

## Genetic and Sequence Analysis of an 8.7-Kilobase *Pseudomonas denitrificans* Fragment Carrying Eight Genes Involved in Transformation of Precorrin-2 to Cobyric Acid

J. CROUZET,<sup>1\*</sup> B. CAMERON,<sup>1</sup> L. CAUCHOIS,<sup>1</sup> S. RIGAULT,<sup>1†</sup> M.-C. ROUYEZ,<sup>1‡</sup> F. BLANCHE,<sup>2</sup>  
D. THIBAUT,<sup>2</sup> AND L. DEBUSSCHE<sup>2</sup>

Laboratoire de Génétique, Institut des Biotechnologies,<sup>1</sup> and Département de Chimie Analytique,<sup>2</sup> Rhône-Poulenc Santé, Centre de Recherche de Vitry, 13 Quai Jules Guesde B.P. 14, 94 403 Vitry sur Seine Cédex, France

Received 6 March 1990/Accepted 11 July 1990

A 8.7-kilobase DNA fragment carrying *Pseudomonas denitrificans* *cob* genes has been sequenced. The nucleotide sequence and the genetic analysis revealed that this fragment carries eight different *cob* genes (*cobF* to *cobM*). Six of these genes have the characteristics of translationally coupled genes. *cobI* has been identified as *S*-adenosyl-L-methionine (SAM):precorrin-2 methyltransferase structural gene because the encoded protein has the same NH<sub>2</sub> terminus and molecular weight as those of the purified enzyme. From protein homology with CobA and CobI, two SAM-dependent methyltransferases of the cobalamin pathway, it is proposed that *cobF*, *cobJ*, *cobL*, and *cobM* code for other methyltransferases involved in the cobalamin pathway. In addition, purified CobF protein has affinity for SAM, as expected for a SAM-dependent methyltransferase. Accumulation of cobalamin precursors in *Agrobacterium tumefaciens* mutants complemented by any of these eight genes suggest that, apart from *cobI*, whose function is identified, the products of all these genes are implicated in the conversion of precorrin-3 into cobyric acid.

We have reported the cloning of at least 14 *Pseudomonas denitrificans* *cob* genes (9) which are grouped into four different clusters on the chromosome. Genetic and biochemical analysis has allowed us to identify five *cob* genes (*cobA* to *cobE*) on a 5.4-kilobase(kb) fragment from complementation group C-pair (10a). *cobA* was found to be the structural gene encoding *S*-adenosyl-L-methionine:uroporphyrinogen III methyltransferase (SUMT), *cobB* codes for cobyric acid *a,c*-diamide synthase, and *cobC* and *cobD* code for proteins involved into the transformation of cobyric acid into cobinamide. The function of protein CobE remains to be identified.

Cobalamin biosynthesis is considered to require between 20 and 30 steps to transform uroporphyrinogen III (urogen III) into cobalamin. Between precorrin-3 and cobyric acid no intermediate or enzymatic activity have been identified so far (20). However, biochemical data support the following reactions: (i) methylations at C-1, C-5, C-12, C-15, and C-17; (ii) decarboxylation of the acetic acid side chain at C-12; (iii) loss of C-20; and (iv) cobalt insertion. The sequential steps are not identified; only the methylation order is known (20). Identification of genes involved in this part of the pathway would be a valuable tool for biochemical studies. We report here the genetic analysis and nucleotide sequence of an 8.7-kb *EcoRI* *P. denitrificans* fragment from complementation group A (9), which is shown to contain eight genes involved into the transformation of precorrin-2 into cobyric acid.

### MATERIALS AND METHODS

**Bacterial strains and plasmids.** Bacterial strains and plasmids used in this study are described in Table 1. The 8.7-kb *EcoRI* fragment studied was cloned into pXL151 (9). This fragment was inserted into the pKT230 or pRK290 *EcoRI* site to give pXL253 and pX367, respectively.

**Media, bacteriological techniques, and chemicals.** Bacteria were grown in LB (21) and PS4 (9) media for routine culturing, complementation test, or cobalamin production, respectively. The growth temperature was 37°C for *Escherichia coli* or 30°C for *Agrobacterium tumefaciens* and *P. denitrificans*. Antibiotic concentrations and culture conditions used for cobalamin synthesis by *A. tumefaciens* strains have been described previously (9). Complementations of *A. tumefaciens* Cob mutants were performed as previously described (9). Cobalamin concentrations were determined by the microbiological assay with the *E. coli* vitamin B<sub>12</sub> auxotroph 113-3 Cbl1 as the indicator strain (9).

**General methods.** Recombinant DNA techniques used in this study include all the methods used previously (9). The procedures to mobilize plasmid DNA from *E. coli* to *A. tumefaciens* have been described previously (9). DNA sequencing was performed as described elsewhere (10a). Concentration of intracellular accumulated corrinoids was performed as described elsewhere (F. Blanche, D. Thibaut, M. Couder, and J. C. Muller, Anal. Chem., in press).

**Mutagenesis.** Tn5 and Tn3lacZ were inserted into the 8.7-kb *EcoRI* fragment cloned into pXL367. Tn5 mutagenesis was performed as described by de Bruijn and Lupski (12), using λ467 to transfect LE392(pXL367). After plasmid DNA purification, 29 Tn5 insertions were identified on the 8.7-kb fragment. Tn3lacZ was inserted as previously described (28). Transposition was performed in strain HB101(pHoHo1, pSShe, pXL367). Plasmid pHoHo1 carries the Tn3lacZ transposon and is defective for transposition; it transposes when *tnpA* is supplied in *trans* with plasmid pSShe. In a first

\* Corresponding author.

† Present address: Résidence le Gilly, 73200 Gilly sur Isère, France.

‡ Present address: INSERM U. 152, Hôpital Cochin, 75674 Paris Cédex 14, France.

TABLE 1. Bacterial strains and plasmids used

Bacterial strains, phage, or plasmid	Marker and replicon	Relevant properties	Reference or source
<i>E. coli</i>			
LE392		F <sup>-</sup> <i>hsdR514 supE44 supF58 lacY1 galK2 galT22 metB1 trypR55</i> λ	12
SF800		Nal <sup>r</sup> <i>polA</i>	27
TG1		Δ( <i>lac-pro</i> ) <i>thi supE hsdD5/F' proAB lacI<sup>q</sup>ZΔM15</i>	T. J. Gibson, Ph.D. thesis <sup>a</sup>
113-3 Cbl1		<i>metE</i> cannot convert cobinamide into cobalamin	9
HB101		F <sup>-</sup> <i>hsdS20 recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1 supE44</i>	7
MC1060		Δ( <i>lacIOPZYA</i> )X74 <i>galU galK strA2 hsdR</i>	10
<i>P. denitrificans</i>			
SC510		High-cobalamin-producing strain	9
SC510 Rif <sup>r</sup>		Rif <sup>r</sup> derivative of SC510	5
Plasmids			
pRK2013	Km <sup>r</sup> ColE1	Carries the <i>tra</i> genes of RK2 helper plasmid for conjugal transfer	14
pRK290	Tet <sup>r</sup> RK2		13
pKT230	Km <sup>r</sup> RSF1010		2
pXL151	Km <sup>r</sup> RSF1010	13-kb <i>P. denitrificans</i> <i>Sau3AI</i> fragment cloned into <i>Bam</i> HI site of pXL59	9
pXL253	Km <sup>r</sup> RSF1010	8.7-kb <i>Eco</i> RI fragment of pXL151 <i>cobF</i> to <i>cobM</i>	This study
pXL367	Tet <sup>r</sup> RK2	8.7 kb <i>Eco</i> RI fragment of pXL151 <i>cobF</i> to <i>cobM</i>	This study
pSShe	Cm <sup>r</sup> P15A	<i>tnpA</i> (Tn3)	27
pHoHo1	Amp <sup>r</sup> ColE1	Tn3 <i>lacZ</i>	27
Phage λ467		λ b221 <i>rex::Tn5 cI857 Oam29 Pam80</i>	12

<sup>a</sup> Thesis from University of Cambridge, Cambridge, England, 1984.

step the transposition occurred in the strain carrying the three plasmids (transposon donor plasmid, transposase-supplying plasmid, and target plasmid). In a second step the plasmids were transferred into SF800 by using strain MC1060(pRK2013). The selected transconjugants contained the target plasmids with Tn3*lacZ* only, since pSShe and pHoHo1 do not replicate in SF800. Thirteen Tn3*lacZ* insertions were selected on the 8.7-kb fragment.

**SAM binding to CobA, CobF, and bovine serum albumin.** The *S*-adenosyl-L-methionine (SAM)-binding test is based on the hypothesis that SAM-dependent methyltransferases exhibit a specific affinity for SAM. The purified protein (about 10 μg) was incubated for 10 min at 30°C in 200 μl of 0.1 M Tris hydrochloride (pH 7.7)–5 nmol of [*methyl*-<sup>3</sup>H] SAM (1 μCi) (Amersham France S.A.). Immediately after incubation, 100 μl of the incubation mixture was applied to a TSK-125 column (Bio-Rad Laboratory) and eluted at 4°C at a flow rate of 1 ml/min with 50 mM Na<sub>2</sub>SO<sub>4</sub>–20 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 6.8). Fractions (0.5 ml) were collected, and the radioactivity present in each fraction was counted. SAM and protein retention times were determined by on-line monitoring of the A<sub>280</sub> of the eluate. Pure SUMT was obtained as described previously (5); CobF was purified from an *E. coli*-overproducing strain (unpublished data).

**Nucleotide sequence accession number.** The sequence of the 8.7-kb fragment has been assigned the GenBank accession number M32224.

