Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 7303–7307, December 1984
Biochemistry

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

(photosynthesis/nucleotide sequence/membrane protein/sequence comparison)

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

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, August 1, 1984

ABSTRACT **The reaction center is an integral membrane protein that, together with several cofactors, mediates the primary photochemical events in bacterial photosynthesis. The amino-terminal sequences of the three subunits, L, M, and H, of the reaction center protein and the sequence of the structural gene encoding the M subunit have been reported previously. In the present study, we found that the 3′ end of the structural gene encoding the L subunit overlaps by eight bases the 5′ end of the gene encoding the M subunit. The primary structure of the L subunit has been determined from the nucleotide sequence of the gene and from analyses of the amino and carboxyl termini of the protein. The sequences of a number of tryptic and chymotryptic peptides were used to corroborate the nucleotide sequence. The L subunit was found to be composed of 281 amino acids (M_r 31,319) and to contain five hydrophobic segments. It is homologous to the M subunit and to a plant thylakoid protein referred to as the Q_B or M_r 32,000 protein.**

The primary energy conversion step of photosynthesis occurs in the reaction center (RC). The RC of the purple nonsulfur bacterium *Rhodopseudomonas sphaeroides* consists of three polypeptide chains, four bacteriochlorophylls, two bacteriopheophytins, two ubiquinones, and one iron atom, and has a molecular weight of ≈100,000 (reviewed in ref. 1). The subunits of the RC, designated L, M, and H, are hydrophobic and are found in the cytoplasmic membrane in a 1:1:1 stoichiometry (2).

The knowledge of amino acid sequences of the RC subunits is required for the determination of the three-dimensional structure by x-ray diffraction and for experiments designed to probe structure–function relationships. The complete amino acid sequences are being determined from the nucleotide sequences of the structural genes, together with amino- and carboxyl-terminal sequences of the polypeptides. The amino acid composition and sequences of a number of peptides were used to corroborate the nucleotide sequence. The genes encoding the L and M subunits were isolated with oligonucleotide probes corresponding to the amino-terminal sequences of the subunits (3). The determination of the sequence of the M subunit has been reported previously (4). The sequence of the L subunit has been determined by a similar method and is presented here.

EXPERIMENTAL PROCEDURES

Materials. Sources for restriction endonucleases, DNA-modifying enzymes, deoxynucleotides, and dideoxynucleotides have been described (4). The restriction enzyme *Pvu* II was from Boehringer Mannheim. M13 mp10, mp11, mp18, and mp19 replicative form DNA were from P-L Biochemicals. Sources for trypsin and chymotrypsin have been described (5). Carboxypeptidase A (treated with diisopropyl fluorophosphate) was from Sigma. Plasmid DNA used for cloning was prepared by an alkaline-NaDodSO$_4$ procedure (6). The DNA used to isolate the L and M subunit genes came from *R. sphaeroides* 2.4.1 (7). The protein used for peptide analysis came from *R. sphaeroides* R-26 (8).

Isolation of Genes. A ≈13-kilobase (kb) *Bam*HI restriction fragment of genomic DNA containing the L and M subunit genes was identified by hybridization with mixed-sequence oligonucleotide probes (4). The sequences of the probes were based on the amino-terminal sequences of the polypeptides (3).

Cloning. The general strategy for cloning was similar to that described (4). Subclones of the *Bam*HI fragment were obtained either from isolated fragments or from mixtures of fragments. Clones containing portions of the L subunit gene were identified either by restriction-site mapping or by hybridization with previously identified clones. The recombinant phage were constructed as follows: a 1.2-kb *Pst* I–*Pvu* II fragment was cloned into the *Pst* I and *Sma* I sites of M13 mp10; a 485-base-pair *Pvu* II–*Sal* I fragment was cloned into the *Sma* I and *Sal* I sites of M13 mp10 and M13 mp11; *Sau*3A fragments were cloned into the *Bam*HI site of M13 mp7; *Hpa* II and *Taq* I fragments were cloned into the *Acc* I site of M13 mp7; *Sau*3A–*Kpn* I fragments were cloned into the *Bam*HI and *Kpn* I sites of M13 mp18 and M13 mp19; and *Taq* I–*Kpn* I fragments were cloned into the *Acc* I and *Kpn* I sites of M13 mp18 and M13 mp19.

