JOURNAL OF BACTERIOLOGY, Sept. 1994, p. 5474–5482
0021-9193/94/$04.00+0

Vol. 176, No. 17

The Control Region of the *pdu/cob* Regulon in *Salmonella typhimurium*

PING CHEN,[1] DAN I. ANDERSSON,[2] AND JOHN R. ROTH[1]*

Department of Biology, University of Utah, Salt Lake City, Utah 84112,[1] and Department of Microbiology, Biomedical Center, Uppsala University, Uppsala S-75123, Sweden[2]

Received 14 April 1994/Accepted 28 June 1994

The *pdu* operon encodes proteins for the catabolism of 1,2-propanediol; the nearby *cob* operon encodes enzymes for the biosynthesis of adenosyl-cobalamin (vitamin B_{12}), a cofactor required for the use of propanediol. These operons are transcribed divergently from distinct promoters separated by several kilobases. The regulation of the two operons is tightly integrated in that both require the positive activator protein PocR and both are subject to global control by the Crp and ArcA proteins. We have determined the DNA nucleotide sequences of the promoter-proximal portion of the *pdu* operon and the region between the *pdu* and *cob* operons. Four open reading frames have been identified, *pduB*, *pduA*, *pduF*, and *pocR*. The *pduA* and *pduB* genes are the first two genes of the *pdu* operon (transcribed clockwise). The *pduA* gene encodes a hydrophobic protein with 56% amino acid identity to a 10.9-kDa protein which serves as a component of the carboxysomes of several photosynthetic bacteria. The *pduF* gene encodes a hydrophobic protein with a strong similarity to the GlpF protein of *Escherichia coli*, which facilitates the diffusion of glycerol. The N-terminal end of the PduF protein includes a motif for a membrane lipoprotein-lipid attachment site as well as a motif characteristic of the MIP (major intrinsic protein) family of transmembrane channel proteins. We presume that the PduF protein facilitates the diffusion of propanediol. The *pocR* gene encodes the positive regulatory protein of the *cob* and *pdu* operons and shares the helix-turn-helix DNA binding motif of the AraC family of regulatory proteins. The mutations *cobR4* and *cobR58* cause constitutive, *pocR*-independent expression of the *cob* operon under both aerobic and anaerobic conditions. Evidence that each mutation is a deletion creating a new promoter near the normal promoter site of the *cob* operon is presented.

Under anaerobic conditions, the bacterium *Salmonella typhimurium* is capable of full synthesis of the cofactor adenosylcobalamin (vitamin B_{12}) (19). Genetic studies have shown that most of the B_{12} synthetic genes are located in the *cob* operon at minute 41 of the genetic map (13, 20). The phenotypic defect of these mutants can be corrected by supplying exogenous cyanocobalamin (CN-B_{12}). Most of the *cob* operon has been cloned, and its nucleotide sequence has been determined. It has been possible to assign functions to many of the *cob* proteins of *S. typhimurium* by comparing their inferred amino acid sequences to those of cobalamin synthetic genes in *Pseudomonas denitrificans* (34).

Only four B_{12}-dependent enzymes in *S. typhimurium* have been identified. Homocysteine methyl transferase (*metH*) transfers a methyl group from *N*-5-methyltetrahydrofolate to convert L-homocysteine to L-methionine (36). Ethanolamine ammonia-lyase cleaves ethanolamine to acetaldehyde and ammonia (32, 33). Queuosine synthetase reduces epoxyqueuosine to queuosine, a modified base found in four tRNAs (17). Propanediol dehydratase rearranges 1,2-propanediol to yield propionaldehyde and H_2O (18, 38).

Enzymes for the use of propanediol are encoded by the *pdu* operon, which is located adjacent to the *cob* operon at minute 41 (19, 20). The *pdu* and *cob* operons are divergently transcribed and map together at minute 41 of the *Salmonella* chromosome. Recent work shows that both operons are induced by propanediol under the same conditions and that this induction is mediated by a single positive regulatory protein, PocR (7, 31). Furthermore, both the *pdu* and *cob* operons and the gene for the positive activator (*pocR*) are subject to positive control by the global regulatory proteins Crp and ArcA (1, 3). The regulatory integration of the *pdu* and *cob* operons suggests that the most important use of B_{12} in *S. typhimurium* is the breakdown of 1,2-propanediol. The produced propionaldehyde can be used as a hydrogen sink when it is reduced to propanol and excreted (27). Propionaldehyde can also provide a carbon and energy source aerobically when it is converted to propionyl-coenzyme A and ultimately to pyruvate (16, 17a, 18).

The nucleotide sequence of the region between the divergent *pdu* and *cob* operons is reported here. These data complete the sequence of the regulatory gene *pocR*, provide the sequence of two *pdu* genes, and reveal a new gene located between the operons, *pduF*, which we infer encodes a transporter of propanediol. We also report that two *cob* regulatory mutations (*cobR4* and *cobR58*) are deletions that extend from the normal *cob* promoter into the region between the *pdu* and *cob* operons.

MATERIALS AND METHODS

Bacterial strains and plasmids. The bacterial strains used in this study (Table 1) are all derived from *S. typhimurium* LT2 except for strains MC1061, TT17507, and TT17509, which are derived from *Escherichia coli* K-12. Plasmid pDA2979 is a pUC19 plasmid vector carrying a 3.7-kb *Eco*RI DNA fragment derived from the *Salmonella* chromosome. The insert includes the control region of the *pdu/cob* regulon. Plasmid pPC3 is a pUC118 vector carrying a 1.4-kb *Eco*47III subfragment from the middle of the original 3.7-kb *S. typhimurium* fragment.

Growth media and β-galactosidase assay. Rich medium was Difco nutrient broth with NaCl added (15 g/liter). Minimal medium was the E medium of Vogel and Bonner (41) with glucose (0.2%) as the carbon source. For β-galactosidase

* Corresponding author.

TABLE 1. List of strains used in this study

Strain	Genotype
TT17508	*metE205 ara-9 cob-24*::Mud*J cobR58*
TT14679	*metE205 ara-9 cob-24*::Mud*J cobR4*
TT10852	*metE205 ara-9 cob-24*::Mud*J*
TT17074	*del1077* (*metE*) *zeb-3718*::Tn*10 d*Tc
TT17075	*del1077* (*metE*) *zeb-3719*::Tn*10 d*Tc
TT17569	*metE205 ara-9 pocR15*::Tn*10 d*Cm
TT14383	*metE205 ara-9 cob-24*::Mud*Q*
DA1307	*metE205 ara-9 zeb-1845*::Tn*10 cobR58 cob-345*::Mud*Q*
DA4095	*metE205 ara-9 zeb-1845*::Tn*10 cobR4 cob-345*::Mud*Q*
MC1061	*araD139* Δ(*ara-leu*)*7697* Δ*lacX74 galU galK rpsL hsr hbm*$^+$
TT17507	Same as strain MC1061/pDA2979 (see Fig. 1)
TT17509	Same as strain MC1061/pPC3 (see Fig. 1)

assays, NCE medium (no citrate E medium) was used with various carbon sources at the following concentrations: D-glucose, 0.2%; sodium pyruvate, 0.44%; sodium fumarate, 0.32%; DL-1,2-propanediol, 0.2%. The final concentrations of antibiotics in rich medium were as follows: ampicillin, 50 μg/ml; chloramphenicol, 20 μg/ml; kanamycin, 50 μg/ml; tetracycline, 20 μg/ml. In minimal medium, the final concentrations of antibiotics were as follows: tetracycline, 10 mg/ml; kanamycin, 12.5 μg/ml; chloramphenicol, 5 μg/ml; ampicillin, 30 μg/ml. The chromogenic β-galactosidase substrate, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal), was used in plates at 20 μg/ml.

