Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 6505–6509, November 1983
Biochemistry

Primary structure of the M subunit of the reaction center from *Rhodopseudomonas sphaeroides*

(photosynthesis/oligonucleotide probe/nucleotide sequence/membrane protein)

J. C. WILLIAMS*†, L. A. STEINER‡, R. C. OGDEN§, M. I. SIMON§, AND G. FEHER†

Departments of *Biology and †Physics, University of California, San Diego, La Jolla, CA 92093; ‡Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and §Agouron Institute, 505 Coast Boulevard South, La Jolla, CA 92037

Contributed by G. Feher, July 25, 1983

ABSTRACT **The reaction center is a membrane-bound bacteriochlorophyll–protein complex that mediates the primary photochemical events in the photosynthetic bacterium *Rhodopseudomonas sphaeroides*. The previously determined amino-terminal sequences of the three subunits of the reaction center protein were used to design synthetic mixed oligonucleotide probes for the structural genes encoding the subunits. One of these probes was used to isolate and clone a fragment of DNA from *R. sphaeroides* that contained the gene encoding the M subunit. The nucleotide sequence of this gene was determined by the dideoxy method. In addition, a number of tryptic and chymotryptic peptides from the M protein were isolated and subjected to sequence analysis, and the sequence of the carboxyl terminus was determined. Together with the amino-terminal sequence, the data establish the primary structure of the M protein. The distribution of hydrophobic residues in the amino acid sequence suggests the presence of five membrane-spanning segments. A significant homology was found between the amino acid sequence of the M subunit and a thylakoid membrane protein (M_r 32,000) from spinach that has been implicated in herbicide and quinone binding.**

The reaction center (RC) is the site of the initial photochemical charge separation in the electron transfer process of photosynthesis (for reviews, see refs. 1 and 2). In the purple bacterium *Rhodopseudomonas sphaeroides*, the RC is part of the photosynthetic apparatus found in the highly invaginated cytoplasmic membrane that is synthesized by the bacterium under anaerobic conditions (3). The RC protein ($M_r \approx 100{,}000$) contains three highly hydrophobic subunits, designated L, M, and H, in a 1:1:1 stoichiometry (4, 5). The cofactors found in the RC are four bacteriochlorophylls, two bacteriopheophytins, two ubiquinones, and one iron.

The determination of the amino acid sequence of the RC polypeptides is essential to the detailed elucidation of the RC structure. The sequences of 25–28 amino-terminal residues of each subunit have been determined previously by automated Edman degradation (6). These were used to design mixed-sequence oligonucleotide hybridization probes for the structural genes of each of the subunits. The probe for the gene encoding the M subunit was used to identify a restriction fragment of *R. sphaeroides* DNA that contained the gene. The nucleotide sequence of this gene was determined by the dideoxy method and is presented together with the amino acid sequence derived from it. The sequences of the carboxyl terminus and of several peptides were determined for corroboration of the sequence obtained from the DNA. Preliminary accounts of this work have been given (7, 8).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

EXPERIMENTAL PROCEDURES

Materials. Large fragment *Escherichia coli* DNA polymerase I, M13mp7 replicative form (RF) DNA, M13 24-base-pair primer fragment, dideoxynucleotides, and the restriction enzymes *Acc* I, *Bam*HI, *Eco*RI, *Hin*dIII, *Hpa* II, *Kpn* I, *Sal* I, *Sma* I, and *Taq* I were obtained from Bethesda Research Laboratories. T4 polynucleotide kinase, T4 DNA ligase, and the restriction enzymes *Nru* I, *Pst* I, and *Sau*3AI were from New England BioLabs. Deoxynucleotides, M13mp8 RF DNA, and M13mp9 RF DNA were from P-L Biochemicals. Pronase was from Calbiochem and lysozyme was from Sigma. [γ-^{32}P]ATP, [α-^{32}P]dATP, and α-^{35}S-labeled deoxyadenosine 5′-[α-thio]triphosphate were from Amersham. Sources for trypsin, chymotrypsin, and carboxypeptidases A and B have been described (9). Plasmid DNA used for cloning was prepared by a cleared lysate procedure (10). *E. coli* HB101 (11) and *E. coli* JM103 (12) were used as hosts for plasmid and phage vectors, respectively. The wild-type *R. sphaeroides* 2.4.1 (13) was used to obtain DNA for cloning due to the ease of aerobic culturing. *R. sphaeroides* R-26 (14) was used to isolate the protein for peptide and carboxyl-terminal analysis. The R-26 strain is a carotenoidless mutant derived from the wild type by ultraviolet mutagenesis; it reverts readily to the wild type and is believed to have a single point mutation in a gene associated with carotenoid synthesis.

