JOURNAL OF BACTERIOLOGY, Nov. 1995, p. 6371–6380 Vol. 177, No. 22
0021-9193/95/$04.00+0
Copyright © 1995, American Society for Microbiology

The Cobalamin (Coenzyme B_{12}) Biosynthetic Genes of *Escherichia coli*

JEFFREY G. LAWRENCE AND JOHN R. ROTH*

Department of Biology, University of Utah, Salt Lake City, Utah 84112

Received 16 June 1995/Accepted 2 September 1995

The enteric bacterium *Escherichia coli* synthesizes cobalamin (coenzyme B_{12}) only when provided with the complex intermediate cobinamide. Three cobalamin biosynthetic genes have been cloned from *Escherichia coli* K-12, and their nucleotide sequences have been determined. The three genes form an operon (*cob*) under the control of several promoters and are induced by cobinamide, a precursor of cobalamin. The *cob* operon of *E. coli* comprises the *cobU* gene, encoding the bifunctional cobinamide kinase-guanylyltransferase; the *cobS* gene, encoding cobalamin synthetase; and the *cobT* gene, encoding dimethylbenzimidazole phosphoribosyltransferase. The physiological roles of these sequences were verified by the isolation of Tn*10* insertion mutations in the *cobS* and *cobT* genes. All genes were named after their *Salmonella typhimurium* homologs and are located at the corresponding positions on the *E. coli* genetic map. Although the nucleotide sequences of the *Salmonella cob* genes and the *E. coli cob* genes are homologous, they are too divergent to have been derived from an operon present in their most recent common ancestor. On the basis of comparisons of G+C content, codon usage bias, dinucleotide frequencies, and patterns of synonymous and nonsynonymous substitutions, we conclude that the *cob* operon was introduced into the *Salmonella* genome from an exogenous source. The *cob* operon of *E. coli* may be related to cobalamin synthetic genes now found among non-*Salmonella* enteric bacteria.

Although commonly touted as one of the largest and oldest of the cofactors participating in modern biochemistry (20, 23, 59, 60), the physiological importance of cobalamin (coenzyme B_{12}), especially among enteric bacteria, has remained enigmatic (58, 63). Cobalamin is derived from uroporphyrinogen III, a precursor in the synthesis of heme, siroheme, and chlorophylls as well as cobamides (Fig. 1). The conversion of uroporphyrinogen III to the complex intermediate cobinamide is referred to as part I of the biosynthetic pathway (hereafter designated CobI). Part II of the pathway (CobII) entails the biosynthesis of dimethylbenzimidazole (DMB) from probable flavin precursors (11, 27, 39). The covalent linkage of cobinamide, DMB, and a phosphoribosyl group donated by nicotinate mononucleotide is achieved by part III of the pathway (CobIII). Although most enteric bacteria can synthesize cobalamin de novo under either aerobic or anaerobic conditions (38), *Escherichia coli* synthesizes cobalamin only when provided with the complex intermediate cobinamide (38, 66); therefore, *E. coli* performs only parts II and III of the cobalamin biosynthetic pathway.

The cobalamin biosynthetic genes in the closely related enteric bacterium *Salmonella typhimurium* have been characterized, and most constitute a large, 20-gene operon located at min 41 on the genetic map (28, 29, 52). It has been proposed that these genes are not ancestral to the *Salmonella* chromosome but have been introduced by horizontal transfer at or following the divergence of the salmonellae from *Escherichia* species; this operon is found in nearly all extant *Salmonella* species (38). These conclusions were based on striking differences in the patterns of cobalamin synthesis and use among the salmonellae in comparison with those of other enteric bacteria. *Salmonella* species synthesize cobalamin de novo only under anaerobic growth conditions, and the cobalamin synthetic genes are induced by propanediol during both aerobic (requiring cyclic AMP [cAMP] receptor protein-cAMP complex [CRP-cAMP]) and anaerobic (requiring CRP-cAMP and/or the ArcA protein) growth conditions (1, 7, 12, 18, 51).

In contrast, species of *Klebsiella* and *Citrobacter* readily produce cobalamin under aerobic growth conditions, and production is not influenced by propanediol or the levels of the CRP-cAMP complex. Moreover, cobalamin-producing enteric species, except for *Salmonella* spp., ferment glycerol in a coenzyme B_{12}-dependent fashion. These factors suggested that cobalamin synthesis and use in *Salmonella* spp. differ markedly from cobalamin metabolism in closely related enteric species. This conclusion is reinforced by the lack of DNA hybridization between *Salmonella* cobalamin genes and the genomes of closely related enteric bacteria, including *E. coli* (38).

To test the hypothesis that the *cob* operon was introduced into the *Salmonella* genome, it is necessary to compare directly the cobalamin biosynthetic genes from various enteric bacteria. We have characterized the genes encoding CobII and CobIII in *E. coli* K-12. The data support the hypothesis that the differences in cobalamin synthesis and use between *E. coli* and *Salmonella* spp. can be attributed to the horizontal transfer of DNA, including the *cob* operon, into the *Salmonella* genome.

MATERIALS AND METHODS

Bacterial strains and culture conditions. Bacterial strains are listed in Table 1. Strains of the ECOR collection of natural isolates of *E. coli* (46) were provided by H. Ochman. An *E. coli* strain (TT17196) harboring a *metE*::Tn*10* insertion was supplied by G. Stauffer. The *Klebsiella pneumoniae* wild-type strain M5a1 (TR7176) was provided by G. Roberts. The rich medium used was Luria-Bertani (LB) medium; defined media included E medium (65) and its derivative NCE, which lacks carbon sources. MacConkey indicator agar was used with the addition of specific carbon sources each at a concentration of 1% (wt/vol). Kanamycin and tetracycline were added to media at final concentrations of 40 and 20 μg/ml, respectively.

Genetic methods. Transductional crosses mediated by bacteriophage P22 were performed as described previously (14) by using the highly transducing bacteriophage mutant P22 HT *int*-105. To prepare bacteriophage P22 lysates from *S. typhimurium* TR6579 (*galE*), cells were grown on LB medium supplemented with 0.2% glucose and 0.02% galactose to permit bacteriophage adsorption. Trans-

* Corresponding author. Phone: (801) 581-3412. Fax: (801) 585-6207. Electronic mail address (Internet): ROTH@BIOSCIENCE.UTAH.EDU.

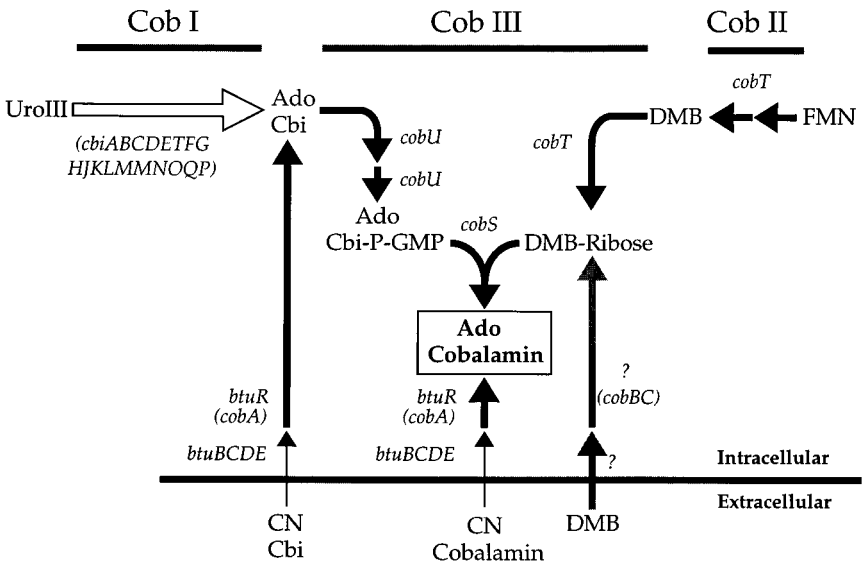

FIG. 1. Outline of cobalamin biosynthesis in enteric bacteria. *E. coli* genes encoding the relevant enzymes are shown; *S. typhimurium* gene designations are given in parentheses. Abbreviations: Ado, adenosyl; Cbi, cobinamide; CN, cyano; FMN, flavin mononucleotide; UroIII, uroporphyrinogen III. A reaction unique to *S. typhimurium* (open arrows), reactions involving genes and/or mutations which have been characterized in this study (heavy arrows), reactions inferred from this study (grey arrows), and reactions previously characterized for *E. coli* and *S. typhimurium* (thin arrows) are shown.

ductional crosses mediated by bacteriophage P1 were performed as described elsewhere (38) by using the bacteriophage mutant P1 *vir*.

