JOURNAL OF BACTERIOLOGY, Oct. 1992, p. 6479–6487
0021-9193/92/206479-09$02.00/0
Copyright © 1992, American Society for Microbiology
Vol. 174, No. 20

# Identification of an Operon Involved in Sulfolipid Biosynthesis in *Rhodobacter sphaeroides*

C. BENNING* AND C. R. SOMERVILLE

*MSU-DOE-Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824-1312*

Received 17 March 1992/Accepted 10 August 1992

**Two new mutants of *Rhodobacter sphaeroides* deficient in sulfolipid accumulation were isolated by directly screening mutagenized cell lines for polar lipid composition by thin-layer chromatography of lipid extracts. A genomic clone which complemented the mutations in these two lines, but not the previously described sulfolipid-deficient *sqdA* mutant, was identified. Sequence analysis of the relevant region of the clone revealed three, in tandem open reading frames, designated *sqdB*, ORF2, and *sqdC*. One of the mutants was complemented by the *sqdB* gene, and the other was complemented by the *sqdC* gene. Insertional inactivation of *sqdB* also inactivated *sqdC*, indicating that *sqdB* and *sqdC* are cotranscribed. The N-terminal region of the 46-kDa putative protein encoded by the *sqdB* gene showed slight homology to UDP-glucose epimerase from various organisms. The 30-kDa putative protein encoded by ORF2 showed very striking homology to rabbit muscle glycogenin, a UDP-glucose utilizing, autoglycosylating glycosyltransferase. The 26-kDa putative protein encoded by the *sqdC* gene was not homologous to any protein of known function.**

Sulfoquinovosyl diacylglycerol, commonly called sulfolipid, is thought to occur in the photosynthetic lamellae of all photosynthetic organisms (14). The abundance varies from about 2.5% of ether-extractable lipids in species such as *Rhodobacter sphaeroides* (26) up to 40% of glycolipids in marine red algae (11). Thus, this compound represents a quantitatively important component of the global sulfur cycle (15), yet we do not know how this lipid is synthesized or what role it plays in photosynthetic organisms.

In an attempt to address the questions about the biosynthesis and the function of the sulfolipid, we have initiated a genetic dissection of the biosynthetic pathway in the purple bacterium *R. sphaeroides*. We have previously described the isolation of sulfolipid-deficient mutants of *R. sphaeroides* by directly analyzing the polar lipid composition of mutagenized cell lines by thin-layer chromatography. In addition a gene, designated *sqdA*, which complemented one of the mutants was isolated and characterized (7).

Here, we describe the isolation of two additional genes which complement two different classes of mutants with defects in sulfolipid biosynthesis in *R. sphaeroides*. These two genes are in two complementation groups different from that previously described for the *sqdA* gene and they flank a third open reading frame for which no mutant has yet been isolated. Two of these three open reading frames have limited deduced amino acid sequence homology with proteins of known function, suggesting possible roles in the pathway of sulfolipid biosynthesis.

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** Bacterial strains and plasmids used in this study are shown in Table 1.

**Media and growth conditions.** *R. sphaeroides* cell cultures were grown in the malate-basal-salt medium described by Ormerod et al. (22) or in Sistrom's succinate-basal-salt medium (30, 31) as described previously (7). When required, 0.8 μg of tetracycline per ml was added.

*Escherichia coli* strains were grown on Luria broth. Antibiotics were added to the medium at the following final concentrations (in micrograms per milliliter): tetracycline, 10; kanamycin, 50; ampicillin, 100; and streptomycin, 20.

**Extraction and analysis of lipids.** Lipid extracts were prepared from cell cultures grown photoheterotrophically in Sistrom's medium, and polar lipids were extracted and separated by one-dimensional thin-layer chromatography on ammonium sulfate-treated silica plates as previously described (7). The lipids were visualized by charring.

**Time course for the incorporation of [$^{35}$S]sulfate into sulfolipid.** Cells grown photoheterotrophically were labelled with [$^{35}$S]sulfate and extracted, and the incorporation of sulfate into sulfolipid was determined as described previously (7).

**Triparental mating and complementation assay.** To test for complementation, individual clones, derivatives of either pLA2917 or pCHB500 harbored by *E. coli* HB101, were transferred into the mutant strains by triparental mating with the helper plasmid pRK2013 in *E. coli* MM294. The sulfolipid phenotype of the exconjugants was tested by extraction of the lipids, followed by thin-layer chromatography analysis on ammonium sulfate-impregnated plates as previously described (7).

**Determination of CAT activity.** *R. sphaeroides* cells harboring various plasmid constructs containing the chloramphenicol acetyltransferase (CAT) gene were grown photoheterotrophically in 200 ml of Sistrom's medium. Cells were harvested by centrifugation at 2,000 × *g* and resuspended in 5 ml of enzyme buffer (0.5 M sucrose, 5 mM ascorbic acid, 0.1 M Tris-Cl [pH 7.8], 5 mM dithiothreitol). After the addition of 0.5 mg of DNase I, the cells were broken in a French press at 18,000 lb/in$^2$. The protein concentration in the supernatant was determined by using the Bradford microassay (8). To assay CAT activity, 40 μl of enzyme was mixed with 5 μl of 0.25 M Tris-Cl (pH 8.0), 1 μl of 10 mM acetyl coenzyme A, and 0.5 μl of [$^{14}$C]chloramphenicol (57 μCi/mmol, 25 μCi/ml; NEN), and the mixture was incubated at 37°C for 1 h. After extraction with 300 μl of ethyl acetate, the organic phase was concentrated and spotted onto a Si250 Silica TLC plate (Baker) and activated for 15 min at 120°C. The plates were developed with chloroform-methanol (95:5 [vol/vol]), and radioactivity was determined by scanning

* Corresponding author.

