of photosystem II reaction centers. In: "Encyclopedia of Plant Physiology: Photosynthesis III." Vol. 19. Staehelin, A.C., Arntzen, C.J., eds. Berlin: Springer Verlag. 1986:371–381.
13. Arntzen, C.J., Pakrasi, H.B. Photosystem II reaction center: Polypeptide subunits and functional cofactors. In: "Encyclopedia of Plant Physiology, New Series: Photosynthesis III." Staehelin, A.C., Arntzen, C.J., eds. Berlin: Springer Verlag. 1986:457–467.
14. Kyle, D. J. The 32000 dalton Q_B protein of photosystem II. Photochem. Photobiol. 41:107–116, 1985.
15. Alt, J., Morris, J., Westhoff, P., Herrman, R.G. Nucleotide sequence of the clustered genes for the 44kd chlorophyll α apoprotein and the "32 kd"-like protein of the photosystem II reaction center in the spinach plastid chromosome. Curr. Genetics 8:597–606, 1984.
16. Holschuh, K., Bottomley, W., Whitfeld, P.R. Structure of the spinach chloroplast genes for the D2 and 44 kd reaction-centre proteins of photosystem II and for $tRNA^{Ser}$ (UGA). Nucleic Acids Res. 12:8819–8834,1984.
17. Rasmussen, O.F., Bookjans, G., Stummann, B.M., Henningsen, K.W. Localization and nucleotide sequence of the gene for the membrane polypeptide D2 from pea chloroplast DNA. Plant Mol. Biol. 3:191–199,1984.
18. Rochaix, J.-D., Dron, M., Rahire, M., Malnoe, P. Sequence homology between the 32K dalton and the D2 chloroplast membrane polypeptides of *Chlamydomonas reinhardii*. Plant Mol. Biol. 3:363–370, 1984.
19. Yanisch-Perron, C., Viera, J., Messing, J. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119, 1985.
20. Elliott, B.W. Jr., Eisen, H.N., Steiner, L.A. Amino acid sequence diversity in mouse $\lambda 2$ variable regions. J. Immunol. 133:2757–2761, 1984.
21. Doolittle, R.F. Similar amino acid sequences: chance or common ancestry? Science 214:149–159, 1981.
22. Feng, D.F., Johnson, M.S., Doolittle, R.F. Aligning amino acid sequences: Comparison of commonly used methods. J. Mol. Evol. 21:112–125, 1985.
23. Needleman, S.B., Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48:443–453,1970.
24. Dayhoff, M.O., Schwartz, R.M., Orcutt, B.C. A model of evolutionary change in proteins. In: "Atlas of Protein Sequence and Structure." Vol. V. Dayhoff, M.O., ed. Washington, DC: National Biomedical Research Foundation. 1978:345–358.
25. Feng, D.F., Doolittle, R.F., Mol. Evol., in press.
26. Fitch, W.M., Margoliash, E. Construction of phylogenetic trees. Science 155:279–284, 1967.
27. Erickson, J., Schneider, M., Vallet, J.-M. Dron, M., Bennoun, P. Rochaix, J.-D. Chloroplast gene function: Combined genetic and molecular approach in *Chlamydomonas reinhardii*. In: "Advances in Photosynthesis Research." Vol IV. Sybesma, C., ed. Boston: Martinus Nijhoff/Dr. W. Junk. 1984:491–500.
28. Zurawski, G., Bohnert, H.J., Whitfeld, P.R., Bottomley, W. Nucleotide sequence of the gene for the M_r 32,000 thylakoid membrane protein from *Spinacia oleracea* and *Nicotiana debneyi* predicts a totally conserved primary translation product of M_r 38,950. Proc. Natl. Acad. Sci. USA 79:7699–7703, 1982.
29. Golden, S.S., Haselkorn, R. Mutation to herbicide resistance maps within the psbA gene of *Anacystis nidulans* R2. Science 229:1104–1107, 1985.
30. Fergusen, K.A. Starch-gel electrophoresis-application to the classification of pituitary proteins and polypeptides. Metabolism 13:985–1002, 1964.
