JOURNAL OF BACTERIOLOGY, Nov. 1988, p. 5352–5359
0021-9193/88/115352-08$02.00/0

Vol. 170, No. 11

Molecular Characterization of the *tdc* Operon of *Escherichia coli* K-12

THOMAS J. GOSS,† HERBERT P. SCHWEIZER, AND PRASANTA DATTA*

Department of Biological Chemistry, The University of Michigan Medical School, Ann Arbor, Michigan 48109-0606

Received 6 April 1988/Accepted 12 August 1988

The nucleotide sequence of a 2-kilobase DNA fragment of the *tdc* region of *Escherichia coli* K-12, previously cloned in this laboratory, revealed two open reading frames, *tdcC* and ORFX, downstream from the *tdcB* gene (formerly designated *tdc*) encoding biodegradative threonine dehydratase. A 24-base-pair sequence separated *tdcC* from the dehydratase coding region, and an untranslated region of 60 nucleotides, which contains a recognizable −10 consensus sequence, was found between *tdcC* and ORFX. The deduced amino acid sequence of *tdcC* showed it to be a large hydrophobic polypeptide of 431 amino acid residues, whereas ORFX coded for a small 135-residue polypeptide lacking glutamine and tryptophan. A computer-assisted sequence analysis revealed no similarity among the *tdcB*, *tdcC*, and ORFX polypeptides, and a search of the GenBank database failed to detect similarity with any other known proteins. The *tdc* genes and ORFX showed similar codon usage and, in analogy with other bacterial genes, showed codon usage typical for genes expressed at an intermediate level. Transcriptional analysis with S1 nuclease indicated two distinct transcription start sites upstream of the *tdcB* gene in regions previously identified as promoterlike elements P_1 and P_2. Interestingly, expression of *tdcB* and *tdcC*, but not ORFX, was contingent upon the presence of P_1. These results taken together tend to suggest that the biodegradative threonine dehydratase is the second gene in a polycistronic transcription unit constituting a novel operon (*tdcABC*) in *E. coli* implicated in anaerobic threonine metabolism.

Biodegradative threonine dehydratase (EC 4.2.1.16) of *Escherichia coli* catalyzes the pyridoxal phosphate-dependent dehydration of L-threonine and L-serine to ammonia and to α-ketobutyrate and pyruvate, respectively (22). The enzyme is normally induced under anaerobic culture conditions in tryptone-yeast extract medium lacking fermentable carbohydrates. Recently, Hobert and Datta (11) devised a synthetic medium consisting of four amino acids, threonine, serine, valine, and isoleucine, plus cyclic AMP and fumarate, which supported higher levels of dehydratase synthesis as compared with that found in the rich medium with or without cyclic AMP and fumarate. Because the requirements for cyclic AMP and fumarate were absolute in terms of enzyme synthesis, it was proposed that not the amino acids themselves but some metabolite(s) derived from these amino acids during anaerobic metabolism was needed for dehydratase induction. In addition, Merberg and Datta (17) reported that various strains of *E. coli* with unrelated genotypes produced various levels of the enzyme in the same medium and that mutations at multiple loci on the chromosome influenced enzyme expression in vivo. These cumulative findings led to the notion that a complex network of regulatory systems controls the synthesis of dehydratase in anaerobic culture conditions.

To understand the mechanism of dehydratase induction at the DNA level, we cloned a 6.2-kilobase-pair (kb) DNA fragment of *E. coli* harboring the *tdc* gene encoding threonine dehydratase (6, 7). Subcloning, deletion analysis, and nucleotide sequencing of a 2.5-kb segment of the cloned DNA (3, 7) revealed that the functional promoter P_1 for *tdc* gene expression was located about 1 kb upstream from its protein coding region, separated by an open reading frame of 936 base pairs (bp), which produced a polypeptide of about 34 kilodaltons (Fig. 1). The amino acid sequence of threonine dehydratase, deduced from the nucleotide sequence of the *tdc* gene (3), showed it to be a polypeptide of 329 amino acid residues with a molecular weight of 35,000. Interestingly, a preliminary examination of the nucleotide sequence flanking the 3′ end of the *tdc* gene suggested a potential open reading frame beyond the dehydratase-coding region. In this report we present the nucleotide sequence of the DNA segment downstream from *tdc*, identify the transcription initiation site of the upstream promoter, and provide evidence for a new *E. coli* operon consisting of at least three separate genes controlled by the distal promoter. In this article, the 5′ open reading frame and the *tdc* gene encoding threonine dehydratase (3, 6, 7) are renamed *tdcA* and *tdcB*, respectively, and the open reading frame downstream from *tdcB* is referred to as *tdcC*.

MATERIALS AND METHODS

Bacterial strains and growth media. The bacterial strains and plasmids are listed in Table 1. Media were prepared as described by Miller (19) and Davis et al. (4). If needed, antibiotics were added at the following concentrations: ampicillin, 100 μg/ml; tetracycline, 10 μg/ml; kanamycin, 50 μg/ml.

Plasmids. pEC61, PTG4B, and pTG122 have been described (7); pTG1A, pTG11, and pBal11 were isolated by the procedure of T. J. Goss (Ph.D. thesis, The University of Michigan, Ann Arbor, 1986). pTG1A was constructed by cloning a 3.7-kb *Sal*I fragment from pEC61 into *Sal*I-digested pUC13 (18, 23) with *lacP* in the opposite orientation of *tdc*. pRS124 was constructed by isolating a 3.4-kb *Hin*dIII fragment from pTG1A and then cloning the purified fragment into pTG122 previously digested with *Hin*dIII and treated with alkaline phosphatase. After transformation of *E. coli* LE392, Ap^r colonies were screened for the presence of

* Corresponding author.
† Present address: School of Environmental and Life Sciences, Murdoch University, Murdoch, Western Australia 6150, Australia.