Oligonucleotide Synthesis. The mixed oligonucleotide probes were synthesized on a polyacrylmorpholide resin by a modified triester method (15). The chains were lengthened by the addition of monomers or dimers and, at positions where the sequence was ambiguous due to synonymous codons, equimolar amounts of the possible nucleotides were added. After removal of the protecting groups and cleavage from the resin, the oligonucleotides were purified by column chromatography and polyacrylamide gel electrophoresis (16). The oligonucleotides were labeled with ^{32}P by using T4 polynucleotide kinase and [γ-^{32}P]ATP.

Isolation of Genomic DNA. *R. sphaeroides* 2.4.1 was grown aerobically to stationary phase in 10 g of tryptone per liter, 10 g of NaCl per liter, and 5 g of yeast extract per liter and collected by centrifugation at 12,000 × *g*. The cells (wet weight, 3 g) were washed and resuspended in 50 ml of 0.15 M NaCl/0.1 M EDTA, pH 8, and treated with lysozyme (10 mg/ml, 30 min, 37°C), $NaDodSO_4$ (5%, 10 min, 60°C), and Pronase (50 mg, 1 hr, 37°C). The DNA was isolated by centrifugation in a cesium chloride equilibrium density gradient (Ti 60 rotor, 40,000 rpm, 40 hr, 15°C). After removal from the gradient, the DNA was dialyzed against 6 mM Tris·HCl, pH 7.4/10 mM NaCl/0.1 mM EDTA, extracted with phenol, precipitated in ethanol, and

Abbreviations: RC, reaction center; kb, kilobase(s); F_3CCOOH, trifluoroacetic acid; RF, replicative form.

M subunit
3′ CT$^{T}_{C}$ AT$^{A}_{G}$ GT$^{T}_{C}$ TT$^{A}_{G}$ TA 5′
Glu Tyr Gln Asn Ile

L subunit
3′ AA$^{A}_{G}$ CT$^{A}_{G}$ AA$^{A}_{G}$ 5′
Phe Asp Phe

H subunit
3′ AA$^{A}_{G}$ ACC TA$^{A}_{T}$G AA$^{A}_{G}$ 5′
Phe Trp Ile Phe

FIG. 1. Sequences of the oligonucleotide probes for the genes encoding the M, L, and H subunits of the RC. All combinations of the sequences for each probe were synthesized simultaneously. The corresponding amino acid sequences (6) are also shown.

stored in the dialysis buffer. For hybridization experiments, the DNA was digested with *Bam*HI restriction endonuclease and separated by electrophoresis in a 0.8% agarose gel.

Hybridization of Oligonucleotides to DNA. After agarose gel electrophoresis, DNA fragments were transferred to nitrocellulose filters (17), incubated with labeled probe for 20 hr at 24°C (16), washed four times with 0.9 M NaCl/0.09 M sodium citrate at 4°C, and autoradiographed at −80°C for 24 hr on Kodak XAR-5 film with an intensifying screen.

Cloning. Preliminary experiments determined the size of the *Bam*HI restriction fragment of *R. sphaeroides* DNA that hybridized to the probe for the M subunit. DNA fragments of approximately this size were eluted from an agarose gel (18), ligated into the plasmid pBR322, and transformed into *E. coli* to create a clone bank enriched for the target fragment. Plasmid DNA from each clone was isolated (19), electrophoresed, and checked for hybridization to the probe.

Subfragments of the cloned *Bam*HI fragment were obtained by digestion with restriction enzymes. The sizes of these fragments were determined by comparing their electrophoretic mobility in agarose gels with that of λ cI857 *Eco*RI–*Hin*dIII fragments. A 4.5-kilobase (kb) *Pst* I subfragment and a 2.5-kb *Sal* I subfragment that hybridized to the probe were identified and isolated. From the 4.5-kb *Pst* I fragment, *Hpa* II and *Taq* I fragments were obtained and cloned into the *Acc* I site of the phage vector M13mp7 (12). The resulting clones were screened with the probe to identify those that contained the amino-terminal coding sequences. From the 2.5-kb *Sal* I fragment, *Hpa* II, *Taq* I, and *Sau*3AI fragments were obtained either directly after digestion or after separation on a 9% polyacrylamide gel; these fragments were cloned into M13mp8 (20). A 0.7-kb *Sal* I–*Nru* I fragment was cloned in both orientations into the *Sal* I and *Sma* I sites of M13mp8 and M13mp9, and a 1.2-kb *Nru* I fragment was cloned into the *Sma* I site of M13mp8.

