Regions specifying transcriptional termination and pausing in the bacteriophage SP01 terminal repeat

Sean M. Brennan* and E. Peter Geiduschek

Department of Biology, B-022 University of California, San Diego, La Jolla, CA 92093, USA

Received 18 January 1983; Revised and Accepted 13 May 1983

ABSTRACT

We have determined the nucleotide sequences of four termination sites recognized by *Bacillus subtilis* RNA polymerase. These sites are located in the terminally repeated segment of the bacteriophage SP01 genome, where most early phage transcription occurs. The SP01 terminators have structures that are similar to those recognized by *Escherichia coli* RNA polymerase, containing a region of dyad symmetry followed by a stretch of HMU residues in the noncoding DNA strand (HMU is substituted for T in SP01 DNA). We note that in a terminator that is only 60% efficient *in vitro*, there is a greater distance between these two conserved elements than exists in more efficient terminators.

We also find that RNA polymerase molecules which elongate a transcript through a partial terminator often pause at this site.

INTRODUCTION

The mechanism of transcription termination has recently been the focus of considerable study. Several common structural features of ρ -independent *E. coli* terminators have been detected (reviewed in 1,2). First, all terminators contain a region of hyphenated dyad symmetry, centered 20 ± 4 nucleotides upstream from the site of termination. A second common feature is a GC-rich region preceding the point of termination by roughly 10 nucleotides. Finally, a string of T residues (usually 4-5 long) is found in the nontranscribed strand immediately following the region of dyad symmetry. Termination occurs heterogeneously across this T-stretch.

Models for the mechanism of termination (1-4) are based primarily on nucleic acid-nucleic acid interactions. It is proposed that the region of the transcript containing the dyad symmetry element adopts a stem-loop conformation, causing the elongating RNA polymerase molecule to pause over the string of T residues immediately downstream. Transcript release is thought to be related to the relative instability of the rU:dA hybrid (5) formed immediately downstream of the dyad symmetry region.

We have been analyzing transcription unit organization in the virulent

Bacillus subtilis bacteriophage SP01. The most heavily-expressed early transcription units are located within the 12.6 kbp terminally redundant segment of the 145 kbp SP01 genome (6). Thirteen promoters are found in this region; transcription from these promoters is overlapping and converges toward a bidirectional termination region (ref. 7; see also Fig. 1). The component of this bidirectional terminator which specifies termination of transcription approaching from the left has been named T_H3 (terminator of heavy strand transcription), while the element responsible for transcripts that enter this region from the right is called T_L1 (terminator of light strand transcription). These assignments were based on the known strand-specificity of transcription in the terminal repeat (7). While termination in both polarities at the bidirectional terminator is over 90% efficient, six partial termination sites, with efficiencies ranging from 10% to 60% in vitro, are found at various sites upstream of the bidirectional terminator.

We were interested to know the structure of the bidirectional terminator in greater detail, particularly whether it consisted of a single site or of two separate terminators. Furthermore, since the SP01 terminal repeat contains a spectrum of terminators ranging in efficiency from 10% to over 90%, it provides an opportunity to ask if structure is related to efficiency. Accordingly, we have compared the sequences of the strong terminators T_H3 and T_L1 to that of T_L2 , a 60% efficient terminator. While all of these sites are similar to terminators recognized by E. coli RNA polymerase in that they contain a region of hyphenated dyad symmetry and a T-stretch in the non-coding strand, the spacing between these two regions differs for partial and efficient terminators.

MATERIALS AND METHODS

A. DNA End-Labeling and Strand-Separation

Labeling at 5' ends was done as described previously (7) except that carrier-free [γ - 32 P]-ATP was used (specific activity ~9000 Ci/mmol, purchased from ICN, Irvine, CA).

Restriction fragments having protruding 5' ends were 3'-labeled by "fill-in" using the large proteolytic fragment (8,9) of E. coli DNA polymerase I. Blunt-ended DNA molecules (produced by AluI digestion) were labeled at their 3' ends by utilizing the combined exonucleolytic and polymerizing activities of T4 DNA polymerase (10,11). Prior to strand separation, labeled restriction fragments were precipitated with ethanol, resuspended and denatured in 80% (v/v) dimethylsulfoxide. Strand separation gels (12) were 5% or

Figure 1. Promoters and terminators in the SPO1 Terminal Repeat.