DNA Sequence Analysis. Sequence data were obtained from the recombinant M13 phage by the dideoxy method and were analyzed as described (4).

Preparation and Sequence Analysis of Peptides. The L subunit was isolated as described (3). For digestion with trypsin, a total of 0.1 mg of enzyme was added in three aliquots to 1.2 mg of protein in 0.8 ml of 0.1 M NH_4HCO_3 over 46 hr at 37°C; the peptides that were soluble in 0.1% trifluoroacetic acid (≈10% of the total material) were fractionated by reversed-phase high-pressure liquid chromatography as described (4). For digestion with chymotrypsin, a total of 0.06 mg of enzyme was added in two aliquots to 0.8 mg of protein in 0.5 ml of 0.05 M NH_4HCO_3 over 24 hr at 37°C; the peptides that were soluble in 0.2 ml of 0.1% trifluoroacetic acid (≈45% of the total material) were fractionated by reversed-phase high-pressure liquid chromatography (4). Selected fractions were subjected to amino acid analysis and sequence analysis by the manual dansyl Edman procedure or by automated Edman degradation [with 6 mg of Polybrene (9) added to the cup], as described (4). In the dansyl Edman procedure, glutamic acid and aspartic acid are not distinguished from the amides.

For carboxyl-terminal analysis of isolated peptides, ≈1 nmol of peptide in 30 μl of 0.1 M NH_4HCO_3 was treated with

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.

Abbreviations: RC, reaction center; kb, kilobase(s).

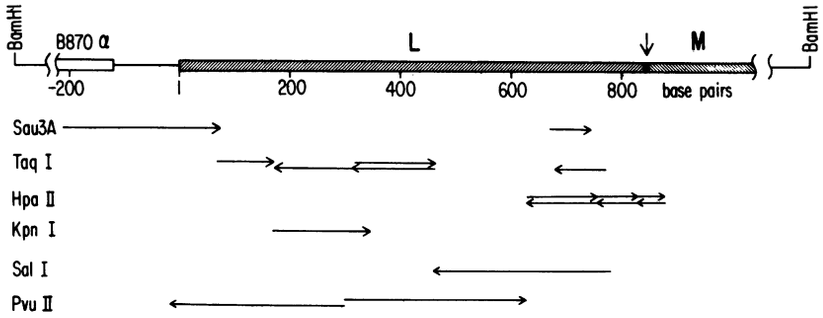

FIG. 1. Partial map of the ≈13-kb *Bam*HI fragment and fragments that were sequenced. The boxes indicate the positions of the genes encoding the L and M subunits of the RC and the α subunit of the B870 light-harvesting complex. The genes of the L and M subunits overlap by 8 bases (see vertical arow). The horizontal arrows indicate the extent and direction of sequence analysis of subcloned fragments.

5 μg of carboxypeptidase A for 10–60 min at room temperature; after lyophilization, the digests were applied directly to the amino acid analyzer.

Digestion of L Subunit with Carboxypeptidase A. Two samples of L (0.15 mg each) in 0.11 ml of 0.1 M NH_4HCO_3 were treated with 5 μg of carboxypeptidase A for 30 min or 6 hr at room temperature; after lyophilization, they were applied directly to the amino acid analyzer. A third sample, also 0.15 mg, was incubated with 5 μg of enzyme for 6 hr, lyophilized, and extracted with 10% trichloroacetic acid; after centrifugation, the supernatant was lyophilized, hydrolyzed for 24 hr in constant-boiling HCl and applied to the amino acid analyzer.