## RESULTS

**DNA sequence of the 8.7-kb *Eco*RI fragment from plasmid pXL151.** The 8.7-kb *Eco*RI DNA fragment from plasmid pXL151 leads to the complementation of all the Cob mutants classified in complementation group A (9). Therefore this fragment must contain all the *cob* genes that complement

these mutants. The sequence of this 8,753-base-pair (bp) fragment (Fig. 1) was analyzed by the program of Staden and McLachlan (28, 29) as described elsewhere (10a). This method allowed the identification of eight open reading frames (ORFs), named ORF1 to ORF8, displaying a high coding probability (Fig. 2). All these ORFs were found on the same strand; on the other strand, no ORFs with a high coding probability could be found (data not shown). The positions of the initiation codon and the termination codon, as well as the predicted molecular weight of the encoded protein, are presented for each ORF in Table 2. A GUG initiation codon has been chosen for ORF4 and ORF6 since the probability of these ORFs is high immediately 3' to the termination of the previous ORF and no AUG codon nearby is present (the next AUG codon being 357 and 564 bp downstream for ORF4 and ORF6, respectively). Translation initiation codons were chosen for ORF1 and ORF2, but translation may start at the following AUG codons, which are several bases downstream (15 and 12 bp for ORF1 and ORF2, respectively) (Fig. 1). Identification of the NH<sub>2</sub> terminus of the encoded protein would allow us to determine the right translation initiation codon.

We found no sequence, upstream of these ORFs, showing a significant homology with the *E. coli* consensus ribosome-binding site (26). On the contrary, potential ribosome-binding-site sequences were found for genes *cobA* to *cobE* (10a). We do not know whether this difference is correlated to the level of gene expression from one cluster to another. Six ORFs (ORF2 to ORF7) show the characteristics of ORFs that are translationally coupled; i.e., the stop codons of ORF2 to ORF6 are either overlapping with the initiation codon of the next ORF or, for ORF2, are only 2 bases apart (Fig. 1). In contrast, a 100-bp intergenic region separates ORF1 from ORF2, and ORF7 ends 69 bp upstream from

EcoRI  
 gaattcgccagcgeectacatggetgacctcaagcagttccctcgtggccrcaagaagcagggccgacagatttccctcgcggccctgaglatlctgcgcgclcgacclgacgcc 116  
 gctcgacaagggtgcgctgggtcattctcggccaggatccctatccaggtgacgccagggccatgggctcgtctcagcgttcgccccggtgtccggaccgcccgtcgetggtea 232  
 acatctacaagggaactgaatccgatctcggatallccgcggcgcgctcagggtttctcgaaagctgggcaaggcagggcgctgctctlltgaaacagctgctgacggtagagcgc 348  
 gggaaagtcgctcacaccaggtcacggltgggaaaagtcacggatgcatccgtgcgggtcaacgagggccagacatcccgctgclctcatgctllggggctcctatgcccag 464  
 aagaagggcgctctgctgcacgctcgcgccatctgtcctgagggccaccacatccgtcgcgcctctcagccatccggctllctcggtgcccgcattlltccagggccaatgc 580  
 ctctcgaagcaaggtctcgatccgatgactggcgctccggaaaatccggctcgggcatcaactgaaggcttggcgcgaatgacggctlltgcctgcctcgggtcct 696  
 gccctggggcggggatccgctcaagcggcccgaacgaa ATG GCG GAG GCG GGC ATG GCG AAA ATT CTG ATC ATC GGC ATC GGT TCG GGC AAT CCC 792  
 H A E A G M A R K I L I I G I G S G M P  
 ORF 1 (cabE)  
 GAA CAC ATG ACC GTG CAG GCG ATC AAC GCG CTG AAC TGC GCC GAC GTG CTC TTT ATC CCG ACC AAG GGA GCG AAG AAG ACC GAG CTT 879  
 E H M T U Q A I N A L M C A D U L F I P T K G A K K T E L  
 GCC GAA GTG GCG GCG GAC ATC TGC GCC GCG TAC GTC ACG GCG AAG GAC AGC GCG ACC GTC GAG TTC GCG GTG CCC GTG GCG GCG ACC 966  
 A E U A R A D I C A R A V U T A K D S R T U E F A U P U A R A T  
 GAA GGC GTC AGC TAT GAC GGC AGC GTC GAT GAC TGG CAC GCC CAG ATC GCT GGG ATT TAC GAA GCG CTT CTA TCG AAG GAG TTG GGC 1053  
 E G U S Y D G S U D W H A Q I A G I V E A L L S K E L G  
 GAA GAG GGA ACT GGC GCG TTT CTC GTC TGG GGC GAC CCG ATG CTC TAT GAC AGC ACC ATT GCG ATC GTC GAG GCG GTC AAG GCA GCG 1140  
 E E G T G A F L U M G D P M L Y D S T I A I U E A U K A R  
 GGT GAG GTC GCC TTC GCC TAC GAC GTC ATT CCC GGG ATC ACC AGT CTG CAG GCG CTT TGC GCC AGC CAC GCG ATT CCG CTG AAC CTC 1227  
 G E U A F A Y D U I P G I T S L Q A L C A S H A I P L N L  
 GTC GGC AAG CCG GTG GAG ATC ACC ACG GGG CGT CCG CTG CAC GAA AGC TTT CCC GAG AAG AGC CAG ACC TCG GTC GTC ATG CTC GAT 1314  
 U G K P U E I T T G R A L H E S F P E K S Q T S U U M L D  
 GGC GAA CAG GCG TTT CAG GCG GTC GAG GAC CCG GAG GCG GAG ATC TAT TGG GGC GCC TAT CTC GGC ACG CCG GAT GAG ATC GTC ATT 1401  
 G E Q A F Q R A U E D P E A E I V M G A Y L G T R A D E I U I  
 TCC GGC GCG GTG GCT GAG GTG AAG GAC CCG ATC CTT GAA ACG CCG GCG GCG GCG GCG AAG ATG GGA TGG ATC ATG GAC ATC TAT 1488  
 S G R U A E U K D R I L E T R A A R A R A K M G M I M D I Y  
 CTC CTG GCG AAG GGC GCC GAC TTC GAC GAG tgcgggggagggccgatcgcgctglttgcctcactcaaggttgcggctgtgtatagcgtcttaagggc 1593  
 L L R K G A D F D E \*\*\*  
 ttcttcaggggaggaacctcaagt ATG ACG GAT TTG ATG ACC AGC TGC GCC CTT CCA TTG ACC GGA GAT GCC GGC ACC GTC GCT TCG ATG 1685  
 M T D L M T S C A R L P L T G D A G T U A S M  
 ORF 2 (cabG)  
 CGC GCG GGC GCC TGC CCG TCC TTG GCA GAG CCG ATG CAG ACC GGC GAC GGC CTG CTC GTG AGG GTG AGG CCA ACG GAT GAC AGC CTG 1772  
 A R G A C P S L A E P M Q T G D G L L U R U R A P T D D S L  
 ACG CTG CCG AAG GTC ATT GCC CTT GCC ACG GCT GCC GAG GCG TTC GGC AAT GGC ATC ATC GAG ATT ACC GCG GCG GGA AAC CTG CAG 1859  
 T L P K U I A L A T A R A E R F G M G I I E I T A R A G M L Q  
 CTT GCG GGC CTG AGC GCG GCT TCG GTG CCA AGG CTG GCG CAG GCG ATC GGC GAT GCG GAG ATC GCC ATT GCC GAG GGG CTC GCG ATC 1946  
 L R G L S A A S U P R L A Q A I G D A E I A I A E G L A I  
 GAG GTG CCG CCC CTG GCC GGC ATC GAC CCG GAC GAG ATC ACC GAT CCG CCG CCG ATT GCC ACT GAG CTT CGT GAA CCG TTG GAT GTG 2033  
 E U D P L A G I D P D E I A G P A P I A T E L R E A L D D  
 CCG CAG GTG CCG TTG AAG CTT GCA CCC AAA TTA TCC GTC GTC ATC GAT AGC GGT GGC CCG TTT GGT CTC GGC GCT GTC GTC GCC GAC 2120  
 A Q U P L K L A P K L S U U I D S G G R F G L G A U U A D  
 ATT GCG CTT CAG GCG GTT TCG ACT GTC GCG GGG GTG GCC TGG GTG CTG TCG CTT GGC GGC ACG TCA ACG AAG GCA TCG AGC GTC GGG 2207  
 I R L Q A U S T U A G U A W U L S L G G T S T K A S S U G  
 ACG TTG GCC GGC AAC GCG GTC GTG CCG GCC CTG ATC ACC ATT CTC GAG AAA CTG GCG AGC CTG GGC ACG ACG ATG GCG GGG GCG GAT 2294  
 T L A G M A U U P A L I T I L E K L A S L G T T M R A G A D  
 CTG GAC CCG TCG GAA ATC GCG GCG CTC TGT GCG TGT GAG ACA TCG TCC GAA CCG CCG GCC GCT CCG CGT TCG GCC GCA ATA CCC GGC 2381  
 L D P S E I R A L C A C E T S S E A P A A P A S A A I P G  
 ATT CAT GCG CTG GGT AAC GCC GAC ACC GTT CTC GGC CTC GGT CTG GCC TTT GCT CAG GTG GAG GCC GCC GCG CTG GCA TCC TAC CTG 2468  
 I H A L G N A D T U L G L G L A F A Q U E A A A L A S Y L  
 CAT CAG GTC CAG GCG CTT GGC GCC AAT GCG ATC CCG CTT GCG CCC GGG CAC GCC TTC TTC GTC CTC GGC CTT TGC CCC GAG ACC GCG 2555  
 H Q U Q A L G A M A I R A L A P G H A F F U L G L C P E T A  
 GCT GTG GCG CAG AGC CTG GCA GCG TCA CAC GGT TTT GCG ATT GCC GAG CAG GAT CCG GCG AAT GCG ATC GCC ACC TGC ACC GGC AGC 2642  
 A U A Q S L A A S H G F A I A E Q D P R A I A T C A G S  
 AAG GGT TGC GCC TCG GCG TGG ATG GAA ACC AAG GGC ATG GCC GAG CCG CTC GTC GAG ACG GCG CCG GAA TTG CTC GAC GGG TCG CTC 2729  
 K G C A S A M H E T K G N A E R L U E T A P E L L D G S L  
 ACC GTG CAT CTC TCC GGC TGC GCC AAG GGC TGC GCC CCG CCG AAG CCG TCC GAA CTG ACG CTT GTC GGT GCG CCA TCA GGA TAC GGG 2816  
 T U H L S G C A K G C A R P K P S E L T L U G A P S G Y G

FIG. 1. Nucleotide sequence of the *EcoRI* 8.7-kb fragment from complementation group A. The positions of the ORF are indicated, along with the predicted amino acid sequence of the encoded polypeptide. Noncoding DNA is represented in lowercase letters.