The top line shows the genomic coordinates of the SPO1 terminal repeat (0-12.6 kb at the left end, 132.4-145 kb at the right end). Below this, a map of promoters (above the line) and terminators (below the line) is given. Transcription is convergent on terminators H3 and L1. (Rightward transcription initiated at P_{E1-8} uses the SPO1 DNA H strand as template and involves terminators designated as T_H . Leftward transcription initiated at P_{E9-12} uses the SPO1 DNA L strand as template and involves terminators designated as T_L . See ref. 7 for further details). The positions of restriction fragments that were used for sequencing or as templates for transcription are shown underneath the promoter map. The enlargement of BstNI/HaeII-550 shows the locations of the three terminators on this fragment; the sizes of transcripts that are terminated at two of these sites are given below.

10% acrylamide with an acrylamide:bisacrylamide ratio of 60:1 and dimensions of 40 x 20 x 0.05 cm. The sample well was 15 cm long; running buffer was 50 mM Tris-Borate, pH 8.3, 1.25 mM EDTA. Gels were run at 4° and prerun at 200V for 0.5-1 hr. As soon as the sample was loaded, 2000-3000 volts were applied for 2 min; the voltage was then reduced to 200-400 volts and electrophoresis was continued for 12-18 hrs.

B. DNA Sequencing

The limited chemical degradation method was used (12). Methylene blue was used as a G-specific reagent (13). All other reactions (A+G, C+T, C) were as described (12), except that occasionally the alternative pyrimidine-specific reactions of Rubin and Schmid (14) were used.

Sequencing gels were either 40 or 80 cm long, 20 cm wide and 0.05 cm thick. They contained 6, 10 or 20% acrylamide (acrylamide to bisacrylamide

ratio = 19:1), 8.3 M urea, 100 mM Tris-Borate, pH 8.3, 2.5 mM EDTA. Electrophoresis was conducted at 2000-4000 volts.

C. S1 Protection Analysis of RNA-DNA Hybrids

Hybrids were formed by incubating RNA and single-stranded, end-labeled DNA in 10 mM PIPES, pH 7, 0.1 mM EDTA, 0.3 M NaCl either at 70° for 3 hours or at 60° overnight. RNA and DNA were first combined in the absence of other components, boiled for 2 min, then the other components were added to a final volume of 20 µl. A typical hybridization contained several thousand cpm of ³²P, corresponding to 1-2 fmole of DNA molecules. Amounts of RNA are indicated in the figure legends.

For S1 digestion, the hybridization mixtures were diluted into 180 µl of 30 mM sodium acetate, pH 5.2, 250 mM NaCl, 1 mM ZnCl₂, 20 µg/ml denatured calf thymus DNA. Twenty units of S1 nuclease (Sigma Chemical Co.) was added and digestion was allowed to occur for 1 hr at 37°. Digestion was stopped by adding EDTA to 5 mM and the DNA was ethanol-precipitated. Samples that contained labeled RNA were resuspended in 0.1 ml of 0.1 M NaOH, boiled 5 min, neutralized and ethanol-precipitated before resuspending in sample buffer. Electrophoresis was performed on the same type of gel that was used for DNA sequencing.

For the mapping at T_L2 (Fig. 6), RNA-DNA hybrids were formed under R-loop conditions (15-17) using a double-stranded DNA probe labeled at one end. RNA and DNA were mixed, precipitated with ethanol, washed with absolute ethanol and resuspended in 20 µl of 80% formamide (recrystallized), 0.4 M NaCl, 50 mM PIPES, pH 7, 1 mM EDTA. Samples were heated to 70° for 10 min, then incubated at 42° for 4 hr. At the conclusion of the incubation, they were diluted into ice-cold S1 digestion mix (noted above).

In vitro RNA was purified as described (18). In vivo RNA was purified either by hot phenol extraction (19) or by CsCl centrifugation (20) and was a gift from J. Greene.

D. In vitro transcription

The reaction mixture used for in vitro transcription has been described (7). B. subtilis RNA polymerase holoenzyme was prepared by B. Chelm. For the experiment shown in Fig. 7, RNA polymerase and DNA were incubated for 5 min at 30°, then synthesis was initiated by adding ATP and GTP to 100 µM and CTP and [³²P]-UTP (20 Ci/mmol) to 2.5 µM. After a 10 min "pulse", further initiation was blocked by the addition of 25 µg/ml rifampicin and the rNTP concentration was increased to 1 mM each (unlabeled). During this "chase" phase, portions of the reaction mixture were withdrawn for gel analysis at the times

specified in the figure legend. Sample preparation and gel electrophoresis of RNA have been described (7).

RESULTS

A. Sequence of the Bidirectional Terminator

The sequence of a 550 nucleotide BstNI/HaeII fragment, spanning from 8.0 to 8.55 kbp from the left end of the 12.6 kbp terminal repeat, was determined. This fragment was known to contain the bidirectional terminator, since cleavage at the BstNI site at 8.0 kbp shortened rightward transcripts and cleavage at the HaeII site at 8.55 kbp shortened leftward RNAs (7). A restriction map of this fragment, as well as its sequence, is presented in Fig. 2.