RESULTS

Nucleotide Sequences. The nucleotide sequence of the gene encoding the L subunit was determined from 17 recombinant

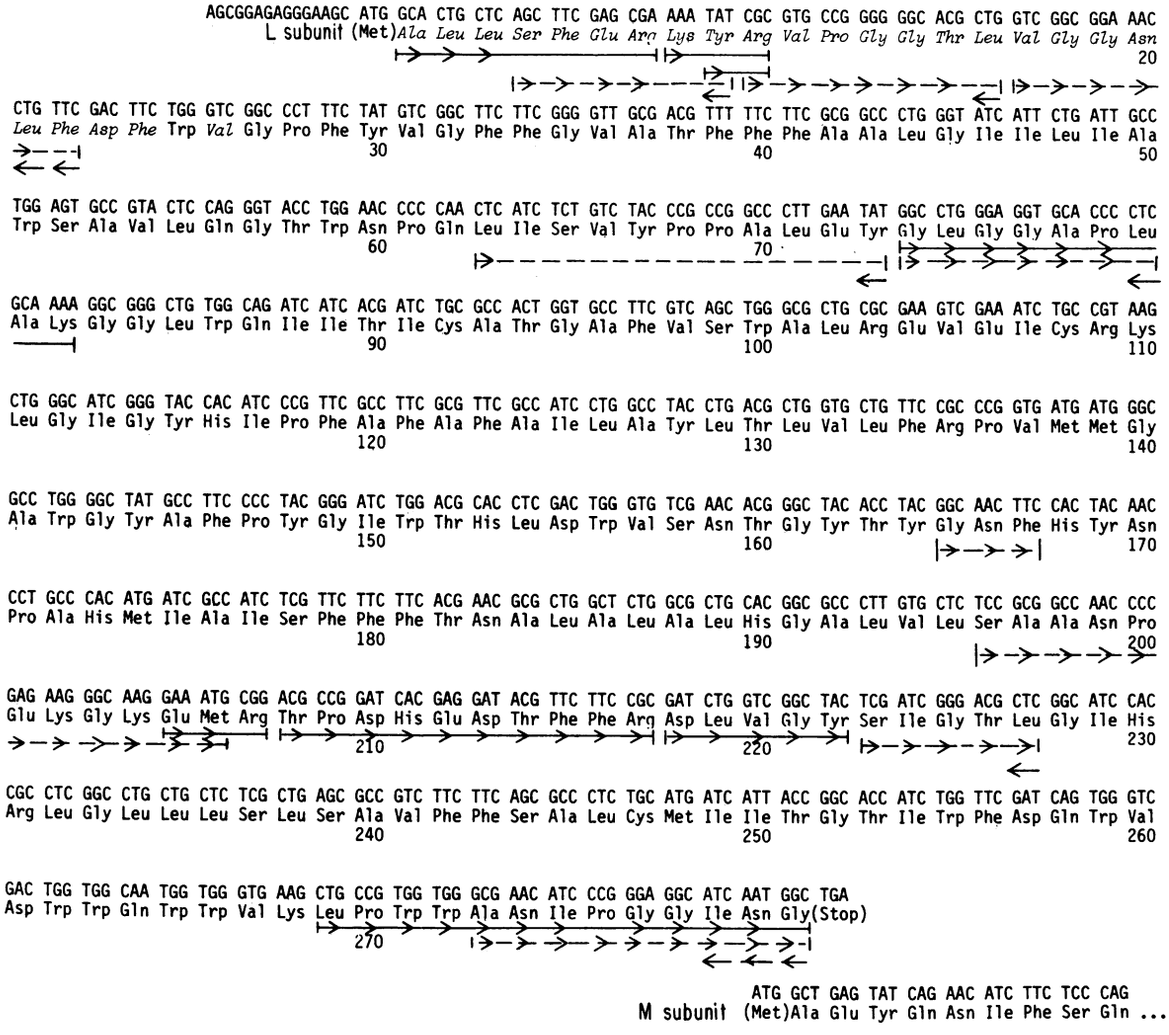

FIG. 2. Nucleotide sequence of the L subunit gene. The noncoding strand of the L subunit gene and the 5′ portion of the M subunit gene are shown. The numbers refer to residue positions of the derived amino acid sequence. Residues that confirm the previously determined amino-terminal sequence of the L subunit (3) are in italics. We had previously made a tentative assignment of histidine at position 25, but it was found in low yield (3); the nucleotide sequence indicates a tryptophan at this position. (|——|), Positions of isolated tryptic peptides; (|---|), positions of chymotryptic peptides; →, residues identified by Edman degradation; ←, residues identified by digestion.

