Light-harvesting II (B800–B850 complex) structural genes from Rhodopseudomonas capsulata

(Douglas C. Youvan*† and Sajida Ismail*)

*Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; and †Exxon Research and Engineering Company, Corporate Research Science Laboratory, Clinton Township Route 22 East, Annandale, NJ 08801

Communicated by George Feher, August 31, 1984

ABSTRACT The light-harvesting II (LHII) structural genes coding for the (B800–B850 complex) β- and α-polypeptides have been cloned and the nucleotide and deduced polypeptide sequences have been determined. This completes the sequencing of all seven structural genes coding for the structural polypeptides of the photosynthetic apparatus that bind the pigments and cofactors participating in the primary light reactions of photosynthesis. Unlike the structural coding genes for the reaction center L, M, and H subunits and the light-harvesting I (LHI) (B870 complex) structural polypeptides, the LHII structural genes are not within the 46-kilobase photosynthetic gene cluster carried by the R-prime plasmid pRPS404. Identical organization of the β and α structural genes for both LHI and LHII and sequence homologies between the two β-polypeptides and between the two α-polypeptides suggests that both complexes arose by gene duplication from a single ancestral light-harvesting complex and that the putative bacteriochlorophyll binding sequence Ala-X-X-His has been absolutely conserved.

The function of the light-harvesting (LH) complex is to absorb visible and near-infrared radiation and to transfer this energy with high efficiency to the reaction center. In the reaction center the light energy is transduced into a charge separated state that, in conjunction with the b5c-oxidoreductase, generates a transmembrane proton gradient. The potential energy of this gradient is utilized by the coupling factor ATPase in the synthesis of ATP. This scheme is essentially the chemiosmotic theory of Mitchell (1) as it pertains to purple nonsulfur photosynthetic bacteria, such as Rhodopseudomonas capsulata.

Bacteriochlorophyll molecules associated with integral membrane proteins are responsible for both light absorption and the primary processes involved in charge separation. The design of the photosynthetic bacteria has been to place a large number of bacteriochlorophyll molecules and associated carotenoid pigments in the light-harvesting antennae and to structurally coordinate this large ensemble of absorbers with relatively few bacteriochlorophyll molecules that are participating in the photochemistry of the reaction center (2). From future x-ray crystallographic studies, we expect to see that the light-harvesting proteins are intimately involved in the precise alignment and spacing of the bacteriochlorophyll molecules in order to facilitate efficient transfer of energy to the reaction center trap.

Our approach to the structure–function problem of the photosynthetic apparatus has been to isolate, clone, and sequence the structural genes coding for the reaction center L, M, and H subunits and the LHI (B870 complex) α-, and β-polypeptides (3–5). In this communication, we describe the cloning, sequencing, and analysis of the LHII antenna (B800–B850 complex) α- and β-polypeptides from R. capsulata. Comparison of the deduced polypeptide sequences and gene organization from both light-harvesting complexes suggests that the antenna arose by a gene duplication and that the sequence Ala-X-X-His, which may be involved in binding bacteriochlorophyll (6, 7), has been absolutely conserved.

MATERIALS AND METHODS

Oligonucleotide Primers and Probes. Deoxyribonucleotides were synthesized on an Applied Biosystems 380-A synthesizer using phosphite triester chemistry and were purified by electrophoresis on polyacrylamide gels (8). Two mixed probes complementary to the amino termini of the β- and α-subunits of LHII were synthesized, each having a complexity of 64 (see Fig. 1). Five oligonucleotides were synthesized to serve as primers in dideoxy-sequencing of the Smal B fragment (designated SP1*, SP1+, SP2*, SP2+, and the M13 universal primer; see Fig. 3). Oligonucleotide probes were 5'-end-labeled as described (8).

Southern Hybrids. Agarose gels (0.8%) were run in buffer A (40 mM Tris acetate, pH 8/1 mM EDTA/0.5 μg of ethidium bromide per ml) at 2 V/cm until the bromophenol blue dye had migrated 16 cm. Gels were shaken in 0.2 M NaOH/0.6 M NaCl for 45 min and then neutralized with 0.5 M Tris–HCl, pH 7.5/1.5 M NaCl/10 mM EDTA for 45 min. Gels were transferred to Schleicher & Schuell BA85 nitrocellulose paper using a wick, in 10× NaCl/Cit buffer overnight (1× NaCl/Cit buffer was 0.15 M NaCl/0.015 M sodium citrate/0.5 mM EDTA, pH 7.2 by HCl). Nitrocellulose was baked in vacuo for 2 hr at 75°C.