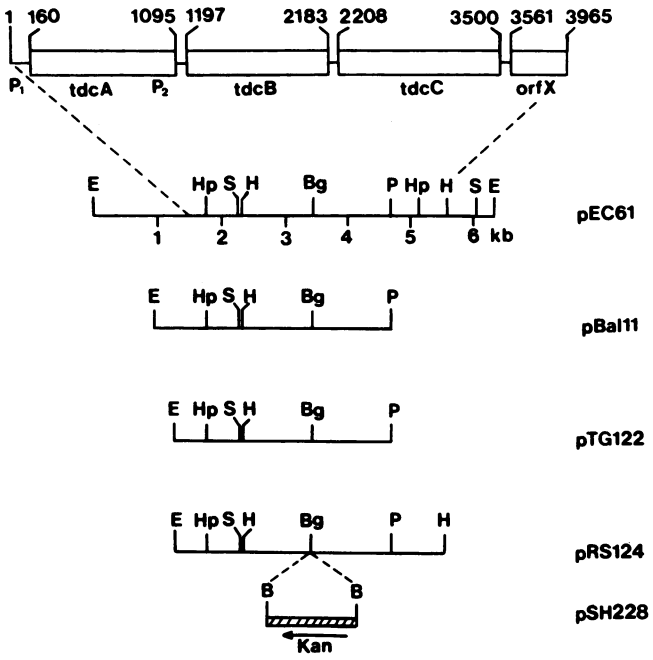

FIG. 1. Schematic diagram of the *tdc* region of *E. coli* DNA. The extent of DNA harbored by various plasmids discussed in the text is indicated. Abbreviations: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; Hp, *Hpa*I; P, *Pst*I; S, *Sal*I; Kan, kanamycin resistance gene cassette. The nucleotide numbers were assigned arbitrarily beginning with the endpoint of pTG122 (7).

threonine dehydratase activity. Plasmid DNAs were isolated (1) from two transformants with enzyme activity and digested with *Eco*RI and *Pst*I to identify the correct orientation of the cloned DNA. The resultant plasmid pRS124 contained a 4-kb *E. coli* insert harboring P_1 and the entire coding region of *tdcABC* and open reading frame ORFX and with *lacP* in opposite orientation to P_1. Plasmid pSH228 was obtained by insertion of the 1.4-kb *Bam*HI fragment of pUC4K (23) containing a Kan^r gene into the single *Bgl*II site of pRS124. Only isolates that contained the Kan^r gene in a transcriptional orientation opposite to that of the *tdc* genes were chosen for further study (Fig. 1). The plasmid pTG11, used for sequencing *tdcC* and its downstream region, contained the 2.6-kb *Bgl*II-*Eco*RI right-arm fragment of pEC61 (7).

TABLE 1. *E. coli* K-12 strains and plasmids

Strain or plasmid	Genotype or insert	Source or reference
E. coli		
JM103	Δ(*lac-pro*) *supE thi strA sbcB15 endA hsdR4*/F′ *traD36 proAB lacI*^{q} ZΔ*M15*/P1 lysogen	26
LE392	F$^-$ *hsdR514* (r^-_K m^+_K) *supE44 supF58 lacY1* or Δ(*lacIZY*)*6 galK2 galT22 metB1 trpR55* (λ$^-$)	16
χ1411	F$^-$ *glnU42 minA1 minB1* (λ$^-$)	2
Plasmids		
pEC61	6.2-kb *Eco*RI fragment	7
pTG4B	4.5-kb *Eco*RI-*Pst*I fragment of pEC61	7
pTG122	2.8-kb *Bal*31 deletion–*Pst*I	7
pTG11	2.6-kb *Bgl*II-*Eco*RI right-arm fragments of pEC61	Goss (Ph.D. thesis)
pBal11	3.1-kb *Bal*31 deletion–*Pst*I	Goss (Ph.D. thesis)
pRS124	4.0-kb *Bal*31 deletion–*Hin*dIII	This study
pTG1A	3.7-kb *Sal*I fragment of pEC61	Goss (Ph.D. thesis)
pSH228	1.4-kb Kan^r *Bam*HI fragment of pUC4K in pRS124	This study

Isolation and sequencing of DNA fragments. The methods for isolation of restriction fragments, cloning into M13 phages, and sequencing by the dideoxy-chain termination procedure were as described previously (3, 7). Restriction enzymes, polynucleotide kinase, and the Klenow fragment of DNA polymerase were purchased from Bethesda Research Laboratories, Inc. (Rockville, Md.), or International Biotechnologies, Inc. (New Haven, Conn.). Reagents for dideoxy DNA sequencing were bought from Amersham Corp. (Arlington Heights, Ill.).

S1 nuclease mapping. S1 mapping was carried out as described by Maniatis et al. (16), except DNA-RNA hybridizations were carried out at 37 or 42°C and samples were electrophoresed on 6% sequencing gels. For some experiments, S1 nuclease (Pharmacia Fine Chemicals, Piscataway, N.J.) was substituted by mung bean nuclease (Pharmacia), and the digestion was carried out in buffer containing 50 mM NaCl, 50 mM sodium acetate (pH 4.6), 4.5 mM zinc sulfate, and 20 μg of denatured calf thymus DNA per ml. Specific DNA fragments used in hybridizations were as follows: for P_1 promoter, a 527-bp *Eco*RI-*Hpa*I fragment (extending 270 bp upstream from the endpoint of pTG122) (7) (Fig. 1) end labeled at the *Hpa*I terminus was isolated from pBal11; for P_2, a 433-bp *Hin*dIII-*Acc*I fragment (from nucleotides 638 through 1070), end labeled at the *Acc*I terminus, was isolated from pTG122 (7). RNA was isolated by the hot phenol method of von Gabain et al. (24), as modified by Lau et al. (15), from strains JM103 and JM103(pEC61), grown anaerobically for 8 h at 37°C in tryptone-yeast extract medium (7, 17). Purified RNA was precipitated with ethanol and dissolved in sterile water.

Purification and labeling of minicells. Minicells were purified from 200 ml of a late-log-phase culture (A_{600} of 1) of the minicell-producing strain X1411 grown in L broth as described by Clark-Curtiss and Curtiss (2). Minicells obtained after three differential low-speed centrifugations and two sucrose gradient fractionations were suspended in buffered saline containing 1% gelatin and 1 mM EDTA (2) to a cell density of ca. 0.5 optical density unit at 600 nm. Purified minicells were stored at −70°C after the addition of glycerol to a final concentration of 30%. Proteins were labeled with [^{35}S]methionine (New England Nuclear Corp., Boston, Mass.) during a 30-min incubation at 37°C in M9 medium (4) as described by Schweizer and Boos (20), except that the chase period was shortened to 5 min and glucose was omitted. Radiolabeled proteins were analyzed on 13% sodium dodecyl sulfate–polyacrylamide slab gels as described by Larson et al. (14) with the buffer of Laemmli (13).