DNA Sequence Analysis. The M13 recombinant phage were used to prepare single-stranded templates for sequence analysis by the dideoxy method (21), with either ^{32}P or ^{35}S (22) as a label. The labeled DNA was electrophoresed at ≈1,500 V and ≈40 mA (maintaining constant power) in 6 or 8% polyacrylamide gels (0.4 mm × 25 cm × 35 cm) that were covalently bound to one of the glass electrophoresis plates (23). The gel plates were backed by 6-mm aluminum plates to reduce temperature gradients. After electrophoresis, the fixed gels were dried onto the glass plate and autoradiographed on Kodak XAR-5 film. The sequences were analyzed with the computer programs of Staden (24).

Preparation and Sequence Analysis of Peptides. For digestion with trypsin and chymotrypsin, the M subunit was isolated as described (6). The M subunit was quite resistant to digestion with trypsin; to 0.4 mg in 0.7 ml of 0.1 M NH_4HCO_3, a total of 30 μg of trypsin was added in four aliquots over 48 hr to achieve ≈20% release of lysine and ≈60% release of arginine after digestion with carboxypeptidase B. After lyophilization the digest was taken up in 0.2 ml of 0.1% trifluoroacetic acid (F_3CCOOH). The soluble peptides (representing ≈20% of the total material) were fractionated by reversed-phase HPLC with a 4.6 × 250 mm Altex ultrasphere ODS column. Elution was at 1 ml/min with 0.1% F_3CCOOH (6 min), followed by a linear gradient to 0.1% F_3CCOOH/60% acetonitrile (60 min). The effluent was monitored at 214 and 280 nm. Selected fractions were subjected to amino acid analysis, determination of amino-terminal residues by the microdansyl procedure, and automated Edman degradation as described (6, 9).

For digestion with chymotrypsin, 0.4 mg of the M subunit in 3.3 ml of 1 mM NH_4HCO_3 was treated with two aliquots of enzyme (20 μg each) for a total of 18 hr at 37°C. After lyophilization, 13% of the digest was soluble in 0.1% F_3CCOOH; this material was fractionated by HPLC. Selected fractions were subjected to amino acid analysis and manual dansyl Edman degradation. Glutamic acid and aspartic acid are not distinguished from amides in this procedure.

Digestion with Carboxypeptidase A. The M subunit was dialyzed into 0.1 M NH_4HCO_3 and aliquots (≈30 μg in ≈0.4 ml) were treated with 1 μg of carboxypeptidase A for 2 and 20 min

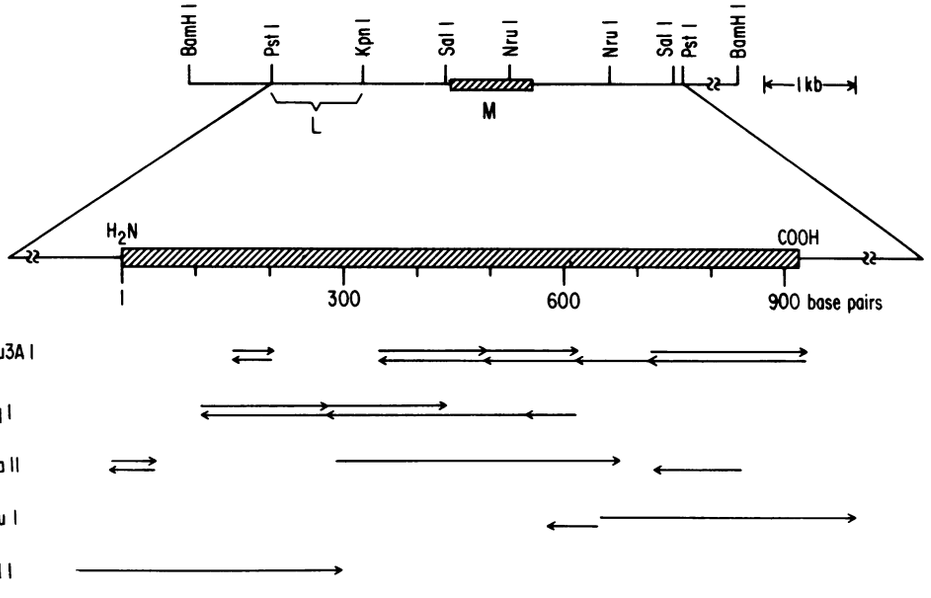

FIG. 2. Restriction map of the ≈13-kb *Bam*HI fragment and sequence analysis strategy. The box indicates the position of the gene encoding the M subunit of the RC. Also shown is the fragment to which the probe from the amino-terminal region of the L subunit hybridizes. The probe for the H subunit did not hybridize to the *Bam*HI fragment. The arrows indicate the extent and direction of sequence analysis of subcloned fragments.

at room temperature. The digests were lyophilized and applied directly to the amino acid analyzer.