The β-galactosidase assay and units of activity used were those described by Miller (24). For these assays, cells were grown either aerobically or anaerobically under conditions described previously (7).

Isolation of the control region of the *pdu/cob* regulon. The nucleotide sequence of the distal end of the *pocR* gene has been reported previously (34). This reported sequence includes an *Eco*RI site just inside the distal end of the *pocR* gene. To clone the rest of the *pocR* gene, an *Eco*RI fragment which includes the *pocR* sequence promoter proximal to this *Eco*RI site was identified. This fragment was cloned from DNA obtained from bacteriophage particles produced by inducing a Mud-P22 lysogen (45). Strain TT14383 (*metE205 ara-9 cob-24*::Mud*Q*) has a Mud-P22 prophage inserted within the *cob* operon. Induction (with mitomycin) causes the prophage to be expressed without excision; chromosomal DNA is then packaged processively in a clockwise direction from the prophage. DNA was prepared from the phage particles in this lysate, and 5 μg was digested with *Eco*RI, separated on a 1% agarose gel, and transferred to a nitrocellulose filter (23). The filters were probed with an oligonucleotide whose sequence (5′-CTCTG GCGGTTTACCCACGC-3′) is found promoter proximal to the *Eco*RI site of the *pocR* gene. A 3.7-kb genomic *Eco*RI fragment was identified. The isolated 3.7-kb fragment was ligated into *Eco*RI-cleaved pUC19 and introduced into strain MC1061. Plasmid DNA from one of the transformants included the sequence of bases present in the probe. The plasmid with the 3.7-kb insert was designated pDA2979 (Fig. 1). A 1.4-kb *Eco*47III fragment of the 3.7-kb *Eco*RI insert was subcloned into vector pUC118 cleaved with *Sma*I. This construct was designated pPC3. Both plasmids were used for DNA sequencing.

Plasmid preparation and DNA sequencing. Cells for plasmid preparation were grown in rich medium. Plasmid DNA

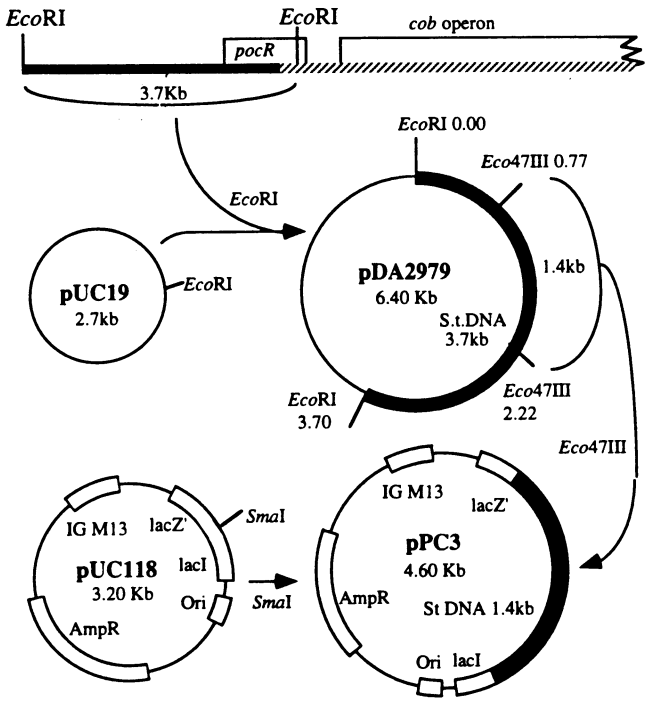

FIG. 1. Cloning and subcloning of a 3.7-kb fragment. A 3.7-kb *Eco*RI fragment isolated from a phage Mud-P22 lysate was cloned into vector pUC19 to form pDA2979. Plasmid pDA2979 was cut with *Eco*47III, and the resulting 1.4-kb *Eco*47III fragment was gel purified. Vector pUC118 was cut with *Sma*I at its polycloning site. Both *Eco*47III and *Sma*I leave a blunt-ended DNA fragment. Digested samples were mixed and ligated. The resulting plasmid was introduced into *E. coli* K-12 strain MC1061. The thick black line represents the cloned DNA, and the thick dashed line represents the previously known sequence.

was prepared by boiling (23) or passage over Qiagen columns. Bacterial transformation, restriction enzyme digestion, ligation, and gel electrophoresis were performed as described by Maniatis et al. (23). Plasmid pPC3 was used to prepare the single-stranded DNA template (40) for the determination of the sequence of one strand of the 1.4-kb *Eco*47III fragment with the Muta-Gene phagemid in vitro mutagenesis kit (Bio-Rad Laboratories); bacteriophage M13K07 was used as a helper phage. Both pUC19 and pUC118 have M13 sequence adjacent to the cloning site; primers complementary to these sequences were used to determine the sequence adjacent to the cloning vector. The second-strand DNA sequence of the 1.4-kb *Eco*47III fragment and the remaining sequences of the *Eco*RI fragment were obtained with custom-made primers and double-stranded plasmid DNA templates by conventional sequencing methods. The nucleotide sequences of both strands were determined completely and compiled with the Genetics Computer Group sequence analysis software package (12).

DNA sequences were determined by the dideoxy-chain termination procedure of Sanger et al. (35). All enzymes and chemicals for sequencing were purchased from United States Biochemical Corporation. For routine sequencing, Sequenase (version 2.0) was used. Reactions were run with 7-deaza-GTP to clarify compressions and other ambiguous sequences. Radiolabeled dATP was purchased from Amersham Corp. Custom-made primers were prepared by Robert Schackmann at the University of Utah Sequence Facility.

