JOURNAL OF BACTERIOLOGY, June 1990, p. 2877–2887
0021-9193/90/062877-11$02.00/0

Vol. 172, No. 6

Structure and Transcription of the Genes Encoding the B1015 Light-Harvesting Complex β and α Subunits and the Photosynthetic Reaction Center L, M, and Cytochrome *c* Subunits from *Rhodopseudomonas viridis*

C. WIESSNER,[1] I. DUNGER,[2] AND H. MICHEL[1]*

Max-Planck-Institut für Biophysik, Heinrich-Hoffmannstrasse 7, D-6000 Frankfurt/Main,[1] and Max-Planck-Institut für Biochemie, D-8033 Martinsried,[2] Federal Republic of Germany

Received 20 September 1989/Accepted 21 February 1990

The genes encoding the β and α subunits of the B1015 light-harvesting complex (LHC) and the L, M, and cytochrome *c* subunits of the photosynthetic reaction center from *Rhodopseudomonas viridis* are organized in an operon, in analogy to other nonsulfur purple bacteria, named the *puf* operon. In photoheterotrophically grown cells, two abundant *puf* operon mRNA species of 3,581 and 621 bases were present. The large transcript encoded the LHC β, LHC α, and reaction center L, M, and cytochrome *c* polypeptides, whereas the small transcript only coded for the LHC β and α polypeptides. Both transcripts share a common 5′ end which is located 115 bases upstream from the initiation codon of the LHC β gene. Two additional low-level transcripts of 3,718 and 758 bases with 5′ ends 254 ± 3 bases upstream from the LHC β gene were detected. Analysis of the DNA sequence preceding the different 5′ ends revealed DNA elements of striking homology. The 3′ ends of the small transcripts were mapped within the α-L intercistronic DNA region downstream from a sequence capable of forming a very stable stem-loop when transcribed into RNA. The 3′ termini of the large transcripts are located immediately downstream from the region coding the cytochrome *c* subunit in two areas resembling *rho*-independent transcription terminators. No open reading frames corresponding to *pufQ* and *pufX* from *Rhodobacter capsulatus* and *Rhodobacter sphaeroides* were present in the flanking DNA regions of the *puf* operon. In contrast, an open reading frame ending 191 base pairs upstream from the LHC β gene showed 50% homology at the amino acid level to the available sequence of the *bchA* gene from *R. capsulatus*. The genes coding for the B1015 LHC subunits had C-terminal extensions of 13 (β) and 10 (α) amino acids which were not present in the proteins isolated from intracytoplasmic membranes.

Rhodopseudomonas viridis is a gram-negative purple nonsulfur bacterium (18) which normally grows under photoheterotrophic anaerobic conditions, although it has been also reported to grow very slowly microaerophilically (24, 37) and aerobically (21).

The photosynthetic apparatus of *R. viridis* has been subject to extensive spectroscopic and structural investigations which ultimately led to the elucidation of the three-dimensional structure of the photosynthetic reaction center (14, 15, 27). The reaction center consists of four protein subunits, L, M, H, and a *c*-type cytochrome (36), with four covalently linked heme groups (32). In the intracytoplasmic membranes, the reaction center is surrounded by a ring of 12 B1015 light-harvesting complexes (LHCs), each consisting of single β, α, and γ subunits (9).

The DNA sequences of the genes encoding the L (*pufL*), M (*pufM*), and cytochrome *c* (*pufC*) subunits were determined recently (28, 29, 38), but until now there has been no detailed study of the transcriptional organization of the genes. This is in contrast to *Rhodobacter capsulatus* (for example, see references 5, 7, and 41) and *Rhodobacter sphaeroides* (for reviews, see references 17 and 23), in which the structure and regulation of genes necessary for photosynthetic growth have been extensively investigated. Very recently, the transcriptional organization of the B880 LHC β and α and L and M subunit genes of *Rhodospirillum rubrum* has been analyzed (6). *R. viridis* differs in many aspects from the above-mentioned organisms; the light-absorbing pigment is bacteriochlorophyll (bCh) *b* instead of bCh *a*, the reaction center includes a tightly bound cytochrome *c*, it lacks a peripheral LHC II (which is also the case for *R. rubrum*), and it grows only very poorly under aerobic and microaerophilic conditions and not at all fermentatively. It thus would be valuable to provide additional data concerning transcription of photosynthetic genes from another member of the nonsulfur purple bacteria.

In this study, for the first time the transcriptional organization of the *puf* DNA region from a nonsulfur purple bacterium with a tightly bound cytochrome *c* subunit in the photosynthetic reaction center was determined. In addition, we determined the DNA sequence of the LHC β and α subunit genes (*pufB* and *pufA*, respectively) and of an additional 958 base pairs (bp) of the upstream DNA region.

MATERIALS AND METHODS

Materials. DNA-modifying enzymes were obtained from Boehringer Mannheim Biochemicals (Mannheim, Federal Republic of Germany) or Bethesda Research Laboratories, Inc. (Gaithersburg, Md.). Avian myeloblastosis virus (AMV) reverse transcriptase was purchased from Boehringer Mannheim. Biodyne transfer membranes were obtained from Pall Corp. (Glenn Cove, N.Y.). All chemicals were of analytical grade.

Bacterial strains and growth conditions. *R. viridis* (DSM 133) was grown in sodium succinate medium 27 (11) in screw-cap bottles with stirring at 30°C under white light (5 to

* Corresponding author.

10 W/m^2). *Escherichia coli* DH5α (33) was used as the host for recombinant plasmids and was grown in LB medium (30) or on LB agar plates supplemented with ampicillin (100 μg/ml).

DNA sequencing. DNA was sequenced by the dideoxy-chain termination technique (34) with the modification that 7-deaza-dGTP was used instead of dGTP. Strands were synthesized at 50°C with Klenow polymerase. A *Pst*I-*Eco*RI restriction fragment of 1,337 bp containing *pufB*, part of *pufA*, and upstream DNA sequences and a *Bam*HI restriction fragment of 466 bp containing the C terminus of *pufA* and the N-terminal part of *pufL* were derived from the plasmid LHC 2 (28). These fragments were cloned into the vector pBluescript (Stratagene, La Jolla, Calif.). Appropriate smaller restriction fragments were isolated and subsequently cloned in the vectors M13mp18 and M13mp19 (26) to allow DNA sequencing with single-stranded recombinant bacteriophage DNA. DNA sequences were analyzed with UWGCG DNA analysis software (16) on a microVAX computer.

RNA isolation. RNA was isolated from photoheterotrophically grown cells in the exponential growth phase essentially as described previously (10). Cells were harvested by centrifugation for 10 min at 10,000 × *g* and washed once in sterile double-distilled water, and the wet weight was determined. The cells were transferred to a mortar, frozen in liquid nitrogen, and then ground to a fine powder. For each gram (wet weight) of cells, 5 ml of RNA isolation buffer (5 M guanidinium isothiocyanate, 0.5 M β-mercaptoethanol, 50 mM sodium acetate) and 1 ml of sterile double-distilled water were added. The resulting viscous homogenate was centrifuged for 30 min at 30,000 × *g*. The supernatant was then carefully loaded on a 5.7 M cesium chloride cushion (3 ml) in Beckman 70.1 Ti tubes and centrifuged for 6 h at 290,000 × *g*. After the supernatant was removed, the RNA pellet was washed twice with 70% ethanol and dissolved in sterile TE buffer (10 mM Tris hydrochloride, 1 mM EDTA [pH 6.5]). The concentration of the RNA solution was determined spectrophotometrically.

RNA electrophoresis, blotting, and hybridization. RNA was size fractionated on 1.2% agarose gels in MOPS buffer (20 mM morpholinepropanesulfonic acid, 5 mM sodium acetate, 1 mM EDTA, pH 7.0). Before electrophoresis, the RNA was denatured by glyoxylation (25). The RNA was transferred to nylon membranes in an LKB 2016 vacuum gene-blotting chamber for 2 h at a vacuum of 50 cm H_2O (5 kPa) with 20× SSC (3 M sodium chloride, 0.3 M trisodium citrate) as the transfer solution. The RNA was cross-linked to the membranes by UV light. Hybridization probes were obtained by in vitro transcription of appropriate DNA restriction fragments cloned in the phage promoter vectors pBluescribe and pBluescript in the presence of [α-^{32}P]UTP as recommended by the supplier (Stratagene). Before hybridization, the nylon membranes were prehybridized for 4 h in hybridization buffer (50% formamide, 5× SSC, 5× Denhardt solution, 200 μg of sonified herring sperm DNA per ml) at 65°C. Fresh hybridization buffer and labeled RNA probe (5×10^5 dpm) were then added, and hybridization was done for 12 h at 65°C. The filters were washed twice in 50% formamide–0.5% sodium dodecyl sulfate–5× SSC for 1 to 2 h at 65 to 75°C. They were exposed to X-ray films (Konica Medical) for 3 h to 5 days at −70°C with an intensifier screen.

