# Localization and structural analysis of the ribosomal RNA operons of *Rhodobacter sphaeroides*

Sylvia C.Dryden and Samuel Kaplan*
Department of Microbiology, University of Texas Health Science Center at Houston, Houston, TX 77225, USA

Received September 28, 1990; Revised and Accepted November 7, 1990 EMBL accession nos X53853–X53855

## ABSTRACT

**We have identified, cloned and sequenced the three ribosomal RNA (rRNA) operons (*rrn*) present in the facultative photoheterotroph *Rhodobacter sphaeroides*. DNA sequence analysis has identified the 16S, 23S, and 5S rRNAs, two tRNAs (ile and ala) in the spacer region between the 16S and 23S rRNAs, and an f-met tRNA immediately following the 5S rRNA gene of all three operons. Physical mapping, genetic analysis, and Southern hybridization data indicate that *rrnA* is contained on a large chromosome and *rrnB* and *rrnC* are contained on a second smaller chromosome. These findings are discussed in relation to the origins of diploidy.**

## INTRODUCTION

*Rhodobacter sphaeroides* is a facultative photoheterotrophic bacterium which responds to changes in its environment by undergoing physiological as well as morphological adaptations (1). Under aerobic conditions, the organism possesses a typical Gram-negative cell envelope structure (2). When the oxygen partial pressure decreases below 2.5%, a new membrane system, referred to as the intracytoplasmic membrane (ICM), is induced and is formed as a series of membranous invaginations arising from the cell membrane (1). The ICM contains all components necessary for light energy capture, subsequent electron transport, and energy transduction (3). Induction of the ICM and synthesis of ICM components is regulated at both the transcriptional and post-transcriptional levels. The cellular content of specific mRNA and protein components unique to photosynthetic growth are either not detectable or present at extremely low levels in aerobically growing cells (3). Because of the apparent translational coupling between the synthesis of specific ICM proteins and the availability of small molecule ligands, eg. bacteriochlorophyll (4), it was of interest to investigate those cellular components which could be involved at the level of post-transcriptional control of gene expression.

One component intimately involved in translation is the ribosome complex and its associated ribosomal RNA (rRNA) operons. In *Escherichia coli*, it has been shown that some promoters associated with ribosomal RNA expression are growth rate regulated (5, 6, 7) and in *Plasmodium*, it has been shown that rRNA expression is stage specific (8). Thus, it was of interest to investigate the regulation of the rRNA operons in *R. sphaeroides* to determine if transcription of these operons differs under varying growth conditions and, as such, might play a role in the control of ICM induction as well as the synthesis of ICM components. Concurrent with these studies, physical mapping of *R. sphaeroides* genomic DNA indicated that the arrangement of the *R. sphaeroides* genome was complex, comprised of a large (containing one *rrn* operon) and small (with two *rrn* operons) chromosome as well as several endogenous plasmids (9). Thus, sequence analysis of these *rrn* operons could shed light on the origin of these two chromosomes, as well as providing information on differential transcription. Additionally, we have shown that several other genes are found on both chromosomes (9). In addition, identifying the promoter regions of the rRNA operons was of interest, because there are indications that *R. sphaeroides* promoters are different from those found in *E. coli* (3). Thus, the disposition of the rRNA operons within this genomic context was of importance, as was the sequence relationships between these *rrn* operons.

In addition to the regulatory aspects, analysis of the rRNA operons might yield information on the assembly of the ribosomal subunits. *R. sphaeroides* (10, 11), and a small number of other eubacteria (12, 13), do not contain a mature 23S ribosomal RNA in the 50S subunit, instead, the 50S subunit consists of fragmented rRNAs of varying sizes. In *R. sphaeroides* the mature 50S subunit contains 16S and 14S rRNA species; no mature 23S RNA molecule is found. It was shown that the 16S and 14S rRNAs found in the 50S ribosomal subunit are derived from the cleavage of a 23S precursor molecule (10).

## MATERIALS AND METHODS

### Strains and growth conditions

*R. sphaeroides* wild type strain 2.4.1 was grown in Sistrom's minimal media (14), with tetracycline (1 μg/ml), kanamycin (25 μg/ml), and/or 50 μg/ml spectinomycin and streptomycin when appropriate. Cells were grown at 30°C either aerobically on a rotary shaker or anaerobically in the light in completely filled screw cap tubes. *E. coli* strains, M15A (15), C600 (16), JM101

* To whom correspondence should be addressed

(17), and S17-1 (18) were grown aerobically on a rotary shaker at 37°C using Luria Broth as the growth media. The same concentration of antibiotics was used as in *R. sphaeroides* with the exception of tetracycline which was increased to 10 μg/ml. Matings between *E. coli* strain S17-1 and *R. sphaeroides* were performed as described previously (19).

### DNA isolation and manipulations

Plasmids were isolated as described (20, 21). Chromosomal DNA was isolated as described (22) or using the following technique. Cells (either *R. sphaeroides* or *E. coli*) were grown to approximately mid-log phase, 10 ml of cells were then pelleted and frozen. After freezing, cell pellets were thawed and resuspended in 1 ml of lysis buffer (50 mM Tris, 20 mM EDTA, pH8.0 and 0.5mg lysozyme). The suspension was then divided between two tubes and incubated 2−3 h at 37°C with gentle shaking. Then 0.1 ml of 1 mg/ml proteinase K and 10 μl 10% SDS were added, mixed well, and incubated at 50−55°C for 3−4 h or until viscous. The samples were phenol extracted until the aqueous phase was clear, then extracted 3 to 4 times with ether. If, at this point the aqueous phase was still turbid, flash spinning in a microcentrifuge, would clear the sample. The samples were then ethanol precipitated overnight, dried and resuspended in 50−100 μl of TE (10 mM Tris, 0.1 mM EDTA, pH 8.0) containing 2 μl of 1 mg/ml RNase. Transformations, transfections, Southern hybridizations and other DNA manipulations are described elsewhere (20).

### Cosmid bank construction

The *cos* vector pLA2917 was used (23) and packaging extracts were obtained from Stratagene, La Jolla, CA. *R. sphaeroides* strain 2.4.1 chromosomal DNA was partially digested with *Mbo*I. The digested DNA was sized to yield 30−40 kb fragments and then ligated into the *Bgl*II site of pLA2917. Packaging of the cosmids into phage heads and their subsequent infection into the *E. coli* strain S17-1 were performed as described by the manufacturer.

### DNA sequencing

Dideoxy sequencing was performed using [$^{35}$S] dATP and Sequenase (United States Biochemicals, Inc., Cleveland, OH) following the manufacturers instructions. All DNA sequencing was performed using single stranded M13mp18 and M13mp19 (17). The rRNA operons were digested with appropriate restriction endonucleases to yield 1 to 2.5 kb fragments which were then subcloned into M13 for sequencing. The −40 primer included with the Sequenase kit, as well as the oligonucleotides listed in Table 1 were used as primers. These oligonucleotides were synthesized by Genosys Biotechnologies, Inc., The Woodlands, TX. Additional primers were kindly provided by C.R. Woese, University of Illinois, Urbana, IL.

### Computer analysis

Individual DNA sequences were analyzed and assembled using the Intelligenetics, Inc. PC/gene software program. tRNA secondary structure analysis was performed using the PC/gene programs. For comparisons with other DNA sequences, the Univ. of Wisconsin GCG program (24) and the GenEMBL database were used. Secondary structural analysis of the 16S and 23S molecules of *R. sphaeroides* and *E. coli* were kindly provided by Robin Gutell, Cangene Corp., Mississauga, Ontario, Canada (25).