Other methods. The DNA sequence was analyzed by using the Delaney program (5) in the Michigan Terminal System of The University of Michigan. Amino acid sequence homology search and other analyses were carried out by using the Beckman MicroGenie software (Beckman Instruments, Palo Alto, Calif.).

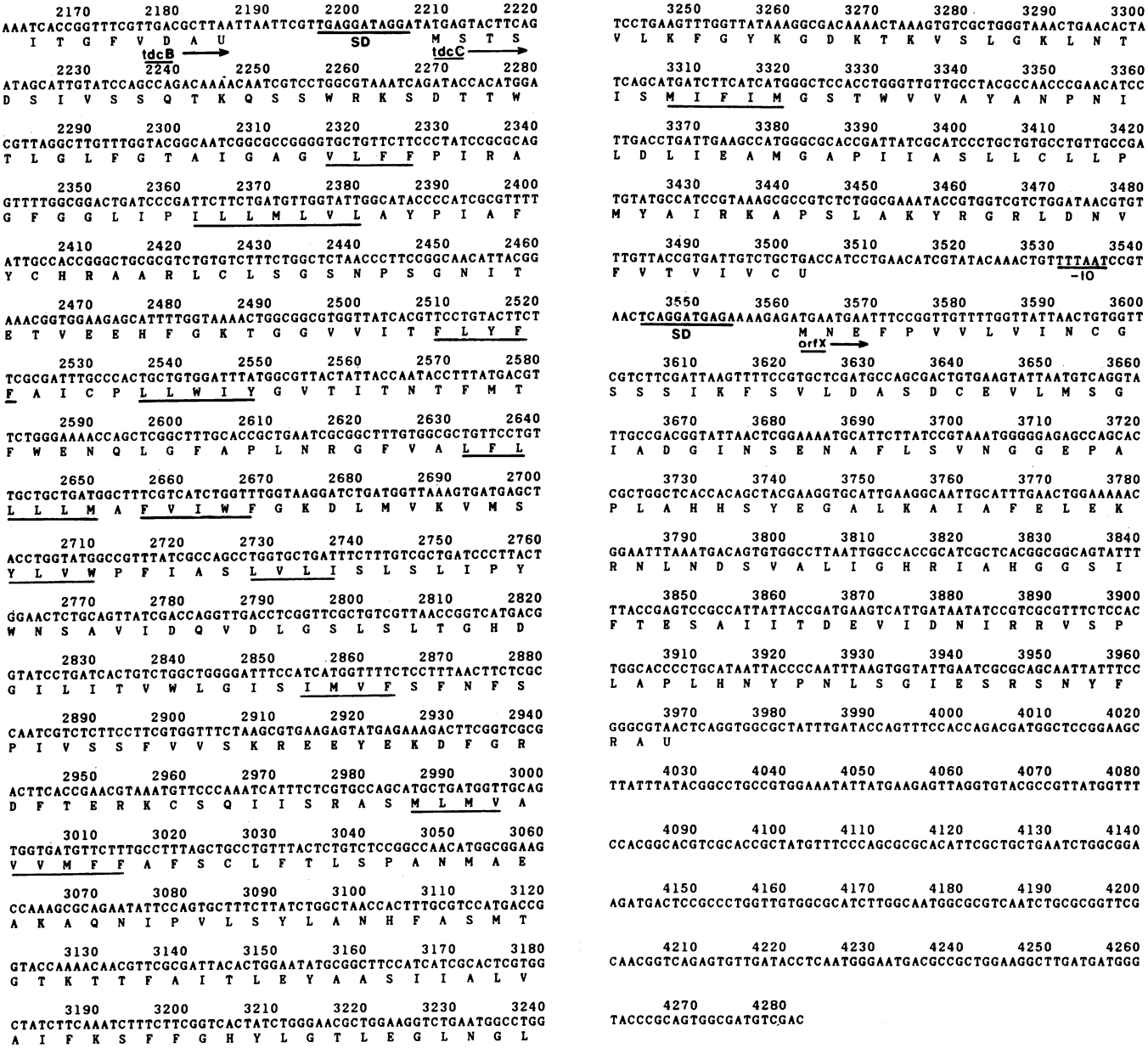

FIG. 2. Nucleotide sequence of the 3'-flanking region of the *tdcB* gene. The nucleotide numbers were assigned beginning with the endpoint of pTG122 DNA (7). The 3' end of the *tdcB* coding region is shown at the top left; the complete sequence of *tdcB* has previously been published (3). The bases underlined and marked SD and −10 represent homology with the Shine-Dalgarno sequence and the consensus Pribnow box, respectively. The deduced amino acid sequences of *tdcC* and ORFX are also shown. Stretches of hydrophobic residues in the *tdcC* open reading frame are underlined.

RESULTS AND DISCUSSION

Nucleotide sequence. In earlier experiments (3, 6, 7), a 6.2-kb *Eco*RI fragment of *E. coli* DNA, cloned in this laboratory, was found to contain the threonine dehydratase gene (Fig. 1). The nucleotide sequence of an approximately 2.5-kb segment of this DNA from plasmid pTG122 (3, 7) revealed an untranslated sequence of 159 nucleotides harboring the functional promoter P_1, followed by the coding regions of *tdcA* (nucleotides 160 through 1095) and *tdcB* (nucleotides 1197 through 2183). The sequence data also suggested a potential open reading frame beyond the *tdcB* gene (Fig. 1). To determine the nucleotide sequence farther downstream, plasmid pTG11 was constructed that contained a 2.6-kb *Bgl*II-*Eco*RI right-arm fragment of pEC61 (Fig. 1). Figure 2 displays the nucleotide sequence of 2,121 bp (from nucleotides 2161 through 4282) that overlaps with seven codons at the 3' end of the *tdcB* reading frame (3). A computer-aided analysis of possible open reading frames in all three phases on both strands revealed two potential coding sequences in this DNA segment: a 1,293-bp region (designated *tdcC*) 24 nucleotides downstream from the end of the *tdcB* gene and a 405-bp sequence (tentatively named ORFX) separated by an untranslated region of 60 nucleotides from *tdcC*. Only one other potential coding sequence of