For phage or plaque hybrids, filters (Schleicher & Schuell BA85) were adsorbed to plates for 20 min, denatured with 0.5 M NaOH for 5 min, and then neutralized with 0.5 M Tris–HCl, pH 7.5/1.5 M NaCl for 5 min. Filters were washed with 2× NaCl/Cit buffer for 5 min and dried and baked as described above. Phage and colony filters were prehybridized for 2 hr at 65°C in 6× NaCl/Cit buffer containing 5× Denhardt's solution and 0.5% NaDodSO4 (1× Denhardt's solution = 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone). Usually, 2 pmol of 32P-labeled oligomer was used per hybridization in a total volume of 15 ml containing 6× buffer B (1× buffer B = 0.15 M NaCl/0.015 M Tris–HCl, pH 7.5/0.001 M EDTA). Filters were hybridized overnight at 37°C and then washed 3 times with 200 ml of 6× NaCl/Cit buffer for 15 min at room temperature. Filters were dried and exposed on Kodak XAR5 film with an intensifying screen at −80°C, typically for 1 day. These hybridization methods are similar to those of Wallace et al. (9).

Plasmid Constructions. λ phage, strain GT7-arab carrying

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.
EcoRI fragments from the chromosome of *R. capsulata* (courtesy of Pablo Scolnik; see ref. 10), bearing the LHII genes were identified by the plaque-hybridization method described above. DNA from positively hybridizing phage was prepared as described (11) and digested with EcoRI. A 5.75-kilobase (kb) EcoRI fragment was prepared by hot phenol extraction from a low melt agarose gel and ligated into the EcoRI site of pBR322. This plasmid (pRPSLH2) was mapped with restriction enzymes by using multiple digest information, partial digests of end-labeled fragments, and by Southern hybridization. The Sma I B fragment was mapped and found to carry DNA hybridizing to the \(\beta\) and \(\alpha\) probes. The Sma I B fragment was purified by hot phenol extraction from low melt agarose gel electrophoresis and ligated into the Sma I site of M13mp9. Sma I B inserts in both orientations were obtained and single-stranded phage DNA was prepared for dideoxy-sequencing.

A kanamycin-resistant derivative of pRPSLH2 was obtained by ligating a 1850-base-pair (bp) fragment specifying kanamycin resistance into the BamHI site of pRPSLH2. The 1850-bp kanamycin fragment was constructed from Tn5 by partial digests of Tn5 from strain DB1706-3 into the BamHI site of pBR322. Ligation of the Sau3A site at sequence position 1153 from the left end of the Tn5 map (12–14) into the BamHI site of pBR322 regenerated a BamHI site. The kanamycin-resistant derivative is designated pRPSLH2KAN. HB101 was used in all constructions.

**Dideoxy-sequencing.** One of the two orientations of the Sma I B fragment in M13mp9 was found to hybridize (dot blots) to the \(\beta\) and \(\alpha\) probes. This orientation, designated the forward orientation, was sequenced starting with the M13 universal primer, through the Sma I site up to the amino terminal of the \(\beta\)-subunit gene (assigning nucleotide number 1), using SP1'- plus SP2'- primers. SP1' was used to sequence back to the Sma I site on the reverse template. This procedure was repeated after reading the SP1'-generated sequence ending in the \(\beta\)-subunit. SP2' extended the sequence past the end of the coding sequences on the forward template, and SP2' was used to obtain the second strand on the reverse template through much of the coding region. Dideoxy-sequencing reactions were essentially that of Sanger et al. (15). The Cold Spring Harbor Laboratory programs (16) were used to compile, map, and translate the DNA sequence.

**Conjugal Matings.** Plasmid pRPSLH2KAN was transferred from HB101 to the LHII mutant MW442 by a triparental conjugal mating using an R-factor-ColE1 helper plasmid, pDPT41, in an HB101 background (17). One volume each of HB101 (pRPSLH2KAN) and HB101 (pDPT41), both strains growing exponentially in L broth, were mixed with 8 vol of an exponentially growing photosynthetic culture of MW442 in PYE medium and 25 \(\mu\)l was spotted and dried on a PYE plate. The spots were incubated at 35°C for 24 hr, resuspended, and washed 2 times in 1 ml of RCV medium, and 25 \(\mu\)l was plated on RCV plates containing 25 \(\mu\)g of kanamycin per ml. These plates were incubated at 35°C for 3 days and an average of 300 colonies per plate was obtained; no colonies were detected on control plates on which one of the three parents was omitted. Approximately one-half of these colonies had a darker pigmentation, characteristic of wild type. These colonies were picked and repurified on PYE plates containing 25 \(\mu\)g of kanamycin per ml.