## RESULTS

### Isolation and subcloning of the rRNA operons

In order to identify the rRNA genes, purified rRNA (10) was used to probe a lambda bank consisting of *R. sphaeroides* chromosomal DNA. One recombinant lambda phage was identified and appeared to contain a portion of an rRNA operon (P. Hallenbeck and S. Kaplan, unpublished results). When a subclone of the lambda phage was utilized as a probe in Southern hybridizations of *R. sphaeroides* bulk genomic DNA, three strong signals were identified in two different restriction endonuclease digestions (Fig. 1). Three signals could be seen in *Bam*HI and *Pvu*II restriction endonuclease digestions as shown in Figure 1, lanes 1 and 4, respectively. Six signals were found in an *Eco*RI digestion (Fig. 1, lane 2), three of these signals were found to be adjacent DNA fragments, as the lambda subclone contained an internal *Eco*RI site in the insert DNA. The fact that there were only two signals in a *Pst*I digestion (Fig. 1, lane 3) could be explained by the fact that there were actually two *Pst*I signals of the same size. The 4.0 kb signal in lane 3 is approximately twice as strong as the 5.5 kb signal in lane 3. A 3.2 kb *Pvu*II fragment (Fig. 1, lane 4), which hybridized with the recombinant lambda phage, was subcloned from bulk *R. sphaeroides* chromosomal DNA and used to screen a cosmid bank of *R. sphaeroides* genomic DNA so that entire rRNA operons could be identified. Numerous cosmids were found that hybridized with the 3.2 kb *Pvu*II fragment as well as the lambda probe, and of these many replicates, three cosmids were chosen for further study. Each of these three cosmids contained one, and only one, of the unique *Bam*HI signals (see Fig. 1, lane 1). The 3.2 kb *Pvu*II DNA fragment which hybridized to the lambda subclone and was common to all three cosmids, was partially sequenced in order to confirm that the rRNA operons had been identified.

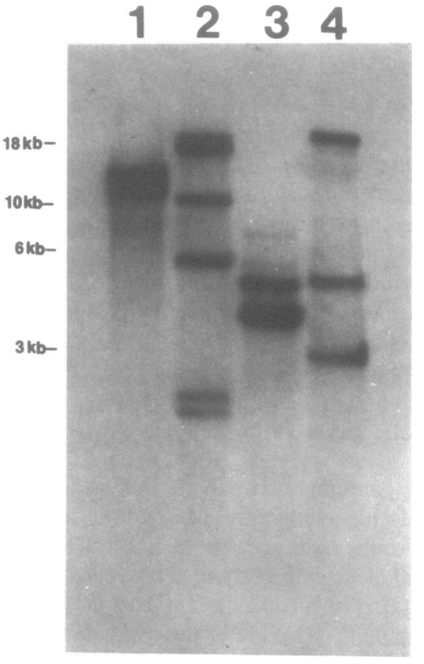



**Figure 1**. Southern hybridization of *R. sphaeroides* chromosomal DNA digested with: lane 1, *Bam*HI; lane 2, *Eco*RI; lane 3, *Pst*I; lane 4, *Pvu*II. Molecular weight standards are indicated on the left side of the figure. The probe utilized for hybridization was a recombinant lambda phage containing a portion of an rRNA operon.

The sequence was then compared to the GenEMBL database to identify and localize the position of this DNA fragment within the rRNA operons. Sequence comparisons indicated that the common *Pvu*II fragment was internal to the rRNA operons, specifically it contained the 5′ end of the 23S rRNA gene. Regions flanking the *Pvu*II fragment were then isolated from the cosmids and subcloned into pUC18 and pUC19 (17). These subclones were digested with a variety of restriction endonucleases to yield DNA fragments between 1 kb and 2.5 kb and ligated into M13 for sequencing of both strands of the DNA, as described in the Materials and Methods.

In addition to sequencing the rRNA operons, the physical map locations of these operons were also identified within the genome of *R. sphaeroides*. The 3.2 kb *Pvu*II DNA fragment, common to all three *rrn*-containing cosmids, was used as a probe and the locations of the *rrn* operons were identified. The operons, designated *rrnA*, *rrnB* and *rrnC*, were found to map onto two different physical linkage groups (9). *rrnA* was located on the large approximately three-megabase chromosome I, whereas *rrnB* and *rrnC* were found to be approximately 35 kb apart and located on the smaller, 914 kb chromosome II. This discovery provided further impetus for a complete sequence determination of all three rRNA operons, since this was the first instance where it might be concluded that a bacterium possesses two chromosomes.

In order to determine if *rrnA*, *rrnB*, and *rrnC* were the only three rRNA operons or if each recognized restriction fragment existed as multiples, Omega cartridges (16), were inserted into each operon (see Fig. 2), introduced into *R. sphaeroides* on the suicide vector pSUP202 (18), and exconjugates were selected and screened for the presence of double cross-over events. The insertion of an approximately 2 kb cartridge into an *rrn* operon resulted in a shift in the electrophoretic mobility of the wild-type signal and the appearance of a signal with a different electrophoretic mobility (Fig. 3, lanes 2–5). Insertions into *rrnB* (Fig. 3, lane 4) or *rrnC* (Fig. 3, lane 3) resulted in a signal 2 kb larger than the wild-type signal due to the presence of the 2 kb omega cartridge. In *rrnA*, due to two *Bam*HI sites within the cartridge, two signals of a lower molecular size were found (Fig. 3, lane 5). *E. coli* chromosomal DNA was used as a control in order to demonstrate that the *R. sphaeroides* 16S rRNA probe used in Fig. 3 would also hybridize to *E. coli* DNA (Fig. 3, lane 1), and the hybridizing signals were presumed to correspond to *E. coli* rRNA operons. The fact that insertion of a cartridge resulted in the loss of the wild type signal, leads to the conclusion that there are only three rRNA operons in *R. sphaeroides*. Thus *rrnA* can be assigned to the 10 kb *Bam*HI signal, *rrnB* to the 14 kb *Bam*HI signal, and *rrnC* to the 13 kb *Bam*HI signal. The biological consequences of the insertion of these cartridges into the *rrn* operon(s) is currently under investigation.

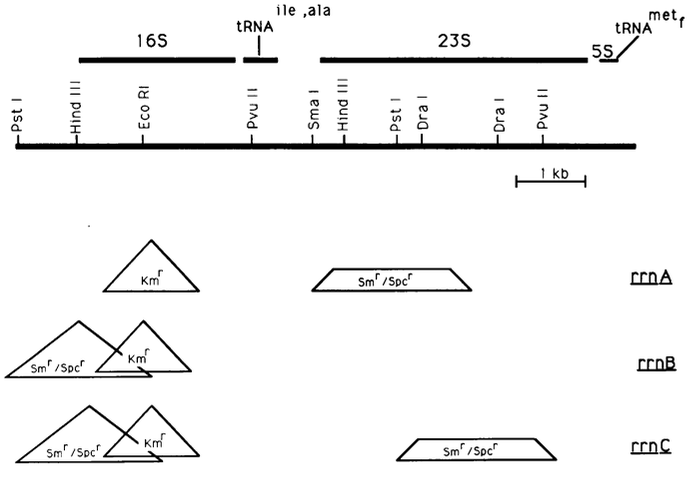



**Figure 2.** Restriction map of the three rRNA operons found in *R. sphaeroides*. The organization of the operons is illustrated above the restriction map. Below are the locations of the different Omega cartridges inserted into each operon. Triangles represent the insertion of a cartridge and the trapezoids represent the deletion of the DNA shown above the trapezoid in the restriction map and then insertion of a cartridge into that region.

## DNA sequence analysis and organization of the *rrn* operons

The double-stranded DNA sequence of each operon was obtained using oligonucleotide primers specific to eubacterial rRNAs (26). In order to sequence both strands of each operon, primers in addition to those provided by C.R. Woese were synthesized (Table 1). The sequences of these oligonucleotides were based on: a) the already elucidated sequence of the opposite strand, or b) *Rhodobacter capsulatus* sequences (GenEMBL Databank, 27). From the DNA sequence (Fig. 4) and comparisons with sequences in the GenEMBL Databank, the organization of each operon was deduced (Fig. 2). Based on comparisons with other eubacterial rRNA sequences, the 5′ and 3′ ends of the mature 16S and 23S rRNAs were identified. In the 16S-23S spacer region two tRNAs, isoleucine and alanine tRNA, were identified. Downstream of the 23S gene, the 5S rRNA gene 5′ and 3′ ends were deduced from secondary structural features by comparison with the proposed structures for *Paracoccus denitrificans* (28) and *R. capsulatus* 5S rRNA (29). Surprisingly, immediately downstream of the 5S rRNA gene, an initiator methionine tRNA