Within the 85 bp MboI fragment (MboI-85, nucleotides 226-309 in Fig. 2) a region bearing distinct homology to an *E. coli* terminator is found. This site runs from nucleotides 246 to 282 (numbering shown in Fig. 2) and contains a region of inverted sequence complementarity followed by 5 consecutive HMU residues (HMU in SPO1 DNA corresponds to T in other DNA). That this site functions as a terminator, both in vitro and in vivo, is indicated by the results of the S1 protection experiments shown in Fig. 3. In these experiments, the 3' end-labeled strands of MboI-85 were separated and each labeled strand was hybridized with various RNAs prior to digestion with the single strand-specific endonuclease S1. The size of the labeled DNA surviving S1 digestion is an indication of the length of the hybrid, which is in turn determined by the distance between the 3' end of the RNA and the labeled 3' end of the DNA (21,22). Fig. 3 shows that in vivo RNA (lanes C and G), a gel-purified in vitro transcript (lanes A and B) and the entire collection of in vitro transcripts (lane F) protect 55-60 nucleotides of the MboI-85 slow strand from S1 digestion. Since the distance from the left end of MboI-85 to the stretch of 5 HMUs at nucleotides 278-282 is 55-60 nucleotides, it is concluded that termination is occurring here. Lesser amounts of 35-38 nucleotide protected product are also observed; this segment of MboI-85 (corresponding to nucleotides 258-261 in Fig. 2) contains no recognizable terminator homology and is not particularly AT-rich. Perhaps these protected products, although they are only approximately 20 ntds. shorter, represent transcripts that are prematurely terminated when an elongating RNA polymerase molecule encounters a stalled termination complex. Lanes C and G of Fig. 3 indicate that these shorter transcripts also exist in vivo.

No protection was observed when RNA was hybridized to the fast strand of

Figure 2. Sequence and restriction map of BstNI/HaeII-550.

The nucleotide sequence of BstNI/HaeII-550 is shown (HMU is represented by T). Numbering is from left to right on the SP01 genome, with number one corresponding to the first nucleotide detected beyond the BstNI site at 8.0 kbp of the terminal repeat. The sequence of the L strand is written 5' → 3' from left to right on the top of each line, the H strand 3' → 5' below.

The dyad symmetry elements of termination sequence homologies are identified by lines above and below the sequence; the T stretches are indicated by wavy lines. The 7 base pair direct repeat immediately to the right of T_{H3} is denoted by arrows. The middle enzyme promoter homology is identified by dotted lines (see text for explanation). The entire sequence, except for nucleotides 200 to 225, was determined on both strands.

The bottom shows a restriction map of BstNI/HaeII-550, the positions of the terminators, and identifies the fragments used as probes in the S1 protection experiments. Numbers refer to the distance, in nucleotide pairs, from the BstNI site.

Figure 3. 3' end mapping of RNAs terminated at T_{H3} .

The 3' labeled slow strand of Mbo I-85 (nucleotides 226-309 in Fig. 2), which corresponds to the H strand of SPO1 DNA, was used as probe in an S1 protection experiment. Hybrids were formed, treated with S1 and the protected DNA was sized on a 10% polyacrylamide gel. In lanes A and B, the 1200 nucleotide, $P_{E7} \rightarrow T_{H3}$ transcript, eluted from a preparative RNA gel, was hybridized. Lane A: 0.4 ng RNA; lane B: 0.8 ng RNA; lane C: 2.6 μ g of 12 min CAM RNA; lane D: no RNA added to hybridization; lane E: no S1 digestion, no RNA added. Numbers to the right of lane E give the sizes, in nucleotides, of DNA markers run on the same gel (a HaeIII digest of pBR322).

A separate experiment is shown in lanes F-I. Lane F: 0.56 ng of RNA purified from an *in vitro* transcription reaction in which full-length SPO1 DNA was used as template; lane G: 2.6 μ g of 12 min CAM RNA; lane H: no RNA added; lane I: no S1 digestion, no RNA added.

12 min CAM RNA is RNA extracted from cells 12 min after infection by wild type SPO1 in the presence of chloramphenicol. It contains only early transcripts, since chloramphenicol blocks phage protein synthesis and, hence, the switch to middle transcription.

MboI-85 prior to S1 digestion, indicating that T_{L1} must be at least 25 nucleotides to the right of T_{H3} .