CTT GTC GTA AAT GGG GCT GCC AAT GGC TTG CCA AGC GCC TAC ACC GAT GAG AAT GGA ATG GGA TCC GCC CTT GCC CGG CTC GGC CGG 2903  
 L U U H G A A H G L P S A V T D E N G H G S A L A R A L G R  
 CTG GTG CGG CAA AAC AAA GAC GCT GGC GAA TCG GCG CAG TCC TGT CTT ACA CGG CTC GGA GCT GCG CGC GTC TCG GCA GCG TTC GAA 2990  
 L U R Q N K D A G E S A Q S C L T R L G A A R U S A A F E  
 CAG GGA <sup>legac</sup> ATG CCT GAG TAT GAT TAC ATT CGC GAT GGC AAC GCC ATC TAC GAG CGT TCC TTC GCC ATC ATC CGC GCC GAG GCC 3076  
 Q G \*\*\* H P E Y D V I R D G H A I Y E R S F A I I R A E A  
 ORF3 (cobl)  
 GAT CTG TCG CGC TTC TCC GAA GAG GAA GCG GAT CTG GCT GTG CCG ATG GTG CAC GCC TGC GGT TCC GTC GAG GCG ACC AGG CAG TTC 3163  
 D L S R A F S E E E A D L A U R A H U H A C G S U E A T A Q F  
 GTG TTT TCT CCC GAT TTC GTA AGC TCG GCC CGT GCG GCG CTG AAA GCC GGT GCG CCG ATC CTC TGC GAT GCC GAG ATG GTT GCG CAC 3250  
 U F S P D F U S S A R A R A L K A G A P I L C D A E H U A H  
 GGT GTC ACC CGC GCC CGT CTG CCG GCC GGC AAC GAG GTG ATC TGC ACG CTG CCG GAT CCT CGC ACG CCC GCA CTT GCG GCC GAG ATC 3337  
 G U T A R A R L P A G H E U I C T L R D P R T P A L A A E I  
 GGC AAC ACC CGC TCC GCC GCA GCC CTG AAG CTC TGG AGC GAG CCG CTG GCC GGT TCG GTG GTC GCG ATC GGC AAC GCG CCG ACG GCG 3424  
 G H T R S A A R A L K L W S E R L A G S U U A I G H A P T A  
 TTG TTC TTC CTC TTG GAA ATG CTG CCG GAC GGC GCG CCG AAG CCG GCA ATC CTC GGC ATG CCC GTC GGT TTC GTC GGT GCG GCG 3511  
 L F F L E H L R A D G A P K P A I L G H P U G F U G A R  
 GAA TCG AAG GAT GCG CTG GCC GAG AAC TCC TAT GGC GTT CCC TTC GCC ATC GTG CCG GGC CCG CTC GGC GGG AGT GCC ATG ACG GCG 3590  
 E S K D A L A E H S Y G U P F A I U R G R L G G S A M T A  
 GCA GCG CTT AAC TCG CTC GCG AGG CCG GGC CTGTGAGC GGC GTC GGC GTG GGG CCG CTG ATC GGT GTT GGG ACC GGC CCC GGT GAT 3684  
 A A L M S L A R P G L \*\*\*  
 N S G U G U G R L I G U G T G P G D  
 ORF4 (cobl)  
 CCG GAA CTT TTG ACG GTC AAG GCG GTG AAG GCG CTC GGG CAA GCC GAT GTG CTT GCC TAT TTC GCC AAG GCC GGG CGA ACG GGT AAC 3771  
 P E L L T U K A U K A L G Q A D U L A V F A K A G A S G H  
 GGC CCG GCG GTG GTC GAG GGT CTG CTG AAG CCC GAT CTT GTC GAG CTG CCG CTA TAC TAT CCG GTG ACG ACC GAA ATC GAC AAG GAC 3850  
 G R A U U E G L L K P D L U E L P L V Y P U T T E I D K D  
 GAT GGC GCC TAC AAG ACC CAG ATC ACC GAC TTC TAC AAT GCG TCG GCC GAA GCG GTA GCG GCG CAT CTT GCC GCC GGG CCG ACG GTC 3945  
 D G A V K T Q I T D F Y H A S A E A U A R A H L A A G A T U  
 GCC GTG CTC AGT GAA GGC GAC CCG CTG TTC TAT GGT TCC TAC ATG CAT CTG CAT GTG CCG CTC GCC AAT CGT TTC CCG GTC GAG GTG 4032  
 A U L S E E G D P L F Y G S Y M H L H U R A L A H A F P U E U  
 ATC CCC GGC ATT ACC GCC ATG TCC GGC TGT TGG TCG CTT GCC GGC CTG CCG CTG GTG CAG GGC GAC GAC GTG CTC TCG GTG CTT CCG 4119  
 I P G I T A M S G C W S L A G L P L U Q G D D U L S U L P  
 GGC ACC ATG GCC GAG GCC GAG CTC GGC CCG AGG CTT GCG GAT ACC GAA GCC GCC GTG ATC ATG AAG GTC GGG CCG ART TTG CCG AAG 4206  
 G T H A E A E L G R A L A D T E A A U I H K U G A N L P K  
 ATC CGT CCG GCG CTC GGT GCC TCC GGC CGT CTC GAC CAG GCC GTC TAT GTC GAA CCG GGC ACG ATG AAG AAC GCG GCG ATG ACG GCT 4293  
 I R A R A L A S G R L D Q A U Y U E R A G T M K H A A M T A  
 CTT CCG GAA AAG GCC GAC GAC GCG CCC TAT TTC TCG CTG GTG CTC GTT CCC GGC TGG AAG GAC CGA CCATGACC GGT ACG CTC 4379  
 L A E K A D D E A P V F S L U L U P G W K D A P \*\*\*  
 N T G T L  
 ORF5 (cobl)  
 TAT GTC GTC GGT ACC GGA CCG GGC AGC GCC AAC CAG ATG ACG CCG GAA ACG GCG GAA GCC GTT GCG GCC GCT CAG GAG TTT TAC GGC 4466  
 Y U U G T G T G P G S A R K Q N T P E T A E A U A R A Q E F Y G  
 TAC TTT CCC TAT CTC GAC CCG CTG AAC CTC AAG CCG GAT CAG ATC CGT GTC GCG AAC CCG GAG GAG CTC GAT CCG GCA CAG 4553  
 Y F P Y L D R L N L A P D Q I R U A S D H A E E L G D A R A Q  
 GTC GCG CTG ACG CCG GCT GCG GCA GGC GTG AAG GTC TGC ATG GTC TCC GGT GGC GAT CCC GGT GTC TTT GCC ATG GCG GCC GCC GTC 4640  
 U A L T A R A A A G U K U C H U S G G D P G U F A N A A R U  
 TGC GAG GCG ATC GAC AAG GGA CCG GCG GAA TGG AAG TCG GTT GAA CTG GTG ATC ACG CCC GGC GTG ACC GCG ATG CTC GCC GTT GCC 4727  
 C E A I D K G P A E H K S U E L U I T P G U T A H L A U A  
 GCC CCG ATC GGC GCG CCG CTC GGT CAT GAT TTC TGT GCG ATC TCG CTT TCC GAC AAT CTG AAG CCC TGG GAA GTC ATC ACC CCG CGT 4814  
 A R I G A P L G H D A I S D H L K P W E U I T C R A R  
 CTC AAG CTG GCG GCG GAA GCG GGC TTC GTC ATT GCC CTC TAC AAT CCG ATC AGC AAG GCG CCG CCC TGG CAG CTC GGT GAG GCC TTC 4901  
 L A L A A E A G F U I A L Y H P I S K A R A P W Q L G E A F  
 GAG CTT CTG CCG AGC GTT CTG CCG GCA AGC GTT CCG GTC ATC TTC GGC CGT GCG GCC GGG CCG CCG GAC GAA CCG ATC GCG GTG ATG 4988  
 E L L R S U L P A S U P U I F G R A A G A P D E R I A U N  
 CCG CTC GGC GAG GCC GAT GCC AAC CCG GCC AAG ATG GCG ACC TGC GTC ATC ATC GGC TCG CCG GAG ACG CCG ATC GTC GAG CCG GAC 5075  
 P L G E A D H A R A D H A T C U I I G S P E T A I G U E A D  
 GGC CAA CCC GAT CTC GTC TAC ACA CCG CCG TTC TAT GCA GGG GCG AGC CAGTGGC GAT GCG GTT GAG TGC CTC GTC GCA ACT GCC 5161  
 G Q P D L U Y T P R F Y A G A S Q \*\*\*  
 N S D A U E C L U A T A  
 ORF 6 (cobl)  
 GAC CGT CCG CAC GTC CCG GGG CTT GCG CCG CTC GAC CAT GAT CAC CTC GAT GCC GAG CCG GCG CCG TGC GGC AAT CTT GCC GTA GGT 5248  
 D R A H U A G L A P L D H D H L D A E P A R C G H L A Y G  
 GGC GCT GCC ACC GCT GTT CTT GGC GAC GAT CAC ATC GAT CTG CCG ACT CCT GAG CAA CCG GGC TTC GTC GGC TTC CCG AAA GGG ACC 5335  
 G A R T A U L G D D H I D L P T P E Q R G F U G F R K G T  
 GGT CCG CAG GAT CCG CTC CTG GTC GGG CAG ATT AAG CCG CCG CGT CAC CCG ATC GAC GCT CCG GAT GAC GTA GCT GTG CTG CCG CCG 5422  
 G R Q D A L L U G Q I K R A R A H A I D A A D D U A U L R A  
 GAC CTC GAA GTG GAA ACG TTC CTG TCG ACC TAT CCG CAG GAA GAC GCG GCG TCG CCG ATC ACC GAG CCG GCT GAC GGC CTC GAC AAC 5509  
 D L E U E S F L S T Y A Q E D A A S P I T E A R A D G C L D H  
 GCT ATC GAC ACG AGT CCA GCG GTC GCC AAG CAG GGG CAC CCA TTC CCG TCG GCG GAG GGC GAT AAG CCG AAC GCC GGT TCT TTG CCG 5596  
 A I D S S P A U A A Q G H P F A S A E G D K R H A G S L R

FIG 1—Continued.

