JOURNAL OF BACTERIOLOGY, Feb. 1991, p. 1193–1200
0021-9193/91/031193-08$02.00/0

Vol. 173, No. 3

A Transcription Terminator in the Thymidylate Synthase (*thyA*) Structural Gene of *Escherichia coli* and Construction of a Viable *thyA*::Km^r Deletion

DEBORAH BELL-PEDERSEN,[1,2] JILL L. GALLOWAY SALVO,[1] AND MARLENE BELFORT[1,3]*

Wadsworth Center for Laboratories and Research, New York State Department of Health,[1] and School of Public Health, State University of New York at Albany,[3] P.O. Box 509, Empire State Plaza, Albany, New York 12201-0509, and Department of Biological Sciences, State University of New York at Albany, Albany, New York 12222[2]

Received 30 August 1990/Accepted 3 December 1990

A transcription terminator has been identified within the coding sequence of the *Escherichia coli thyA* gene. Fusion of a relevant segment of the *thyA* structural gene to *galK* sequences showed that the terminator functions in vivo. Primer extension and Northern hybridization (RNA blot) analysis of *thyA* RNA suggested that the terminator acts as the transcription stop signal for an upstream gene and for *thyA*-specific transcripts. Results from antitermination studies utilizing a λp_L-*thyA* fusion also offer evidence that the terminator is capable of attenuating *thyA* expression by reducing the amount of full-length *thyA* transcripts. This gene arrangement suggested that previous unsuccessful attempts to create a chromosomal *thyA* deletion in *E. coli* were attributable to the presence of the overlapping transcript. Introducing a deletion into the nonoverlapping portion of the cloned *thyA* gene and inserting a gene encoding kanamycin resistance produced a (Δ*thyA*::Km^r) that was easily transferred to the chromosome of a *recD* host by marker replacement. This Δ*thyA*::Km^r allele provides a useful and readily transducible chromosomal marker.

Thymidylate synthase (TS) (EC 2.1.1.45), which converts dUMP to TMP, is present in low amounts in *Escherichia coli*, there being only about 250 molecules per actively growing cell (6). Although its regulation is poorly understood, it has been demonstrated that both thymine starvation and excess are mutagenic in bacteria and in eucaryotic cells (3, 11, 12, 19, 20). The mutagenic effect of thymine deprivation is apparently due to the high rate of dUMP incorporation into DNA during replication. It is likely that thymine excess causes imbalances of the nucleotide pools at the replication fork and thereby impairs the fidelity of DNA replication (20). For these reasons, the precise regulation of *thyA* expression is believed to be essential for maintaining a balanced supply of TMP.

The nucleotide sequence of the *E. coli thyA* gene has revealed a number of features that may influence its expression (6). First, the Shine-Dalgarno (SD) sequence (GAGGA) is situated only 3 nucleotides (nt) upstream from the AUG start codon (Fig. 1C). This close proximity is found only rarely in the *E. coli* genome (the usual spacing being 6 to 8 nt) and may be a means to influence the translational efficiency of *thyA* mRNA (16). Second, there is an open reading frame immediately 5′ to the *thyA* structural gene, with its UGA translational stop codon overlapping the *thyA* SD sequence (Fig. 1C). Codon usage analysis indicated that this open reading frame is likely to be functional (17). Third, 60 bp downstream from the *thyA* translation start site is a sequence characteristic of a rho-independent transcription terminator, with a G-C-rich region of dyad symmetry followed by consecutive U residues (Fig. 1C and E). These observations point to the presence of a gene 5′ to *thyA*, with the 3′ end of its transcript terminating within the *thyA* structural gene. Indeed the essential *umpA* (unidentified membrane protein) gene has recently been shown to be located immediately upstream of *thyA*, with its stop codon overlapping the *thyA* SD sequence (Fig. 1C and D) (31, 32). This gene arrangement creates the potential for another level of regulation of *thyA* expression, as is evidenced by a reduced level of TS when *umpA* expression is impaired (31, 32).

In this communication we demonstrate the existence of multiple transcripts extending into the *thyA* gene and show that the transcription terminator within *thyA* coding sequences is active in vivo. In addition, these studies facilitated construction of a viable gene with a *thyA* deletion marked by a Km^r insert (Δ*thyA*::Km^r), an accomplishment that has been elusive in the past (7, 10). The Δ*thyA*::Km^r allele provides a useful and stable marker that is readily transducible into different strain backgrounds, facilitating a variety of genetic studies.

MATERIALS AND METHODS

Bacterial strains. *E. coli* cells used in this study include the *recA* strain Rue10, a *thyA* derivative of HB101 (from R. Lester) (26), and UC6183 (*galK rpsL*) (from D. Wulff), the host for galactokinase assays. The N$^+$/N$^-$ λ lysogens used were 6405 [C600 *thr leu pro bio uvrB* r_K^- m_K^+ (λ *c*I857 *N*$^+$ Δ*Bam*HI)] and 6590 [C600 *thr leu pro bio uvrB lacZ*::Tn*10* r_K^- m_K^+ (λ *c*I857 *N* Δ*Bam*HI*bioT76*)], hereafter referred to as 6405(N$^+$) and 6590(N$^-$), respectively (from M. Gottesman). Strain V355 (*recD1014*) (from G. Walker) was used for chromosomal marker replacement (27) to generate V355K5 and V355K6, which carry the Δ*thyA*::Km^r allele (this work). Strains AT713 (F$^-$ *thi*-1 *argA21 cysC43 lysA22 mtl-2 xyl-7 malA1 rpsL104* λ^r λ$^-$ *supE44*) (CGSC 4529 from B. Bachmann) and NK5992 (W3110 *argA*::Tn*10*) (from N. Kleckner) were used as *argA* recipients in cotransduction experiments.

Plasmids. Plasmids include pBTAH1.2, which is pBR322 carrying a 1,164-nt *Hin*dIII fragment (referred to interchangeably as a 1.2-kb fragment) containing the *thyA* gene, and pBTA7.8, which is pBR322 with a 7.8-kb *thyA* fragment

* Corresponding author.