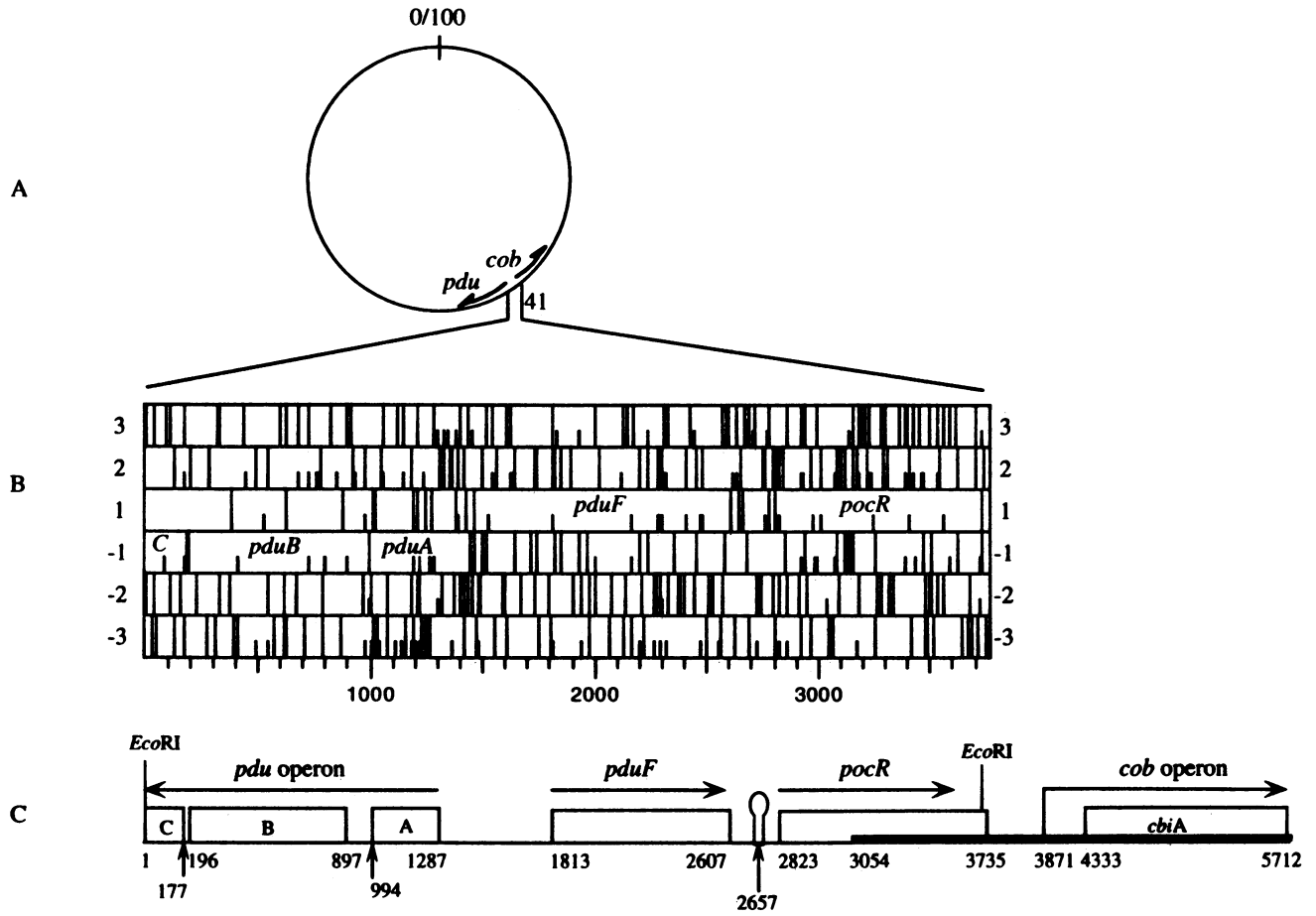

FIG. 2. Open reading frames in the determined sequence. (A) Genetic map of the *Salmonella* chromosome. (B) Translation (six frames) of the 3.7-kb DNA sequence with DNA Strider (version 1.2) software. The long vertical lines indicate the positions of termination codons; the short vertical lines indicate the positions of AUG codons (potential initiation codons). (C) Genetic map of the *pdu/cob* control region. The thin line represents the new sequence, and the thick line represents the known sequence from previous studies (28). The rectangles are open reading frames. The stem-loop structure is the putative *pduF* terminator. The arrows show the directions of gene transcription. The numbers indicate positions in the sequence, numbered from the left end of the determined sequence.

Characterization of *cobR4* and *cobR58* mutations. The *cobR4* and *cobR58* mutations cause constitutive high expression of the *cob* operon (4). Each mutation was transduced (with a linked Tn*10* element) into a strain carrying a nearby Mud-P22 prophage (strains DA1307 and DA4095). These two strains were induced to yield a phage lysate enriched for the chromosome region including the *cobR4* and *cobR58* mutations. DNA was isolated from phage particles by the method of Youderian et al. (45). DNA obtained from the phage lysate was sequenced directly to determine the sequence alteration of the two mutations. An oligonucleotide of sequence 5′-CT TACCGGTATATTGACG-3′ was used as the primer for sequencing. This oligonucleotide corresponds to the complement of bases +72 to +89 in the *cob* leader sequence as described previously (30). The *cobR4* and *cobR58* mutations are deletions known from genetic data to remove the *pocR* gene and to extend clockwise from the *cob* promoter region.

Computer analysis. Nucleotide sequence data were assem-

TABLE 2. Genes in the control region of the *pdu/cob* regulon

Gene	Upstream region[a]	Codon		Position		Peptide length (aa)[b]	Mol wt (kDa)
		Start	Stop	First base	Last base		
pocR	CTGAGGGGTTTTATC	ATG	TGA	2823	3735	304	34.4
pduF	CTCAGAAGGTGTCAC	ATG	TAA	1813	2607	265	27.7
pduA	CATGCGAGGGTCTTT	ATG	TGA	1287	994	98	10
pduB	ATACGAGAGACGGCT	ATG	TGA	897	196	234	24
pduC	CCACGAGGCTGATTC	ATG	Unknown	177	Unknown	>59	>6.9

[a] Region immediately before the start codon. Putative ribosome binding sites are underlined.
[b] aa, amino acids.

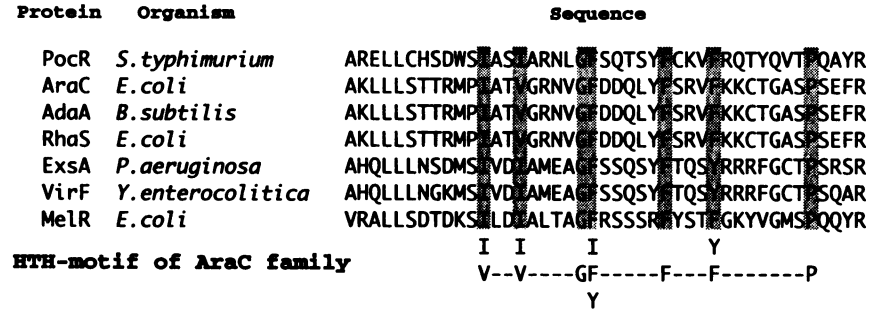

FIG. 3. Helix-turn-helix motif of the PocR protein and its alignment with those of other members of the AraC family. The following sequences are compared: arabinose operon regulatory protein of *E. coli* (AraC; GenBank accession number J01041), methylphosphotriester-DNA alkyltransferase of *Bacillus subtilis* (AdaA; GenBank accession number X53399), L-rhamnose operon regulatory protein of *E. coli* (RhaS; GenBank accession number X06058), exoenzyme S regulatory protein of *Pseudomonas aeruginosa* (ExsA; GenBank accession number M64975), virulence regulon transcriptional activator of *Yersinia enterocolitica* (VirF; GenBank accession number M22781), and melibiose operon regulatory protein of *E. coli* (MelR; GenBank accession number M18425). The helix-turn-helix motif is highlighted by gray boxes. The consensus motif is shown below the aligned sequences.

bled and analyzed with the sequence analysis software package of Genetics Computer Group, Madison, Wis. (12). By using the TFASTA program (28), which translates each database sequence in all six frames, the entire GenBank and EMBL databases were searched for proteins similar to those inferred from the sequence data. Gene Works version 2.1 (17b) was used for GC content analysis. The DNA Strider program version 1.2 (23a) was used for hydrophobicity plots according to the method of Kyte and Doolittle (21). OLIGO version 4 primer analysis software (34a) was used for the analysis and selection of primers for sequencing. Loopviewer (17a) was used to help draw RNA secondary structures.