TGC GTC CGC GGC GTT GTG CGA AAT GCG TGC GGC AAA GGG GTG CGT CGC ATC GAC CAG CAG CGC GAT GTT TTC GTC ATG CAC GAA ATG 5683  
C U R G U U R A N A C G K G U R A R I D Q Q R D U F U H H E N

CGC CAG CCC ATC GGC GCC GCC AAA GCC GCC GAT GCG CGT CTT GAC CGG CTG CCG CCG GTC CGC GGT GCG GCC GGC CAG CGA GAT 5770  
R Q P I G A R A K A R A D A R L D A L R P R U R G A A G Q R D

GGC GGT GTC GTA GCG GAC ATC TTC GGC CAA GCG GCG CGC GAG TTC GCG TGC CTC GGT GGT GCC ACC CAG AAT CAG AAT ACG AGG TTT 5857  
G G U U A D I F G Q A A R E F A C L G G A T Q N Q N T R F

TTC CATGGCTGAC GTG TCG AAC AGC GAA CCC GCC ATA GTC TCC CCC TGG CTG ACC GTC ATC GGT ATC GGT GAG GAT GGT GTA GCG 5942  
F H G \*\*\*  
M A D U S H S E P A I U S P M L T U I G I G E D G U A  
OFA 7 (subL)

GGT CTC GGC GAC GAG GCC AAG CGG CTG ATC GCC GAA GCG CCG GTC GTC TAC GGC GGC CAT CGT CAT CTG GAG CTC GCC GCC TCC CTC 6029  
G L G D E A K A L I A E A P U U V G G H A H L E L A A S L

ATC ACC GGC GAA GCG CAC AAT TGG CTA AGC CCC CTC GAA GCG TCG GTC GTC GAG ATC GTC GCG CGT CCG GGC AGC CCG GTG GTG GTG 6116  
I T G E A H H W L S P L E A S U U E I U A R A G S P U U U

CTT GCC TCG GGC GAC CCG TTC TTC TTC GGC GTC GGC GTG ACG CTG CCG CCG CCG ATC GCC TCG GCC GAA ATA CCG ACG CTT CCG GCG 6203  
L A S G D P F F F G U T L A R A I A S A E I A T L P A

CCG TCG TCG ATC AGT CTT GCC GCC TCG CCG CTC GGC TGG GCG CTG CAG GAT GCG ACG CTC GTC TCC GTA CAT GGG CCG CCG CTG GAT 6290  
P S S I S L A A S A L G W A L Q D A T L U S U H G A P L D

CTG GTG CGA CCG CAT TTG CAT CCG GGG GCG CGT GTG CTT ACG CTC ACG TCG GAC GGT GCG GGT CCG CGA GAC CTT GCC GAG CTT CTG 6377  
L U R P H L H P G A R U L T L T S D G A G P A D L A E L L

GTT TCA AGC GGC TTC GGT CAG TCG CGA CTG ACC GTG CTC GAA GCG CTG GGC GGC GCC GGC GAA CCG GTG ACG ACG CAG ATC GCC GCG 6464  
U S S G F F G Q S A L T U L E A L G G A G E R U T T Q I A A

CGC TTC ATG CTC GGC CTC GTG CAT CCT TTG AAC GTC TGC GCC ATT GAG GTG GCG GCC GAC GAG GGC GCG CGC ATC CTG CCG CTT GCC 6551  
A F H L G L U H P L H W C A I E U A A D E G A A I L P L A

GCC GCC CCG GAC GAT GCG CTG TTC GAA CAT GAC GGG CAG ATC ACC AAG CCG GAG GTG CCG GCG CTG ACG CTG TCG GCA CTC GCA CCG 6638  
A G A D D A L F E H D G Q I T K R E U R A L T L S A L A P

CGC AAG GGC GAA CTG CTA TGG GAC ATC GGC GGC GGC TCC GGC TCG ATC GGC ATC GAA TGG ATG CTC GCC GAT CCG ACC ATG CAG GCG 6725  
A K G E L L M D I G G G S G S I G I E W H L A D P T H Q A

ATC ACC ATC GAG GTT GAG CCG GAG CCG GCA GCG CGC ATC GGC CCG AAC GCG ACG ATG TTC GGC GTG CCC GGG CTG ACG GTT GTC GAA 6812  
I T I E U E P E R A R A R I G R A N A T M F G U P G L T U V E

GGC GAG GCG CCG GCG GCG CTT GCC GGC CTG CCA CAA CCG GAC GCG ATC TTC ATC GGC GGC GGC ACC GAA GGC GGC ATG ATG GAA 6899  
G E A P A A L A G L P Q P D A I F I G G G G S E D G U H E

GCA GCG ATC GAG GCG CTC AAG TCA GGC GGA CCG CTG GTT GCC AAC GCG GTG ACG ACG GAC ATG GAA GCG GTG CTG CTC GAT CAT CAC 6986  
A A I E A L K S G G R L U A N A U T T D N E A U L L D H H

GCG CCG CTC GGC GGT TCG CTG ATC CCG ATC GAT ATC GCG CGT GCA GGA CCC ATC GGC GGC ATG ACC GGC TGG AAG CCG GCC ATG CCG 7073  
A R L G G S L I R I D I A R A G P I G G H T G W K P A H P

GTC ACC CAA TGG TCG TGG ACG AAG GGC ~~taaaacagttccagcgaagtgtagcgggttttgcgtccggcaactgcgcaaaaaagaagagtaacct~~ ATG ACG 7177  
U T Q W S H T K G \*\*\*  
M T  
ORF 8 (subH)

GTA CAT TTC ATC GGC GCC GGC CCG GGA GCC GCA GAC CTG ATC ACG GTG CGT GGT CCG GAC CTG ATC GGC CCG TGC CCG GTC GCT 7264  
U H F I G A G P G A A D L I T U A G A D L I G A C P U C L

TAC GCC GGC TCG ATC GTC TCG CCG GAG CTG CTG CGA TAT TGC CCG CCG GGC GCC CCG ATT GTC GAT ACG GCG CCG ATG TCC CTC GAC 7351  
Y A G S I U S P E L L A V C P P G A A I U D T A P H S L D

GAG ATC GAG GCG GAG TAT GTG AAG GCC GAA GCC GAA GGG CTC GAC GTG GCG CCG CTT CAT TCG GGC GAC CTT TCG GTC TGG AGT GCT 7438  
E I E A E Y V K A E A E G L D U A R A L H S G D L S U W S A

GTG GCC GAA CAG ATC CCG CCG CTC GAG AAG CAT GGC ATC GCC TAT ACG ATG ACG CCG GGT CCT TCC TTT CCG GCG GCG GCT TCA 7525  
U A E Q I R A L E K H G I A Y T N T P G U P S F A R A S

GCG CTC GGT CCG GAA TTG ACC ATT CCG GCC GTG GCC CAG ACG CTG GTG CTG ACC CCG GTT TCG GGC CCG GCC TCG CCG ATG CCG AAC 7612  
A L G A E L T I P A U A Q S L U L T R U S G A A S P H P H

TCA GAA ACG CTT TCC GCT TTC GGC GCT ACG GGA TCG ACG CTG GCA ATC CAC CTT GCG ATC CAT GCG CTT CAG CAG GTG GTC GAG GAA 7699  
S E T L S A F G A T G S T L A I H L A I H A L Q Q U U E E

CTG ACG CCG CTC TAC GGT GCC GAC TGC CCG GTC GCC ATC GTC GTC AAG GCC TCC TGG CCG GAC GAA CCG GTG GTG CCG GGC ACG CTC 7786  
L T P L V G A D C P U A I U U K A S M P D E A U U R G T L

GGT GAC ATC GCC GCC AAG GTG GCG GAA GAG CCG ATC GAG CCG ACG GCG CTG ATC TTC GTC GGT CCG GGC CTC GAA GCC TCC GAT TTC 7873  
G D I A A K U A E E P I E A T A L I F U G P G L E A S D F

CGT GAA AGC TCG CTC TAC GAT CCC GCC TAT CAG CCG CCG TTC AAG GGG CCG GCC GAA ~~taggcgcactccctcgggggtcggectaagttcccg~~ 7969  
A E S S L V D P A V Q A R A F R G A G E \*\*\*

tgagaggggttttgaacctatctcggcgttctcgcgcggcgccgctgcttagcgggacgcgcgcttttccctcagcggctcgcggtagagcgtgcttccccaagcagcat 8085  
cagcgtcaccggcgtgggtggcgaagcgaagcagatgatcaggattctgtggaataccagcggctctgcagcagggcaagcagatgatagagcggcgcagatcatcagtaacg 8201  
cgccgctggctcgcagcgtcgggtgctgcagcgcctctagaagctggtaaccggagcgaagccagcggagccagcagcgcactcggcgcagaggcggtagcggccagcagc 8317  
agaacggctcaccagcgggaaggtcgggtgaggtggctcattcgatgatctcccgcgcatacgaaccttgcgaagggcagtcgagagcgaagccgatcaaaagccagatcagg 8433  
gcggctcgaatagagcaggttggcgtcggatccgaaggtcaagagcagcagatggcgttgcataagagcgtcgcagcggcagggatcgggtcctggcgcgcggtcccc 8549  
caccatgcgatagaagcaaaagccatcgcagcggcagcagatgatcgggcaatcaggatcgaccagatgatgaaagttccatcccaatctccttcaggcggctcctata 8665  
gcgcttagccgtatcgagcagatgtctctgttctccatgtcagcagcgtggaagcagcgggacttgcggcgcagatccggggaattc 8753

FIG. 1—Continued.