phage (Fig. 1). These recombinant phage encompassed the entire gene on both the coding and noncoding strands. Two regions of the sequence (corresponding to residues 101–103 and 145–147) showed band compression on sequencing gels, and sequence data in these regions were obtained on only one strand. Amino acid sequence and composition data obtained from the protein and isolated peptides confirmed 32% of the sequence. The complete nucleotide and derived amino acid sequences are shown in Fig. 2. A nucleotide sequence of 84 bases corresponding to the carboxyl-terminal portion of the α subunit of the B870 light-harvesting complex (10) was identified beginning 204 bases upstream of the start of the gene encoding the L subunit (Fig. 1).

Isolation and Sequence Analysis of Tryptic and Chymotryptic Peptides. After digestion of the L subunit with trypsin, eight peptides, varying in length from 2 to 13 residues, were isolated in 5–60% yield; their sequences were determined completely or in part by automated Edman degradation or the manual dansyl Edman procedure. The results are summarized in Fig. 2. Two of the peptides contained neither lysine nor arginine; one of these peptides was in the carboxyl-terminal position and the other, obtained in 13% yield and corresponding to positions 218 to 222, was presumably the result of chymotrypsin-like cleavage. After digestion with chymotrypsin, 11 peptides, varying in length from 2 to 11 residues, were isolated in 15–45% yield; their sequences were determined completely or in part by automated Edman degradation or by the manual dansyl Edman procedure. The carboxyl-terminal residue(s) in some of the peptides was identified by digestion with carboxypeptidase A. The results, except for two dipeptides, Gly-Tyr and Thr-Tyr, are indicated in Fig. 2.

Carboxyl-Terminal Analysis of L Subunit. Amino acid analysis of carboxypeptidase A digests of the L subunit showed elution peaks in the position of serine, glycine, and isoleucine. However, the ratio of absorbance of the ninhydrin derivatives at 590 and 440 nm indicated that the material eluted in the position of serine was probably asparagine (see ref. 4). This identification was confirmed by showing that aspartic acid and not serine was found after hydrolysis of the released amino acids. The rates of release of asparagine and glycine were similar, but isoleucine was released more slowly. Thus, the carboxyl-terminal sequence of the L subunit is Ile-(Asn,Gly)-COOH. That asparagine precedes glycine was shown by sequence analysis of isolated tryptic and chymotryptic peptides (see Fig. 2).

DISCUSSION

The structural gene encoding the L subunit of the RC from *R. sphaeroides* has been shown previously, by hybridization with an oligonucleotide probe, to be on the *Bam*HI restriction fragment that contained the gene encoding the M subunit. The nucleotide sequence of this gene has now been determined, and its identity was confirmed by comparison of the derived amino acid sequence with sequences obtained by Edman degradation of the amino terminus of the L subunit and of selected peptides. The data show that the L subunit is composed of 281 amino acid residues corresponding to a M_r of 31,319. This value is considerably higher than that obtained by $NaDodSO_4$/polyacrylamide gel electrophoresis (M_r 21,000) (2). Anomalous behavior in $NaDodSO_4$/polyacrylamide gel electrophoresis was observed also with the M subunit [M_r 34,265 determined from the derived amino acid sequence (4) and M_r 24,000 determined by $NaDodSO_4$/polyacrylamide gel electrophoresis (2)] and other membrane proteins (11). The amino- and carboxyl-terminal ends of the polypeptide correspond to the residues immediately after a start codon and before a stop codon in the gene, indicating that there is no post-translational cleavage of the polypeptide other than the presumptive removal of a formyl-methionine residue. The codon usage for the L subunit gene shows a large preference for codons ending in G and C (85% of the codons) as was also observed in the M subunit gene (4).

