

# Structure and Organization of *Escherichia coli* Genes Involved in Biosynthesis of the Deazaguanine Derivative Queuine, a Nutrient Factor for Eukaryotes

KLAUS REUTER, ROBERT SLANY, FRANK ULLRICH, AND HELGA KERSTEN\*

*Institut für Biochemie, Universität Erlangen-Nürnberg, Fahrstrasse 17, D-8520 Erlangen, Germany*

Received 1 August 1990/Accepted 23 January 1991

The plasmid pPR20 contains the gene *tgt*, which encodes tRNA guanine transglycosylase (Tgt), on a 33-kbp DNA insert from a region around 9 min on the *Escherichia coli* linkage map. The plasmid was subcloned to determine the sequence and organization of the *tgt* gene. Tgt is a unique enzyme that exchanges the guanine residue with 7-aminomethyl-7-deazaguanine in tRNAs with GU(N) anticodons. After this exchange, a cyclopentendiol moiety is attached to the 7-aminomethyl group of 7-deazaguanine, resulting in the hypermodified nucleoside queuosine (Q). Here we give the complete sequence of a 3,545-bp *StuI-BamHI* DNA fragment where we found the *tgt* gene and three previously unknown genes encoding proteins with calculated molecular masses of 42.5 (Tgt), 14, 39, and 12 kDa. The gene products were characterized on sodium dodecyl sulfate gels after synthesis in a combined transcription-translation system. The mRNA start sites of the open reading frames (ORFs) were determined by primer extension analysis. Plasmids containing the ORF encoding the 39-kDa protein (ORF 39) complemented a mutation in Q biosynthesis after the Tgt step. This gene was designated *queA*. The genes are arranged in the following order: ORF 14 (transcribed in the counterclockwise direction), *queA*, *tgt*, and ORF 12 (all transcribed in the clockwise direction). The organization of the promoter sequences and the termination sites suggests that *queA*, *tgt*, and ORF 12 are localized on a putative operon together with the genes *secD* and *secE*.

The hypermodified nucleoside queuosine [Q; 7-(((4,5-*cis*-dihydroxy-2-cyclopentene-1-yl)-amino)-methyl)-7-deazaguanosine] usually occurs in the first position of the anticodon of eubacterial and eukaryotic tRNAs, specifying the amino acids aspartic acid, asparagine, tyrosine, and histidine. In *Escherichia coli*, the biosynthesis of Q commences with GTP, which is converted to 7-aminomethyl-7-deazaguanine by unknown mechanisms and enzymes that are not yet characterized. This precursor of Q is then inserted into the first position of the anticodons of the tRNAs by exchange with guanine. The reaction is catalyzed by the tRNA guanine transglycosylase Tgt (EC 2.4.2.29), the sole enzyme of the biosynthetic pathway so far purified (19). The biosynthesis of Q is completed at the level of tRNA. A cyclopentendiol moiety, probably derived from a sugar, is attached, and epoxy-queuosine [oQ; 7-(*N*-(2,3-epoxy-4,5-*cis*-dihydroxycyclopent-1-yl)-aminomethyl)-7-deazaguanosine] is formed (12). Finally, oQ is reduced to Q by a cobamide-dependent enzyme system.

In *E. coli*, the tRNAs of the Q family are fully modified with Q when the cells are grown under aerobic or strict anaerobic conditions, provided that the growth medium is supplemented with appropriate amounts of iron ions and vitamin B<sub>12</sub> (6, 9; reference 10 and references therein). *E. coli* mutants with a defect in *tgt* have been described (18), but unfortunately these mutants harbor an additional mutation in the anaerobic regulator gene *fnr* and proved to be unsuitable to elucidate the function of Q (5). Transduction of the mutated *tgt* locus into another *E. coli* recipient has been performed to reexamine the consequences of the *tgt* mutation. F' plasmids containing a fused *lacI-lacZ* gene with the nonsense codon UAG at different positions in the *lacI* part

were transferred to *E. coli* strains with a mutated or nonmutated *tgt* locus. A twofold increase in the frequency of incorrect readthrough of the UAG codon dependent on the codon context has been found in the *tgt* mutant. Increased readthrough appears to be caused by tRNA<sup>Tyr</sup> with G instead of Q when the adjacent tRNA on the ribosome also belongs to the Q family (5). Whether this slight increase in error frequency is of any biological significance is presently not known.

Only eubacteria synthesize Q de novo, while lower and higher eukaryotes obtain the corresponding base, queuine, from nutrients or from the intestinal flora. With the exception of yeast cells and plant leaf cells, all eukaryotic cells contain Q as the free base or incorporated into the respective cytosolic and mitochondrial tRNAs. The extent of Q modification of the respective tRNAs and the occurrence of free queuine changes according to the proliferative and metabolic state of eukaryotic cells and tissues (summarized in references 9, 10, and 17). Previously, we have reported that queuine stimulates the proliferation of HeLa cells at high cell density or under hypoxic conditions (13). Queuine reduces slightly, but significantly, the overall synthesis of proteins in intact mammalian cells and in corresponding cell-free systems and modulates the phosphorylation of specific proteins (13, 14).

The importance of queuine as a growth modulator for higher eukaryotes, especially under hypoxic conditions, and the limited knowledge of the biosynthesis and function of Q in eubacteria (5, 9, 17) led us to study the genes and enzymes involved in Q biosynthesis in *E. coli*. Cloning and characterization of Q biosynthesis genes, described in this article, permit (i) overproduction of the corresponding enzymes and study of their structure and interaction with tRNA and (ii) construction of defined deletion mutants and characterization of changes in their phenotypes.

\* Corresponding author.