RESULTS

Location of Structural Genes. The probe for the gene encoding the M subunit consisted of a mixture of 16 oligonucleotides, each 14 bases long. Probes for the gene encoding the L subunit and the gene encoding the H subunit were also constructed, as shown in Fig. 1. The sequences of the probes were complementary to the messenger RNA. The M probe hybridized to a single band obtained by electrophoresis of *R. sphaeroides* DNA that had been digested with *Bam*HI. An *E. coli* clone containing a plasmid with this fragment inserted was isolated from a clone bank constructed from size-fractionated DNA fragments. This fragment was found to be ≈13 kb long and was mapped with respect to a number of restriction enzymes (Fig. 2). The L probe hybridized to a 1.0-kb *Pst* I–*Kpn* I subfragment of this *Bam*HI fragment (Fig. 2). The H probe did not hybridize to the *Bam*HI fragment.

Nucleotide Sequence. The nucleotide sequence of the gene encoding the M subunit was determined from 12 restriction fragments cloned into M13 phage vectors (Fig. 2). These clones allowed 94% of the sequence to be determined on both the coding and noncoding strands. Several regions of the sequencing gels (corresponding to residues 45–46, 136–137, and 247–249) showed band compression, and sequence data in these regions were also obtained from only one strand. Amino acid sequence data obtained from the protein confirmed 24% of the sequence. The complete nucleotide and derived amino acid sequences are shown in Fig. 3.

Isolation and Sequence Analysis of Tryptic and Chymotryptic Peptides. After digestion of the M subunit with trypsin, four arginine-containing peptides, varying in length from 2 to 16 residues, were isolated by HPLC in 30–40% yield; their sequence was determined by automated Edman degradation. The

```
GTCGACTGGTGGCAATGGTGGGTGAAGCTGCCGTGGTGGGCGAACATCCCGGGAGGCATCA ATG GCT GAG TAT CAG AAC ATC TTC TCC CAG GTC
                                                                (Met)Ala Glu Tyr Gln Asn Ile Phe Ser Gln Val
                                                                                                        10

CAG GTC CGC GGA CCG GCC GAC CTG GGG ATG ACC GAA GAC GTC AAC CTG GCC AAC CGT TCG GGC GTC GGT CCC TTC TCG ACC CTG CTC GGC
Gln Val Arg Gly Pro Ala Asp Leu Gly Met Thr Glu Asp Val Asn Leu Ala Asn Arg Ser Gly Val Gly Pro Phe Ser Thr Leu Leu Gly
                  20                                                  30                                  40

TGG TTC GGC AAC GCC CAG CTC GGC CCG ATC TAT CTC GGC TCG CTC GGC GTC CTG TCC CTC TTC TCG GGC CTG ATG TGG TTC TTC ACC ATC
Trp Phe Gly Asn Ala Gln Leu Gly Pro Ile Tyr Leu Gly Ser Leu Gly Val Leu Ser Leu Phe Ser Gly Leu Met Trp Phe Phe Thr Ile
                  50                                                  60                                  70

GGG ATC TGG TTC TGG TAT CAG GCG GGC TGG AAC CCG GCC GTC TTC CTG CGC GAC CTG TTC TTC TTC TCG CTC GAG CCG CCG GCA CCC GAA
Gly Ile Trp Phe Trp Tyr Gln Ala Gly Trp Asn Pro Ala Val Phe Leu Arg Asp Leu Phe Phe Phe Ser Leu Glu Pro Pro Ala Pro Glu
                  80                                                  90                                  100

TAC GGT CTG TCC TTC GCG GCT CCG CTG AAG GAA GGC GGG CTG TGG CTG ATC GCG TCG TTC TTC ATG TTC GTC GCG GTC TGG TCC TGG TGG
Tyr Gly Leu Ser Phe Ala Ala Pro Leu Lys Glu Gly Gly Leu Trp Leu Ile Ala Ser Phe Phe Met Phe Val Ala Val Trp Ser Trp Trp
                  110                                                 120                                 130

GGC CGC ACC TAT CTC CGC GCT CAG GCG CTG GGC ATG GGC AAG CAC ACC GCC TGG GCG TTC CTC TCG GCC ATC TGG CTG TGG ATG GTG CTG
Gly Arg Thr Tyr Leu Arg Ala Gln Ala Leu Gly Met Gly Lys His Thr Ala Trp Ala Phe Leu Ser Ala Ile Trp Leu Trp Met Val Leu
                  140                                                 150                                 160

GGC TTC ATC CGT CCG ATC CTC ATG GGG TCC TGG TCG GAA GCG GTT CCC TAC GGC ATC TTC TCG CAC CTC GAC TGG ACG AAC AAC TTC TCG
Gly Phe Ile Arg Pro Ile Leu Met Gly Ser Trp Ser Glu Ala Val Pro Tyr Gly Ile Phe Ser His Leu Asp Trp Thr Asn Asn Phe Ser
                  170                                                 180                                 190

CTC GTC CAC GGC AAC CTG TTC TAC AAC CCC TTC CAC GGT CTC TCG ATC GCC TTC CTC TAC GGG TCG GCC CTG CTC TTC GCG ATG CAC GGT
Leu Val His Gly Asn Leu Phe Tyr Asn Pro Phe His Gly Leu Ser Ile Ala Phe Leu Tyr Gly Ser Ala Leu Leu Phe Ala Met His Gly
                  200                                                 210                                 220

GCG ACC ATC CTC GCG GTC TCC CGC TTC GGC GGC GAG CGC GAG CTG GAG CAG ATC GCC GAC CGC GGG ACG GCA GCG GAG CGG GCC GCC CTC
Ala Thr Ile Leu Ala Val Ser Arg Phe Gly Gly Glu Arg Glu Leu Glu Gln Ile Ala Asp Arg Gly Thr Ala Ala Glu Arg Ala Ala Leu
                  230                                                 240                                 250

TTC TGG CGC TGG ACC ATG GGT TTC AAC GCC ACG ATG GAA GGC ATC CAC CGC TGG GCC ATC TGG ATG GCG GTC CTC GTG ACC CTC ACC GGC
Phe Trp Arg Trp Thr Met Gly Phe Asn Ala Thr Met Glu Gly Ile His Arg Trp Ala Ile Trp Met Ala Val Leu Val Thr Leu Thr Gly
                  260                                                 270                                 280

GGC ATC GGC ATC CTG CTC TCG GGC ACG GTC GTG GAC AAC TGG TAC GTC TGG GGC CAG AAC CAC GGC ATG GCG CCG CTG AAC TGA
Gly Ile Gly Ile Leu Leu Ser Gly Thr Val Val Asp Asn Trp Tyr Val Trp Gly Gln Asn His Gly Met Ala Pro Leu Asn Stop
                  290                                                 300
```