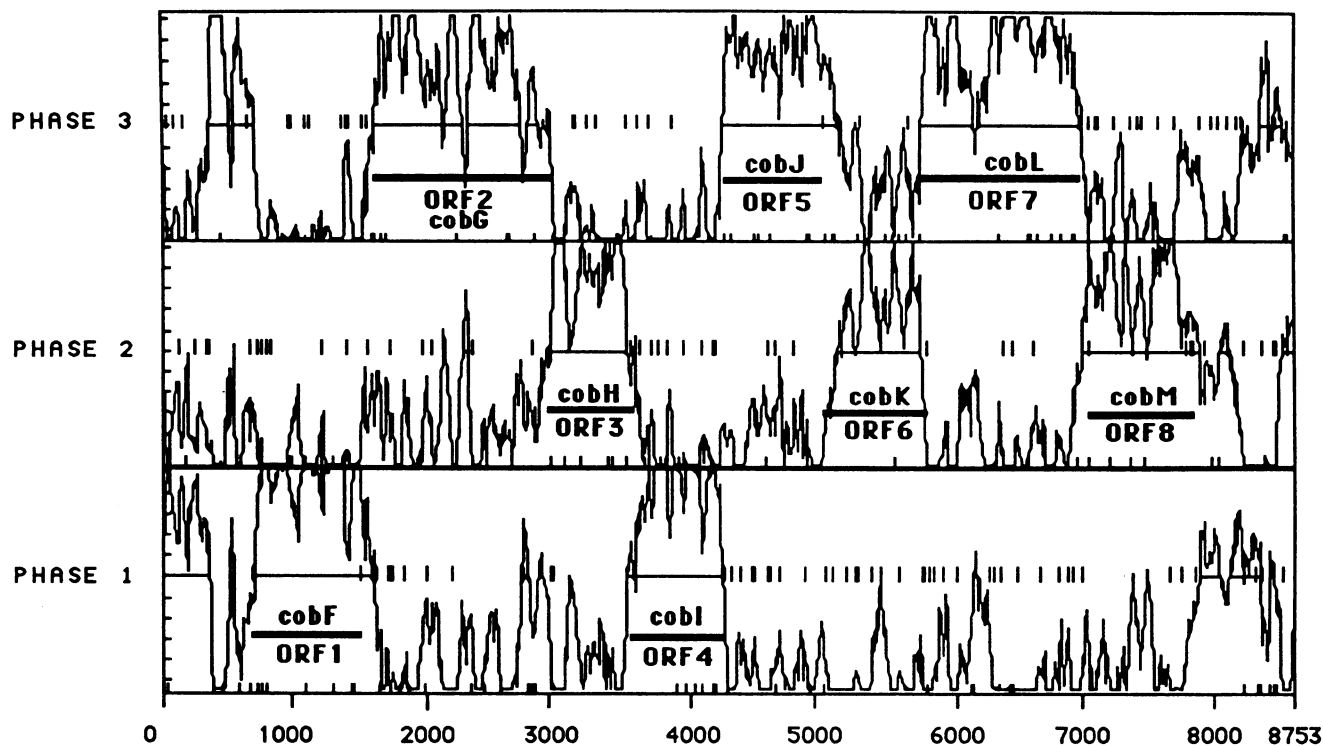


FIG. 2. Codon preference plot of the 8.7-kb DNA sequence by the codon usage method. The window size is 25 codons. This analysis is shown for the three frames of the strand going 5'→3' (Fig. 3). Lines across the half-point of each frame correspond to 50% coding probability. Vertical dashes across the half-points represent stop codons, and those at the bottom of each box represent ATG codons. Heavy lines represent the ORFs.

ORF8. No sequence with a dyad symmetry capable of forming a  $\rho$ -independent terminator (23) appears in the intergenic regions between ORF1 and ORF2 and between ORF7 and ORF8, or after ORF8. The same G+C content was found in this 8.7-kb fragment and in the *P. denitrificans* 5.4-kb fragment carrying *cobA* to *cobE* (10a), i.e., 65.4 and 65.7%, respectively.

**Genetic analysis of the 8.7-kb fragment.** A mutation analysis of the 8.7-kb *EcoRI* was carried out to localize the *cob* genes among the eight identified ORFs. The 8.7-kb fragment was cloned into the pRK290 *EcoRI* site. The resulting plasmid, pXL367, was mutagenized with Tn5 and Tn3lacZ transposons. A total of 29 Tn5 and 13 Tn3lacZ insertions were selected in the 8.7-kb fragment and mapped (Fig. 3). These pXL367 derivatives were introduced by bacterial mating into *A. tumefaciens* Cob mutants already described

(9). The mutants (G164, G609, G610, G611, G612, G613, G614, G615, G616, G620, and G638) were chosen because they represent every complementation class previously identified on group A (9). With the exception of the Tn5 mutant, G164, the mutants were obtained by *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine mutagenesis and are probably nonpolar mutants (9). They are expected to be blocked in a single step of the biosynthesis. Complementation of the Cob mutants by the pXL367 derivatives was examined. As a result, the insertions are classified into nine groups (Fig. 3). One group contains all the insertions that still lead to the complementation of all the mutants, and the other groups, 1 to 8, correspond to insertions for which the plasmids do not complement specific mutants (these insertions are inactivating insertions). Inactivating insertions which no longer complement specific Cob mutants define a group; the mutants and the corresponding groups are described in Fig. 3. Insertions belonging to the same group are always mapped close to each other and are never disrupted by insertions from any other groups. There are as many inactivating insertions groups as ORFs, and for each group, transposons are always inserted within a single ORF, where transposons from other groups are never mapped. All the inactivating insertions are within ORFs. Most of the noninactivating insertions are mapped outside ORFs, with a few exceptions (i.e., Tn5 insertions 955 and 982 might be at the very end of ORF1 or at the beginning of ORF2, and Tn5 insertion 984 seems to be at the very end of ORF8), but these uncertainties are in the range of the mapping accuracy (100 bp). This study allowed the identification of eight genes, involved in cobalamin synthesis, named *cobF* to *cobM* and corresponding to

TABLE 2. ORFs of the 8.7-kb *EcoRI* fragment from complementation group A

ORF	Position of:		Mol wt of encoded polypeptide
	First codon	Last codon	
ORF1	736	1518	28,900
ORF2	1620	2996	46,700
ORF3	3002	3631	22,000
ORF4	3631	4365	25,800
ORF5	4365	5126	27,100
ORF6	5126	5866	26,800
ORF7	5862	7100	42,900
ORF8	7172	7930	26,800

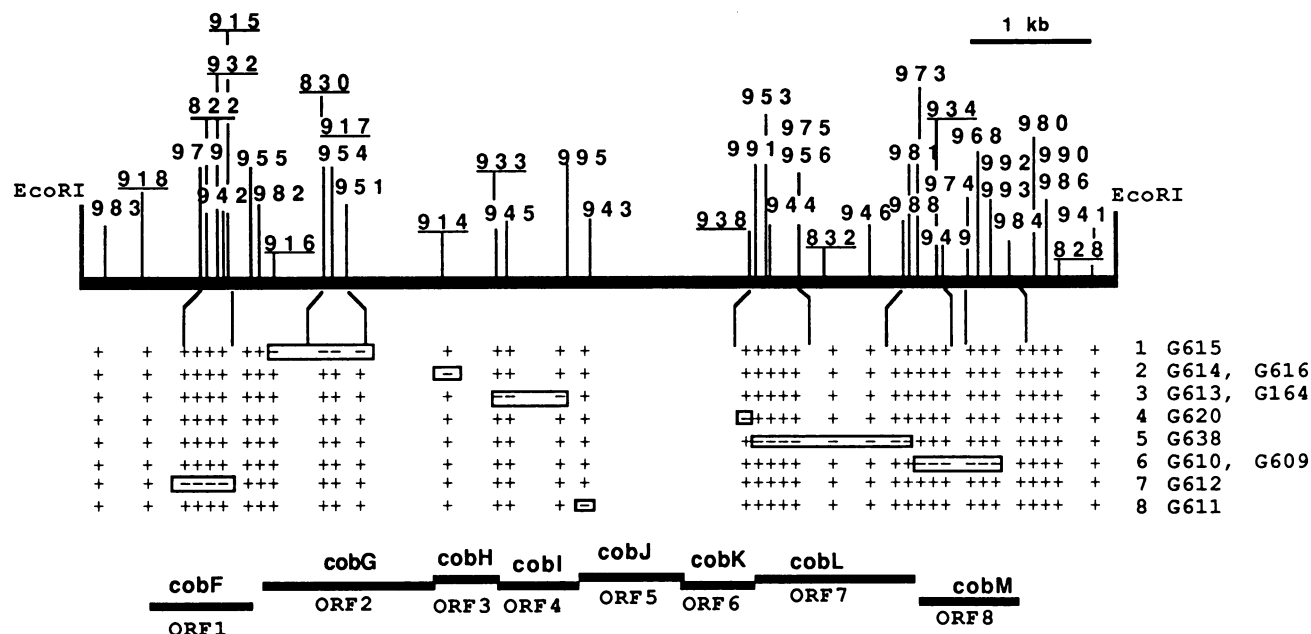


FIG. 3. Tn5 and Tn3lacZ insertions into the 8.7-kb *EcoRI* fragment from complementation group A. They are inserted in plasmid pXL367. The name of each transposon insertion and the site of insertion are indicated. Tn3lacZ insertions are underlined. ORF1 to ORF8 deduced from the DNA sequence are represented in the middle of the figure. Plus and minus signs are aligned vertically with the insertion site to indicate that the plasmid carrying the corresponding insertion does or does not, respectively, complement one of the *A. tumefaciens* Cob mutants, whose names are aligned horizontally with the symbol; the number of each group of inactivating insertions is indicated with the mutant name. Minus signs for mutants belonging to the same groups are boxed. The groups and the associated Cob mutant(s) are as follows: 1, G615; 2, G614 and G616; 3, G613 and G164; 4, G620; 5, G638; 6, G610 and G609; 7, G612; 8, G611.