TABLE 1. *E. coli* strains and plasmids

Strain or plasmid	Genotype or relevant characteristics	Source or reference(s) <sup>a</sup>
<b>Strains</b>		
K-12	Wild type	K-12 from DSM (strain 498)
DH5 $\alpha$	<i>endA1 hsdR17 supE44 thi-1 recA1 gyrA96 relA1 <math>\Delta</math>(argF-lac)U169 <math>\Phi</math>80dlacZ <math>\Delta</math>M15 <math>\lambda^-</math></i>	B. J. Bachmann (CGSC)
SJ1505	<i>araD139 <math>\Delta</math>(argF-lac)U169 thi deoC relA1 rpsL150 <i>tgt</i></i>	6
JE 10651	<i>queA thrA1 leuB6 trp-1 his-1 thyA argH1 thiA1 lacY1 malA1 mtl-2 xyl-7 fhuA1 supE44 rpsL150</i>	20
<b>Plasmids</b>		
pACYC184	Cm <sup>r</sup> Tc <sup>r</sup>	3
pUC19	Ap <sup>r</sup> , <i>lac<math>\alpha</math></i> peptide multiple cloning site	34
pPR20	pRSF2124 [ <i>phoBR</i> Tn10:: <i>tsx</i> (Tc <sup>r</sup> ) ( <i>Bam</i> HI)] (Tn10 is incompletely present)	28, 30
pHH1	pACYC184 ( <i>tgt</i> containing <i>Hind</i> III fragment from pPR20)	This work
pHBI	pACYC184 ( <i>tgt</i> containing <i>Hind</i> III- <i>Bam</i> HI fragment from pPR20)	This work
pUSB12	pUC19 ( <i>tgt</i> containing <i>Sal</i> I- <i>Bam</i> HI fragment from pPR20)	This work
pHS1	pUC19 ( <i>Hind</i> III- <i>Sma</i> I fragment upstream from and partly overlapping insert of pUSB12)	This work
pSH17	pUC19 ( <i>Hind</i> III- <i>Sal</i> I subclone of pHS1 insert)	This work

<sup>a</sup> DSM, Deutsche Sammlung von Mikroorganismen, Göttingen, Germany; CGSC, *E. coli* Genetic Stock Center, Yale University, New Haven, Conn.

The *tgt* gene is localized around 9 min on the *E. coli* map (1, 18) between the *phoBR* operon and *tsx*. The *phoBR* operon has been cloned (30), and J. P. Tommassen and coworkers kindly provided plasmid pPR20, which contains the *tgt* gene. Here we describe the structure and organization of a putative operon containing *tgt* and a new Q biosynthesis gene designated *queA*.

## MATERIALS AND METHODS

**Materials.** Restriction endonucleases and DNA-modifying enzymes were purchased from Boehringer, Mannheim, Germany, and from GIBCO Laboratories, Bethesda, Md. The [<sup>3</sup>H]guanine sulfate, [ $\alpha$ -<sup>32</sup>P]ATP, and [ $\gamma$ -<sup>32</sup>P]ATP were obtained from Amersham Corp., Little Chalfont, United Kingdom. Culture media were from Difco Laboratories, Detroit, Mich. Quiagen columns were purchased from Diagen, Düsseldorf, Germany. The Nested Deletion Kit was from Pharmacia, Uppsala, Sweden.

**Bacterial strains and plasmids.** The bacterial strains and plasmids used in these studies are listed in Table 1. The bacteria were transformed with the plasmids by the standard transformation protocol (15).

**Media.** Strains were subcultured in LB medium (16) or LB medium with 1.5% agar added. Antibiotics were added to the growth media in the following concentrations: chloramphenicol, 10  $\mu$ g/ml; ampicillin, 100  $\mu$ g/ml; tetracycline, 20  $\mu$ g/ml; and streptomycin, 25  $\mu$ g/ml.

**Cloning procedures.** DNA manipulations and restriction fragment isolations were performed according to the methods described in reference 15.

**Characterization of the *tgt* and *queA* mutations.** The respective tRNAs from *tgt* mutants contain a guanosine residue in the anticodon in place of queuosine (Q). The total amount of the Q-deficient specific tRNAs in bulk tRNA was determined by using a partially purified tRNA guanine transglycosylase from *E. coli* (5, 19). The enzyme not only replaces the guanine residue in the anticodon with the queuine precursor, 7-aminomethyl-7-deazaguanine but also can exchange the specific guanine residue in Q-deficient tRNAs by guanine. Once the synthesis of queuosine is completed, the enzyme cannot replace it with guanine (19). The crude enzyme preparation incorporated at a maximum of 90 to 120 pmol of [<sup>3</sup>H]guanine per A<sub>260</sub> unit of bulk tRNA from the *tgt* mu-

tants. The mutant JE 10651 is unable to complete the synthesis of queuine after insertion of the precursor 7-aminomethyl-7-deazaguanine. In our laboratory, this precursor can be partially exchanged in the tRNA transglycosylase assay. A maximum of about 50 to 60 pmol of [<sup>3</sup>H]guanine was incorporated per A<sub>260</sub> unit of bulk tRNA from mutant JE 10651.

**DNA sequence determination.** Sequence analysis was performed by the dideoxy-chain termination method (24). The respective DNA fragments were cloned in pUC19. The sequence was analyzed with double-stranded DNA as a template. The DNA was purified with Quiagen-pack 100 columns, according to the protocol of the producer. Both the universal M13 primer and the reverse M13 primer were used for the modified T7 polymerase (Sequenase) reaction (29). Occasional band compression effects were resolved by the use of dITP instead of dGTP in the Sequenase reaction. The nucleotide sequence was determined from both directions. Nested deletions were obtained by using a Nested Deletion Kit with exonuclease III and S1 nuclease.

**Computational analysis.** The sequence was analyzed by applying the Genetics Computer Group Sequence Analysis Software Package, version 6.1 (April 1990) (4) with regard to (i) open reading frames (ORFs), (ii) similarity to other known sequences, and (iii) the codon frequency of the new ORFs that were discovered.

**Determination of mRNA start sites by primer extension.** Oligonucleotides (22-mers) complementary to nucleotides 812 to 833, 1919 to 1940, and 3147 to 3168 were labeled with <sup>32</sup>P at the 5' end. The labeled primer (0.5 pmol) was hybridized with 20  $\mu$ g of total RNA from *E. coli* K-12 for 1 h at 42°C and then extended with avian myeloblastosis virus reverse transcriptase in the presence of unlabeled deoxynucleoside triphosphates for 1 h at 42°C. The sizes of the extended products were determined on a sequencing gel. Sequencing ladders of the DNA from plasmid pUSB12 or pHS1 which were established with the respective primers (33) were used as size markers.

**In vitro expression.** The in vitro expression of the plasmid-encoded genes was performed with a Procaryotic, DNA-Directed Translation Kit from Amersham. The gene products were labeled with [<sup>35</sup>S]methionine and separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis

(SDS-PAGE), running parallel to a molecular weight standard.

**Nucleotide sequence accession numbers.** The nucleotide sequence data in Fig. 2 are accessible in the EMBL Data Bank under the accession numbers M35286 and M37702.