TABLE 1. Bacterial strains and plasmids

| Strain or plasmid | Description or construction | Source or reference |
|---|---|---|
| Strains | | |
| *R. sphaeroides* 2.4.1 | Wild type | R. L. Uffen |
| *R. sphaeroides* CHB16 | Sulfolipid deficient, *sqdB* | This study |
| *R. sphaeroides* CHB18 | Sulfolipid deficient, *sqdC* | This study |
| *E. coli* HB101 | $F^-$ *proA2 recA13 mcrB* | 21 |
| *E. coli* DH5αF′ | F′ *endA1 recA1* Δ(*lacZYA-argF*)*U169* [φ80d*lac*Δ(*lacZ*)*M15*] | 27 |
| *E. coli* MM294 | $F^-$ *endA1 hsdR17 thi-1* | 2 |
| Plasmids | | |
| pBS-KS$^+$, pBS-KS$^-$ | Amp$^r$, Bluescript | Stratagene |
| pRK2013 | Kan$^r$ Tra$^+$ RK2-ColE1$_{rep}$ | 12 |
| pLA2917 | Kan$^r$ Tc$^r$ *cos* | 1 |
| pCHB500 | Tc$^r$; expression vector for *R. sphaeroides* | 7 |
| pBR328 | Cm$^r$ Amp$^r$ Tc$^r$ | 3 |
| pHP45Ω | Amp$^r$ Sm$^r$ Spc$^r$; Ω cassette | 25 |
| pCHB1601 | Cosmid clone complementing the *sqdB* and *sqdC* mutants | This study |
| pCHB1611 | pCHB1601 digested with *Hin*dIII and recircularized | This study |
| pCHB1621 | 11-kb *Hin*dIII fragment of pCHB1601 in *Hin*dIII of pLA2917 | This study |
| pCHB16014 | 2.4-kb fragment of pCHB1601 in pLA2917 complementing the *sqdB* mutant | This study |
| pCHB16014D-B | pCHB16014 cut with *Bgl*II and recircularized | This study |
| pCHB16014D-AB | pCHB16014 cut with *Bam*HI and *Bgl*II and recircularized | This study |
| pCHB16201 | 2-kb *Sal*I-*Eco*RI partial fragment of pCHB16014 cloned into pBS-KS$^-$, excised with *Kpn*I and *Eco*RI and cloned into pCHB500 | This study |
| pCHB16202 | 2-kb *Sal*I-*Eco*RI partial fragment of pCHB16014 cloned into pBS-KS$^-$; excised with *Kpn*I and *Xba*I and cloned into pCHB500 | This study |
| pCHB16151 | 1.5-kb *Sal*I-*Eco*RI fragment of pCHB16014 cloned into pBS-KS$^-$; excised with *Kpn*I and *Eco*RI and cloned into pCHB500 | This study |
| pCHB16152 | 1.5-kb *Sal*I-*Eco*RI fragment of pCHB16014 cloned into pBS-KS$^-$; excised with *Kpn*I and *Xba*I and cloned into pCHB500 | This study |
| pCHB16014CAT-1 | 1-kb *Sau*3AI fragment of pBR328 containing the CAT gene coding sequence, cloned into the *Bam*HI site of pCHB16014 | This study |
| pCHB16014CAT-2 | Same as pCHB16014CAT-1, except opposite orientation of the CAT coding sequence | This study |
| pCHB181 | 4.2-kb *Bgl*II fragment of pCHB1601 cloned into the *Bam*HI site of pCHB500 | This study |
| pCHB182 | Same as pCHB181, except opposite orientation of the insert | This study |
| pCHB183 | 1.2-kb *Xba*I-*Bam*HI partial fragment of pCHB181 cloned into pBS-KS$^+$; excised with *Kpn*I and *Xba*I and cloned into pCHB500 | This study |
| pCHB184 | 1.5-kb *Xba*I-*Bam*HI partial fragment of pCHB181 cloned into pBS-KS$^+$; excised with *Kpn*I and *Xba*I and cloned into pCHB500 | This study |
| pCHB185 | 1.5-kb *Xba*I-*Bam*HI partial fragment of pCHB181 cloned into pBS-KS$^+$; excised with *Kpn*I and *Xba*I and cloned into pCHB500 | This study |
| pCHB186 | 2.3-kb *Pst*I-*Eco*RI fragment of pCHB181 cloned into pCHB500 | This study |
| pCHB187 | 1.2-kb *Pst*I-*Bam*HI partial fragment of pCHB181 cloned into pCHB500 | This study |
| pCHB188 | 2.0-kb *Eco*RI-*Bam*HI partial fragment of pCHB181 cloned into pBS-KS$^+$; excised with *Xba*I and *Eco*RI and cloned into pCHB500 | This study |
| pCHB189 | 1.7-kb *Eco*RI-*Bam*HI partial fragment of pCHB181 cloned into pBS-KS$^+$; excised with *Xba*I and *Eco*RI and cloned into pCHB500 | This study |
| pCHB160018 | 2.0-kb partial *Pst*I fragment of pCHB185 ligated in proper orientation with 22-kb *Pst*I fragment of pCHB16014 | This study |
| pCHB160018::Ω | 2.0-kb *Bam*HI Ω cassette of pHP45Ω inserted into the *Bam*HI site of pCHB16014; the 24-kb *Pst*I fragment of this clone was ligated in proper orientation with the 2.0-kb *Pst*I partial fragment of pCHB185 | This study |

individual lanes with a System 200 Imaging Scanner (Bioscan). The enzyme concentration was adjusted, such that only monoacetylated product was formed and not more than 80% of the substrate was converted.

**DNA manipulations.** Restriction maps were constructed by analysis of CsCl gradient-purified DNA according to established methods (21). DNA fragments separated on agarose gels were blotted onto Hybond-N$^+$ (Amersham) nylon membranes by using the alkaline transfer method, and the DNA was hybridized by the method of Reed and Mann (28).

**Sequence analysis.** DNA sequences cloned into pBS-KS plasmids were determined by employing the chain termination reaction method (29) with a modified T7 DNA polymerase (32) or sequencing grade *Taq* polymerase (17). To reduce sequencing artifacts due to the high GC content of *R. sphaeroides* DNA, GTP was replaced by 7-deaza-dGTP (4). The labelled DNA fragments were separated on 6% polyacrylamide–8 M urea slab gels (23). Oligonucleotides used as sequencing primers were synthesized on an Applied Biosystems 380A DNA synthesizer by using the phosphoramidite method (5).