Primer extension and RNA sequencing. Primer extension was performed as described previously (3). Cellular RNA (10 μg) was hybridized to the 5′-end-labeled 32-mer oligodeoxynucleotide 5′-CGCCTCTTCTTCCGTCAACCCTGTCAGG CTCG-3′ (5×10^4 dpm) at 30°C for 12 h. After extension of the hybridized primer with AMV reverse transcriptase (40 U), the mixture was phenol extracted and precipitated with ethanol. The pellet was dissolved in formamide loading buffer (90% formamide, 0.1% [wt/vol] bromphenol blue, 0.1% xylene cyanol) and analyzed on a 6% polyacrylamide–urea gel. Cellular RNA was sequenced as described in reference 19 with some modifications. Cellular RNA (10 μg) was ethanol precipitated and dissolved in 12 μl of annealing buffer (250 mM potassium chloride, 10 mM Tris hydrochloride [pH 8.3]). Labeled primer (5 ng, see above) was added, and the mixture was incubated for 3 min at 80°C followed by incubation at 37°C for 45 min. Aliquots (2 μl) of the primer-RNA mixture were added to 1.5-ml reaction tubes containing 3 μl of RT buffer (24 mM Tris hydrochloride [pH 8.3], 16 mM magnesium chloride, 8 mM dithiothreitol, 0.4 mM dATP, 0.4 mM dCTP, 0.8 mM dGTP, 0.4 mM dTTP) and 1 μl of 0.1 mM ddATP, 0.1 mM ddCTP, 0.1 mM ddGTP, or 0.2 mM ddTTP. The termination reaction was done at 50°C for 45 min with 40 U of AMV reverse transcriptase and stopped by the addition of 4 μl of formamide loading buffer (see above). The samples were analyzed on 6% polyacrylamide–urea sequencing gels.

Mung bean nuclease mapping. Appropriate DNA restriction fragments with one protruding 5′ end were labeled with either [α-^{32}P]dATP (3,000 Ci/mmol) with Klenow polymerase or [γ-^{32}P]ATP with T4 polynucleotide kinase by the method of Maniatis et al. (25). Labeled DNA restriction fragments (5×10^5 dpm) were coprecipitated with 40 μg of cellular RNA, dissolved in 30 μl of hybridization buffer (80% formamide, 40 mM PIPES [piperazine-*N*,*N*′-bis(ethanesulfonic acid)] [pH 6.4], 0.4 M sodium acetate, 1 mM EDTA), heated for 5 min at 85°C, and hybridized at 60°C for 12 h. The samples were digested with 200 U of mung bean nuclease for 30 min at 37°C after 300 μl of ice-cold mung bean nuclease buffer (50 mM sodium chloride, 1 mM zinc sulfate, 1 mM L-serine, 5% glycerol [vol/vol], 50 mM sodium acetate [pH 5.0]) was added. They were phenol extracted, ethanol precipitated, dissolved in formamide loading buffer (see above), and analyzed on 6% polyacrylamide–urea sequencing gels. An M13mp18 sequence ladder was used as a size marker.

RESULTS

Nucleotide sequence of upstream region of *pufL* gene. Recently, the complete sequences of the *pufL*, *pufM*, and *pufC* genes from *R. viridis* have been determined (28, 37). The sequence from an upstream *Pst*I restriction site to the ATG start codon of the *pufL* gene extending over 1,519 bp is presented in Fig. 1. This DNA region is contained in the two plasmids pRCPE1337 and pRCB466, which were constructed by inserting a *Pst*I-*Eco*RI DNA restriction fragment of 1,337 bp (Fig. 2, fragment K) and a *Bam*HI restriction fragment of 466 bp (Fig. 2, fragment C), respectively, into the vector pBluescript. Both DNA fragments were derived from the plasmid pLHC 2 (28).

Computer analysis of the sequence revealed two open reading frames (ORFs) (Fig. 1, *pufB* and *pufA*) which could be identified as the genes for the LHC β and α polypeptides by comparison with the previously published sequences obtained by protein sequencing (9). The translated amino acid sequences of *pufB* and *pufA* are in perfect agreement with the previously published amino acid sequences and in addition show the existence of a C-terminal extension of 13 amino acids for the β polypeptide and 10 amino acids for the α polypeptide. The prepolypeptides consist of 69 amino

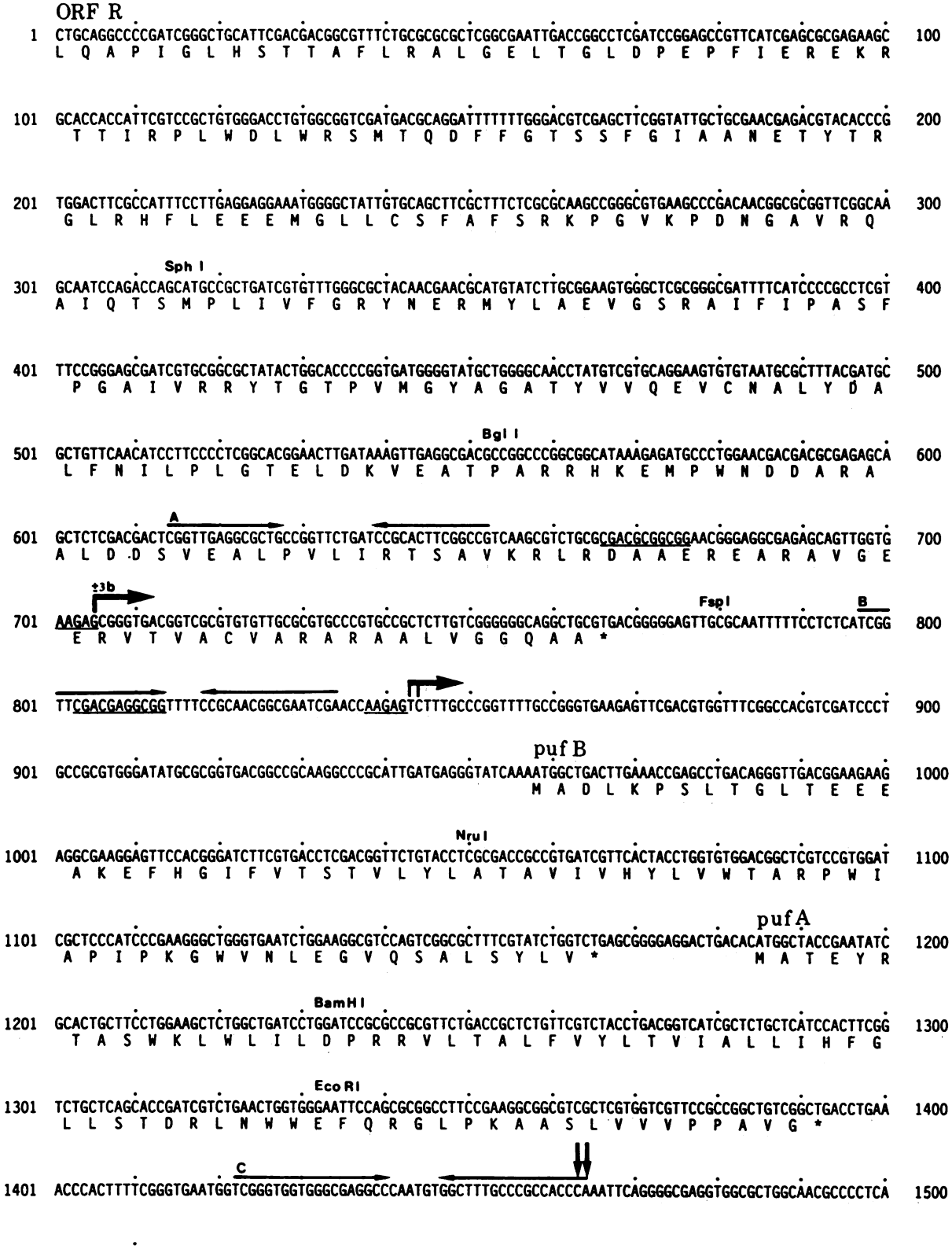

FIG. 1. DNA sequence of the LHC β and α subunit genes and the 5′ upstream DNA region between a *Pst*I restriction site and the ATG codon of the *pufL* gene. The translated sequences of the *pufB* gene, the *pufA* gene, and ORF R are indicated below the DNA sequence. Horizontal arrows indicate potential stem-loop structures. Stem-loop A has a ΔG of −11.4 kcal/mmol, stem-loop B has a ΔG of −13.4 kcal/mmol, and stem-loop C has a ΔG of −32.2 kcal/mmol. The locations of 5′ ends of *puf* operon transcripts are indicated by short horizontal arrows; the 3′ ends of the small transcripts are indicated by vertical arrows. Similar sequence elements upstream of the 5′ ends and putative Shine-Dalgarno sequences of the *pufB*, *pufA*, and *pufL* genes (which starts at nucleotide 1517) are underlined. Recognition sequences for DNA restriction enzymes are indicated above the DNA sequence.