## RESULTS

**Subcloning of the *tgt* gene.** Plasmid pPR20 (30) contains a 33-kbp insert of *E. coli* chromosomal DNA that spans the whole region where the *tgt* gene is located (18). This plasmid complemented the *tgt* mutation when transformed into the *tgt* mutant SJ1505. The tRNA from the transformed strain was fully modified with respect to Q. This was demonstrated by analysis of the tRNA from the transformant SJ1505/pPR20 by the transglycosylase assay and by high-performance liquid chromatography analysis of the tRNA nucleosides (details of tRNA analysis are described in Materials and Methods and in reference 6) (results are not shown). The insert in pPR20 was subcloned and resulted a *Sall*-*Bam*HI insert of 1,823 bp (pUSB12 in Fig. 1) that fully complemented the *tgt* mutation.

**Nucleotide sequence of *tgt* and its flanking regions.** The nucleotide sequence of the *Sall*-*Bam*HI fragment was determined for both strands from overlapping deletion subclones. The sequence data that we obtained led us to suggest that the *tgt* gene might be located together with other genes on one operon. We therefore determined the sequence for the whole *Hind*III-*Bam*HI insert of the plasmid pHB1, from which the *Sall*-*Bam*HI fragment was originally subcloned. The strategy for subcloning is shown in Fig. 1. We present the sequence of the pHB1 insert from the *Stu*I site to the *Bam*HI site in Fig. 2. Computational translation of the nucleotide sequence in the three possible reading frames and for both strands revealed four ORFs.

The first reading frame (ORF 14) was oriented opposite to the other three reading frames. It consisted of 114 codons from nucleotide 483 to 142. Therefore, it is depicted on the minus strand in the sequence. A putative ribosome binding site (27), GGA (TCC on the minus strand), was centered 4 bp upstream of the start codon (Fig. 2). According to the codon usage in *E. coli* (8), the putative gene was suggested to be expressed at a relatively low level. The molecular mass of the gene product was calculated to be 14 kDa.

The second reading frame (ORF 39) consisted of 356 codons from nucleotides 812 to 1879 and ended with a TAA stop codon. A potential ribosome binding site (GAG) started 8 bp upstream of the initiator codon ATG. According to the codon frequency, this gene is expected to be moderately expressed. The calculated molecular mass of the gene product was 39 kDa.

The third open reading frame, *tgt*, 58 bp downstream of ORF 39, was composed of 375 codons from nucleotides 1938 to 3062 and was terminated by a TAA stop codon. This was followed in frame by a sense codon and three further stop codons, TAA TAA TGA. The reading frame corresponded to the *tgt* gene, as identified by nested deletions of pUSB12 from both sides and subsequent complementation analysis using pUSB12 derivatives (Table 2). A putative ribosome binding site, GGAG, was positioned 5 bp upstream of the initiator codon ATG. Also, *tgt* appeared to be moderately expressed with respect to the codon usage in *E. coli*. The calculated molecular mass of the gene product was 42.5 kDa, which differs somewhat from the molecular mass of 46 kDa estimated for the purified enzyme by SDS-PAGE.

The fourth ORF (ORF 12) was found 25 bp downstream of

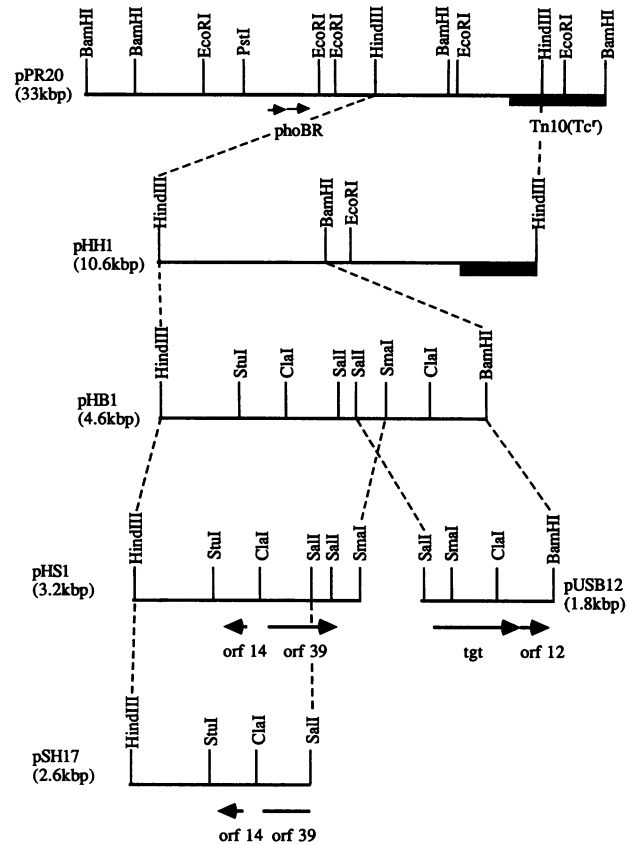


FIG. 1. Strategy of subcloning of pPR20, which contains the *tgt* region. Going from left to right, the map corresponds to the clockwise orientation of the *E. coli* linkage map. Only the inserts of the constructed plasmids are shown; sizes of the inserts are indicated in parentheses.

the last triplet of *tgt*, from nucleotides 3088 to 3417. This reading frame comprised 110 codons and also terminated with a TAA stop codon. The codon usage pointed to a more strongly expressed gene. A possible ribosome binding site (GAG) was present upstream at a distance of 8 bp. The molecular mass of the corresponding gene product was calculated to be 12 kDa.

No DNA structures resembling rho-independent terminators were found downstream of any of the open reading frames. All four reading frames were analyzed at the amino acid level for sequence similarities with the protein sequences in the GenBank and EMBL data bases by using a VAX computer and the Genetics Computer Group program TFasta (4). Significant homologies or identities were not observed.

The proximity and the same transcriptional orientation of ORF 39, *tgt*, and ORF 12 suggested that these genes are arranged in one operon. While this manuscript was being prepared, Gardel and coworkers published the nucleotide sequence of the *E. coli* genes *secD* and *secF*, which are located on the *E. coli* chromosome in the same region as *tgt* (7). The products of *secD* and *secF* are membrane proteins involved in the translocation of secreted proteins. Here we show that the sequence upstream of *secD* overlaps ORF 12 and partially overlaps the sequence of *tgt*. The sequence data presented here indicate that *secD* follows ORF 12 at a distance of 30 or 63 nucleotides. Gardel et al. (7) speculate

