

Acinetobacter calcoaceticus Genes Involved in Biosynthesis of the Coenzyme Pyrrolo-Quinoline-Quinone: Nucleotide Sequence and Expression in *Escherichia coli* K-12

NORA GOOSEN,* HAROLD P. A. HORSMAN, RENÉ G. M. HUINEN, AND PIETER VAN DE PUTTE
Department of Molecular Genetics, University of Leiden, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands

Received 23 June 1988/Accepted 26 September 1988

Synthesis of the coenzyme pyrrolo-quinoline-quinone (PQQ) from *Acinetobacter calcoaceticus* requires the products of at least four different genes. In this paper we present the nucleotide sequence of a 5,085-base-pair DNA fragment containing these four genes. Within the DNA fragment three reading frames are present, coding for proteins of M_r 10,800, 29,700, and 43,600 and corresponding to three of the PQQ genes. In the DNA region where the fourth PQQ gene was mapped the largest possible reading frame encodes for a polypeptide of only 24 amino acids. Still, the expression of this region is essential for the biosynthesis of PQQ. A possible role for this DNA region is discussed. Sandwiched between two PQQ genes an additional reading frame is present, coding for a protein of M_r 33,600. This gene, which is probably transcribed in the same operon as three of the PQQ genes, seems not required for PQQ synthesis. Expression of the PQQ genes in *Acinetobacter lwoffii* and *Escherichia coli* K-12 led to the synthesis of the coenzyme in these organisms.

Pyrrolo-quinoline-quinone (PQQ) is used by a variety of organisms as cofactor in dehydrogenase reactions. In *Pseudomonas* species and *Acinetobacter calcoaceticus* PQQ is associated with glucose dehydrogenase (GDH) (1, 5, 22). In *Pseudomonas testosteroni* (7) and *Pseudomonas stutzeri* (M. van Kleef, personal communication) and PQQ-dependent alcohol dehydrogenase is present. *Thiobacillus versutus* (19) and *Paracoccus denitrificans* (2) synthesize methylamine dehydrogenase, to which PQQ is covalently linked.

Also in mammals the presence of PQQ as cofactor has been demonstrated. In bovine plasma (10, 15) and porcine kidney (20) PQQ-dependent amine oxidases have been identified, and it has been suggested (4, 15) that amine oxidases from other mammals, *Saccharomyces cerevisiae*, *Aspergillus*, and plants also use PQQ as a coenzyme.

In some bacterial species, viz., *A. calcoaceticus*, PQQ is synthesized constitutively in the absence of the enzyme's substrate (i.e., glucose), whereas in *P. stutzeri* PQQ can be detected only when ethanol is present in the culture medium as an inducing agent (M. van Kleef, personal communication). *Escherichia coli* K-12 and *Acinetobacter lwoffii* do not produce PQQ. These organisms synthesize the apo-GDH enzyme and are dependent on uptake of PQQ from the culture medium for the constitution of the holo-GDH (8, 21).

The biosynthetic pathway of PQQ has not yet been elucidated. Our approach to this problem was to clone the genes and to study the encoded gene products that are involved in PQQ synthesis. Recently we reported the isolation of a plasmid containing four cloned PQQ genes from *A. calcoaceticus* on a 5,000-base-pair (bp) fragment (6). In this paper we report the complete nucleotide sequence of this fragment and show that the genes from *A. calcoaceticus* can direct the synthesis of the cofactor in other microorganisms.

MATERIALS AND METHODS

Bacterial strains and culture conditions. The following bacterial strains have been previously described: *E. coli* JM101 ($\Delta lac-pro thi F' traD36 proAB lacI^q \Delta M15$) (13), *E.*

coli PPA41 (*thi ptsI*) (8), and *A. lwoffii* (21). These strains were cultured in L broth, on L plates, or in defined minimal medium (14) as indicated. Bacterial matings were performed as described previously (6). Acid production by *A. lwoffii* was tested on L plates containing 0.4% glucose and a few drops of phenol red. Acid production resulted in yellow plates, whereas the plates of non-acid-producing strains remained red.

Chemicals and reagents. Restriction endonucleases were from Bethesda Research Laboratories, Inc. T4 DNA ligase and DNA polymerase (Klenow enzyme) were from Pharmacia Fine Chemicals. All recombinant DNA techniques were essentially as described previously (12). Purified PQQ was a gift from M. van Kleef. The universal primer for DNA sequencing (5'-GTAAAACGACGGCCAGT-3') was purchased from Pharmacia. The oligonucleotide primer corresponding to the Tn5 end sequences (5'-GGTTCCGTTTCAG GACGCTAC-3') was kindly synthesized by J. H. van Boom. Radiochemicals and in vitro coupled transcription-translation kits were from Amersham Corp.

DNA sequencing. Transformation of strain JM101 was performed as described previously (13). M13 phage particles were purified, and the single-stranded DNA was extracted by the method of Sanger et al. (16). For the sequencing reaction the primer extension method (17) was used. Sequences were compiled and analyzed by using the sequence analysis software package of the University of Wisconsin Genetics Computer Group (3).

Sequence strategy. A schematic representation of the sequence strategy is indicated in Fig. 1. The *EcoRI-BamHI* fragment of pSS2, containing the four PQQ genes, was subcloned in M13mp19, resulting in pSS144. Deletion derivatives for DNA sequencing were isolated by an adaptation of the method of Hong (9). pSS144 was partially digested with *Sau3A*, *RsaI*, or *DraI*. Linear fragments were isolated, digested with *BamHI* (for *Sau3A*) or *HincII* (for *RsaI* and *DraI*), recircularized, and transformed to JM101. Thus the different *Sau3A*, *RsaI*, and *DraI* sites were brought close to the primer for sequencing.

For sequencing of the opposing strand, different subclones

* Corresponding author.

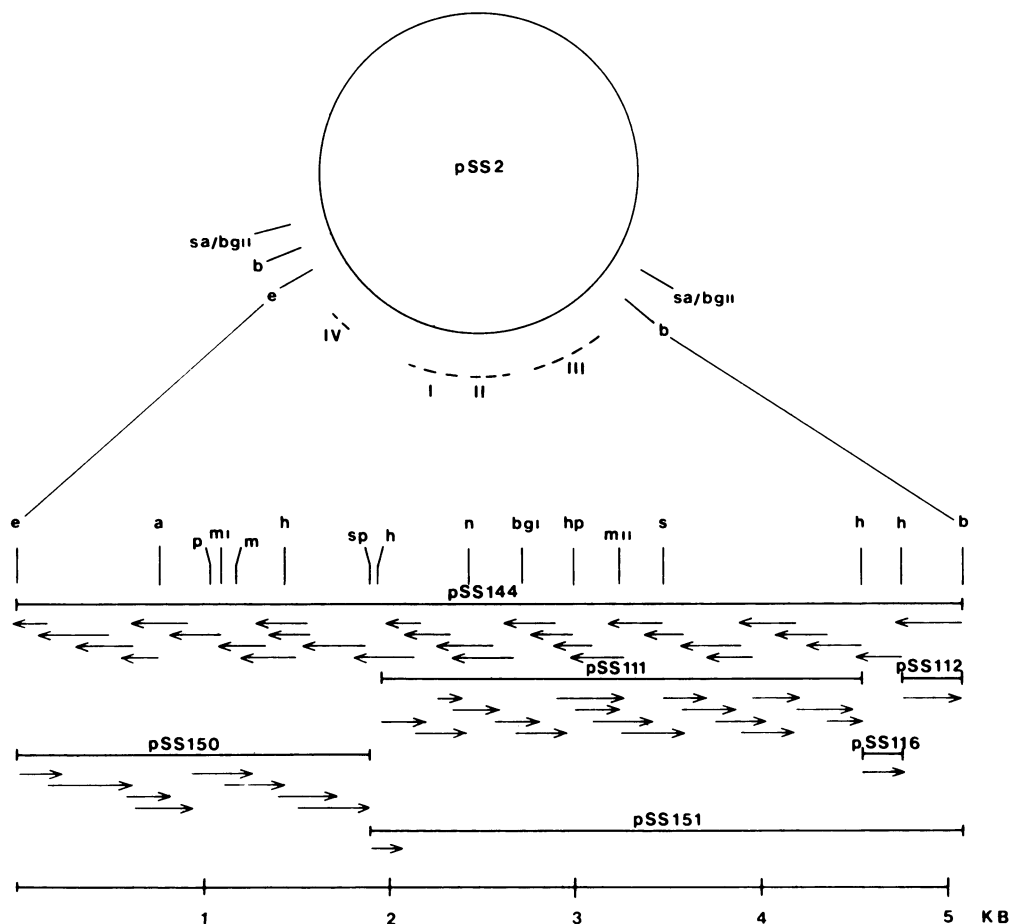


FIG. 1. Schematic representation of the sequencing strategy. Plasmid pSS2, containing a 5.5-kb *Sau3A* insert in the *Bgl*III site of vector pRK290, has been described previously (6). Only restriction sites that occur with low frequency are indicated: a, *Acc*I; b, *Bam*HI; bgi, *Bgl*II; bgii, *Bgl*III; e, *Eco*RI; h, *Hind*III; hp, *Hpa*I; m, *Mlu*I; mi, *Mst*I; mii, *Mst*II; n, *Nco*I; p, *Pvu*II; s, *Sac*I; sa, *Sau*3A; sp, *Sph*I. pSS144 contains the *Eco*RI-*Bam*HI fragment in M13mp19, pSS111 contains the *Hind*III fragment in M13mp19, pSS112 has the *Hind*III-*Bam*HI fragment inserted in M13mp18, pSS150 has the *Eco*RI(*Sal*I)-*Sph*I fragment in M13mp19, pSS116 contains the *Hind*III fragment in M13mp18, and pSS151 carries the *Sph*I-*Bam*HI fragment in M13mp18. The arrows show the location, direction, and length of nucleotide sequence obtained from each template.