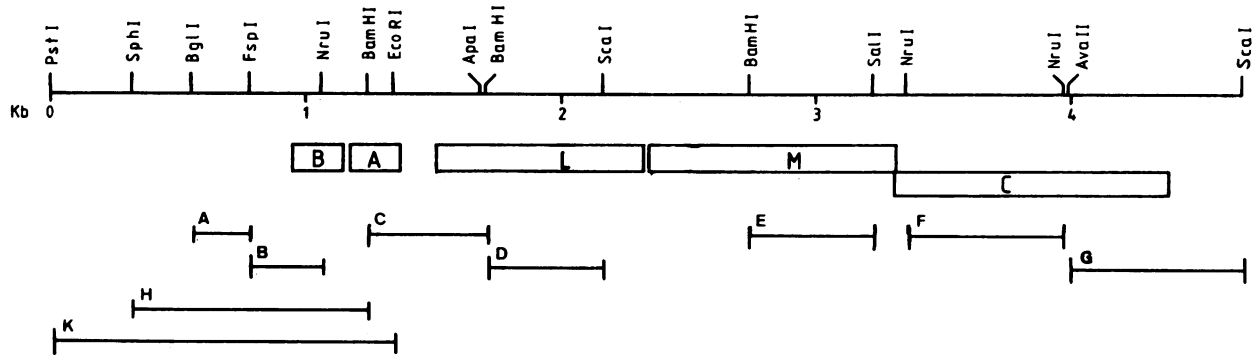

FIG. 2. Restriction and genetic map of the DNA region bearing the *pufB*, *pufA*, *pufL*, *pufM*, and *pufC* genes. The DNA region shown extends over 4.6 kilobase pairs from a *PstI* to a *ScaI* restriction site and is contained in the cosmid clone BD3 (38). Restriction sites used to obtain DNA fragments needed to perform Northern blot analysis (see Fig. 3) and mung bean nuclease mapping (see Fig. 5 and 6) are shown on the upper straight line, whereas the corresponding DNA fragments are shown at the bottom of the figure. Open bars indicate the location of the genes *pufB*, *pufA*, *pufL*, *pufM*, and *pufC*. Kb, Kilobases.

acids each. Both genes are transcribed in the same direction as the *pufL*, *pufM*, and *pufC* genes. The *pufA* gene is separated from the *pufL* gene by a 122-bp intergenic region. Within this region, a very stable potential stem-loop structure with a ΔG of -32.4 kcal/mol is located. The *pufB* and *pufA* genes are separated by 17 bp and are preceded by putative Shine-Dalgarno sequences. The resulting organization of the genes in the *puf* DNA region is shown in Fig. 2.

In addition to the *pufB* and *pufA* genes, a third ORF was detected. This ORF has the same orientation as the LHC genes and is separated by a 191-bp intergenic region from the *pufB* gene. Its 5'-terminal part is yet to be sequenced. Computer analysis revealed an amino acid homology of about 50% to the available sequence of the *bchA* gene from *R. capsulatus* (1, 39). No equivalent to *pufQ*, which recently has been reported to belong to the *puf* operon from *R. capsulatus* (4, 40) and to reside in the upstream DNA region of the *puf* operon from *R. sphaeroides* (12, 17), was detected in the DNA region upstream from the *pufB* gene. The search for an ORF corresponding to the B1015 γ polypeptide, which has been reported to be a component of the B1015 antenna complex in *R. viridis* (9), also was not successful.

Northern (RNA) blot analysis. The arrangement of the genes *pufB*, *pufA*, *pufL*, *pufM*, and *pufC* indicated that they were likely to be organized in an operon (Fig. 2). To test this hypothesis, cellular RNA was size fractionated on vertical agarose gels, blotted onto nylon membranes, and hybridized with radioactive RNA probes specific for different parts of the coding strand from the *puf* DNA region.

Probes specific for the genes *pufL*, *pufM*, and *pufC* (Fig. 2, fragments D, E, and F, respectively) all hybridized to an RNA species of approximately 3,700 bases (Fig. 3, lanes D, E, and F). The minimal length for an mRNA coding for the LHC β and α and reaction center L, M, and cytochrome *c* subunits would be 3,437 bases. Therefore, the length of the detected mRNA species (referred to as the large transcript) is obviously sufficient to represent a polycistronic mRNA coding for these five proteins.

When probes that included at least part of the gene *pufB* or *pufA* (Fig. 2, fragments B and C, respectively) were used in addition to the large RNA species, a transcript of about 600 bases (referred to as the small transcript) was detected (Fig. 3, lanes B and C). It was clearly more abundant than the large transcript in cells growing under photoheterotrophic conditions, although the relative intensities of the hybridization signals vary from about four- to sevenfold as determined by densitometric measurements (data not shown).

In addition, a minor RNA species of about 750 nucleotides was observed with probes B and C (Fig. 3, lanes B and C, respectively). An RNA species of identical size was detected with probe A, which is complementary to the coding strand in the DNA region 177 to 399 bp upstream from the LHC β gene (Fig. 3, lane A). When the exposure times of strips hybridized with probe A were extended to 5 days (Fig. 3, lane A'; exposure time for all other lanes was 3 h), it became evident that in addition to the 750-base RNA a very rare RNA species of about 3,700 bases was detected.

An additional notable feature of the Northern blot analysis was the detection of discrete intermediate-sized RNA molecules with probes E (ca. 2.0 kilobases) and F (2.0, 1.4, 1.3, and 1.2 kilobases; see small arrows in Fig. 3). These RNA molecules are likely to represent degradation products of the large *puf* operon transcripts.

Fine mapping of 5' ends of *puf* operon transcripts. Primer extension analysis and RNA sequencing were done with a 32-mer oligodeoxynucleotide complementary to a sequence of the coding strand in the 5'-terminal region of the *pufB* gene. It was 5' end labeled and hybridized to cellular RNA. The extension reaction was started by adding AMV reverse transcriptase to the annealed mixture. For direct RNA sequencing, appropriate amounts of dideoxynucleoside triphosphates were added to the reaction mixtures. The resulting extension products were size fractionated on a denaturing 6% polyacrylamide gel. A sequencing reaction of plasmid pRCPE1337 with the same 32-mer oligodeoxynucleotide as a primer was run in parallel as a size marker.

In primer extension experiments, two prominent products of 162 and 163 bases (Fig. 4, lanes 3 and 4) corresponding to 5' ends 114 and 115 bases upstream from the *pufB* gene were observed. In addition to these very abundant DNA molecules, several minor products of smaller size were observed. The ends of these molecules map in regions where potential stem-loops are located in the RNA sequence. We therefore consider these bands to be experimental artifacts. In addition to these minor smaller products, even less abundant products larger than 163 bases were detected on autoradiographs with extended exposure times. These products could have been due to RNA molecules possessing a more upstream 5' end as would be expected from the Northern blot analysis. Direct RNA sequencing, however, revealed that

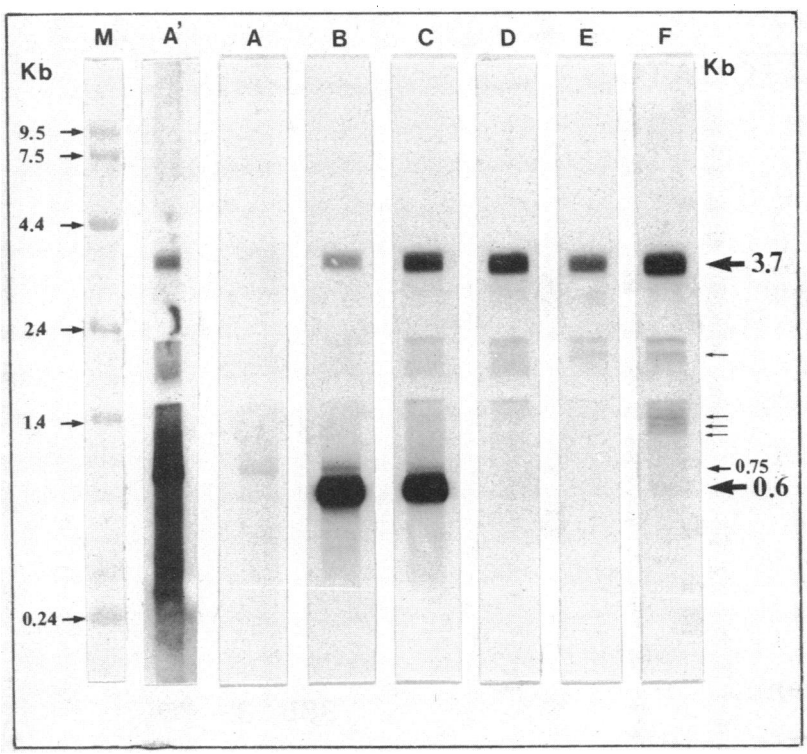

FIG. 3. Northern blot analysis of cellular RNA from *R. viridis* with probes specific for the *puf* DNA region. Cellular RNA was isolated from photosynthetically grown cells, size fractionated on 1.2% vertical agarose gels, and blotted onto nylon membranes. Hybridization was performed with [α-^{32}P]UTP-labeled RNA probes derived from in vitro transcription of recombinant phage promoter vectors bearing the DNA fragments depicted in Fig. 2. The designation of the lanes (A to F) corresponds to the nomenclature of the DNA restriction fragments in Fig. 2. Lane A' shows the autoradiography of a filter strip which was exposed to an X-ray film for 5 days at −70°C with intensifier screen, whereas exposure time for all the strips in lanes A to F was 3 h. An RNA ladder (Bethesda Research Laboratories) was used as a molecular weight standard (lane M), also transferred to nylon membranes, and stained with methylene blue (31). Numbers on the left (marker RNAs) and right (*puf* operon mRNAs) show size in kilobases (Kb).