ORF1 to ORF8, respectively (Fig. 3). Based on the organization of the ORFs, it is probable that most or all of these genes are part of the same operon, but this remains to be confirmed. Transposon insertions on plasmid pXL367 were not polar for the complementation of *A. tumefaciens* Cob mutants. This result might be explained by a promoter activity on the ends of Tn5 that allows the expression of the downstream genes as previously discussed (12).

The mutants classified in groups 1 to 8 are therefore mutated into genes *cobF* to *cobM*, respectively. The mutants are *A. tumefaciens* strains complemented by *P. denitrificans* genes. These bacteria are both aerobic gram-negative rods, and it is likely that they share the same pathway for cobalamin synthesis. This would imply that the Cob mutants complemented by *P. denitrificans* genes are blocked in the step catalyzed by the product of the complementing gene. Therefore, each mutant may be considered to be mutated in the gene homologous to the *P. denitrificans* *cob* gene, allowing the complementation presented in Fig. 3.

The proteins encoded by *cobF* to *cobM*, named CobF to CobM, respectively, were analyzed by the program of Hopp and Woods (16), which draws hydrophilicity plots. The hydrophilicity plots of all these Cob proteins are typical of cytoplasmic proteins (data not shown). The codon usage of these genes is similar to the one observed for *cobA* to *cobD* (10a), indicating that all *P. denitrificans* *cob* genes share the same codon usage, although one of them, *cobE*, is slightly different. This result confirms that *P. denitrificans* codon usage presents some differences from *P. aeruginosa* codon usage (31), as already discussed (10a).

***cobI* is the structural gene encoding SP<sub>2</sub>MT.** *P. denitrificans* SP<sub>2</sub>MT has been purified (D. Thibaut, M. Couder, J. Crouzet, L. Debussche, B. Cameron, and F. Blanche,

submitted for publication). The enzyme is a homodimer for which the estimated subunit molecular weight is 26,000. The enzyme activity has been purified from strain SC510 Rif<sup>r</sup> pXL253, since this plasmid allowed a significant amplification of this activity in strain SC510 Rif<sup>r</sup>. pXL253 has been obtained by cloning the sequenced 8.7-kb fragment into the *EcoRI* site of pKT230. The purified SP<sub>2</sub>MT has been subjected to NH<sub>2</sub>-terminal sequencing (Thibaut et al., submitted), and the first 17 identified amino acids are identical to the amino-terminal sequence of the CobI protein (Fig. 1), except that the amino-terminal methionine has been removed. Removal of the amino-terminal methionine by *E. coli* methionyl aminopeptidase has been proposed to occur when the penultimate amino acid is serine (15). Therefore, it is probable that *P. denitrificans* methionyl aminopeptidase obeys to the same rules as the *E. coli* enzyme (4, 15). The molecular weight of the protein encoded by *cobI* is 25,800, which is very close to the estimated size of an SP<sub>2</sub>MT monomer. It was concluded that *cobI* is the structural gene for SP<sub>2</sub>MT.

A search for homologous proteins in the EMBL bank by using Kanehisa's program (19) did not reveal any significant homology with the CobF to CobM proteins. A computer search for similarity between CobA and CobI led to the identification of three regions of homology (Fig. 4). Interestingly, these three regions correspond exactly to the CobA regions showing homology with *E. coli* CysG protein (10a). The CysG protein is expected to catalyze the same reaction as CobA (i.e., the transfer of two methyl groups from SAM to urogen III), as suggested previously (18). New data have been presented (A. I. Scott, Abstr. Pacifichem '89, 41, p. 205, 1989) demonstrating that CysG has SUMT activity. Since the CobI substrate is slightly different from the CobA

COBI	SGVGVGRLIGVGTGPGDPELLTVKAVKALGQADV	(1-35)
	-- = -	
COBA	PALEKGSVWLVGAGPGDPLLTLHAANALRQADVI	(10-44)
	= -	
CYSG	PLDHRGEVVLVAGPGDAGLLTLKGLQQIQADV	(210-244)
----- REGION 1		
COBI	AAGRTVAVLSEGDPFLFYGSYMLHVRLANR-FPVEVIPGITAMSGCWSLAGLPL	(100-152)
	-- = - - = -	
COBA	AGNRVLR-LKGGDPFVFRGEGEALTLVEHQVFRIVPGITAGIGGLAYAGIPV	(91-143)
	= -	
CYSG	KGKRVVR-LKGGDPFIFGRGEELETLNAGIPFSVVPGITAAASGCSAYSIGIPL	(291-343)
	= -	
-----		
REGION 2		----- REGION 3

FIG. 4. Alignment of amino acid sequences of *P. denitrificans* CobA and CobI and *E. coli* CysG proteins. Similarities are indicated below the sequences as follows: =, same amino acid; -, amino acids belonging to the same group (hydroxyl/small aliphatic: A, G, S, T; acid and acid amide: N, D, E, Q; basic H, R, K; aliphatic: M, I, L, V; aromatic: F, Y, W). Dashes within the sequences indicate gaps in the alignment. Regions 1, 2, and 3 are indicated.

and CysG substrates the homologies between these three proteins might reflect (i) either similar structural domains between SAM-binding proteins, catalyzing SAM-dependent methyl transfer reaction in the cobalamin pathway, or (ii) structural similarities between urogen III, precorrin-1, and precorrin-2.

**Sequence homology among six Cob proteins.** Five other methyl transfers occur in the pathway, and the corresponding enzymes have not been identified. We looked for homology among other Cob proteins and CobA, CobI, and CysG. In the 5.4-kb fragment carrying five *cob* genes (10a), none of the four proteins, CobB to CobE, had homology with CobA or CysG. Of the CobF to CobM proteins, CobF, CobJ, CobL, and CobM showed homologies within the three regions of homologies previously presented (Fig. 5). No significant homology was found within other parts of the proteins and CobA, CobI, and CysG. The homologies between CobF, CobJ, CobL, and CobM are not very strong, and most of the time the existing homologies refer to amino acids belonging to the same group as previously defined (11). For each region, a consensus sequence was raised (data not shown) and compared with proteins in GenBank, by using Kanehisa's program (19). For regions 1 and 2 the scores were higher between the consensus sequence and sequences shown in Fig. 5 than with any protein segment in GenBank, with the exception of CobJ and CobL in region 1 and CobF and CobM in region 2. Homologies within the third region were not significant. These observations indicate that although the observed homologies are not very strong, some of them are significant and all four proteins have significant homologies with CobA and/or CobI within conserved region 1 or 2. In addition, the regions involved in the homologies are located in the same parts of the proteins, i.e., amino acids 1 to 45 (region 1), 60 to 120 (region 2), and 85 to 171, (region 3), with the exception of CysG, which has been proposed to have two domains (10a). Since the homologies involve regions conserved among three SAM methyltransferases, it is postulated that CobF, CobJ, CobL, and CobM are probably SAM methyltransferases, but we cannot exclude the possibility that these sequence homologies also reflect structural similarities between different substrates of the pathway. All these Cob proteins have similar molecular weights (around 28,000), except CobL with a molecular weight of 42,000. CobL may have a second domain catalyzing another reaction, because it is a longer protein and the region of homologies is limited to the first two-thirds of the protein.

Homology with other C-methyltransferases was investigated. We studied bacterial DNA C-methyltransferases catalyzing the formation of 5-methylcytosine and found that 13 DNA methyltransferases present at least five highly conserved regions (24). We also found that CobA presents some homology with one of the conserved regions and a higher homology with *EcoRII* methyltransferase (Fig. 5). Part of this homology is included in region 1.

**CobF affinity to SAM.** One way to test the hypothesis that proteins CobF, CobJ, CobL, and CobM might be SAM-dependent methyltransferases is to purify one of them and show that it exhibits SAM affinity to a comparable extent to CobA or CobI. We purified CobF. *cobF* was placed under the control of the *E. coli trp* promoter and a ribosome-binding site derived from the lambda *cII* gene (unpublished data). Under derepression conditions of the *trp* promoter, overexpression of a protein was observed in *E. coli* by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. This protein has a molecular weight comparable to the CobF predicted molecular weight and was purified on a molecular weight basis. NH<sub>2</sub>-terminal sequencing of the purified preparation was performed. We concluded that such a preparation was a mixture of two proteins, one being initiated at the proposed CobF ATG initiation codon presented in Fig. 1 and the second being translated from the initiation codon which is 15 bp downstream in the sequence. The second protein was the predominant form, since it represented 66% of the NH<sub>2</sub> terminus. This result will be discussed elsewhere (unpublished data). This purified preparation could be used for the SAM-binding assay. Purified CobA (5) and bovine serum albumin were also analyzed under the same conditions. Both CobA and CobF have affinity for SAM (Fig. 6), because a radioactivity peak was eluted at the same retention time as the purified proteins alone (around 7.5 min). This radioactivity peak did not correspond to free SAM, which was eluted later (10.5 min). In contrast to CobA and CobF, bovine serum albumin did not bind SAM. Therefore, we concluded that CobF was binding SAM to a significant extent. The hypothesis based on sequence homology seems to be confirmed. We propose that CobF, CobJ, CobL, and CobM are SAM methyltransferases. This implies that these proteins play a role in the pathway between precorrin-3 and cobyrinic acid, since the methyl groups (those that are transferred to the macrocycle nucleus, apart from those that are transferred by SUMT and SP<sub>2</sub>MT) are introduced between precorrin-3 and cobyrinic acid (20).