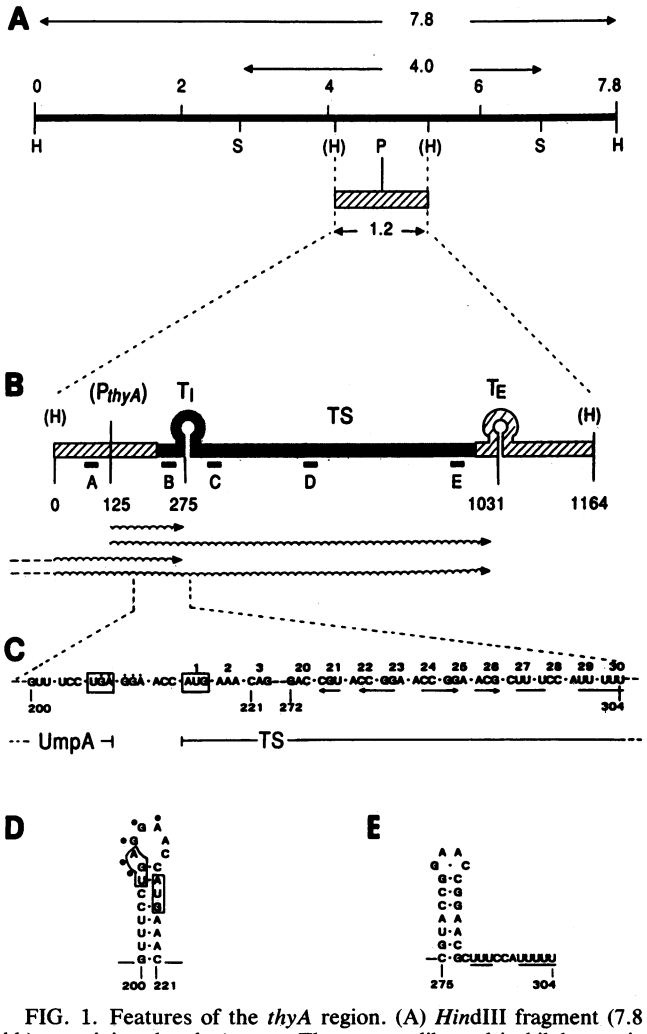

FIG. 1. Features of the *thyA* region. (A) *Hin*dIII fragment (7.8 kb) containing the *thyA* gene. The map, calibrated in kilobases, is shown with relevant restriction enzyme sites. Abbreviations: H, *Hin*dIII; (H), H sites created with linkers; S, *Sma*I. The *thyA* gene is contained on the 1.2-kb (1,164-nt) *Hin*dIII fragment (▨), which contains a unique *Pvu*II (P) site, used to generate an internal deletion (see Fig. 6). (B) Transcription signals of the 1.2-kb fragment. (P_{thyA}) indicates the putative *thyA* promoter (see text), whereas T_I and T_E represent the internal and external *thyA* transcription terminators, respectively. These are indicated in the context of the TS coding region (■) with nucleotide numbers below the map. Oligonucleotides A through E are shown under the map, with natural transcripts in the *thyA* region depicted (⟿). Transcripts originating at an upstream promoter, likely to be P_{umpA}, (----) are shown. (C) 5′ region of the *thyA* gene. The relevant amino acid residues of TS are numbered above the sequence, and the nucleotides are numbered below the sequence. The *thyA* ribosome-binding site, GAGGA, is shown (· · ·) with an overlapping UGA stop codon (boxed) at the 3′ end of *umpA*. The proposed transcription terminator within the *thyA* coding region (T_I) begins at nt 275, with arrows below the sequence identifying the inverted repeats followed by two strings of U residues (underlined). (D) Potential RNA secondary structure at the 3′ *umpA-thyA* translation start region. The *thyA* ribosome-binding site (· · ·) and the initiation codon and *umpA* termination codon (both boxed) are shown. Residue numbers correspond to those used in panel C. (E) Hairpin structure representing T_I. Numbering of residues is as in panel C.

containing an additional 4 kb of DNA 5′ to the *thyA* gene (including *umpA*) and 2.6 kb 3′ to the *thyA* gene (Fig. 1A) (5). Terminator probe plasmid pKG1900 (22) and its *thyA*-containing derivatives pKG364 and pKG364R were used in [^{14}C]galactokinase assays and are described in the legend to Fig. 2. Plasmid pKTAH1.2, which was used in the S1 oligonucleotide-protection experiment, is pKC30 containing the 1.2-kb *thyA* fragment in transcriptional alignment with the p_L promoter (5). M13TAH1.2 contains the 1.2-kb *thyA* fragment cloned into the *Hin*dIII site of M13mp9.

Plasmid pKOTAS4 carries the *thyA* gene on a 4-kb *Sma*I fragment (Fig. 1A), inserted into the *Sma*I site of pKO100 (25). This plasmid was linearized at the unique *Pvu*II site within *thyA*, digested with BAL 31 nuclease, and religated with *Xho*I linkers as previously described (6). One of these plasmids, pKOTAΔ283, had lost 283 bp between positions 466 and 750 of the *thyA* gene (numbering as in reference 6), as determined by DNA sequence analysis. This deletion, which abolishes the Thy$^+$ phenotype, does not interfere with the overlapping transcript (Fig. 1 and 6).

The kanamycin resistance gene was obtained on a 1,252-bp *Sal*I fragment from pUC4K (28, 30). After ligation into the compatible *Xho*I site of pKOTAΔ283, Ap^r Km^r transformants were screened by restriction enzyme analysis. Plasmids pKOTAΔ283K5 and pKOTAΔ283K6 contained the insert in opposite orientations, with the transcriptional orientation of the Km^r gene identical to that of the *thyA* gene in pKOTAΔ283K6. These plasmids were used for allele replacement studies to transfer the Δ*thyA*::Km^r marker to the chromosome (see Results).

Transduction of the Δ*thyA*::Km^r allele into *recA* recipients. To facilitate transfer of the Δ*thyA*::Km^r allele to *recA* recipients, which are not readily transduced by P1, the following strategy was employed. A pBR322-derived Tc^r *recA*$^+$ plasmid (pDR1453, provided by C. Radding) was transformed into strain Rue11, a Thy$^+$ revertant of *recA* host strain Rue10 (26) (provided by R. Lester). This Rec$^+$ transformant was easily transduced by P1 to Δ*thyA*::Km^r. To cure the cells of the Tc^r *recA*$^+$ plasmid, the medium described by Bochner et al. (8) containing 6 μg of fusaric acid per ml was used to select the cells which had become Tc^s. Cells that were Tc^s and UV sensitive had been cured of the *recA*$^+$ plasmid and maintained the chromosomal *recA* marker.

Growth media. Cells were grown in TBYE (1% tryptone [Difco], 0.5% NaCl, 0.5% yeast extract) and supplemented, when required, with 50 μg of thymine per ml. When cells harbored plasmids, ampicillin was added to a concentration of 50 μg/ml in liquid culture and to 100 μg/ml in solid medium containing 1.5% Bacto-Agar.