A deletion of the *metE* gene of *E. coli* was created by selecting a tetracycline-sensitive mutant of strain TT18495 (*metE163*::Tn*10*) by the method of Bochner et al. (8). Tetracycline-sensitive mutants were screened via DNA hybridization for the presence of *tetA* and *tetR* genes; strains lacking these sequences were inferred to bear deletions removing adjacent *metE* sequences. Strains bearing the *metE571* deletion depend upon the cobalamin-dependent MetH enzyme for methionine synthesis.

Insertion mutations in *E. coli cob* genes were created by introducing Tn*10d*Tc elements into a *metE* strain (TT18496) via bacteriophage λ as described previously (32). Tetracycline-resistant transductants were screened for mutants that had lost the ability to use cobinamide but whose loss could be corrected by exogenous cobalamin or methionine. The tetracycline resistance determinants were transduced via bacteriophage P1 into the parent *E. coli* strain, and the linkages of the Tet^r and Cob$^-$ phenotypes were determined. Insertion mutations with 100% linkage of the Tet^r and Cob$^-$ phenotypes were assumed to be located in genes required either for cobinamide transport or for transformation of cobinamide into cobalamin.

Insertion mutants were separated into the CobII and CobIII nutritional classes by determining if the strains could grow with exogenous DMB and cobinamide. If so, the mutations were classified as CobII; otherwise, they were classified as CobIII. Tests were performed under both aerobic and anaerobic conditions.

TABLE 1. Strains used

Species and strain	Description or genotype
E. coli	
ECOR 1–72[a]	Natural isolates
TR7177	Wild-type K-12 W3310
TT18495	*metE163*::Tn*10*
TT18496	Δ*metE571*
TT18497[b]	*supE44 thi-1* Δ(*lac-proAB*)
TT18498[b]	*supE44 thi-1* Δ(*lac-proAB*) pZT385
TT18499[b]	*supE44 thi-1* Δ(*lac-proAB*) pZT386
TT18500	Δ*metE571 cobT14*::Tn*10d*Tc
TT18501	Δ*metE571 cobT15*::Tn*10d*Tc
TT18502	Δ*metE571 cobT16*::Tn*10d*Tc
TT18503	Δ*metE571 cobS6*::Tn*10d*Tc
TT18504	Δ*metE571 cobS7*::Tn*10d*Tc
TT18505	Δ*metE571 cobS12*::Tn*10d*Tc
TT18506	Δ*metE571 btuR510*::Tn*10d*Tc
TT18507	Δ*metE571 btuR511*::Tn*10d*Tc
S. typhimurium	
TR6579	*metA22 metE551 trpD2 ilv-452 hsdA29 hsdB strA120 galE leu pro*
TT10858	*metE205 ara-9 cob-66*::MudJ
TT18508	*metE205 ara-9 cob-66*::MudJ pZT387
TT11855	*metE205 ara-9* DEL299 (*his cob*)
S. arizonae TR6331	Wild type
S. seminole TR6332	Wild type
K. pneumoniae TR7176	Wild-type M5a1

[a] The ECOR strains are described by Ochman and Selander (46).

[b] Strain TT18497 (JM101) and all its derivatives contain the fertility factor F′ *traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15.

Molecular methods. Plasmid DNA was prepared by Qiagen column purification according to the manufacturer's instructions, and bacteriophage λ DNA was prepared as described elsewhere (26). Chromosomal DNA was prepared from *E. coli* as described previously (57). Agarose gel electrophoresis and DNA hybridization followed standard protocols (55), and DNA fragments were labeled to high specific activity as described previously (21). Cells used for DNA electroporation were grown to stationary phase overnight in rich medium. The cells were recovered by centrifugation and prepared for electroporation by the method of Dower et al. (16). DNA was amplified by the PCR as described elsewhere (53, 54). DNA sequencing of plasmid and PCR-amplified DNAs was performed as described previously (36).

Sequences flanking Tn*10d*Tc elements were amplified with oligonucleotides specific to the tetracycline resistance determinant which prime DNA synthesis directed out of the transposon. When used with oligonucleotide primers in the flanking DNA, specific sequences are amplified. The primer facing out of the *tetR* gene was R635 (5′TCCATTGCTGTTGACAAAGGGAAT3′); the primer facing out of the *tetA* gene was F3376 (5′ACCTTTGGTCACCAACGCTTTTC C3′).

Promoter mapping. RNA transcript mapping followed the protocol of Kofoid et al. (33). Cells were grown to early log phase (net, 70 Klett units), placed on ice, recovered by centrifugation, and resuspended in one-half volume of distilled, deionized water. Oligonucleotides were labeled with [γ-^{33}P]ATP as described elsewhere (55). Cells were mixed with the radiolabeled DNA primer and *T*th DNA polymerase, reverse transcription buffer, and $MnCl_2$ supplied by Promega. RNA was released from the cells by incubation at 94°C for 30 s and reverse transcription was performed for 99 cycles in an Idaho Technologies air-driven thermal cycler. Bands were visualized by size fractionating the reaction mixture by polyacrylamide gel electrophoresis in lanes adjacent to appropriate DNA sequencing reaction mixtures. The transcript amounts were quantified by using a Molecular Dynamics PhosphorImager.

Nucleotide sequence accession number. DNA sequence analysis employed the Genetics Computer Group program package (15). The *E. coli cob* operon sequence reported here has been submitted to GenBank and has been assigned accession no. U33333.

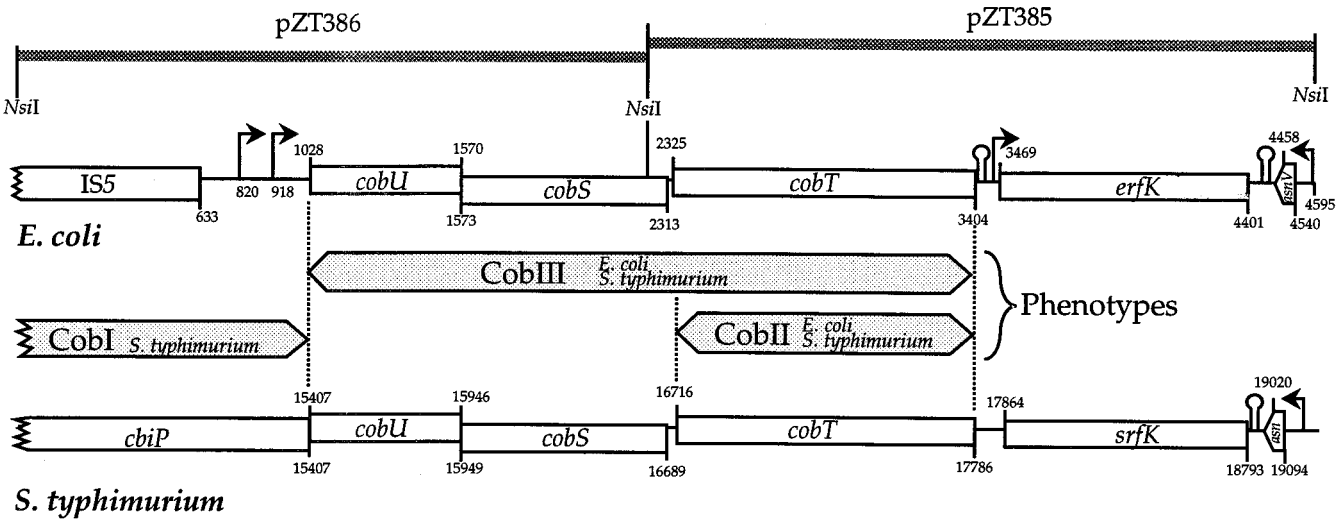

FIG. 2. Structure of the *E. coli cob* operon and the promoter-distal portion of the *S. typhimurium cob* operon. Coordinates indicate bases in the *E. coli cob* operon sequence or bases in the *Salmonella cob* operon sequence as specified by Roth et al. (52) (GenBank accession no. L12006), as extended by Chen et al. (11) (GenBank accession no. L35477). Coordinates for the translation start site and the translation stop site are indicated above and below the physical map, respectively. Regions carried by the specified plasmids (dark grey bars), putative promoters (arrows), and putative terminators (hairpins) are indicated.

