Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 189–192, January 1984
Genetics

Reaction center and light-harvesting I genes from *Rhodopseudomonas capsulata*

(photosynthesis/DNA sequence/enhanced fluorescence mutants/R-prime plasmid/genetic map)

DOUGLAS C. YOUVAN*†, MARIE ALBERTI*, HELMUT BEGUSCH*‡, EDWARD J. BYLINA*‡, AND JOHN E. HEARST*‡

*Division of Chemical Biodynamics, Lawrence Berkeley Laboratory, and ‡Department of Chemistry, University of California, Berkeley, CA 94720

Communicated by George C. Pimentel, September 2, 1983

ABSTRACT **Five structural genes coding for the reaction center (RC) L, M, and H subunits and the two light-harvesting (LH) I polypeptides, B870α and B870β, have been mapped on two restriction fragments from the R-prime plasmid pRPS404. It has been recently shown that enhanced near-infrared fluorescence mutants of *Rhodopseudomonas capsulata* typically lack RC or LH I polypeptides and that these lesions are marker-rescued by two restriction fragments from the R-prime plasmid: the 7.5-kilobase-pair *Eco*RI F fragment and the 4.75-kilobase-pair *Bam*HI C–*Eco*RI fragment. We have now determined the nucleotide sequence of two restriction fragments and have found that the *Bam*HI C–*Eco*RI B fragment carries the structural genes for the RC L and M subunits and both LH I polypeptides. Forty kilobase pairs away from this locus, the *Bam*HI F fragment (within the *Eco*RI F fragment) carries the RC H subunit. The structural genes on the *Bam*HI C–*Eco*RI B fragment are probably transcribed as part of a polycistronic mRNA. All of the structural genes begin with a consensus Shine–Dalgarno sequence and separate AUG start codons, indicating that the structural polypeptides are not cleaved from larger precursor polypeptides.**

The R-prime plasmid pRPS404 was generated by R-factor mobilization of the photosynthetic gene cluster from *Rhodopseudomonas capsulata* (1). This plasmid carries a 50-kilobase-pair (kb) insert of DNA indigenous to the *R. capsulata* chromosome that codes for most or all of the genes necessary for the differentiation of the oxidative-respiratory membrane into the photosynthetic membrane in response to low oxygen tension. The 50-kb gene cluster codes for biosynthetic enzymes required for carotenoid and bacteriochlorophyll biosynthesis and for the structural polypeptides of the reaction center (RC) and light-harvesting (LH) complexes. Approximately 15 genes have been mapped by a variety of techniques, including cotransduction of markers utilizing the gene transfer agent (2), marker-rescue using subclones of the R-prime (3), and R-prime site-directed transposon mutagenesis (4).

Recently, 20 enhanced fluorescence mutants were isolated from a tetracycline suicide procedure, the chromatophore membranes were characterized by $NaDodSO_4$/polyacrylamide gel electrophoresis, and the point mutations were mapped by marker-rescue using plasmid subclones of the R-prime (5). These studies showed that point mutations may cause enhanced LH II near-infrared fluorescence due to the loss or reduction of RC subunits or LH I polypeptides (or both). All of these point mutations map either to the *Bam*HI C–*Eco*RI B fragment or to the *Eco*RI F fragment from the R-prime plasmid pRPS404. The restriction map of the R-prime plasmid has been recently aligned with the genetic map for genes encoding carotenoid and bacteriochlorophyll biosynthetic enzymes (3).

Here we describe a preliminary analysis of the nucleotide sequence of the *Bam*HI C–*Eco*RI B fragment (bearing *rxcA*) and the *Bam*HI F fragment (contained within the *Eco*RI F fragment and bearing *rxcB*). Recently, the LH I polypeptides have been subjected to sequence analysis (M. H. Tadros, H. Zuber, and G. Drews, personal communication) and the RC L, M, and H subunit amino-terminal sequences are available from *R. sphaeroides* (6). These data have been used to scan the nucleotide sequence data to find the corresponding structural genes in the *Bam*HI F and *Bam*HI C–*Eco*RI B fragments. In this preliminary communication we report the complete nucleotide sequence of the structural genes for B870α and B870β [previously designated as the 12- and 8-kilodalton (kDa) LH I polypeptides (7), respectively]. Here we report only the position and deduced amino-terminal sequences of the RC subunits because fragmentary internal polypeptide data (not currently available) are required to verify the deduced RC polypeptide sequences.