Nucleotide sequence accession number. The nucleotide sequence of the 3.7-kb *Eco*RI fragment reported here has been

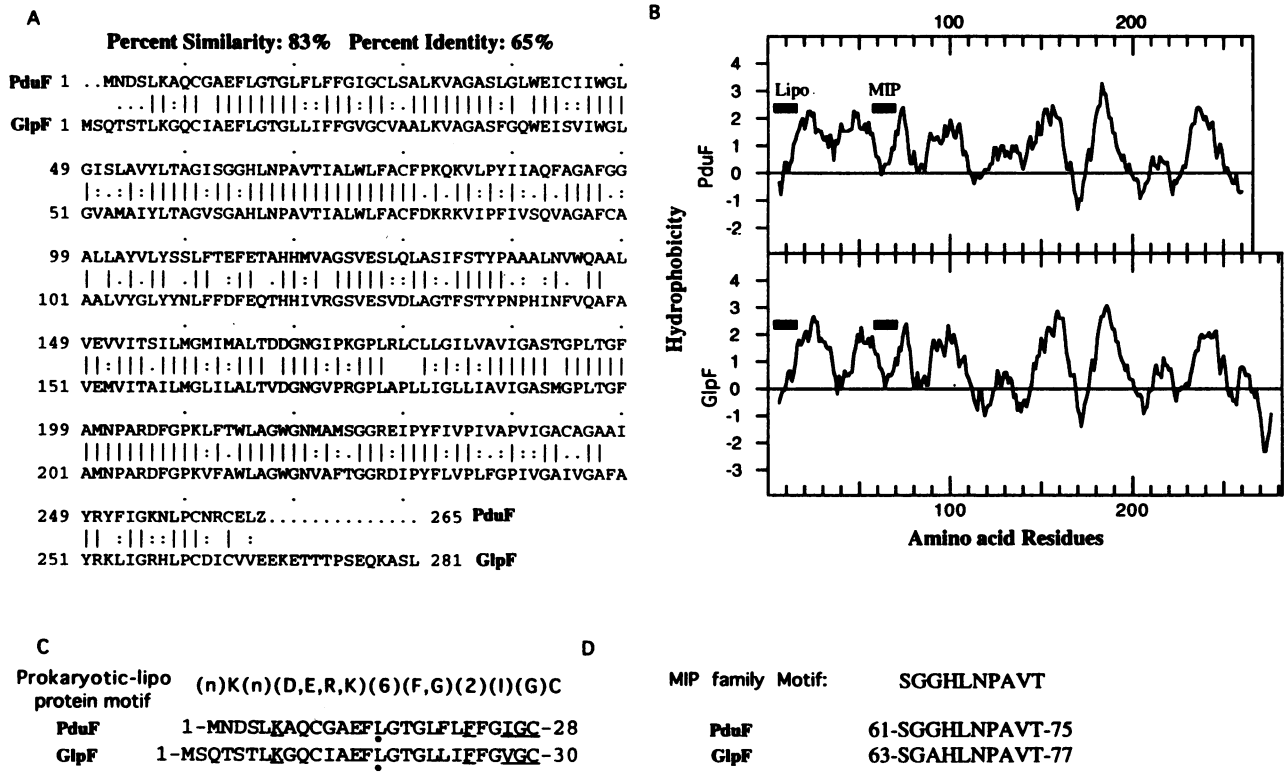

FIG. 4. Comparison of the PduF and GlpF proteins. (A) Alignment of PduF and GlpF amino acid sequences. Identical residues are indicated by vertical lines, and similar residues are indicated by dots. One dot indicates a similarity value between 0.0 and 0.5 on the Dayhoff scale; two dots indicate a value between 0.5 and 1.5 on that scale (35a). The two proteins share 65% overall identity. (B) Hydrophobicity profiles of PduF and GlpF. Hydropathy values were calculated by the method of Kyte and Doolittle (21). Black bars show the positions of the prokaryotic lipoprotein motif and MIP family motif. (C) Prokaryotic lipoprotein motifs of PduF and GlpF. n, variable base number. Conserved residues are underlined. The only mismatch with the consensus is indicated by a dot. (D) The MIP motifs of PduF and GlpF.

A

```
Proteins    Alignment of PduA with related proteins                    Residue numbers   Identity to PduA   Similarity to PduA
PduA          MQQEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVIVRGDVGAVKA   1-56
CcmK           MPIAVGMIETLGFPAVVEAADAMVKAARVTLVGYEKIGSGRVTVIVRGDVSEVQA   1-55          (over the entire
AnaI     MATRSHQNVGAIGLIETNGFPALVGAADAMLKSANVKLICYEKTGSGLCTAIVQGTVSNVTV   1-63            98 residues)
SyoI           MPIALGMVEVLGHPPALAVADVMVKAARVTLVGYEVVSGARLTIIVRGDVSEVQI   1-55
                  *.*..*. * .... .**.*.* * * *..**.....  *.**.*.*. * .

PduA       ATDAGAAAARNV--GEVK------AVHVIPRPHTDVEKILPKELANEQQZ            57-98
CcmK       SVSAGLDSAKRVAGGEVL------SHHIIARPHENLEYVLPIRYTEAVEQFRMZ        55-102          55%          70%
AnaI       AVEAGMYAAERI--GQLN------AIMVIPRPLDDLMDSLPEPQSDSEAAQPLQLPLR-- 64-112(160)     46%          64%
SyoI       AVAAGVEAAKKIPAQSPKEKTLYLSSTVIPRPHENLEAVFPKMRFQYGDGWERFLVZ     56-112          43%          61%
           .. **  .*. .  .            .*.**.  ..  .*
```

B

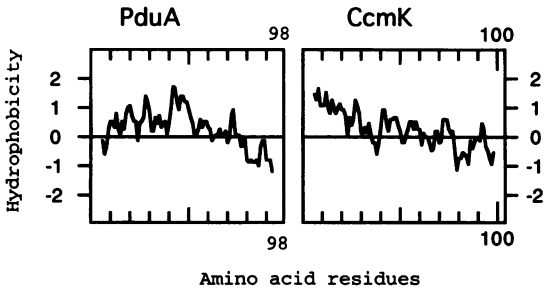

FIG. 5. Comparison of the PduA and CcmK proteins. (A) Sequence alignment of PduA and related proteins. These sequences were aligned with the new protein alignment program of Gene Works version 2.1. The CcmK sequence is from *Synechococcus* sp. strain PCC7942 (26), AnaI is from *Anacystis nidulans* 6301 (33), and SyoI is an open reading frame from *Synechococcus elongatus* (GenBank accession number D16540). Symbols: ★, invariant residue; ., conserved residue; –, gap. (B) Hydrophobicity profiles of the PduA and CcmK sequences. Hydropathy values were calculated by the method of Kyte and Doolittle (21).

submitted to GenBank and been assigned the locus STYP DUC with accession number L31414.