RESULTS

Cloning the *E. coli cob* genes by complementation. Chromosomal DNA from *E. coli* K-12 (strain W3110) was partially digested with the restriction endonuclease *Sau*3AI and size fractionated by agarose gel electrophoresis. DNA fragments 5 to 12 kb long were isolated and ligated into purified pACYC177 DNA cleaved with *Bam*HI. The plasmids were introduced into *S. typhimurium* TR6579 by electroporation, and kanamycin-resistant colonies were selected. Transformant colonies were pooled, and the plasmids from this pool were transduced into *S. typhimurium* TT10858 (CobIII$^-$) by bacteriophage P22 with selection for kanamycin resistance. Strains with plasmid-borne CobIII genes were identified by their ability to grow on minimal medium when provided with exogenous cobinamide. Five plasmids were isolated, including pZT387. Strains bearing these plasmids were found to have severe growth defects (doubling times of over 120 min). Relaxation of the selection for cobalamin synthesis resulted in rapid loss of the plasmids or in plasmid rearrangement, as determined by restriction fragment analysis of plasmid DNA. It was inferred that the plasmids included a sequence that was detrimental to cell growth when present at high levels on a multicopy plasmid.

Cloning the *E. coli cob* genes by hybridization. The DNA sequence of a 1,251-bp fragment of the *E. coli* chromosome bearing portions of the *cobS* and *cobT* genes was provided by Carleen Collins (GenBank accession no. L25054). That nucleotide sequence was used to design oligonucleotide primers that allowed amplification of a 1,019-bp fragment via the PCR (53, 54). The amplified DNA was labeled to high specific activity and used to probe a membrane bearing bound DNA from the collection of Kohara et al. cloned DNA (34). Strong hybridization was shown by adjacent clones 7D4 (miniset 346) and 9B2 (miniset 347), located at min 43.5 on the *E. coli* genetic map (4). DNA was prepared from these bacteriophages and digested to completion with the restriction endonuclease *Nsi*I, which cleaved once within the described 1,251-bp DNA sequence at the 5′ end of the *cobT* homolog. The upstream and downstream fragments were cloned into plasmid pUC19 cleaved with the restriction endonuclease *Pst*I. Two clones, pZT385 and pZT386, bearing the upstream and downstream fragments, respectively, were isolated.

DNA sequence of the *E. coli cob* operon. Since the cloned insert in plasmid pZT387 was known to rearrange at high frequency, it was not employed for initial DNA sequence analysis. Instead, the nucleotide sequences of both strands of plasmids pZT385 and pZT386 were determined. The junction sequence between the two plasmids was determined by sequencing PCR-generated templates from *E. coli* K-12 chromosomal DNA. The contiguous DNA sequence spanned 4,595 bp. Upon completion of the DNA sequence, an additional 1,800 bp of DNA sequence, generated from six different primers, was determined from plasmid pZT387. This plasmid complemented the CobIII$^-$ phenotype of *S. typhimurium* TT10858. The DNA sequences from pZT387 were 100% identical to the DNA sequences generated from pZT385 and pZT386. These data show that the sequences cloned by hybridization complement CobIII$^-$ mutations and are likely to encode cobalamin biosynthetic genes.

Analysis of the nucleotide sequence revealed five open reading frames over 300 bp long. Other open reading frames were disregarded since they were insubstantial in length, exhibited no codon usage bias (codon adaptation index of <0.2 as determined by the method of Sharp and Li [62]), and did not show the G+C content bias by codon position (second codon position showing lower percent G+C) typical of genes with comparable nucleotide contents (43). Three of the inferred open reading frames were clearly homologous to the *cobU*, *cobS*, and *cobT* genes in the *S. typhimurium cob* operon which provide CobII and CobIII gene functions (11, 52). The *E. coli* genes were termed *cobU*, *cobS*, and *cobT* to indicate their similarity to the *Salmonella* homologs. Not surprisingly, database searches revealed that the *E. coli cobUST* genes, like the *S. typhimurium cobUST* genes, are similar to the homologous cobalamin synthetic genes (*cobPVU*) from *Pseudomonas denitrificans* (6, 10, 13).

The physical map of the *cob* operon region is shown in Fig. 2. Sequences highly similar to the consensus Shine-Dalgarno sequence for ribosome binding were found at appropriate lo-

cations upstream of all three genes. The *cobU*, *cobS*, and *cobT* genes all lie close to each other, with the 5′ end of the *cobS* gene overlapping the 3′ end of the *cobU* gene. The proximity of the genes suggested that they may form an operon expressed by promoters upstream of the *cobU* gene.

There is an 11-base spacer between the end of the *cobS* gene and the first possible translation initiation site of the *cobT* gene. Roth et al. (52) described a translation initiation site for the *Salmonella cobT* gene which overlapped the stop codon of the *cobS* gene; this start site is not conserved in the *E. coli* sequence. However, the putative start site for the *E. coli cobT* gene can also be identified in the *Salmonella cobT* gene 30 bases downstream of the start site suggested by Roth et al. (52). In addition, the peptide sequences of the two CobT proteins upstream of these start sites are not similar. We propose that the translation start site of the *cobT* gene identified by Roth et al. (52) is in error. The new translation start site for the *S. typhimurium cobT* gene has a potential Shine-Dalgarno ribosome binding site showing a stronger match to the consensus sequence than that of the previously identified translation start site. These corresponding start sites yield 11- and 26-base spacers upstream of the *E. coli* and *S. typhimurium cobT* genes, respectively.

The fourth open reading frame lay immediately downstream of the *E. coli cobT* gene and was homologous to the open reading frame located downstream of the *S. typhimurium cobT* gene (11). In this paper, the *Salmonella* and *E. coli* homologs of this gene will be referred to as the *srfK* and *erfK* genes, respectively. A good consensus ribosome binding sequence is found upstream of the *E. coli erfK* gene. The fifth open reading frame is part of an IS*5* transposon; the 5′ 633 bases of the sequence reported here were identified as IS*5* specific (Fig. 2).

An asparagine-accepting-tRNA gene and its associated promoter were identified downstream of the *erfK* sequence. Three asparagine-accepting-tRNA genes are found in the *E. coli* chromosome (4). The sequence downstream of the *erfK* gene was identified specifically as the *asnV* gene since the flanking DNA sequence, as well as the tRNA sequence itself, was 100% identical over 200 bases to the nucleotide sequence of the *asnV* locus (GenBank accession no. X52792). The location of the *asnV* gene on the *E. coli* genetic map corresponds to that location of the Kohara (34) λ clones from which the *cob* locus was cloned. The restriction map of the *cob* operon region corresponds to the region from kb 2072 to 2076 on the Kohara map. A homologous asparagine-accepting-tRNA gene (100% identity) is found downstream of the *srfK* sequence in *S. typhimurium* (11).

Transcription of the *E. coli cob* operon. The promoters of the *E. coli cob* operon were mapped by the method of Kofoid et al. (33). A promoter was identified at position 918 in the nucleotide sequence reported here (Fig. 3) and is designated P1. This transcript would provide a 109-base leader RNA prior to the translation initiation site of the *cobU* gene. Putative consensus −10 and −35 sites for the σ^{70} RNA polymerase holoenzyme were identified upstream of this transcription start site and match 6 of 12 bases in the combined consensus sequence (40) with a 15-base spacer between the −35 and −10 sites. This spacing and the similarity to the consensus sequence are consistent with a low-level constitutive promoter.

An additional promoter was identified at position 820 (designated P2), providing a 204-base leader RNA prior to the *cobU* translation initiation site. Putative consensus −10 and −35 sites for the σ^{70} RNA polymerase were identified upstream of this site and match 5 of 12 bases in the combined consensus sequence (40), with a 16-base spacer between the −35 and −10 sites. Unlike promoter P1, promoter P2 is induced at least 100-fold by cobinamide (Fig. 3); under our conditions, maximal induction of P2 is approximately equal to the basal level of promoter P1.

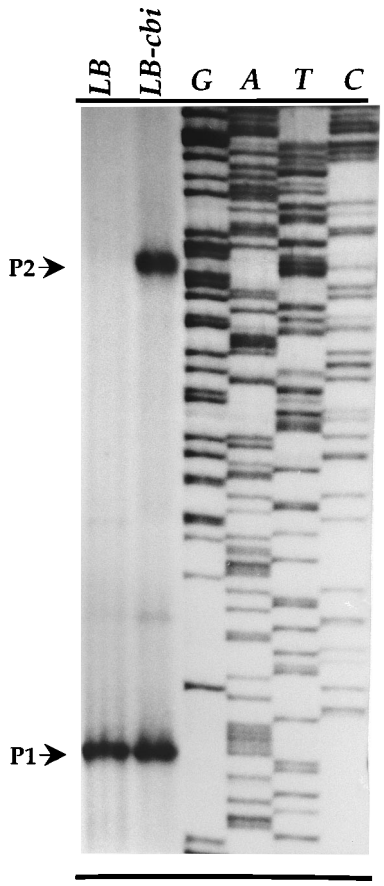

FIG. 3. RNA transcript mapping of *cob* operon promoter. Lanes: LB, cells grown in LB medium; LB-cbi, cells grown in LB medium plus 100 nM dicyanocobinamide; G, A, T, and C, chain-terminating DNA sequencing reactions using the same oligonucleotide primer.