Enzymes, biochemicals, and radiochemicals. Restriction enzymes were purchased from Bethesda Research Laboratories or New England BioLabs and used according to manufacturers' instructions. The Klenow fragment of DNA polymerase I and S1 nuclease were supplied by Bethesda Research Laboratories, and T4 DNA ligase was supplied by New England BioLabs. T4 polynucleotide kinase was obtained from Boehringer Mannheim, and reverse transcriptase was from Life Sciences Inc. Radiochemicals included [α-^{32}P]dATP (3,000 Ci/mmol) and D-[1-^{14}C]galactose (40 to 60 mCi/mmol) from Amersham and [γ-^{32}P]ATP (3,000 Ci/mmol) from New England Nuclear.

Oligonucleotides. Oligonucleotides were synthesized on an Applied Biosystems automated DNA synthesizer. Those used for RNA analysis were purified from a 20% acrylamide–7 M urea sequencing gel. The oligonucleotides, which are complementary to *thyA* mRNA, contained the following

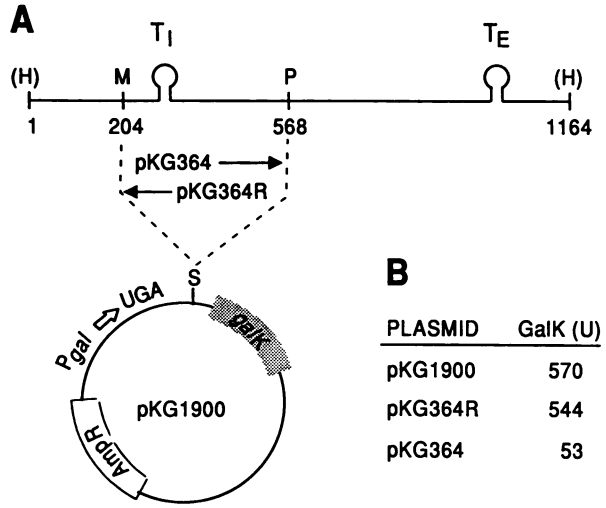

FIG. 2. Test of termination function. (A) Cloning of the putative terminator sequences into pKG1900. A 364-bp *Mst*II-*Pvu*II fragment containing the putative terminator was inserted in both transcriptional (pKG364) and reverse (pKG364R) orientations into the *Sma*I site of pKG1900 between P_{gal} and the *galK* structural gene. The resulting plasmids were used for galactokinase measurements in UC6183 to assess termination events. Abbreviations: (H), *Hin*dIII (linker site); M, *Mst*II; P, *Pvu*II; S, *Sma*I. (B) Galactokinase assays. Galactokinase units are expressed as nanomoles of [^{14}C]galactose phosphorylated per minute per milliliter (10^8 cells per ml), in an assay described by McKenney et al. (22).

parts of the 1,164-nt sequence (6): nt 67 to 85 (A), nt 245 to 265 (B), nt 368 to 388 (C), nt 525 to 544 (D), and nt 860 to 880 (E) (Fig. 1B).

S1 nuclease oligonucleotide protection. The S1 nuclease oligonucleotide protection assay was performed as previously described (4). Briefly, RNA (50 μg) extracted from strain 6590(N^+) carrying pKTAH1.2 and strain 6405(N^-) carrying pKTAH1.2 was mixed with 80 pmol of end-labeled oligonucleotide. After ethanol precipitation, the pellets were resuspended in 30 μl of a solution of 40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES) (pH 6.4), 1 mM EDTA, and 0.4 M NaCl and incubated at 49°C for 30 min. Following the addition of 300 μl of S1 buffer (0.25 M NaCl, 50 mM sodium acetate [pH 5.4], 4 mM $ZnSO_4$, 5% glycerol, 20 μg of single-stranded DNA per ml) containing 100 U of S1 nuclease per ml, the reactions were incubated for 30 min at 23°C. Samples were reprecipitated and separated on a 20% acrylamide–7 M urea sequencing gel.

RNA methods. RNA extraction, hybridization, and primer extension analyses were performed as previously described (4).

RESULTS

Transcription termination by *thyA* coding sequences. To test functionality of the putative transcription terminator (T_I) within the *thyA* structural gene, a 364-nt fragment containing the relevant sequence was inserted between the *gal* promoter and the *galK* structural gene in plasmid pKG1900 (Fig. 2). Expression of *galK* was reduced 10-fold when the 364-bp test fragment was inserted in its normal transcriptional orientation (pKG364) in relation to the *gal* promoter. This suggests that 90% of transcription initiated at the *gal* promoter terminated in the inserted sequence rather than entering the galactokinase gene. When the same DNA sequence was inserted into pKG1900 in the opposite orientation (pKG364R), no appreciable termination was detected, implying that the insertion of 364 nt between the promoter and the *galK* coding sequence does not directly affect *galK* expression. These data support the hypothesis that the putative terminator sequences within the coding region of the *thyA* gene can function in vivo.

Multiple transcripts in the *thyA* region. Plasmids pBTAH-1.2, containing the *thyA* gene, and pBTA7.8, harboring *thyA*, *umpA*, and neighboring sequences, were utilized for RNA hybridization analysis to characterize the mRNA species in the *umpA-thyA* region. Total RNA extracted from cells containing either pBTAH1.2 or pBTA7.8 was probed (Fig. 3A, lanes 1 and 2, respectively) with a whole gene probe (the 1.2-kb *Hin*dIII fragment containing the entire *thyA* gene [Fig. 1A and B]). Two bands that appear in common to pBTAH1.2 and pBTA7.8 are of ca. 1 kb (band c) and ca. 0.2 to 0.3 kb (band d). The 1-kb band corresponds well in size (0.93 kb) to that of a transcript originating at the putative *thyA* promoter (see below) and terminating at T_E, the external *thyA* terminator. (Length estimates in parentheses were derived from *thyA* sequence data [6], the 5′ ends of transcripts inferred from primer extension analysis [Fig. 4], and the presumed termination sites at the consecutive T residues of both terminators.) The smaller band approximates the size of a prematurely terminated transcript originating from the putative P_{thyA} promoter and terminating at T_I within *thyA* (0.17 kb).

Two additional transcripts with sizes of ca. 1.1 kb (band e) and 0.3 to 0.4 kb (band f) are unique to pBTAH1.2 (Fig. 3A, lane 1). These are in the size range of RNAs originating from a promoter within the tetracycline resistance gene of pBR322 (9) and terminating at T_E (1.05 kb) or at T_I within *thyA* (0.30 kb), respectively (Fig. 3B and C). Transcription begins near the *Hin*dIII site of the 1.2-kb *thyA* fragment (see also Fig. 4), in the sense opposite to that of the P_{tet} promoter (9). Bands e and f are absent from pBTA7.8, in which the *thyA* gene is remote from this pBR322 promoter (Fig. 3A, lane 2).