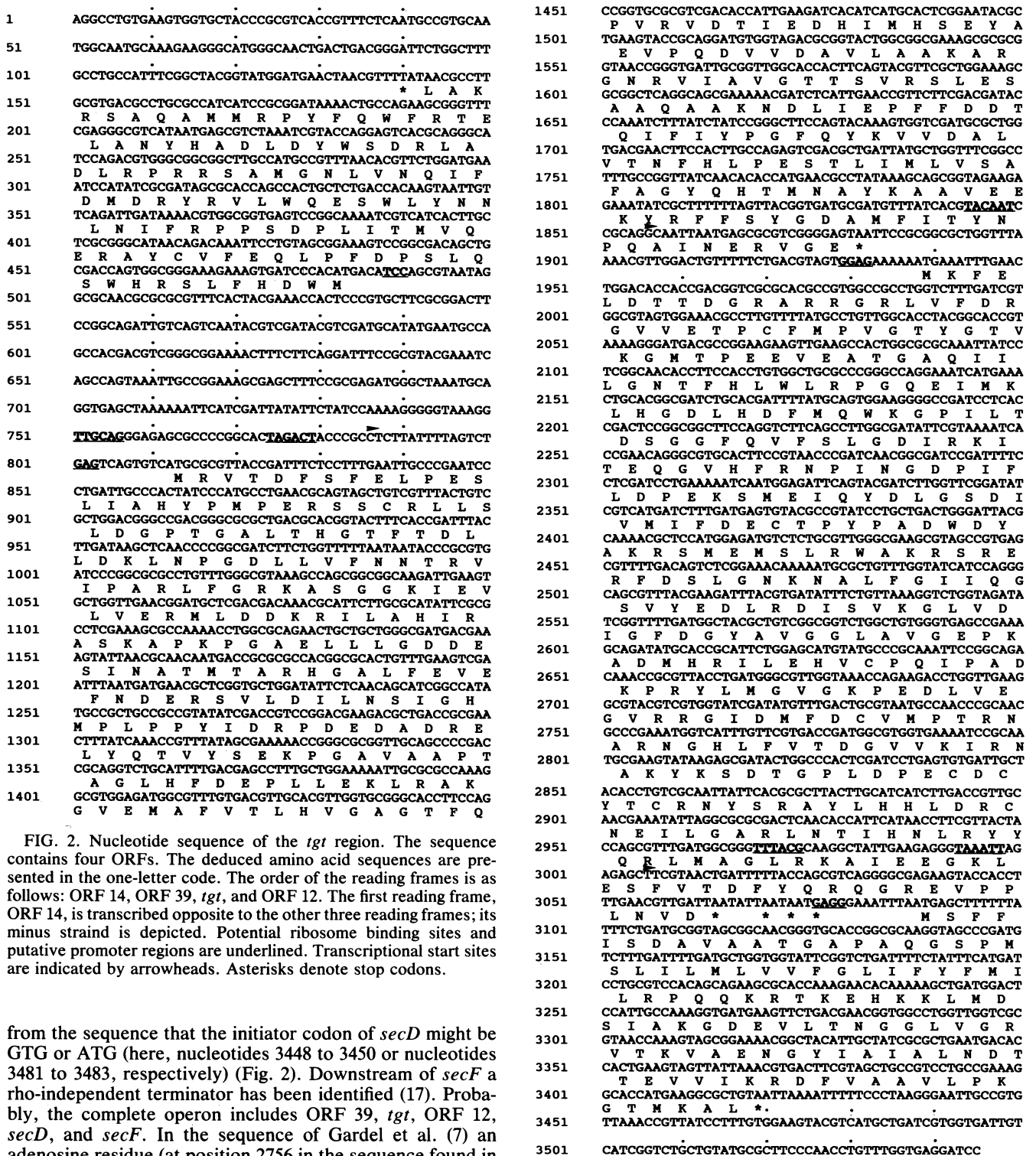


FIG. 2. Nucleotide sequence of the *tgt* region. The sequence contains four ORFs. The deduced amino acid sequences are presented in the one-letter code. The order of the reading frames is as follows: ORF 14, ORF 39, *tgt*, and ORF 12. The first reading frame, ORF 14, is transcribed opposite to the other three reading frames; its minus strand is depicted. Potential ribosome binding sites and putative promoter regions are underlined. Transcriptional start sites are indicated by arrowheads. Asterisks denote stop codons.

from the sequence that the initiator codon of *secD* might be GTG or ATG (here, nucleotides 3448 to 3450 or nucleotides 3481 to 3483, respectively) (Fig. 2). Downstream of *secF* a rho-independent terminator has been identified (17). Probably, the complete operon includes ORF 39, *tgt*, ORF 12, *secD*, and *secF*. In the sequence of Gardel et al. (7) an adenosine residue (at position 2756 in the sequence found in our study) is lacking. Therefore, the *tgt* gene was not recognized by Gardel et al. as a coherent reading frame. Furthermore, the GC at positions 3331 and 3332 within ORF 12 was stated to be CG (7). The 84th amino acid coded for by ORF 12 must be alanine and not arginine.

**Mapping of the 5' ends of the putative mRNAs.** The 5' ends of the respective mRNAs were mapped by primer extension analysis. To elucidate the start site of ORF 39 mRNA, an oligonucleotide complementary to nucleotides 812 to 833,

5'-AGGAGAAATCGGTAACGCGCAT-3' was annealed to total RNA from *E. coli* K-12. Primer extension with reverse transcriptase was performed, and the size of the product was determined by electrophoresis on a DNA sequencing gel. The DNA sequence of plasmid pHS1, analyzed with the same primer, served as a size marker (Fig. 3a). The 5' end of

TABLE 2. pUSB12 derivatives and their ability to complement the *tgt* mutation<sup>a</sup>

Plasmid	Genes encoded by insert	Nucleotide positions	[ <sup>3</sup> H]guanine incorporation (pmol/A <sub>260</sub> unit of tRNA) <sup>b</sup>
pDE1b	Complete <i>tgt</i> ; complete ORF 12	1844–3545	2 ± 0.2
pDE2a <sup>c</sup>	<i>tgt</i> lacking first 2 amino-terminal codons; complete ORF 12	1944–3545	5 ± 0.5
pDE4-12	<i>tgt</i> lacking 30 amino-terminal codons; complete ORF 12	2027–3445	115 ± 19
pBSB25	Complete <i>tgt</i> ; ORF 12 lacking 94 carboxy-terminal codons	1723–3135	0
pBSB5	<i>tgt</i> lacking 2 carboxy-terminal codons	1723–3057	11 ± 2
pBSB12	<i>tgt</i> lacking 49 carboxy-terminal codons	1723–2916	120 ± 17

<sup>a</sup> Deletions of the pUSB12 insert were produced by exonuclease III treatment for sequence analysis from both sides. All inserts were under control of the vector-encoded *lac* promoter.

<sup>b</sup> Low values of [<sup>3</sup>H]guanine incorporation into tRNA (up to 5 pmol per A<sub>260</sub> unit of tRNA) in the transglycosylase assay (see Materials and Methods) represent almost complete modification of the corresponding tRNAs with respect to Q. Data are means ± standard deviations.