these longer products most likely result from self-priming of the 163-base extension product via a hairpin loop and subsequent extension by AMV reverse transcriptase in the backward direction; when RNA-sequencing gels were exposed for longer periods to X-ray films, very rare chain termination products appeared above the 163-base products (Fig. 4), but the sequence did not fit with that of plasmid pRCPE1337. The sequence NNNNNACGTGGTTTCGGC CACGTCGAT was discerned, which corresponds exactly to the sequence from nucleotides 875 to 896 in Fig. 1. The occurrence of this sequence was concordant with the formation of a hairpin loop by base pairing of the 3'-terminal bases 3'-AGAAACGGGCC-5' of the 163-base extension product with its internal sequence 5'-TCTTCACCCGG-3' and subsequent extension by AMV reverse transcriptase.

Therefore, a second experimental approach was used to determine the 5' ends of the *puf* operon transcripts. Mung bean nuclease mapping was done with a 914-bp *Sph*I-*Bam*HI restriction fragment (Fig. 2, fragment H) which was selectively labeled at the protruding 5' end of the *Bam*HI restriction site, hybridized to cellular RNA, and digested with mung bean nuclease as described in Materials and Methods. Figure 5 shows protected DNA molecules of 386 ± 3 bases and 532 ± 3 bases (lane 1). The size of the smaller, more abundant protected DNA molecules corresponds to a 5' end at a position 107 bp upstream from the LHC β gene. This result is in agreement with primer extension analysis, taking into account the size determination uncertainty of three bases and that the proximal five bases of the transcript contain four adenine residues. This can result in low stability of the RNA-DNA duplex, which may be susceptible to mung bean nuclease attack.

The observation of protected DNA molecules of 532 ± 3 nucleotides indicates the existence of *puf* operon transcripts with 5' ends corresponding to a position 254 ± 3 nucleotides upstream from the *pufB* gene. Considering the detection of a low level of 750-base RNA in the Northern blot analysis with probes B and C (Fig. 2 and 3), the observation that probe A (Fig. 3, lanes A and A') hybridizes only to the 750-base RNA and a low-level RNA of about 3.7 kilobases, and the 3'-end determination of the small *puf* operon transcripts (see below), we suggest the existence of a minor fraction of large and small *puf* operon transcripts with 5' ends 254 ± 3 bases upstream from the LHC β gene. The major fraction of the large and small transcripts is supposed to possess a leader sequence of 115 bases. The hybridization signal corresponding to an RNA of approximately 3.7 kilobases in Northern blot analysis observed with probes B, C, D, and E (Fig. 3) should therefore be due to two RNA species of slightly different size (139 bases), which were not resolved on the agarose gels.

The same pattern of protected DNA molecules of 386 and 532 bases was obtained when either a 677-bp *Bgl*I-*Bam*HI restriction fragment labeled at the *Bam*HI site (Fig. 1, nucleotides 560 to 1237) or a 1,232-bp *Pst*I-*Bam*HI restriction fragment (Fig. 1, nucleotides 5 to 1237) was used for mung bean nuclease mapping (data not shown), confirming the result described above.

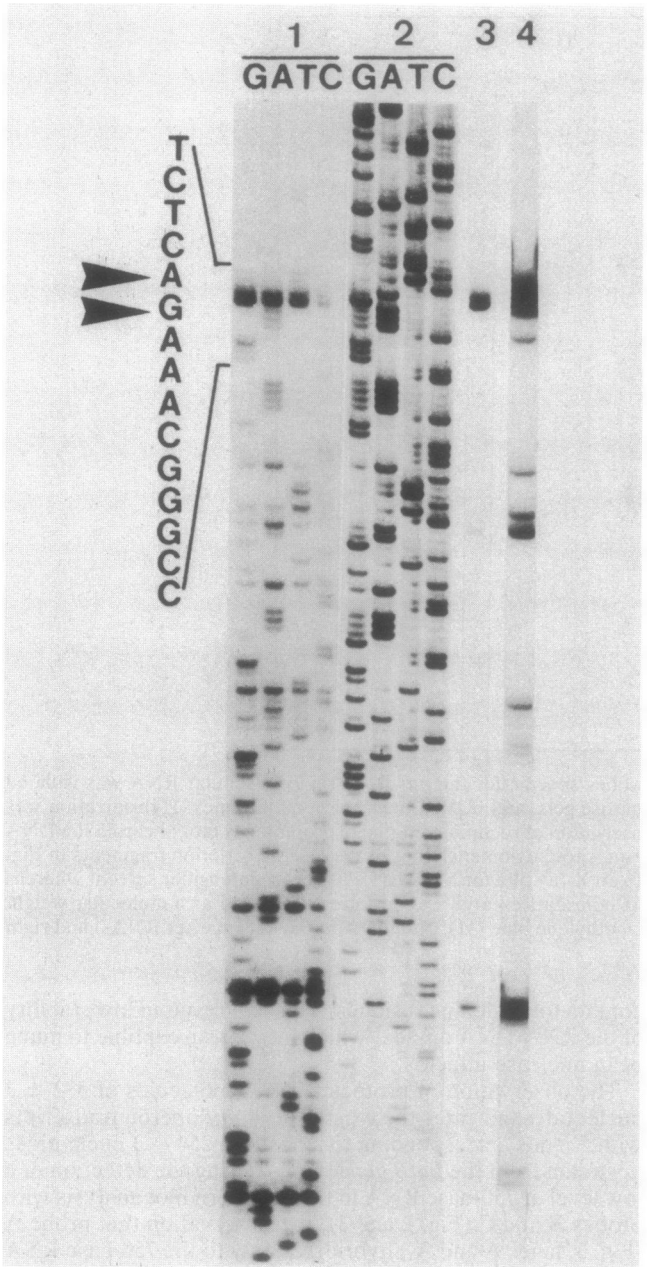

FIG. 4. Mapping of 5′ ends of *puf* operon transcripts by RNA sequencing and primer extension. A radioactive 5′-end-labeled (5×10^4 dpm) 32-mer oligodeoxynucleotide (5′-CGCCTCTTCTTCCGT CAACCCTGTCAGGCTCG-3′, see Fig. 1, bases 974 to 1005) was hybridized to cellular RNA (10 μg) and extended in the presence of either dideoxynucleotides (panel 1) or deoxynucleotides (lanes 3 and 4) with 40 U of AMV reverse transcriptase. In panel 2, the plasmid pRCPE1337 bearing a 1,337-bp *Pst*I-*Eco*RI restriction fragment (Fig. 2, fragment K) was used as a template for dideoxynucleotide sequencing (34), using the same oligodeoxynucleotide as above as a primer to obtain a size marker. Lane 4 shows the same track as lane 3 but exposed for 2 days to an X-ray film, in contrast to a exposure time of 12 h for all other lanes.