Two additional bands are unique to pBTA7.8, which contains the entire *umpA* sequence on its 7.8-kb insert (Fig. 3A, lane 2). A 1.4-kb species (band b) originates at an upstream promoter, which most likely corresponds to P_{umpA}, and terminates within the *thyA* structural gene at T_I. A 2.0-kb species (band a) is thought to originate at P_{umpA}, read through T_I, and end at T_E. This 2.0-kb transcript, which is absent when upstream sequences are missing (pBTAH1.2) therefore represents a readthrough polycistronic message.

The identification of bands a through f was corroborated with oligonucleotide probes A, B, and C, which hybridize to upstream non-*thyA* sequences (A), to *thyA* sequences downstream of T_I (C), or to both (B). The data are summarized in tabular form in Fig. 3C. Thus, only transcripts that extend beyond T_I (a, c, and e) were evident with probe C, whereas only transcripts originating upstream of P_{thyA} (a, b, e, and f) were apparent with probe A. The hybridization to these latter oligonucleotide probes complementary to the mRNA also shows directionality of the transcripts, confirming that they are all synthesized from the same DNA strand.

Transcript mapping. To further demonstrate the existence of overlapping transcripts and to map the 5′ end of the monocistronic *thyA* mRNA, primer extension analyses were undertaken (Fig. 4). A reverse transcriptase stop was evident at adenine 125 (numbering of residues as in reference 6) in cDNA extended from an oligonucleotide complementary to the overlapping region (Fig. 1 and 3, probe B), using

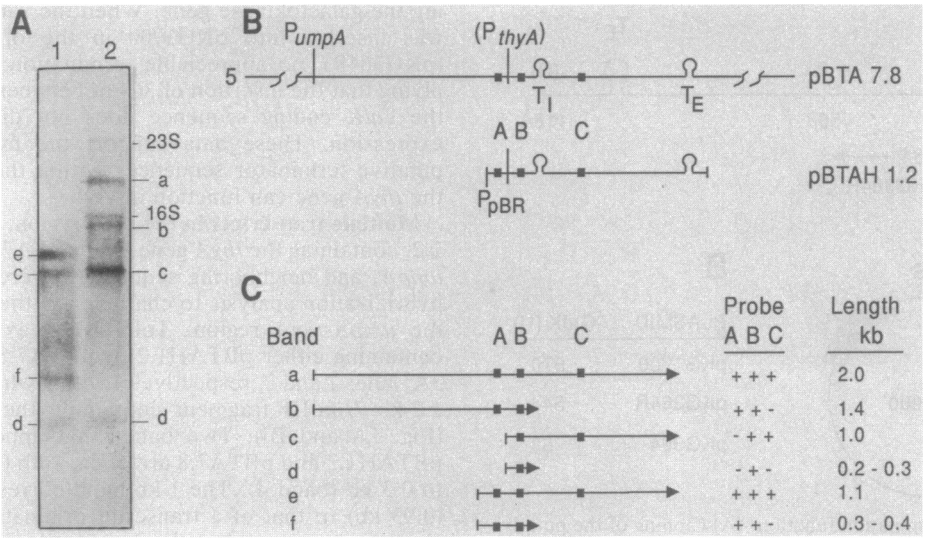

FIG. 3. Hybridization analysis of transcripts in the *thyA* region. (A) RNA blot with whole gene probe. Total RNA from *E. coli* Rue10 containing plasmid pBTAH1.2 (lane 1) or pBTA7.8 (lane 2) was size separated and transferred to a nitrocellulose filter. Hybridization was done with the nick-translated 1.2-kb *Hin*dIII *thyA* fragment. The DNA size markers were 7.6, 4, 1.2, 0.8, and 0.4 kb (data not shown). Bands a to f are labeled according to the schematic in panel C, while 23S and 16S represent rRNA. No well-resolved *thyA*-specific transcripts from the chromosome were detectable in untransformed cells. (B) Maps of pBTA7.8 and pBTAH1.2. Sites at which oligonucleotides A, B, and C anneal are indicated (■). (C) Transcript identification. Bands were identified by using end-labeled oligonucleotide probes A, B, and C, with the results tabulated (+, band present; −, band absent). The lengths were determined from multiple autoradiograms. The pBR322-promoted transcripts start at P_{pBR} near the *Hin*dIII site and are detected only from pBTAH1.2 (see text).

pBTAH1.2 RNA as the template (Fig. 4A). A125 is 90 nt upstream of the *thyA* initiation codon. A second stop, mapping to within the *Hin*dIII site of the 1.2-kb fragment, was detected higher up on the gel, in agreement with the hybridization analysis which revealed promoter activity from within the tetracycline resistance gene of pBR322 (Fig. 3).

A comparison of cDNAs synthesized from pBTAH1.2 and pBTA7.8 RNA showed that the A125 band is present in both cases, as expected if it represents the 5′ end of the monocistronic *thyA* transcript (Fig. 4B). Predictably, the stop associated with the external promoter at the *Hin*dIII site is seen only for pBTAH1.2. Furthermore, cDNA products made from pBTA7.8 RNA alone continued past the stops from pBTAH1.2 (Fig. 4B). This cDNA reflects sequences from *umpA* mRNA that extend into the *thyA* structural gene.

Additional primer extension experiments ruled out the proposed proximal −35 and −10 polymerase recognition sites that were previously suggested for *thyA* (6), while they corroborated a 5′ end of the *thyA* monocistronic transcript corresponding to A125. These data agree well with the size of the *thyA* message determined by hybridization analysis (Fig. 3), but we have not eliminated the possibility that the 5′ end represents a processing site, rather than a transcription start site.

Antitermination studies. To further investigate the effect of premature termination on *thyA* transcription, the effect of antitermination mediated by the phage lambda N gene product was measured (24). N protein was supplied from a prophage to effect antitermination from p_L-promoted *thyA* transcripts in plasmid pKC30 (construct pKTAH1.2). To examine the effect of T_I on *thyA* mRNA synthesis, S1 nuclease protection analysis was performed with oligonucleotides B and C (Fig. 5) (4). Total RNA extracted from strain 6405(N^+) carrying pKTAH1.2 and strain 6590(N^-) carrying pKTAH1.2 was hybridized to an excess of either oligonucleotide B or C, which anneal to sequences upstream and downstream of T_I, respectively. The amount of oligonucleotide protected was estimated in two independent experiments by a direct determination of radioactivity from equivalent-sized gel samples or by densitometric scans of the autoradiograms. The quantity of protected oligomer B was essentially the same with RNA from N^+ and N^- lysogens, consistent with annealing of this oligomer to *thyA* RNA sequences 5′ to T_I. In contrast, when oligonucleotide C, which anneals 3′ to T_I, was used, a ca. 40% increase in full-length *thyA* transcripts was observed in the N^+ over the N^- strain. This increase is presumably due to N-mediated antitermination and suggests that the terminator reduces the amount of full-length *thyA* mRNA by 40%, if one assumes that N is 100% efficient at this terminator. These experiments also provide direct evidence that T_I functions as a termination site, rather than an RNA processing site.