The *cob* genes were also found to be transcribed from a very weak promoter within the upstream IS*5* transposon. Given that IS*5* is known to bear an outbound −35 promoter site (22), this is not surprising. Transcription from this promoter would yield a leader mRNA region of nearly 400 bases. The level of transcription from this promoter amounts to <0.1% of the basal level of promoter P1 (data not shown).

It is likely that the *cobT* gene is the most promoter-distal cobalamin synthetic gene in this cluster. The *srfK* gene of *S. typhimurium* does not participate directly in cobalamin synthesis and is not coregulated with the *cob* operon (11). Despite this, no Rho-independent transcription terminator is evident between these two *Salmonella* genes. In contrast, there is a sequence strongly suggestive of a factor-independent terminator between the *cobT* and *erfK* genes in *E. coli*. The sequence bears an 8-base stem with one mismatch and a 4-base loop. The sequence is preceded by four adenine residues and followed by three uracil residues. These features are common for Rho-independent terminators. If transcription termination occurs between these two genes in both organisms, it suggests that the *S. typhimurium* region includes a terminator not recognizable from inspection of the DNA sequence alone. In addition, good consensus matches to σ^{70} −35 and −10 sites were identified upstream of the *erfK* gene.

TABLE 2. Phenotypes of *E. coli* Cob$^-$ insertion mutations

Strain	Genotype	Aerobic growth with[a]:												Anaerobic growth with[a]:					
		CN-B_{12}			Ado-B_{12}			Cbi			Cbi + DMB[b]			Cbi			Cbi + DMB[b]		
		0	0.13	0.5	0	0.13	0.5	0	4	8,192	0	4	8,192	0	4	8,192	0	4	8,192
TR7177	Wild type	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
TT18496	Δ*metE571*	−	+	+	−	+	+	−	+	+	−	+	+	−	+	+	−	+	+
TT18500	Δ*metE571 cobT14*	−	+	+	−	+	+	−	−	−	−	+	+	−	−	−	−	+	+
TT18503	Δ*metE571 cobS6*	−	+	+	−	+	+	−	−	−	−	−	−	−	−	−	−	−	−
TT18506	Δ*metE571 btuR510*	−	−	+	−	+	+	−	−	−	−	−	−	−	+	+	−	+	+

[a] Compounds (CN-B_{12}, cyanocobalamin; Ado-B_{12}, adenosylcobalamin; and Cbi, dicyanocobinamide) were added to the medium at the indicated nanomolar concentrations. Strains were incubated at 37°C under aerobic or anaerobic conditions as indicated.
[b] DMB was added to a final concentration of 50 μM.

Without a terminator, transcription of the *erfK* gene would be likely to interfere strongly with converging transcription from the downstream *asnV* gene (Fig. 2). There is a strong candidate for a Rho-independent terminator in the short (57-bp) sequence between these two genes. The sequence bears a 7-base, GC-rich stem with no mismatches and a 4-base loop. The sequence is preceded by five adenine residues and followed by six uracil residues. It is likely that this sequence functions bidirectionally as a terminator for transcripts originating from both the *erfK* and the *asnV* promoters.

Distribution of the *cob* promoter-proximal IS*5* among ECOR strains. The presence of an IS*5* transposon upstream of the *cob* operon raises an interesting question: is this transposon responsible for the loss of CobI gene function in *E. coli*? This appears unlikely, as the partial nucleotide sequence of this IS*5* is nearly identical to IS*5* sequences found in GenBank. Therefore, it is likely that the element has recently inserted upstream of the *cob* operon in *E. coli* K-12. A prediction of this hypothesis is that no other strain of *E. coli* will carry an IS*5* in this location. To test the hypothesis, chromosomal DNA was prepared from natural isolates of *E. coli* from the ECOR collection (46), including 26 isolates harboring at least one IS*5* element (57) and 5 isolates lacking sequences homologous to IS*5*.

By using oligonucleotide primers specific to the *cobU* gene and to IS*5*, the PCR was employed to amplify an 800-bp fragment including the junction of IS*5* and the *cob* operon. While the sequences specific to either the *cob* operon or IS*5* were amplified in IS*5*-bearing strains of *E. coli*, the junction between IS*5* and the *cob* operon was detected only in *E. coli* K-12. Therefore, we conclude that the IS*5* transposon upstream of the *cob* operon is unique to *E. coli* K-12 and played no role in the evolution of the *cob* operon in *E. coli.*

Localization of mutations within the *E. coli cob* operon. Insertion mutagenesis of *E. coli* was performed by introducing the Tn*10d*Tc minitransposon from bacteriophage λ vectors as described in Materials and Methods. Tetracycline-resistant mutant strains were screened for their inability to convert cobinamide into cobalamin. Ten such mutants whose tetracycline resistance determinants were 100% linked to the Cob$^-$ phenotype were isolated. Five of these strains could convert cobinamide into cobalamin when DMB was provided (Table 2). It was inferred that these insertion mutants lacked CobII gene functions. PCR amplification of portions of the *E. coli cob* operon was attempted to determine if the Tn*10d*Tc insertions mapped within that region. All five CobII$^-$ insertions mapped within the *cobT* gene. The sites of insertion were determined for three CobII$^-$ insertions (Table 3). The DNA sequences of the insertion sites revealed that all three Tn*10d*Tc elements had inserted 647 bases downstream of the *cobT* translation start site, approximately in the middle of the *cobT* gene. Two of the insertions (*cobT14*::Tn*10d*Tc and *cobT15*::Tn*10d*Tc) are oriented with the *tetA* gene transcription downstream in the *cob* operon; the other insertion (*cobT16*::Tn*10d*Tc) is oriented in the opposite direction. The target sequence 5′TGTTGTG CG3′ was duplicated at the insertion site, with the canonical Tn*10d*Tc sequence lying between the two repeats. These data show that the *cobT* gene is required for CobII functions in *E. coli* but is not essential for CobIII functions.

The remaining five Cob$^-$ insertion mutants lacked CobIII functions, since the addition of both cobinamide and DMB did not allow cobalamin production under aerobic growth conditions (Table 2). The PCR was employed as described above to determine if these insertions mapped within the *cob* operon. Three of these insertions mapped within the *cobS* gene, 311, 551, and 556 bp from the translation start codon (Table 3). Since insertions in the *cobT* gene yielded CobII$^-$ phenotypes, the CobIII$^-$ phenotypes of *cobS* insertions could not result from polarity on the *cobT* gene. Rather, the *cobS* gene is inferred to provide CobIII functions. The homologous genes in *S. typhimurium* (*cobS*) and *P. denitrificans* (*cobU*) also provide CobIII functions.

Two of the *E. coli* CobIII$^-$ insertions were shown to map outside the *cob* operon by the PCR screen described above. Since previous studies have shown that cobalamin and cobinamide are both transported via the BtuB protein (9, 31), it is unlikely that these mutations lay within the *btuB* gene. Other possible locations for these mutations included the gene for a transport function specific for cobinamide, since cobalamin was transported efficiently, or the *btuR* gene. These insertions

TABLE 3. Locations of *E. coli* Cob$^-$ Tn*10d*Tc insertion mutations

Strain or sequence	Allele	Target duplication[a]	
		Position	Sequence
TT18500	*cobT14*	655–647	CGCACAACA
TT18501	*cobT15*	655–647	CGCACAACA
TT18502	*cobT16*	647–655	TGTTGTGCG
TT18503	*cobS6*	559–551	AGCCGAGTG
TT18504	*cobS7*	564–556	GGCTAAGCC
TT18505	*cobS12*	319–311	TATCACGCA
TT18506	*btuR510*	276–268	GGTAAAGCC
TT18507	*btuR511*	209–201	GGCCAGGTG
Consensus			-RYY-RRY-

[a] Numbers indicate the location of the target duplication in the reported sequence. The *btuR* coordinates are numbered from the translation initiation codon. The *tetR* gene is proximal to the first listed base, and the *tetA* gene is proximal to the second listed base.