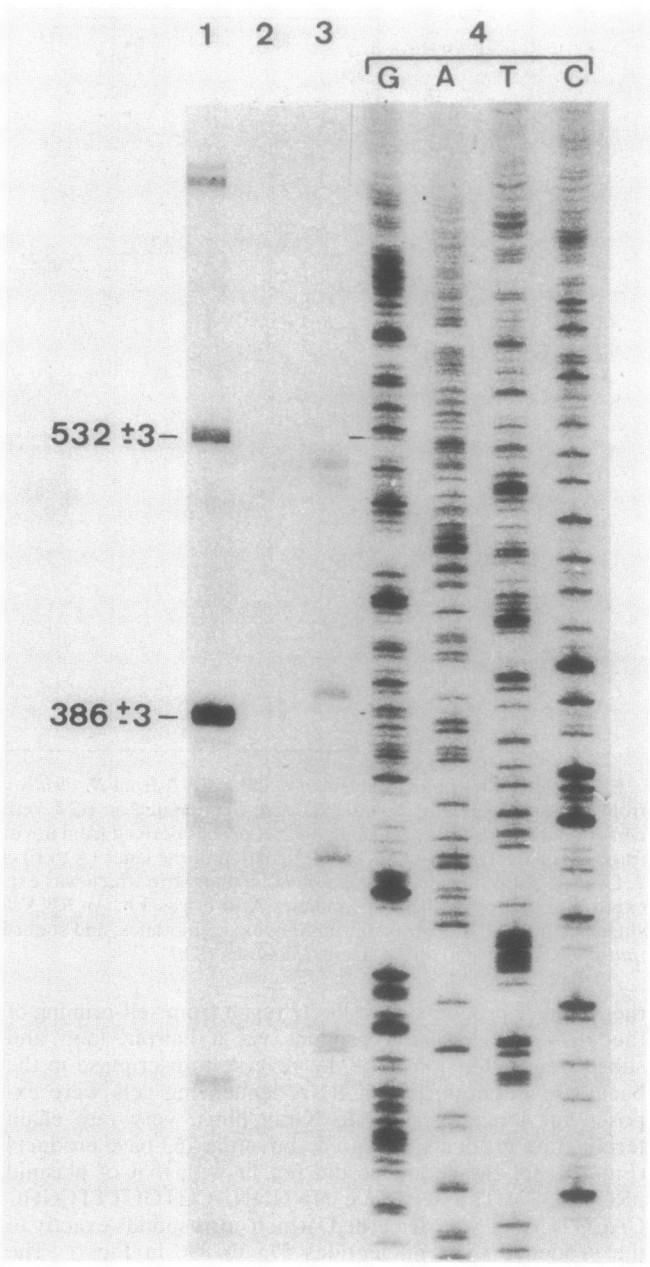

FIG. 5. Determination of 5′ ends of *puf* operon transcripts by mung bean nuclease mapping. A 914-bp *Sph*I-*Bam*HI restriction fragment (Fig. 2, fragment H) labeled with [γ-^{32}P]ATP by using T4 polynucleotide kinase (5×10^4 dpm) was hybridized to 40 μg of cellular RNA (lane 1) or 50 μg of *E. coli* tRNA as a negative control (lane 2) and digested with 250 U of mung bean nuclease. The samples were analyzed on a 5% polyacrylamide–urea gel with radioactively labeled *Hin*fI restriction fragments from plasmid pBR322 (lane 3) and an M13mp18 sequence ladder (panel 4) as size standards. Two protected fragments corresponding to 532 ± 3 and 386 ± 3 bases were observed. The uppermost band results from rehybridization of the denatured 914-bp *Bam*HI-*Sph*I restriction fragment and subsequent protection from mung bean nuclease digestion.

A thorough examination of the DNA region upstream from the determined 5′ ends of the *puf* operon transcripts revealed some striking features (Fig. 1; and see Fig. 9A). The 5′ terminus corresponding to position −115 is preceded by the sequence AAGAG, which is also found in the region where the 5′ end of the low-level transcripts map. More strikingly, in both cases, 24 bp upstream a sequence of 11 bp is found which is nearly identical (10 of 11 bp). In addition, both regions are preceded by sequences of dyad rotational sym-

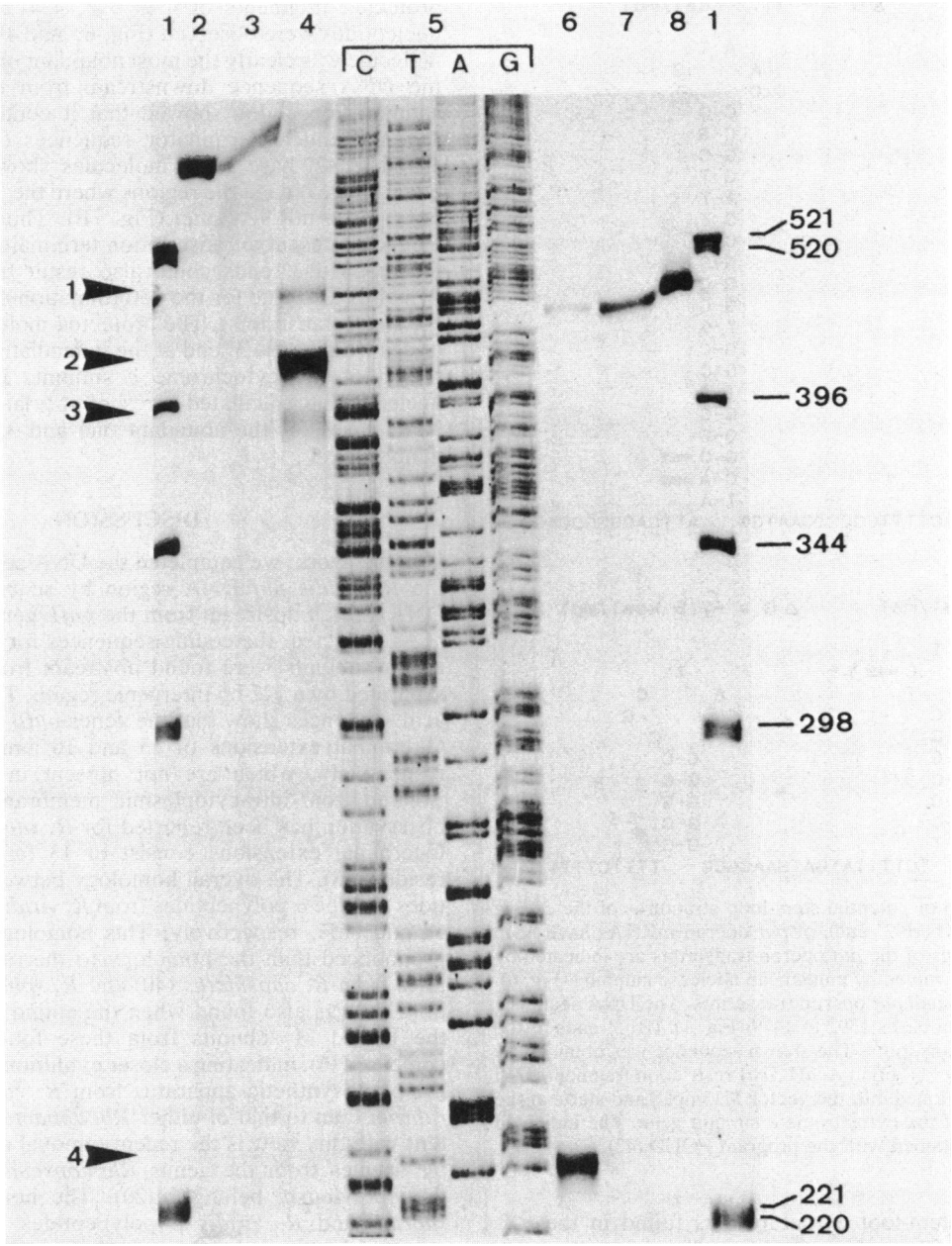

FIG. 6. Mung bean nuclease mapping of 3′ ends of *puf* operon transcripts. Cellular RNA (40 μg) was hybridized to 3′-end-labeled DNA restriction fragments (5×10^5 dpm), and the samples were digested with 200 U of mung bean nuclease as described in the legend to Fig. 5 and analyzed on 6% polyacrylamide–urea gels. Lane 4 shows the result of a digestion assay when a 685-bp *Ava*II-*Sca*I restriction fragment (Fig. 2, fragment G) was used to determine the 3′ ends of large *puf* operon transcripts. Three protected fragments of 479 ± 5 bases (arrowhead 1), 434 ± 5 bases (arrowhead 2), and 404 ± 5 bases (arrowhead 3) could be observed. The 434-base fragment is obviously the most abundant one. Lane 6 shows the result of a similar experiment with a 466-bp *Bam*HI-*Apa*I restriction fragment (Fig. 2, fragment C) as a protection probe to determine the 3′ end of the 600-base small *puf* operon transcript. In this case, a double band corresponding to 231 and 230 bases (arrowhead 4) is visible. In lanes 2 and 8, the labeled fragments (fragments C and G, respectively) are shown, whereas lanes 3 and 7 (fragments C and G, respectively) represent samples in which the labeled fragments were hybridized to 50 μg of *E. coli* tRNA and subsequently digested with 200 U of mung bean nuclease as a negative control. As size markers, radioactively labeled *Hin*fI restriction fragments from plasmid pBR322 (lane 1) and an M13mp18 sequence ladder (panel 5) were used. Numbers on right show size in nucleotides.

metry which seem to share some homology (Fig. 1, stem-loops A and B). The symmetry of the sequence upstream of the 5′ end at position −115 is much more pronounced.

Mapping the 3′ ends of *puf* operon transcripts. To map the 3′ ends of the small transcripts encoding only the LHC β and α polypeptides, we performed mung bean nuclease digestion protection experiments using a 466-bp *Bam*HI restriction fragment (Fig. 2, fragment C). This fragment was recut with the restriction enzyme *Apa*I to allow selective labeling at the *Bam*HI site.

Two protected DNA molecules of 230 and 231 nucleotides (Fig. 6, lane 6) were observed, which showed the small transcripts to have 3′ termini mapping immediately downstream from the α-L intercistronic stem-loop structure (Fig.