**Nucleotide sequence accession number.** The DNA sequence was deposited in GenBank with the accession number M89780. Protein coding regions were predicted from the DNA sequence by codon preference analysis by the method of Gribskov et al. (13). A codon usage table for *R. sphaeroi-*

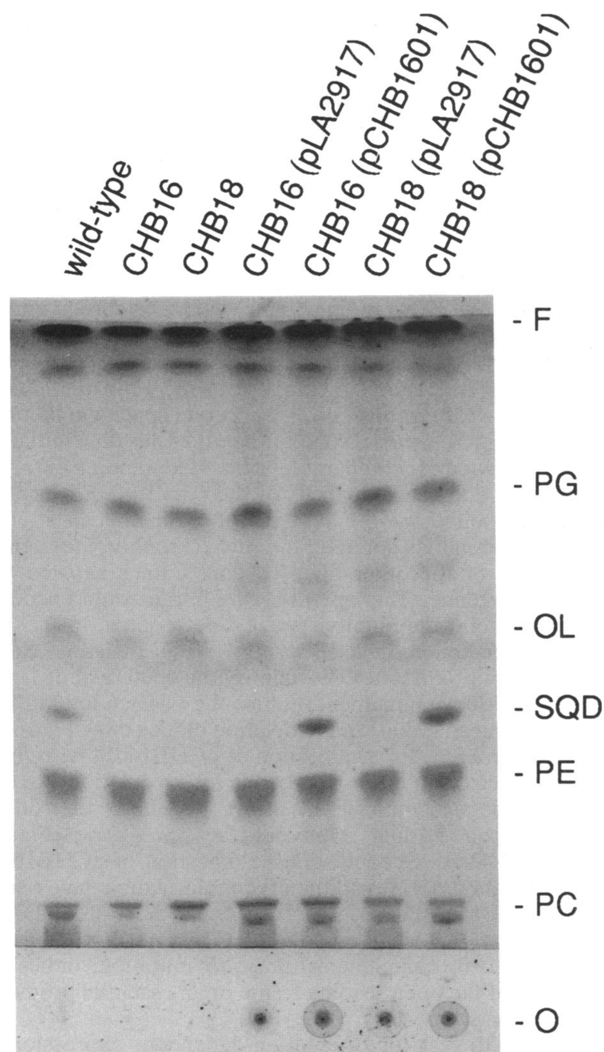


FIG. 1. Lipid composition of the wild-type and the CHB16 and CHB18 mutant lines. Lipid extracts were prepared from different strains lacking or harboring either the cosmid vector pLA2917 or the complementing cosmid pCHB1601 as indicated above each lane. The cells were grown photoheterotrophically, and the lipids were separated by thin-layer chromatography on an ammonium sulfate-impregnated silica gel plate. The lipids were visualized by charring. F, solvent front; PG, phosphatidylglycerol; OL, sulfolipid; PE, phosphatidylethanolamine; PC, phosphatidylcholine; O, origin.

des was prepared by translating *R. sphaeroides* DNA sequences available from GenBank, including a total number of 3,865 codons.

## RESULTS

**The primary phenotype of two sulfolipid-deficient mutant lines of *R. sphaeroides*.** During the initial screen for sulfolipid-deficient mutants of *R. sphaeroides* (7), two lines, designated CHB16 and CHB18, which both have strongly reduced amounts of sulfolipid as revealed by thin-layer chromatography of the lipid extracts (Fig. 1), were identified. However, upon labelling of the cells with [$^{35}$S]sulfate, low but significant incorporation of sulfate into sulfolipid was observed for the mutant line CHB18, while incorporation of labelled

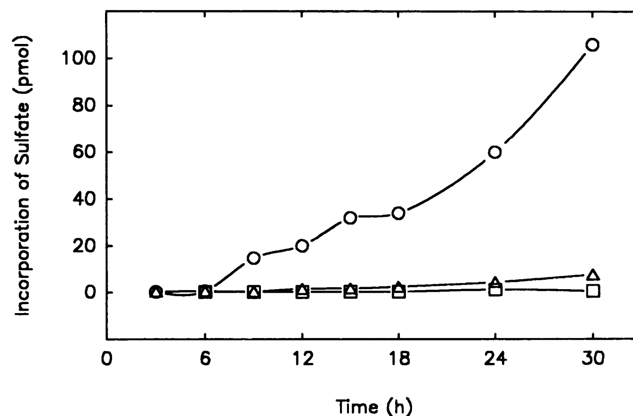


FIG. 2. Time course for the incorporation of $^{35}$S-labelled sulfate into sulfolipid by the wild type (○), the *sqdB* mutant (□), and the *sqdC* mutant (△).

sulfate into sulfolipid by mutant line CHB16 was not detectable (Fig. 2). Therefore, mutant line CHB18 carries a leaky mutation, which strongly reduces the rate of sulfate incorporation into sulfolipid.

Both mutants are capable of photoheterotrophic growth on Sistrom's and Ormerod's media which do not contain any reduced sulfur compounds such as cysteine or methionine. In addition, [$^{35}$S]sulfate was incorporated into protein at a similar rate compared with that for the wild type (6), indicating that these two mutants are not deficient in the sulfur assimilation pathway or the biosynthesis of sulfur-containing amino acids. In this respect, these two new sulfolipid-deficient mutants resemble the previously described *sqdA* mutant (7). Therefore, both mutants appear to carry mutations which specifically affect the biosynthesis of sulfolipid.

**Isolation and characterization of a cosmid clone complementing the two mutants.** Individual cosmids from a library of wild-type *R. sphaeroides* (7) were transferred by triparental mating from *E. coli* into the sulfolipid-deficient mutant line CHB16. The polar lipid composition of approximately 800 exconjugants was examined by thin-layer chromatography in order to identify one cosmid (pCHB1601), which complemented the sulfolipid deficiency in the CHB16 mutant line (Fig. 1). To test whether cosmid pCHB1601 could complement other sulfolipid-deficient mutants, it was mated into the previously described *sqdA* mutant and into mutant line CHB18. Testing of the exconjugants for restoration of the ability to synthesize sulfolipid revealed that the cosmid pCHB1601 also complemented the CHB18 mutant line (Fig. 1) but the *sqdA* mutant was not complemented. This result indicated that the *sqdA* gene does not belong to the same complementation group as the defective gene (or genes) in the two new mutants. The question of whether the two mutations in the mutant lines CHB16 and CHB18 are allelic was answered through the isolation of two distinctive classes of subclones of pCHB1601 complementing the two mutants as described below.

On the basis of restriction analysis, the length of the cosmid pCHB1601 was estimated to be about 48 kb. A centrally located *Hin*dIII site allowed the isolation of the right half of the cosmid as an 11-kb *Hin*dIII fragment, which was subcloned into pLA2917 to give pCHB1621 (Fig. 3A). Recircularization of the remaining plasmid gave rise to pCHB1611, which contained the left half of pCHB1601 (Fig.