The sequence at T_{H3} contains two features that are not commonly found in *E. coli* terminators and are not shared by any of the other SPO1 terminators that were examined (Fig. 9). The first is the unusually large loop size. With one exception, the 14-nucleotide loop at T_{H3} is at least twice as large as any found at other terminators (1). In spite of its abnormally large loop, the thermodynamic stability of T_{H3} is comparable with those of the other SPO1 terminators (Fig. 9) and T_{H3} is a highly efficient terminator *in vitro* (7). The second distinguishing characteristic of the sequence around T_{H3} is that, beginning 5 bases to the 3' side of the T stretch, a seven base pair sequence, AGTTGAG, is directly repeated (nucleotides 287-300 in Fig. 2). There is no precedent for believing that this repeated sequence plays a role in termination. Nevertheless, it is curious that this structure should be located so close to the termination site.

The location of T_{L1} was determined using AluI-164 (nucleotides 374-538, Fig. 2) as probe. Surprisingly, two *in vivo* termination sites were detected, separated by 56 nucleotides (Fig. 4). A typical termination sequence homology is found at each site (Figs. 2 and 4). The site further to the right (closer to the HaeII site) is named $T_{L1'}$. The distance between T_{L1} and $T_{L1'}$ is close to the size difference between two transcripts that were previously mapped to the region between P_E9 and T_{L1} (the 510- and 470-nucleotide transcripts described in ref. 7, see also Fig. 1). We now conclude that both transcripts are initiated at P_E9 : the 470-nucleotide RNA is terminated at $T_{L1'}$ and the 510-nucleotide transcript at T_{L1} . We will describe further properties of $T_{L1'}$ in a separate communication.

It has been observed by others (26,27) that S1 cleavage can sometimes occur within an RNA-DNA hybrid structure at tracts of rU:dA base pairs. Since rU:dA stretches of 5 bp (in T_{H3} and T_{L1}) and 7 out of 8 bp (in $T_{L1'}$) occur at the sites that have been mapped as terminators, we must consider the possibility (Figs. 3 and 4) of S1 cleavage at regions of rU:dA pairing in the interior of a hybrid, rather than at RNA 3' ends. However, as mentioned above, the analysis of *in vitro* transcription by RNA gel analysis (7) has already indicated that termination sites lie within the 550 nucleotides whose sequence is presented in Fig. 2.

Moreover, within the Alu I-164 probe there is a 4-nucleotide T-stretch at nucleotides 429-432, as well as a 9-nucleotide stretch of A:T pairs at residues 412-420 (Fig. 2). If transcripts which read through T_{L1} were present one

Figure 4. Mapping T_{L1} and $T_{L1'}$.

The fast strand of AluI-164 (ntds 374-538, Fig. 2), labeled at its 3' end, was hybridized with 12 min CAM RNA. Lane 1: 18.2 μ g RNA; lane 2: 39 μ g RNA. (See Fig. 3 legend for a description of 12 min CAM RNA.) Hybrids were treated with S1 and the protected DNA was analyzed on an 8% acrylamide gel alongside a sequence ladder of the AluI-164 fast strand. The base specificity of the sequencing reaction is indicated above the appropriate lane. The two pictures are of a single 80 cm gel, which was cut in half for autoradiography. The top half of the gel is on the left; the bottom half is on the right. One nucleotide is lost at the point where the gel is cut. Every tenth nucleotide is identified; numbering corresponds to that in Fig. 2.

The bottom shows the sequence of the top (L; template) strand of BstNI/HaeII-550 in the region of T_{L1} and $T_{L1'}$. Polarity is 5' to 3' left to right, numbering as in Fig. 2. Transcription proceeds from right to left. The dyad symmetry regions are underlined, major termination points are indicated by arrows. These were determined from the positions of the bands in lanes 1 and 2. The production of a series of bands in these two lanes most likely results from a combination of heterogeneous termination and "nibbling" by S1 nuclease at the ends of RNA-DNA hybrid duplexes. Since the sequence of the coding strand (whose polarity is opposite to that of the message) is shown, there is a stretch of A residues at the termination site.

might expect that the (rU:dA)₄ tracts which would form at these sites might be susceptible to digestion by S1 nuclease. However, no signal corresponding to either of these sites is observed (Fig. 4). We therefore conclude that leftward RNA 3' ends do not extend leftward of nucleotide 432 (Fig. 2), which is only 8-9 ntds downstream of the 3' ends that we map at T_L1.