31. Banker, G.A., Cotman, C.W. Measurement of free electrophoretic mobility and retardation coefficient of protein-sodium dodecyl sulfate complexes by gel electrophoresis. J. Biol. Chem. 247:5856–5861, 1972.
32. Rosen, D. The isolation and characterization of the subunits of the photosynthetic reaction center and their interaction with c-type cytochromes. Ph.D. Dissertation (1979). University of California, San Diego.
33. Okamura, M.Y., Steiner, L.A., Feher, G. Characterization of reaction centers from photosynthetic bacteria. I. Subunit structure of the protein mediating the primary photochemistry in *Rhodopseudomonas sphaeroides* R-26. Biochemistry 13:1394–1403, 1974.
34. Steiner, L.A., Okamura, M.Y., Lopes, A.D., Moskowitz, E.,

Feher, G. Characterization of reaction centers from photosynthetic bacteria. II. Amino acid composition of the reaction center protein and its subunits in *Rhodopseudomonas sphaeroides* R-26. Biochemistry 13:1403–1410,1974.
35. Straley, S.C., Parson, W.W., Mauzerall, D.C., Clayton, R.K. Pigment content and molar extinction coefficients of photochemical reaction centers from *Rhodopseudomonas sphaeroides*. Biochim. Biophys. Acta 305:597–609, 1973.
36. Marrs, B. Mobilization of the genes for photosynthesis from *Rhodopseudomonas capsulata* by a promiscuous plasmid. J. Bacteriol. 146:1003–1012, 1981.
37. Mandel, M., Leadbetter, E.R., Pfennig, N., Trüper, H.G. Deoxyribonucleic acid base compositions of phototrophic bacteria. Int. J. Syst. Bacteriol. 21:222–230, 1971.
38. Bernardi, G., Bernardi, G. Codon usage and genome composition. J. Mol. Evol. 22:363–365, 1985.
39. Bachman, R.C., Gillies, K., Takemoto, J.Y. Membrane topography of the photosynthetic reaction center polypeptides of *Rhodopseudomonas sphaeroides*. Biochemistry 20:4590–4596, 1981.
40. Hall, R.L. Doorley, P.F., Niederman, R.A. Trans-membrane localization of reaction center proteins in *Rhodopseudomonas sphaeroides* chromatophores. Photochem. Photobiol. 28:273–276, 1978.
41. Valkirs, G.E., Feher, G. Topography of reaction center subunits in the membrane of the photosynthetic bacterium, *Rhodopseudomonas sphaeroides*. J. Cell Biol. 95:179–188, 1982.
42. Wiemken, V., Bachofen, R. Probing the topology of proteins in the chromatophore membrane of *Rhodospirillum rubrum* G-9 with proteinase K. Biochim. Biophys. Acta 681:72–76, 1982.
43. Zürrer, H., Snozzi, M., Hanselmann, K., Bachofen, R. Localisation of the subunits of the photosynthetic reaction centers in the chromatophore membrane of *Rhodospirillum rubrum*. Biochim. Biophys. Acta 460:273–279, 1977.
44. Tanford, C. The hydrophobic effect and the organization of living matter. Science 200:1012–1018, 1978.
45. Marinetti, G.V., Cattieu, K. Lipid analysis of cells and chromatophores of *Rhodopseudomonas sphaeroides*. Chem. Phys. Lipids 28:241–251, 1981.
46. Kyte, J., Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105–132, 1982.
47. Deisenhofer, J., Epp, O., Miki, K.,. Huber, R., Michel, H. X-ray structure analysis of a membrane protein complex. Electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodopseudomonas viridis*. J. Mol. Biol. 180:385–398, 1984.
48. Nabedryk, E., Tiede, D., Dutton, P.L., Breton, J. Polarized infrared spectroscopy of bacterial reaction centers: The LMH and LM complexes in reconstituted membrane. In: "Advances in Photosynthesis Research." Vol. II. Sybesma, C., ed. The Netherlands: Martinus Nijhoff/Dr. W. Junk. 1984:II.3.177-II.3.180.