<sup>c</sup> See "Characterization of gene products" in Results.

ORF 39 corresponded to nucleotide 786, which is located 26 bp upstream of the initiator codon. A putative Pribnow-Schaller box (22), TAGACT, differing from the eubacterial consensus sequence TATAAT in two bases, was found 6 bp further upstream. At the correct interval of 17 bp further upstream, a -35 region, TTGCAG, was present, differing from the consensus sequence TTGACA in three bases.

The results indicate that at least ORF 39, *tgt*, and ORF 12 compose one operon. Complementation analysis, however, revealed that the *tgt* gene can be expressed independently of the ORF 39 promoter (data not shown). In accordance with this result, an additional promoter site was detected by using

the 22-mer oligonucleotide 5'-CATTTCCTCCACTACGT CAG-3', which is complementary to nucleotides 1919 to 1940. This oligonucleotide was annealed to total RNA from *E. coli* K-12. The mRNA start site was found to be located within ORF 39 at nucleotide 1856 and 81 nucleotides upstream of the *tgt* initiator codon ATG (Fig. 3b). Six base pairs further upstream a -10 box with the sequence TAC AAT was present; however, a -35 region was not detectable within a reasonable distance.

To elucidate whether an additional promoter also precedes ORF 12, the 22-mer oligonucleotide with the sequence 5'-CAGCATCAAATCAAATCAAAGACATC-3', comple-

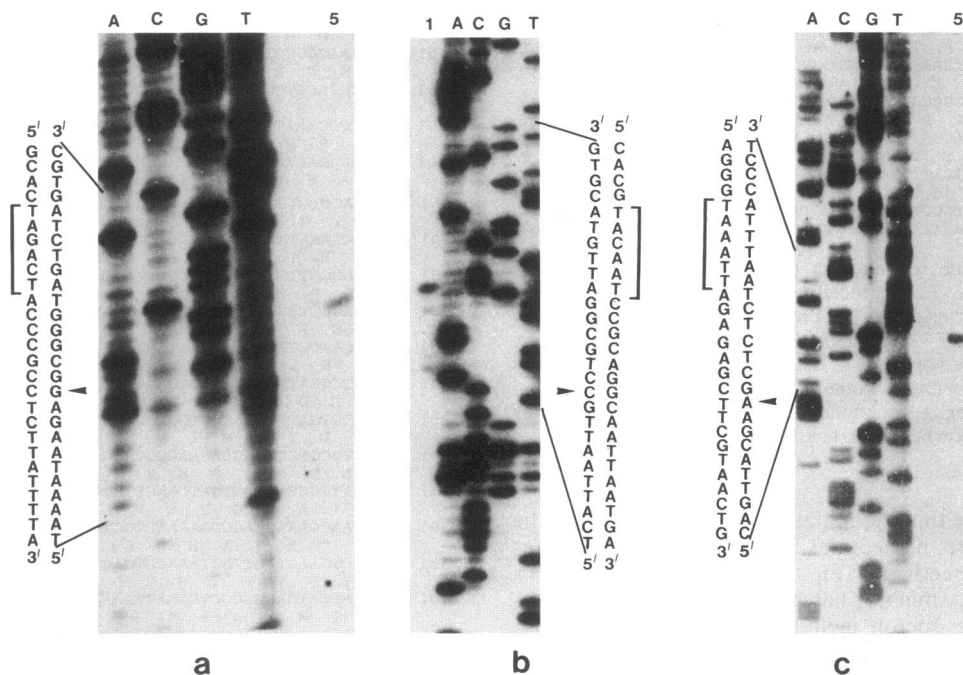


FIG. 3. Primer extension analysis of mRNA start sites. (a) ORF 39. An oligonucleotide complementary to nucleotides 812 to 833 was annealed to total RNA from *E. coli* K-12. The oligonucleotide was extended with reverse transcriptase, and the products were analyzed on a DNA sequencing gel. The primer extension product is shown in lane 5. The sequencing ladder of the DNA from plasmid pHS1, starting with the same oligonucleotide, served as a size marker. The relevant region of the sequence is shown. The mRNA start site is indicated by an arrowhead, and the -10 region is indicated by a bracket. (b) *tgt*. Analysis was performed with total RNA from *E. coli* K-12 with an oligonucleotide complementary to nucleotides 1919 to 1940. The primer extension product is shown in lane 5. The sequencing ladder of the DNA from plasmid pUSB12, starting with the same oligonucleotide, served as a size marker. The nucleotide sequence of the corresponding size marker is included as in panel a. (c) ORF 12. Analysis was performed with total RNA from *E. coli* K-12 with an oligonucleotide complementary to nucleotides 3147 to 3168. The primer extension product is shown in lane 5. The sequencing ladder of the DNA from plasmid pUSB12, starting with the same oligonucleotide, served as a size marker.