T_L1 represents the leftmost point of rightward transcription in the terminal repeat. Its location, 440 nucleotides rightward of the Bst N1 site, indicates that there are 158 nucleotides, located between T_H3 and T_L1, that are not transcribed by holoenzyme. We do not know if this region is transcriptionally silent; closer inspection of the sequence reveals ambiguous clues (see figure 2). A reasonably good middle enzyme (E.gp28) promoter homology can be seen near T_L1'. [This SP01-modified middle RNA polymerase recognizes specific and characteristic sequences centered 10 (TTTXXX) and 35 (AGGAGA) nucleotides upstream from the startpoint of transcription (23,24)]. Two overlapping -10 homology regions lie adjacent to the T-stretch of T_L1' at nucleotides 490-496 and 491-497, while the -35 sequence is located within the dyad symmetry element. If SP01-modified middle RNA polymerase initiates here, the transcript probably starts around AGG at positions 484-486, only 40 nucleotides ahead of T_L1. Moreover, an excellent ribosomal binding site corresponding to residues 458-467 is located very far upstream of a start codon at residues 437-439, but the intervening 18 residues have some weak secondary structure possibilities. The start codon is followed by a short open reading frame with the potential to encode 17 amino acids. It is not obvious that E.gp28 would terminate at the promoter-proximal T_L1. It is known that gp28 leaves the RNA polymerase core less readily than does σ (25), so that termination at T_L1 by E.gp28 should not be prejudged. However, the ribosome binding site should quickly be occluded by formation of the residue 447-454/459-466 stem (figure 2). On balance, it seems probable that little or no protein comes from this region, but an untranslated RNA might be produced during the middle stage of infection.

B. Sequence of T_L2

T_L2 is a partial terminator located 3.2 kbp from the right end of the terminal redundancy. Transcripts initiated at P_E10, P_E11 and P_E12 (Fig. 1) terminate here with an average efficiency of about 60%. It was hoped that a comparison of the sequence of this terminator with those of the strong terminators T_H3 and T_L1 might help explain why termination at T_L2 is less efficient. Accordingly, the sequence of a 450 nucleotide TaqI/BstEII fragment (9.1 to 9.55 kbp of the terminal repeat), known to contain T_L2, was determined

```

      10      20      30      40      50
5' CTGTTACTTT GCCTTCCTTC ATGTATTCTT GGAATATCCA CTGGAATTC

      60      70      80      90     100
CCATATTCTA GGCTGTAGTT CACCACTGCT GGGGTGTCAA TAATCATATG

      110     120     130     140     150
ATATCCTCCA TGCTCATACT CAGTGTTCTT TAATCCAGCT TTCTTCATCG

      160     170     180     190     200
TGCCACCCTC CAATAAAACC TGCCTAACC AGTGCATAAA TGTACGTTGC
              3' TGG ACGCGATTGG TCACGTATTT ACATGCAACG

      210     220     230     240     250
GATTGAGATA AAGTGTGTCA TTATTTTCTT CCTCCTTATA TGTAATTATG
CTAACTCTAT TTCACACAGT AATAAAAAGA GGAGGAATAT ACATTAATAC
      .....

      260     270     280     290     300
ACCAACAAGGA AATAATTTCC TTTGCTAGAA ACAAAGCATC AAAGGCCAAA
TGGTGTTCCT TTATTAAAGG AAACGATCTT TGTTCGTAG TTTCCGTTTT
      .....

      310     320     330     340     350
AGATCCCAAG CTGTAAGCTA GCGTGGGGCT GTGTTATAAT CTTATGCTGA
TCTAGGTGTC GACATTCGAT CGCACCCCGA CACAATATTA GAATACGACT
      .....

      360     370     380     390     400
ACGTAAGACC TCTCCTAGAC GCTTCTTTGA TACACATTCA AAGCCGTCTT
TGCATTCTCG AGAGGATCTG CGAAGAAACT ATGTGTAAGT TTCGGCAAGA

      410     420     430     440     450
CCAGATACTT TCCTAGTATG CTGTCAAGTG TTTGTCTTT GTCTGTCTCA 3'
GGTCTATGAA AGGATCATAC GACAGTTCAC AAAACAGAAA CAGACAGAGT 5'

```

Figure 5. Sequence and restriction map of TaqI/BstEII-450.

The sequence reads from left to right along the SP01 map, with nucleotide 1 being the first detected beyond the TaqI site at 9.1 kbp of the terminal repeat. The region of hyphenated dyad symmetry is indicated by lines above and below the sequence; the T stretch is indicated by wavy lines. The elements of middle promoter homology (see text) are indicated by dotted lines. The sequence of only one strand was determined for nucleotides 1-167; nucleotides 168-450 were determined for both strands.

The bottom shows a resents of middle promoter homology (see text) are indicated by dotted lines. The sequence of only one strand was determined for nucleotides 1-167; nucleotides 168-450 were determined for both strands.