Generating a chromosomal *thyA* deletion. We reasoned that the presence of the overlapping transcript may have impeded attempts to generate complete *thyA* deletions in the *E. coli* chromosome (7, 10). The *umpA* gene, which has been identified upstream of *thyA* and appears to encode the overlapping transcript, has been shown to be essential for viability (32). We therefore sought to generate an internal *thyA* deletion without disrupting the overlapping transcript. To this end, we generated a 283-nt deletion with BAL 31 nuclease at the *Pvu*II site of the *thyA* gene in plasmid pKOTAS4, to yield pKOTAΔ283. A 1.2-kb fragment encoding kanamycin resistance was cloned in both orientations at the *Xho*I site of the deletion joint in pKOTAΔ283 to generate plasmids pKOTAΔ283K5 and pKOTAΔ283K6 (Fig. 6A).

To move the Δ*thyA*::Km^r construct into the chromosome, we used the approach of Shevell et al. (27). *E. coli* strains that are *recD1014* lack exonuclease V activity and can

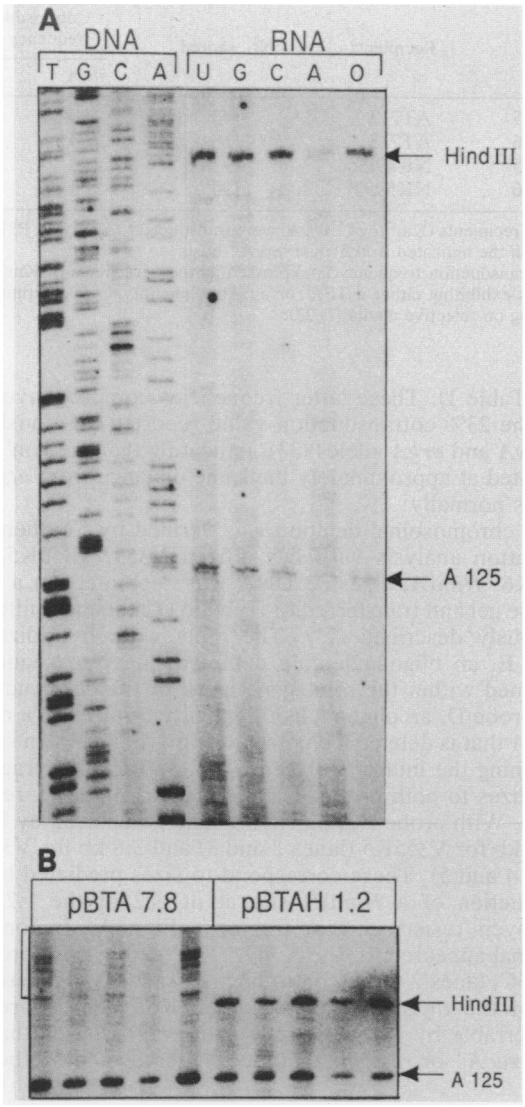

FIG. 4. Primer extension analysis of *thyA* transcripts. (A) Assignment of the 5′ end of monocistronic *thyA* mRNA. Total RNA from Rue10(pBTAH1.2) served as the template for reverse transcription from end-labeled oligonucleotide B (RNA lanes). Dideoxy sequencing of single-stranded DNA (M13TAH1.2) was conducted simultaneously (DNA lanes). T or U, G, C and A represent the complement of the dideoxynucleoside triphosphate added to either the sequencing (DNA) or primer extension (RNA) reactions, with 0 indicating no dideoxynucleoside triphosphate addition. The DNA sequence was used to determine the apparent 5′ end of *thyA* mRNA, seen as a reverse transcriptase stop in all the RNA lanes at nucleotide A125. A cDNA stop is also apparent near the top of the gel, at the *Hin*dIII site of the 1,164-nt fragment-vector junction. The promoter within the tetracycline gene of pBR322, which initiates transcription in the orientation opposite to that of P_{tet}, is responsible for the transcript that directs this cDNA product (see text). (B) Comparison of RNA from pBTA7.8 and pBTAH1.2. Primer extension of oligonucleotide B reveals the reverse transcriptase stop at A125 for both plasmids and the stop that maps to the *Hin*dIII site in pBTAH1.2 only. The bracket on the left identifies cDNAs in pBTA7.8 that are absent from pBTAH1.2 and that represent sequences of the overlapping transcript that extend into *thyA*.

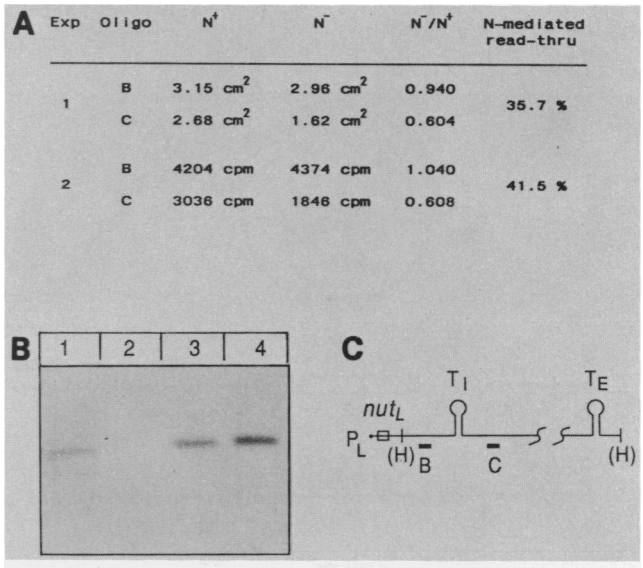

FIG. 5. S1 nuclease oligonucleotide protection by RNA isolated from N^+ and N^- λ lysogens. (A) N-mediated readthrough. Total RNA from strain 6590(N^-) carrying pKTAH1.2 or strain 6405(N^+) carrying pKTAH1.2 was annealed to oligonucleotide B or C. After digestion with S1 nuclease, the protected oligomers were separated on a 20% acrylamide–7 M urea sequencing gel. Relative band intensities from two separate experiments are shown in this panel. The data from experiment 1 were quantitated by measuring the area under the peak from densitometric scans of the autoradiogram, whereas those from experiment 2 were quantitated by direct radioactivity measurements from equivalent-sized gel slices. N-dependent readthrough represents the percentage increase of full-length *thyA* transcripts under N^+ relative to N^- conditions. (B) Representative autoradiogram. Bands represent purified oligonucleotide B (lane 1), S1 nuclease-digested oligonucleotide B (lane 2), oligonucleotide B protected with RNA extracted from strain 6590(N^-) carrying pKTAH1.2 (lane 3) or from strain 6405(N^+) carrying pKTAH1.2 (lane 4). (C) Diagrammatic representation of hybridization sites of oligonucleotides B and C on p_L-initiated transcripts. The N utilization site, nut_L, is shown in relation to the 1.2-kb *Hin*dIII (H) *thyA* fragment in pKTAH1.2 (6).