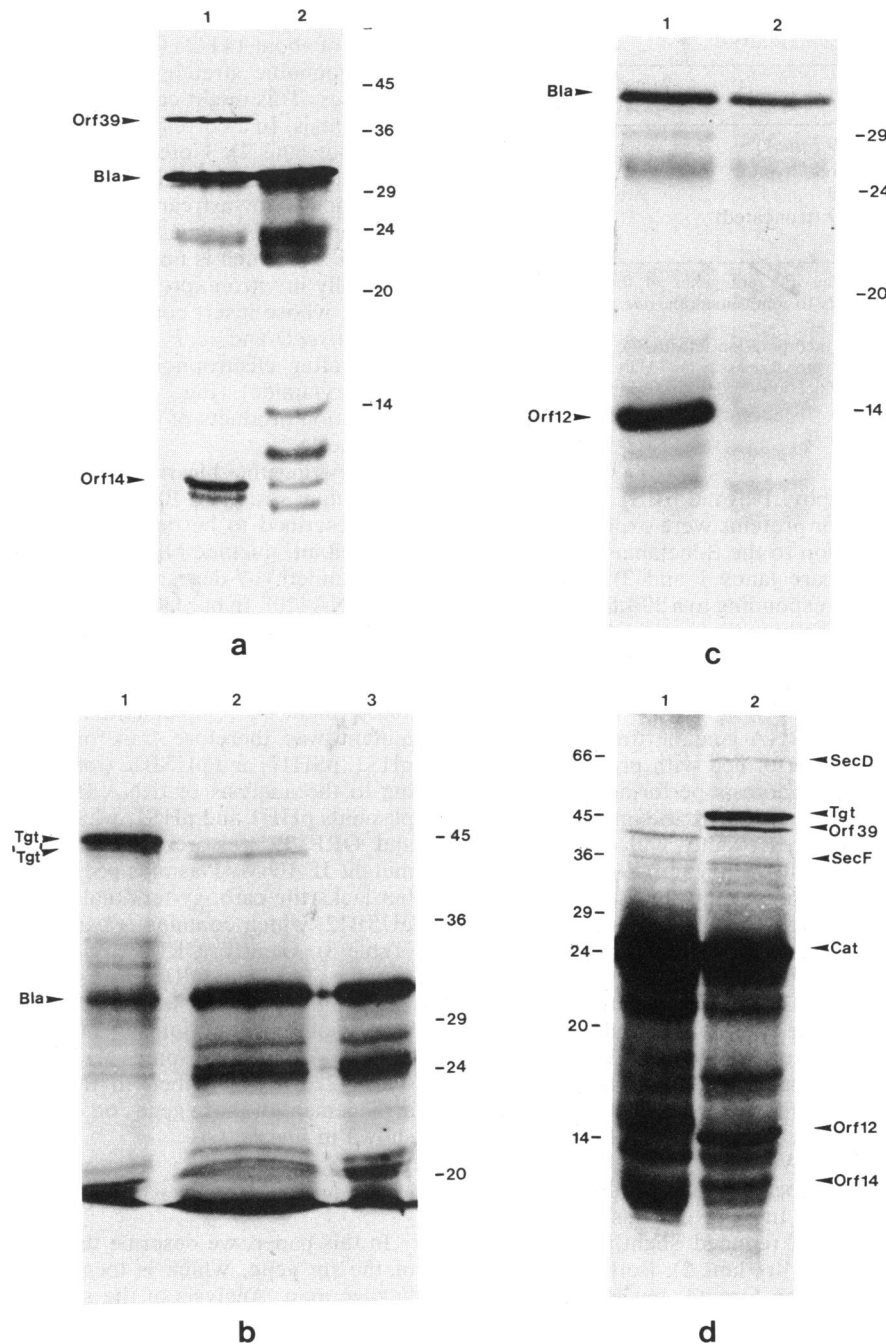


FIG. 4. Analysis of [<sup>35</sup>S]methionine-labeled proteins synthesized in vitro from various plasmids containing the sequenced genes. (a) Gene products of pHS1, which contains ORF 14 and ORF 39 (lane 1), compared with those of the vector pUC19 (lane 2). The proteins were separated on an SDS-PAGE gel (15% polyacrylamide). Bla, β-Lactamase. (b) Gene products of pUSB12, which contains *tgt* (lane 1), and of pDE2a, which contains the truncated *tgt* (lane 2), compared with those of pUC19 (lane 3). The proteins were separated on an SDS-PAGE gel (10% polyacrylamide). (c) Gene products of pDE4-12 (derivative of pUSB12 containing ORF 12) (lane 1) compared with those of pUC19 (lane 2). The proteins were separated on an SDS-PAGE gel (15% polyacrylamide). (d) Gene products of pHH1, which contains the putative operon. The gene products of pHH1 are shown in lane 2; the gene products of the vector pACYC184 are shown in lane 1. Cat, Chloramphenicol acetyltransferase. The proteins were separated on an SDS-PAGE gel (15% polyacrylamide).

mentary to nucleotides 3147 to 3168, was used. The mRNA start site was found at nucleotide 3006 (Fig. 3c); this site was preceded by a -10 region, TAAATT, and by a -35 region, TTTACG, separated from the putative Pribnow-Schaller box by 18 bp.

No mRNA start site for ORF 14 was identified.

**Characterization of gene products.** Various plasmids that carried sequenced genes were tested for in vitro expression in a combined transcription-translation assay. The [<sup>35</sup>S]methionine-labeled proteins were separated by SDS-PAGE and

TABLE 3. Complementation analysis of the *queA* mutation of strain JE 10651<sup>a</sup>

Plasmid	Genes encoded by insert	[ <sup>3</sup> H]guanine incorporation (pmol/A <sub>260</sub> unit of tRNA) <sup>b</sup>
pHH1	ORF 14, ORF 39 ( <i>queA</i> ), <i>tgt</i> , ORF 12, <i>secD</i> , <i>secF</i>	0
pHS1	ORF 14, ORF 39	2 ± 0.3
pSH17	ORF 14, ORF 39 (truncated)	63 ± 8
pUSB12	<i>tgt</i> , ORF 12	59 ± 3

<sup>a</sup> Various subclones containing different genes of the putative *tgt-sec* operon were tested for their ability to complement the *queA* mutation of strain JE 10651.

<sup>b</sup> Low values of [<sup>3</sup>H]guanine incorporation into tRNA (up to 5 pmol per A<sub>260</sub> unit of tRNA), in the transglycosylase assay (see Materials and Methods) represent almost complete modification of the corresponding tRNAs with respect to Q.

analyzed by autoradiography. Plasmid pHS1 encoded ORF 14 and ORF 39. Two major proteins were expressed and are visible in Fig. 4a in addition to the β-lactamase encoded by the pUC19 vector (compare lanes 1 and 2). One protein migrated to a position corresponding to a 39-kDa protein; the molecular mass agreed exactly with the predicted molecular mass of the ORF 39 gene product. The other protein, which migrated to the molecular mass region of about 12 kDa, must be the product of ORF 14.

The product of *tgt*, the tRNA guanine transglycosylase, was obtained in an assay performed with plasmid pUSB12 (Fig. 4b, lane 1). In electrophoresis performed with a 10% polyacrylamide gel, Tgt migrated to the molecular mass region of 46 kDa. This agrees exactly with the molecular mass described for the purified enzyme (19). Both values differ somewhat from the molecular mass of 42.5 kDa deduced from the nucleotide sequence, indicating differences between the overall distribution of amino acids of Tgt and that of the marker protein.

Surprisingly, plasmid pDE2a, which contains nucleotides 1944 to 3545 but lacks the ribosomal binding site, the promoter, and the first two codons of the *tgt* gene, complemented the *tgt* mutation in part (Table 2). Q modification of tRNA was completed only in the stationary phase. This indicates that the modifying enzyme was present in growing cells at a very low level. When the truncated gene was cloned into a low-copy-number plasmid, pACYC184, complementation did not occur. In vitro expression of pDE2a resulted a *tgt* gene product reduced slightly in size and considerably in amount (Fig. 4b, lane 2). Both results show that an incomplete Tgt can be formed, but in rather small amounts. Anor codon might be the GTG at positions 2007 to 2009. It is not known whether the alternate expression is of any biological significance. A similar phenomenon was observed in the case of the *mcrB* gene (23). The *mcrB* locus of *E. coli* K-12 mediates sequence-specific restriction of cytosine-modified DNA. For comparison, the proteins derived from the vector pUC19 are shown (Fig. 4b, lane 3).