were constructed. First the internal 2,600-bp *Hind*III fragment of pSS2 was cloned into M13mp19, the correct orientation was selected (pSS111), and deletion derivatives were isolated as described above. A *Sal*I linker fragment was inserted into the *Eco*RI site of pSS2, and then the 1,900-bp *Sal*I-*Sph*I fragment was subcloned in M13mp19 (pSS150). Also from pSS150 deletions were constructed by the method described above. The remaining right part of the *Eco*RI-*Bam*HI fragment was sequenced by subcloning the 210-bp *Hind*III fragment and the 320-bp *Hind*III-*Bam*HI fragment in M13mp18 (pSS116 and pSS112, respectively). Finally, the gap between *Sph*I and *Hind*III was sequenced by insertion of the *Sph*I-*Bam*HI fragment in M13mp18 (pSS151).

RESULTS

Nucleotide sequence of the PQQ genes. Plasmid pSS2 contains a 5.5-kilobase insert from the *A. calcoaceticus* chromosome which complements four different classes of chromosomal PQQ⁻ mutants and which is therefore expected to carry four different genes involved in PQQ synthesis (6). With Tn5 insertions these cloned genes (I, II, III, and IV) were approximately mapped within a 5-kilobase *Eco*RI-

*Bam*HI fragment (6). We determined the complete nucleotide sequence of this fragment as described in Materials and Methods. It was comprised of 5,087 bp and showed a rather high A+T content (61%). Within the sequenced fragment different open reading frames (ORFs) were indicated. The complete nucleotide sequence and the deduced amino acid sequence of the ORFs are reported in Fig. 2, and a schematic representation of the ORFs is given in Fig. 3.

To determine whether these ORFs correspond to the different PQQ genes, we also determined the precise insertion points of the Tn5 insertions. Therefore the pSS2-Tn5 plasmids were restricted with *Eco*RI and *Pst*I. Since *Eco*RI cuts at the left border of the 5-kilobase fragment (Fig. 1) and *Pst*I cuts within Tn5, a fragment was generated containing one end of Tn5 and the PQQ sequences located to the left of the insertion point. This fragment was inserted into M13mp18, and the fusion point was sequenced by using a synthetic oligonucleotide (5'-GGTTCGGTTCAGGACGC TAC-3') that was complementary to the end of Tn5 as a primer. As a control also the *Bam*HI-*Pst*I fragment containing the other end of Tn5 fused to PQQ sequences located to the right of the insertion point was cloned and sequenced. In

50

G AAT TOC ACC ATA AGT TTT GAC TTT ATA AGC AGG GAA GAT TGT TTT ACT GGT GTC GTT TTT CAT TGC TAC GTC AGC AGC GAG GAA ATC AAA TTT TGC ACC
 C TTA AGG TGG TAT TCA AAA CTG AAA TAT TGG TOC CTT CTA ACA AAA TGA OCA CAG CAA AAA GTA AGC ATG CAA TGG TGG CTC CTT TAG TTT AAA AGC TGG
 Ile Gly Gly Tyr Thr Lys Val Lys Tyr Ala Pro Phe Ile Thr Lys Ser Thr Asp Asn Lys Met Ala Val Asn Ala Ala Leu Phe Asp Phe Lys Ala Gly

150

ACT AAA GTT TTT ATT AAT TTG ACA AGG TAC AGT GCT GGT ATA TTG CTG ACA TOC ACC ATT TTT TAA ACG GCG TAA TTC GTC TGT ACC ACG GTC AAA TTC
 TGA TTT CAA AAA TAA TTA AAC TGT TOC ATG TCA OGA OCA TAT AAC GAC TGT AGG TGG TAA AAA ATT TGC OGC ATT AAG CAG ACA TGG TGC CAG TTT AAG
 Ser Phe Asn Lys Asn Ile Gln Cys Pro Val Thr Ser Thr Tyr Gln Gln Cys Gly Gly Asn Lys Leu Arg Arg Leu Glu Asp Thr Gly Arg Asp Phe Glu

200

ATG GAT ACC TAC AGC ATC AAA GTC AAT TTG AAT ATC ATT CAT GAC TTC AAT GGT TGG TTC ATC TAA AAA TAA AGA AGA GGT GAG AGG AGA AGC GCT AAT
 TAC CAA TGG ATG TGG TAG TTT CAG TTA AAC TTA TAA TAA TAA GTA CTG AAG TTA OCA ACC AAG TAG ATT TTT ATT TCT TCT OCA CTC TOC TCT TOG OGA TTA
 His Asn Gly Val Ala Asp Phe Asp Ile Gln Ile Asp Asn Met Val Glu Ile Thr Pro Glu Asp Leu Phe Leu Ser Ser Thr Leu Pro Ser Ala Ser Ile

300

TAA ATC GCC TGC CGA AAC CAC AGC ATT ATT TGG ATT CTC TGC TTT AAG TTT TTT AAT CCG ATC GCC GAA GTA ACT CAC TOC TOC TAC AGG AAT ACG AAC
 ATT TAG CCG ACG GCT TTG GIG TCG TAA TAA ACC TAA GAG ACG AAA TTC AAA AAA TTA GCG TAG CCG CTT CAT TGA GTG AGG ACG ATC TOC TTA TGC TTG
 Leu Asp Gly Ala Ser Val Val Ala Asn Asn Pro Asn Glu Ala Lys Leu Lys Lys Ile Ala Asp Ala Phe Tyr Ser Val Gly Gly Val Pro Ile Arg Val

400

AGA TTG AGC AGC ATC ATT TGG ATT TGG TGC CTC GAC ATA ACG CTT AGG TGG CTC AAG ATT TOC GIG AAA GTC ATT AAA ACG TAA AAT ATT TAC GGT TTG
 TCT AAC TGG TGG TAG TAA ACC TAA ACC ACG GAG CTG TAT TGC GAA TOC ACC GAG TTT TAA AGG CAC TTT CAG TAA TTT TGG ATT TTA TAA ATG OCA AAC
 Ser Gln Ala Ala Asp Asn Pro Asn Pro Ala Glu Val Tyr Arg Lys Pro Pro Glu Leu Asn Gly His Phe Asp Asn Phe Ala Leu Ile Asn Val Thr Gln

500

ATT GCT TCT TTG AGC AGG TGG TGC TGT GTC ATT ATC ATC ATC GTC ATT ACA TOC CAC CAA TAA TGC TAC AGC CAG AGA AAG TGT TGT TAC TTT AAA AAG
 TAA OGA AGA AAC TOG TOC ACC ACG ACA CAG TAA TAG TAG TAG CAG TAA TGT AGG GTC GTT ATT ACG ATC TOG GTC TCT TTC ACA ACA ATG AAA TTT TTC
 Asn Ser Arg Gln Ala Pro Pro Ala Thr Asp Asn Asp Asp Asp Asn Cys Gly Val Leu Leu Ala Val Ala Leu Ser Leu Thr Val Lys Phe Leu

600

TAT TGT TAA ATT TTT CAT TTTTATCATTTGATCAATTAAGACCCATAAAAAATACGCTTTTAGTGGTAAATGTTTGAACCAACCAATAAGCAAAATATGTCGAAATATTAATTTGGACAA
 ATA ACA ATT TAA AAA GAA AAAAAATGATAAAGTAGTAAATTCCTGGTATTTTTATGCGAAAATACACACTTACAAACTACTTTGGTGGTATTCGTTTTATACAAAGCTTTAATATTAACCTGTT
 Ile Thr Leu Asn Lys Met