therefore be transformed with linear DNA (27). Since linear molecules do not replicate, only markers transferred to the chromosome by recombination can be expressed. Plasmids pKOTAΔ283K5 and pKOTAΔ283K6 were digested with *Eco*RI to excise Δ*thyA*::Km^r and adjacent sequences on a 5.6-kb fragment and transformed into strain V355 (*recD1014*) (Fig. 6A). Of the Km^r transformants from pKOTAΔ283K5 and pKOTAΔ283K6, 96 of 107 and 76 of 76, respectively, were Thy^- and Ap^s (11 of 107 were Thy^+ Ap^r and presumably contained plasmid). One Δ*thyA*::Km^r Ap^s strain derived from each plasmid was used in subsequent experiments. These strains are called V355K5 and V355K6.

Genetic and molecular analysis of the chromosomal Δ*thyA*::Km^r allele. To confirm that the chromosomal *thyA* gene had been replaced by the Δ*thyA*::Km^r construct, we first measured cotransduction frequencies between Km^r, *thyA*, and a linked marker, *argA* (Fig. 6A; see reference 13 for a correlation of the genetic and physical map of the *thyA* and *argA* region). P1 lysates prepared on V355K5 or V355K6 were used to transduce one of two *argA* recipients, *E. coli* NK5992 or *E. coli* AT713. All (100%) of the Km^r transductants were Thy^-, while Km^r-*argA* cotransduction frequencies for both constructs in both strains were similar at 15 to

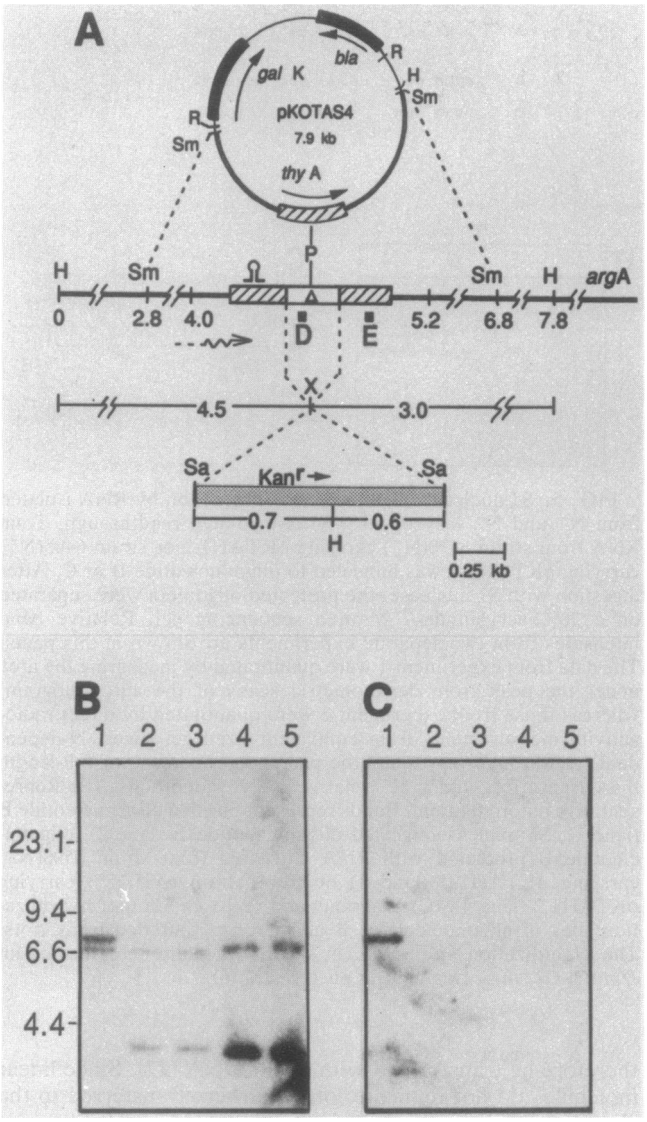

FIG. 6. Chromosomal Δ*thyA*::Km^r allele. (A) Construction of Δ*thyA*::Km^r. Plasmid pKOTAS4 contains the *thyA* gene (▨) on a 4-kb *Sma*I fragment (—) oriented as shown relative to the *galK* and *bla* genes. The extent of the *thyA* deletion in pKOTAΔ283 and the 3′ end of the overlapping transcript are indicated below the plasmid on a linear map of the 7.8-kb *thyA* fragment. The *Sal*I Km^r fragment was inserted at the *Xho*I site of the *thyA* deletion. The positions of probes D and E are shown (■). Abbreviations: H, *Hin*dIII; R, *Eco*RI; Sm, *Sma*I; Sa, *Sal*I; X, *Xho*I; P, *Pvu*II. (B and C) Chromosomal DNA blots of Δ*thyA*::Km^r-containing strains. *Hin*dIII digests of total *E. coli* DNA were blotted and probed with probes E (B) and D (C). The position of markers is shown on the left of panel B. Lanes 1, V355 DNA (parental strain); lanes 2 and 3, V355K5 DNA (two independent isolates); lanes 4 and 5, V355K6 DNA (two independent isolates). The 6.7-kb band present in all lanes in panel B represents nonspecific hybridization to probe E. Identical blots on strains containing the intact *thyA* gene plus the plasmid pKOTAΔ283K5 or pKOTAΔ283K6 showed that no plasmid is present in V355K5 or V355K6 (data not shown).