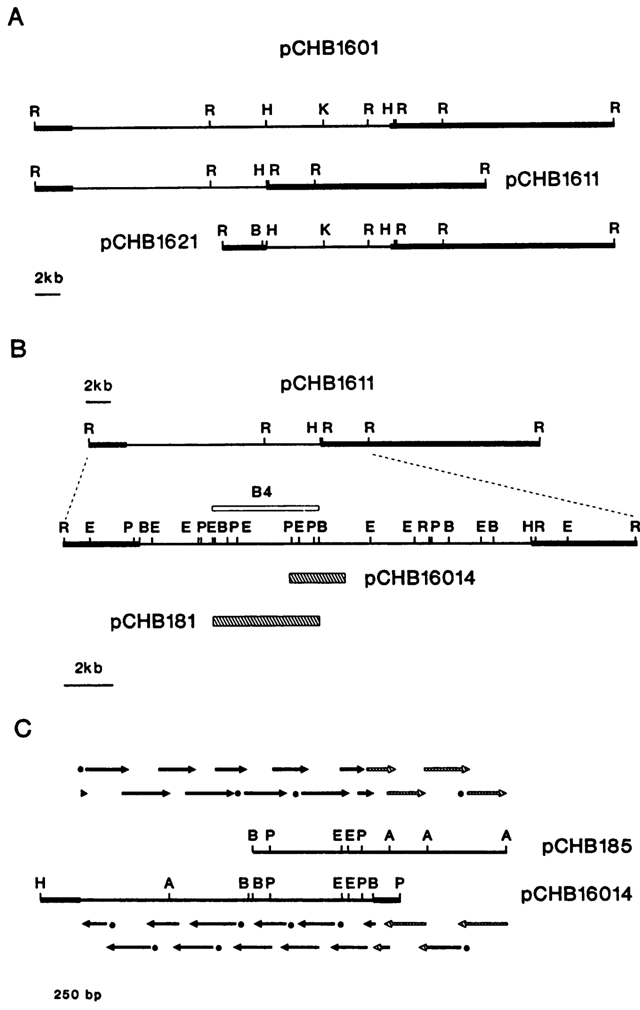


FIG. 3. Restriction maps of cosmid pCHB1601 and its derivatives. (A) Overview of restriction maps for pCHB1601 and the two subclones pCHB1611 and pCHB1621. (B) Detailed restriction map of pCHB1611; the locations of the two subclones pCHB16014 and pCHB181 complementing the mutant lines CHB16 and CHB18, respectively, are indicated as cross-hatched boxes. B4 (open box) represents the location of a 4.3-kb *Bgl*II fragment carrying the *sqdC* gene. (C) Sequencing strategy used to sequence pCHB16014 and parts of pCHB185. Closed arrows indicate sequencing runs obtained by sequencing the pCHB16014 template and cross-hatched arrows indicate runs obtained by sequencing the pCHB185 template. Sequencing runs preceded by a dot were obtained by using synthetic oligonucleotide primers. Vector sequences are drawn with thick lines. Restriction sites: A, *Bam*HI; B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; R, *Eco*RV.

3A). These two subclones were subjected to further restriction and Southern analysis to provide detailed maps for pCHB1611 (Fig. 3B) and pCHB1621.

**Isolation and sequence analysis of a small subclone of pCHB1601, complementing the CHB16 mutant line.** To isolate the gene, tentatively designated *sqdB*, which is defective in the CHB16 mutant line, 2- to 3-kb fragments of pCHB1601 from a *Sau*3A partial digest were cloned into pLA2917. Of 96 clones from this sublibrary, 5 restored the wild-type phenotype when mated into the CHB16 mutant line. The smallest positive clone, pCHB16014, contained an insert of about 2.4 kb. A 2.9-kb fragment containing the insert as well as flanking pLA2917 sequences was excised by using a *Hin*dIII-*Pst*I partial digest, cloned into pBluescript and sequenced (Fig. 3C).

Analysis of the DNA sequence of pCHB16014 (Fig. 4, bases 1 to 2380) revealed that four of the possible six frames contained a relatively large number of stop codons distributed throughout the sequence so that only relatively small open reading frames were present; one frame lacked a stop codon over the entire length of the pCHB16014 insert, and one frame contained one stop codon resulting in a large open reading frame. This reading frame was considered promising, because one of the in-frame ATG initiation codons was preceded by a putative ribosome binding site, showing the AGGAGG motif complementary to the 3′ end of the 16S rRNA of *R. sphaeroides* as well as correct spacing relative to the ATG (Fig. 4, *sqdB*). None of the other possible initiation codons showed a comparable ribosome binding site. In addition, codon usage analysis of the putative *sqdB* protein coding sequence by the method of Gribskov et al. (13) revealed codon bias common to other *R. sphaeroides* genes, while none of the other open reading frames showed this codon preference. This open reading frame could encode a 46-kDa protein, the putative *sqdB* gene product.

**Complementation analysis using various deletion clones derived from pCHB16014.** To identify the open reading frame responsible for complementation of the CHB16 mutant line by genetic means, two simple deletion clones of pCHB16014 were constructed. First, digestion of pCHB16014 with *Bgl*II and then recircularization resulted in a 1-kb deletion expanding from the right end of the insert into the 3′ end of the putative open reading frame for the *sqdB* gene (Fig. 5, pCHB16014D-B). Second, double digestion of pCHB16014 with *Bgl*II and *Bam*HI and then recircularization gave rise to clone pCHB16014D-AB, which carried a deletion of 1.7 kb beginning at the right insert end (Fig. 5). Both deletion clones failed to complement the *sqdB* mutation, indicating that parts of the open reading frame or an essential promoter sequence was lost.

In order to test for orientation-dependent expression of the putative *sqdB* open reading frame, various fragments of the pCHB16014 insert were cloned behind a cytochrome *c* promoter from *Rhodobacter capsulatus* located on the pCHB500 vector (7). Two *Sal*I sites, one upstream of the putative initiation codon and one downstream in the putative coding region, were particularly useful for the analysis. A 1.9-kb fragment spanning the region from the upstream *Sal*I site to the nearest *Eco*RI site downstream of the putative *sqdB* open reading frame was obtained by a partial digest. Insertion of the fragment in both orientations into pCHB500 behind the *cycA* promoter resulted in two clones, pCHB16201 and pCHB16202 (Fig. 5), which were both able to complement the *sqdB* mutant. On the other hand, insertion of a shorter fragment of 1.6 kb from the second *Sal*I site to the *Eco*RI site nearest to the 3′ end of the putative *sqdB* open reading frame in both orientations into pCHB500 gave rise to two clones, pCHB16151 and PCHB16152 (Fig. 5), which were unable to complement the defect in the *sqdB* mutant. Therefore, the defective gene in the *sqdB* mutant must be located within the 1.9-kb *Sal*I-*Eco*RI fragment of pCHB16014. In addition, this fragment must contain an endogenous promoter driving the expression of the putative gene, since complementation was independent of the orientation of the insert relative to the external promoter provided by the pCHB500 vector. The fact that the smaller *Sal*I-*Eco*RI fragment cloned in both orientations into pCHB500

FIG. 4. Composite nucleotide sequence of pCHB16014 and pCHB185. The deduced amino acid sequences of the putative *sqdB* and *sqdC* genes and ORF2 are shown beneath the coding strand of the nucleotide sequence. The designations for the open reading frames are indicated above the proposed initiation codons. *Bam*HI sites which were used for subcloning are listed above the sequence. A putative ribosome binding site (SD) for the *sqdB* gene is indicated above the sequence.

failed to provide sufficient sequence for complementation was consistent with the idea that this fragment did not include the N-terminal portion of the putative open reading frame as shown in Fig. 5 and that it was missing a putative promoter proposed to be located on the 350-bp fragment between the two *Sal*I sites.