750

800

850

AAAAACAATATTOGAA CAATATCTAGCTAGGGGGTGGCTACAAAACCAATTTTTTTATTTAGATAATAATTTTTTAATTTGOCITATAAATCTTAATCAAGTATTCATTTCATAAGA TTAATATGAACCGG
 TTTTGTGTTAATAGCTTGTATAGAAATGATGATCCACAGATGTTTGGTGTAAAAAAAATAAACTTATTTAAAAAATAAAGGAAATTTAGAATTAGTCTACTAAGTAAAGTATCTAATATTAACCTGGCC
 TTTTGTGTTAATAGCTTGTATAGAAATGATGATCCACAGATGTTTGGTGTAAAAAAAATAAACTTATTTAAAAAATAAAGGAAATTTAGAATTAGTCTACTAAGTAAAGTATCTAATATTAACCTGGCC

Tn 5-18

M Q W T K P A F T D L R I G F E V T M Y F E A R *

950

AGATAATTTATGCAATGAGCACTAAACCACTTTTATTTACCCATGGTGTGAGAGTTACAAATGTTTGAAGCAAGTTAATCAGTTTCATCAAAATATGTCATAGCOCTAGTCTTGTCTAGGGCTT
 TCTATTAATATGATGATCCTGATTTGGTGGAAATGCTAAATGCGTAACCAAACTTCAATGTTACATGAAACTTGTGCAATTTAGTCAAAAGTAGTTATACAGTATTOGGGATCAGAAAGATCOOGAA

1000

1050

Met Tyr Ile Tyr Val Leu Gly Ser Ala Ala Gly Gly Gly Phe Pro Gln Trp Asn Cys Asn Cys Pro Asn Cys His Gly
 ATCTCTTCTAATAAGATATTAATTTTATG TAT ATT TAT GTT TTA GGT TCA GCT GCT GGA GGC GGG TTT CCG CAG TGG AAT TGT AAT TGC CCG AAT TGT CAT GGT
 TAGAGAAGATTAATTTATTAATTAATAA TAC ATA TAA ATA CAA AAT CCA AGT CGA CGA CCT CCG CCC AAA GGC GTC ACC TTA ACA TTA ACG GGC TTA ACA GTA CCA

1100

1150

Val Arg Thr Gly Thr Ile Gln Ala Lys Ala Arg Thr Gln Ser Ser Ile Ala Val Ser Glu Asn Gly Thr Asp Trp Val Leu Leu Asn Ala Ser Pro Asp
 GTC CCG ACA GGT ACA ATC CAA GCC AAA GCC GGT ACT CAA TCA TCG AAT GCT GGT TCT GAA AAT GGA ACA GAC TGG GIT TTA TTA AAC GCG TCA CCT GAC
 CAC GCG TGT CCA TGT TAG GTT CCG TTT CCG GCA TGA GTT AGT AGC TAA OGA CAA AGA CTT TTA OCT TGT CTG ACC CAA AAT AAT TTG CCG AGT GGA CTG

1200

1250

Ile Arg Gln Gln Leu Phe Glu Phe Lys Ala Ala Gln Pro Ala Arg Lys Leu Arg Asp Thr Gly Ile Ile Ser Val Ile Leu Met Asp Ser Gln Leu Asp
 AIC GGT CAG CAA CTC TTT GAA TTT AAA GCA GCG CAA CCA GCG GGT AAA TTG GGT GAT ACA GGA ATA ATA AGC GTT ATT TTA ATG AAC GGC TTT OCA GTA TTT
 TAG GCA GTC GGT GAG AAG CTT AAA TTT CGT CCG GTT GGT CCG GCA TTT AAC GCA CTA TGT CCT TAT TAT TGG CAA TAA AAT TAC CTG TCA GTC AAT CTG

1300

Tn 5-16

1350

His Thr Thr Gly Leu Leu Thr Leu Arg Glu Gly Cys Pro Met Asn Val Trp Cys Thr Glu Met Val His Gln Asp Leu Thr Asn Gly Phe Pro Val Phe
 CAT ACC ACT GGG CTT TTA ACC TTA CCGT GAA GGT TGT CCA ATG AAT GTA TGG TGT ACA GAA ATG GTC CAC CAA GAT TTA ACG AAC GGC TTT OCA GTA TTT
 GTA TGG TGA CCC GAA AAT TGG AAT GCA CTT CCA ACA GGT TAC TTA CAT ACC ACA TGT CTT TAC CAG GIG GIT CTA AAT TGC TTG CCG AAA GGT CAT AAA

1400

1450

Asn Met Leu Lys His Trp Asn Gly Gly Leu Gln Tyr His Glu Ile Asn Pro Lys Gln Ala Phe Lys Ile Asp Gly Phe Glu Asn Leu Glu Phe Leu Pro
 AAT ATG CTC AAA CAC TGG AAT GGT GGT CTT CAA TAT CAT GAG ATT AAC CCT AAA CAA GCT TTT AAA ATT GAT GGT TTT GAA AAT TTA GAA TTT TTA CCC
 TTA TAC GAG TTT GIG ACC TTA OCA CCA GAA GAT ATA GTA CTC TAA TTG GGA TTT GTT OGA AAA TTT TAA CTA CCA AAA CTT TTA AAT CTT AAA AAT GGG

1500

1550

Leu Ile Ile Lys Ser Ala Ala Pro Pro Tyr Ser Pro His Arg Asn Asn Pro His Asp Gly Asp Asn Ile Ala Leu Ile Ile Lys Asp His Lys Thr Gln
 TTA ATT AAT AAA AGT GCA GCT CCA CCA TAT TCA CCG CAT OGA AAT AAC CCA CAT GAT GGC GAT AAT AAT GCT TTA ATT ATC AAA GAT CAT AAG ACA CAA
 AAT TAA TAA TTT TCA CCG OGA GGT GGT ATA AGT GGC GTA GCT TTA TTG GGT GTA CTA CCG CTA TTA TAA OGA AAT TAA TAG TTT CTA GTA TTC TGT GIT

1600

1650

Tn 5-20

Lys Gln Leu Phe Tyr Ala Pro Gly Leu Gly Lys Ile Asp Asp Gln Ile Met Gln Ile Met Gln Ser Ser Asp Cys Val Met Ile Asp Gly Thr Leu Trp
 AAA CAG TTG TTC TAT CCG CCG GGT CTT GGA AAA ATT GAC GAT CAA ATC ATG CAG ATT ATG CAG AAT TCA GAC TGT GTC ATG ATT GAT GGT ACG CTT TGG
 TTT GTC AAC AAG ATA CCG GGC CCA GAA OCT TTT TAA CTG CTA GTT TAG TAC GTC TAA TAC GTC TCA AGT CTG ACA CAG TAC TAA CTA CCA TGC GAA ACC

1700

1750

Thr Asp Asp Glu Met Gln Gln Thr Gly Val Gly Thr Lys Thr Gly Arg Glu Met Gly His Leu Tyr Ile Ser Gly Glu Gly Gly Ser Leu Ser Tyr Leu
 ACA GAC GAT GAA ATG CAA CAA ACA GGG GGT GGA ACT AAA ACT GGC CCG GAA ATG GGG CAT TTA TAT ATT AGT GGC GAA GGT GGT TCA CTG TCT TAT TTA
 TGT CTG CTA CTT TAC GTT GTT TGT CCC CAA OCT TGA TTT TGA CCG GCG CTT TAC CCC GTA AAT ATA TAA TCA CCG CTT CCA CCA AGT GAC AGA ATA AAT

1800

1850

Asn Lys Leu Ser Thr Pro Lys Lys Val Leu Ile His Ile Asn Asn Thr Asn Pro Ile Leu Asn Glu Asn Ser Ser Gln Phe Ala Glu Leu Lys Ala Asn
 AAC AAG CTG AGT ACA OCT AAA AAA GTC CTG ATT CAT ATT AAC AAT ACC AAT CCA ATT TTA AAT GAA AAC TCT AGT CAG TTT GCT GAG CTT AAA GCA AAT
 TTG TTC GAC TCA TGT GGA TTT TTT CAC GAC TAA GTA TAA TTG TTA TGG TTA GGT TAA AAT TTA CTT TTG AGA TCA GTC AAA OGA CTC GAA TTT GGT TTA

this way the insertion points of nine different Tn5 insertions were determined (Fig. 2). The phenotypes of the pSS2 plasmids containing these Tn5 insertions (13) are listed in Table 1.

Transposon insertions Tn5-16, Tn5-20, and Tn5-7 were located within an ORF (indicated as gene V) starting with

ATG (position 1010) and ending with TAA (position 1921), which could code for a protein of M_r 33,600. It has been shown that chromosomal deletion mutants that lack the complete PQQ region (including gene V) can still be complemented by these Tn5 insertions (13). Apparently gene V is not involved in PQQ biosynthesis.