TABLE 1. Cotransduction frequencies of Km^r with *thyA* and *argA*

Donor	Recipient[a]	No. scored	Cotransduction frequency (%)[b]	
			thyA	*argA*
V355K5	AT713	143	100	15
V355K6	AT713	118	100	15
V355K5	NK5992	177	100	18
V355K6	NK5992	194	100	17

[a] The recipients (Km^s *thyA*$^+$ *argA*) were transduced to Km^r with P1 lysates grown on the indicated Δ*thyA*::Km^r-*argA*$^+$ hosts.

[b] Cotransduction frequency is expressed as the percentage of Km^r transductants exhibiting either a Thy$^-$ or an Arg$^+$ phenotype as determined by screening on selective media (7, 23).

18% (Table 1). These latter frequencies compare favorably with the 23% cotransduction value reported previously for the *thyA* and *argA* alleles (33), indicating that the Km^r gene is located at approximately the same distance from *argA* as *thyA* is normally.

The chromosomal deletion was verified by Southern hybridization analysis with DNA from V355, V355K5, and V355K6. *Hin*dIII-digested DNA was separated on a 0.7% agarose gel and transferred to Hybond (Amersham) filters as previously described (21). The blots were hybridized to probe E, an oligonucleotide complementary to sequences contained within the *thyA* gene in all of the constructs, or with probe D, an oligonucleotide that hybridizes to a region of *thyA* that is deleted from Δ*thyA*::Km^r (Fig. 6A). In strains containing the intact *thyA* gene, a 7.8-kb *Hin*dIII fragment hybridizes to both probes E and D (Fig. 6B and C, respectively). With probe E, this 7.8-kb band is replaced by bands of 3.9 kb for V355K5 (lanes 2 and 3) and 3.8 kb for V355K6 (lanes 4 and 5). These correspond to sizes predicted by the introduction of a *Hin*dIII site at nt 682 of the 1,252-bp kanamycin resistance gene fragment (Fig. 6A). In contrast, no signal appeared for the Δ*thyA*::Km^r strains with probe D (Fig. 6C, lanes 2 to 5), confirming the successful deletion of part of the *thyA* gene. The Δ*thyA*::Km^r allele is readily transferrable by P1 transduction to *recD*$^+$ strains that are either *recA*$^+$ or *recA* mutant (see Materials and Methods).

DISCUSSION

The results described above demonstrate the presence of a transcription terminator within the coding region of the *thyA* gene. A significant fraction of monocistronic *thyA* messages and polycistronic transcripts originating upstream of *thyA* appear to be terminated at this site. Antitermination studies argue against T_I being an RNA processing site and show that the terminator within the *thyA* gene reduces the amount of full-length *thyA* message by ca. 40%. A similar termination efficiency may be inferred from band intensities in Northern hybridization (RNA blot) analysis, while experiments with the *galK* vector pKG1900 suggest 90% termination. The efficiency of termination in this latter system likely reflects the lack of translation across the terminator in this plasmid construct (note the stop codon preceding the *thyA* insert in Fig. 2), as opposed to active translation at the time of RNA extraction in both the other experiments. Alternatively, the different termination efficiencies may be an effect of the different promoters (p_L versus P_{gal}) driving transcription (29). In either event, all these experiments support the presence of a functional terminator, which reduces the

production of full-length TS mRNA while it serves as a punctuation signal at the end of the upstream overlapping transcript.

In light of the apparent 5′ end of the monocistronic *thyA* transcript mapping to A125, we sought promoter sites that most closely resemble canonical −35 (TTGACA) and −10 (TATAAT) sequences (18). The best-match −35 and −10 sequences near A125 are TTTACC at nt 85 to 90, separated by 15 nt from TACATC at nt 96 to 91 on the 1.2-kb fragment (6). Identities to the consensus sequences are underlined. The deviations from the consensus sequence and from the optimal spacing of 17 nt would suggest that at best this is a weak transcription initiation signal (2). Such an apparently weak promoter, along with a surprisingly long leader region (90 nt) may be a consequence of the sequence constraints imposed by *umpA*. Alternatively, the deviations from the consensus sequence may reflect an unusual class of promoters that is regulated by some yet undefined mechanism. Yet another possibility is that the 5′ end at A125 reflects a processing site. This interpretation would, however, not account for *umpA*-independent promoter activity of the *thyA* gene (32; also our unpublished results).

The RNA analyses (Fig. 2 and 3) indicate multiple *thyA* transcripts and suggest that the *thyA* gene is expressed from both monocistronic and polycistronic mRNAs. The transcript originating upstream of *thyA* apparently extends from the *umpA* promoter into the *thyA* gene (31). The overlap between the stop codon of *umpA* and the putative SD sequence of *thyA* (Fig. 1C) raises the possibility of translational coupling (15). A potential stem-loop containing the *thyA* translation-initiation signals further suggests that such coupling may have regulatory importance for *thyA* expression from the polycistronic RNA (Fig. 1D).

It would appear, therefore, that there is a potential for regulation of *thyA* expression at four different levels: initiation of transcription at the upstream polycistronic promoter, P_{umpA}; initiation of transcription or processing of the monocistronic *thyA* transcript; attenuation of transcription at the T_I terminator; and translational coupling with *umpA* in the polycistronic message. It will be intriguing to unravel the complex interplay of these potential regulatory mechanisms. Undoubtedly, their delicate balance is required to boost TS levels in response to rapid cell division and elevated DNA replication, while preventing a large TS surplus that would result in the mutagenic condition of TMP excess.

The existence of an overlapping transcript may provide insight into some unusual observations previously associated with *thyA*. A number of temperature-sensitive mutations were found to cluster at the 5′ region of the *thyA* gene (1, 7) and may actually be a consequence of the overlapping transcript. That is, the temperature-sensitive phenotype may reflect altered properties of the *umpA* transcript (e.g., stability), rather than being related to the mutation in *thyA* per se. Furthermore, since the overlapping transcript appears to code for the vital *umpA* product (32), this might explain why past attempts to construct chromosomal *thyA* deletions, which extended into the 5′ region of *thyA*, were frustrated (7, 10). The successful isolation of the chromosomal *thyA*::Km^r marker, which was constructed with the intent of leaving the *umpA* transcript intact, supports this hypothesis. Additionally, the Δ*thyA*::Km^r allele is useful for transferring a stable, marked Δ*thyA* construct into different genetic backgrounds. Such strains will facilitate the cloning of heterologous TS genes that can be expressed in *E. coli*, as well as the selection of TS$^+$ revertants from TS$^-$ clones (see, for example, reference 14), without background reversion of the chromosomal *thyA* marker.