TABLE 4. Comparison of cobalamin synthetic genes from *E. coli* and *S. typhimurium*

Species	Gene	Length (bp)	CAI[a]	GC content[b]					% Identity[c]		Divergence[d]	
				Gene	Position			Δ3				
					1	2	3		Nucleotide	Protein	NS	S
E. coli	*cobU*	543	0.248	50.6	63.5	39.2	50.8					
S. typhimurium	*cobU*	543	0.286	55.1	65.2	43.1	56.9	6.1	74.0	82.3	0.12	2.55
E. coli	*cobS*	744	0.267	53.5	62.5	44.4	53.6					
S. typhimurium	*cobS*	744	0.257	56.7	62.1	46.4	61.7	8.1	74.3	80.6	0.13	1.34
E. coli	*cobT*	1,080	0.239	51.3	62.5	44.4	46.9					
S. typhimurium	*cobT*	1,071	0.260	59.8	66.7	47.9	64.7	17.8	70.3	76.1	0.20	1.68
E. coli	*erfK*	933	0.264	49.6	58.5	45.3	45.0					
S. typhimurium	*srfK*	930	0.276	52.5	61.6	43.9	51.9	6.9	71.6	78.4	0.15	1.70
E. coli	*asnV*	83	—	45.7	—	—	—	—				
S. typhimurium	*asnV*	83	—	45.7	—	—	—	—	100.0	—	—	
E. coli	Average[e]			52	57	44	57					
S. typhimurium	Average[e]			53	58	44	58	2.0	84.4	93.0	0.039	0.94

[a] CAI, codon adaptation index (calculated as described elsewhere [62]).
[b] G+C contents (percentages) are reported for the coding sequence (gene) and for the first, second, and third codon positions. The difference between the third codon position G+C contents of homologous *E. coli* and *S. typhimurium* loci is also reported (Δ3).
[c] Calculated by using the GAP program from the Genetics Computer Group program package, version 8 (15).
[d] Nucleotide divergences at nonsynonymous (NS) and synonymous (S) sites were calculated by using the DIVERGE program of the Genetics Computer Group program package (15). A divergence of 1.0 represents an average of one nucleotide change per codon position; divergence values of >1.0 indicate that nucleotide differences reflect more than one nucleotide change per codon position.
[e] Average divergences are reported for a set of 67 homologous *E. coli* and *S. typhimurium* loci (61). Average nucleotide compositions included sets of 200 *E. coli* and *S. typhimurium* genes (45).

were mapped to the *btuR* gene by the PCR method described above. Both mutations had inserted in the middle of the *btuR* coding sequence, 201 and 268 bp from the translation start codon (Table 3). Like homologous *Salmonella cobA* mutants, these *E. coli btuR* mutants showed no impairment of cobalamin transport under anaerobic growth conditions (Table 2).

The *btuR* gene of *E. coli* and the homologous *cobA* gene of *S. typhimurium* provide functions which add the adenosyl moiety to cobamides to make active cofactors (42). The *cobA* gene is also required for the aerobic conversion of cobinamide to cobalamin (19). It is inferred that exogenous cobinamide must be adenosylated by CobA before the CobIII reactions can occur. Under anaerobic conditions, however, *Salmonella cobA* mutants can convert cobinamide to cobalamin, presumably circumventing or replacing the CobA function. *E. coli btuR* mutants can also convert cobinamide to cobalamin under anaerobic conditions and may circumvent or replace BtuR function in a similar manner.

Regulation of the *E. coli cob* operon. The *cob* operon of *S. typhimurium* is induced by propanediol; induction is mediated by both the CRP-cAMP and the ArcAB complexes and occurs only during growth on poor carbon sources (1, 2, 7, 18, 51). According to current models, the *Salmonella cob* operon is induced for the purpose of using the cofactor for the degradation of propanediol under starvation conditions (1). In addition, both the *btuB* gene of *E. coli* and the *cob* operon of *S. typhimurium* were found to be repressed by the addition of exogenous cobalamin (2, 30, 41, 42, 50). Specifically, both the *btuB* and the *cob* operon transcripts are repressed by adenosylcobalamin. The regulation of the *E. coli cob* operon might be expected to differ substantially since *E. coli* does not metabolize propanediol in a cobalamin-dependent fashion (38), and expression of the *E. coli cob* genes would be useless without cobinamide.

To determine the relative amounts of transcription provided by the two *cob* promoters (Fig. 2), cells were grown on LB medium (poor carbon source), LB medium plus propanediol (inducing conditions for the *S. typhimurium cob* operon), LB medium plus glucose (noninducing conditions for the *S. typhimurium cob* operon), and LB medium plus cobalamin (repressing conditions for the *S. typhimurium btuB* and *cob* operons); a parallel set of media also contained cobinamide. Similar amounts of mRNA originated from the P1 promoter under all growth conditions. The P2 promoter made very little product unless cobinamide was present. In all media containing cobinamide, the ratio of the two transcripts remained constant. These data show that transcription from the *cob* P1 and P2 promoters is unaffected by global growth conditions known to induce and repress the *S. typhimurium cob* operon. Instead, the *E. coli cob* operon is induced by cobinamide, its presumed substrate. We conclude that the regulation of the *E. coli* cobalamin synthetic genes differs substantially from that of the *S. typhimurium* homologs.

DISCUSSION

Comparison of *E. coli* and *Salmonella* homologs. The three *E. coli cob* genes show between 69 and 72% nucleotide identity with the *S. typhimurium cobU*, *cobS*, and *cobT* genes (Table 4). The predicted proteins are between 76 and 82% identical. Thus, the *E. coli cob* genes are likely to encode cobalamin synthetic proteins. Typical pairs of homologous genes in *E. coli* and *S. typhimurium* show much higher nucleotide identities (84 to 95%) and encode proteins with more than 91% amino acid identity (61) (Table 4). As shown in Fig. 4, the *E. coli btuR* gene hybridizes with homologs in the *Salmonella* and *Klebsiella* genomes. This behavior is typical of genes shared among these taxa. In contrast, under the same hybridization conditions, the

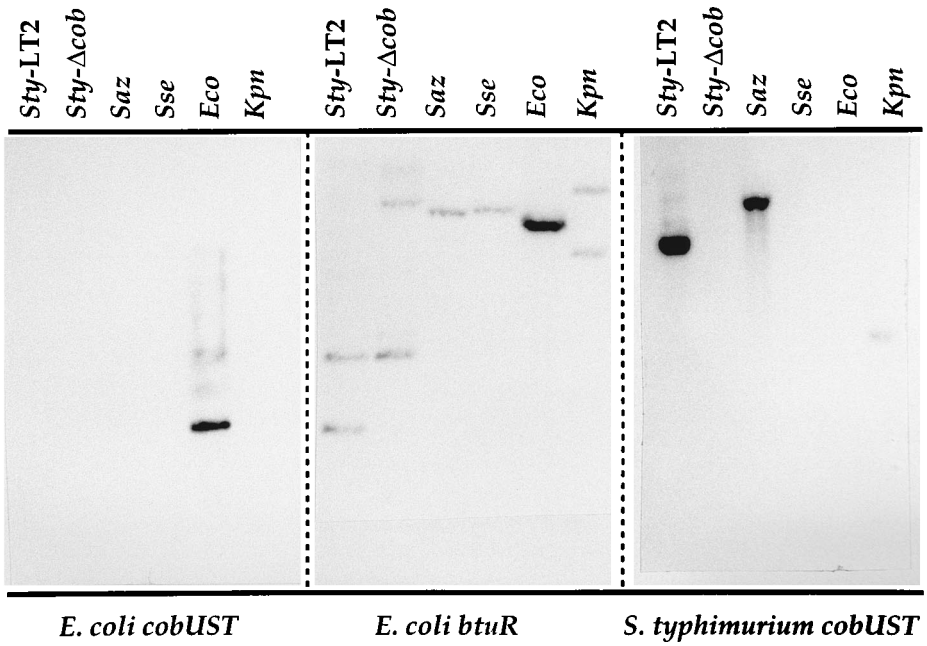

FIG. 4. DNA hybridization using probes from the *E. coli cobU*, *cobS*, and *cobT* genes, the *E. coli btuR* gene, and the *S. typhimurium cobU*, *cobS*, and *cobT* genes. Each lane contains total genomic DNA. Lanes: *Sty*-LT2, *S. typhimurium* LT2; *Sty*-Δ*cob*, *S. typhimurium* TT11855, which lacks the entire *cob* operon; *Saz*, *Salmonella arizonae* TR6331, a Cob^+ natural isolate; *Sse*, *Salmonella seminole* TR6332, a Cob^- natural isolate; *Eco*, *E. coli* TR7177; *Kpn*, *K. pneumoniae* TR7176. The DNAs in lanes *Sty*-Δ*cob* were incompletely digested; in complete digestions, lanes *Sty*-LT2 and *Sty*-Δ*cob* show identical hybridization patterns for the *btuR* probe.

S. typhimurium cobUST genes detect no homologs outside *Salmonella* genomes; the *E. coli cobUST* genes also do not hybridize with the *Salmonella* or *Klebsiella* genomes.