The gene encoding the L subunit is located upstream of the gene encoding the M subunit. A similar arrangement of the genes for the L and M subunits of the RC was found in a closely related bacterium, *Rhodopseudomonas capsulata*, for which it was postulated that the two genes are transcribed as part of the same operon that also includes the genes for the B870 light-harvesting polypeptides (12). Our preliminary sequence data of the gene for the α subunit of the B870 protein indicate the same arrangement in *R. sphaeroides* (Fig. 1). The overlap of eight bases between the 3′ end of the L subunit gene and the 5′ start of the M subunit gene may have a role in assuring the 1:1 stoichiometry of the L and M subunits. For the *trpD* and *trpE* genes in the *trp* operon of *Escherichia coli*, a translational coupling was observed in which efficient translation of the second gene was shown to be dependent on the efficient translation of the first gene (13). An overlap of a single base in the stop and start codons of these two genes is potentially involved in this coupling.

The RC is an integral membrane protein that has been shown to span the cytoplasmic membrane (ref. 14; ref. 15 and references therein). Analysis by the method of Kyte and Doolittle (16) showed that the M subunit contains five segments that are rich in hydrophobic amino acids (4). A similar analysis indicates that the L subunit also contains five hydrophobic domains (Fig. 3). These segments may be membrane-spanning helices. This is consistent with data obtained

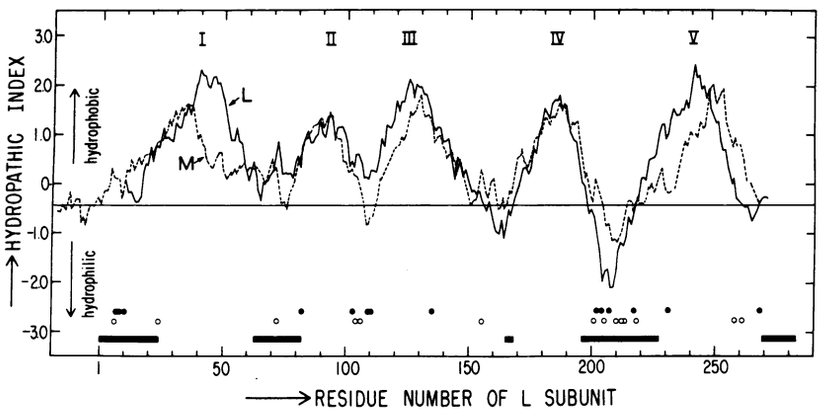

FIG. 3. Hydropathy profiles of the L and the M subunits. The average hydropathy value (16) of a moving window of 19 amino acids is plotted at the midpoint of the window. —, Profile of L subunit; ---, profile of M subunit, ● and ○, positions of basic (Lys, Arg) and acidic (Glu, Asp) residues in the L subunit, respectively; ■, positions of peptides isolated from the L subunit.

FIG. 4. Comparison of sequences of L and M subunits from *R. sphaeroides* and *R. capsulata* (21) and spinach (*Spinacia oleracea*) Q_B protein (22). Gaps were introduced into sequences to improve the homology. Boxed residues are identical in all four RC subunit sequences or in the *R. sphaeroides* L subunit and the spinach Q_B protein. Numbers at the end of the lines refer to positions of the residues in the L subunit.

by circular dichroism and polarized infrared spectroscopy, indicating that the RC contains a large amount of α-helical structure that is approximately perpendicular to the plane of the membrane (17). As expected, most of the charged residues occur in the more hydrophilic segments, and the soluble tryptic and chymotryptic peptides were also derived from these segments.

Proteolysis experiments with chromatophores from another photosynthetic bacterium, *Rhodospirillum rubrum*, show that the amino terminus of the L subunit is on the cytoplasmic side of the membrane (18). This leads to the prediction that two of the hydrophilic segments (between peaks II and III and peaks IV and V) are on the cytoplasmic side of the membrane and that the other two hydrophilic segments and the carboxyl terminus are on the periplasmic side. The components of the electron transfer process are arranged asymmetrically across the membrane, from cytochrome *c* on the periplasmic side (19) to the quinones on the cytoplasmic side (20). Thus, knowing the orientation of the protein in the membrane facilitates the localization of the cofactors by narrowing down the possible portions of the protein involved.