The product of ORF 12 was identified upon in vitro

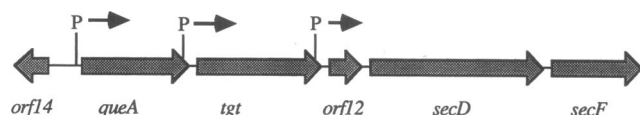


FIG. 5. Schematic representation of the putative *tgt-sec* operon. Regions determined to be promoters are indicated by P.

expression of plasmid pDE4-12. Migration was found in the region of about 14 kDa (Fig. 4c, lane 1). The protein contains a hydrophobic stretch of about 30 amino acids at the N terminus. This might cause a slight retardation during electrophoresis. In a corresponding control, pUC19 was applied (Fig. 4c, lane 2). Note that the protein with an apparent molecular mass of 14 kDa on the gel is the product of ORF 12, which is downstream of *tgt*, and that the protein with an apparent molecular mass of 12 kDa on the gel is the product of ORF 14, which is upstream of ORF 39.

Finally in vitro expression was investigated with plasmid pHH1, whose insert contains all of the sequenced genes as well as *secD* and *secF*. The expected proteins were recognized after electrophoresis on an SDS-PAGE gel (15% polyacrylamide) (Fig. 4d, lane 2). For comparison, the translation products of plasmid pACYC184 are shown (Fig. 4d, lane 1).

**A new queuine biosynthesis gene, *queA* (ORF 39).** An *E. coli* mutant strain, JE 10651 (20), a derivative of *E. coli* K-12, was described to be defective in queuine biosynthesis. In this mutant, queuine biosynthesis stops after the precursor, 7-aminomethyl-7-deazaguanine, has been incorporated into the tRNA (20). In our laboratory, tRNAs from this strain can be characterized by the tRNA transglycosylase assay (see Materials and Methods). The proximity of ORF 39 and ORF 12 to *tgt* led us to suggest that the genetic defect of mutant JE 10651 might be complemented by one of these genes. The mutant was therefore transformed with plasmids pHH1, pHS1, pSH17, and pUSB12 (for details, see Fig. 1). According to the analysis of tRNA from the transformants, only plasmids pHH1 and pHS1, which encode the intact ORF 14 and ORF 39, restored the biosynthesis of queuosine in mutant JE 10651. Plasmids pSH17, which contains ORF 14, but lacks the carboxy-terminal 140 codons of ORF 39, and pUSB12, which contains *tgt* and ORF 12, were ineffective (Table 3). Clearly, ORF 39 encodes a new Q biosynthesis enzyme. The gene ORF 39 was designated *queA*.

**Chromosomal organization of *tgt* and adjacent genes.** Comparison of the restriction map of the insert of pPR20 with that of the *E. coli* chromosome (11) indicated that *queA*, *tgt*, and ORF 12 are transcribed in the clockwise direction. The arrangement of these genes on the putative *tgt-sec* operon is shown in Fig. 5.

## DISCUSSION

In this paper, we describe the structure and organization of the *tgt* gene, which is located at 9 min on the *E. coli* linkage map. Analysis of the sequence and organization of the *tgt* region supports the view that *tgt* is part of a complex operon. The putative operon comprises five genes, and is preceded by a gene transcribed in the opposite orientation. Upstream, close to *tgt*, a new Q biosynthesis gene designated *queA* was discovered. Complementation studies using strain JE 10651, which has a mutation in Q biosynthesis beyond the Tgt step, suggest that the gene is responsible for the synthesis and/or attachment of the cyclopentadiol moiety to the 7-aminomethyl group of 7-deazaguanine at the level of tRNA.

Further analysis of the *tgt* region indicated that *queA* and *tgt* are followed by an ORF encoding a 12-kDa protein whose function is as yet unknown. Downstream of ORF 12, the genes *secD* and *secF* are located. The sequences of *secD* and *secF* were published recently by Gardel et al. (7). Here we show that part of the *tgt* sequence and that of ORF 12 overlap the 5' adjacent sequence described for *secD*. The *tgt*

gene was not recognized as a coherent reading frame; Gardel et al. did not detect the adenosine residue in *tgt* at position 2756 in the presented sequence. Also, the ORF 12 shown here and in reference 7 are at variance. The two reading frames ORF 1 and ORF 2 described in reference 7 are incorrect; they do not code for the respective proteins. Therefore, we include the overlapping sequences in Fig. 2.

The proposal that the five genes are part of one operon is based on the following results. (i) The five genes occur close to each other and are transcribed in the same orientation. (ii) A typical prokaryotic promoter region, including a Pribnow-Schaller box and a  $-35$  consensus sequence exactly 17 bp further upstream, was found in the 5' sequence preceding the first gene, *queA*. (iii) The next gene upstream of *queA*, namely, ORF 14, is transcribed counterclockwise. (iv) A typical rho-independent terminator sequence was found only at the 3' end, downstream of *secF*.

The putative *tgt* operon contains additional promoters. Two promoters were found, one in front of the *tgt* gene and another in front of ORF 12. From these promoters, *tgt* and ORF 12, respectively, can be transcribed independently of the *queA* promoter. Each of these promoters was found to be localized within the preceding gene, so that termination between *queA* and *tgt* or between *tgt* and ORF 12 is rather unlikely. The promoter region for *tgt* included a Pribnow-Schaller box, while a  $-35$  consensus sequence was not present, a characteristic feature for positively regulated promoters (21). This might be a hint that the *tgt* promoter requires binding of an activator protein.

The  $-35$  and the  $-10$  regions of the *queA* promoter are spaced by a GC-rich sequence (14 out of 17 bp are GC). A stretch of GC base pairs resembling a discriminator region precedes the transcription start point, a typical quality of stringently controlled promoters (31, 32). We assume that *secD* and *secF* are part of this operon. This agrees well with the fact that genes encoding products which are involved in the transport of proteins are located on complex operons that are subject to mechanisms of growth control. Thus, *secA* occurs in a cluster of regulated genes involved in cell fission and the biosynthesis of the murein sacculus (26). The genes *secY* and *secE* are organized on a stringently controlled operon, encoding ribosomal proteins (2, 25).

ORF 14, which precedes the putative *tgt-sec* operon, appears to be rather weakly expressed. Possibly the product of ORF 14 is a regulatory protein that controls the expression of the *tgt* operon.