any significant length (90 nucleotides, between nucleotides 4143 and 4232) was found, 178 bp downstream from ORFX. Based on the sequence data, the small untranslated region between *tdcB* and *tdcC* does not appear to have a strong promoterlike element, although it contains a consensus Shine-Dalgarno sequence (21) at nucleotides 2196 through 2206 that is capable of binding ribosomes. On the other hand, the noncoding region between *tdcC* and ORFX contains a recognizable −10 sequence (around nucleotide 3535), typical of a procaryotic promoter (10), and a purine-rich strong ribosome-binding site (nucleotides 3544 through 3554). In addition, a 38-bp sequence was noted between nucleotides 3969 and 4006 immediately after the stop codon of ORFX, which is capable of forming a stem-loop structure in the transcribed RNA (data not shown). However, the role of this structure in the regulation of transcription of the preceding coding regions remains to be deciphered.

S1 nuclease mapping. From the initial sequence data and deletion studies of the 5′-flanking region of the *tdcB* gene, Goss and Datta (7) postulated the presence of two promoterlike elements: P_1, near nucleotide 120, approximately 40 bp upstream from the translation-start site of *tdcA*; and P_2, around nucleotide 900, located within the coding region of *tdcA* near the 3′ end (Fig. 1). Further, P_1 was shown to be necessary and sufficient for dehydratase expression, and P_2 by itself had no apparent significant role in TdcB synthesis. These results were interpreted to mean that transcription of mRNA for both *tdcA* and *tdcB* was initiated from P_1. The 24-bp sequence between *tdcB* and *tdcC* lacked a consensus promoter (Fig. 2), raising the possibility that P_1 and/or P_2 might be involved in transcription of the *tdcC* gene. In view of this, we examined the in vivo function of these putative promoters by S1 nuclease mapping experiments using appropriate probes specific for P_1 and P_2 (see Materials and Methods). The results showed one major transcription start site from two adjacent nucleotides in P_1 (Fig. 3A), and two closely spaced regions of transcription initiation in P_2 (Fig. 3B), in addition to several weak bands representing transcription of a few RNA molecules, especially from the P_2 region. A comparison of relative band intensities on autoradiographs of the protected DNA fragments from P_1 and P_2 (having equivalent radioactivity; see legend to Fig. 3) tends to suggest that the amount of in vivo transcripts originating from the P_2 promoter is quite small. Control experiments revealed relatively weak bands at the same positions with RNA extracted from cells without the plasmid (Fig. 3, lane 3) and lack of S1 protection when tRNA was used for hybridization (Fig. 3A, lane 5; Fig. 3B, lane 1). The use of mung bean nuclease in place of S1 appeared to result in slightly longer protected DNA fragments (Fig. 3A, lane 1; Fig. 3B, lane 2), which is consistent with the milder digestion characteristics of this nuclease of single- or double-stranded junctions (16).

Figure 4 displays the nucleotide sequence of DNA around promoter P_1 and P_2 (7). Based on the relative banding patterns with both mung bean and S1 nucleases (see above), we can tentatively assign the transcription initiation sites at G-129 and C-130 in P_1 and at T-876, T-884, and T-887 in P_2; the exact nucleotides involved in the initiation of RNA synthesis from these promoters, however, must be confirmed by other means. It is noteworthy that both of these transcription start sites are preceded by the consensus −10 and −35 regions presumed to bind RNA polymerase molecules. Because transcription initiation at pyrimidines, especially at a cluster of sites as seen in P_2 and in several other *E. coli* promoters (10), is generally indicative of weak promoter-polymerase interactions (25), it is not surprising that relatively few in vivo transcripts were found to originate from P_2 (Fig. 3B). In any event, the significance of the two promoters, especially P_2, which is located within the coding sequence of *tdcA*, remains unclear, particularly when P_2 by itself is not sufficient for *tdcB* expression (7). It is possible to envisage that active transcription from P_2 requires simultaneous transcription from P_1; alternatively, the region around