The L and M subunits show considerable homology in their amino acid sequences (Fig. 4). The introduction of three gaps in the sequence of the L subunit and two gaps in

Table 1. Homologies among L and M subunits and Q_B protein

	% identical residues*				
Polypeptide compared	*R. sphaeroides* L subunit	*R. capsulata* L subunit	*R. sphaeroides* M subunit	*R. capsulata* M subunit	Spinach Q_B protein
R. sphaeroides L subunit†	100	78	33	32	23
R. capsulata L subunit‡	78	100	29	32	25
R. sphaeroides M subunit§	33	29	100	77	24
R. capsulata M subunit‡	32	32	77	100	22
Spinach Q_B protein¶	23	25	24	22	100

*The number of identical residues in the aligned sequences (of two respective polypeptides) divided by the number of residues in the shorter sequence.
†This work.
‡Ref. 21.
§Ref. 4.
¶Ref. 22.

the sequence of the M subunit produces an alignment in which 33% of the amino acid residues are identical. In addition, the general shape of the hydropathy profiles is very similar (Fig. 3). Inspection of the most highly conserved regions of the sequences reveals that these regions contain a high proportion of glycine, proline, and histidine residues. These residues may have important structural roles in the RC. For example, the conserved proline residues are in hydrophilic regions or at the edges of hydrophobic regions and may be involved in the formation of turns outside the membrane. The conserved structural features may include binding sites for cofactors to the RC. The binding sites for the bacteriochlorophylls and bacteriopheophytins have not been determined, although they must be on the L or M subunits, since the H subunit can be removed without the loss of photochemical activity (23). Histidine residues have been shown to form ligands to bacteriochlorophyll in a water-soluble bacteriochlorophyll *a*-containing protein (24) and have been implicated in this role in membrane-bound light-harvesting proteins (25, 26). Histidine residues also have been suggested as ligands to the iron and the quinones in RCs (27, 28). The conservation of amino acid residues and the similarity between the hydropathy profiles suggest that the L and M subunits have a similar tertiary structure.

The homology between the L and M subunits that has been found in *R. capsulata* (21), as well as in *R. sphaeroides* (Fig. 4), suggests that the genes for these subunits have arisen by the duplication of an ancestral gene. There is a weaker but significant homology of the L and M sequences with a thylakoid protein referred to as the M_r 32,000 or Q_B protein (Fig. 4). The sequence of the gene encoding the thylakoid protein has been determined in several plant species (22, 29–31) and in a green alga (32). The Q_B protein binds triazine herbicides, which block electron transfer from the primary to the secondary electron acceptors in both plant chloroplasts and in bacterial RCs; it is thought to be involved in binding plastoquinone, which functions as the secondary electron acceptor of photosystem II (33). The hypothesis that two conserved regions containing histidine residues comprise the quinone binding site is discussed by Hearst and Sauer (34). Table 1 summarizes the relationships among the L subunit, the M subunit, and the Q_B protein in terms of the percentage of identical residues when optimally aligned. Identity of 20% or greater provides strong evidence for common ancestry (35). If one assumes that there has been a constant rate of change in the amino acid sequences, then the data show that the species divergence between *R. sphaeroides* and *R. capsulata* occurred after the gene duplication leading to the L and M subunits. That the L and M subunits differ by the same amounts in both species of bacteria supports the assumption of a constant rate of change. The data also indicate that the divergence of the ancestor of the chloroplast and the ancestor of *R. sphaeroides* and *R. capsulata* occurred before the gene duplication giving rise to the L and M subunits.

We thank E. Abresch for the preparation of the RCs, L.-P. Li for assistance in peptide preparation and analysis, R. F. Doolittle for the computer alignments of homologous protein sequences, and R. Theiler and D. C. Youvan for access to unpublished material. This work was supported by grants from the National Institutes of Health (GM 07313 and 13191), the National Science Foundation (PCM 82-02811), and the Office of Naval Research (N00014-83-K-0079).