1900
Tn 5-7
 Gly Val Glu Val Ala Tyr Asp Gly Met Gln Ile Glu Leu * Met Thr Gln Thr Pro Glu Ala Leu Thr Thr Glu Gln Phe Lys Gln Ala Ile
 ATC GAT AAA GGC CAG TAT TAT CAT ATC TAT CAT CCA TTT CAT GIG ATG ATG TAT GAA GGT AAA GGC ACC CAG CAA CAA ATT CAG GCT TGG GTT GCA AAC
 CGC GTA GAA GGT GCC TAT TAT CAT ATC TAT CAT CCA TTT CAT GIG ATG ATG TAT GAA GGT AAA GGC ACC CAG CAA CAA ATT CAG GCT TGG GTT GCA AAC
 CGC CAT CTT CAA CCG ATA CTG CCG TAC GTC TAA CTT GAG ATT CCACCTTGG TAC TGA GTT TGT GGA CTT CGA AAT TGC TGA CTT GTT AAG TTT GTT CCG TAA

2000
 Ile Asp Lys Gly Gln Tyr Tyr His Ile Tyr His Pro Phe His Val Met Met Tyr Glu Gly Lys Ala Thr Gln Gln Gln Ile Gln Ala Trp Val Ala Asn
 ATC GAT AAA GGC CAG TAT TAT CAT ATC TAT CAT CCA TTT CAT GIG ATG ATG TAT GAA GGT AAA GGC ACC CAG CAA CAA ATT CAG GCT TGG GTT GCA AAC
 TAG CTA TTT CCG GTC ATA ATA GTA TAG ATA GTA GGT AAA GTA CAC TAC TAC ATA CTT CCA TTT CCG TGG GTC GTT GTT TAA GTC CGA ACC CAA CGT TTG

2100
 Arg Tyr Tyr Tyr Gln Ile Asn Ile Pro Leu Lys Asp Ala Ala Ile Met Ala Asn Cys Pro Asp Gln Arg Val Arg Gln Glu Trp Ile Gln Arg Met Ile
 CGA TAT TAC TAT CAA ATT AAT ATT CCG CTT AAA GAT GGG GGG ATT ATG GCA AAT TGC CCT GAT CAG CGA GTC CGT CAA GAA TGG ATT CAA CGC ATG ATC
 GCT ATA ATG ATA GTT TAA TTA TAA GGC GAA TTT CTA CCG CCG TAA TAC CGT TTA ACG GGA CTA GTC GCT CAG GCA GGT CTT CAA TAA GAT GGT TAC TAC

2200
 Asp Gln Asp Gly Glu Tyr Pro Asp Gly Gly Gly Arg Glu Ala Trp Leu Arg Leu Ala Glu Ala Val Gly Leu Ser Arg Glu Gln Val Ile Ser Glu Glu
 GAT CAA GAT GGT GAA TAT CCA GAT GGC GGT GGT CGA GAA GCA TGG CTA CCG TTA GCT GAA GCT GIG GGC TTG AGT CCG GAA CAA GAT ATT TCT GAA GAG
 CTA GTT CTA CCA CTT ATA GAT CTA CCG CCA CCA GCT CTT CGT ACC GAT GGG AAT CGA CTT CGA CAC CCG AAC TCA GCG CTT GTT CAA TAA AGA CTT CTT

2300
Tn 5-17
 Leu Val Leu Pro Gly Val Arg Phe Ala Val Asp Ala Tyr Val Asn Phe Arg Arg Arg Ala Ser Trp Arg Glu Ala Ala Ser Ser Ser Leu Thr Glu Leu
 TTA GTT TTA CCG GGC GTT CCG TTC GGG GTA GAT GCT TAT GTA AAC TTT CCA CCG CCG CCG TCA TGG CCG GAA GCA GCA AGC AGC TCT TTA ACT GAA CTT
 AAT CAA AAT GGC CCG CAA GGG AAG CCG CAT CTA CCG ATA CAT TTG AAA GCT CCG CCA CGA AGT ACC CGA CTT CGT CGT TCG TCG AGA AAT TGA CTT GAA

2400
 Phe Ala Pro Gln Ile His Gln Ser Arg Leu Asp Ser Trp Pro Gln His Tyr Pro Trp Ile Asp Asp Lys Gly Tyr Glu Tyr Phe Arg Ser Arg Leu Ser
 TTT GCT CCA CAA ATT CAT CAA TCA AGA CTT GAT TCA TGG CCA CAA CAT TAC CCA TGG ATT GAT GAT AAA GGC TAT GAG TAT TTT CGT TCT CGT TTG AGT
 AAA CGA GGT GTT TAA GTA GTT AGT TCT GAA CTA AGT ACC GGT GTT GTA ATG GGT ACC TAA CTA CTA TTT CCG ATA CTC ATA AAA GCA AGA GCA AAC TCA

2500
 Gln Ala Arg Arg Asp Val Glu His Gly Leu Thr Ile Thr Leu Asp Ser Phe Thr Thr Phe Glu Gln Gln Gln Glu Arg Met Leu Glu Ile Leu Gln Phe Lys
 CAG CCA CCG CCG GAC GTT GAA CAT GGT CTA AGG ATT AGC CTT GAT TCA TTT ACA ACT TTT GAG CAA CAA GAA CCA ATG CTT GAA ATA TTG CAG TTT AAA
 GTC CCG GCG GCA CTG CAA CTT GTA CCA GAT TGC TAA TGC GAA CTA AGT AAA TGT TGA AAA CTC GTT GTT CTT CCG TAC GAA CTT TAT AAC GTC AAA TTT

2600
 Leu Asp Ile Leu Trp Ser Ile Leu Asp Ala Leu Thr Leu Ala Tyr Val His Asn Glu Ala Pro Tyr His Ser Val Thr Ser Lys Arg Val Trp His Lys
 CTG GAT ATT TTA TGG AGC ATT TTA GAT GCT TTA ACT TTG GCA TAT GTT CAC AAT GAA GCA CCA TAT CAC AGC GIG ACA AGC AAA CGA GAT TGG CAC AAA
 GAC CTA TAA AAT ACC TCG TAA AAT CTA CGA AAT TGA AAC CGT ATA CAA GIG TTA CTT CGT GGT ATA GIG TCG CAC TGT TCG TTT CCT CAA ACC GIG TTT

2700
 Met Asn Lys Glu Gln Phe Asp Leu Asn Leu Val Pro Thr Trp Arg Gln Gly Tyr Arg Phe Gln Phe Glu Pro Ala Gln Asn Gly Phe
 Gly Leu Phe Lys *
 GGA CTA TTT AAA TGAAT AAA GAG CAG TTT GAT TTA AAC CTT GIG CCA ACA TGG CGA CAA GGT TAC CGT TTT CAG TTC CAG CCA GCT CAA AAT GGT TTT
 CCT GAT AAA TTT ACTTA TTT CTC GTC AAA CTA AAT TTG GAA CAC GGT TGT ACC GCT GTT CCA ATG GCA AAA GTC AAG CTC GGT CGA GTT TTA CCA AAA

2800
 Val Ile Leu Tyr Pro Glu Gly Met Ile Lys Leu Asn Glu Ser Ala Gly Ala Ile Gly Gln Tyr Ile Asp Gly Gln Gln Asn Val Ser Ala Ile Ile Ala
 GIG ATT TTA TAT OCT GAA GGC ATG ATT AAG TTA AAT GAG AGT GCA GGG GCA ATC GGG CAA TAT ATT GAT GGG CAG CAA AAT GTT TCT GCA ATT ATT GCT
 CAG TAA AAT ATA GCA CTT CCG TAC TAA TTC AAT TTA TIC CCA CCG CCG TAT CCG CCG GTT ATA TAA CCA CCG GTC GTT TTA CAA AGA CCG TAA TAA CGA

2900
 Gln Leu Lys Gln Lys Phe Gly Asp Ile Ser Glu Ile Asp Gln Asp Val Val Asp Tyr Met Leu Val Ala Lys Gln Gln His Trp Ile Asp Leu Val *
 CAG TTA AAG CAA AAA TTT GGT GAT ATT TCT GAG ATT GAT CAA GAT GTA GIG GAT TAT ATG CTG GTT GCA AAA CAG CAA CAC TGG ATT GAT TTA GTA TGACA
 GTC AAT TTC GTT TTT AAA CCA TAA AGA CTC TAA CTA GTT CTA CAT CAC CTA ATA TAC GAC CAA CCG TTT GTC GTT GIG ACC TAA CTA AAT CAT ACTGT

3000
Tn 5-9
 Glu Gly Val Gly Leu Pro Leu Trp Leu Leu Ala Glu Leu Thr Tyr Arg Cys Pro Leu Gln Cys Pro Tyr Cys Ser Asn Pro Leu Asp Tyr Ala Gln His
 GAA AAA TTT GGC CCG TCG TCA TGG TTA TTA TTA GCG GAG TTA ACT TAT CPT TGT CCG CTA CAA TGC CCA TAT TGC TCA AAC CCG TTG GAC TAT GCT CAG CAT
 CTT CCA CAA CCG GAC CGA AAT ACC AAT AAT CCG CTC AAT TGA ATA GCA ACA GGC GAT GTT ACG GGT ATA ACG AGT TTG GGG AAC CTG ATA CGA GTC GTA