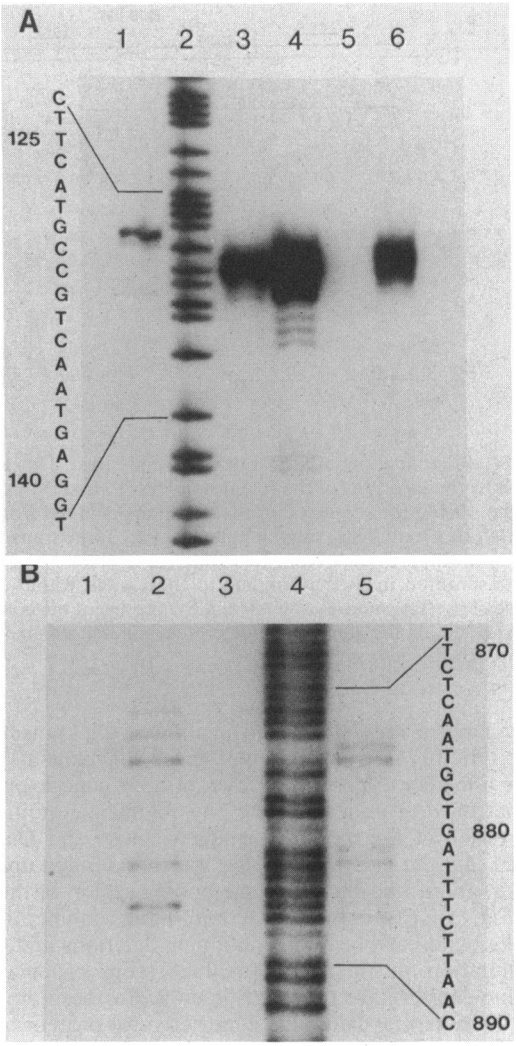

FIG. 3. S1 mapping of transcription start-sites. (A) For the P_1 promoter, 0.03 pmol (about 30,000 cpm) of a 527-bp *Eco*RI-*Hpa*I fragment was hybridized at 42°C with 100 μg of RNA and treated with nucleases as follows (lanes): 1, 4, or 6, RNA from JM103(pEC61) and 200 U of mung bean nuclease, 200 U of S1, or 50 U of S1, respectively; 3, RNA from JM103 and 200 U of S1; 5, yeast RNA and 200 U of S1; 2, A+G sequence ladder of labeled DNA (60,000 cpm). (B) For the P_2 promoter, 0.03 pmol (about 30,000 cpm) of a 433-bp *Hin*dIII-*Acc*I fragment was hybridized at 37°C with 100 μg of RNA and treated with nucleases as follows (lanes): 1, yeast RNA and 200 U of S1; 2 or 5, JM103(pEC61) RNA and 200 U of mung bean nuclease or S1, respectively; 3, JM103 RNA and 200 U of S1; 4, A+G sequence ladder of labeled DNA with 60,000 cpm. After hybridization and nuclease treatment, the samples were denatured and electrophoresed through 6% polyacrylamide sequencing gel and subjected to autoradiography. A correction of one nucleotide downward was made in comparing the nuclease protected bands with the sequence ladder displaying the noncoding strand.

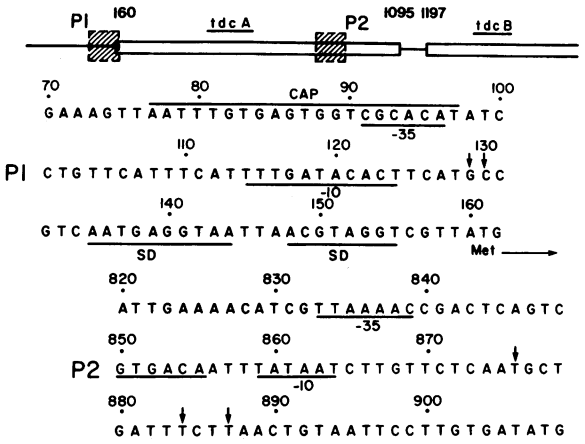

FIG. 4. Identification of transcription start sites. The hatched boxes show the locations of the two promoters, P_1 and P_2, in the *tdc* DNA; the nucleotide sequence of these regions is taken from Goss and Datta (7). From the S1 mapping data of Fig. 3, the transcription of mRNA in P_1 may occur from two adjacent nucleotides, G_{129} and C_{130}; transcription in P_2 may initiate at three sites within a 12-bp sequence, T_{876}, T_{884}, and T_{887}. CAP and SD represent the consensus cyclic AMP-catabolite gene activator protein-binding site and Shine-Dalgarno sequence, respectively.

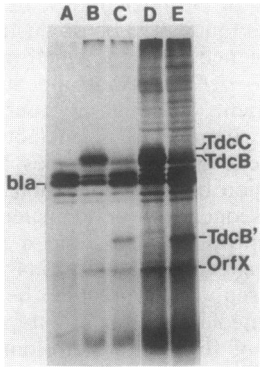

FIG. 5. Plasmid-directed polypeptide synthesis in minicells. Purification and labeling of minicells are described in Materials and Methods. Lanes: A, vector pUC13; B and D, pRS124; C and E, pSH 228. bla, β-Lactamase. The positions of the TdcB and TdcC polypeptides and of a truncated TdcB polypeptides are indicated. Lanes D and E are the same as B and C but were overexposed to visualize the TdcC polypeptide. Marker proteins (not shown) included β-galactosidase (M_r 116,000), phosphorylase *b* (M_r 97,000), bovine serum albumin (M_r 67,000), ovalbumin (M_r 45,000), carbonic anhydrase (M_r 29,000), and lysozyme (M_r 14,400).

P_2 may have a regulatory role in influencing transcription from P_1. If this is true, either of these mechanisms might provide a means to control of level of *tdcB* gene expression over that of *tdcA* under specific physiological conditions.

Expression of the *tdc* gene products. Goss and Datta (7) reported that the plasmid pTG4B, which extended up to the *Pst*I site, about one-third into the coding region of the *tdcC* gene (Fig. 1), expressed TdcA and TdcB polypeptides in minicells, whereas plasmids containing deletions and frameshift mutation in the *tdcA* reading frame produced the threonine dehydratase polypeptide only. Furthermore, deletion of the potential catabolite gene activator protein-binding site in P_1 in pTG107 (7) prevented dehydratase synthesis. These and other results led to the notion that both *tdcA* and *tdcB* are transcribed as a single transcription unit from P_1. The nucleotide sequence in Fig. 2 revealed that both *tdcC* and ORFX are preceded by strong Shine-Dalgarno sequences for ribosome binding and as such are capable of initiating translation of polypeptides of 47,000 and 14,000 daltons, respectively, as calculated from the deduced amino acid sequences of their coding regions. Plasmid-directed expression of various polypeptides in minicells was examined to decide whether these polypeptides are indeed synthesized in vivo, and whether *tdcC* and ORFX belong to the same transcriptional unit with *tdcA* and *tdcB*.

Figure 5 displays the patterns of polypeptide synthesis in minicells by plasmids pRS124 and pSH228; the extent of chromosomal DNA carried by these plasmids is shown in Fig. 1. It is evident that pRS124 directed the synthesis of three polypeptides (lane B) of 16, 38, and 45 kilodaltons (16K, 38K, and 45K polypeptides, respectively), in addition to the vector-encoded products (lane A). On the other hand, plasmid pSH228, carrying an insertion within the *tdcB* gene, did not express the 38K and 45K polypeptides but synthesized the 16K polypeptide and a new polypeptide of about 20K (lane C). We can make several conclusions from these results. (i) As reported earlier (3, 7), the 38K polypeptide appears to be the *tdcB* gene product. Its position on the gel is in good agreement with the calculated molecular weight of 35,238 from the amino acid composition of threonine dehydratase (3); furthermore, the absence of this polypeptide and the appearance of a new 20K polypeptide, seen with pSH228, is consistent with the presence of a truncated product of this size due to interruption of the *tdcB* reading frame at the *Bgl*II site by the Kan^r cassette (Fig. 1). (ii) The 45K polypeptide expressed in pRS124 most likely represents the *tdcC* gene product, a highly hydrophobic polypeptide of 47,000 daltons. The plasmid pSH228 did not synthesize this polypeptide in minicells, presumably because insertion of the Kan^r cassette in *tdcB* (in opposite transcriptional orientation) is likely to exert a strong polar effect on *tdcC* gene expression. These results, taken together with the previous findings that both *tdcA* and *tdcB* are transcribed from P_1 (7), suggest that all three genes, *tdcA*, *tdcB*, and *tdcC*, belong to the same operon. It should be mentioned in this context that the *tdcA*-encoded polypeptide of 35,000 daltons cannot be detected in this gel system because of its relatively weak expression in minicells (7) and the relative positions of the vector-directed polypeptides. (iii) The 16K polypeptide observed in both pRS124 and pSH228 appears to derive from ORFX as judged from its molecular size calculated from the deduced amino acid sequence (Fig. 2). Because the synthesis of this polypeptide was not reduced in pSH228, it most likely does not represent a *tdc* gene product.