MATERIALS AND METHODS

DNA Sequence and Computer Analyses. The nucleotide sequence of the *Bam*HI F and *Bam*HI C–*Eco*RI B fragments from the photosynthetic gene insert on the R-prime plasmid pRPS404 was determined by the shotgun dideoxy method (8). *Taq* I, *Sau*3A, and *Hpa* II fragments from the pBR322 derivatives carrying these fragments (pRPSE2 and pRPSEB2, respectively) were ligated into M13mp8 and M13mp9 vectors (Bethesda Research Laboratories) and primed with a 15-nucleotide primer (P-L Biochemicals). A consensus sequence was obtained by comparing the results from an average number of six individual sequence determinations from a total of ≈300 recombinant templates, each having an average of three restriction fragment inserts. Approximately 50 kb of sequence were read to obtain a consensus on the *Bam*HI F and *Bam*HI C–*Eco*RI B fragments, together totaling 9 kb. These nucleotide sequences were scanned by computer using a high-speed homology program (9) with RC amino-terminal sequence data from *Rhodopseudomonas sphaeroides* (6) and LH I polypeptide data from *R. capsulata* (M. H. Tadros, H. Zuber, and G. Drews, personal communication). We are regarding the current nucleotide sequence as tentative until internal polypeptide data from the *R. capsulata* RC proteins are obtained and used to verify that the computer-predicted polypeptide sequences are in the proper reading frame throughout their entire length.

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; LH, light-harvesting; kb, kilobase pair(s); kDa, kilodalton.

†Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.