**Spectroscopy.** Absorption spectroscopy on chromatophore membranes was performed as described (3). Spectra from a Perkin Elmer 330 spectrophotometer were normalized to the 375-nm absorption band, using a Model 3600 data station. Continuous scans were taken from 1150 nm to 350 nm with detector change set at 900 nm. Constant gain (setting 2) was used in the near infrared and a constant slit width (2 nm) was used in the visible range.

**RESULTS**

**Cloning the LHII Genes.** Two 17-nucleotide oligonucleotide probes were synthesized with a degeneracy of 64, based on the amino-terminal sequences (18, 19) of the \(\alpha\) and \(\beta\)-subunits of LHII (Fig. 1). In Southern hybridizations with total digests of the bacterial chromosome the \(\beta\) probe hybridized with only one band in any given digest; the \(\alpha\) probe was less specific. However, the \(\alpha\) bands always included the single \(\beta\) band. We interpreted these data as indicating that the \(\alpha\) and \(\beta\) genes were closely linked and not separated by any of the seven-hexamer recognition restriction enzymes tested. Gel bands ranged from 4 kb for *Pvu* I to 20 kb for BamHI digests. An EcoRI band of \(\approx 6\) kb was detected in the genomic Southern hybridizations. Since this falls within the size range of EcoRI fragments carried by the \(\lambda\) genomic library of Scolnik (10), we proceeded with plaque hybridization, using this library and the \(\beta\) probe.

Positively hybridizing \(\lambda\) plaques were found with a frequency of \(\approx 1\) in 5000 plaques. DNA from several phage was prepared and, by comparison to the genomic Southern hybridization, was found to carry the 6-kb EcoRI fragment, which hybridized to the \(\alpha\) and \(\beta\) probes. The 6-kb fragment was subcloned into the EcoRI site of pBR322; this plasmid (pRPSLH2) was mapped by restriction analysis using double and triple digests, partial digests with end-labeled inserts, and by Southern hybridizations. These results combined with data from DNA sequencing are shown in Fig. 2.

**LHII Gene Organization.** Using Southern hybridizations, the Sma I B fragment of pRPSLH2 was found to carry the \(\alpha\) and \(\beta\) structural genes. This fragment was subcloned into the Sma I site of M13mp9, and the nucleotide sequence was determined from the left Sma I B site through the structural genes (600 bp). The asymmetric position of the *Nru I* site within the Sma I B fragment was used to confirm the orientation of the \(\alpha\) and \(\beta\) structural genes within the EcoRI insert of pRPSLH2.

The nucleotide sequence of the first 600 bp of the Sma I B fragment from pRPSLH2 is given in Fig. 3. The nucleotide sequence was translated in all six reading frames, and two polypeptides were identified that are in complete agreement with the known polypeptide sequence of \(\alpha\) and the partial sequence of \(\beta\). Alternative reading frames in the carboxyterminal end of the \(\beta\)-subunit utilize forbidden codons (5). The ATG start codon of \(\beta\) begins at sequence position 198 (assigning nucleotide number 1 to the first guanine in the left Sma I B site) and terminates at position 347. There is a non-

\[
\begin{align*}
\text{5'} & \quad \text{A} \quad \text{G} \quad \text{C} \quad \text{G} \quad \text{A} \quad \text{T} \\
\beta & 18 \quad \text{K} \quad \text{D} \quad \text{C} \quad \text{C} \quad \text{A} \\
\text{3'} & \quad \text{C} \quad \text{A} \quad \text{C} \quad \text{G} \quad \text{T} \quad \text{A} \\
\end{align*}
\]

\[
\begin{align*}
\text{5'} & \quad \text{A} \quad \text{T} \quad \text{G} \quad \text{A} \quad \text{C} \quad \text{G} \\
\alpha & 8 \quad \text{K} \quad \text{D} \quad \text{C} \quad \text{C} \quad \text{A} \\
\text{3'} & \quad \text{G} \quad \text{T} \quad \text{A} \quad \text{C} \quad \text{G} \quad \text{A} \\
\end{align*}
\]

**FIG. 1.** Oligonucleotide probes based on amino-terminal sequences of the \(\beta\) and \(\alpha\)-polypeptides from the LHII antenna. These 17-nucleotide oligomers have a degeneracy of 64 and were used to probe the \(\lambda\) genomic library of *R. capsulata* and to map the \(\beta\) and \(\alpha\) structural genes on pRPSLH2. The 8-kDa and 10-kDa labeling refer to the apparent molecular sizes of the subunits that were used in the older literature. Amino acids are represented by standard one-letter abbreviations.
coding gap of 21 nucleotides, and the \( \alpha \) gene begins at 368 and terminates at 550. Both genes are preceded by the consensus Shine-Dalgarno (20) sequences \((\alpha, A-G-G-A-G; \beta, G-G-A-G-G)\), which are identical to the sequences used in the \( \text{LHI} \) genes (4, 5).