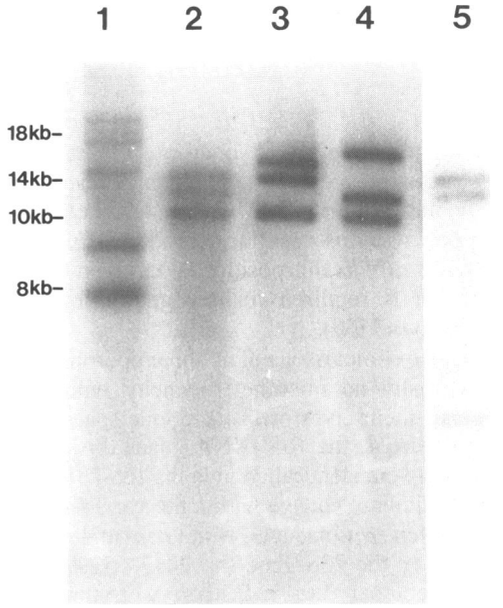



**Figure 3.** Southern hybridization of *E. coli* (lane 1) and *R. sphaeroides* (lanes 2–5) chromosomal DNA probed with a DNA fragment internal to the 16S rRNA genes of the *R. sphaeroides* rRNA operons. Molecular weight standards are indicated on the left side of the figure. DNA in all lanes was digested with *Bam*HI. The strains used are: lane 1, *E. coli* strain M15A; lane 2, *R. sphaeroides* strain 2.4.1 (wild type); lane 3, *R. sphaeroides* 2.4.1 in which a cartridge had been inserted into the chromosomal copy of *rrnC*; lane 4, *R. sphaeroides* 2.4.1 with a cartridge inserted into *rrnB*; lane 5, *R. sphaeroides* 2.4.1 with a cartridge inserted into *rrnA*.

**Table I.** Oligonucleotides synthesized for use as DNA sequencing primers

| Oligo | Direction | Oligo Sequence | Location |
|---|---|---|---|
| 1 | Reverse | (A/G)ATCCGGAAAGAGCA(G/A) | Upstream of 16S |
| 2 | Forward | TAGCTTCGGCGATGAT | Upstream of 16S |
| 3 | Reverse | ATCATCGCCGAAGCTA | Upstream of 16S |
| 4 | Forward | CGCTGGCGGCAGGCCTAA | 16S |
| 5 | Forward | GGGGGAAGATTTATCGGC | 16S |
| 6 | Forward | GGATGATCAGCCACAC | 16S |
| 7 | Forward | GGTAC(G/C)AC(G/C)AGAAGAAG | 16S |
| 8 | Forward | GAACACCAGTGGCGAAG | 16S |
| 9 | Forward | GATCGCACACAGGTGCT | 16S |
| 10 | Forward | GCCGGAGGAAGGTGTG | 16S |
| 11 | Forward | GTAATCGCGTAACAGC | 16S |
| 12 | Forward | GTCGTAACAAGGTAGCCG | 16S |
| 13 | Forward | CCAGGTGATCTAGCCATG | 16S-23S spacer |
| 14 | Reverse | CATGGCTAGATCACCTGG | 16S-23S spacer |
| 15 | Reverse | GCGCTTGCTCTGTCTTG | 16S-23S spacer |
| 16 | Forward | GAATGTACATGCTTCTG | 16S-23S spacer |
| 17 | Reverse | CAGAAGCATGTACATTC | 16S-23S spacer |
| 18 | Forward | AGAGTGACTGGAATGGC | 23S |
| 19 | Reverse | GCTCTACCCCCCGAGG | 23S |
| 20 | Reverse | CACGGAACGTCCGCTACC | 23S |
| 21 | Reverse | CAACATTCTCGCTTC | 23S |
| 22 | Reverse | CCCCAGGGAATATTAA | 23S |
| 23 | Reverse | CTTGGTATTCTCTAC | 23S |
| 24 | Forward | CTCCGTAAGTTCGCG | 23S |
| 25 | Forward | AAGGAGGAGTGCAAGCT | 23S |
| 26 | Forward | CAAGATGCACACTTCCCG | 23S |
| 27 | Forward | GACGCCAGTTCCCGTG | 23S |
| 28 | Forward | AGTTTGACTGGGGCGGTC | 23S |
| 29 | Reverse | GACCGCCCCAGTCAAACT | 23S |
| 30 | Forward | AGGGCCATCGCTCAAC | 23S |
| 31 | Reverse | GTTGAGCGATGGCCCT | 23S |
| 32 | Forward | TATGGCTGTTCGCCATTT | 23S |
| 33 | Reverse | AAATGGCGAACAGCCATA | 23S |
| 34 | Reverse | GGCAACTCTTCTCAAGTC | 23S |
| 35 | Reverse | CTATCGACGTGGTGGTC | 23S |
| 36 | Reverse | GCTTTTGCTTGTTCCGGG | 23S-5S spacer |
| 37 | Forward | CTGCGTCTAAGACGTGGAGAG | 5S |
| 38 | Reverse | CTCTCCACGTCTTAGACGCAG | 5S |
| 39 | Forward | GCGCCAGACCTGAAAAGAAACGAAC | 5S-tRNA spacer |

was found for all three rRNA operons. Figure 5 shows the secondary structure of the f-met tRNA, which revealed a striking feature in the anticodon stem, where an A-U base-pair and not a G-C base-pair was observed in the middle of the stem. In *E. coli* and *Bacillus subtilis* this position is occupied by a third G-C base-pair which is required for recognition by the *E. coli* formylating enzyme (30).

DNA sequence comparisons of all three operons revealed that there was virtually no microheterogeneity within the rRNA structural genes themselves or in intergenic spacer regions. As indicated in Figure 4, the 16S rRNA genes (bp 286−1751) of all three operons were identical. Within the 16S-23S spacer region there was a single base change within the variable region of the ile tRNA, in which *rrnC* had a G, while *rrnA* and *rrnB* contained a C (bp 2021). In the 23S gene (bp 2417−5299), when *rrnA* and *rrnC* were compared to *rrnB*, there were three single base deletions in *rrnB* (shadowed bases in Fig. 4). When *rrnA* and *rrnB* were compared to *rrnC*, there was a single base change in *rrnC* that was not found in either *rrnA* or *rrnB* (bp 5540). There appeared to be more heterogeneity in the 5S rRNA species in which four bases present in the 5S rRNA derived from *rrnA* and *rrnC* were not found in *rrnB*. In the 5S-f-met tRNA spacer region, *rrnA* and *rrnB* were identical, but *rrnC* contained three single base changes when compared to the other two operons. Thus, quantitatively, there was found to be 99.3% homology between operons *A* and *B*, 99.8% between *rrnA* and *rrnC*, and 99.3% homology between *rrnB* and *rrnC*.

The homology between each of the three *rrn* operons extended 285 bp upstream of the start of the mature 16S rRNA. A boxA sequence (31) was identified in this region, but no *E. coli* promoter-like sequences have yet been identified (Fig. 4, bp 1−285). Interestingly, in *rrnA* and *rrnC*, the DNA sequences showed considerable similarity for an additional 118 bp upstream of the 285 bp, as indicated in Figure 6, whereas the analogous region in *rrnB* was entirely different from the other operons. The DNA sequences immediately following the end of the f-met tRNA gene in all three operons also showed no similarity (Fig. 7). However, within these downstream regions, potential stem-loop structures indicative of rho-independent transcription terminators have been identified for all three operons (Fig. 7). Finally, the upstream and downstream regions of the *rrn* operons were analyzed for open reading frames (ORF) and only one possible ORF was identified, located immediately downstream of the f-met tRNA gene present in *rrnA*. Analysis of this region by the method of Fickett (32) yielded a probability for coding of 40%, however no likely Shine-Dalgarno sequence was identified upstream of this ORF. In addition, the proposed transcription terminator of *rrnA* was within this potential ORF. Because of these observations, we feel that the probability of translation of this ORF is low.