As expected, the *E. coli cobUST* and *S. typhimurium cobUST* genes are significantly more different from each other than typical gene pairs. We have previously presented evidence that the *cob* operon of *S. typhimurium* was acquired by horizontal transfer (38). This conclusion was based on unusual cobalamin metabolic phenotypes, the atypical G+C content, dinucleotide frequencies, and codon usage bias of the *cob* genes, as well as the lack of DNA hybridization of *cob* genes to the genomes of cobalamin-producing enteric bacteria (38). Analysis of the *E. coli cob* operon shows that its sequence features do not differ significantly from those of typical *E. coli* chromosomal genes with respect to G+C content, dinucleotide frequencies, or codon usage bias (Table 4). Hence, the unusual features of the *Salmonella cob* genes are not general properties of cobalamin synthetic genes. These data support the notion that the *cob* genes are not native to the *Salmonella* chromosome. The *cps* genes of *E. coli* are proposed to have been introduced by lateral transfer; this conclusion was based, in part, on the unusual nucleotide composition of the *E. coli* and *Salmonella enterica cps* homologs (3).

Alternatively, the large differences between the *S. typhimurium cob* genes and the *E. coli cob* genes may be the result of a relaxed selection for function. Genes under strong selection show few differences between *E. coli* and *S. typhimurium*; genes with less critical functions show more differences (61). Therefore, the large difference between the *E. coli* and *S. typhimurium cob* genes may reflect very weak selection for cobalamin synthesis. The nucleotide sequence of the *E. coli cob* operon allows the horizontal transfer model to be tested directly.

The *E. coli* and *S. typhimurium* CobUST proteins show significant differences, being only 76 to 82% identical (Table 4). If weak selection for protein function allowed these differences to accumulate, one might expect substantially different amino acids among the resultant substitutions. Substitutions involving only very similar amino acids are indicative of strong selection for protein function. The substitutions among the *cob* genes do not appear to have resulted from the indiscriminant accumulation of nucleotide changes. For example, of the 22 amino acid differences between the *E. coli* and *S. typhimurium* CobS proteins, 8 are highly conservative changes (e.g., a V-to-L mutation at position 22 [V22L]), and 13 differences involve similar amino acids (e.g., T53S). Only one difference involves residues that are physically and/or chemically distinct (R46A). This pattern of changes is inconsistent with a rapid and indiscriminant accumulation of nucleotide substitutions. Rather, it reflects an accumulation of substitutions that do not substantially alter the activity of the encoded peptide.

Inspection of Table 2 shows that the *E. coli* and *S. typhimurium cob* genes also show excesses of synonymous substitutions, that is, nucleotide changes that do not alter the encoded protein. Synonymous substitution rates approach 1% per million years (35, 47). Lower rates are observed among genes showing strong codon usage bias due to strong selection (62); however, higher rates are not observed among genes under weak selection. The numbers of synonymous substitutions among the three cobalamin synthetic genes are, on average, more than twice that observed for typical chromosomal genes. This excess of synonymous substitutions shows that the *cob* genes must have diverged prior to the divergence of typical *E. coli* and *S. typhimurium* genes.

The nucleotide differences between the *E. coli* and *S. typhimurium cob* operons bear a strong bias in G+C content (Table 4). The *S. typhimurium* genes are significantly more GC rich than the corresponding *E. coli* genes. Moreover, this difference is greatest in the third, typically synonymous codon position. If the differences between these operons resulted from relaxed selection for function, a stipulation that the substitutions have actively biased the genic nucleotide content must be included.

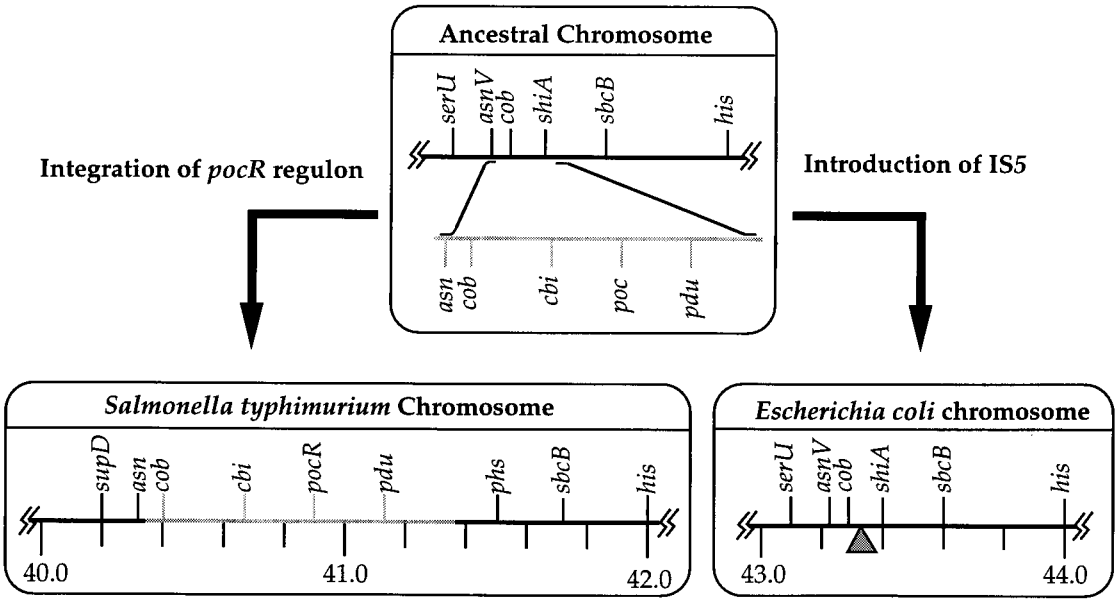

FIG. 5. Model for the introduction of the *cob* operon into the *Salmonella* chromosome. The genetic maps of corresponding regions of the specified chromosomes (*E. coli* after Bachmann [4]; *S. typhimurium* after Sanderson and Roth [56] except for data of Bensen and Roth [5] for the *sbcB* locus) (black lines), the introduced DNA fragment (grey line), and insertion sequence IS*5* (triangle) are shown. The *E. coli serU* gene is homologous to the *S. typhimurium supD* gene, which is designated *serU* in this figure.

More importantly, this bias has avoided nonsynonymous sites and has acted most rapidly on the synonymous codon positions. This stipulation is not consistent with any known biological process.

The *E. coli cob* operon is located at min 43.5 on the genetic map. The *E. coli cob* genes were cloned from bacteriophage λ clones corresponding to the chromosomal region counterclockwise of the *his* operon and clockwise of the *asnV* asparagine-accepting-tRNA gene on the *E. coli* genetic map (4). This location corresponds well with the location of the *cob* operon, which is 1 min counterclockwise of the *his* operon at min 41 on the *S. typhimurium* genetic map (29, 56) and is also adjacent to a $tRNA^{Asn}$ locus (11). Superficially, this may seem surprising in the face of the supposition that the *cob* operon was introduced into the *Salmonella* genome by horizontal transfer. Horizontally transferred genes might have been expected to be at an unrelated location.

However, legitimate recombination events involving somewhat divergent sequences are more frequent than illegitimate integration events. We propose that the *cob* operon was introduced into the *Salmonella* chromosome by legitimate homologous recombination, using the *asnV* gene as a region of identity to allow one crossover (Fig. 5). As noted above, the promoter-distal ends of the two *cob* operons are adjacent to asparagine-accepting-tRNA genes, bearing 100% nucleotide identity. This conclusion makes a strong prediction: we expect that pairs of homologous genes located on the other side of *asnV* in the *E. coli* and *S. typhimurium* chromosomes will be more similar to each other than are the homologous *cob* genes.

If the *asnV* locus delimits the counterclockwise endpoint of the recombined region, the clockwise endpoint must lie between the *cob* and *his* operons in both taxa (Fig. 5). It is noteworthy that the *phs* operon and the cobalamin-dependent *pdu* operon are located in this region of the *Salmonella* chromosome; these functions are both absent from *E. coli*. The nucleotide sequences of the *phs* genes (25), however, show no evidence of atypical composition. The *E. coli shiA* (shikimate permease) gene is located in this region and is not found in *S. typhimurium*. Our model (Fig. 5) supposes that the *pdu* and *cob* operons were introduced into the ancestral *Salmonella* chromosome by a single event, eliminating the *shiA* locus and replacing the ancestral *cobUST* genes with foreign homologs.

A prediction of this model is that the *E. coli cob* genes may have homologs in the *Klebsiella* genome, for which no horizontal transfer event is postulated. It is puzzling, therefore, that the *E. coli cob* genes did not detect homologs in the *Klebsiella* genome, while the *btuR* gene hybridized well (Fig. 4). Resolution of this question awaits the characterization of the *Klebsiella* CobII and CobIII genes.