**pRPS404 and the \( \text{LHII} \) Genes.** Southern hybridizations were performed (Fig. 4) to compare the cloned \( \text{LHII} \) genes on pRPSLH2 with the corresponding chromosomal copy and with pRPS404. R-prime plasmid pRPS404 carries most of the genes necessary for the expression of the photosynthetic apparatus (21), including the structural genes for the reaction center and \( \text{LHII} \) (3–5). Hybridization of the \( \beta \) probe to \( \text{EcoRI}, \text{EcoRI} \) and \( \text{PvuI} \), and \( \text{SmaI} \) digests of plasmid, chromosomal and R-prime indicates that no rearrangements have occurred during the cloning of the \( \text{LHII} \) genes (identical hybridization to the chromosomal and pRPSLH2 lanes). Furthermore, the absence of pRPS404 bands at the position of the pRPSLH2 bands indicates that the \( \text{LHII} \) structural genes are not within the 46-kb photosynthetic gene cluster carried by pRPS404. Nor are the structural genes within one \( \text{EcoRI} \) fragment from the ends of pRPS404, because the R-prime and chromosomal junction bands A and B are 3 and 10 kb (22), respectively, whereas the \( \text{LHII} \) EcoRI band is 5.75 kb. This places the \( \text{LHII} \) structural genes at least 11 kb away from the reaction center \( \text{L} \) and \( \text{M} \) structural genes, or at least 7 kb away from the \( \text{H} \) gene of the reaction center.

**pRPSLH2KAN Complementation of MW442.** MW442 lacks only the \( \text{LHII} \) complex (23). This mutant has a characteristic absorption spectrum where the light harvesting 800-nm absorption band is absent and the longer wavelength peak is red-shifted to 872 nm, characteristic of the \( \text{LHI} \) contribution (Fig. 5). Wild-type strain SB1003 has peaks at 800 and 855 nm in approximately a 1:1.5 ratio. Introduction of pRPSLH2KAN into MW442 restores the wild-type absorption maxima wavelengths and approximate peak ratios. These data are consistent with the explanation that pRPSLH2KAN is complementing the chromosomal point mutation in MW442, which must be within or near the \( \text{LHII} \) structural genes. The high frequency of complementation observed in the transconjugants implies that the \( \text{LHII} \) promoter on the plasmid is functioning in *trans*. We have not tested whether the expression of \( \text{LHII} \) from the plasmid can be inhibited by high light intensity (24).

**DISCUSSION**

**Light-Harvesting Structures.** Zuber and co-workers (6, 7) have modeled the light-harvesting polypeptides as containing a single transmembrane \( \alpha \) helix of \( \approx 20 \) residues that spans the membrane once. Alignment of the primary sequence data for maximum homology (such as we have done for *R. capsulata* in Fig. 6) revealed a conserved histidine residue in the putative transmembrane \( \alpha \) helix. This histidine has been modeled as the bacteriochlorophyll binding site wherein a lone pair electron from the nitrogen in the imidazole ring coordinates as the fifth ligand to the porphyrin magnesium. They have also proposed an \( \alpha-\beta \) dimeric structure for the higher-order structure of the light-harvesting peptides that predicts specific excitonic coupling of bacteriochlorophyll molecules. In the case of the *R. capsulata* sequences shown in Fig. 6, the B870 absorption band is the consequence of a bacteriochlorophyll dimer coordinated by \( \alpha \)-histidine-32 and \( \beta \)-histidine-39; B850 dimer is coordinated by \( \alpha \)-histidine-31 and \( \beta \)-histidine-38. The B800 absorption band is
nonexcitonic and may be coordinated by β-histidine-20. No spectral band has been assigned to B870 α-histidine-51, although this histidine is within an Ala-X-X-His sequence.