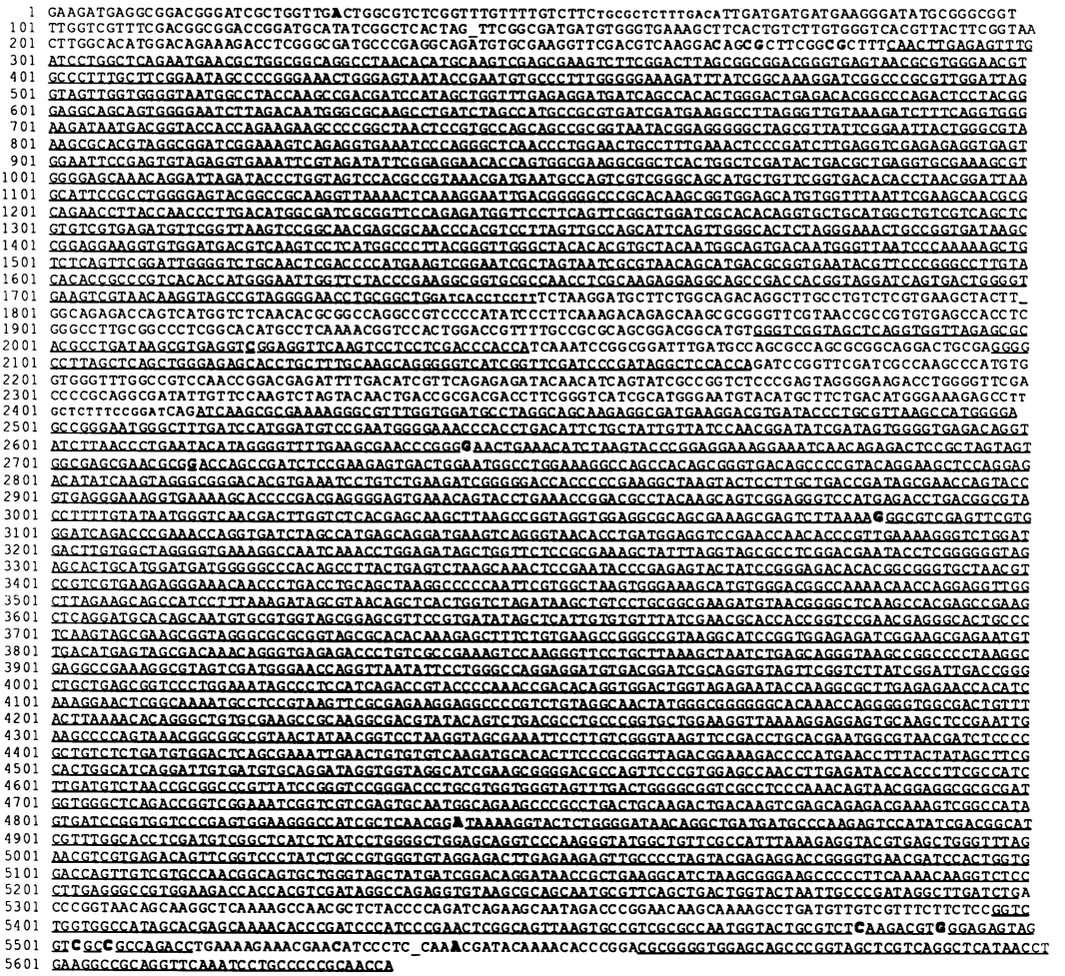


**Figure 4**. DNA sequence of *rrnA* from *R. sphaeroides* strain 2.4.1. Shown here is the region of homology between the 3 operons beginning 287 bp 5′ to the start of the mature 16S rRNA and ending with the last base of the f-met tRNA. Nucleotides in bold print are different in *rrnB* or *rrnC*: at bp30 *rrnB* has a T, at bp59 *rrnB* has a T, at bp270 *rrnB* has a T, at bp273,274 *rrnB* has TC, at bp281,282 *rrnB* has GA, at bp2021 *rrnC* has G, at bp2715 *rrnC* has a C, at bp5529 *rrnC* has a T. A (−) between 2 bases indicates an insertion in an operon: at bp143−144, insertion of C in *rrnB* and *rrnC*, at bp5536−5537, insertion of T in *rrnC*. A single base that is shadowed indicates that this base has been deleted, all deletions occur in *rrnB* except for the deletion at bp5540 which occurs in *rrnC*. The boxA sequences (upstream of the 16S and 23S genes) and the anti-Shine-Dalgarno at the end of the 16S are in small print. The genes within the operons are underlined: bp286−1751 16S, bp1973−2049 ile tRNA, bp2097−2172 ala tRNA, bp2417−5299 23S, bp5397−5514 5S, bp5559−5635 f-met tRNA. EMBL Data Library accession numbers for the complete sequence of each operon are: X53853 for *rrnA*, X54854 for *rrnB* and X53855 for *rrnC*.

## 16S, 23S and 5S rRNA Secondary Structure

In addition to analyzing the primary sequences of the three *rrn* operons, secondary structures of the 16S and 23S rRNAs were derived by comparative analysis. When secondary structures of the 16S and 23S rRNAs were generated, differences between the typical eubacterial rRNAs and *R. sphaeroides* rRNAs could be visualized. Figure 8 shows the proposed secondary structure of the 16S rRNA generated from the sequence of *rrnB*. When this secondary structure was compared with the corresponding *E. coli* secondary structure, the regions displaying variation were all at common sites of structural variation, as determined by comparative analysis of eubacterial 16S rRNAs (33−35), except perhaps for the 2 regions boxed in Figure 8. In the case of box 1 (Fig. 8), the helix is usually observed to be shorter or single stranded as compared to the large stem-loop seen in *R. sphaeroides*. Comparative analysis in the region where box 2 is located has indicated that a region of conserved base-pairing is deleted in the *R. sphaeroides* 16S secondary structure (34).

Structural variations are much more apparent for the secondary

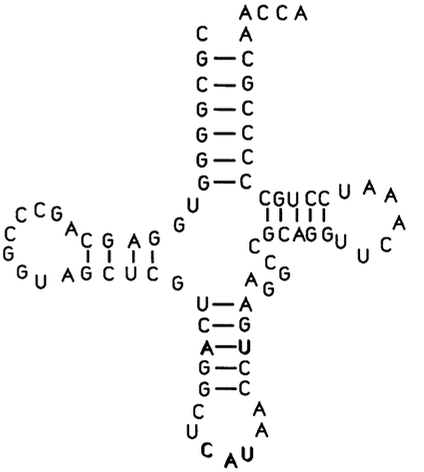


**Figure 5**. Secondary structure of the f-met tRNA. The unusual A-U basepair is shadowed, as is the anticodon.

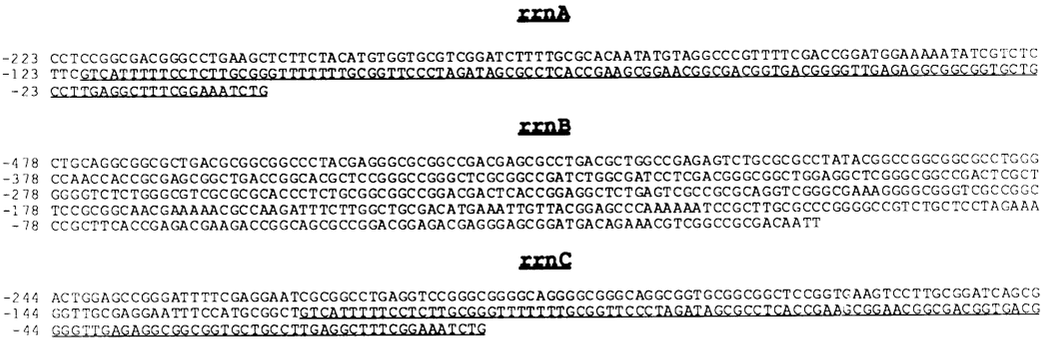



**Figure 6.** DNA sequences 5′ of nucleotide +1 in the DNA sequence shown in Figure 4. The underlined sequences in *rrnA* and *rrnC* are a region of additional homology which is not present in *rrnB*.