RESULTS AND DISCUSSION

Cloning and subcloning the control region of the *pdu/cob* regulon. The previously reported sequence of the *S. typhimurium cob* operon (34) includes the promoter-distal end of a gene inferred to encode the regulatory gene *pocR*. An oligonucleotide from the known end of the *pocR* gene was used as a probe to identify a 3.7-kb *Eco*RI fragment of the *S. typhimurium* genome. The cloned fragment overlaps the previously known sequence by less than 690 bp and thus includes 3 kb of previously uncharacterized DNA derived from the chromosomal region just clockwise of the *cob* operon. The DNA nucleotide sequence of this fragment is reported here. The subcloning and sequencing procedures used are described in Materials and Methods.

Detected open reading frames. Figure 2 depicts the open reading frames (four complete and one partial) found in the 3.7-kb DNA fragment. The inferred genes are designated *pduC* (partial), *pduB*, *pduA*, *pduF*, and *pocR* (left to right). The *pduABC* genes are read clockwise, which is consistent with the orientation of the *pdu* operon as determined by genetic studies which suggest that the entire *pdu* operon includes six or seven genes (18, 43). An open reading frame (designated *pduF*) (see below) between the *pdu* operon and the *pocR* gene was found. Multiple insertion mutations in this region showed no striking phenotype (7). Several less precisely mapped insertions in this general region resulted in slightly impaired induction of the *cob* operon by propanediol, consistent with either a regulatory or transport role for this region (31). The *pduF* open reading frame has several possible start codons. However, several lines of evidence suggest that translation starts with the ATG codon at position 1813. First, there is a good Shine-Dalgarno ribosome binding site near this start codon; none are found near the other potential start codons. Second, nucleotide content analysis (see below) (see Fig. 7) shows that the GC percentage rises sharply at bp 1813 and is high across the entire subsequent open reading frame, suggesting the existence of an AT-rich intergenic sequence to the left of that position. Finally, alignment of the PduF amino acid sequence with that of the GlpF protein starts at base 1813.

The *pocR* gene was initially identified genetically (7, 31). Part of this gene was sequenced previously (34). The new sequence extends and completes the DNA sequence of the *pocR* gene. We will present evidence elsewhere that *pocR* regulatory mutations alter the sequence that includes this open reading frame (10). The 3.7-kb *Eco*RI fragment whose cloning and sequencing are described here includes all but the last 21 bp of the *pocR* gene; this explains the failure of this plasmid to complement *pocR* mutations. The *pduF* and *pocR* genes are both transcribed counterclockwise.

All open reading frames described here have an ATG start codon which is preceded by a reasonable match to the Shine-Dalgarno ribosome binding site (Table 2). The *pduF* gene has a TAA stop codon; all the other genes terminate with a TGA stop codon (Table 2).

The *pocR* gene appears to encode a member of AraC family of regulatory proteins. Genetic studies identified a *trans*-acting

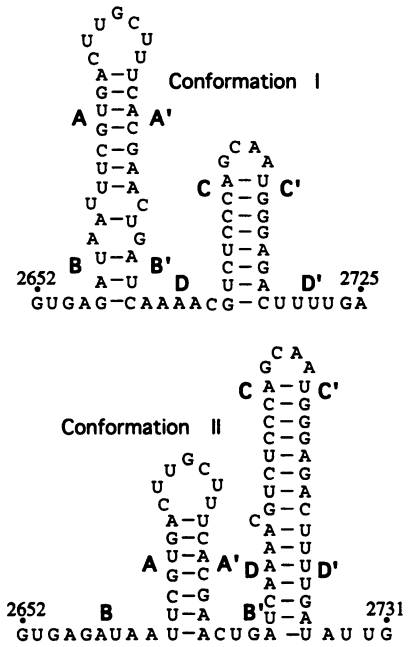

FIG. 6. Putative terminator of the *pduF* gene. The putative terminator shown in conformation I includes a stretch of uridine residues following the stem-loop. This sequence may also adopt the alternative, conformation II.

regulatory gene located between the *pdu* and *cob* operons. This gene, *pocR*, coregulates the expression of both operons (7, 31). Deletions that fail to recombine with *pocR* mutations remove the sequences that we have designated *pocR* (4, 7). The *pocR* gene encodes a 34.4-kDa protein with 304 amino acids. The inferred amino acid sequence at the C-terminal end of the PocR protein (amino acids 191 to 295) is similar to the sequences of transcriptional regulators of the AraC family (55% similarity and 33% identity compared with AraC); there is no notable similarity between the N-terminal portion of the PocR sequence (from amino acid 1 to 190) and any other protein in the database. The region of similarity between the PocR protein and proteins of the AraC family includes the helix-turn-helix motif shown in Fig. 3. This similarity suggests that PocR protein could bind specific DNA sites to activate transcription and is consistent with the regulatory effects of mutations in this region (7).

The *pduF* gene may encode a propanediol diffusion facilitation protein. The inferred amino acid sequence of the PduF protein shows 65% identity to the GlpF protein of *E. coli* (Fig. 4A) (25). The *glpF* gene has been extensively characterized and encodes a protein that facilitates the diffusion of glycerol across the cytoplasmic membrane (37, 46). The similarity of the PduF and GlpF proteins spans the entire protein sequence with no gaps. The similarity of the PduF and GlpF proteins (and the chemical similarity of propanediol and glycerol) suggests that the PduF protein may facilitate the diffusion of propanediol. This function for the gene would account for the observation that mutants in this region show impaired induction of the *cob* operon by propanediol (31). The existence of an exit facilitator for propanediol in *E. coli* has been predicted (46). Hydrophobicity plots of the PduF protein (Fig. 4B) indicate that it is a very hydrophobic protein with multiple membrane-spanning domains at positions similar to those of the GlpF protein. The PduF and GlpF proteins both contain good candidates for a prokaryotic membrane lipoprotein-lipid attachment site (42). The PduF and GlyF sequences match this motif with only one mismatched amino acid (marked by a dot in Fig. 4C). In the GlpF and PduF sequences, as in other examples, the motif is found within the first 30 amino acids of the N terminus (Fig. 4C). A lysine (K) residue should be present between this motif and the N terminus. The cysteine (C) residue of the consensus is the lipid attachment site. This cysteine residue is amino acid 28 of the PduF sequence and residue 30 of the GlpF sequence.