**Biochemical study of Cob mutants complemented by *cobF* to**



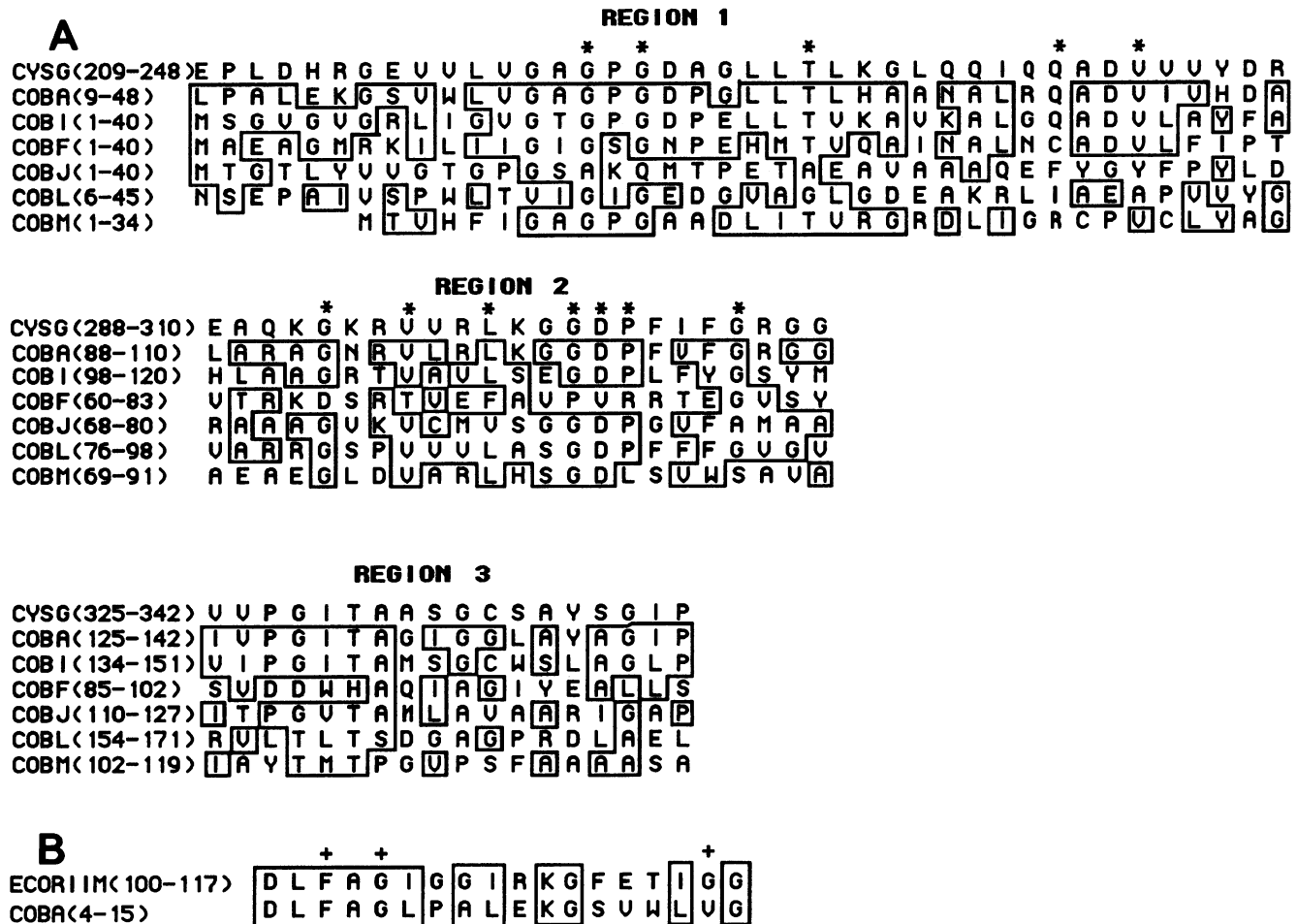


FIG. 5. Alignment of amino acid sequences of CysG, CobA, CobI, CobF, CobJ, CobL and CobM (A) and EcoRII methyltransferase (ECORIIM) and CobA (B). The sequences were aligned manually, and amino acids are boxed when more than three residues belong to the same group as defined in Fig. 4. Regions refer to highly conserved regions between CobA and CysG as defined previously (10a). Unlike the other sequences, the CysG sequence is not boxed. Asterisks represent identical residues for proteins showing significant homology among regions 1 and 2. Plus signs refer to highly conserved amino acids among cytosine methyltransferases.

**cobL.** Mutants used in the genetic study were analyzed for their intracellular corrinoid content as described previously (10a). With the exception of mutant G612, no mutant accumulated any corrinoid (from cobyrinic acid to coenzyme B<sub>12</sub>) above the µg/liter threshold, which was significant, since other mutants already studied under similar conditions did accumulate corrinoids (10a). The absence of cobyrinic acid and downstream intermediates in the mutants indicates that they are blocked before cobyrinic acid. Mutant G612 synthesized significant amount of coenzyme B<sub>12</sub> (34% of the parent strain) and low level of cobinamide (2% of equivalent coenzyme B<sub>12</sub> synthesized by the parent strain). This mutant is probably not blocked between cobinamide and coenzyme B<sub>12</sub>, since the accumulation of cobinamide is very low and CobF is a protein with SAM affinity. SAM is unlikely to play a role in the last steps of cobalamin synthesis apart from 5,6-dimethylbenzimidazole synthesis (20), for which this mutant is not deficient (9). Mutants G615 and G616 synthesized coenzyme B<sub>12</sub> to a low level (17 and 13% of the parent level, respectively), indicating that they are partial mutants. The observation that these mutants are blocked before cobyrinic acid is consistent for mutant G613, which is a *cobI* mutant, since it is complemented by the *P. denitrificans cobI*

gene. Since *cobI* is the structural gene for SP<sub>2</sub>MT, this mutant is blocked before cobyrinic acid.

## DISCUSSION

Eight *cob* genes (*cobF* to *cobM*) have been identified on the 8.7-kb *P. denitrificans* fragment from complementation group A. *cobI* is the structural gene for SP<sub>2</sub>MT, which transfers one methyl group from SAM to precorrin-2. The other proteins encoded by these genes are involved in the transformation of precorrin-3 into cobyrinic acid, since the corresponding mutants do not accumulate cobyrinic acid or any other intermediates between cobyrinic acid and coenzyme B<sub>12</sub>. Five SAM-dependent methyl transfers occur between precorrin-3 and cobyrinic acid (20). On the basis of protein homology with SAM methyltransferases of the pathway (CobI and CobA) and SAM affinity for one of them (CobF), we propose that CobF, CobJ, CobL, and CobM carry four methyltransferase activities. We do not know whether another *cob* gene codes for a fifth methyltransferase or whether this activity is included in one of the potential or identified methyltransferases (as it was found that SUMT catalyzes two different methyl transfers [5]). SUMT trans-

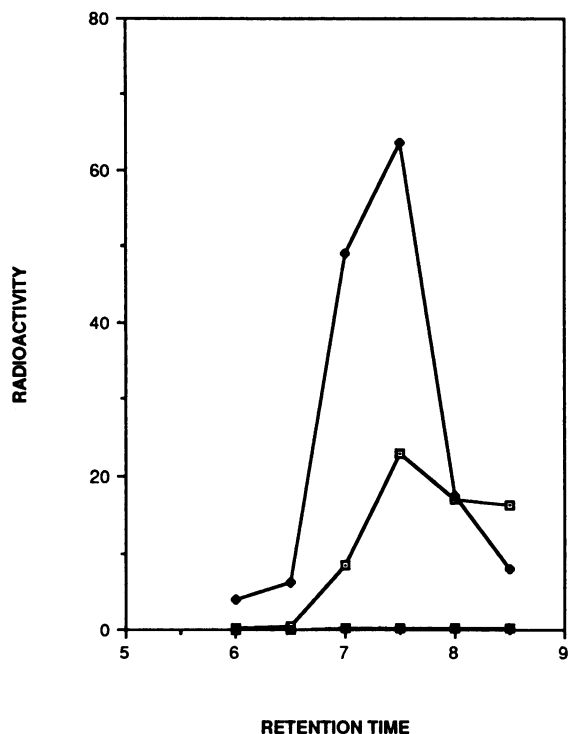


FIG. 6. SAM-binding affinity of CobA (♦) and CobF (◻). The retention times on the column are indicated in minutes; radioactivity of fractions representing 0.5 min of elution time is indicated in arbitrary units. Pure bovine serum albumin (◻) has been included in the test. Free SAM is eluted at 10.5 min. Proteins CobA and CobF are eluted at around 7.5 min.

fers two methyl groups at positions C-2 and C-7, which are at homologous  $\beta$ -positions on pyrroles A and B. Some of the methylations between precorrin-3 and cobyrinic acid are also at homologous positions, for instance methylations at C-5 and C-15. Therefore, one identified methyltransferase might carry out two methyl transfer reactions. CobL is proposed to be a methyltransferase and has, in addition, another domain which might be involved in one of the other reactions between precorrin-3 and cobyrinic acid; these reactions are decarboxylation at C-12, extrusion of C-20, and cobalt insertion (3). Similar hypotheses are postulated for CobG, CobH, and CobK, which are not a priori methyltransferases but are involved in this part of the pathway.

A class of Cob mutants (named CobI) in *Bacillus megaterium* has been identified as being blocked before cobalt insertion (8). Cobalt insertion has been proposed to occur just before cobyrinic acid (25). Therefore, most of these mutants might correspond to *cobF* to *cobM* mutations. Although these mutations are linked on the chromosome, the complementing inserts are not overlapping (32). It suggests that in *B. megaterium*, genes which are involved in the pathway before cobyrinic acid are not as tightly linked as in *P. denitrificans*. In *Salmonella typhimurium*, genes involved into the transformation of precorrin-2 into cobinamide are organized into a single operon (18). It is not known whether the genes involved in the part of the pathway before cobyrinic acid are as tightly linked as in *P. denitrificans*.

The eight *cob* genes, clustered on the 8.7-kb fragment, code for proteins involved in the same part of the pathway. This is very different from the *cobA* to *cobD* genes (10a),

which are coding for proteins implicated in different parts of the pathway (SUMT, cobyrinic acid *a,c*-diamide synthase, and transformation of cobyrinic acid into cobinamide).