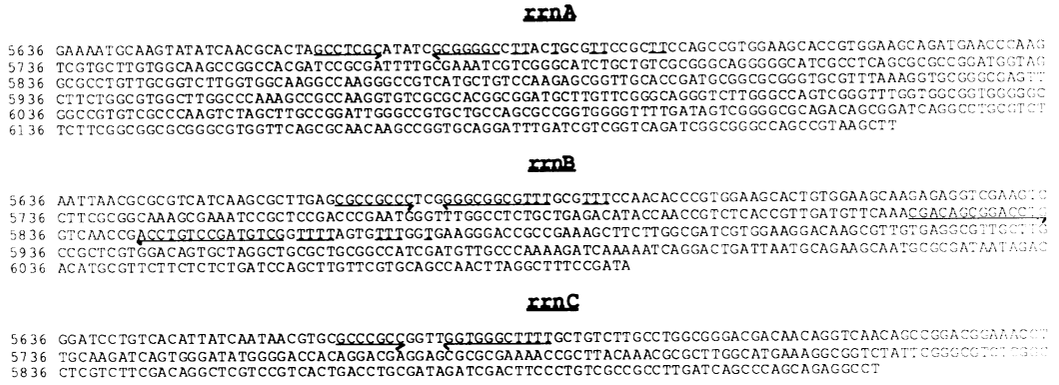



**Figure 7.** DNA sequences 3′ of the f-met tRNA in the rRNA operons. Potential rho-independent terminators are indicated by the arrows, the T's which normally follow these terminators are underlined.

structures generated from the 23S rRNA from *R. sphaeroides* when compared to that of *E. coli* (Fig. 9, A and B), especially in the 5′ half of the molecule. The function and/or biological significance of many of these regions of dissimilarity are unknown, however, the possible functions of two of the stem-loop regions (boxed regions in Fig. 9) are proposed below. In *R. sphaeroides*, there is an additional stem-loop structure located at what is the equivalent of position 1200 in the *E. coli* secondary structure. Previous data suggest that this extra stem-loop is a processing site for the cleavage of the precursor 23S molecule into a 16S and 14S species; these smaller species are found in the mature 50S ribosomal subunit (10). Oligonucleotide mapping has shown that a 14S specific oligonucleotide hybridizes just upstream of the stem-loop region and a 16S specific oligonucleotide hybridizes about 60 bp following the stem-loop structure (36). The second stem-loop structure of interest is located from 131 to 148 on the *E. coli* secondary structure. In *R. sphaeroides*, this is an extended stem of 55 bp vs. the 17 bp found in *E. coli*. There has been some suggestion that this may be another processing site, which would yield a third small RNA species of about 5.8S (C.R. Woese, personal communication). Mapping of the ends of the 14S, 16S, and putative 5.8S species is currently underway. Many of the other structural differences, such as the lack, in *R. sphaeroides*, of the extended stem from bp 1405 to 1600 found in *E. coli*, are also shared by *R. capsulatus*, a very closely related bacterium (25).

Due to the fact that *rrnB* had a deletion of four basepairs in the 5S rRNA gene when compared to *rrnA* and *rrnC*, two secondary structures for the 5S rRNA could be derived (Fig. 10, A and B). The secondary structure thus derived for *rrnB* contained two shortened helices due to the base deletions (Fig. 10 B).

## DISCUSSION

The data presented here show that *Rhodobacter sphaeroides* has only three ribosomal RNA operons, unlike *E. coli* or *B. subtilis* which contain seven and ten rRNA operons, respectively (37, 38). There is variability in numbers of rRNA operons as well as genome size, and there appears to be no correlation between number of rRNA operons within a bacterial species and the physical size of the genome. Physical mapping of genomic DNA of *Anabaena* strain PCC 7120 indicates that there are only two rRNA operons contained within a 6.37 Mb genome (39), whereas mapping of *Haemophilus influenzae* Rd indicates that within the 1.9 Mb genome there are six rRNA operons (40). The operons of *R. sphaeroides* have a typical eubacterial organization of 16S, spacer region containing tRNAs, 23S, and 5S genes. The distinctive aspects of the operons are: first, within all three operons, the two tRNAs encoded in the 16S-23S spacer region are ile and ala, which differs from *E. coli* where different operons contain different tRNAs in the 16S-23S spacer regions (37). Second, all three operons encode an f-met tRNA immediately downstream of the 5S rRNA gene. There are no apparent rho-independent termination signals (37) found in the 90 nt spacer region between the 5S rRNA and f-met tRNAs, suggesting that