**Expression of a promoterless CAT gene inserted into pCHB16014.** To determine the predominant orientation of gene expression and, thereby, the orientation of the *sqdB* gene on pCHB16014, a 0.9-kb *Sau*3AI fragment containing the CAT coding sequence but not a promoter was excised from pBR328 and cloned in both orientations into the unique *Bam*HI site of pCHB16014. The two resulting clones, pCHB16014CAT-1 and -2 (Fig. 6A), were mated into the *sqdB* mutant along with the pLA2917 vector and pCHB16014. Upon testing of the lipid composition of the resulting strains, it was noted that both clones containing the insertions lost the ability to complement the defective *sqdB* gene, indicating that the insertion resulted in the disruption of the *sqdB* gene located on pCHB16014. Measurements of the CAT activity in extracts obtained from the four cell lines described above revealed that the presence of the clone pCHB16014CAT-1 in the *sqdB* mutant resulted in 10-fold-higher activity compared with that for pCHB16014CAT-2. No CAT activity could be detected in extracts obtained from cells harboring pLA2917 or pCHB16014 (Fig. 6B). Since in pCHB16014CAT-1 the orientation of the CAT coding sequence was the same as for the putative *sqdB* coding

FIG. 5. Complementation analysis of the *sqdB* mutant by using various subclones of pCHB16014. The putative open reading frame for the *sqdB* gene is indicated above the pCHB16014 sequence (the pCHB prefix of the clone names was omitted). The complementation result is indicated as absence (−) or presence (+) of the sulfolipid (SL). Vector sequences derived from pLA2917 are drawn with a thicker line, while sequences derived from the pBluescript polylinker are shown as open boxes. The location and direction of the cytochrome *c* promoter (Pcyc) located on pCHB500 are indicated by an arrow. Restriction sites: A, *Bam*HI; B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; O, *Xho*I; P, *Pst*I; S, *Sal*I; X, *Xba*I; *, sites which were lost during the cloning procedure.

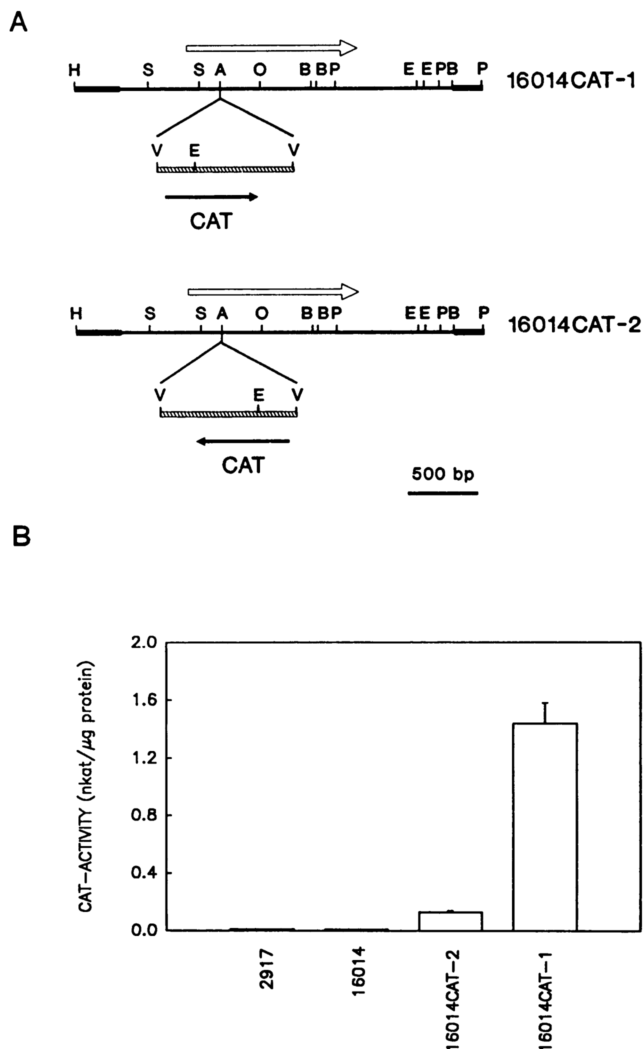


FIG. 6. CAT insertion clones of pCHB16014 and enzyme activity observed in cells containing these clones. (A) Location and orientation of the fragment containing the CAT coding sequence (filled box) in clones pCHB16014CAT-1 and -2. The exact location and orientation of the putative *sqdB* open reading frame and the coding sequence for the CAT gene are indicated by arrows. Vector sequences derived from pLA2917 are drawn with a thicker line. Restriction sites: A, *Bam*HI; B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; O, *Xho*I; P, *Pst*I; V, *Sau*3AI. (B) CAT activity measured in extracts from *sqdB* mutant cells containing the plasmids pLA2917, pCHB16014, pCHB16014CAT-1, and pCHB16014CAT-1. Three replicates were measured and averaged for each of the extracts, and the calculated standard errors are indicated.