3100
Tn 5-2
 Lys Asn Glu Leu Thr Thr Gln Glu Trp Phe Asp Val Phe Asp Gln Ala Arg Gln Met Gly Ala Val Gln Leu Gly Phe Ser Gly Gly Glu Pro Leu Val
 AAA AAT GAA CTA ACC ACT CAA GAG TGG TTT GAC GTT TTT GAT CAG GCA CCG TCA ATG GGT GCT GTT CAA CTC GGC TTT TCT GGC GGT GAA CCA CTG GTA
 TTT TTA CTT GAT TGG TGA GTT CTC ACC AAA CTG CAA AAA CTA GTC CCG CGA GAT TAC CCA CGA CAA GTT GAG CCG AAA AGA CCG CCA CTT GGT GAC CAT

3200
 Arg Gln Asp Leu Glu Gln Leu Val Ala His Ala His Gln Gln Gly Phe Tyr Thr Asn Leu Ile Thr Ser Gly Met Gly Leu Thr Glu Gln Arg Ile Ala
 CGT CAA GAT TTA GAA CAA CTC GTT GCA CAT GCT CAT CAA CAA GGG TTT TAT ACC AAC CTC ATT ACC TCA GGC ATG GGT CTT ACC GAA CAA CCG ATT GCT
 GCA GTT CTA AAT CTT GTT GAG CAA CCG GTA CGA GTA GTT GTT CCG AAA ATA TGG TTG GAG TAA TGG AGT CGC TAC CCA GAA TGG CTT GTT GCA TAA CGA

3300
 Asp Leu Lys Gln Ala Gly Leu Asp His Ile Gln Val Ser Phe Gln Ala Ser Asp Pro Val Val Asn Asp Ala Leu Ala Gly Ser Lys His Ala Phe Glu
 GAT CTA AAA CAA CCG GGT TTA GAC CAC AAT CAA GIG AGC TTC CAA GCT AGC GAT CCT GIG GIG AAT GAT GCC TTA GCA GGT TCA AAA CAT GCT TTT GAA
 CTA GAT TTT GGT CCG CCA AAT CTG GIG TAA GTT CAC TCG AAG GTT CGA TCG TTA GGA CAC CAC TTA CTA CCG AAT CGT CCA AGT TTT GTA OGA AAA CTT

3400
 Gln Lys Tyr Glu Met Cys Arg Leu Val Lys Lys Tyr Asp Tyr Pro Met Val Leu Asn Phe Val Ile His Arg His Asn Ile Asp Gln Ile Glu Gln Ile
 GAA AAA TAT GAA ATG TGC CGA TTG GTT AAG AAA TAC GAT TAT CCA ATG GIG CTT AAG TTT GTT ATT CAT CGA CAT AAC ATT GAC CAG ATT GAA CAA ATT
 GTT TTT ATA CTT TAC ACG GCT AAC CAA TTC TTT ATG CTA ATA GGT TAC CAC GAA TTG AAA CAA TAA GTA GCT GTA TTG TAA CTG GTC TAA CTT GTT TAA

3500
 Ile Glu Leu Cys Leu Glu Leu Asn Ala Asp Thr Val Glu Leu Ala Ile Cys Gln Phe Tyr Gly Trp Ala Phe Leu Asn Arg Gln Gly Leu Val Pro Thr
 ATT GAA CTT TGC CTT GAC CAC CCA GCA GAC ACA GTC GAG TTG GCA ATT TGT CAG TTT TAT GGC TGG GCT TTT TTA AAC CCG CAA GGT TTA TTA CCG ACA
 TAA CTT GAA ACG GAA CTC GAG TTG CGT CTG TGT CAG CTC AAC CGT TAA ACA GTC AAA ATA CCG ACC CGA AAA AAT TTG GCG GTT CCA AAT AAT GGG TGT

Transposon insertion Tn5-17 inhibited the expression of both gene I and gene II (Table 1). We have suggested (6) that these two genes are located in the same operon and that, due to a polar effect on transcription, Tn5 affects expression of both genes. Indeed in the DNA sequence two ORFs were indicated that were closely linked. The first ORF (gene I)

started at ATG (position 1930), ended at TGA (position 2688), and coded for a protein of M_r 29,700. Insertion Tn5-17 was located within this ORF. The stop codon (TGA) of gene I overlapped with the start ATG (position 2685) of a second ORF (gene II) that stopped at TGA (position 2969), coding for a protein of M_r 10,800. This overlap of start and stop

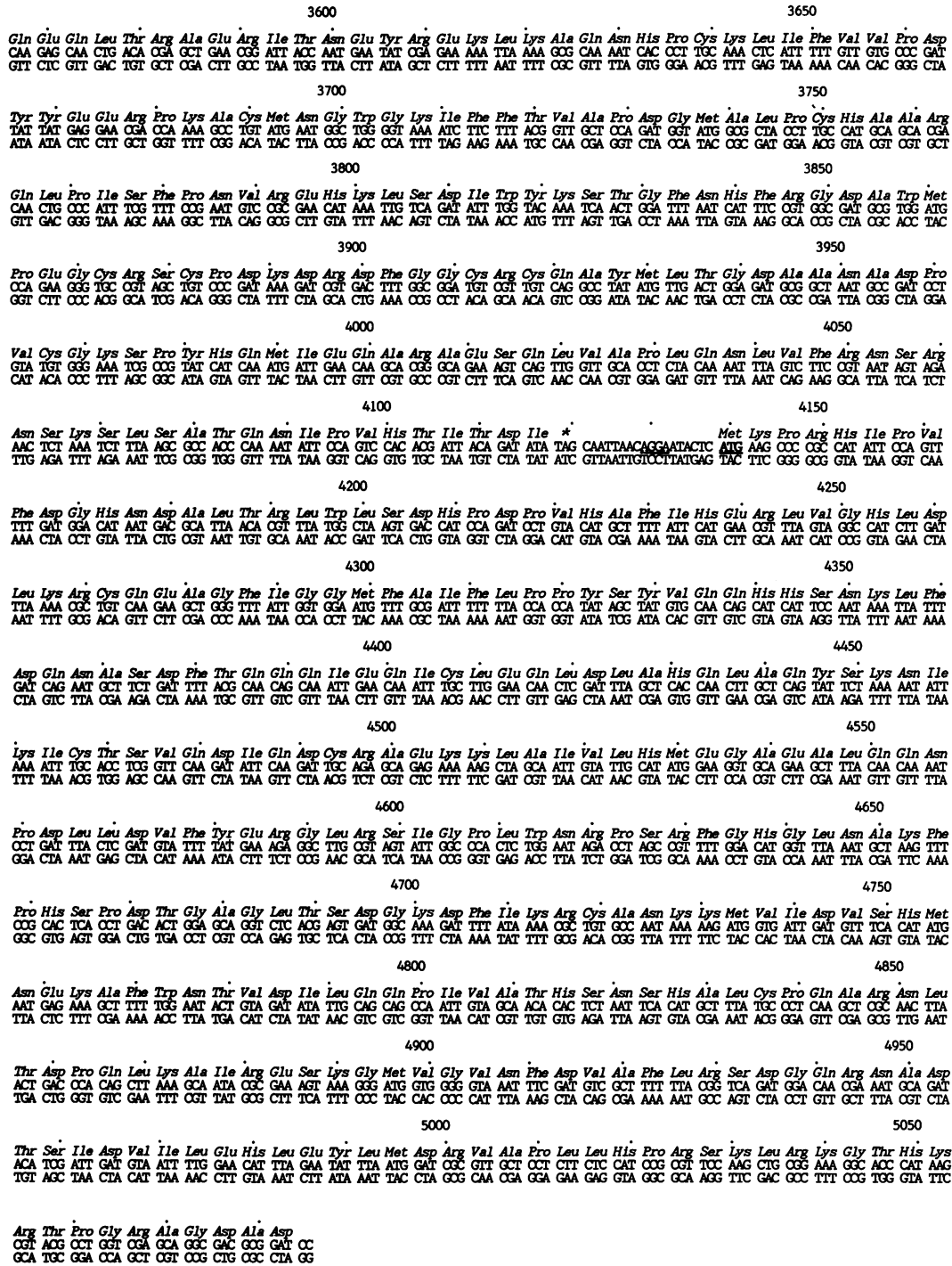


FIG. 2. Sequence of the 5,085-nucleotide *EcoRI-BamHI* insert of pSS2. The predicted amino acid sequence of the different reading frames is indicated as follows (positions): gene L (613 through 1), gene V (1010 through 1921), gene I (1930 through 2688), gene II (2685 through 2969), gene III (2966 through 4118 or 3125 through 4118), and gene R (4139 through 5085). The putative 24-amino-acid polypeptide coded for by gene IV is indicated by one-letter symbols (positions 861 through 935). The ATG translation start sites and the putative ribosome-binding sites are underlined. The insertion sites of the different Tn5 insertions are indicated by a box, representing the 9 bp that are duplicated upon Tn5 insertion.