ACKNOWLEDGMENTS

We thank Sidney Kushner and Michael Williams for providing sequences upstream of *thyA* and Sidney Kushner for helpful discussions and useful comments on the manuscript. We thank H. Engelberg-Kulka for suggesting the use of the Rec$^+$ plasmid; Barbara Bachmann, Max Gottesman, Nancy Kleckner, Charles Radding, Graham Walker, and Dan Wulff for providing strains and plasmids; Renee Schroeder for curing the Rue11 Δ*thyA*::Km^r strain; and Maryellen Carl and Carolyn Wieland for meticulous preparation of the manuscript.

This work was supported by NSF grant DMB8502961 and NIH grants GM39422 and GM44844 to M.B. J.L.G.S. was the recipient of an American Cancer Society postdoctoral fellowship.

REFERENCES

1. **Alikhanian, S. I., T. S. Iljina, E. S. Kaliaeva, S. V. Kameneva, and V. V. Suchodolec.** 1966. A genetical study of thymineless mutants of *E. coli* K12. Genet. Res. **8**:83–100.
2. **Aoyama, T., M. Takanami, E. Ohtsuka, Y. Taniyama, R. Marumoto, H. Sato, and M. Ikehara.** 1983. Essential structure of *E. coli* promoter: effect of spacer length between the two consensus sequences on promoter function. Nucleic Acids Res. **11**:5855–5864.
3. **Barclay, B. J., B. A. Kunz, J. G. Little, and R. H. Haynes.** 1982. Genetic and biochemical consequences of thymidylate stress. Can. J. Biochem. **60**:172–194.
4. **Belfort, M., K. Ehrenman, and P. S. Chandry.** 1990. Genetic and molecular analysis of splicing in *Escherichia coli*. Methods Enzymol. **181**:521–539.
5. **Belfort, M., G. F. Maley, and F. Maley.** 1983. Characterization of the *Escherichia coli thyA* gene and its amplified thymidylate synthetase product. Proc. Natl. Acad. Sci. USA **80**:1858–1861.
6. **Belfort, M., G. Maley, J. Pedersen-Lane, and F. Maley.** 1983. Primary structure of the *Escherichia coli thyA* gene and its thymidylate synthase product. Proc. Natl. Acad. Sci. USA **80**:4914–4918.
7. **Belfort, M., and J. Pedersen-Lane.** 1984. Genetic system for analyzing *Escherichia coli* thymidylate synthase. J. Bacteriol. **160**:371–378.
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. **Brosius, J.** 1984. Plasmid vectors for the selection of promoters. Gene **27**:151–160.
10. **Chung, S.-T., and G. R. Greenberg.** 1973. Loss of an essential function of *Escherichia coli* by deletions in the *thyA* region. J. Bacteriol. **116**:1145–1149.
11. **Coughlin, C. A., and E. A. Adelberg.** 1956. Bacterial mutation induced by thymine starvation. Nature (London) **178**:531–532.
12. **DeVries, J. K., and S. S. Wallace.** 1982. Reversion of bacteriophage T4rII mutants by high levels of pyrimidine deoxyribonucleosides. Mol. Gen. Genet. **186**:101–105.
13. **Dykstra, C. C., D. Prasher, and S. R. Kushner.** 1984. Physical and biochemical analysis of the cloned *recB* and *recC* genes of *Escherichia coli* K-12. J. Bacteriol. **157**:21–27.
14. **Galloway Salvo, J. L., T. Coetzee, and M. Belfort.** 1990. Deletion-tolerance and trans-splicing of the bacteriophage T4 *td* intron: analysis of the P6-L6a region. J. Mol. Biol. **211**:537–549.
15. **Gold, L.** 1988. Posttranscriptional regulatory mechanisms in *Escherichia coli*. Annu. Rev. Biochem. **57**:199–233.
16. **Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo.** 1981. Translational initiation in prokaryotes. Annu. Rev. Microbiol. **35**:365–403.
17. **Grosjean, H., and W. Fiers.** 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199–209.
18. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic

Acids Res. **11**:2237–2255.
19. **Kunz, B. A.** 1982. Genetic effects of deoxyribonucleotide pool imbalances. Environ. Mutagen. **4**:695–725.
20. **Kunz, B. A., and B. W. Glickman.** 1985. Mechanism of mutation by thymine starvation in *Escherichia coli*: clues from mutagenic specificity. J. Bacteriol. **162**:859–864.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **McKenney, K., H. Shimatake, D. Court, U. Schmeissner, C. Brady, and M. Rosenberg.** 1981. A system to study promoter and terminator signals recognized by *Escherichia coli* RNA polymerase, p. 383–415. *In* J. G. Chirikjian and T. S. Papas (ed.), Gene amplification and analysis, vol. II. Structural analysis of nucleic acids. Elsevier/North-Holland Publishing Co., New York.
23. **Okada, T., K. Yanagisawa, and F. J. Ryan.** 1961. A method for securing thymineless mutants from strains of *E. coli*. Z. Vererbungsl. **92**:403–412.
24. **Roberts, J. W.** 1988. Phage lambda and the regulation of transcription termination. Cell **52**:5–6.
25. **Rosenberg, M., A. B. Chepelinsky, and K. McKenney.** 1983. Studying promoters and terminators by gene fusion. Science **222**:734–739.
26. **Rubin, E. M., G. A. Wilson, and F. E. Young.** 1980. Expression of thymidylate synthetase activity in *Bacillus subtilis* upon integration of a cloned gene from *Escherichia coli*. Gene **10**:227–235.
27. **Shevell, D. E., A. M. Abou-Zamzam, B. Demple, and G. C. Walker.** 1988. Construction of an *Escherichia coli* K-12 *ada* deletion by gene replacement in a *recD* strain reveals a second methyltransferase that repairs alkylated DNA. J. Bacteriol. **170**:3294–3296.
28. **Taylor, L. A., and R. E. Rose.** 1988. A correction in the nucleotide sequence of the Tn903 kanamycin resistance determinant in pUC4K. Nucleic Acids Res. **16**:358.
29. **Telesnitsky, A. P. W., and M. J. Chamberlin.** 1989. Sequences linked to prokaryotic promoters can affect the efficiency of downstream termination sites. J. Mol. Biol. **205**:315–330.
30. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
31. **Williams, M., M. Aldea, and S. R. Kushner.** Unpublished data.
32. **Williams, M. G., M. Fortson, C. C. Dykstra, P. Jensen, and S. R. Kushner.** 1989. Identification and genetic mapping of the structural gene for an essential *Escherichia coli* membrane protein. J. Bacteriol. **171**:565–568.
33. **Willetts, N. S., and D. W. Mount.** 1969. Genetic analysis of recombination-deficient mutants of *Escherichia coli* K-12 carrying *rec* mutations cotransducible with *thyA*. J. Bacteriol. **100**: 923–934.