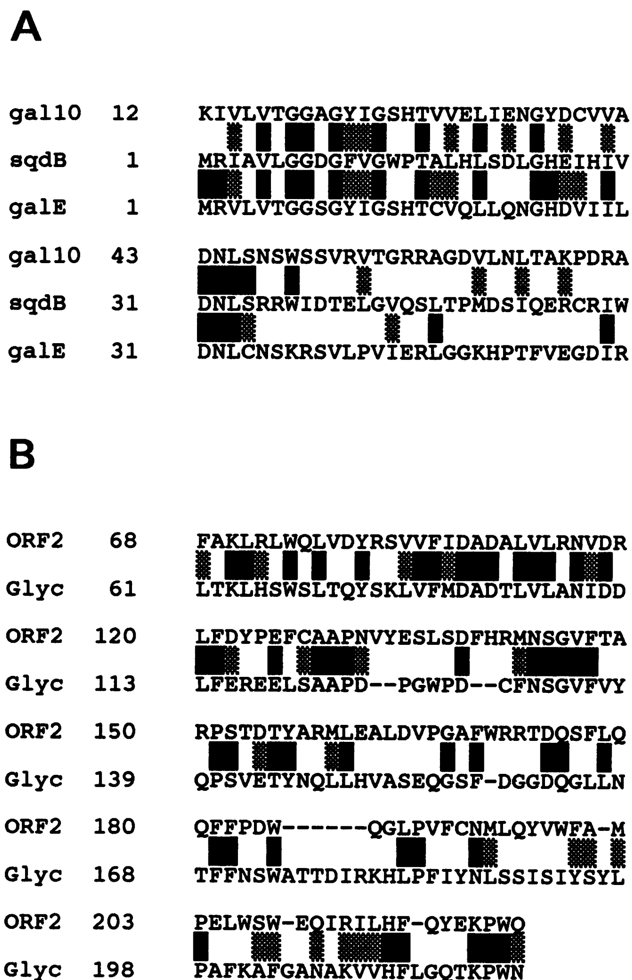


FIG. 7. Sequence homology between the predicted amino acid sequences of the *sqdB* gene and ORF2 and proteins of known function. (A) Comparison of the N-terminal amino acid sequence of the putative *sqdB* gene product and UDP-glucose epimerases from *Saccharomyces cerevisiae* (10) and *E. coli* (20). The amino acid sequences were translated from the nucleotide sequence of the *gal10* gene of *S. cerevisiae* (top), the *sqdB* gene (center), and the *galE* gene of *E. coli* (bottom). (B) Comparison of the putative amino acid sequence of ORF2 and rabbit skeletal muscle glycogenin (9). Dotted boxes indicate conservative substitutions, while solid boxes represent perfect identity.

sequence, it was concluded that the coding sequence depicted in Fig. 4 represents the *sqdB* gene. It must be noted, however, that small but significant amounts of CAT activity were observed when the CAT coding sequence and the putative *sqdB* coding sequence were oriented in opposite direction in pCHB16014CAT-2 (Fig. 6).

**The *sqdB* gene product shows homology to the N terminus of UDP-glucose epimerase from different organisms.** The putative open reading frame for the *sqdB* gene encodes a 404-amino-acid-residue polypeptide with an aggregate molecular mass of 46 kDa. A hydropathy analysis by the method of Kyte and Doolittle (19) indicated that the putative protein is probably soluble rather than membrane bound, because no major hydrophobic domains typical for membrane spanning regions could be detected.

The deduced amino acid sequence of the *sqdB* gene product was used in a homology search of two protein data bases (NBRF-PIR release 26 and SWISS-PROT release 15). The enzyme UDP-glucose epimerase from various organisms showed significant sequence homology (28.3% identity over 60 amino acids for the *E. coli* protein or 43.3% if identical amino acids and conservative substitutions were considered) to the N-terminal region of the putative *sqdB* gene product (Fig. 7A).

**Mutant line CHB18 marks a gene involved in sulfolipid biosynthesis located adjacent to the *sqdB* gene.** The defective gene in the sulfolipid deficient mutant CHB18 was complemented by the cosmid pCHB1601 and its subclone

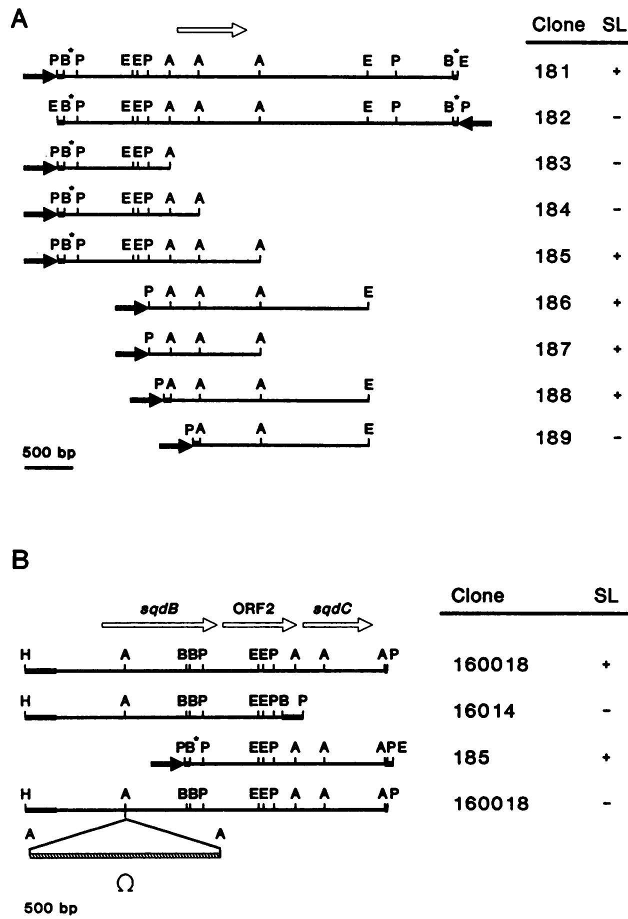


FIG. 8. Deletion and complementation analysis of derivatives of pCHB181 (A) and reconstruction and insertional inactivation of the *sqd* operon (B). (A) The open arrow indicates the location of the *sqdC* open reading frame on pCHB181. (B) The open arrows above the pCHB160018 sequence indicate the location of the *sqd* genes and ORF2. The Ω cassette is shown as a cross-hatched box. Clone designations (without the pCHB prefix) and the presence (+) or absence (−) of sulfolipid when the respective clone is mated into the *sqdC* mutant are indicated. Vector sequences are drawn with a thicker line. The solid arrows indicate the position and orientation of the *cycA* promoter of pCHB500. Restriction sites: A, *Bam*HI; B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; *, sites lost during subcloning.

pCHB1611 (Fig. 3B). A 4.3-kb *Bgl*II fragment (Fig. 3B, B4) of pCHB1611 beginning inside the *sqdB* coding sequence was cloned in both directions into the *Bam*HI site of the pCHB500 expression vector, giving rise to clones pCHB181 and pCHB182 (Fig. 8A). Plasmid pCHB181, but not pCHB182, complemented the mutation in CHB18. This result clearly located the gene on a defined fragment of pCHB1611 which overlapped with the 3′ portion of the putative *sqdB* gene. The orientation dependence of the fragment relative to the cytochrome *c* promoter in pCHB500 for complementation indicated that this gene was expressed in the same orientation as the *sqdB* gene. In addition, the dependence of expression on the *cycA* promoter suggested that the gene was not directly preceded by a promoter. This gene was tentatively designated *sqdC*.