codons suggests that the polar effect that Tn5-17 exerted on the expression of gene II might also be due to a coupled translation of both genes.
 The stop codon of gene II again overlapped with a putative start ATG (position 2966), followed by an ORF ending with

TAG (position 4118) which could code for a protein of M_r 43,600. This ORF might correspond to gene III, since Tn5-2 and Tn5-22, which abolished gene III expression, were located within this reading frame. However Tn5-9, which was mapped 20 bp downstream of this putative start of gene

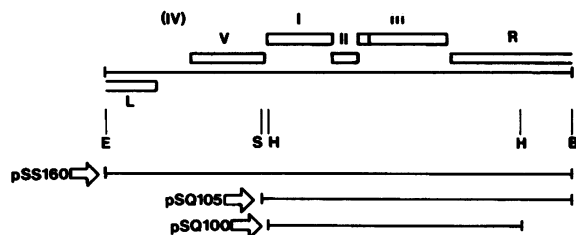


FIG. 3. Schematic representation of the different ORFs (indicated by bars) present in the 5,085-bp *EcoRI*-*Bam*HI fragment. The gene IV region is indicated with brackets. Relevant restriction sites: E, *EcoRI*; S, *Sph*I; H, *Hind*III; B, *Bam*HI. The inserts of the plasmids used for analysis of the proteins encoded by the PQQ genes are shown. pSS160 contains the *EcoRI*-*Bam*HI fragment inserted in pUC12, pSQ100 contains the *Hind*III fragment inserted in pUC12, and pSQ105 contains the *Sph*I-*Bam*HI fragment inserted in pUC19. The arrows represent the *P lac* promoter from which the PQQ genes are expressed.

III (Fig. 2), did not affect PQQ expression. An alternative ATG start could be indicated at position 3125 from which a protein of M_r 37,400 could be synthesized. If this second ATG were used for gene III translation, there would be an intracistronic region of 150 bp between genes II and III. To test whether a promoter for gene III expression was located within this region we isolated two deletion derivatives of pSS2 containing Tn5-9. First, all the sequences to the left of Tn5-9 were deleted (Fig. 4a). The resulting plasmid (Tn5-9/*Bg*III) still complemented gene III mutants, indicating that no sequences to the left of Tn5-9 are essential for gene III expression. Then 14 bp to the right of Tn5-9 were deleted (Fig. 4a). Complementation studies of the resulting plasmid (Tn5-9/*Hpa*I) showed that as a result of this deletion gene III was no longer expressed. A possible explanation is that the promoter (or part of the promoter) for gene III was located within these 14 bp of the PQQ sequence. However, since in Tn5-9/*Hpa*I also most of the Tn5 insertion had been deleted (Fig. 4a), it is more likely that a promoter was present in the inverted repeat of Tn5-9 which directed expression of gene

TABLE 1. Complementation of different PQQ⁻ mutants by Tn5 insertion plasmids (6)

Tn5 insertion	Complementation of PQQ ⁻ mutant			
	I	II	III	IV
Tn5-2	+	+	-	+
Tn5-7	+	+	+	+
Tn5-9	+	+	+	+
Tn5-14	+	+	+	+
Tn5-16	+	+	+	+
Tn5-17	-	-	+	+
Tn5-18	+	+	+	-
Tn5-20	+	+	+	+
Tn5-22	+	+	-	+

III and which was deleted in Tn5-9/*Hpa*I. Evidence for a promoter in the inverted repeats of Tn5 from which transcription can proceed into sequences adjacent to the Tn5 insertion has already been obtained in *E. coli* (11). This promoter was mapped between the *Ball* and *Hpa*I sites of the inverted repeat (Fig. 3). Recently, by subcloning the *gdh* gene of *A. calcoaceticus* we obtained evidence that this gene can be expressed in this organism from a promoter located within Tn5 (A. M. Cleton-Jansen et al., manuscript in preparation).

If indeed the Tn5 promoter is also active in *A. calcoaceticus*, the presence of a Tn5 insertion could be nonpolar for transcription, since all genes downstream from the Tn5 insertion could be expressed under the control of this promoter. In this case it is possible that genes I, II, and III are located in one operon. The presence of Tn5-17 in gene I might block expression of gene II by a polar effect on translation, whereas gene III could still be expressed under the control of the Tn5 promoter.

If the second ATG at position 3125 were the correct start site for gene III translation, Tn5-2 would be located between the ribosomal binding site and the translation start site (Fig. 2). The deficiency in gene III expression in this case would then probably be due to an inhibition of translation initiation.

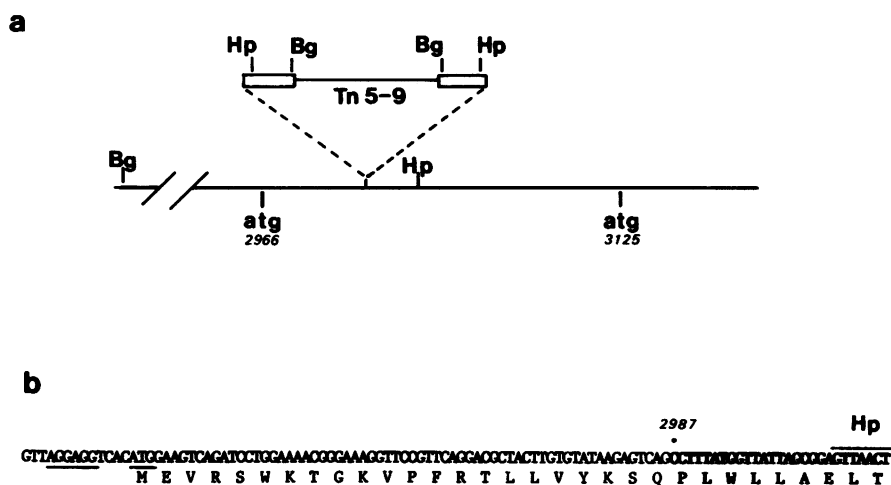


FIG. 4. (a) Schematic representation of the translation start region of gene III. The possible translation start sites for gene III are ATG (position 2966) and ATG (position 3125). The position of Tn5-9 is shown. Tn5-9/*Bg*III was constructed by deletion of the *Bg*III fragment (*Bg*III cuts to the left of the 5,085-bp insert and near the inside ends of the inverted repeats of Tn5). Tn5-9/*Hpa*I was constructed by deletion of the *Hpa*I fragment (*Hpa*I cuts 14 bp to the right of Tn5-9 and 190 bp from the outside ends of the inverted repeats of Tn5). Relevant restriction sites: *Bg*III; *Hpa*I, *Hpa*I. (b) Nucleotide sequence of the junction between Tn5-9 and the gene III region of pSS2. Base 2987 is the first nucleotide of the PQQ sequence. Putative ribosomal binding sites and the translation start site within the Tn5 sequence are underlined. The amino acids of the proposed fusion protein are shown.

However, arguments could be given for the first ATG at position 2966 being the correct start codon. Examination of the nucleotide sequence of the Tn5 region that is transcribed by the putative Tn5 promoter (Fig. 4b) revealed the presence of an ATG codon that was preceded by a good Shine and Dalgarno sequence (5'-AGGAGGT-3') and followed by an ORF that continued until the end of the transposon. As far as we know, no evidence in the literature has been presented that this sequence can serve as a translation start site, but if it does so in *A. calcoaceticus* the presence of Tn5-9 might generate a fusion protein (Fig. 4b). This fusion protein should contain 22 amino acids from Tn5 and the amino acids of the gene III product starting with Pro (position 2987). If the ATG at position 2966 were the correct start site for gene III translation, the fusion protein would lack only the seven N-terminal amino acids of the gene III product and would therefore be very likely to retain activity. In conclusion, our results still leave two possible translation start sites for gene III, one generating a protein of M_r 37,400 and one generating a protein of M_r 43,600.