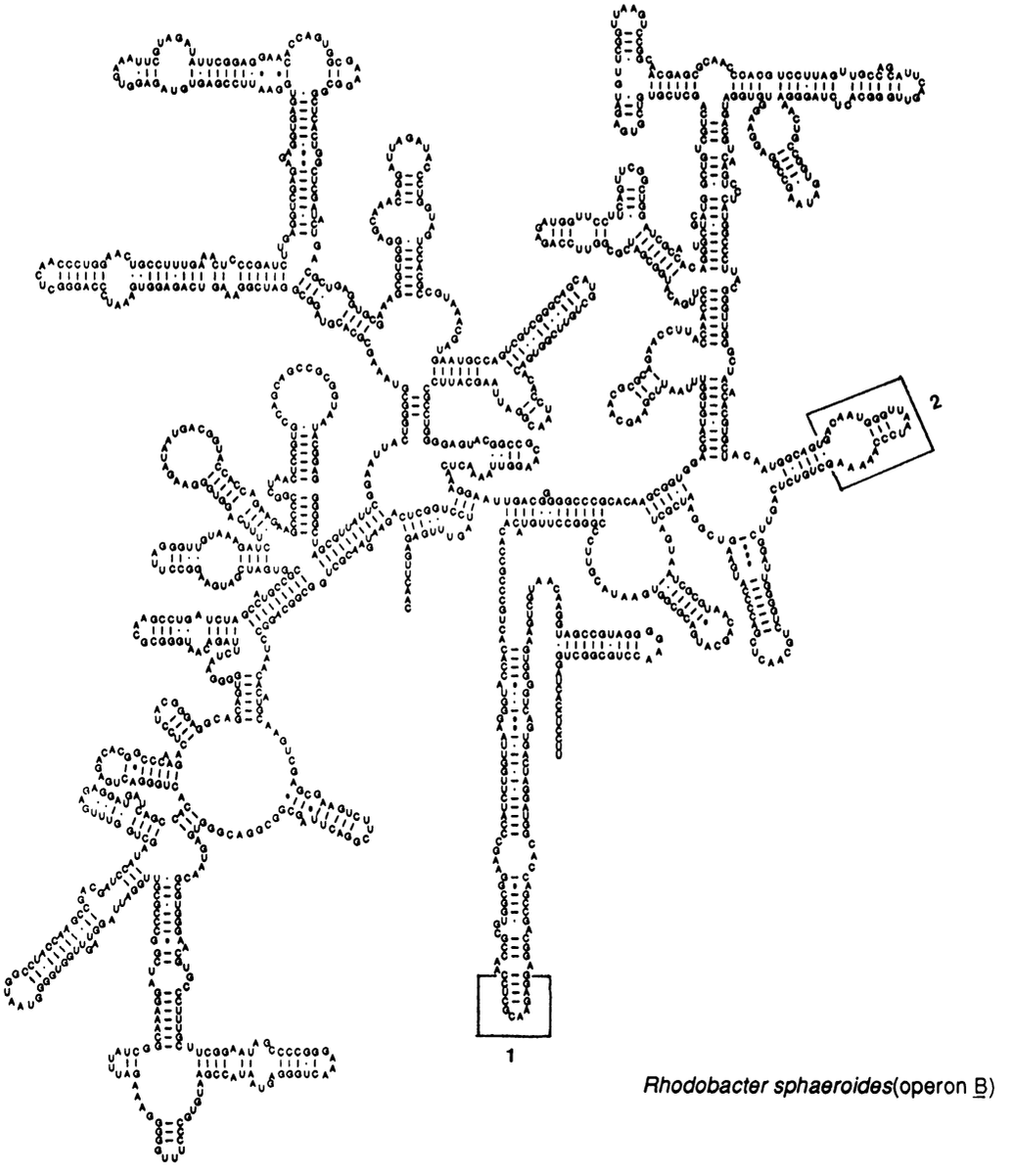


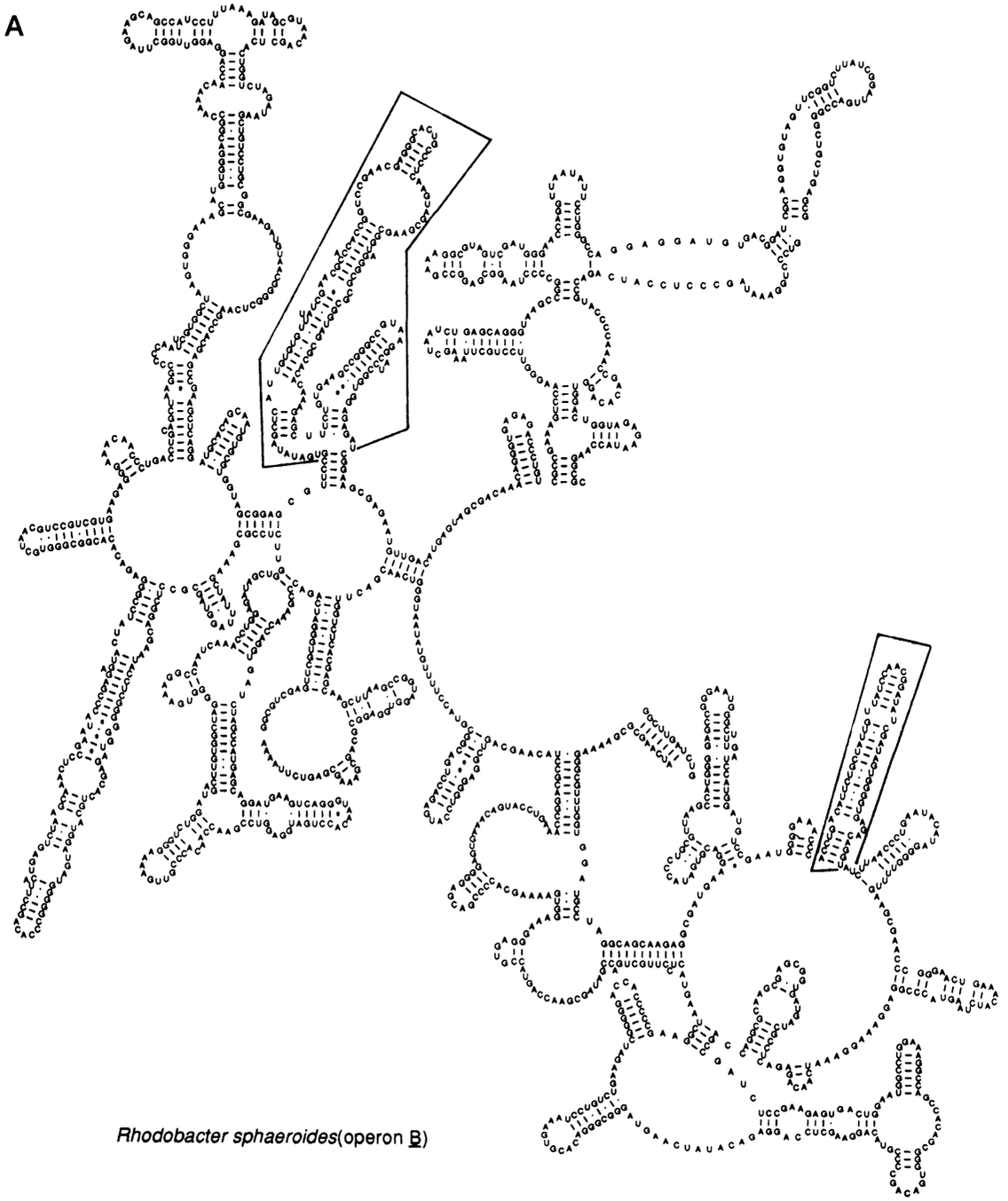

**Figure 8**. Secondary structure of 16S rRNA of the rRNA operons. The boxed regions may be uncommon structural variations.

the f-met tRNA is a part of each of the rRNA operons and thus also a part of that transcriptional unit. In other eubacteria so far studied, there are no known f-met tRNAs that comprise a part of an rRNA operon. In *E. coli* three rRNA operons contain one (*rrnD* and *rrnH*), or two (*rrnC*) tRNAs, located immediately downstream of the 5S rRNA gene, which are believed to be transcribed with the rRNA genes (37). *B. subtilis* contains large clusters of tRNA genes downstream of some of the rRNA operons, one of which does contain an f-met tRNA, however, these tRNA gene clusters are believed to be transcribed seperately from the rRNA operons (41). Third, it appears that *R. sphaeroides* is very unusual in that physical mapping and Southern hybridization data show that two of the rRNA operons (*rrnB* and *rrnC*) are contained within a 914 kb circular linkage group, while the third operon (*rrnA*) is found on the large, 3.045 megabase chromosome I (9). Studies are currently in progress to delete one or more operons in various combinations involving the small chromosome, as well as the large chromosome. Initial data, in which Omega cartridges have been inserted into each operon and then recombined back into the chromosome (Fig.2 and Fig. 3) indicate that there are phenotypical changes associated with the mutants.

At the primary sequence level, the *R. sphaeroides* 16S and 23S rRNA genes show little microheterogeneity between the operons, with all three 16S rRNA genes being identical and, of the few differences between the 23S rRNA genes, only one (at bp 2715 in *rrnC*, indicated in Fig. 4, the equivalent position in E. coli is position 261; Fig. 9B) disrupts the base pairing of a helical structure. The 5S rRNA gene of *rrnB* shows increased heterogeneity when compared to the 5S rRNA genes from operons *A* and *C*, but this is not unusual since in *E. coli* there are base substitutions found when comparing the different 5S genes within a strain (42). Although operons *A* and *C* have identical 5S rRNA gene sequences, operons *A* and *B* have identical 5S-f-met tRNA spacer regions, whereas *rrnC* contains a base change, an insertion and a deletion within the same spacer

region. The observation that the very few differences between the operons are primarily deletions or insertions of a single base is intriguing, since the diversity described in 5S rRNA in *E. coli* strains are simple base changes (42) which do not disrupt base-pairing in helical regions. Deletions and insertions may be more likely to disrupt secondary and/or tertiary interactions. No information was available in the GenEMBL database to allow primary sequence comparisons of 16S and 23S rRNA genes in order to analyze the degree of microheterogeneity among the 7 *E. coli* rRNA operons as a comparison with the number and type (substitution, deletion, or insertion) of base changes found in *R. sphaeroides*.

The presence of virtually identical rRNA operons on different circular linkage groups in *R. sphaeroides* suggests that their derivation might have been attributable to extensive recombination between homologs, resulting in the absence of micro-heterogeneity, if we assume no sequence-associated selective advantage. Recombination between two distinct chromosomes would be less likely to result in the deleterious loss of intervening DNA than would recombination between regions on the same chromosome. Similar to the arrangement of rRNA operons, a number of genetic regions in *R. sphaeroides* have counterparts located on both linkage groups (9). Therefore, we might consider the possibility that the origin of 'diploidy' reflects the advantageous opportunity for diversity created by extensive chromosomal duplication, while limiting the possible loss of critical DNA were duplications to be maintained within a single chromosome. In addition, if homologs on different chromosomes have different physiological roles, then the duplication of genes might be required for normal cell growth, and maintenance of