FIG. 3. Nucleotide sequence of the M subunit gene. The noncoding strand is shown, along with 61 nucleotides of the sequence that precedes the structural gene. The box contains the sequence complementary to the probe. Amino acid residues that confirm the previously determined amino-terminal sequence (6) are in italics. The positions of isolated tryptic (|——|) and chymotryptic (|----|) peptides are indicated; residues identified by Edman degradation are indicated by >; residues identified by digestion with carboxypeptidase A are indicated by ←.

results are shown in Fig. 3, superimposed on the sequence derived from the DNA sequence. An additional peptide that was obtained in lower (≈10%) yield appeared, from its amino acid composition, to correspond to the amino-terminal 13 residues of the M subunit; its sequence was not determined. After digestion with chymotrypsin, four pure peptides were obtained and subjected to sequence analysis by the manual dansyl-Edman method; the resulting sequences are indicated in Fig. 3.

Carboxyl-Terminal Analysis. Amino acid analysis of carboxypeptidase A digests of the M subunit showed elution peaks corresponding in position to leucine and serine. However, under the conditions of the analysis, asparagine and glutamine coelute with serine. Asparagine can be distinguished from both glutamine and serine by the ratio of the absorbance of the ninhydrin derivatives at 590 and 440 nm, and the absorbance ratios obtained by this method indicated that the material in the "serine" peak was largely asparagine. This was confirmed by showing that aspartic acid was present after hydrolysis of the supernatant obtained by precipitating the digested protein with trichloroacetic acid. Results obtained after digestion for 2 and 20 min indicated that the carboxyl-terminal sequence of the M subunit is Leu-Asn-COOH. The residue on the amino-terminal side of leucine is proline (see Fig. 3) which is not released by carboxypeptidase A.