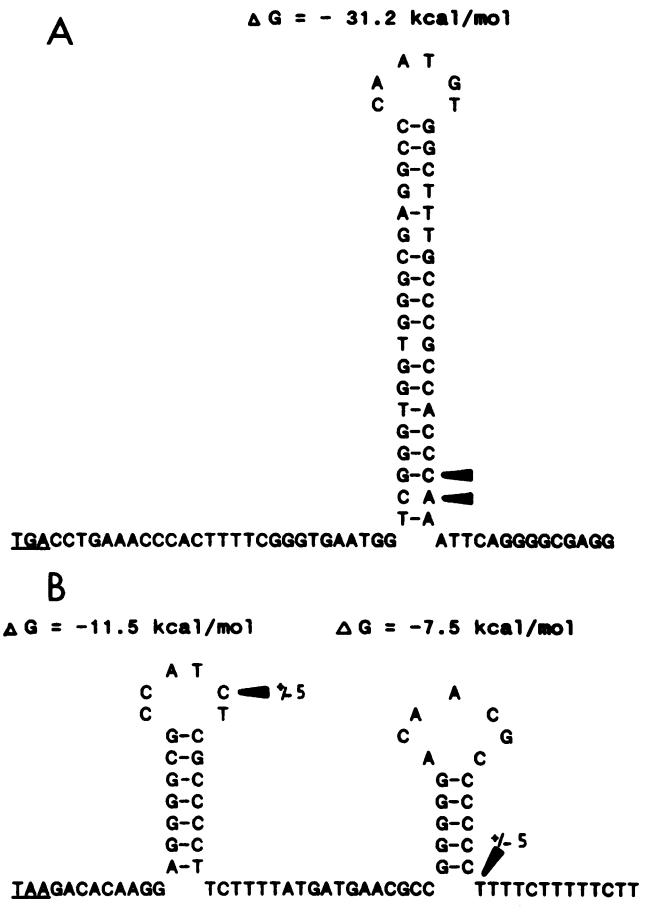

FIG. 7. Diagram of potential stem-loop structures of the coding strand in regions where 3′ ends of *puf* operon mRNAs have been mapped. The 3′ ends of the *puf* operon transcripts are indicated by arrowheads as determined by mung bean nuclease mapping (Fig. 6). (A) 3′ ends of the small *puf* operon transcripts. The DNA sequence shown extends from bases 1392 to 1479 (Fig. 1). (B) 3′ ends of the large *puf* operon transcripts. The shown sequence was obtained by dideoxy sequencing (34) of an *Ava*II-*Sca*I restriction fragment (G in Fig. 2) which was cloned into the vector M13mp18 and starts at the TAA stop codon of the cytochrome *c* subunit gene. The indicated ΔG values were obtained with the program FOLD (42).

7A). A similar stem-loop structure was found in the α-L intergenic region from other common sulfur purple bacteria (7, 8, 13) and was shown to be essential for the formation of the 0.5-kilobase *puf*-specific transcript in *R. sphaeroides* (13). The observation of protected DNA molecules of 457 bases results in this case not only from rehybridization of the denatured DNA fragment but also from RNA-DNA hybrid formation between the large *puf* operon mRNA and the labeled strand of the DNA fragment. The fact that no protected molecules with sizes between 230 and 457 bases were detected shows that the minor 750-base RNA (encoding LHC β and α subunits) observed in Northern analysis has no 3′ end extending in the L-subunit-coding region. This finding supports our suggestion of two fractions of large and small *puf* operon transcripts with different 5′ ends. The exact sizes of the small transcripts are 621 and 758 nucleotides.

The same experimental approach was used to determine the terminator region of the *puf* operon with a 685-bp *Ava*II-*Sca*I restriction fragment (Fig. 2, fragment G) which was labeled at the *Ava*II site as a probe. In this case, three protected fragments of 479 ± 5, 434 ± 5, and 404 ± 5 nucleotides were observed (Fig. 6, lane 4), and the 434-base molecule was clearly the most abundant one. Examination of the DNA sequence downstream from the cytochrome *c* subunit stop codon showed that it contains two potential rho-independent terminator sequences (2). The protected 434- and 479-base DNA molecules showed that these sequences are indeed the regions where the 3′ ends of the large *puf* operon mRNAs map (Fig. 7B). Thus, this region may indeed represent a transcription terminator region, although the mapped 3′ ends could also result from 3′ processing (which is also true for the determination of the 3′ ends from the small transcripts). The protected molecule of 404 nucleotides would set a 3′ end at the immediate end of the coding region of the cytochrome *c* subunit. Its origin remains unclear. The calculated sizes of the large transcripts are 3,581 bases for the abundant one and 3,718 bases for the minor one.

DISCUSSION

In this work, we completed the DNA sequence analysis of the *R. viridis puf* DNA region by sequencing a 1,519-bp DNA stretch upstream from the *pufL* gene.

As expected, the coding sequences for the B1015 LHC β and α subunits were found upstream from the *pufL* gene, separated by a 122-bp intergenic region. The deduced amino acid sequences show that the genes *pufB* and *pufA* code for C-terminal extensions of 13 and 10 amino acid residues, respectively, which are not present in the polypeptides isolated from intracytoplasmic membranes (9). The same observation has been reported for *R. rubrum*, in which the C-terminal extensions consist of 13 (*pufB*) and 9 (*pufA*) residues (8). The overall homology between the β polypeptides and the α polypeptides from *R. viridis* and *R. rubrum* is 68 and 60%, respectively. This homology is clearly more pronounced than the homology to the respective polypeptides from *R. capsulatus* (40) and *R. sphaeroides* (22) (Fig. 8). This was also found when the amino acid sequences of the L and M subunits from these four organisms were compared (6), indicating a closer evolutionary relationship of the photosynthetic apparatus from *R. viridis* to that of *R. rubrum* than to that of either *Rhodobacter* species. Consistent with this view is the recent removal of both *Rhodobacter* species from the genus *Rhodopseudomonas* to which they previously belonged (20). The homology of the *R. rubrum* and *R. viridis* β polypeptides is also very pronounced between the C-terminal extensions, in contrast to those of the α polypeptides (Fig. 8). The C-terminal extensions of the LHC precursor proteins might be needed for their correct insertion in the membrane resulting in the N terminus protruding in the cytoplasm (35).

Although the arrangement of the genes in the *puf* DNA region from the nonsulfur purple bacteria studied so far is similar in general, our sequence analysis demonstrates significant differences in the DNA regions flanking the *puf* operon from *R. viridis*.

(i) No equivalent to *pufQ* is present in the immediate-upstream DNA region of the *R. viridis puf* operon. The existence of *pufQ* has been noted for both *R. capsulatus* and *R. sphaeroides*, encoding a putative protein of 74 (4) or 77 (17) amino acid residues. It has been shown that *pufQ* is essential for bCh *a* synthesis in *R. capsulatus*, and the putative gene product is supposed to be a carrier of bCh *a* biosynthetic intermediates (4). Since the bCh *a* and bCh *b* biosynthetic pathways are likely to be very similar, it is

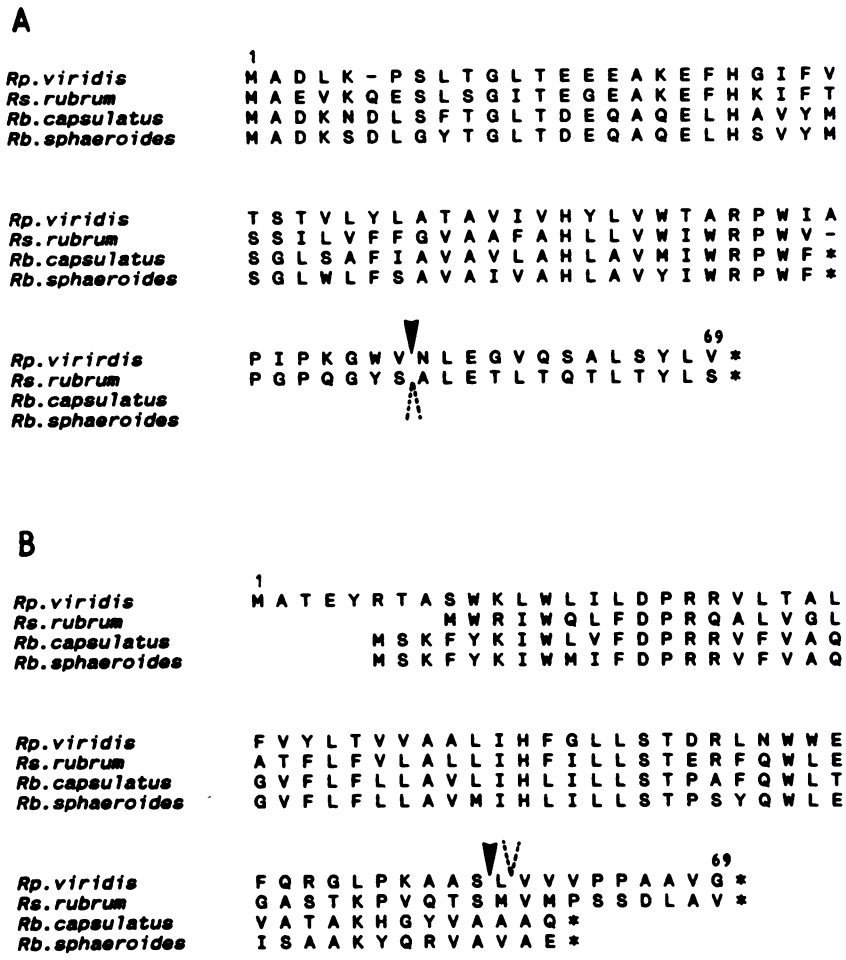

FIG. 8. Comparison between the amino acid sequences of the LHC β (A) and α (B) subunits from different purple nonsulfur bacteria. The DNA sequences of both LHC subunits from *R. viridis* and *R. rubrum* (8) indicate the existence of C-terminal extensions which are not present in the polypeptides isolated from intracytoplasmic membranes, whereas neither *Rhodobacter* (22, 40) species has such extensions. The putative processing sites of the precursor proteins are indicated by solid arrowheads (*R. viridis*) or dotted arrowheads (*R. rubrum*).