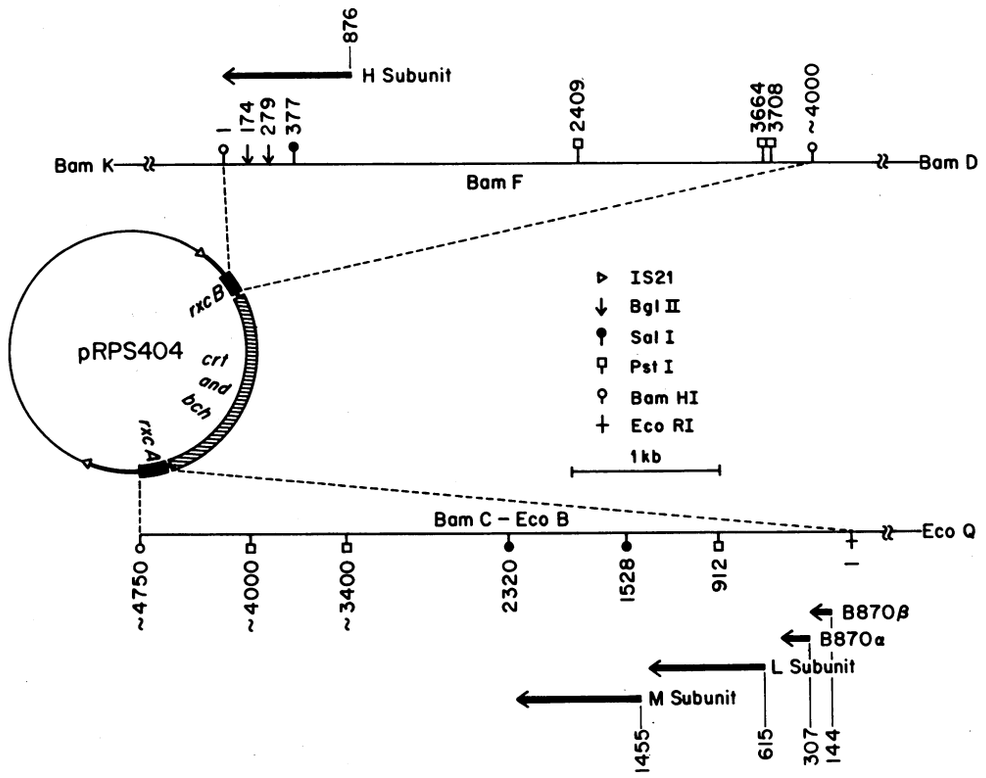

FIG. 1. Organization of the photosynthetic apparatus genes coding for the RC and LH I (B870) structural polypeptides. Preliminary sequence data from the *Bam*HI F and *Bam*HI C–*Eco*RI B restriction fragments were analyzed by computer to find nucleotide sequences homologous to the LH I and RC structural polypeptides. The H subunit is coded by the *Bam*HI F fragment near the junction with *Bam*HI K. The RC L and M subunits and the LH I polypeptides were found by computer analysis on the *Bam*HI C–*Eco*RI B fragment. B870β and B870α are also known by their apparent molecular masses: 8- and 12-kDa LH I polypeptides, respectively.

RESULTS AND DISCUSSION

Fig. 1 illustrates the organization of the photosynthetic apparatus structural genes with respect to restriction fragments from the R-prime plasmid pRPS404. On the *Bam*HI C–*Eco*RI B fragment, the gene organization in the direction of transcription is B870β and B870α (also known as the 8-and 12-kDa LH I polypeptides, respectively), followed by the RC L and M subunits. All four genes begin with an AUG codon and are probably transcribed as part of the same operon. Transcription is away from the two closest markers, *bchA* and *bchC*. It is possible that the promoter for this putative operon lies close to the *Eco*RI Q and *Bam*HI C–*Eco*RI B junction, unless other genes in the operon lie upstream from the LH I structural genes. We have identified the structural gene for the H subunit of the RC on the *Bam*HI F fragment. This gene begins with an AUG codon and transcription is oriented toward the *Bam*HI K fragment.

All five structural genes are preceded by an average space of six nucleotides and the consensus Shine–Dalgarno sequence (A or c)G-G-A-G(G or a). This consensus is complementary to the 3′ terminal sequence from 16S rRNA in both *R. capsulata* and *Escherichia coli* (10, 11), as shown in Fig. 2. Comparison of the nucleotide and mature peptide sequences indicates the absence of post-translational processing of precursor polypeptides or amino-terminal leader sequences.

The nucleotide sequence of the LH I genes is given in Fig. 3. The deduced polypeptide sequence is in complete agreement with the polypeptide sequence determined by automated Edman degradation.

The amino-terminal polypeptide sequence of the RC subunits is strongly conserved between *R. capsulata* and *R. sphaeroides*. Fig. 4 shows the DNA sequence and the deduced amino-terminal polypeptide sequences for *R. capsulata* and compares this with the amino-terminal data from *R. sphaeroides*, which were determined by Edman degradation (6). Seventy-five percent of the amino-terminal residues in the three RC subunits are conserved. Because these data are compared across species, it can only be suggested that the methionine residues on the deduced amino termini of the L and M subunits are cleaved post-translationally.

Enhanced fluorescence mutants have pleiotropic loss or

B870 β	5′	AUC CGGAGG UUGUUAUG 3′
B870 α		CUG AGGAGA AACUGAAAAUG
RC L Subunit		CAG CGGAGA CAGGGCAUG
RC M Subunit		GGC AGGAGG CAUCAAUG
RC H Subunit		CAA AGGAGG ACCAACAUG
mRNA Consensus	5′	$^{C}_{A}$GGAG$^{A}_{G}$ 3′
E. coli 16S rRNA		HO^{AU} UCCUCC ACUA...5′
R. caps. 16S rRNA		HO^{UCUU} UCCUCC ACUA...5′