DISCUSSION

The structural gene encoding the M subunit of the RC of *R. sphaeroides* was identified by using a synthetic mixed oligonucleotide probe based on the amino-terminal sequence. The identity of the gene was confirmed by comparison of the derived amino acid sequence with sequences obtained by Edman degradation of the amino-terminal region and of several peptides as well as the determination of the carboxyl-terminal residues. The data show that the M subunit is composed of 307 amino acid residues corresponding to a M_r of 34,265. This value is in fair agreement with that calculated from the amino acid composition (M_r 32,000) (25) but disagrees considerably with the value obtained by $NaDodSO_4$/polyacrylamide gel electrophoresis (M_r 24,000) (4). The discrepancy with the latter value is thought to result from the high hydrophobicity of the protein which increases the amount of bound $NaDodSO_4$.

The coding sequence for the mature protein begins immediately after an ATG codon, suggesting that there is no leader sequence. Although hydrophobic leader sequences have been found in many excreted and some membrane proteins of bacteria, they are not essential for insertion of proteins into the membrane (reviewed in ref. 26). The two carboxyl-terminal residues of the protein (Leu-Asn) correspond to the codons preceding a stop codon, indicating that there is no post-translational modification of the carboxyl terminus. The sequence G-G-A-G-G, starting 10 bases upstream from the initiating ATG, is complementary to a C-C-U-C-C sequence at the 3′ terminus of the 16S RNA of *R. sphaeroides* (27) and is a potential ribosome binding site (28). Analysis of the codon usage in the gene shows that the third position is a G or a C in 92% of the codons. Overall, the G + C content of the gene is 65%, similar to the G + C composition of the total DNA (69%) (29). A similar preference for codons ending in G and C was found in the bacteriorhodopsin gene of *Halobacterium halobium* (30).

The hybridization position of the L probe indicates that the start of the L gene is 1–2 kb from the start of the gene encoding the M subunit. Partial sequence analysis of the L gene shows that it has a high degree of homology with M and that it is contiguous with it. The H gene is at least 3 kb away from the M gene.

The M subunit has been shown to span the cytoplasmic membrane by several labeling techniques (31, 32). The sequence obtained from the gene was examined for potential membrane-spanning regions by the method of Kyte and Doolittle (33). In this method, each amino acid is rated on a hydropathy scale, which ranges from 4.5 for isoleucine, judged to be the most hydrophobic amino acid, to −4.5 for arginine, the most hydrophilic residue. The average hydropathy value of a moving segment (19 residues) is calculated over the entire sequence. A hydropathy profile generated in this manner is shown in Fig. 4. This analysis suggests that there are five hydrophobic segments, having an average hydropathy index ≥ 1.5, that are long enough (at least 20 residues) to span the membrane in an α helix (see peaks I–V in Fig. 4). The prediction of helices spanning the membrane is in agreement with circular dichroism and polarized infrared spectroscopy data, which indicate that the RC protein consists largely of α helices that are approximately perpendicular to the plane of the membrane (34). The segments that separate the putative membrane-spanning regions contain almost all of the charged residues. They were analyzed for β turns by the method of Chou and Fasman (35); each of these segments contained at least one tetrapeptide predicted to form a β turn (having a $p_t \geq 1.5 \times 10^{-4}$). The soluble tryptic and chymotryptic peptides were derived from these less hydrophobic segments. Because an odd number of membrane-spanning segments are observed, the amino and carboxyl ends of the protein must be on opposite sides of the membrane. Their

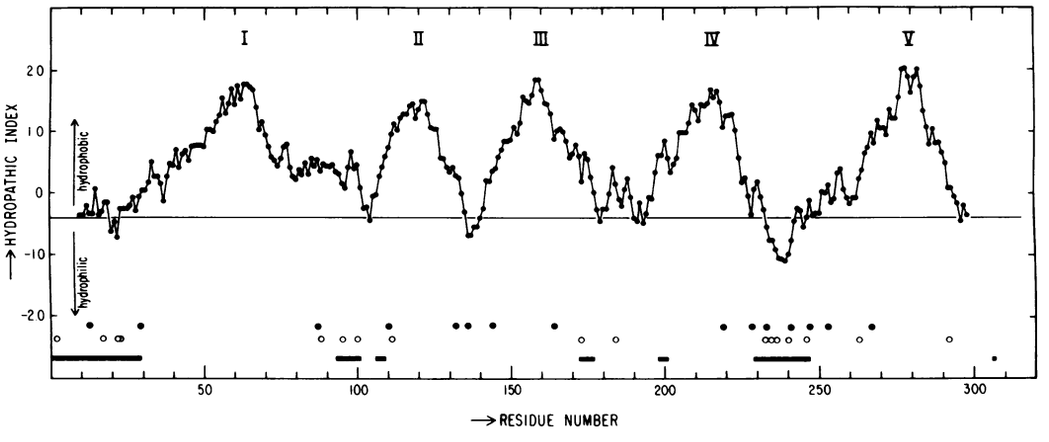

FIG. 4. Hydropathy profile of the sequence of the M subunit. The average hydropathy value (33) of a moving window of 19 amino acids is plotted at the midpoint of the window. Positions of basic (Lys, Arg) and acidic (Glu, Asp) residues are denoted by ● and ○, respectively. Solid bars indicate the positions of isolated peptides and residues identified by digestion with carboxypeptidase A.

location may be established by selective labeling and proteolysis experiments.