The bottom shows a restriction map, the position of the probe used for S1 protection experiments and the location of T₂. Numbers at the bottom indicate distance from the TaqI site, in nucleotide pairs.

and is presented in Fig. 5. Within this sequence, at positions 297 to 327, is a region that resembles the already identified terminators (indicated in Fig. 5). An S1 protection experiment (Fig. 6) using *in vitro* RNA yields a signal

Figure 6. Mapping *in vitro* RNA 3' ends at T_L2.

HhaI/BstEII-282 (Fig. 5), 3' labeled at the BstEII end, was incubated with RNA under R-loop conditions, treated with S1 nuclease and subjected to electrophoresis in parallel with sequencing reaction products of the same fragment. Leftmost lane: 0.3 µg RNA purified from an *in vitro* transcription reaction in which full-length SP01 DNA was used as template. The base specificity of each sequencing reaction is indicated above the appropriate lane. Every tenth nucleotide is marked. This gel was cut into sections before autoradiography; several nucleotides were lost at the junction.

The bottom shows the sequence of the top (coding) strand in the region of T_L2, reading 5' to 3' from left to right. Transcription of this region is from right to left. The region of dyad symmetry is underlined and the major points of termination are indicated by arrows (subject to the same uncertainty as discussed in the legend to Fig. 4).

within this region, at nucleotides 297-300 (Fig. 5). However, the failure to obtain full-length protection with *in vitro* RNA, which is known to contain transcripts reading through T_L2 (7), suggests that S1 nuclease cleaves within the rU:dA tract (nucleotides 297-301) that would be present in a hybrid duplex

of readthrough RNA. Since T_{L2} has been defined as a partial terminator by in vitro transcription mapping (7), we conclude that the signal seen in Figure 6 is contributed both by transcripts whose 3' ends map at T_{L2} and by transcripts which read through T_{L2} .

In vivo RNA yields a signal similar to that seen with in vitro RNA (i.e. a signal at nt ds. 297-300 but no full length protection; data not shown). However, the above-mentioned limitations of the S1 protection assay with this particular probe leave us unable to determine the in vivo efficiency of T_{L2} . Thus, the question of whether T_{L2} actually functions in vivo is not clearly answered, despite its sequence homology with terminators recognized by E. coli RNA polymerase.

Examination of the sequence of T_{L2} shows that it differs from those of T_{H3} and T_{L1} in one respect. In T_{H3} and T_{L1} , as well as in all but one of the E. coli terminators examined to date (1), the T-stretch is contiguous with the last base pair of the potential RNA stem. However, at T_{L2} , there is a 3 nucleotide gap between the stem and the T-stretch. It is possible that the increased distance between these two conserved elements is responsible for the reduced efficiency of termination at T_{L2} .

The sequence in the T_{L2} region also reveals the presence of homology to the -10 consensus sequence for E.gp28 (24) at nucleotides 279-285, but without a corresponding -35 homology, and a near-canonical -10 sequence, TCTATT, at residues 206-212, associated with a -35 homology (GGAGGA, residues 230-236). We have observed transcription initiation in vitro by E.gp28 in this region (unpublished data).

C. Pausing at T_{L2}

Termination at T_{L2} is only 60% efficient in vitro (7). To determine whether RNA polymerase molecules that elongate through T_{L2} pause there, we examined the time-course of ApG-primed transcription of TaqI-11, which selects for initiation at P_{E10} (Fig. 7). The T_{L2} -terminated, 1000-ntd transcript (designated "T" in the figure) was present within 30 sec after the start of the chase (lane A) and the 1300-ntd readthrough RNA (designated "RT") was also detectable after 30 sec (lane A) and was abundant by one minute (lane B). In a dinucleotide-primed reaction, approximately 400 nt ds are polymerized by the fastest RNA polymerase molecules during the 10 min pulse (28, "pulse" and "chase" are defined in Materials and Methods). Thus, the fastest enzyme molecules elongated RNA chains during the chase at an average rate approaching 30 ntd/sec. For these RNA polymerase molecules, no significant pausing could have occurred anywhere within the transcription unit. However, when subse-

Figure 7. Characterization of Pausing at T_{L2} by RNA Gel Analysis.