**Identification and characterization of the *sqdC* open reading frame.** To identify the *sqdC* open reading frame on clone pCHB181, a series of deletion clones of pCHB181 was constructed in the expression vector pCHB500 and tested for complementation of the *sqdC* mutant (Fig. 8A). The fact that pCHB183 and pCHB184 do not complement as opposed to pCHB185 indicated that the *sqdC* gene is located in the left half of pCHB181. The start of the gene is located within the 307-bp *Bam*HI fragment present in the complementing clone pCHB188 and absent in the noncomplementing clone pCHB189 (Fig. 8A).

The insert of pCHB185 was cloned into pBluescript, and the sequence from the right end of the overlapping clone pCHB16014 to the right end of pCHB185 (Fig. 3C) was determined (Fig. 4, composite sequence of pCHB16014 and pCHB185). Open reading frame analysis in combination with codon usage analysis revealed a putative *sqdC* open reading frame which can encode a protein with a molecular mass of 26 kDa. The proposed initiation codon as well as two other in-frame initiation codons further upstream are not preceded by a recognizable ribosome binding site. However, since the first initiation codon is located just outside the 307-bp *Bam*HI fragment determined to contain the start site for the *sqdC* gene and the second initiation codon already falls into a region with increased unusual codons, the third initiation codon was proposed to represent the start of the *sqdC* gene.

The deduced amino acid sequence from the putative *sqdC* open reading frame showed no homology to known proteins listed in two data bases (NBRF-PIR release 26 and SWISS-PROT release 15). The result of a hydropathy analysis according to the method of Kyte and Doolittle (19) did not reveal major hydrophobic domains typical for membrane spanning regions of membrane-bound proteins.

**The translated amino acid sequence of an open reading frame flanked by the *sqdB* and *sqdC* genes shows homology to glycogenin.** Analysis of the composite sequence of pCHB16014 and pCHB185 (Fig. 4) revealed an additional reading frame, designated ORF2, proposed to begin 57 bp downstream of the *sqdB* stop codon and ending 68 bp upstream of the proposed initiation codon for the *sqdC* gene. The result of codon usage analysis indicated that the codon usage of ORF2 is consistent with it being the protein coding sequence for another *R. sphaeroides* gene. The initiation codon proposed on the basis of the result of the reading frame and codon usage analysis of the DNA sequence was not found to be preceded by a recognizable ribosome binding site. The deduced amino acid sequence of ORF2 was used to search NBRF-PIR release 26 and SWISS-PROT release 15 for homologous sequences. Significant homology (36.6% identity over 134 amino acids or 53.0% if identical amino acids and conservative substitutions were considered) to rabbit muscle glycogenin was discovered (Fig. 7B). Glycogenin is thought to be an autoglycosylating, UDP-glucose-dependent glycosyltransferase involved in priming of glycogen biosynthesis in mammals (9, 24).

**Creation of a polar mutation by insertion of an Ω cassette into the *sqdB* gene.** A common feature of bacterial genes organized in a transcriptional unit or operon is that mutations, particularly transposon insertions, in the genes located at the 5′ end of the operon can exert polar effects by inhibition of expression of the genes located in the 3′ direction of the insertion. In order to examine whether the *sqdB* and *sqdC* genes are cotranscribed, the colinear sequence of the three open reading frames on the cosmid pCHB1601 was reconstructed by combining parts of the two subclones pCHB16014 and pCHB185, which complement the *sqdB* and *sqdC* mutants, respectively. The plasmid pCHB16014 was digested with *Pst*I, and the small fragments

were replaced by a 2.0-kb *Pst*I partial fragment of pCHB185, giving rise to pCHB160018 containing all three open reading frames in their original order (Fig. 8B).

A polar mutation in the *sqdB* gene was created by inserting the 2.0-kb *Bam*HI fragment of pHP45Ω carrying an Ω cassette, which contains a spectinomycin and streptomycin resistance gene flanked by transcriptional terminators and stop codons in all three frames (25), into the 5'-proximal *Bam*HI site of pCHB160018, giving rise to pCHB160018::Ω (Fig. 8B). The Ω cassette causes termination of transcription and translation independently of the orientation. In order to test whether this mutation would affect the expression of the *sqdC* gene located at the 3' end of the region, pCHB160018 and pCHB160018::Ω were mated into the *sqdC* mutant. Plasmid pCHB160018 complemented the mutation, whereas pCHB160018::Ω did not (Fig. 8B). This result indicated that the *sqdB* and *sqdC* genes are cotranscribed and that the two genes *sqdB* and *sqdC* together with ORF2 are organized in an operon.

## DISCUSSION

On the basis of the complementation analysis described here, it can be concluded that the two sulfolipid-deficient mutant lines of *R. sphaeroides*, CHB16 and CHB18, carry mutations in two different genes involved in sulfolipid biosynthesis. These two novel genes have been preliminarily designated *sqdB* and *sqdC*, respectively. Together with the recently identified *sqdA* gene (7), the number of genes presently known to be required for sulfolipid biosynthesis is three. In addition, an open reading frame, designated ORF2, located between the *sqdB* and *sqdC* genes may represent another gene in this pathway. Since we have not yet isolated a mutant defective in this putative gene, additional experiments are required to test whether ORF 2 represents an additional *sqd* gene required for sulfolipid biosynthesis. Since we have not yet isolated the same mutant twice, it is assumed that the pathway of sulfolipid biosynthesis is not saturated with mutations.

Unfortunately, we do not know what the functions of any of the gene products are, and further experiments will be required to address this question. However, the apparent amino acid sequence identity between regions of the UDP-glucose epimerase from various organisms and the putative *sqdB* gene product permits speculation about the function of the putative *sqdB* gene product. The enzyme UDP-glucose epimerase catalyzes the conversion of UDP-glucose to UDP-galactose (18). UDP-sulfoquinovose has recently been demonstrated to be a precursor of the plant sulfolipid (16). It is therefore conceivable that the *sqdB* gene product may be involved in sulfolipid biosynthesis at the level of synthesis of UDP-sulfoquinovose or some other related sugar nucleotide. Similarly, although the involvement of ORF2 in sulfolipid biosynthesis has not yet been demonstrated, the apparent sequence similarity to glycogenin, a UDP-glucose-dependent glycosyltransferase, raises the possibility that ORF2 represents a *sqd* gene coding for a glycosyltransferase, which has been postulated to catalyze the last step in the pathway (16).