Another motif, called the major intrinsic protein (MIP) family signature, is found between positions 61 and 72 of the PduF protein and between residues 63 and 77 of the GlpF sequence (Fig. 4D), a position similar to that for the motif in other proteins. While the function of this motif is unclear, all proteins with this motif seem to be transmembrane channel proteins and contain six transmembrane segments. This motif is found in the mammalian MIP, a major component of the lens fiber gap junction which mediates the direct exchange of ions and small molecules from one cell to another; it is also found in the soybean nodulin-26 protein (6). It has been suggested that the MIP, soybean nodulin-26 protein, and GlpF protein all have a common ancestor, although these proteins are distributed in widely different organisms (6, 44). In the GlpF and PduF proteins, the MIP motif is located in an inferred cytoplasmic loop between the second and third transmembrane regions (Fig. 4B).

The *pduA*, *pduB*, and *pduC* genes are the most promoter-proximal genes in the *pdu* operon. Genetic studies have shown that the *cob* operon is transcribed counterclockwise and that the *pdu* operon is transcribed clockwise. The *pdu* region is likely to represent a single operon since regulatory mutants and insertions between the *pduA* gene and its promoter affect the expression of *lac* fusions throughout the region (7, 10, 43). Nucleotide sequence analysis showed that the *pduA*, *pduB*, and *pduC* genes are transcribed clockwise and thus are likely to represent the upstream end of the *pdu* operon. We will present evidence elsewhere that *pdu* mutations alter this sequence (10).

A TFASTA search of the GenBank and EMBL databases showed that the 98-amino-acid PduA protein is similar to several proteins found in photoautotrophic cyanobacteria (Fig. 5). Among these proteins, the PduA protein is most similar to the CcmK protein (carbon dioxide concentrating mechanism protein) of *Synechococcus* sp. strain PCC7942, with 55% identity and 70% similarity (29). The CcmK protein has 102 amino acids and is involved in the assembly and function of the carboxysome of synechococcus. Carboxysomes are small polyhedral subcellular structures enclosed by a thin protein shell composed of at least 12 structural proteins. The structure contains most of the cell's ribulose-1,5-bisphosphate carboxylase/oxygenase activity and serves as a microcompartment in which HCO_3^- is concentrated and converted to CO_2 by carbonic anhydrase (29). The CcmK and PduA proteins are compared in Fig. 5. Propanediol dehydratase from *Klebsiella pneumoniae* (ATCC 8724) has five subunits (39). We propose that PduA protein may assemble a similar complex of Pdu proteins in *S. typhimurium* and that this complex could serve to protect the Pdu enzymes from oxygen.

The PduB and PduC proteins showed no significant matches to any sequences in the database. There is a 98-bp noncoding sequence between the end of the *pduA* gene and the start of the *pduB* gene. No promoter-like sequence could be found in this region. The space between the *pduB* and *pduC* genes is 18 bp.

Features of the intergenic regions. Two intergenic regions

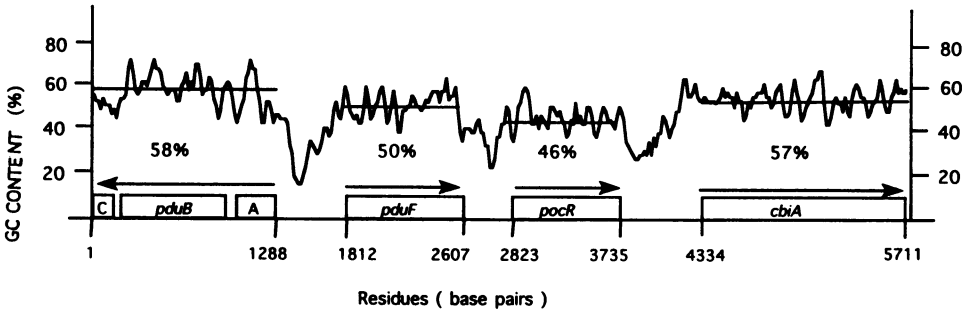

FIG. 7. GC content of the control region. The GC content was calculated by using 25-bp windows. Rectangles represent open reading frames. The arrows indicate the directions of gene transcription. Lines across the GC content curve indicate the average GC contents of the covered regions, with the corresponding percentages indicated underneath.

are described here. One lies between the *pocR* and *pduF* genes (bp 2608 to 2822); the other is between the *pduA* and *pduF* genes (bp 1288 to 1812). Both regions are AT rich, 66 and 64%, respectively. One stretch of 120 bp between the *pduF* and *pduA* genes is 73% AT. While AT richness is common for intergenic regions, it contrasts sharply in this case with the adjacent GC-rich coding sequences described below.

Outside the promoter-distal end of the *pduF* gene is a potential stem-loop structure which may adopt two conformations (Fig. 6). Conformation I has a hairpin structure with five GC pairs followed by a stretch of U residues. This hairpin structure fits the criteria for a terminator (8, 11); the energy of the structure for conformation I is about −20 kcal (1 cal = 4.184 J). The conformation I structure may terminate *pduF* transcription. The alternate mRNA structure (conformation II; −26 kcal) could play a regulatory role, allowing transcription to proceed from *pduF* into the immediately adjacent *pocR* gene under some conditions.

GC content of the control region. The GC content of the 3.7-kb *Eco*RI fragment, the *cob* leader sequence, and the entire *cbiA* gene (first gene of the *cob* operon) is shown in Fig. 7. The GC content of the *pocR* gene is 46%, and that of the *pduF* gene is 49%; the entire known sequence of the *pdu* operon (*pduA*, *pduB*, and part of *pduC*) is about 57% GC. The GC content of the *cbiA* gene and the rest of the *cob* operon is also high (58% GC) (34). The typical GC content for the genes of *S. typhimurium* and *E. coli* is 53% (26). Since neither the *pdu* operon nor the *cob* operon is present in *E. coli*, this unusual GC content suggests that *S. typhimurium* may have acquired the entire *cob/pdu* region by horizontal transfer from another organism (22, 34). The difference in GC content between the coding sequences and the spacers fits well with the positions of inferred open reading frames (Fig. 7).

The regulatory mutants *cobR4* and *cobR58* are deletions that may create new constitutively expressed promoters. The *cob* operon is maximally induced by propanediol under anaerobic conditions. During aerobic growth, induction occurs only on a very poor carbon source when cyclic AMP levels are high (7, 14). Thus, the expression of the normal *cob* operon in cells grown aerobically on glucose is minimal. Regulatory mutants that express the *cob* operon at high levels under these conditions were isolated (4, 5). Two of these mutations, *cobR4* and *cobR58*, are described here.

The *cobR4* mutation was previously shown to be a dominant, *cis*-acting mutation which (in the absence of B_{12}) increases expression of the *cob* operon about 90-fold (compared with that of wild-type cells) during aerobic growth on either glucose or glycerol (4). The *cobR58* mutation was isolated in the same way and is phenotypically similar to the *cobR4* mutation as shown in Table 3. Both mutants express the *cob* operon regardless of the global growth conditions, with or without propanediol; however, both mutants retain a sensitivity to repression by B_{12}. Repressibility by B_{12} appears to be conferred by the long *cob* leader sequence immediately upstream of the first *cob* open reading frame (30). It should be noted that these deletion mutants show better repression by B_{12} than does the fully induced wild-type operon. This has been noted previously and appears to be due to the fact that the removal of the *pocR* gene prevents the expression of the *pdu* operon, which includes genes that antagonize repression by B_{12} (2).