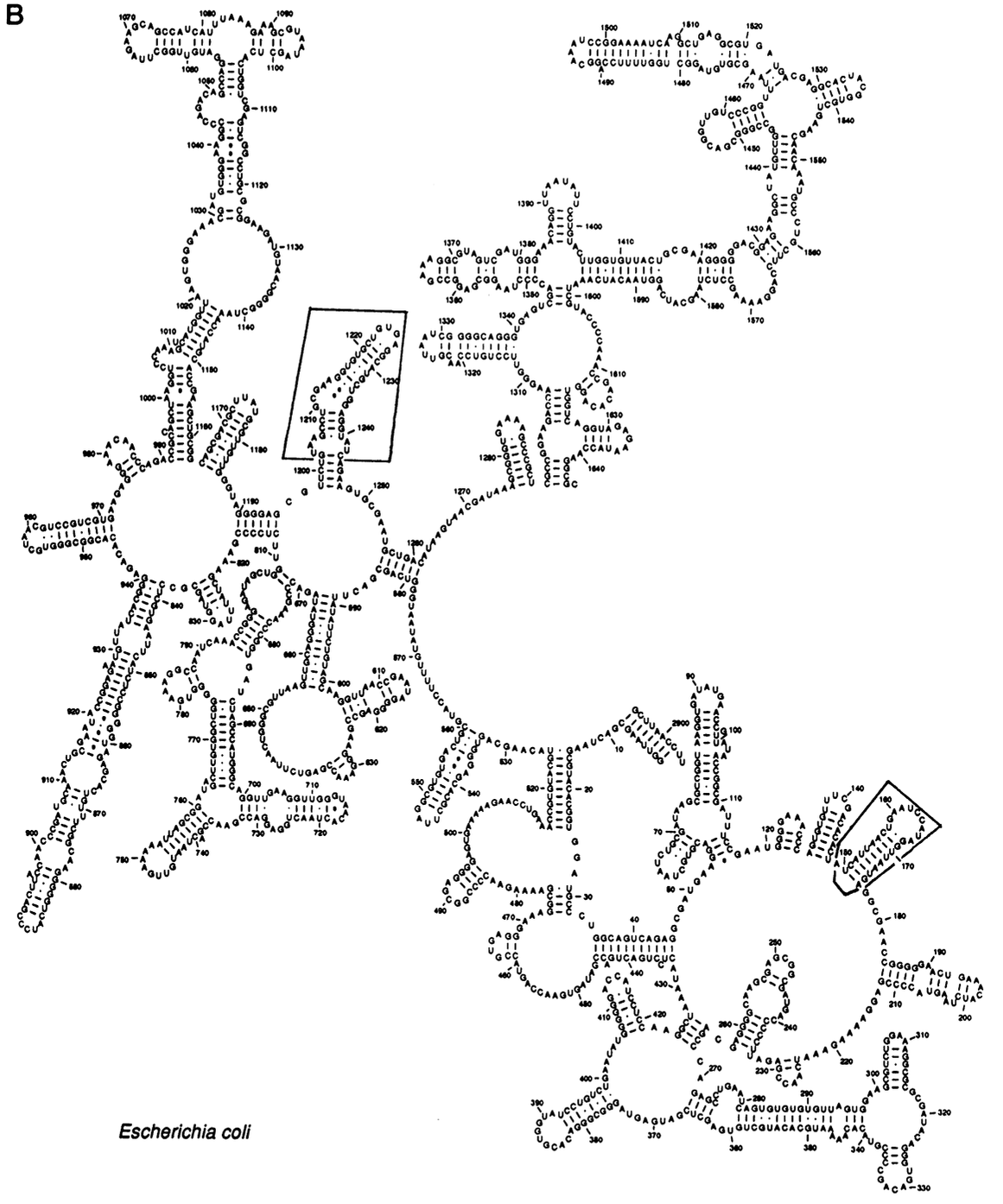



**Figure 9**. Secondary structures of the 5′ half of the 23S rRNA from *R. sphaeroides* (A) and *E. coli* (B). The 2 boxed regions are potential RNase processing sites in *R. sphaeroides*, the analogous non-processed regions found in *E. coli* are boxed for reference.

these genes would be under constant selection. However, if drift in the DNA sequence homology between duplicate genes occurs, then the existence of duplicate genes can be maintained within a single linkage group. *R. sphaeroides*, because of its skewed codon usage (70% G+C), is probably constrained in its ability to undergo DNA sequence drift. Thus, the origin of a second linkage group is a way out of the dilemma.

The fact that identity between the three operons continues for an additional 287 bp upstream of the start of the mature 16S rRNA, strongly suggests that regulatory sequences may be present in this region. In addition, *rrnA* and *rrnC* show identity for an additional 118 bp 5′ to the possible regulatory region. This additional DNA sequence similarity in what could be additional regulatory sequences or a second promoter region, raises the possibility that operons *A* and *C* may share similar regulatory activity (note that these 2 operons are on different chromosomes), whereas regulation of transcription of *rrnB* might differ. Preliminary results have identified the probable 5′ end of the primary rRNA transcript (Dryden and Kaplan, unpublished results). Upstream of the transcriptional start site potential −10 and −35 sequences have been identified. These sequences, CCTAGA (−10) and TTGCG (−35) are spaced identically in all

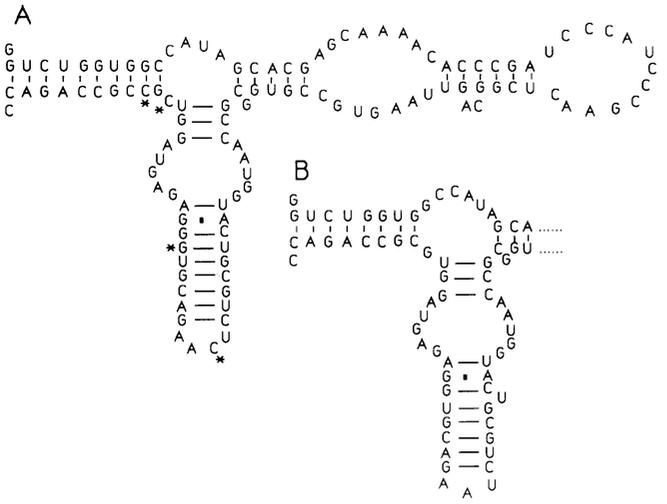



**Figure 10**. Secondary structures of the 5S rRNA. G/A basepairs are joined by a square instead of a (−). Panel A: The secondary structure derived for *rrnA* and *rrnC*. The bases that are deleted in *rrnB* are starred (*). Panel B: Secondary structure of the first 3 helices in *rrnB*, which are different due to the deleted bases.

three rRNA operons upstream of the transcriptional start site. Studies are currently underway to confirm these observations.

Analysis of *R. sphaeroides* rRNA secondary structures is also revealing. The structure of the 16S rRNA conforms with the very typical eubacterial structure, placing this organism within the alpha subdivision of the purple bacteria (or as Stackebrandt et al. have suggested, the Proteobacteria [43]), whereas *E. coli* falls into the gamma subdivision (33). However, this is the first analysis at the DNA sequence level of all the operons present in *R. sphaeroides*. Analysis of the two 5S rRNA structures (Fig. 10) yield similarities in the first four helices to the 5S rRNA structure proposed for *P. denitrificans* (28), especially in the length of the single stranded regions surrounding these helices. However, the 5th helix of *R. sphaeroides* is more similar in both primary and secondary sequence to the 5S rRNA of *R. capsulatus*. These three bacterial species are related, in that *R. capsulatus* is the type species for the genus *Rhodobacter* (44), and *P. denitrificans* is a nonphotosynthetic member of the purple bacterial subgroup alpha-3 to which the *Rhodobacter* genus belongs (33). Very preliminary results suggest that *P. denitrificans*, like *R. sphaeroides* contains two large linkage groups, while *R. capsulatus* may only have a single chromosome (A. Suwanto and S. Kaplan, unpublished data). These data need to be confirmed with physical mapping of their respective genomes, but these data and the similarities in 5S primary and secondary structures may suggest a closer relationship between *P. denitrificans* and *R. sphaeroides* than previously thought.

Analysis of the f-met tRNA located at the end of each rRNA operon in *R. sphaeroides* (Fig. 5) indicates that most of the sites which are unique to met-tRNAs which can be formylated with *E. coli* enzyme are present in *R. sphaeroides* f-met tRNA (45). The one site that is different in the *R. sphaeroides* f-met tRNA is the A-U base-pair in the anticodon stem shadowed in Fig. 5. Previously, it was shown that mutations, in which one, two, or all three of the G-C base-pairs found in this helix in *E. coli* were mutated, decreased initiation and also affected conformation of the anticodon loop (30). Studies in which the initiation of protein synthesis by the *R. sphaeroides* f-met tRNA is analyzed both in *R. sphaeroides* and other cell systems may yield more information on the specificity of f-met tRNA and the requirement for three G-C basepairs in the anticodon stem of the *R. sphaeroides* f-met tRNA.

The 23S rRNA secondary structure of *R. sphaeroides* has some very interesting features. First, sequence analysis has identified a probable RNase processing site which supports earlier observations that there are two species of RNA present in the 50S ribosome, a 16S and a 14S RNA species, and that the 23S rRNA is merely a precursor molecule (10, 36). Furthermore, a possible second processing site which would yield yet a third RNA species of approximately 5.8S has been identified. This additional processing site is very reminiscent of eukaryotic cytoplasmic rRNAs, in which the 5′ portion of the large RNA species is cleaved to yield a 5.8S species and remains a part of the large ribosomal subunit (34). A comparison of the 23S rRNA secondary structures of *R. capsulatus* (25) and *R. sphaeroides* shows that they are very similar and *R. capsulatus* 23S rRNA also contains the two probable processing sites. When primary sequence comparisons were performed, *R. capsulatus* showed 92.7% similarity to the *R. sphaeroides* DNA sequence, except in the two processing sites which had little similarity in DNA sequence. Thus in *R. capsulatus*, as well as in *R. sphaeroides*, there may be three RNA species noncovalently associating in the 50S ribosomal subunit, all resulting from processing of the 23S rRNA precursor. It should be noted that previously described processing sites in large subunit rRNA molecules (12, 46−48) are not located in the same position as the 14S-16S processing site found in *R. sphaeroides* and *R. capsulatus*. Further studies, such as mapping of the ends of these smaller rRNA species may yield new information on the maturation and assembly of the 50S ribosomal RNA subunit in *R. sphaeroides*.