surprising that no *pufQ* gene is found in the *R. viridis puf* operon upstream DNA region. It might be worthwhile to look for an equivalent to *pufQ* in the *R. viridis* chromosome by hybridizing genomic DNA with heterologous probes from *R. capsulatus* or *R. sphaeroides*.

(ii) Another ORF, *pufK*, which has been supposed to be crucial for selective transcription initiation in *R. sphaeroides* (17) is also not present in the *R. viridis puf* operon upstream DNA region.

(iii) The *puf* operons from both *R. sphaeroides* and *R. capsulatus* have been reported to include a *pufX* gene (17, 39). The function of this gene is as yet unclear (17). Preliminary sequence analysis of about 500 bp downstream from the *R. viridis puf* operon showed that no equivalent to *pufX* is present in this DNA region (C. Wiessner, unpublished data). This observation is consistent with the determination of the 3′ ends of the large *puf* operon transcript in a region immediately downstream from the cytochrome *c* gene. It is interesting that 3′-end-mapping experiments of the large transcript from *R. rubrum* indicate that this *puf* operon also does not include a *pufX* gene (6).

In contrast, an ORF R was found in the *R. viridis puf* operon upstream DNA region which has the same transcriptional orientation as the *puf* operon genes. This ORF R shares a high homology with the *bchA* gene (39) from *R. capsulatus*. A comparison of the translated sequences of the 3′-terminal part of the *bchA* sequence (derived from reference 1) and the partial sequence of the *R. viridis* ORF R shows 50% identical amino acid residues. It is likely that the putative gene product of the *R. viridis* ORF R has a function very similar to that of the *bchA* gene product.

By combining the results of the Northern analysis and the 5′-end-mapping and 3′-end-mapping experiments, we conclude that the *R. viridis puf* operon gives rise to two abundant mRNAs of 3,581 and 621 nucleotides with a leader region of 115 bases and to two minor mRNAs of 3,718 and 758 nucleotides with a leader region of 254 bases. In each instance, the smaller transcript is severalfold more abundant than the larger transcript (Fig. 3, lanes A, A′, and B). This suggestion is supported by the following observations. (i) A minor RNA species of about 750 bases was detected by Northern blot analysis with probes A, B, and C (Fig. 2 and 3). Probe A did not hybridize to the abundant 600-base RNA but to a very rare RNA of about 3,700 bases. (ii) 5′-end-mapping experiments revealed two different 5′ ends for the *puf* operon transcripts. RNA molecules with 5′ ends extended further upstream were significantly less abundant than RNA molecules with 5′ ends more proximal to the LHC

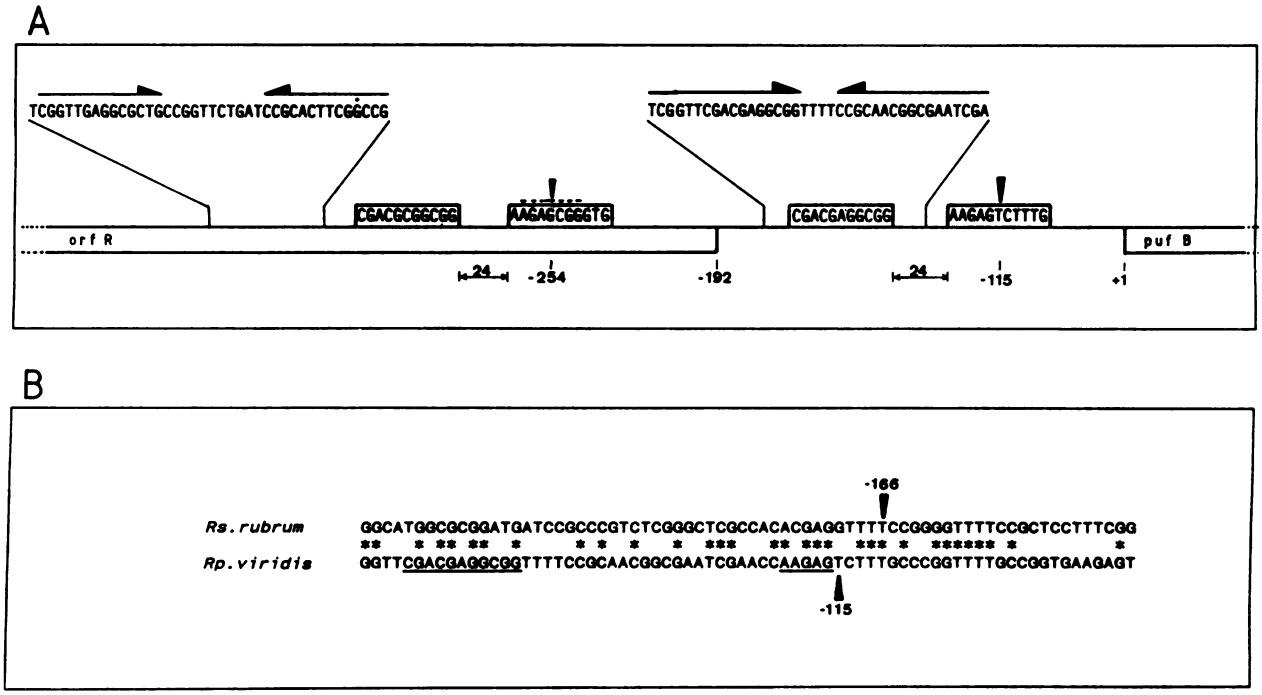

FIG. 9. Diagram of sequence similarities in the DNA regions where the 5′ ends of *puf* operon transcripts have been mapped. (A) Sequence similarities in the DNA regions 115 and 254 nucleotides upstream the *pufB* gene from *R. viridis*. The nucleotides corresponding to the 5′ ends of stable mRNAs are marked with a vertical arrowhead. (B) DNA sequence alignment of the DNA regions from the *R. viridis* and *R. rubrum* *puf* operons in which 5′ ends of stable mRNAs have been mapped.

β gene (Fig. 5, lane 1). (iii) The 3′-end-mapping experiments designed to detect the ends of the small transcripts revealed only one size class of protected molecules (Fig. 6, lane 6). If the 750-base RNA would include part of the L-subunit gene coding region, a larger protected fragment should have been observed, since the region between position −115 (relative to the LHC β gene) and the *Apa*I site of the used fragment C extends over 846 bp.

These results are consistent with the proposed existence of four *puf* operon transcripts. However, our experiments do not rule out the possibility that all transcripts initially share one common 5′ end and are then subject to rapid 5′ processing events, as is proposed for the *puf* operon from *R. capsulatus* (1, 5).

The DNA regions immediately upstream from the positions which correspond to the two different determined 5′ ends of the *puf* operon transcripts show striking similarities (Fig. 9A). In both cases, the motif CGACGMGGCGG-N_{24}-AAGAG (M means C or A) is found. It is very unlikely that these sequence elements occur by accident at the determined positions. Thus, it is possible that these defined regions are transcription initiation sites, although the observed sequence elements also could represent recognition sites for specific processing events (see above). Additional studies are necessary to define the promoter region of the *R. viridis puf* operon.

When we compared the DNA sequences upstream from the determined mRNA 5′ ends, no apparent homologies to common procaryotic consensus sequences recognized by different σ factors of RNA polymerase or regulating proteins were found. This was also noted for the *puf* operon transcripts from *R. capsulatus* (7) and *R. sphaeroides* (22), although more recently a sequence which resembles the consensus sequence recognized by the σ^N (*ntrA*) subunit of RNA polymerase has been reported in the oxygen-regulated *puf* operon promoter from *R. capsulatus* (5).