FIG. 2. Consensus Shine–Dalgarno sequences predicted from the DNA sequence. The deduced mRNA sequences coding for all five RC and LH I structural genes include an AUG translational start codon preceded by a consensus Shine–Dalgarno sequence. This sequence is complementary to a sequence within the 3′ terminus of 16S rRNA in *E. coli* and *R. capsulata* (*R. caps.*).

```
        140         150         160         170         180         190
         *           *           *           *           *           *
CCG GAG GTT GTT ATG GCT GAT AAG AAC GAC CTG AGC TTC ACA GGT CTT ACC GAC GAG CAA
            Met Ala Asp Lys Asn Asp Leu Ser Phe Thr Gly Leu Thr Asp Glu Gln

        200         210         220         230         240         250
         *           *           *           *           *           *
GCG CAA GAA CTG CAT GCC GTC TAC ATG AGC GGG CTG TCG GCG TTC ATC GCC GTC GCG GTG
Ala Gln Glu Leu His Ala Val Tyr Met Ser Gly Leu Ser Ala Phe Ile Ala Val Ala Val

        260         270         280         290         300         310
         *           *           *           *           *           *
CTC GCC CAT CTG GCG GTC ATG ATC TGG CGT CCG TGG TTC TGA GGAGAAACTGAAA   ATG TCC
Leu Ala His Leu Ala Val Met Ile Trp Arg Pro Trp Phe ---                 Met Ser

        320         330         340         350         360         370
         *           *           *           *           *           *
AAG TTC TAC AAA ATC TGG CTC GTT TTC GAT CCC CGT CGC GTG TTC GTG GCC CAG GGC GTG
Lys Phe Tyr Lys Ile Trp Leu Val Phe Asp Pro Arg Arg Val Phe Val Ala Gln Gly Val

        380         390         400         410         420         430
         *           *           *           *           *           *
TTC CTG TTC CTG CTC GCG GTG CTG ATC CAC CTG ATC CTG CTC TCG ACC CCC GCT TTC AAC
Phe Leu Phe Leu Leu Ala Val Leu Ile His Leu Ile Leu Leu Ser Thr Pro Ala Phe Asn

        440         450         460         470         480         490
         *           *           *           *           *           *
TGG CTG ACC GTT GCC ACC GCC AAG CAT GGC TAC GTG GCT GCT GCC CAG TAA GCG TTA CTG
Trp Leu Thr Val Ala Thr Ala Lys His Gly Tyr Val Ala Ala Ala Gln ---
```

FIG. 3. Nucleotide sequence analysis and deduced polypeptide sequences of the LH I structure polypeptides. The sequence begins 132 nucleotides from the *Eco*RI end of the *Bam*HI C–*Eco*RI B fragment and proceeds toward the *Bam*HI end. The structural gene coding for B870β (also known as the LH I 8-kDa polypeptide) begins at sequence position 144 and ends at position 293. The structural gene coding for B870α (also known as the LH I 12-kDa polypeptide) begins at sequence position 307 and ends at position 483. The amino acid sequences are derived from the photochemically active mutant strain Ala$^+$ of *R. capsulata* GSC938 (German strain collection, Göttingen) = 37b4 (12). The nucleic acid sequences reported here are on the mobilized plasmid pBLM2, which is derived from *R. capsulata* strain SB1003 (St. Louis) (1).

```
H Subunit:

DNA     ATG GTC GGT GTC AAT TTC TTC GGA GAC TTC GAT CTG GCC AGT
R.c.    Met Val Gly Val Asn Phe Phe Gly Asp Phe Asp Leu Ala Ser
R.s.    --- --- --- --- Thr Ala --- --- Asn --- --- --- --- ---

DNA     CTT GCA ATC TGG TCG TTC TGG GCG TTT CTC GCC TAT CTG ATC
R.c.    Leu Ala Ile Trp Ser Phe Trp Ala Phe Leu Ala Tyr Leu Ile
R.s.    --- --- --- Tyr --- --- --- Ile --- --- ---  X  --- ---

M Subunit:

DNA     ATG GCT GAG TAT CAG AAC TTC TTT AAC CAG GTT CAG GTC GCC GGC
R.c.    Met Ala Glu Tyr Gln Asn Phe Phe Asn Gln Val Gln Val Ala Gly
R.s.     X  --- --- --- --- --- Ile --- Ser --- --- --- --- Arg ---

DNA     GCG CCC GAA ATG GGC CTC AAG GAA GAC GTC GAC ACC TTC GAA
R.c.    Ala Pro Glu Met Gly Leu Lys Glu Asp Val Asp Thr Phe Glu
R.s.    Pro Ala Asp Leu --- Met Thr --- --- --- Asn Leu Ala Asn

L Subunit:

DNA     ATG GCT TTG CTC AGC TTC GAA CGA AAA TAC CGT GTG CCG GGC
R.c.    Met Ala Leu Leu Ser Phe Glu Arg Lys Tyr Arg Val Pro Gly
R.s.     X  --- --- --- --- --- --- --- --- --- --- --- --- ---

DNA     GGC ACC TTG ATT GGC GGG AGC CTT TTC GAC TTC TGG GTC
R.c.    Gly Thr Leu Ile Gly Gly Ser Leu Phe Asp Phe Trp Val
R.s.    --- --- --- Val --- --- Asn --- --- --- ---(His)---
```