Six of the eight genes present the characteristics of translationally coupled genes (22), i.e., overlapping or very close spacing between the initiation codon and the stop codon. Such an arrangement, which is rather unusual in gram-negative bacteria, suggests that these genes are part of the same operon. We do not know whether they are expressed at different or comparable levels, as reported for translationally coupled genes in *E. coli* (e.g., *trpA* and *trpB* [1]). Translational coupling occurs more frequently in *P. denitrificans* than in *E. coli*. So far, 10 of 13 *P. denitrificans* *cob* genes sequenced show characteristics of translationally coupled genes. It is not known whether such an organization is peculiar to *cob* genes in *P. denitrificans* or to overall gene organization in *P. denitrificans*. The close spacing of these genes or the translational coupling between six of them suggests that they are part of the same operon sharing the same regulation for the expression of enzymes acting in the same part of the cobalamin pathway; nevertheless, the presence of an internal promoter cannot be excluded.

Of the 13 *P. denitrificans* *cob* genes, 3 (*cobI*, *cobK*, and *cobD*) are proposed to have GUG as an initiation codon. The SP<sub>2</sub>MT NH<sub>2</sub>-terminal sequence allowed us to identify GUG as the initiation codon of *cobI*. Only the coding probability suggests that transcription in *cobK* and *cobD* is initiated at the proposed GUG codons; this remains to be demonstrated. These results suggest that GUG is used more frequently as an initiation codon in *P. denitrificans* than in *E. coli*, where it represents around 8% of the initiation codons (30). A possibility is that high-G+C organisms use GUG as translation initiation codons more frequently than other organisms do. In *Streptomyces* species, which are also high-G+C bacteria, GUG makes up more than 30% of the initiation codons, also a high-frequency level (17).

*P. denitrificans* is an obligate aerobic bacterium, and the intermediates between urogen III and cobyrinic acid are suggested to be very sensitive to oxygen (3). It is probable that aerobic cobalamin-synthetizing bacteria have evolved particular systems to protect the oxygen-sensitive intermediates. The intermediates might be channeled from urogen III to cobyrinic acid without being released in the cytoplasm, which would avoid contact with molecular oxygen. Biochemical studies may answer this question and lead to identification of intermediates. The finding that *cobF* to *cobM* are implicated between precorrin-2 and cobyrinic should help the understanding of this part of the coenzyme B<sub>12</sub> pathway. For instance, overexpression in *E. coli* and purification of CobF to CobM would allow us to test whether one of these proteins catalyzes the transformation of precorrin-3. It would answer the question concerning the next step in the pathway, i.e., methylation at C-17 or decarboxylation at C-12 as postulated (6). Gene cloning technology combined with biochemistry is therefore expected to bring a better understanding of the cobalamin pathway.

#### ACKNOWLEDGMENTS

We thank M. Couder for excellent technical assistance. J. Lunel, P. E. Bost, and J.-C. Brunie are acknowledged for their support during this work. We express our gratitude to J.-F. Mayaux and S. Lévy-Schil for critical reading of the manuscript and helpful discussions. We also thank the Centre de Traitement Interuniversitaire d'Informatique à Orientation Biomédicale, for nucleic and protein sequences analysis programs. F. de Bruijn, G. Ditta, and E. Nester are acknowledged for their gifts of strains, plasmids, or phages.

## LITERATURE CITED

1. Askoy, S., C. L. Squires, and C. Squires. 1984. Translational coupling of the *trpA* and *trpB* genes in the *Escherichia coli* tryptophan operon. *J. Bacteriol.* **157**:363–367.
2. Bagdasarian, M., R. Lurz, B. Rückert, F. C. Franklin, M. M. Bagdasarian, J. Frey, and K. Timmis. 1981. Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host vector system for gene cloning in *Pseudomonas*. *Gene* **16**:237–247.
3. Battersby, A. R., and E. MacDonald. 1982. Biosynthesis of the corrin macrocycle, p. 107–144. *In* D. Dolphin (ed.), B12, vol. 1. John Wiley & Sons, Inc., New York.
4. Ben-Bassat, A., K. Bauer, S.-Y. Chang, K. Mijamo, A. Boosman, and S. Chang. 1987. Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J. Bacteriol.* **169**:751–757.
5. Blanche, F., L. Debussche, D. Thibaut, J. Crouzet, and B. Cameron. 1989. Purification and characterization of *S*-adenosyl-L-methionine:uroporphyrinogen III methyltransferase from *Pseudomonas denitrificans*. *J. Bacteriol.* **171**:4222–4231.
6. Blanche, F., S. Handa, D. Thibaut, C. L. Gibson, F. J. Leeper, and A. R. Battersby. 1988. Biosynthesis of vitamin B<sub>12</sub>: when is the 12 $\beta$ -methyl group of the vitamin generated by acetate decarboxylation? *J. Chem. Soc. Chem. Commun.* **1988**:1117–1119.
7. Boyer, H. W., and D. Roulland-Dussoix. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.* **41**:459–472.
8. Brey, R. N., C. D. B. Banner, and J. B. Wolf. 1986. Cloning of multiple genes involved with cobalamin (vitamin B<sub>12</sub>) biosynthesis in *Bacillus megaterium*. *J. Bacteriol.* **167**:623–630.
9. Cameron, B., K. Briggs, S. Pridmore, G. Brefort, and J. Crouzet. 1989. Cloning and analysis of genes involved in coenzyme B<sub>12</sub> biosynthesis in *Pseudomonas denitrificans*. *J. Bacteriol.* **171**:547–557.
10. Casadaban, M. J., A. Martinez-Arias, S. T. Shapira, and J. Chou. 1983.  $\beta$ -Galactosidase gene fusion for analysing gene expression in *Escherichia coli* and yeast. *Methods Enzymol.* **100**:293–308.
- 10a. Crouzet, J., L. Cauchois, F. Blanche, L. Debussche, D. Thibaut, M.-C. Rouyez, S. Rigault, J.-F. Mayaux, and B. Cameron. 1990. Nucleotide sequence of a *Pseudomonas denitrificans* 5.4-kilobase DNA fragment containing five *cob* genes and identification of structural genes encoding *S*-adenosyl-L-methionine:uroporphyrinogen III methyltransferase and cobyrinic acid *a,c*-diamide synthase. *J. Bacteriol.* **172**:5968–5979.
11. Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt. 1979. A model of evolutionary change in proteins, p. 345–352. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.
12. De Bruijn, F. J., and J. R. Lupski. 1984. The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids—a review. *Gene* **27**:131–149.
13. Ditta, G., S. Stanfield, D. Corbin, and D. R. Helinski. 1980. Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* **77**:7347–7351.
14. Figurski, D. H., and D. R. Helinski. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. *Proc. Natl. Acad. Sci. USA* **76**:1648–1652.
15. Hirel, P.-H., J.-M. Schmitter, P. Dessen, and S. Blanquet. 1989. Extent of N-terminal methionine excision within *E. coli* proteins is governed by the side chain of the penultimate aminoacids. *Proc. Natl. Acad. Sci. USA* **86**:8247–8251.
16. Hopp, T. P., and K. R. Woods. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78**:3824–3828.
17. Hopwood, D. A., M. J. Bibb, K. F. Chater, G. R. Janssen, F. Malpartida, and C. F. Smith. 1986. Regulation of gene expression in antibiotic-producing *Streptomyces*, p. 251–276. *In* I. R. Booth and C. F. Higgins (ed.), Regulation of gene expression, 25 years on. Cambridge University Press, Cambridge.
18. Jeter, R. M., and J. R. Roth. 1987. Cobalamin (vitamin B<sub>12</sub>) biosynthetic genes of *Salmonella typhimurium*. *J. Bacteriol.* **169**:3189–3198.
19. Kanehisa, M. 1984. Use of statistical criteria for screening potential homologies in nucleic acids sequences. *Nucleic Acids Res.* **12**:203–215.
20. Leeper, F. J. 1989. The biosynthesis of porphyrins, chlorophylls and vitamin B<sub>12</sub>. *Nat. Prod. Rep.* **6**:171–203.
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. Normark, S., S. Bergström, T. Edlund, T. Grundström, B. Jaurin, F. Lindberg, and O. Olsson. 1983. Overlapping genes. *Annu. Rev. Genet.* **17**:499–525.
23. Platt, T. 1986. Transcription termination and the regulation of gene expression. *Annu. Rev. Biochem.* **55**:339–372.
24. Posfai, J., A. S. Bhagwat, G. Posfai, and R. J. Roberts. 1989. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **17**:2421–2435.
25. Scott, I. A., N. E. Mackenzie, P. J. Santander, P. E. Fagerness, G. Muller, E. Schneider, R. Seldmeier, and G. Worner. 1984. Biosynthesis of vitamin B<sub>12</sub>—timing of the methylation steps between uro'gen III and cobyrinic acid. *Bioorg. Chem.* **12**:356–362.
26. Shine, J., and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to non-sense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**:1342–1346.
27. Stachel, S. E., G. An, C. Flores, and E. W. Nester. 1985. A Tn3*lacZ* transposon for the random generation of  $\beta$ -galactosidase gene fusions: application to the analysis of gene expression in *Agrobacterium*. *EMBO J.* **4**:891–898.
28. Staden, R. 1984. Measurements of the effects that coding for a protein has on a DNA sequence and their use for finding genes. *Nucleic Acids Res.* **12**:551–567.
29. Staden, R., and A. D. McLachlan. 1982. Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucleic Acids Res.* **10**:141–156.
30. Stormo, G. D. 1986. Translation initiation, p. 195–224. *In* W. Reznikoff and L. Gold (ed.), Maximizing gene expression. Butterworth, Boston.
31. West, S. E. H., and B. Iglewski. 1988. Codon usage in *Pseudomonas aeruginosa*. *Nucleic Acids Res.* **16**:9323–9335.
32. Wolf, J. B., and R. N. Brey. 1986. Isolation and genetic characterization of *Bacillus megaterium* cobalamin biosynthesis-deficient mutants. *J. Bacteriol.* **166**:51–58.