In addition, we did not find significant homologies to DNA sequences from defined *puf* operon promoters or to DNA sequences upstream from determined 5′ ends from *puf* operon transcripts from both *R. capsulatus* (1, 7) or *R. sphaeroides* (22). In contrast, a significant homology to the DNA region from *R. rubrum* where the 5′ ends of the *puf* operon transcripts map was detected (Fig. 9B). This finding is likely to emphasize again the closer evolutionary relationship of *R. viridis* to *R. rubrum* than to either *Rhodobacter* species and suggests a similar transcriptional regulation of the *puf* operons from *R. viridis* and *R. rubrum*.

Despite the general similarity of the photosynthetic apparatus and the genetic organization of the respective genes of the nonsulfur purple bacteria, every member of this group of bacteria studied so far has revealed additional special features. It will be interesting to see whether the oxygen-dependent regulation of the photosynthetic apparatus in these organisms shares one underlying principle or is achieved by different mechanisms.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (Leibniz Programm to H.M.) and the Max-Planck-Gesellschaft.

The 32-mer oligodeoxynucleotide was synthesized on an Applied Biosystems 3A DNA synthesizer by J. Engels. We are grateful to H. Reiländer, R. Grisshammer, and S. Vasudevan for helpful comments on the manuscript.

LITERATURE CITED

1. **Adams, C. W., M. E. Forrest, S. N. Cohen, and T. J. Beatty.** 1989. Structural and functional analysis of transcriptional con-

trol of the *Rhodobacter capsulatus puf* operon. J. Bacteriol. **171**:473–482.
2. **Adhya, S., and M. Gottesmann.** 1978. Control of transcription termination. Annu. Rev. Biochem. **47**:967–996.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1987. Current protocols in molecular biology, p. 4.8.1–4.8.3. John Wiley & Sons, Inc., New York.
4. **Bauer, C. E., and B. L. Marrs.** 1988. *Rhodobacter capsulatus puf* operon encodes a regulatory protein (*puf* Q) for bacteriochlorophyll synthesis. Proc. Natl. Acad. Sci. USA **85**:7074–7078.
5. **Bauer, C. E., D. A. Young, and B. L. Marrs.** 1988. Analysis of the *Rhodobacter capsulatus puf* operon. J. Biol. Chem. **263**:4820–4827.
6. **Belanger, G., and G. Gingras.** 1988. Structure and expression of the *puf* operon messenger RNA in *Rhodospirillum rubrum*. J. Biol. Chem. **263**:7639–7645.
7. **Belasco, J. G., J. T. Beatty, C. W. Adams, A. von Gabain, and S. N. Cohen.** 1985. Differential expression of photosynthesis genes in *R. capsulatus* results from segmental differences in stability within the polycistronic rxcA transcript. Cell **40**:171–181.
8. **Berard, J., G. Belanger, P. Corriveau, and G. Gingras.** 1986. Molecular cloning and sequence of the B880 holochrome gene from *Rhodospirillum rubrum*. J. Biol. Chem. **261**:82–87.
9. **Brunisholz, R. A., F. Jay, F. Suter, and H. Zuber.** 1985. The light-harvesting polypeptides of *Rhodopseudomonas viridis*. Biol. Chem. Hoppe-Seyler **336**:87–89.
10. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
11. **Claus, D., and C. Schaab-Engels (ed.).** 1977. German collection of microorganisms. Catalogue of strains, 2nd ed., p. 279–280. Gesellschaft für Strahlen-und Umweltforschung mbH, Munich.
12. **Davis, J., T. J. Donohue, and S. Kaplan.** 1988. Construction, characterization, and complementation of a Puf$^-$ mutant of *Rhodobacter sphaeroides*. J. Bacteriol. **170**:320–329.
13. **DeHoff, B. S., J. K. Lee, T. J. Donohue, R. I. Gumport, and S. Kaplan.** 1988. In vivo analysis of *puf* operon expression in *Rhodobacter sphaeroides* after deletion of a putative intercistronic transcription terminator. J. Bacteriol. **170**:4681–4692.
14. **Deisenhofer, J., O. Epp, K. Miki, R. Huber, and H. Michel.** 1984. X-ray structure analysis of a membrane protein complex: electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction centre from *Rhodopseudomonas viridis*. J. Mol. Biol. **180**:385–398.
15. **Deisenhofer, J., O. Epp, K. Miki, R. Huber, and H. Michel.** 1985. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3 Å resolution. Nature (London) **318**:618–624.
16. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
17. **Donohue, T. J., P. J. Kiley, and S. Kaplan.** 1988. The *puf* operon region in *Rhodobacter sphaeroides*. Photosynth. Res. **19**:39–61.
18. **Drews, G., and P. Giesbrecht.** 1966. *Rhodopseudomonas viridis*, nov. spec., ein neu isoliertes obligat phototrophes Bakterium. Arch. Mikrobiol. **53**:255–262.
19. **Geliebter, J.** 1987. Dideoxynucleotide sequencing of RNA and uncloned cDNA. BRL Focus **9**:5–8.
20. **Imhoff, J. F., H. G. Trüper, and N. Pfennig.** 1984. Rearrangement of the species and genera of the phototrophic "purple nonsulfur bacteria". Int. J. Syst. Bacteriol. **34**:340–343.
21. **Kämpf, C., R. M. Wynn, R. W. Shaw, and D. B. Knaff.** 1987. The electron transfer chain of aerobically grown *Rhodopseudomonas viridis*. Biochim. Biophys. Acta **894**:228–238.
22. **Kiley, P. J., T. J. Donohue, W. A. Havelka, and S. Kaplan.** 1987. DNA sequence and in vitro expression of the B875 light-harvesting polypeptides of *Rhodobacter sphaeroides*. J. Bacteriol. **169**:742–750.
23. **Kiley, P. J., and S. Kaplan.** 1988. Molecular genetics of photosynthetic membrane biosynthesis in *Rhodobacter sphaeroides*. Microbiol. Rev. **52**:50–69.
24. **Lang, F. S., and D. Oesterhelt.** 1989. Microaerophilic growth and induction of the photosynthetic reaction center in *Rhodopseudomonas viridis*. J. Bacteriol. **171**:2827–2834.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:22–79.
27. **Michel, H.** 1982. Three dimensional crystals of a membrane protein complex. The photosynthetic reaction centre from *Rhodopseudomonas viridis*. J. Mol. Biol. **158**:567–572.
28. **Michel, H., K. A. Weyer, H. Gruenberg, I. Dunger, D. Oesterhelt, and F. Lottspeich.** 1986. The 'light' and 'medium' subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: isolation of the genes, nucleotide and amino acid sequence. EMBO J. **5**:1149–1158.
29. **Michel, H., K. A. Weyer, H. Gruenberg, and F. Lottspeich.** 1985. The 'heavy' subunit of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: isolation of the gene, nucleotide sequence and amino acid sequence. EMBO J. **4**:1667–1672.
30. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Monroy, A. F.** 1988. Staining immobilized RNA ladder. BRL Focus **10**:14.
32. **Prince, R. C., J. S. Leigh, Jr., and P. L. Dutton.** 1976. Thermodynamic properties of the reaction center of *Rhodopseudomonas viridis*: in vivo measurement of the reaction center bacteriochlorophyll primary acceptor intermediate electron carrier. Biochim. Biophys. Acta **440**:622–636.
33. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Stark, W., F. Jay, and K. Muehlenthaler.** 1986. Localisation of reaction center and light-harvesting complexes in the photosynthetic unit of *Rhodopseudomonas viridis*. Arch. Microbiol. **146**:130–133.
36. **Thornber, J. P., R. J. Cogdell, R. J. Seftor, and G. D. Webster.** 1980. Further studies on the composition and spectral properties of the photochemical reaction centers of bacteriochlorophyll b containing bacteria. Biochim. Biophys. Acta **593**:60–75.
37. **Trüper, H. G., and N. Pfennig.** 1978. Taxonomy of *Rhodospirillales*, p. 19–27. *In* R. K. Clayton and W. R. Sistrom (ed.), The photosynthetic bacteria. Plenum Publishing Corp., New York.
38. **Weyer, K. A., F. Lottspeich, H. Gruenberg, F. Lang, D. Oesterhelt, and H. Michel.** 1987. Amino acid sequence of the cytochrome c subunit of the photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis*. EMBO J. **6**:2197–2202.
39. **Young, D. A., C. E. Bauer, J. C. Williams, and B. Marrs.** 1989. Genetic evidence for superoperonal organization for pigments and pigment binding proteins in *Rhodobacter capsulatus*. Mol. Gen. Genet. **218**:1–12.
40. **Youvan, D. C., E. J. Bylina, M. Alberti, H. Begusch, and J. E. Hearst.** 1984. Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-center, B870 antenna and flanking polypeptides from *R. capsulata*. Cell **37**:949–957.
41. **Zhu, Y. S., and J. Hearst.** 1986. Regulation of expression of genes for light-harvesting antenna proteins LHI and LHII; reaction center polypeptides RC-L, RC-M, and RC-H; and enzymes of bacteriochlorophyll and carotenoid biosynthesis in *Rhodobacter capsulatus* by light and oxygen. Proc. Natl. Acad. Sci. USA **83**:7613–7616.
42. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.