For the mapping of the gene corresponding to the fourth complementation group in PQQ synthesis (gene IV), the position of Tn5-18, which blocks the expression of this gene (Table 1), was determined. Surprisingly, only two very small ORFs were indicated in this region. One ORF started with ATG (position 861) and ended with TAA (position 933), which would code for a protein of 24 amino acids (Fig. 2). The second ORF was located in the opposing strand starting with ATG (position 898) ending with TGA (position 832) and coded for a protein of 22 amino acids. Even if we consider the unlikely possibility that start sites other than ATG, GTG, or TTG are used in *A. calcoaceticus*, the largest possible ORF is from ACC (position 846) to TAG (position 965), which would code for a protein of 39 amino acids. Tn5-18 could also block expression of gene IV if it is not located within the ORF but in the leader of the transcript of gene IV. If we consider the ORFs of more considerable length near Tn5-18, however, this possibility also seems very unlikely. The first ORF to the right of Tn5-18 was the ORF of gene V, which was not involved in PQQ expression (see above). The only considerable ORF located to the left of Tn5-18 started with ATG (position 613) and proceeded until the *EcoRI* site. Within this reading frame (indicated as gene L) Tn5-14 was mapped (Fig. 2). Since Tn5-14 did not affect the complementation of PQQ⁻ mutants (Table 1), gene L also is not involved in PQQ synthesis. Thus, the effect of Tn5-18 on PQQ synthesis cannot be due to an influence on expression of these larger ORFs. Since Tn5-18 did not affect the expression of genes I, II, and III (Table 1), the conclusion seems justified that the DNA region in which Tn5-18 is located has to code for a product that plays a role in PQQ synthesis. A closer examination of the gene IV DNA region revealed the presence of a G+C-rich hairpin structure (positions 956 through 984) followed by an A+T-rich stretch of DNA. This structure is very likely to be a terminator for transcription initiated within or to the left of the gene IV region. Two possibilities for the gene IV product can now be considered. First the RNA that terminates at the hairpin structure might code for the 24-amino-acid polypeptide starting at ATG (position 861). Second, the RNA itself might play a role in PQQ synthesis. In both cases an unusual pathway for PQQ biosynthesis can be expected.

The last ORF present in the sequence (indicated as gene R) started at position 4139 (ATG) and proceeded beyond the *BamHI* site. Since no other classes of PQQ mutants like the

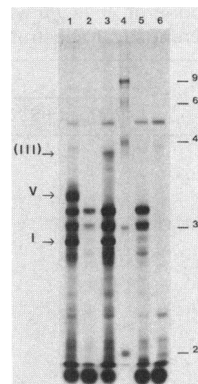


FIG. 5. Autoradiogram of [³⁵S]methionine-labeled proteins produced by in vitro transcription-translation of different DNA templates. Lanes: 1, pSS160; 2, pSQ100; 3, pSQ105; 4, markers; 5, pUC18; 6, no DNA. The positions of the marker bands are indicated. The arrows represent the protein bands corresponding to translation products of genes I, III, and V.

four classes described above have been found (6), this reading frame is not likely to play a role in PQQ synthesis.

Analysis of the proteins encoded by the PQQ genes. To test whether proteins corresponding to the different ORFs could be detected, we used an in vitro coupled transcription-translation system derived from *E. coli*. When pSS2 was used as template in this system, no specific proteins could be detected. This could be due to an improper recognition of *A. calcoaceticus* promoters by *E. coli* RNA polymerase. Therefore we placed the PQQ genes under control of the *E. coli lac* promoter by cloning the 5,085-bp *EcoRI-BamHI* fragment of pSS2 in pUC12, resulting in pSS160. The autoradiogram of the labeled proteins produced by this plasmid and run on a 12% polyacrylamide gel is shown in Fig. 5. Two specific protein bands that were not produced when the vector pUC12 was used as template could be detected. The first had a molecular weight of about 34,000, which might very well correspond to the product of gene V (M_r 33,600). The second protein band migrated at the position of approximately M_r 29,000 and is therefore likely to be the product of gene I (M_r 29,700). To test whether these protein bands indeed correspond to the translation products of genes I and V, we constructed two other plasmids. pSQ105 contained the 3,175-bp *SphI-BamHI* fragment in pUC19 (Fig. 4) and was expected to express genes I, II, and III under control of the *lac* promoter. pSQ100 contained the 2,600-bp *HindIII* fragment in pUC12 (Fig. 4). Since *HindIII* cuts 15 bp downstream the postulated translation start of gene I, this plasmid was expected to express only genes II and III. Indeed, the M_r 34,000 protein band (gene V) was missing when both plasmids were translated (Fig. 5). The M_r 29,000 band was still present in pSQ105 but absent in pSQ100. So the M_r 29,000 protein was very likely to be the translation product of gene I, and the postulated translation start of this gene indeed was located between the *SphI* and *HindIII* sites.

Because the translation product of gene II was very small (M_r 10,800), no band corresponding to this protein is visible in Fig. 5. The protein of gene III, however, should be large enough to be detectable. We have shown above that two possible starts for gene III translation can be considered, one resulting in a protein of M_r 37,400 and the other resulting in a protein of M_r 43,600. With pSQ105 a weak protein band of about M_r 44,000 was visible. If this band corresponded to the gene III product, this would mean that translation of gene III

TABLE 2. Acid production of *A. lwoffii* with or without pSS2 on different sugar substrates

Substrate	Acid production	
	Without pSS2	With pSS2
Glucose	—	+
Glucose + PQQ	+	+
Galactose	—	+
Galactose + PQQ	+	+
Arabinose	—	+
Arabinose + PQQ	+	+
Xylose	—	+
Xylose + PQQ	+	+
Lactose	—	—
Lactose + PQQ	—	—

starts with the first ATG (position 2966). However, the gene III translation product was not observed when pSS160 or pSQ100 were used as a template. Possibly the expression of gene III, which is already very low in pSQ105, is for an unknown reason even more reduced in the other two plasmids.

Expression of the PQQ genes in *A. lwoffii* and *E. coli* K-12. *A. lwoffii* and *E. coli* K-12 produce a PQQ-dependent GDH (8, 21). However, these bacterial species do not synthesize PQQ. Therefore the holo-GDH can be formed only when PQQ is supplied from the culture medium. To test whether our four cloned PQQ genes contain sufficient information for the biosynthesis of PQQ, we introduced these genes into *A. lwoffii* and *E. coli* and tested whether these bacteria produced an active GDH enzyme. First pSS2 was introduced in *A. lwoffii* by conjugation, and then the transconjugants were tested for acid production on different sugar substrates as described in Materials and Methods. The introduction of pSS2 in *A. lwoffii* resulted in acid production when GDH-specific sugars were used as substrates (Table 2). This indicates that in the presence of pSS2 a sufficient amount of PQQ is synthesized for the reconstitution of the holo-GDH enzyme. To test whether *A. lwoffii* is normally lacking all the PQQ enzymes and not just carrying a mutation in one of the PQQ genes, we also investigated whether the pSS2 plasmids with the different Tn5 insertions produced PQQ. *A. lwoffii* with pSS2 carrying the insertion Tn5-2, Tn5-17, or Tn5-18 did not produce acid on GDH-specific sugars, whereas upon introduction of pSS2 with Tn5-7, Tn5-9, Tn5-16, or Tn5-20 acid production due to GDH activity could be shown. Apparently all four genes are required for PQQ synthesis in *A. lwoffii*.

Next we tested whether also *E. coli* K-12 could synthesize PQQ upon introduction of the four genes. For this experiment an *E. coli* K-12 strain carrying a *ptsI* mutation (PPA41) was used. Due to the *ptsI* mutation, this strain no longer grows on glucose as the sole carbon source. In the presence of PQQ, however, the GDH enzyme can be used, and growth on glucose is restored (8). Previously we have shown (6) that PPA41 carrying pSS2 still does not grow on minimal medium plates with glucose. Since this might be due to an improper expression of the PQQ genes by the *E. coli* RNA polymerase, we also transformed pSS160 (carrying the PQQ genes under control of the *lac* promoter) to PPA41. The resulting transformants did form colonies on minimal medium plates with glucose after incubation at 37°C for 2 days. No growth was observed on control plates containing mannitol as the carbon source, indicating that the transformants were still mutated in *ptsI* and that growth on glucose had to

be the consequence of PQQ synthesis. The PQQ production, however, was apparently low, since in the presence of additional PQQ in the culture medium colony formation on glucose plates was already observed after incubation for 1 day. We showed above that at least the *in vitro* synthesis of the gene III protein with *E. coli* extracts was very low, which could be the reason for the low level of PQQ production *in vivo*. To test whether in *E. coli* the expression of genes I, II, and III is sufficient for PQQ production, we also introduced pSQ105 (containing these three genes under control of the *lac* promoter) into PPA41. The resulting transformants did not form colonies on minimal medium plates with glucose even after prolonged incubation, meaning that also in *E. coli* the product of gene IV (which might be an RNA or a small polypeptide) is essential in the biosynthetic pathway of PQQ.