The polar effect of an Ω cassette insertion into the *sqdB* gene on the expression of the *sqdC* gene suggests that these two genes are organized in an operon. Thus, it may be possible to identify other sulfolipid-deficient mutants by directed mutagenesis of the regions of the chromosome adjacent to the *sqdC* gene. It must be pointed out that the previously described *sqdA* (7) mutant is not closely linked to this operon. Therefore, directed mutagenesis of the regions of DNA adjacent to the *sqdA* gene may also reveal other genes involved in sulfolipid biosynthesis in *R. sphaeroides*.

## ACKNOWLEDGMENTS

We thank R. L. Uffen for providing the *R. sphaeroides* wild-type strain and R. S. Hanson for providing pLA2917.

This work was supported by grants from the USDA/NSF/DOE Plant Science Center Program and the U.S. Department of Energy (DE-FG02-90ER20021).

## REFERENCES

1. **Allen, L. N., and R. S. Hanson.** 1985. Construction of broad-host-range cosmid cloning vectors: identification of genes necessary for growth of *Methylobacterium organophilum* on methanol. J. Bacteriol. **161**:955–962.
2. **Bachmann, B. J.** 1987. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p. 1190–1219. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
3. **Balbas, P., X. Soberon, E. Merino, M. Zurita, H. Lomeli, F. Valle, N. Flores, and F. Bolivar.** 1986. Plasmid vector pBR322 and its special-purpose derivatives—a review. Gene **50**:3–40.
4. **Barr, P. J., R. M. Thayer, P. Laybourn, R. C. Najarian, F. Seela, and D. R. Tolan.** 1986. 7-Deaza-2'-deoxyguanosine-5'-triphosphate: enhanced resolution in M13 dideoxy sequencing. BioTechniques **4**:428–432.
5. **Beaucage, S. L., and M. H. Caruthers.** 1981. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett. **22**:1859–1862.
6. **Benning, C.** 1991. Genetic analysis of the pathway for the biosynthesis of the plant sulfolipid in the purple bacterium *Rhodobacter sphaeroides*. Dissertation for the degree of Ph.D. Michigan State University, East Lansing.
7. **Benning, C., and C. R. Somerville.** 1992. Isolation and genetic complementation of a sulfolipid-deficient mutant of *Rhodobacter sphaeroides*. J. Bacteriol. **174**:2352–2360.
8. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–252.
9. **Campbell, D. G., and P. Cohen.** 1989. The amino acid sequence of rabbit skeletal muscle glycogenin. Eur. J. Biochem. **185**:119–125.
10. **Citron, B. A., and J. E. Donelson.** 1984. Sequence of the *Saccharomyces GAL* region and its transcription in vivo. J. Bacteriol. **158**:269–278.
11. **Dembitsky, V. M., E. E. Pechenkina-Shubina, and O. A. Rozentsvet.** 1991. Glycolipids and fatty acids of some seaweeds and marine grasses from the Black Sea. Phytochemistry **30**:2279–2283.
12. **Figurski, D., and D. R. Helinski.** 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function in *trans*. Proc. Natl. Acad. Sci. USA **76**:1648–1652.
13. **Gribskov, M., J. Devreux, and R. R. Burgess.** 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12**:539–549.
14. **Harwood, J. L.** 1980. Sulfolipids, p. 301–320. *In* P. K. Stumpf (ed.), The biochemistry of plants, vol. 4. Academic Press, Inc., New York.
15. **Harwood, J. L., and R. G. Nicholls.** 1979. The plant sulfolipid—a major component of the sulphur cycle. Biochem. Soc. Trans. **7**:440–447.
16. **Heinz, E., H. Schmidt, M. Hoch, K. H. Jung, H. Binder, and R. R. Schmidt.** 1989. Synthesis of different nucleoside 5'-diphospho-sulfoquinovoses and their use for studies on sulfolipid biosynthesis in chloroplast. Eur. J. Biochem. **184**:445–453.
17. **Innis, M. A., K. B. Myambo, D. H. Gelfand, and M. A. D. Brow.** 1988. DNA sequencing with *Thermus aquaticus* DNA polymer-

ase and direct sequencing of polymerase chain reaction-amplified DNA. Proc. Natl. Acad. Sci. USA **85**:9436–9440.

18. **Kalckar, H. M.** 1958. Uridine diphosphogalactose: metabolism, enzymology, and biology. Adv. Enzymol. **20**:111–134.
19. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
20. **Lemaire, H. G., and B. Muller-Hill.** 1986. Nucleotide sequences of the *galE* gene and the *galT* gene of *E. coli*. Nucleic Acids Res. **14**:7705–7711.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Ormerod, J. G., K. S. Ormerod, and H. Gest.** 1961. Light-dependent utilization of organic compounds and photoreduction of molecular hydrogen by photosynthetic bacteria: relationships with nitrogen metabolism. Arch. Biochem. Biophys. **94**:449–463.
23. **Parkinson, C., and S. Y. Cheng.** 1989. A convenient method to increase the number of readable bases in DNA sequencing. BioTechniques **7**:828–829.
24. **Pitcher, J., C. Smythe, and P. Cohen.** 1988. Glycogenin is the priming glycosyltransferase required for the initiation of glycogen biogenesis in rabbit skeletal muscle. Eur. J. Biochem. **176**:391–395.
25. **Prentki, P., and H. M. Krisch.** 1984. In vitro insertional mutagenesis with a selectable DNA fragment. Gene **29**:303–313.
26. **Radunz, A.** 1969. Ueber das Sulfochinovosyl-diacylglycerin aus hoeheren Pflanzen, Algen und Purpurbakterien. Hoppe-Seyler's Z. Physiol. Chem. **350**:411–417.
27. **Raleigh, E. A., K. Lech, and R. Brent.** 1989. *Escherichia coli*, plasmids, and bacteriophages, 1.4.8. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.), Current protocols in molecular biology. Wiley Interscience, New York.
28. **Reed, K. C., and D. A. Mann.** 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. **13**:7207–7221.
29. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
30. **Sistrom, W. R.** 1960. A requirement of sodium in the growth of *R. sphaeroides*. J. Gen. Microbiol. **22**:778–785.
31. **Sistrom, W. R.** 1962. The kinetics of the synthesis of photopigments in *R. sphaeroides*. J. Gen. Microbiol. **28**:607–616.
32. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4774.