Characterization of the Tdc polypeptides. The amino acid composition of the deduced polypeptides encoded by *tdcC* and ORFX are given in Table 2 (along with those of TdcA and TdcB polypeptides for comparison). The predicted TdcC polypeptide with 431 amino acid residues appears highly hydrophobic, with 46% of the total amino acid residues that are hydrophobic in nature; interestingly, these amino acid residues are not randomly distributed but are found in small clusters mostly in the N-terminal half of the polypeptide (underlined residues in Fig. 2). A hydropathy plot of TdcC made by the method of Kyte and Doolittle (12) (Fig. 6) supports this notion and further suggests that this protein with several alternating domains of hydrophobic and hydrophilic stretches is most likely to be found in the *E. coli* membrane. The amino acid compositions of the other polypeptides, on the other hand, do not reveal any unusual

TABLE 2. Amino acid composition of various polypeptides

Amino acid	No. (%) of residues in:			
	tdcA[b]	*tdcB*[c]	*tdcC*	ORFX
Ala	18 (5.8)	28 (8.5)	35 (8.1)	14 (10.4)
Arg	12 (3.8)	16 (4.8)	12 (2.8)	6 (4.4)
Asn	14 (4.5)	14 (4.2)	15 (3.5)	11 (8.1)
Asp	8 (2.6)	21 (6.4)	11 (2.6)	6 (4.4)
Cys	3 (1.0)	6 (1.8)	7 (1.6)	2 (1.5)
Gln	13 (4.2)	9 (2.7)	6 (1.4)	0 (0)
Glu	24 (7.7)	18 (5.5)	12 (2.8)	10 (7.4)
Gly	18 (5.8)	33 (10.0)	32 (7.4)	10 (7.4)
His	2 (0.6)	6 (1.8)	5 (1.2)	5 (3.7)
Ile	21 (6.7)	34 (10.3)	37 (8.6)	13 (9.6)
Leu	32 (10.2)	21 (6.4)	53 (12.2)	12 (8.9)
Lys	16 (5.1)	18 (5.5)	18 (4.2)	3 (2.2)
Met	10 (3.2)	9 (2.7)	16 (3.7)	2 (1.5)
Phe	15 (4.8)	9 (2.7)	35 (8.1)	6 (4.4)
Pro	12 (3.8)	10 (3.0)	15 (3.5)	5 (3.7)
Ser	31 (9.9)	23 (7.0)	38 (8.8)	16 (11.9)
Thr	23 (7.4)	17 (5.2)	28 (6.5)	2 (1.5)
Trp	2 (0.6)	0 (0)	9 (2.1)	0 (0)
Tyr	9 (2.9)	8 (2.4)	14 (3.2)	3 (2.2)
Val	29 (9.3)	29 (8.8)	33 (7.7)	9 (6.7)

[a] Results for hydrophobic amino acids (Ile, Leu, Met, Phe, Trp, Tyr, and Val) were as follows: for *tdcA*, 37.8%; for *tdcB*, 33.4%; for *tdcC*, 45.7%; for ORFX, 33.3%.
[b] See reference 7.
[c] See reference 3.

features except that the putative polypeptide encoded by ORFX lacks glutamine and tryptophan; it may be recalled that the threonine dehydratase polypeptide (TdcB) also lacks tryptophan but is posttranslationally modified by this amino acid (3).

The efficiency of translational elongation of the polypeptides was estimated by examining the bias in codon usage among synonymous codons (8, 9). Table 3 shows the codon usage for each of the four open reading frames. In general, similar codon preferences were observed among these reading frames, with some notable exceptions: for example, UUA and CUG for Leu were used most frequently in all but *tdcC*, where CUG was preferred; AUC for Ile was the dominant codon in *tdcC*; AAU was the most frequently used codon for Asn in *tdcA* and ORFX, whereas AAC was preferred in *tdcB* and *tdcC*; and GGU for Gly was used largely in *tdcA*. A comparison of average codon preference values for the five genes of the *E. coli trp* operon (27) with that of the three *tdc* reading frames and of ORFX shows some differences in the average frequency of codon usage (Table 3). A calculation of codon preference statistics, *P* (which is a measure of bias in codon usage), with the bacterial highly expressed codon usage data of Grantham et al. (8), yielded *P* values of 0.870, 0.967, and 0.973 for *tdcA*, *tdcB*, and *tdcC*, respectively. When these are compared with *P* values of 1.658 for the highly expressed *lpp* gene, 0.865 for the intermediately expressed *lacY* gene, and 0.618 for the weakly expressed *trpR* gene, it is evident that the efficiency of translation of the *tdc* mRNAs was at the intermediate level of expression.

The four polypeptides coded by the open reading frames in the *tdc* region appear to bear no structural resemblance with each other. From the amino acid composition it is clear that only the *tdcC* polypeptide is highly hydrophobic. Efforts to find similarity of the deduced amino acid sequences among these polypeptides with the MicroGenie program yielded negative results, indicating that no gene duplication or recombination occurred within or between their coding sequences. A computer search of the GenBank database (updated in 1987) failed to detect similarity of *tdcA*, *tdcC*, or ORFX products with any other known proteins. On the other hand, as described earlier (3), a significant sequence homology exists between the biodegradative threonine dehydratase (coded by the *tdcB* gene) and the biosynthetic threonine dehydratases of *E. coli* (*ilvA*) and *Saccharomyces cerevisiae* (*ILV1*) and the *E. coli* D-serine dehydrase (*dsdA*). Thus, the polypeptides coded by *tdcA*, *tdcC*, and ORFX represent three new proteins in *E. coli* whose catalytic characteristics and metabolic roles remain to be uncovered.