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., Steiner, L. A. & Feher, G. (1974) *Biochemistry* **13,** 1394–1403.
3. Sutton, M. R., Rosen, D., Feher, G. & Steiner, L. A. (1982) *Biochemistry* **21,** 3842–3849.
4. Williams, J. C., Steiner, L. A., Ogden, R. C., Simon, M. I. & Feher, G. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6505–6509.
5. Steiner, L. A., Garcia Pardo, A. & Margolies, M. N. (1979) *Biochemistry* **18,** 4068–4080.
6. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
7. Cohen-Bazire, G., Sistrom, W. R. & Stanier, R. Y. (1957) *J. Cell. Comp. Physiol.* **49,** 25–68.
8. Clayton, R. K. & Smith, C. (1960) *Biochem. Biophys. Res. Commun.* **3,** 143–145.
9. Tarr, G. E., Beecher, J. F., Bell, M. & McKean, D. J. (1978) *Anal. Biochem.* **84,** 622–627.
10. Theiler, R., Suter, F., Wiemken, V. & Zuber, H. (1984) *Hoppe-Seyler's Z. Physiol. Chem.* **365,** 703–719.
11. Tanforol, C. & Reynolds, J. A. (1976) *Biochim. Biophys. Acta* **457,** 133–170.
12. Youvan, D. C., Alberti, M., Begusch, H., Bylina, E. J. & Hearst, J. E. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 189–192.
13. Oppenheim, D. S. & Yanofsky, C. (1980) *Genetics* **95,** 785–795.
14. Zürrer, H., Snozzi, M., Hanselmann, K. & Bachofen, R. (1977) *Biochim. Biophys. Acta* **460,** 273–279.
15. Valkirs, G. E. & Feher, G. (1982) *J. Cell Biol.* **95,** 179–188.
16. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
17. Nabedryk, E., Tiede, D. M., Dutton, P. L. & Breton, J. (1982) *Biochim. Biophys. Acta* **682,** 273–280.
18. Brunisholz, R. A., Wiemken, V., Suter, F., Bachofen, R. & Zuber, H. (1984) *Hoppe-Seyler's Z. Physiol. Chem.* **365,** 689–701.
19. Prince, R. C., Baccarini-Melandri, A., Hauska, G. A., Melandri, B. A. & Crofts, A. R. (1975) *Biochim. Biophys. Acta* **387,** 212–227.
20. Petty, K. M. & Dutton, P. L. (1976) *Arch. Biochem. Biophys.* **172,** 335–345.
21. Youvan, D. C., Bylina, E. J., Alberti, M., Begusch, H. & Hearst, J. E. (1984) *Cell* **37,** 949–957.
22. Zurawski, G., Bohnert, H. J., Whitfeld, P. R. & Bottomley, W. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 7699–7703.
23. Debus, R. J., Okamura, M. Y. & Feher, G. (1981) *Biophys. J.* **33,** 19a (abstr.).
24. Matthews, B. W., Fenna, R. E., Bolognesi, M. C., Schmid, M. F. & Olson, J. M. (1979) *J. Mol. Biol.* **131,** 259–285.
25. Brunisholz, R. A., Cuendet, P. A., Theiler, R. & Zuber, H. (1981) *FEBS Lett.* **129,** 150–154.
26. Theiler, R. & Zuber, H. (1984) *Hoppe-Seyler's Z. Physiol. Chem.* **365,** 721–729.
27. Bunker, G., Stern, E. A., Blankenship, R. E. & Parson, W. W. (1982) *Biophys. J.* **37,** 539–551.
28. Eisenberger, P., Okamura, M. Y. & Feher, G. (1982) *Biophys. J.* **37,** 523–538.
29. Spielmann, A. & Stutz, E. (1983) *Nucleic Acids Res.* **20,** 7157–7167.
30. Hirschberg, J. & McIntosh, L. (1983) *Science* **222,** 1346–1349.
31. Hirschberg, J., Bleecker, A., Kyle, D. J., McIntosh, L. & Arntzen, C. J. (1984) *Z. Naturforsch.* **39,** 412–420.
32. Erickson, J. M., Rahire, M., Bennoun, P., Delepelaire, P., Diner, B. & Rochaix, J.-D. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3617–3621.
33. 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.
34. Hearst, J. E. & Sauer, K. (1984) *Z. Naturforsch.* **39,** 421–424.
35. Doolittle, R. F. (1981) *Science* **214,** 149–159.