FIG. 4. Interspecies comparison of the RC polypeptide amino termini between *R. capsulata* (R.c.) and *R. sphaeroides* (R.s.). DNA sequence data are given for the L, M, and H subunits of the RC structural genes from *R. capsulata*. The deduced amino-terminal polypeptide data are listed immediately below each DNA sequence for *R. capsulata*. This is to be compared with the bottom line in each set of data, where a comparison is made with the amino termini of subunits from *R. sphaeroides* as determined by Edman degradation (6). Blanks indicate that the residue is conserved between the two species and differences are indicated by the specification of the *R. sphaeroides* residue (X indicates an unspecified residue or deletion in the latter).

reduction of all three RC subunits in a constant stoichiometry (5). Because the L and M subunits are transcriptionally separated from the H subunit, the cause of this pleiotropy cannot be entirely due to transcriptional polarity. Rather, the pleiotropy may be due to the lack of assembly and incorporation of the remaining subunits into the membrane. An alternative explanation would have the H subunit controlling the translation or degradation of the mRNA that codes for the L and M subunits. Clearly, more experiments are essential for an understanding of this pleiotropy.

We thank Dr. Herbert Zuber for communication of the LH I polypeptide sequences prior to their publication, Dr. Barry Marrs for providing plasmid subclones of the R-prime plasmid, Kris Zsebo and Drs. Edgar David and Gregg Clark for transposon and restriction site information, Mr. John Hubbard for assistance in computer analysis of the sequence data, and Ms. Eva Penzes and Mr. Jeremy Gelbwaks for technical assistance. This work was supported by the office of Energy Research, Office of Basic Energy Sciences, Biological Energy Research Division of the U.S. Department of Energy, under Contract DE-ACO3-76SF00098, and by the National Institutes of Health, under Grant GM30786.

1. Marrs, B. L. (1981) *J. Bacteriol.* **146,** 1003–1012.
2. Scolnik, P. A., Walker, M. A. & Marrs, B. L. (1980) *J. Biol. Chem.* **255,** 2427–2432.
3. Taylor, D. P., Cohen, S. N., Clark, W. G. & Marrs, B. L. (1983) *J. Bacteriol.* **154,** 580–590.
4. Youvan, D. C., Elder, J. T., Sandlin, D. E., Zsebo, K., Alder, D. P., Panopoulos, N. J., Marrs, B. L. & Hearst, J. E. (1982) *J. Mol. Biol.* **162,** 17–41.
5. Youvan, D. C., Hearst, J. E. & Marrs, B. L. (1983) *J. Bacteriol.* **154,** 748–755.
6. Sutton, M. R., Rosen, D., Feher, G. & Steiner, L. A. (1982) *Biochemistry* **21,** 3842–3849.
7. Drews, G. (1978) *Curr. Top. Bioenerg.* **8,** 161–207.
8. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
9. Pustell, J. & Kafatos, F. C. (1982) *Nucleic Acids Res.* **10,** 4765–4782.
10. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 1342–1346.
11. Gibson, J., Stackebrandt, E., Zablen, L. B., Gupta, R. & Woese, C. R. (1979) *Curr. Microbiol.* **3,** 59–64.
12. Feick, R. & Drews, G. (1978) *Biochim. Biophys. Acta* **501,** 499–513.