Transcription of TaqI-11 by *E. subtilis* RNA polymerase was primed by ApG and aliquots were removed from the reaction mixture at various times after initiation was blocked. Lane A: 30 sec, lane B: 1 min, lane C: 1.5 min, lane D: 2 min, lane E: 3 min, lane F: 4 min, lane G: 5 min, lane H: 7.5 min, lane I: 10 min. "T" represents the transcript terminated at T_{L2} , "RT" refers to a readthrough RNA which extends to the end of the template. Minor paused transcripts are indicated by the letter "p" at the left hand side. In this experiment rifampicin, rather than NaCl, was used to block initiation, so as to avoid secondary structure isomerizations of the P_{E10} -initiated transcripts

quent time points are examined by densitometry of the appropriate gel lanes (Fig. 8), a somewhat more complicated conclusion emerges. The amount of T_{L2} -arrested transcript decreased, while the amount of readthrough transcript increased. Pausing at sites upstream from T_{L2} (denoted by "p" to the left of lane A in Fig. 7) also occurred, the amount of these paused transcripts decreasing during the course of the reaction. The fact that the amount of T_{L2} -terminated transcript decreased with time, in the face of continued transcription from RNA polymerase molecules elongating out of upstream pause sites, argues that some RNA polymerase molecules that stopped at T_{L2} eventually resumed elongation (i.e., extensive pausing occurred at T_{L2}).

We offer two speculations about the variability in pausing behavior at T_{L2} . 1) One might imagine that the population of RNA polymerase molecules is heterogeneous with respect to susceptibility to pausing. In these terms, the most rapidly-elongating RNA polymerase molecules, which are also the least likely to pause, give rise to the readthrough transcripts seen at 30 sec into the chase. Those RNA polymerase molecules that do pause prior to T_{L2} (produc-

Figure 8. Pausing at T_{L2} .

Densitometer tracings of several of the lanes from Fig. 7 are shown. The length of time allowed for transcription after initiation was blocked is indicated at the left of each tracing. "RT", "T" and "p" are defined in the text and Figure 7 legend.

ing transcripts designated "p" in Fig. 7) are likely to pause again at T_{L2} .

2) Alternatively, this heterogeneity could arise during transcription: enzyme molecules that pause once may become prone to linger and pause again.

While this experiment establishes that some transcription is temporarily halted at T_{L2} and can resume, other experiments show that after 20 min of transcription in the presence of 1 mM rNIPs, transcripts terminated at T_{L2} are still present (7). A possible explanation for these results is that some fraction of the transcripts that stop at T_{L2} are released; those that are not

released continue, after a pause. On the other hand, if previous pausing makes an RNA polymerase molecule more susceptible to a subsequent pause site (as suggested above), it may be that T_{L2} represents an insurmountable barrier to such "laggard" enzyme molecules.

DISCUSSION

The SPO1 terminal repeat contains nine transcription termination sites which range, in efficiency, from 10% to 95% (7). Four of these terminators have been characterized at the nucleotide sequence level and their sequences are similar to those of terminators recognized by *E. coli* RNA polymerase (1). Until now, only two other sites for transcriptional termination in *Bacillus* have been sequenced. One of these is at the 3' end of a precursor to 5S RNA (29); since only the mature RNA has been sequenced, this is an inferred termination site. The other case involves a region of *Staphylococcus aureus* DNA contained on a plasmid that confers erythromycin-induced antibiotic resistance on *E. subtilis*. Sequencing studies have identified a site, downstream from the coding region for the protein specifying drug resistance, that bears homology to *E. coli* terminators (30). However, an RNA 3' end has not yet been mapped to this site.

One of our goals was to examine the structural difference between completely and partially efficient terminators. The 60% efficient terminator T_{L2} has a potential RNA stem-loop with a greater thermodynamic stability than either T_{H3} or T_{L1} and all three terminators have a 5-nucleotide T-stretch (Fig. 9). The only distinguishing feature of T_{L2} is that there is a space of 3 nucleotides between the base of the potential stem and the start of the T-stretch. If distinct sites on the RNA polymerase molecule interact with these two regions to effect termination, increasing the distance between them may weaken the interaction and, hence, favor readthrough. Such "spacing" effects have been demonstrated for initiation at the *lac* promoter by *E. coli* RNA polymerase (31,32). On the other hand, no comparable analysis of spacing effects in terminators has yet been presented.

One must inevitably confront the problem of distinguishing a terminator from a pause site. Strictly speaking, a terminator should be defined as a site where release of nascent transcript and RNA polymerase occurs, since these represent the final events in the transcription cycle. Our experiments have not addressed the question of release; therefore, we infer the existence of a terminator from the presence of a band on a gel after 20-min of *in vitro* transcription at high rNTP concentrations (7). This does not exclude the pos-

sibility that what are defined as terminators might actually represent sites at which extremely long pauses occur. We have shown directly that T_L2 (Fig. 7) is either a terminator or a pause site with a half-life of greater than 20 min at high rNTP concentrations. True pause sites (marked by "p" in Fig. 7) were observed in this experiment. Thus, we conclude that the terminators identified here represent, if not release sites, at least sites where irreversible cessation of elongation occurs.

4173

tion. This provides a further reason for supposing that a nascent transcript halted at T_{L2} either resumes elongation soon after the pause, or is released.