The *E. coli cob* operon shows no homology with *S. typhimurium* CobI genes. While the two CobIII genes and one CobII gene of *E. coli* were easily identified in the cloned DNA fragment, no sequence homologous to *S. typhimurium* CobI genes was detected. In *S. typhimurium*, the *cbiP* gene, encoding an amidase responsible for cobyric acid production, is located immediately upstream of the *cobU* gene; in fact, the coding sequences overlap (Fig. 2) (52). In *E. coli*, there is no sequence similar to the *cbiP* gene upstream of the *cobU* homolog. Rather, there appears to be a constitutive promoter located 100 bp upstream of the *cobU* start codon (Fig. 3); in *S. typhimurium*, no promoter sequence is observed in this location. In addition, an IS*5* element was found 400 bp upstream of the *cobU* gene. These data do not support the hypothesis that cryptic CobI genes are present in *E. coli* at this location, and these functions have not been detected in the laboratory. If cryptic CobI functions are present in the *E. coli* genome, they are encoded elsewhere in the chromosome.

It is unlikely that the insertion of IS*5* upstream of the *E. coli cobU* gene is responsible for the lack of CobI function among *E. coli* strains. Since no *E. coli* strain has been found to encode CobI gene functions, i.e., de novo cobalamin synthesis (38), it is likely that the loss of these functions predates the divergence of tested *E. coli* strains and the appearance of the IS*5* insertion. If the IS*5* insertion were responsible for this loss of function, its

DNA sequence would bear evidence of its age in the accumulation of nucleotide substitutions. However, the nucleotide sequence of the IS*5* element upstream of the *cobU* gene is nearly identical to the sequences of other IS*5* elements found in the GenBank database. In addition, no other strain of *E. coli* bears an IS*5* insertion at this location. These data are consistent with a recent insertion of the IS*5* transposon in this strain. Current models of transposon population dynamics are inconsistent with long-term stability of individual transposon insertions in prokaryotes (24, 37, 44).

The *E. coli cobT* gene provides CobII function. Insertion mutations in the *E. coli cobT* gene yielded a CobII$^-$ phenotype (Table 2); mutations in the homologous *S. typhimurium cobT* gene exhibit a similar phenotype (11). This phenotype is somewhat surprising. The *E. coli* CobT protein is homologous to the *S. typhimurium* CobT and *P. denitrificans* CobU proteins, which both perform the nicotinate mononucleotide::DMB phosphoribosyltransferase reaction (10, 64). Loss of this function should result in a CobIII phenotype, since DMB and cobinamide would not correct the defect. Yet in both *E. coli* and *S. typhimurium*, lesions in this gene provide CobII$^-$ phenotypes (Fig. 2).

Several explanation have been proffered (11, 17, 48, 49). In any model, the phosphoribosyltransferase reaction required for CobIII function must be performed by some other protein(s). Escalante-Semerena and coworkers have posited that the enzymes for this alternate pathway for phosphoribosyl transfer are encoded by the *Salmonella cobB* and/or *cobC* gene (Fig. 1) (48, 49, 64); this explanation is supported by the findings of Chen et al. (12). Since *E. coli cobT* mutations also fail to yield CobIII$^-$ phenotypes, *E. coli* may harbor homologs to the *Salmonella cobB* and *cobC* genes. The question why *E. coli cobT* mutations yield CobII phenotypes remains to be answered. One possibility is that internal DMB synthesis provides insufficient DMB for the CobB and CobC proteins, which are presumably not acting on their preferred substrates (49, 64). More likely, the *E. coli* CobT protein participates directly in the synthesis of DMB, as the data of Chen et al. (11) suggest.

ACKNOWLEDGMENTS

This work was supported by grants GM 15868 (J.G.L.) and GM 34804 (J.R.R.) from the National Institutes of Health.

We thank M. Ailion, N. Bensen, T. Bobik, P. Chen, N. Clark, D. Dykhuizen, T. Galitski, K. Haack, P. Higgins, E. Kofoid, L. Miesel, R. Milkman, H. Ochman, E. Presley, and D. Walter for helpful and enlightening discussions and P. Higgins, H. Ochman, G. Roberts, and G. Stauffer for providing bacterial strains and bacteriophages. We are grateful to Carleen Collins for providing unpublished data and preliminary sequences of portions of the *cobS* and *cobT* genes.

REFERENCES

1. **Ailion, M., T. A. Bobik, and J. R. Roth.** 1993. Two global regulatory systems (Crp and Arc) control the cobalamin/propanediol regulon of *Salmonella typhimurium*. J. Bacteriol. **175:**7200–7208.
2. **Andersson, D. I., and J. R. Roth.** 1989. Mutations affecting regulation of cobinamide synthesis in *Salmonella typhimurium*. J. Bacteriol. **171:**6726–6733.
3. **Aoyama, K., A. M. Haase, and P. R. Reeves.** 1994. Evidence for effect of random genetic drift on G+C content after lateral transfer of fucose pathway genes to *Escherichia coli* K-12. Mol. Biol. Evol. **11:**829–838.
4. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
5. **Benson, N., and J. R. Roth.** 1994. Suppressors of *recB* mutations in *Salmonella typhimurium*. Genetics **138:**11–29.
6. **Blanche, F., L. Debussche, A. Famechon, D. Thibaut, B. Cameron, and J. Crouzet.** 1991. A bifunctional protein from *Pseudomonas denitrificans* carries cobinamide kinase and cobinamide phosphate guanylyltransferase activities. J. Bacteriol. **173:**6052–6057.
7. **Bobik, T. A., M. Ailion, and J. R. Roth.** 1992. A single regulatory gene integrates control of vitamin B_{12} synthesis and propanediol degradation. J. Bacteriol. **174:**2253–2266.
8. **Bochner, B. R., H.-C. Huang, G. L. Schieven, and B. N. Ames.** 1980. Positive selection for loss of tetracycline resistance. J. Bacteriol. **143:**926–933.
9. **Bradbeer, C., J. S. Kenley, D. R. Di Masi, and M. Leighton.** 1978. Transport of vitamin B_{12} in *Escherichia coli*. Corrinoid specificities of the periplasmic B_{12}-binding protein and of energy-dependent B_{12} transport. J. Biol. Chem. **253:**1347–1352.
10. **Cameron, B., F. Blanche, M.-C. Rouyez, D. Bisch, A. Famechon, M. Couder, L. Cauchois, D. Thibaut, L. Debussche, and J. Crouzet.** 1991. Genetic analysis, nucleotide sequence, and products of two *Pseudomonas denitrificans cob* genes encoding nicotinate mononucleotide: dimethylbenzimidazole phosphoribosyltransferase and cobalamin (5′-phosphate) synthase. J. Bacteriol. **173:**6066–6073.
11. **Chen, P., M. Ailion, N. Weyland, and J. R. Roth.** 1995. The end of the *cob* operon: evidence that the last gene (*cobT*) catalyzes the synthesis of the lower ligand of vitamin B_{12}, dimethylbenzimidazole. J. Bacteriol. **177:**1461–1469.
12. **Chen, P., D. I. Andersson, and J. R. Roth.** 1994. The control region of the *cob*/*pdu* regulon in *Salmonella typhimurium*. J. Bacteriol. **176:**5474–5482.
13. **Crouzet, J., S. Levy-Schil, B. Cameron, L. Cauchois, S. Rigault, M.-C. Rouyez, F. Blanche, L. Debussche, and D. Thibaut.** 1991. Nucleotide sequence and genetic analysis of a 13.1-kilobase-pair *Pseudomonas denitrificans* DNA fragment containing five *cob* genes and identification of structural genes encoding cob(I)alamin adenosyltransferase, cobyric acid synthase, and bifunctional cobinamide kinase-cobinamide phosphate guanylyltransferase. J. Bacteriol. **173:**6074–6087.
14. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. A manual for genetic engineering: advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
16. **Dower, W. J., J. F. Miller, and C. Ragsdale.** 1988. High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. **16:**6127–6145.
17. **Escalante-Semerena, J. C., M. G. Johnson, and J. R. Roth.** 1992. The CobII and CobIII regions of the cobalamin (vitamin B_{12}) biosynthetic operon of *Salmonella typhimurium*. J. Bacteriol. **174:**24–29.
18. **Escalante-Semerena, J. C., and J. R. Roth.** 1987. Regulation of cobalamin biosynthetic operons in *Salmonella typhimurium*. J. Bacteriol. **169:**2251–2258.
19. **Escalante-Semerena, J. C., S.-J. Suh, and J. R. Roth.** 1990. *cobA* function is required for both de novo cobalamin biosynthesis and assimilation of exogenous corrinoids in *Salmonella typhimurium*. J. Bacteriol. **172:**273–280.
20. **Eschenmoser, A.** 1988. Vitamin B_{12}: experiments concerning the origin of its molecular structure. Angew. Chem. Int. Ed. Engl. **27:**5–39.
21. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
22. **Galas, D. J., and M. Chandler.** 1989. Bacterial insertion sequences, p. 109–162. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
23. **Georgopapadakou, N. H., and A. I. Scott.** 1977. On B_{12} biosynthesis and evolution. J. Theor. Biol. **69:**381–384.
24. **Hartl, D. L., E. R. Lozovskaya, and J. G. Lawrence.** 1992. Nonautonomous transposable elements in prokaryotes and eukaryotes. Genetica **86:**47–53.
25. **Heinzinger, N. A., S. Y. Fujimoto, M. A. Clark, M. S. Moreno, and E. L. Barrett.** 1995. Sequence analysis of the *phs* operon in *Salmonella typhimurium* and the contribution of thiosulfate reduction to anaerobic energy metabolism. J. Bacteriol. **177:**2813–2820.
26. **Helms, C., M. Y. Graham, J. E. Dutchik, and M. V. Olson.** 1985. A new method for purifying lambda DNA from phage lysates. DNA **4:**39–49.
27. **Hörig, J. A., and P. Renz.** 1978. [5-^{15}N]riboflavin as precursor in the biosynthesis of the 5,6-dimethylbenzimidazole moiety of vitamin B_{12}. J. Biol. Chem. **253:**7410–7414.
28. **Jeter, R. M., B. M. Olivera, and J. R. Roth.** 1984. *Salmonella typhimurium* synthesizes cobalamin (vitamin B_{12}) de novo under anaerobic growth conditions. J. Bacteriol. **159:**206–213.
29. **Jeter, R. M., and J. R. Roth.** 1987. Cobalamin (vitamin B_{12}) biosynthetic genes of *Salmonella typhimurium*. J. Bacteriol. **169:**3189–3198.
30. **Kadner, R. J.** 1978. Repression of synthesis of the vitamin B_{12} receptor in *Escherichia coli*. J. Bacteriol. **136:**1050–1057.
31. **Kenley, J. S., M. Leighton, and C. Bradbeer.** 1978. Transport of vitamin B_{12} in *Escherichia coli*. Corrinoid specificity of the outer membrane receptor. J. Biol. Chem. **253:**1341–1346.
32. **Kleckner, N., J. Bender, and S. Gottesman.** 1991. Uses of transposons with emphasis on Tn*10*. Methods Enzymol. **204:**139–180.
33. **Kofoid, E., J. G. Lawrence, and J. R. Roth.** 1995. Unpublished data.
34. **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.
35. **Lawrence, J. G., D. L. Hartl, and H. Ochman.** 1991. Molecular considerations in the evolution of bacterial genes. J. Mol. Evol. **33:**241–250.
36. **Lawrence, J. G., D. L. Hartl, and H. Ochman.** 1993. Sequencing products of