Computational comparison of the sequenced genes, ORF 14, *queA*, *tgt*, and ORF 12, with other known genes by using the GenBank and EMBL data bases did not reveal any significant homologies. With respect to *tgt* and *queA* this is not surprising, because the tRNA-modifying enzymes are highly specific with respect to the tRNA species, the specific base to be modified, and the position in the tRNA molecule.

#### ACKNOWLEDGMENTS

S. Nishimura provided us with the *E. coli* mutants; we acknowledge his support of this work and many fruitful discussions with gratitude. We also thank B. Frey for engagement in this work and J. Tommassen for providing plasmid pPR20. We thank Sonja Noeth for patient help with the manuscript.

This work was supported by the Deutsche Forschungsgemeinschaft (grant Ke 98/19-3) and by the Fond der Chemischen Industrie.

#### REFERENCES

- Bachmann, B. J. 1990. Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol. Rev.* **54**:130-197.
- Cerretti, D. P., D. Dean, G. R. Davis, D. M. Bedwell, and M. Nomura. 1983. The *spc* ribosomal protein operon of *E. coli*: sequence and cotranscription of the ribosomal protein genes and a protein export gene. *Nucleic Acids Res.* **11**:2599-2616.
- Chang, A. C. Y., and S. N. Cohen. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**:1141-1145.
- Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387-395.
- Frey, B., G. Jänel, U. Michelsen, and H. Kersten. 1989. Mutations in the *Escherichia coli* *fnr* and *tgt* genes: control of molybdate reductase activity and the cytochrome *d* complex by *fnr*. *J. Bacteriol.* **171**:1524-1530.
- Frey, B., J. McCloskey, W. Kersten, and H. Kersten. 1988. New function of vitamin B<sub>12</sub>: cobamide-dependent reduction of epoxycytosine to cytosine in tRNAs of *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.* **170**:2078-2082.
- Gardel, C., K. Johnson, A. Jacq, and J. Beckwith. 1990. The *secD* locus of *E. coli* codes for two membrane proteins required for protein export. *EMBO J.* **9**:3209-3216.
- Gribskov, M., J. Devereux, and R. R. Burgess. 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. *Nucleic Acids Res.* **12**:539-549.
- Kersten, H. 1988. The nutrient factor queuine: biosynthesis, occurrence in transfer RNA and function. *BioFactors* **1**:27-29.
- Kersten, H., and W. Kersten. 1990. Biosynthesis and function of queuine and queuosine tRNAs, p. B69-B108. In C. W. Gehrke and K. C. T. Kuo (ed.), *Chromatography and modification of nucleosides*, part B. Biological roles and function of modification. Elsevier, Amsterdam.
- Kohara, Y., K. Akiyama, and K. Isono. 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**:495-508.
- Kuchino, Y., H. Kasai, K. Nihel, and S. Nishimura. 1976. Biosynthesis of the modified nucleoside Q in transfer RNA. *Nucleic Acids Res.* **3**:393-398.
- Langgut, W., T. Reisser, and H. Kersten. 1990. Queuine modulates growth of HeLa cells depending on oxygen availability. *BioFactors* **2**:245-249.
- Mahr, U., P. Böhm, and H. Kersten. 1990. Possible involvement of queuine in control mechanisms of protein synthesis and protein phosphorylation in eukaryotes. *BioFactors* **2**:185-192.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Miller, J. H. 1972. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Nishimura, S. 1983. Structure, biosynthesis, and function of queuosine in transfer RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **28**:49-73.
- Nogushi, S., Y. Nishimura, Y. Hirota, and S. Nishimura. 1982. Isolation and characterization of an *Escherichia coli* mutant lacking tRNA guanine transglycosylase. *J. Biol. Chem.* **257**:6544-6550.
- Okada, N., and S. Nishimura. 1979. Isolation and characterization of a guanine insertion enzyme, a specific tRNA transglycosylase from *Escherichia coli*. *J. Biol. Chem.* **254**:3061-3066.
- Okada, N., S. Nogushi, S. Nishimura, T. Ohgi, T. Goto, P. F. Crain, and J. A. McCloskey. 1978. Structure determination of a nucleoside Q precursor isolated from *E. coli* tRNA: 7-(aminomethyl)-7-deazaguanosine. *Nucleic Acids Res.* **5**:2289-2296.
- Raiband, O., and M. Schwartz. 1984. Positive control of transcription initiation in bacteria. *Annu. Rev. Genet.* **18**:173-206.
- Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13**:319-353.
- Ross, T. K., E. C. Achberger, and H. D. Braymer. 1989. Nucleotide sequence of the McrB region of *Escherichia coli* K-12 and evidence for two independent translational initiation



- sites at the *mcrB* locus. *J. Bacteriol.* **171**:1974–1981.
24. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
  25. Schatz, P. J., P. D. Riggs, A. Jacq, M. J. Fath, and J. Beckwith. 1989. The *secE* gene encodes an integral membrane protein required for protein export in *E. coli*. *Genes Dev.* **3**:1035–1044.
  26. Schmidt, M. G., E. E. Rollo, J. Grodberg, and D. B. Oliver. 1988. Nucleotide sequence of the *secA* gene and *secA*(Ts) mutations preventing protein export in *Escherichia coli*. *J. Bacteriol.* **170**:3404–3414.
  27. Shine, J., and L. Dalgarno. 1975. Determinants of cistron specificity in bacterial ribosomes. *Nature (London)* **254**:34–38.
  28. So, M., R. Gill, and S. Falkow. 1975. The generation of a ColE1-Ap<sup>r</sup> cloning vehicle which allows detection of inserted DNA. *Mol. Gen. Genet.* **142**:239–249.
  29. Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified T7 DNA polymerase. *Proc. Acad. Natl. Sci. USA* **84**:4761–4771.
  30. Tommassen, J., P. De Geus, and B. Lugtenberg. 1982. Cloning of the regulatory genes *phoB* and *phoR* and identification of their gene products. *J. Mol. Biol.* **157**:265–274.
  31. Travers, A. 1984. Conserved features of coordinately regulated *E. coli* promoters. *Nucleic Acids Res.* **12**:2605–2618.
  32. Travers, A. A. 1980. Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. *J. Bacteriol.* **141**:973–976.
  33. Williams, J. G., and P. J. Mason. 1985. Hybridization in the analysis of RNA, p. 139–160. *In* B. D. Hames and S. J. Higgins (ed.), *Nucleic acids hybridisation—a practical approach*. IRL Press, Oxford.
  34. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:103–119.