The *cobR4* and *cobR58* mutations are deletions (Fig. 8A). The *cobR58* mutation removes bp 2734 to 3885, and the *cobR4* mutation removes bp 2340 to 3837. Both deletions retain the intact *cob* operon leader sequence required for B_{12} repression (30); this explains why *cobR4* and *cobR58* mutants retain a B_{12} repression phenotype (Table 3). Both deletions remove the entire *pocR* gene, so their expression of the *cob* operon must be independent of the normal activator protein. The upstream

TABLE 3. Expression of the *cob* operon in *cobR4* and *cobR58* mutants

Strain	Genotype	β-Galactosidase activity in cells grown under the indicated conditions[a]											
		Aerobic					Anaerobic						
		Glucose			Pyruvate		Glucose			Pyruvate-fumarate			
		None	PD	B_{12}	None	PD	None	PD	B_{12}	None	B_{12}	PD	PD + B_{12}
TT10852	*cob-24*::MudJ	3	10	2	10	110	8	25	3	60	5	550	270
TT14679	*cob-24*::MudJ *cobR4*	300	310	5	380	350	340	330	6	410	16	420	15
TT17508	*cob-24*::MudJ *cobR58*	320	350	7	370	390	370	380	8	450	20	440	24

[a] Enzyme activities are expressed in Miller units (17). Cells were grown at 37°C in NCE medium with methionine and the indicated sources of carbon and energy. PD, propanediol. Cobalamin (B_{12}) was added to a final concentration of 1.5×10^{-7} M.

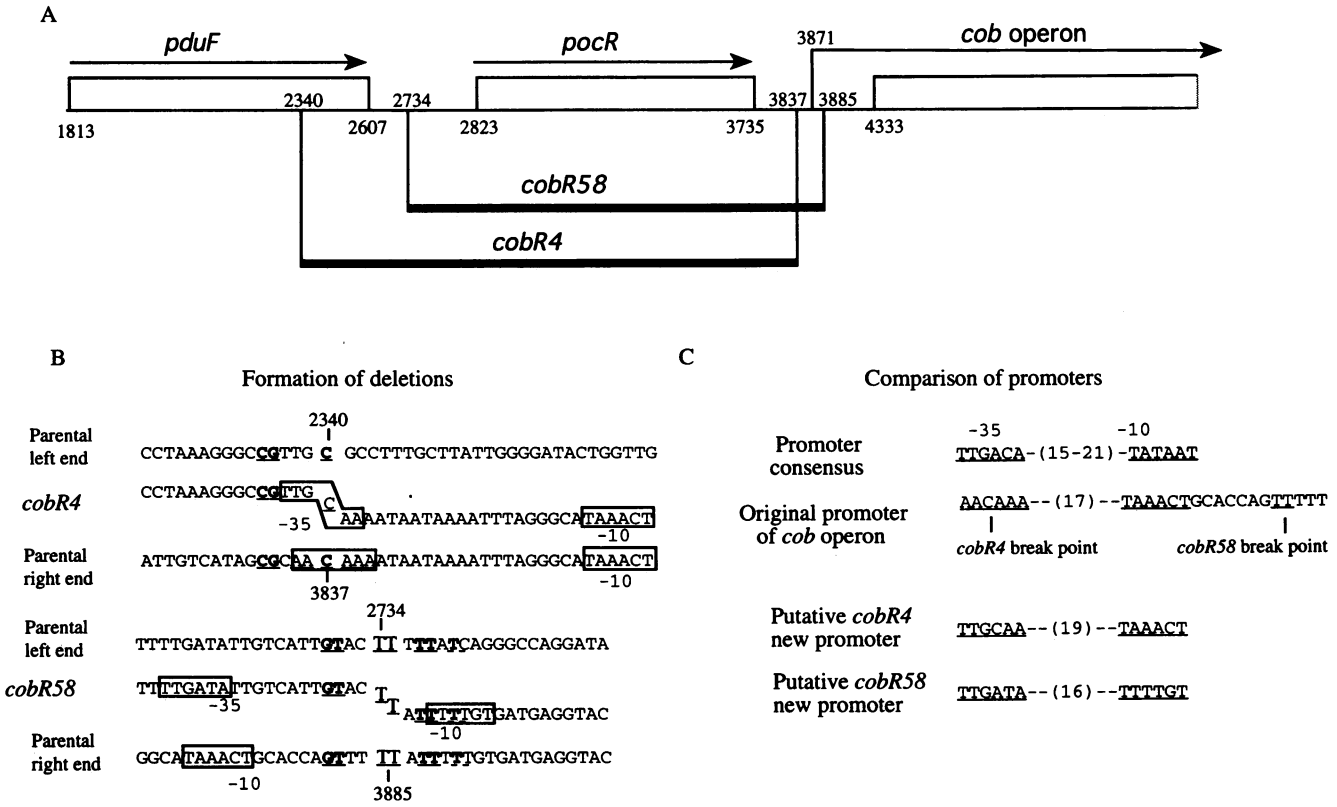

FIG. 8. Molecular characterization of the *cobR4* and *cobR58* deletions. (A) Extent of *cobR4* and *cobR58* deletions. Rectangles represent open reading frames. The arrows indicate the directions of gene transcription. The numbers are base pair positions. Heavy black bars indicate the deleted sequences. (B) Ends of the *cobR4* and *cobR58* deletions. The left and right break sites are indicated, and the original and putative new promoters are boxed. Repeated bases that may have contributed to deletion formation are underlined and in boldface type. (C) Comparison of the original and putative new promoters with the conserved consensus promoter. The −10 and −35 sequences are underlined.

end of the *cobR4* deletion lies within the 3′ end of the *pduF* gene; the upstream end of the *cobR58* deletion lies downstream of the putative terminator of the *pduF* gene. The downstream end of the *cobR4* deletion falls within the −35 region of the normal *cob* promoter, and the *cobR58* deletion ends 11 bp downstream of the −10 TATA box of the normal *cob* promoter (Fig. 8B). In Fig. 8B, the base pairs that may have contributed to the formation of these deletions are in boldface type and underlined (15).

To ensure that the expression of the *cob* operon seen in these deletion mutants is not due to transcription from the *pduF* promoter, two Tn*10 d*Tc insertions in the *pduF* gene were crossed into *cobR58* and *cobR4* deletion strains (TT17074 and TT17075). Although these Tn*10 d*Tc insertions should have terminated the transcripts from the *pduF* promoter(s), they did not reduce the high expression of the *cob* operon caused by the *cobR4* and *cobR58* deletions (data not shown).

The normal *cob* promoter matches the consensus −35 box very poorly (Fig. 8C), and its expression depends on activation by the PocR protein. The two new promoters are PocR independent, and both appear to have acquired a better −35 region. For the *cobR4* deletion, the new −35 sequence is TTGCAA; for the *cobR58* deletion, the new −35 sequence is TTGATA. Both new promoters match the −35 consensus sequence better than the wild-type −35 sequence does (Fig. 8C).