the polymerase chain reaction. Methods Enzymol. **218:**26–35.
37. **Lawrence, J. G., H. Ochman, and D. L. Hartl.** 1992. The evolution of insertion sequences within enteric bacteria. Genetics **131:**9–20.
38. **Lawrence, J. G., and J. R. Roth.** Genetics, in press.
39. **Lingens, B., T. A. Schild, B. Vogler, and P. Renz.** 1992. Biosynthesis of vitamin B_{12}: transformation of riboflavin ^{2}H-labeled in the 1′*R* position or 1′*S* position into 5,6-dimethylbenzimidazole. Eur. J. Biochem. **207:**981–985.
40. **Lisser, S., and H. Margalit.** 1993. Compilation of *E. coli* mRNA promoter sequences. Nucleic Acids Res. **21:**1507–1516.
41. **Lundrigan, M. D., L. C. DeVeaux, B. J. Mann, and R. J. Kadner.** 1987. Separate regulatory systems for the repression of *metE* and *btuB* by vitamin B_{12} in *Escherichia coli*. Mol. Gen. Genet. **206:**401–407.
42. **Lundrigan, M. D., and R. J. Kadner.** 1989. Altered cobalamin metabolism in *Escherichia coli btuR* mutants affects *btuB* gene regulation. J. Bacteriol. **171:**154–161.
43. **Muto, A., and S. Osawa.** 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. USA **84:**166–169.
44. **Naas, T., M. Blot, W. M. Fitch, and W. Arber.** 1994. Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. Genetics **136:**721–730.
45. **Ochman, H., and J. G. Lawrence.** The amelioration of bacterial genomes. *In Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, 2nd ed., in press. American Society for Microbiology, Washington, D.C.
46. **Ochman, H., and R. K. Selander.** 1984. A reference collection of natural isolates of *Escherichia coli*. J. Bacteriol. **157:**517–524.
47. **Ochman, H., and A. C. Wilson.** 1988. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26:**74–86.
48. **O'Toole, G. A., M. R. Rondon, and J. C. Escalante-Semerena.** 1994. Analysis of mutants of *Salmonella typhimurium* defective in the synthesis of the nucleotide loop of cobalamin. J. Bacteriol. **175:**3317–3326.
49. **O'Toole, G. A., J. R. Trzebiatowski, and J. C. Escalante-Semerena.** 1994. The *cobC* gene of *Salmonella typhimurium* codes for a novel phosphatase involved in the assembly of the nucleotide loop of cobalamin. J. Biol. Chem. **269:**26503–26511.
50. **Richter-Dahlfors, and D. I. Andersson.** 1992. Cobalamin (vitamin B12) repression of the *cob* operon in *Salmonella typhimurium* requires sequences within the leader and the first translated open reading frame. Mol. Microbiol. **6:**743–749.
51. **Rondon, M. R., and J. C. Escalante-Semerena.** 1992. The *poc* locus is required for 1,2-propanediol-dependent transcription of the cobalamin biosynthetic (*cob*) and propanediol utilization (*pdu*) genes of *Salmonella typhimurium*. J. Bacteriol. **174:**2267–2272.
52. **Roth, J. R., J. G. Lawrence, M. Rubenfield, S. Kieffer-Higgins, and G. M. Church.** 1993. Characterization of the cobalamin (vitamin B_{12}) biosynthetic genes of *Salmonella typhimurium*. J. Bacteriol. **175:**3303–3316.
53. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:**487–491.
54. **Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. A. Arnheim.** 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230:**1350–1354.
55. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
56. **Sanderson, K. E., and J. R. Roth.** 1988. Linkage map of *Salmonella typhimurium*: edition VII. Microbiol. Rev. **52:**485–532.
57. **Sawyer, S., D. Dykhuizen, R. F. DuBose, L. Green, T. Mutangadura-Mhlanga, D. F. Wolczyk, and D. L. Hartl.** 1987. Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. Genetics **115:**51–63.
58. **Schneider, Z., and A. Stroinski.** 1987. Methylcobamide-dependent reactions, p. 259–266. *In* Z. Schneider and A. Stroinski (ed.), Comprehensive B_{12}. de Gruyter, Berlin.
59. **Scott, A. I.** 1990. Mechanistic and evolutionary aspects of vitamin B_{12} biosynthesis. Acc. Chem. Res. **23:**308–317.
60. **Scott, A. I.** 1993. How nature synthesizes vitamin B_{12}—a survey of the last four billion years. Angew. Chem. Int. Ed. Engl. **32:**1223–1243.
61. **Sharp, P. M.** 1991. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*. J. Mol. Evol. **33:**23–33.
62. **Sharp, P. M., and W.-H. Li.** 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15:**1281–1295.
63. **Stroinski, A.** 1987. Adenosylcobamide-dependent reactions, p. 226–259. *In* Z. Schneider and A. Stroinski ed., Comprehensive B_{12}. de Gruyter, Berlin.
64. **Trzebiatowski, J. R., G. A. O'Toole, and J. C. Escalante-Semerena.** 1994. The *cobT* gene of *Salmonella typhimurium* encodes the NaMN:5,6-dimethylbenzimidazole phosphoribosyl-transferase responsible for the synthesis of N^1-(5-phospho-α-D-ribosyl)-5,6-dimethyl-benzimidazole, an intermediate in the synthesis of the nucleotide loop of cobalamin. J. Bacteriol. **176:**3568–3575.
65. **Vogel, H. J., and D. M. Bonner.** 1956. Acetylornithase of *Escherichia coli*: partial purification and some properties. J. Biol. Chem. **218:**97–106.
66. **Volcani, B. E., J. I. Toohey, and H. A. Barker.** 1961. Detection of cobamide coenzymes in microorganisms by the ionophoretic bioautographic method. Arch. Biochem. Biophys. **92:**381–391.