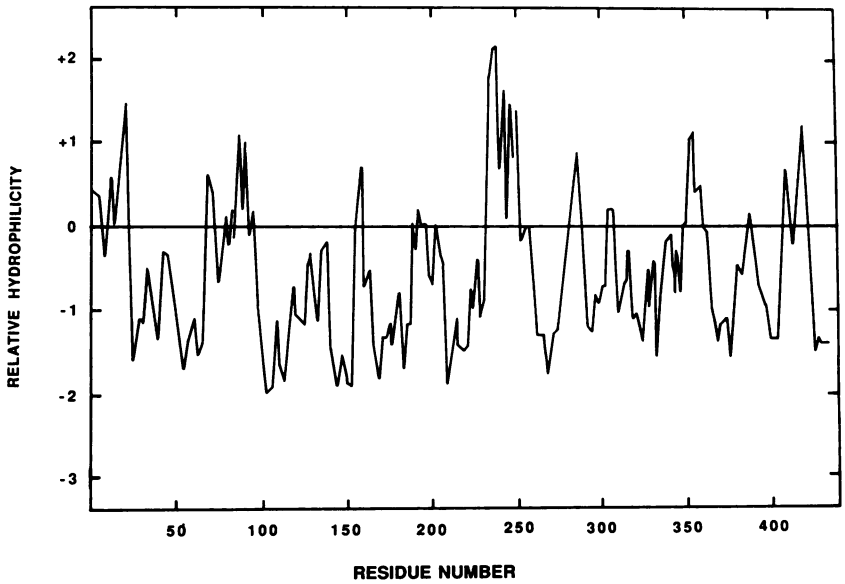

FIG. 6. Computer-generated hydropathy plot (12) of the *tdcC*-encoded polypeptide. Positive values indicate hydrophilicity, and negative values represent hydrophobicity.

TABLE 3. Codon usage in *tdcABC* and in ORFX of *E. coli*[a]

Amino acid	Codon	% Utilization in:				
		tdcA	*tdcB*	*tdcC*	ORFX	*trpEDCBA*
Phe	UUU	67 (10)	56 (5)	46 (16)	67 (4)	52
	UUC	33 (5)	44 (4)	54 (19)	33 (2)	48
Leu	UUA	40 (13)	38 (8)	4 (2)	42 (5)	11
	UUG	16 (5)	0 (0)	11 (6)	17 (2)	13
	CUU	16 (5)	5 (1)	8 (4)	0 (0)	8
	CUC	6 (2)	10 (2)	6 (3)	8 (1)	9
	CUA	0 (0)	0 (0)	0 (0)	0 (0)	4
	CUG	22 (7)	47 (10)	71 (38)	33 (4)	54
Ile	AUU	63 (13)	71 (24)	38 (14)	85 (11)	60
	AUC	28 (6)	24 (8)	62 (23)	15 (2)	38
	AUA	10 (2)	5 (2)	0 (0)	0 (0)	2
Met	AUG	100 (10)	100 (9)	100 (16)	100 (2)	100
Val	GUU	28 (8)	24 (7)	33 (11)	45 (4)	26
	GUC	17 (5)	34 (10)	15 (5)	11 (1)	20
	GUA	17 (5)	7 (2)	10 (3)	22 (2)	12
	GUG	38 (11)	34 (10)	42 (14)	22 (2)	42
Ser	UCU	23 (7)	26 (6)	29 (11)	12 (2)	15
	UCC	16 (5)	35 (8)	29 (11)	19 (3)	20
	UCA	19 (6)	9 (2)	5 (2)	6 (1)	10
	UCG	16 (5)	0 (0)	16 (6)	25 (4)	18
	AGU	16 (5)	17 (4)	3 (1)	19 (3)	13
	AGC	10 (3)	13 (3)	18 (7)	19 (3)	23
Pro	CCU	42 (5)	20 (2)	20 (3)	0 (0)	12
	CCC	0 (0)	0 (0)	7 (1)	20 (1)	14
	CCA	8 (1)	40 (4)	20 (3)	40 (2)	17
	CCG	50 (6)	40 (4)	53 (8)	40 (2)	57
Thr	ACU	30 (7)	18 (3)	25 (7)	0 (0)	14
	ACC	26 (6)	47 (8)	32 (9)	100 (2)	48
	ACA	22 (5)	12 (2)	14 (4)	0 (0)	11
	ACG	22 (5)	23 (4)	29 (8)	0 (0)	27
Ala	GCU	11 (2)	32 (9)	14 (5)	14 (2)	14
	GCC	28 (5)	7 (2)	29 (10)	36 (5)	34
	GCA	33 (6)	25 (7)	26 (9)	43 (6)	17
	GCG	28 (5)	36 (10)	31 (11)	7 (1)	35
Tyr	UAU	78 (7)	50 (4)	57 (8)	33 (1)	61
	UAC	22 (2)	50 (4)	43 (6)	67 (2)	39
His	CAU	50 (1)	50 (3)	40 (2)	20 (1)	51
	CAC	50 (1)	50 (3)	60 (3)	80 (4)	49
Gln	CAA	62 (8)	56 (5)	33 (2)	0 (0)	36
	CAG	38 (5)	44 (4)	67 (4)	0 (0)	64
Asn	AAU	71 (10)	36 (4)	27 (4)	82 (9)	40
	AAC	29 (4)	64 (9)	73 (11)	18 (2)	60
Lys	AAA	75 (12)	94 (17)	83 (15)	33 (1)	82
	AAG	25 (4)	6 (1)	17 (3)	67 (2)	18
Asp	GAU	75 (6)	62 (13)	36 (4)	50 (3)	64
	GAC	25 (2)	38 (8)	64 (7)	50 (3)	36
Glu	GAA	67 (16)	100 (18)	75 (9)	80 (8)	68
	GAG	33 (8)	0 (0)	25 (3)	20 (2)	32
Cys	UGU	67 (2)	50 (3)	29 (2)	100 (2)	45
	UGC	33 (1)	50 (3)	71 (5)	0 (0)	55
Trp	UGG	100 (2)	0 (0)	100 (9)	0 (0)	100
Arg	CGU	25 (3)	38 (6)	67 (8)	17 (1)	30
	CGC	8 (1)	31 (5)	25 (3)	50 (3)	59
	CGA	25 (3)	19 (3)	0 (0)	0 (0)	5
	CGG	8 (1)	0 (0)	8 (1)	33 (2)	3
	AGA	25 (3)	12 (2)	0 (0)	0 (0)	3
	AGG	8 (1)	0 (0)	0 (0)	0 (0)	1
Gly	GGU	61 (11)	46 (15)	44 (14)	50 (5)	38
	GGC	17 (3)	42 (14)	44 (14)	30 (3)	42
	GGA	0 (0)	6 (2)	6 (2)	10 (1)	10
	GGG	22 (4)	6 (2)	6 (2)	10 (1)	11
End	UAA		100 (1)		100 (1)	60
	UAG	100 (1)				
	UGA			100 (1)		40