ACKNOWLEDGEMENTS

We thank T. Friedmann, P. Deininger and P. LaPorte for patient and expert instruction in DNA sequencing and are grateful to J. Greene, D. Stillman and B. Chelm for gifts of materials. Thanks also to M. Brennan for helping with the preparation of Figures and to L. Bass and R. Cann for preparing the manuscript. S.B. was supported by a Cell and Molecular Biology Training Grant from the USPHS. The research was supported by grants from the National Science Foundation and the American Cancer Society.

*Present Address: MRC Laboratory of Molecular Biology
Hills Road
Cambridge, CB2 2QH
England

REFERENCES

1. Rosenberg, M. and Court, D. (1979) *Ann. Rev. Genet.* 13, 319-353.
2. Platt, T. (1981) *Cell* 24, 10-23.
3. Lee, F. and Yanofsky, C. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4365-4369.
4. Adhya, S. and Gottesman M. (1978) *Ann. Rev. Biochem.* 47, 967-996.
5. Martin, F. and Tinoco, I. Jr. (1980) *Nuc. Acids Res.* 8, 2295-2299.
6. Romeo, J., Brennan, S., Chelm, B. and Geiduschek, E.P. (1981) *Virology* 111, 588-603.
7. Brennan, S., Chelm, B., Romeo, J. and Geiduschek, E.P. (1981) *Virology* 111, 604-628.
8. Brutlag, D., Atkinson, M., Setlow, P. and Kornberg, A. (1969) *Biochem. Biophys. Res. Comm.* 37, 982-989.
9. Klenow, H. and Henningsen, I. (1970) *Proc. Natl. Acad. Sci. USA* 65, 168-179.
10. Englund, P. (1971) *J. Biol. Chem.* 246, 3269-3276.
11. O'Farrell, P., Kutter, E. and Nakanishi, M. (1980) *Mol. Gen. Genet.* 179, 421-435.
12. Maxam, A. and Gilbert, W. (1980) in "Methods in Enzymology, Grossman, L. and Moldave, K., Eds., Vol. 65, pp. 499-560, Academic Press, New York/London.
13. Friedmann, T. and Brown, D. (1978) *Nuc. Acids Res.* 5, 615-622.
14. Rubin, C. and Schmid, C. (1980) *Nuc. Acids Res.* 8, 4613-4619.
15. Thomas, M., White, R. and Davis, R. (1976) *Proc. Natl. Acad. Sci. USA* 73, 2294-2298.
16. White, R. and Hogness, D. (1977) *Cell* 10, 177-192.
17. Casey, J. and Davidson, N. (1977) *Nuc. Acids Res.* 4, 1539-1552.
18. Chelm, B., Romeo, J., Brennan, S. and Geiduschek, E.P. (1981) *Virology* 112, 572-588.
19. Gage, L.P. and Geiduschek, E.P. (1971) *J. Mol. Biol.* 57, 279-300.
20. Glisin, V., Crkvenjakov, R. and Byus, C. (1974) *Biochemistry* 13, 2633-2637.
21. Berk, A. and Sharp, P. (1977) *Cell* 12, 721-732.
22. Weaver, R. and Weissman, C. (1979) *Nuc. Acids Res.* 7, 1175-1193.
23. Talkington, C. and Pero, J. (1979) *Proc. Natl. Acad. Sci. USA* 76, 5465-5469.
24. Lee, G. and Pero, J. (1981) *J. Mol. Biol.* 152, 247-265.

25. Chelm, B.K., Beard, C. and Geiduschek, E.P. (1981) *Biochemistry* 20, 6564-6569.
26. Hansen, U., Tenen, D., Livingston, D. and Sharp, P. (1981) *Cell* 27, 603-612.
27. Miller, K. and Sollner-Webb, B. (1981) *Cell* 27, 165-174.
28. Brennan, S. (1982) Ph.D. Thesis. University of California, San Diego.
29. Sogin, M., Pace, N., Rosenberg, M. and Weissman, S. (1976) *J. Biol. Chem.* 251, 3480-3488.
30. Gryczan, T., Grandi, G., Hahn, J., Grandi, R. and Dubnan, D. (1980) *Nuc. Acids Res.* 8, 6081-6097.
31. Stefano, J. and Gralla, J. (1982) *Proc. Natl. Acad. Sci. USA* 79, 1069-1072.
32. Mandecki, W. and Reznikoff, W. (1982) *Nuc. Acids Res.* 10, 903-912.
33. Tinoco, I., Borer, P., Dengler, B., Levine, M., Uhlenbeck, O., Crothers, D. and Gralla, J. (1973) *Nature New Biol.* 246, 40-41.