In conclusion, the DNA sequence of the three rRNA operons found in *R. sphaeroides* can further the identification of conserved helices in 16S and 23S rRNA secondary structure through comparative analysis and further define minimal structures of these molecules (34, 35). In addition, further analysis of the primary and secondary structure of these three operons, as well as further investigation of the presence of two chromosomes in *R. sphaeroides* (9), may yield more information on the relationship of the alpha-3 subdivision of the purple bacteria and eukaryotic mitochondria (33). Since the alpha-purples have been postulated as the precursor organisms to mitochondria (26, 33), such information could lead to a further understanding of the steps involved in the evolution of the mitochondrion. Further investigation of the RNase processing of the 23S rRNA of *R. sphaeroides* and other organisms (12) which results in noncovalently associated RNA species in the 50S ribosome, as is found in eukaryotic cytoplasmic, mitochondria and chloroplast rRNAs (46−48), may help in understanding of the evolution of the ribosome and perhaps yield information on the evolution of intervening sequences and introns. In addition, the construction and characterization of mutants in the rRNA operons of *R. sphaeroides*, as well as analysis of the putative promoter regions of these operons may add to the understanding of the regulatory control governing the transcription and processing of the rRNA genes in *R. sphaeroides*.

## ACKNOWLEDGEMENTS

We would like to thank R. Gutell for generating the secondary structures of the 16S and 23S rRNAs and C.R. Woese for providing oligonucleotide primers for DNA sequencing, as well

as helpful discussions. In addition, we thank A. Varga, E. Neidle, A. Suwanto and P. Hallenbeck for their helpful discussions and assistance in making this work possible. This work was supported by a grant from the US National Institutes of Health (GM31667) to S.K.

## REFERENCES

1. Kaplan,S. (1981) *Photochem. Photobiol.*, **34**, 769–774.
2. Baumgardner,D., Deal,C. and Kaplan,S. (1980) *J. Bacteriol.*, **143**, 265–273.
3. Kiley,P.J. and Kaplan,S. (1988) *Microbiol. Reviews*, **52**, 50–69.
4. Takemoto,J. and Lascelles,J. (1973) *Proc. Natl. Acad. Sci. USA*, **70**, 799–803.
5. Dickson,R.R., Gaal,T., DeBoer,H.A., DeHaseth,P.L. and Gourse,R.L. (1989) *J. Bacteriol.*, **171**, 4862–4870.
6. Gourse,R.L., de Boer,H.A., and Nomura,M. (1986) *Cell*, **44**, 197–205.
7. Nomura,M., Gourse,R. and Baughman,G. (1984) *Ann. Rev. Biochem.*, **53**, 75–117.
8. Gunderson, J.H., Sogin,M.L., Wollet,G., Hollingdale,M., de la Cruz,V.F., Water,A.P. and McCutchan,T.F. (1987) *Science*, **238**, 933–937.
9. Suwanto,A. and Kaplan,S. (1989) *J. Bacteriol.*, **171**, 5850–5859.
10. Marrs,B., and Kaplan,S. (1970) *J. Mol. Biol.*, **49**, 297–317.
11. Lessie,T.G. (1965) *J. Gen. Microbiol.*, **39**, 311–320.
12. Burgin,A.B., Parodos,K., Lane,D.J. and Pace,N.R. (1990) *Cell*, **60**, 405–414.
13. Pace,N.R. (1973) *Bacteriol. Rev.*, **37**, 562–603.
14. Sistrom,W.R. (1962) *J. Gen. Microbiol.*, **28**, 607–616.
15. Ruther,U., Koenen,M., Otto,K. and Huller-Lill,B. (1981) *Nucleic Acids Res.*, **9**, 4087–4097.
16. Prentki,P. and Krisch,H.M. (1984) *Gene*, **29**, 303–313.
17. Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
18. Simon,R., Priefer,U., and Puhler,A. (1983) *Bio/Technology*, **1**, 37–45.
19. Davis,J., Donohue,T.and Kaplan,S. (1988) *J. Bacteriol.*, **170**, 320–329.
20. Maniatis,T., Fritsch,E.F., and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor.
21. Morelle,G. (1989) *Focus*, **11**, 7–8.
22. Giuliano,G., Pollock,D., Stapp,H., and Scolnik,P. (1988) *Mol. Gen. Genet.*, **213**, 78–83.
23. Allen,L.N. and Hanson,R.S. (1985) *J. Bacteriol.*, **161**, 955–962.
24. Devereaux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
25. Gutell,R.R., Schnare,M.N. and Gray,M.W. (1990) *Nucleic Acids Res.*, **18**, 2319–2330.
26. Yang,D., Oyaizu,Y., Oyaizu,H., Olsen,G.J. and Woese,C.R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4443–4447.
27. Hopfl,P., Ludwig,W. and Schleifer,K.H. (1988) *Nucleic Acids Res.*, **16**, 2343.
28. MacKay,R.M., Salgado,D., Bonen,L., Stackebrandt,E. and Doolittle,W.F. (1982) *Nucleic Acids Res.*, **10**, 2963–2970.
29. Van den Eynde,H., Vandenberghe,A. and De Wachter,R. (1986) *Nucleic Acids Res.*, **14**, 8688.
30. Seong,B.L. and RahBhandary,U.L. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 334–338.
31. Berg,K.L., Squires,C. and Squires,C.L. (1989) *J. Mol. Biol.*, **209**, 345–358.
32. Fickett, J.W. (1982) *Nucleic Acids Res.*, **10**, 5303–5318.
33. Woese,C.R. (1987) *Microbiol. Rev.*, **51**, 221–271.
34. Noller,H.F. (1984) *Ann. Rev. Biochem.*, **53**, 119–162.
35. Woese,C.R., Gutell,R., Gupta,R. and Noller,H.F. (1983) *Microbiol. Rev.*, **45**, 621–669.
36. MacKay,R.M., Zablen,L.B., Woese,C.R., and Doolittle,W.F. (1979) *Arch. Microbiol.*,**123**, 165–172.
37. Jinks-Robertson,S. and Nomura,M. (1987) In Neidhardt,F.C.(Ed.), *Escherichia coli and Salmonella Typhimurium, Cellular and Molecular Biology*, American Society for Microbiology, Washington, DC, Vol. 2, pp.1358–1385.
38. Jarvis,E.D., Widom,R., LaFauci,G., Setoguche,U., Richter,I. and Rudner,R. (1988) *Genetics*, **120**, 625–635.
39. Bancroft,I., Wolk,C.P. and Oren,E.V. (1989) *J. Bacteriol.*, **171**, 5990–5948.
40. Lee,J.L., Smith,H.O. and Redfield,R.J. (1989) *J. Bacteriol.*, **171**, 3016–3024.
41. Wawrousek,E.F., Narasimhan,N. and Hansen,J.N. (1984) *J. Biol. Chem.*, **259**, 3694–3702.
42. Jarry,B. and Rosset,R. (1971) *Mol. Gen. Genetics*, **113**, 43–50 41.
43. Stackebrandt,E., Murray,R.G.E. and Truper,H.G. (1988) *Int. J. Syst. Bacteriol.*, **38**, 321–325.
44. Imhoff,J.F., Truper,G.H. and Pfennig,N. (1984) *Int. J. Syst. Bacteriol.*, **34**, 340–343.
45. Yamada,Y., Kucheno,Y. and Ishikura,H. (1980) *J. Biochem.*, **87**, 1261–1269.
46. Boer,H.P. and Gray,M.W. (1988) *Cell*, **55**, 399–411.
47. Campbell,D.A., Kubo,K., Clark,C.G. and Boothroyd,J.C. (1987) *J. Mol. Biol.*, **196**, 113–124.
48. Spencer,D.F., Collings,J.C., Schnare,M.N. and Gray,M.W. (1987) *EMBO J.*, **6**, 1063–1071.