[a] Percentage utilization in each codon family is shown; numbers in parentheses represent absolute values. For comparison, the average frequencies for *trpE*, *trpD*, *trpC*, *trpB*, and *trpA* coding regions of *E. coli* (27) are given.

In summary, nucleotide sequence, transcript mapping, expression of polypeptides, and other data clearly indicate that the transcription of *tdc* mRNA encoding three distinct genes is initiated from a single functional promoter, P_1, and that *tdcB*, the structural gene of the biodegradative threonine dehydratase, is the second gene in this polycistronic mRNA. This was an unexpected finding, and the chromosomal organization of these genes as well as their common transcriptional regulation implicates a novel operon in *E. coli* whose gene products may cooperate biochemically to constitute a pathway for anaerobic threonine metabolism. Experiments are in progress to analyze the physiological functions of these proteins in the context of anaerobic energy metabolism in the coliaerogenes group of bacteria.

ACKNOWLEDGMENTS

We thank G. Sanai for technical assistance.

This work was supported by Public Health Service grant GM 21436 from the National Institutes of Health.

LITERATURE CITED

1. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
2. **Clark-Curtiss, J. E., and R. Curtiss III.** 1983. Analysis of recombinant DNA using *Escherichia coli*. Methods Enzymol. **101**:347–362.
3. **Datta, P., T. J. Goss, J. R. Omnaas, and R. V. Patil.** 1987. Covalent structure of biodegradative threonine dehydratase of *Escherichia coli*: homology with other dehydratases. Proc. Natl. Acad. Sci. USA **84**:393–397.
4. **Davis, R. W., D. Botstein, and J. Roth (ed.).** 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
5. **Delaney, A. D.** 1982. A DNA sequence handling program. Nucleic Acids Res. **10**:61–67.
6. **Goss, T. J., and P. Datta.** 1984. *Escherichia coli* K-12 mutation that inactivates biodegradative threonine dehydratase by transposon Tn*5* insertion. J. Bacteriol. **158**:826–831.
7. **Goss, T. J., and P. Datta.** 1985. Molecular cloning and expression of the biodegradative threonine dehydratase gene (*tdc*) of *Escherichia coli*. Mol. Gen. Genet. **201**:308–314.
8. **Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9**:r43–r74.

9. **Gribskov, M., J. Devereux, and R. R. Burgess.** 1984. The codon preference *plot*: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12**:539–549.
10. **Hawley, D. K., and R. M. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
11. **Hobert, E. H., and P. Datta.** 1983. Synthesis of biodegradative threonine dehydratase in *Escherichia coli*: role of amino acids, electron acceptors, and certain intermediary metabolites. J. Bacteriol. **155**:586–592.
12. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
13. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–683.
14. **Larson, T. J., G. Schumacher, and W. Boos.** 1982. Identification of the *glpT*-encoded *sn*-glycerol-3-phosphate permease of *Escherichia coli*: an oligomeric, integral membrane protein. J. Bacteriol. **152**:1008–1021.
15. **Lau, E., V. Vaughn, and F. C. Neidhardt.** 1985. Isolation of RNA from *Escherichia coli*. Recomb. DNA Tech. Bull. **6**:21.
16. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 207–209. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. **Merberg, D., and P. Datta.** 1982. Altered expression of biodegradative threonine dehydratase in *Escherichia coli* mutants. J. Bacteriol. **150**:52–59.
18. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–79.
19. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Schweizer, H., and W. Boos.** 1984. Characterization of the *ugp* region containing the genes for the *phoB* dependent sn-glycerol-3-phosphate transport system of *Escherichia coli*. Mol. Gen. Genet. **197**:161–168.
21. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal *RNA*: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
22. **Umbarger, H. E.** 1978. Amino acid biosynthesis and its regulation. Annu. Rev. Biochem. **47**:533–606.
23. **Viera, J., and J. Messing.** 1982. The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
24. **Von Gabain, A., J. G. Belasco, J. L. Schottel, and A. C. Y. Chang.** 1983. Decay of mRNA in *Escherichia coli*: investigation of the fate of specific segments of transcripts. Proc. Natl. Acad. Sci. USA **80**:653–657.
25. **Von Hippel, P. H., D. G. Bear, W. D. Morgan, and J. A. McSwiggen.** 1984. Protein-nucleic acid interactions in transcription. Annu. Rev. Biochem. **53**:389–446.
26. **Yanisch-Perron, C., J. Viera, and J. Messing.** 1985. Improved M13 phage cloning vectors and host *strains*: nucleotide sequences of the M13 mp18 and pUC 19 vectors. Gene **33**:103–119.
27. **Yanofsky, C., T. Platt, I. P. Crawford, B. P. Nichols, G. E. Christie, H. Horowitz, M. VanCleemput, and A. M. Wu.** 1981. The complete nucleotide sequence of the tryptophan operon of *Escherichia coli*. Nucleic Acids Res. **9**:6647–6668.