All seven of the histidine residues in the light-harvesting sequences of R. capsulata are preceded by an alanine residue one turn of the helix (in α-3.6 conformation) to the amino-terminal side: Ala-X-X-His. We propose that alanine is also part of the general bacteriochlorophyll binding site and that it has been selected for steric reasons. Alanine has the smallest R group, which is appreciably hydrophobic. Reaction center histidines (5, 25) do not occur in Ala-X-X-His structure, although there are histidine residues within hydrophobic domains. Workers in our laboratory are currently using in vitro oligonucleotide-mediated mutagenesis techniques (8) on the cloned light-harvesting genes to test the involvement of specific residues in bacteriochlorophyll binding. By targeting on or near the Ala-X-X-His sequences, we hope to affect red or blue shifts in single absorption peaks.

Evolution of the Photosynthetic Apparatus. Gene duplications may have played an important role in the evolution of the photosynthetic apparatus structural genes. The L and M subunits of the reaction center are extensively homologous (5) and may have arisen by gene duplication of an ancestral reaction center sequence. Similarly, the percentage homology in residues between light-harvesting β-subunits (34%) and between light-harvesting α-subunits (27%) suggests that both β genes had a common ancestor and that both α genes had a common ancestor. Comparison of the gene organization derived from primary nucleotide sequence data implies that the ancestral LH1 gene had β-α organization. Both antennae have identical arrangements of Shine–Dalgarno sequences (β, G-G-A-G-G; α, A-G-G-A-G). Furthermore, the putative bacteriochlorophyll binding site Ala-X-X-His is absolutely conserved. We may speculate further and suggest that the LHII genes arose from an ancestor more like LH1, because the LH1 antenna is more central to the reaction center in the sense of structural organization, energy flow, and gene regulation. Second, the LHII genes are apparently outside of the photosynthetic gene cluster, a possible consequence of gene duplication involving a transposition to a second chromosomal locus.

The position of the structural gene encoding the 14 kDa

---

**FIG. 5.** Restoration of wild-type chromatophore absorption spectrum by complementation of MW442 (LHII− mutant) with pPRSLH2KAN. Lower spectrum, MW442; middle spectrum, MW442 bearing pPRSLH2KAN; upper spectrum, wild-type strain SB1003.
Fig. 6. Maximum homology alignment of the amino acid sequences for the β-polypeptides and for the α-polypeptides from LHI (B870 complex) and LHII (B800–B850 complex). Conserved putative bacteriochlorophyll binding sequences Aα-X-X-His are boxed. Hydropathy is displayed by coding above or below each residue: *, polar charged; ω, polar noncharged; *, hydrophobic; □, histidine. Structure and phylogeny of the LH sequences are discussed in the text. Amino acids are represented by standard one-letter abbreviations.

(apparent molecular size from NaDodSO4/PAGE) polypeptide is not known. It is indeed possible that the structural gene is encoded by either pRPSLH2 or pRPS404. The 14-kDa subunit probably does not bind bacteriochlorophyll (26), although transposon mutations in the R-prime plasmid pRPS404 have been characterized that result in the loss of both the 800-nm absorption band and the 14-kDa polypeptide (22). These data imply that the 14-kDa polypeptide may act during assembly or in the stabilization of the 800-nm bacteriochlorophyll binding site in LHII.

The phylogeny of the Rhodospirillaceae (27) has been addressed by several groups using 16S rRNA catalogs (28, 29) and from cytochrome structures (30–32). This work has revealed a great diversity in the purple nonsulfur photosynthetic bacteria, wherein species such as R. capsulata and Paracoccus denitrificans are more closely related than R. capsulata and Rhodospseudomonas viridis. Photosynthetic and nonphotosynthetic bacteria are intermingled in the phylogenetic tree. This is consistent with proposals (27, 30) that cyclic photoreaction centers in purple nonsulfur bacteria gave rise to linear electron transport utilizing oxygen in higher bacteria and mitochondria. The degree to which lateral evolution (32) may have led to the spread of photosynthesis among diverse bacteria is currently an area of controversy.

Lateral evolution of a photosynthetic gene cluster is a possible mechanism for the propagation of photosynthesis through diverse bacteria. In this evolutionary scheme, a plasmid bearing the genes for photosynthesis may have spread through a diverse group of nonphotosynthetic species by genetic transfer. Seemingly, this argument has been strengthened by the recent findings that most of the genes necessary for the differentiation of the respiratory membrane into the photosynthetic membrane are clustered in R. capsulata and that this gene cluster may be mobilized for conjugation by broad host range conjugative plasmids (21). Comparison of a phylogenetic tree based on reaction center and LH sequences with the current tree based on rRNA catalogs and cytochrome data should reveal the extent of lateral evolution.

We thank Pablo Scolnik for the λ genomic library of R. capsulata and Barry Marrs for helpful discussions. This work was supported under a collaborative agreement between Cold Spring Harbor Laboratory and Exxon Research and Engineering Liaison.