Little is known about the detailed structure of the binding sites of the cofactors of the RC. One of the histidine residues (219) is in a hydrophobic segment of M and is a candidate for a bacteriochlorophyll or quinone binding site. The binding sites for both quinones have been localized on the M subunit (36, 37). More information about these sites may be obtained by binding photoaffinity-labeled quinones to the M subunit and analyzing the peptides to which they are attached. Cytochrome *c* (the secondary donor) has been shown to interact with the L and M subunits (38); carboxyl groups on the L and M subunits were implicated in the binding (39). Labeling and peptide analysis of the M subunit with and without cytochrome *c* present may identify these groups. Thus, the sequence provides a framework for determining the structure of binding sites on the RC.

A computer search for sequences homologous to the M subunit was performed by Doolittle using a data base of ≈1,400 proteins (40). A region of the M sequence (residues 163–220) was found to be similar to a region (residues 160–216) of the thylakoid membrane protein in spinach (*Spinacea oleracea*) referred to as M_r 32,000 (41). If one residue is deleted at position 194 in the sequence of the M subunit, 21 of the amino acids in the 57-residue region are identical. In addition, the hydropathy profile of the M_r 32,000 protein bears strong resemblance to that of the M subunit. The M_r 32,000 protein is thought to be involved in herbicide and quinone binding (42). The observed homology may indicate conservation of a functionally analogous region of the two proteins.

The primary structure establishes that the M subunit is a complex membrane protein, with several membrane-spanning segments, similar to bacteriorhodopsin and other ion transport proteins (reviewed in ref. 43). The determination of the primary structures of the L and H subunits will give a more complete picture of the RC and will provide necessary information for obtaining the three-dimensional structure by x-ray crystallography. Crystals of RCs have been obtained from *R. viridis* by Michel (44) and from *R. sphaeroides* (unpublished data).¶ The isolation of the RC genes provides a basis for analyzing previously characterized mutations in the RC (45), the introduction of specific amino acid alterations into the RC by oligonucleotide-directed mutagenesis (46), the study of the regulation of the synthesis of the protein (47), and the evaluation of the evolutionary relationships among photosynthetic organisms (48).

¶ Allen, J. P. & Feher, G., Meeting of the Biophysical Society, Feb. 13–16, 1983, San Diego, CA.

We thank E. Abresch for the preparation of the RCs, E. Capuno-Tichy and L.-P. Li for assistance in peptide preparation and analysis, and R. F. Doolittle for the computer search for homologous protein sequences. This work was supported by grants from the National Institutes of Health (GM 13191 and 07313), the National Science Foundation (PCM 82-02811), the Office of Naval Research (N00014-83-K-0079), and the U.S. Department of Agriculture (5901-0410-9-0378-0).