DISCUSSION

Biosynthesis of the coenzyme PQQ is thought to be a multistep process involving at least five or six different enzymes. In this paper we have shown that expression of four different PQQ genes from *A. calcoaceticus* in *E. coli* or *A. lwoffii* is sufficient for the synthesis of the coenzyme in these organisms. Sequence analysis of the PQQ genes showed that only three of them (genes I, II, and III) code for proteins of a size (M_r 29,700, 10,800, and 43,600, respectively) that can be expected for proteins with an enzymatic function. The fourth gene (gene IV), however, does not seem to code for an enzyme, since the most probable reading frame comprises only 24 amino acids. An explanation for the presence of so few PQQ-specific enzymes might be that a precursor resembling the mature coenzyme structure is already present, not only in *A. calcoaceticus* but also in *A. lwoffii* and *E. coli*.

The intriguing question is what the role of gene IV in PQQ synthesis might be. Complementation studies have shown that the gene IV region is coding for a *trans*-acting product. Since among 40 independently isolated PQQ⁻ mutants of *A. calcoaceticus* no gene IV point mutations were found (6; unpublished data), this product was already predicted to be relatively small. The 3' end of the gene IV region contains a G+C-rich hairpin followed by an A+T-rich stretch of DNA, which is characteristic of a transcription terminator. This suggests that a small transcript complementary to only gene IV DNA is formed. Two possible functions for this gene IV RNA can be considered.

First, the gene IV RNA itself might in some way play a role in PQQ synthesis. We consider it unlikely that this RNA would play a direct structural role, but a function as a positive regulator cannot be excluded. RNA has been shown to act in *trans* as a negative regulator in the replication of plasmid Cole1 (18). However until now it has not been reported that RNA can also act as a positive regulator.

As a second possibility the gene IV RNA might be translated in a 24-amino-acid polypeptide, which in turn might have a function in PQQ synthesis. Recently it has been shown that the amino acids tyrosine and glutamate are used as precursors for PQQ biosynthesis in *Hyphomicrobium* sp. strain X (20a) and *Methylobacterium* sp. strain AM1 (C. J. Unkefer, personal communication). Since both amino acids are present in the 24-amino-acid polypeptide, it is possible that this polypeptide is used as a precursor substrate. Through the tertiary structure of the polypeptide both amino acids could be brought together, the enzymes could make the bonds, and finally the complete (or almost complete)

PQQ molecule could be cut out of the polypeptide. Still, when such a polypeptide is used as a precursor it remains unlikely that only three enzymes are sufficient for the complete synthesis of PQQ, since the chemical bonds that have to be made will require multiple enzymatic steps. However, enzymes could be used that also play a role in other biochemical processes.

Finally, the 24-amino-acid polypeptide might also play a role in the transport of PQQ across the cytoplasmic membrane. The active site of the PQQ-dependent GDH enzyme is located at the periplasmic side of the inner membrane (4). Therefore eventually also the PQQ has to be present in the periplasmic space. This could be achieved in two ways. First, the PQQ enzymes could be transported through the inner membrane, and the PQQ could be synthesized in the periplasm. The predicted amino acid sequences of the different PQQ enzymes, however, do not reveal the presence of N-terminal signal peptides (a stretch of hydrophobic amino acids flanked by hydrophilic residues). Alternatively, the PQQ could be synthesized in the cytoplasm and subsequently translocated to the periplasm. The 24-amino-acid polypeptide might be involved in this process, although it also is not highly hydrophobic.

Besides the putative transcription terminator of gene IV, no transcription terminator signals seem to be present at the end of the other PQQ genes. Moreover, the sequence data have shown that the translation start of gene II overlaps with the end of gene I, the start of gene III might overlap with the end of gene II, and the ORFs of genes V and I are separated by only 8 bp. Taken together, these results indicate that it is very likely that genes V, I, II, and III are transcribed from one promoter (although it is still possible that gene III has its own promoter). Gene V, however, which would be the first gene in this operon, is not involved in the synthesis of PQQ. Gene V also does not play a role in the transport of electrons from PQQ to the electron transport chain, since deletion mutants lacking the complete PQQ region (including gene V) show normal GDH activity upon addition of PQQ (13). A reason for the concerted expression of gene V and the PQQ genes might be that gene V codes for another PQQ-dependent enzyme.

ACKNOWLEDGMENTS

We thank Mario van Kleef for fruitful discussions.

Part of this work was supported by the Netherlands Technology Foundation (S.T.W.).

LITERATURE CITED

- Ameyama, M., K. Matsushita, Y. Ohno, E. Shinegawa, and E. Adachi. 1981. Existence of a novel prosthetic group, PQQ, in membrane-bound electron transport chain-linked, primary dehydrogenases of oxidative bacteria. *FEBS Lett.* **130**:179-183.
- Davidson, V. L., and J. W. Neher. 1987. Evidence for two subclasses of methylamine dehydrogenases with distinct large subunits and conserved PQQ-bearing small subunits. *FEMS Microbiol. Lett.* **44**:121-124.
- Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387-395.
- Duine, J. A., J. Frank, and J. A. Jongejan. 1986. PQQ and quinoprotein enzymes in microbial oxidations. *FEMS Microbiol. Rev.* **32**:165-178.
- Duine, J. A., J. Frank, and J. K. van Zeeland. 1979. Glucose dehydrogenase from *Acinetobacter calcoaceticus*: a quinoprotein. *FEBS Lett.* **108**:443-446.
- Goosen, N., D. A. M. Vermaas, and P. van de Putte. 1987. Cloning of the genes involved in synthesis of coenzyme pyrroloquinoline-quinone from *Acinetobacter calcoaceticus*. *J. Bacteriol.* **169**:303-307.
- Groen, B. W., M. A. G. van Kleef, and J. A. Duine. 1986. Quinohaemoprotein alcohol dehydrogenase apoenzyme from *Pseudomonas testosteroni*. *Biochem. J.* **234**:611-615.
- Hommel, R. W. J., P. W. Postma, O. M. Neysse, D. W. Tempest, P. Dokter, and J. A. Duine. 1984. Evidence of a quinoprotein glucose dehydrogenase apoenzyme in several strains of *Escherichia coli*. *FEMS Microbiol. Lett.* **24**:329-333.
- Hong, G. F. 1982. A systematic DNA sequencing strategy. *J. Mol. Biol.* **158**:539-549.
- Lobenstein-Verbeek, C. L., J. A. Jongejan, J. Frank and J. A. Duine. 1984. Bovine serum amine oxidase: a mammalian enzyme having covalently-bound PQQ as prosthetic group. *FEBS Lett.* **170**:305-309.
- Lupski, J. R., S. J. Projan, L. S. Ozaki, and G. N. Godson. 1986. A temperature-dependent pBR322 copy number mutant resulting from a Tn5 position effect. *Proc. Natl. Acad. Sci. USA* **83**:7381-7385.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Messing, J., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. *Nucleic Acids Res.* **9**:309-321.
- Miller, J. H. 1972. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Moog, R. S., M. A. McGuirl, C. E. Cote, and D. M. Dooley. 1986. Evidence for methoxatin (pyrroloquinolinequinone) as the cofactor in bovine plasma amine oxidase from resonance Raman spectroscopy. *Proc. Natl. Acad. Sci. USA* **83**:8435-8439.
- Sanger, F., A. R. Coulson, B. G. Barrell, and A. J. H. Smith. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161-178.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
- Tomizawa, J., and T. Itoh. 1981. Plasmid ColE1 incompatibility determined by interaction of RNAI with primer transcript. *Proc. Natl. Acad. Sci. USA* **78**:6096-6100.
- Van der Meer, R. A., J. A. Jongejan, and J. A. Duine. 1987. Phenylhydrazine as probe for cofactor identification in amine oxidoreductases. *FEBS Lett.* **221**:299-304.
- Van der Meer, R. A., J. A. Jongejan, J. Frank, and J. A. Duine. 1986. Hydrazone formation of 2,4-dinitrophenylhydrazine with pyrroloquinoline quinone in porcine kidney diamine oxidase. *FEBS Lett.* **206**:111-114.
- van Kleef, M. A. G., and J. A. Duine. 1988. L-Tyrosine is the precursor of PQQ biosynthesis in *Hyphomicrobium X*. *FEBS Lett.* **237**:91-97.
- van Schie, B. J., K. J. Hellingwerf, J. P. van Dijken, M. G. L. Elferink, J. M. van Dijk, J. G. Kuenen, and W. N. Konings. 1985. Energy transduction by electron transfer via a pyrroloquinoline-quinone-dependent glucose dehydrogenase in *Escherichia coli*, *Pseudomonas aeruginosa* and *Acinetobacter calcoaceticus*. *J. Bacteriol.* **163**:493-499.
- van Schie, B. J., J. P. van Dijken, and J. G. Kuenen. 1984. Non-coordinated synthesis of glucose dehydrogenase and its prosthetic group in *Acinetobacter* and *Pseudomonas* species. *FEMS Microbiol. Lett.* **24**:133-138.