49. Brunisholz, R.A., Wiemken, V., Suter, F., Bachofen, R., Zuber, H. The light-harvesting polypeptides of *Rhodospirillum rubrum*. II. Localisation of the amino-terminal regions of the light-harvesting polypeptides B 870-α and B 870-β and the reaction centre subunit L at the cytoplasmic side of the photosynthetic membrane of *Rhodospirillum rubrum* G-9. Hoppe-Seyler's Z. Physiol. Chem. 365:689–701, 1984.
50. Okamura, M.Y., Feher, G. Labeling the cytochrome c_2 binding site in reaction centers from *Rhodopseudomonas sphaeroides*. Biophys. J. 41:122a, 1983.
51. Gibson, J., Stackebrandt, E., Zablen, L.B., Gupta, R., Woese, C.R. A phylogenetic analysis of the purple photosynthetic bacteria. Curr. Microbiol. 3:59–64, 1979.
52. Olson, J.M. The evolution of photosynthesis. Science 168:438–446, 1970.
53. Trebst, A. The topology of the plastoquinone and herbicide binding peptides of photosystem II in the thylakoid membrane. Z. Naturforsch. 41c:240–245, 1986.
54. Okamura, M.Y., Satoh, K., Isaacson, R.A., Feher, G. Evidence of the primary charge separation in the D_1 D_2 complex of photosystem II from spinach: EPR of the triplet state. Progress in Photosynthesis Research, Vol. 1, I.4.379–I.4.381 (Proceedings of the VIIth International Conference on Photosynthesis, Brown University, August 1986, ed. J. Biggins), Martinus Nijhoff, Boston, 1987.
55. Satoh, K., Nanba, O. Isolation of a photosystem II reaction center consisting of γ and δ subunits (D-1 and D-2) and cytochrome b-559. Progress in Photosynthesis Research, Vol. 2, II.1.69–II.1.72 (Proceedings of the VIIth International Conference on Photosynthesis, Brown University, August 1986, ed. J. Biggins), Martinus Nijhoff, Boston, 1987.
56. Brown, A.E., Gilbert, C.W., Guy, R., Arntzen, C.J. Triazine herbicide resistance in the photosynthetic bacterium *Rhodopseudomonas sphaeroides*. Proc. Natl. Acad. Sci. USA 81:6310–6314, 1984.
57. Pfister, K., Steinback, K.E., Gardner, G., Arntzen, C.J. Photoaffinity labeling of an herbicide receptor protein in chloroplast membranes. Proc. Natl. Acad. Sci. USA 78:981–985, 1981.
58. de Vitry, C., Diner, B.A. Photoaffinity labeling of the azidoatracine receptor site in reaction centers of *Rhodopseudomonas sphaeroides*. F.E.B.S. Letts. 167:327–331, 1984.
59. Gilbert, C.W., Williams, J.G.K., Williams, K.A.L., Arntzen, C.J. Herbicide action in photosynthetic bacteria. In: "Molecular Biology of the Photosynthetic apparatus." Cold Spring Harbor, New York: Cold Spring Harbor Laboratory. 1985:67–71.
60. Schenk, C.C., Sistrom, W.R., Bunzow, J.R., Rambousek, E.L., Capaldi, R.A. An herbicide-resistance mutation localizes the low-affinity quinone binding site in reaction centers from the genus *Rhodopseudomonas*. Biochemistry, in press, 1986.
61. Paddock, M.L., Williams, J.C., Rongey, S.H., Abresch, E.C., Feher, G., Okamura, M.Y. Characterization of three herbicide resistant mutants of *Rhodopseudomonas sphaeroides* 2.4.1: Structure-function relationship. Progress in Photosynthesis Research, Vol. 3, III.11.775–III.11.778 (Proceedings of the VIIth International Conference on Photosynthesis, Brown University, August 1986, ed. J. Biggins), Martinus Nijhoff, Boston, 1987.
62. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B. Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., Sugiura, M. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. 5:2043–2049, 1986.
63. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.-i., Inokuchi, H., Ozeki, H. Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. Nature 322:572–574, 1986.
64. Doolittle, R.F. "Of URFS and ORFS", Ch. 1, University Science Books, Mill Valley, California, 1987.