1. Feher, G. & Okamura, M. Y. (1978) in *The Photosynthetic Bacteria*, eds. Clayton, R. K. & Sistrom, W. R. (Plenum, New York), pp. 349–386.
2. Okamura, M. Y., Feher, G. & Nelson, N. (1982) in *Photosynthesis: Energy Conversion by Plants and Bacteria*, ed. Govindjee (Academic, New York), Vol. 1, pp. 195–272.
3. Oezle, J. & Drews, G. (1972) *Biochim. Biophys. Acta* **265**, 209–239.
4. Okamura, M. Y., Steiner, L. A. & Feher, G. (1974) *Biochemistry* **13**, 1394–1403.
5. Steiner, L. A., Okamura, M. Y., Lopes, A. D., Moskovitz, E. & Feher, G. (1974) *Biochemistry* **13**, 1403–1410.
6. Sutton, M. R., Rosen, D., Feher, G. & Steiner, L. A. (1982) *Biochemistry* **16**, 3842–3849.
7. Williams, J. C., Feher, G. & Simon, M. I. (1983) *Biophys. J.* **41**, 122a (abstr.).
8. Williams, J. C., Feher, G. & Simon, M. I. (1983) *Photochem. Photobiol.* **37**, S112 (abstr.).
9. Steiner, L. A., Garcia Pardo, A. & Margolies, M. N. (1979) *Biochemistry* **18**, 4068–4080.
10. Kahn, M., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Remaut, E. & Helinski, D. R. (1979) *Methods Enzymol.* **68**, 268–280.
11. Boyer, H. W. & Roulland-Dussoix, D. (1969) *J. Mol. Biol.* **41**, 459–472.
12. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9**, 309–321.
13. Cohen-Bazire, G., Sistrom, W. R. & Stanier, R. Y. (1957) *J. Cell. Comp. Physiol.* **49**, 25–68.
14. Clayton, R. K. & Smith, C. (1960) *Biochem. Biophys. Res. Commun.* **3**, 143–145.
15. Miyoshi, K., Huang, T. & Itakura, K. (1980) *Nucleic Acids Res.* **8**, 5491–5505.
16. Cohn, D. H., Ogden, R. C., Abelson, J. N., Baldwin, T. O., Nealson, K. H., Simon, M. I. & Mileham, A. J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 120–123.
17. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
18. Zieg, J., Hilmen, M. & Simon, M. (1978) *Cell* **15**, 237–244.
19. Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem.* **114**, 193–197.
20. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
22. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
23. Garoff, H. & Ansorge, W. (1981) *Anal. Biochem.* **115**, 450–457.
24. Staden, R. (1980) *Nucleic Acids Res.* **8**, 3673–3694.
25. Rosen, D., Feher, G. & Steiner, L. (1980) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **39**, 1801 (abstr.).
26. Michaelis, S. & Beckwith, J. (1982) *Annu. Rev. Microbiol.* **36**, 435–465.
27. Gibson, J., Stackebrandt, E., Zablen, L. B., Gupta, R. & Woese, C. R. (1979) *Curr. Microbiol.* **3**, 59–64.
28. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.
29. Mandel, M., Leadbetter, E. R., Pfennig, N. & Trüper, H. G. (1971) *Int. J. Syst. Bacteriol.* **21**, 222–230.
30. Dunn, R., McCoy, J., Simsek, M., Majumdar, A., Chang, S. H., RajBhandary, U. L. & Khorana, H. G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6744–6748.
31. Bachmann, R. C., Gillies, K. & Takemoto, J. Y. (1981) *Biochemistry* **20**, 4590–4596.
32. Valkirs, G. E. & Feher, G. (1982) *J. Cell Biol.* **95**, 179–188.
33. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
34. Nabedryk, E., Tiede, D. M., Dutton, P. L. & Breton, J. (1982) *Biochim. Biophys. Acta* **682**, 273–280.
35. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.* **47**, 45–148.
36. Marinetti, T. D., Okamura, M. Y. & Feher, G. (1979) *Biochemistry* **18**, 3126–3133.
37. Debus, R. J., Valkirs, G. E., Okamura, M. Y. & Feher, G. (1982) *Biochim. Biophys. Acta* **682**, 500–503.
38. Rosen, D., Okamura, M. Y., Abresch, E. C., Valkirs, G. E. & Feher, G. (1983) *Biochemistry* **22**, 335–341.
39. Okamura, M. Y. & Feher, G. (1983) *Biophys. J.* **41**, 122a (abstr.).
40. Doolittle, R. F. (1981) *Science* **214**, 149–159.
41. Zurawski, G., Bohnert, H. J., Whitfeld, P. R. & Bottomley, W. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7699–7703.
42. Arntzen, C. J., Darr, S. C., Mullet, J. E., Steinback, K. E. & Pfister, K. (1982) in *Function of Quinones in Energy Conserving Systems*, ed. Trumpower, B. (Academic, New York), pp. 443–452.
43. Engleman, D. M. & Steitz, T. A. (1981) *Cell* **23**, 411–422.
44. Michel, H. (1982) *J. Mol. Biol.* **158**, 567–572.
45. Okamura, M. Y. (1983) in *Biosynthesis of the Photosynthetic Apparatus: Molecular Biology, Development and Regulation*, eds. Hallick, R., Staehelin, L. A. & Thornber, J. P. (Liss, New York), in press.
46. Dalbadie-McFarland, G., Cohen, L. W., Riggs, A. D., Morin, C., Itakura, K. & Richards, J. H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6409–6413.
47. Chory, J. & Kaplan, S. (1982) *J. Biol. Chem.* **257**, 15110–15121.
48. Beatty, J. T. & Cohen, S. N. (1983) *J. Bacteriol.* **154**, 1440–1445.