Apparently, the *cobR4* and *cobR58* deletions each create a new independent promoter rather than fusing *cob* operon expression to the preexisting *pduF* promoter. Since repression by B_{12} is retained by strains with new constitutive promoters, the structures of these deletions provide strong support for the idea that B_{12} "repression" is exerted at a level other than the variation of the frequency of promoter starts (30).

ACKNOWLEDGMENTS

This work was supported in part by NIH grant GM34804.

We thank Tom Bobik, Jeffrey Lawrence, and Michael Ailion for helpful suggestions and technical advice.

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. **Ailion, M., and J. R. Roth.** Unpublished data.
3. **Andersson, D. I.** 1992. Involvement of the Arc system in redox regulation of the Cob operon in *Salmonella typhimurium*. Mol. Microbiol. **6:**1491–1494.
4. **Andersson, D. I., and J. R. Roth.** 1989. Mutations affecting regulation of cobinamide biosynthesis in *Salmonella typhimurium*. J. Bacteriol. **171:**6726–6733.
5. **Andersson, D. I., and J. R. Roth.** 1989. Redox regulation of the genes for cobinamide biosynthesis in *Salmonella typhimurium*. J. Bacteriol. **171:**6734–6739.
6. **Baker, M. E., and M. H. J. Saier.** 1990. A common ancestor for bovine lens fiber major intrinsic protein, soybean nodulin-26 protein and E. coli glycerol facilitator. Cell **60:**185–186.

6a. **Bobik, T.** Unpublished results.
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. **Brendel, V., and E. N. Trifonov.** 1984. A computer algorithm for testing prokaryotic terminators. Nucleic Acids Res. **12:**4411–4427.
9. **Chen, P., and J. R. Roth.** Unpublished data.
10. **Chen, P., and J. R. Roth.** Unpublished data.
11. **Cheng, S. C., K. R. Leason, D. L. Court, B. A. Shapiro, and D. I. Friedman.** 1991. Functional importance of sequence in the stem-loop of a transcription terminator. Science **245:**1205–1207.
12. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
13. **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.
14. **Escalante-Semerena, J. C., and J. R. Roth.** 1987. Regulation of cobalamin biosynthetic operons in *Salmonella typhimurium*. J. Bacteriol. **169:**2251–2258.
15. **Farabaugh, P. J., and J. H. Miller.** 1978. Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *LacI* gene of *Escherichia coli*. J. Mol. Biol. **126:**847–863.
16. **Fernández-Briera, A., and A. Garrido-Pertierra.** 1988. A degradation pathway of propionate in *Salmonella typhimurium* LT-2. Biochimie (Paris) **70:**757–768.
17. **Frey, B., J. McCloskey, W. Kersten, and H. Kersten.** 1988. New function of vitamin B_{12}: cobamide-dependent reduction of epoxyqueuosine to queuosine in tRNAs of *Escherichia coli* and *Salmonella typhimurium*. J. Bacteriol. **170:**2078–2082.
17a. **Gilbert, D. G.** 1990. Loopviewer program.
17b. **Glynias, Manual J.** 1992. Gene Works version 2.1.
18. **Jeter, R. M.** 1990. Cobalamin-dependent 1,2-propanediol utilization by *Salmonella typhimurium*. J. Gen. Microbiol. **136:**887–896.
19. **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.
20. **Jeter, R. M., and J. R. Roth.** 1987. Cobalamin (vitamin B_{12}) biosynthetic genes of *Salmonella typhimurium*. J. Bacteriol. **169:** 3189–3198.
21. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for computing the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
22. **Lawrence, J., and J. R. Roth.** Unpublished data.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
23a. **March, C.** 1992. DNA Strider program version 1.2.
24. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Muramats, S., and T. Mizuno.** 1989. Nucleotide sequence of the region encompassing the *glpKF* operon and its upstream region containing bent DNA sequence of *Escherichia coli*. Nucleic Acids Res. **17:**4378.
26. **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.
27. **Obradors, N., J. Badía, L. Baldomà, and J. Aguilar.** 1988. Anaerobic metabolism of the L-rhamnose fermentation product 1,2-propanediol in *Salmonella typhimurium*. J. Bacteriol. **170:**2159–2162.
28. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
29. **Price, G. D., S. M. Howitt, K. Harrison, and M. R. Badger.** 1993. Analysis of a genomic DNA region from the cyanobacterium *Synechococcus* sp. strain PCC7942 involved in carboxysome assembly and function. J. Bacteriol. **175:**2871–2879.
30. **Richter-Dahlfors, A. A., 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.
31. **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.
32. **Roof, D. M., and J. R. Roth.** 1988. Ethanolamine utilization in *Salmonella typhimurium*. J. Bacteriol. **170:**3855–3863.
33. **Roof, D. M., and J. R. Roth.** 1989. Functions required for vitamin B_{12}-dependent ethanolamine utilization in *Salmonella typhimurium*. J. Bacteriol. **171:**3316–3323.
34. **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.
34a. **Rychlik, W.** 1993. OLIGO version 4 primer analysis software.
35. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
35a. **Schwartz, R. M., and M. O. Dayhoff.** 1979. Dayhoff table, p. 353–358. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, D.C.
36. **Smith, D. A., and J. D. Childs.** 1966. Methionine genes and enzymes of *Salmonella typhimurium*. Heredity **21:**265–286.
37. **Sweet, G., C. Gandor, R. Voegele, N. Wittekindt, J. Beuerle, V. Truniger, E. C. C. Lin, and W. Boos.** 1990. Glycerol facilitator of *Escherichia coli*: cloning of *glpF* and identification of the *glpF* product. J. Bacteriol. **172:**424–430.
38. **Toraya, T., S. Honda, and S. Fukui.** 1979. Fermentation of 1,2-propanediol and 1,2-ethanediol by some genera of *Enterobacteriaceae*, involving coenzyme B_{12}-dependent diol dehydratase. J. Bacteriol. **139:**39–47.
39. **Toraya, T., and F. Saburo.** 1982. Diol dehydrase, p. 234–262. *In* D. Dolphin (ed.), B_{12}, vol. 2. Wiley Interscience, New York.
40. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153:**3–11.
41. **Vogel, H. J., and D. M. Bonner.** 1956. Acetylornithase of *Escherichia coli*: partial purification and some properties. J. Biol. Chem. **218:**97–106.
42. **von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14:**4683–4690.
43. **Walter, D., and J. R. Roth.** Unpublished data.
44. **Yamamoto, Y. T., C. L. Cheng, and A. M. A. Conkling.** 1990. Root-specific genes from tobacco and Arabidopsis homologous to an evolutionarily conserved gene family of membrane channel proteins. Nucleic Acids Res. **18:**7449–7449.
45. **Youderian, P., P. K. Sugiono, K. L. Brewer, N. P. Higgins, and T. Elliott.** 1990. Packaging specific segments of the salmonella chromosome with locked in Mud-P22 prophages. Genetics **118:**581–592.
46. **Zhu, Y., and E. C. C. Lin.** 1989. L-1,2-Propanediol exits more rapidly than L-lactaldehyde from *Escherichia coli*. J. Bacteriol. **171:**862–867.