THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 45, Issue of November 9, pp. 41850–41855, 2001

Printed in U.S.A.

Alternate Paradigm for Intrinsic Transcription Termination in Eubacteria*[S]

Received for publication, July 5, 2001, and in revised form, September 6, 2001
Published, JBC Papers in Press, September 10, 2001, DOI 10.1074/jbc.M106252200

Shyam Unniraman‡, Ranjana Prakash‡§, and Valakunja Nagaraja‡¶‖

From the ‡Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560012, India and ¶Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560064, India

Intrinsic transcription terminators are functionally defined as sites that bring about termination *in vitro* with purified RNA polymerase alone. Based on studies in *Escherichia coli*, intrinsic termination requires a palindromic stretch followed by a trail of T (or U) residues in the coding strand. We have developed a highly efficient algorithm to identify hairpin potential sequences in bacterial genomes in order to build a general model for intrinsic transcription termination. The algorithm was applied to analyze the *Mycobacterium tuberculosis* genome. We find that hairpin potential sequences are concentrated in the immediate downstream of stop codons. However, most of these structures either lack the U trail entirely or have a mixed A/U trail reflecting an evolutionarily relaxed requirement for the U trail in the mycobacterial genome. Predicted atypical structures were shown to work efficiently as terminators both inside the mycobacterial cell and *in vitro* with purified RNA polymerase. The results are discussed in light of the kinetic competition models for transcription termination. The algorithm identifies >90% of experimentally tested terminators in bacteria and is an invaluable tool in identifying transcription units in whole genomes.

The interaction of the template DNA, RNA polymerase, and the nascent RNA chain has evolved so as to minimize the release of the transcript prematurely (1). At certain sequences, the release occurs at a rate comparable with that of elongation either spontaneously or in the presence of assisting factor(s). Based on exhaustive work in *Escherichia coli*, terminators are classified into two groups (2). Functionally, if a sequence can bring about transcript release in an *in vitro* system with purified RNA polymerase alone, it is defined as an intrinsic terminator. These are also referred to as simple or factor-independent terminators. Terminators that require the presence of additional factors are classified as complex or factor-dependent terminators. In *E. coli*, most complex terminators depend on the action of the Rho termination factor (3). These two classes of terminators are not sharply defined as the efficiency of many intrinsic terminators is enhanced by the presence of additional factors (4).

Intrinsic terminators are characterized by the presence of a G/C-rich (interrupted) palindromic region followed by a trail of A residues on the template strand (5–7). There is evidence from multiple sources that the palindromic region extrudes out as a hairpin in the nascent transcript (8–12). Furthermore, there is a strong, although not absolute, correlation between the predicted stability of the stem-loop structure and termination efficiency (12). The stem-loop structure is believed to cause pausing of the polymerase (13–15) and weaken the interaction of the polymerase with the nascent RNA and template DNA (16, 17). The release is facilitated by the presence of a U trail (10, 17) probably due to the unusually weak hybrid formed by the dA·rU base pairing (18). In addition to these two primary determinants, sequences further downstream have also been shown to affect the efficiency of termination probably by being an impediment to transcription elongation (19).

Although intrinsic terminators have been studied extensively in *E. coli*, little is known about their orthologs, if any, in other bacteria. Intrinsic terminators from *E. coli* have been shown to function in many bacteria. However, recent theoretical analysis indicates that only a minority of bacteria may employ this mechanism of transcription termination (20, 21). In an attempt to formulate a general model for intrinsic transcription termination in eubacteria, we have developed an algorithm to identify hairpin potential sequences in bacterial genomes and have applied it to the *Mycobacterium tuberculosis* genome. Such sequences appear to be concentrated in the immediate downstream region of stop codons, a feature one would expect of intrinsic terminators. Surprisingly, nearly 90% of these structures lack a U trail entirely or possess a mixed A/U trail. We show experimentally that these atypical structures work efficiently as terminators both *in vivo* and *in vitro*. Based on our results, the present algorithm represents the most efficient and accurate software for the identification of intrinsic terminators in eubacteria.

EXPERIMENTAL PROCEDURES

Algorithm—The GeSTer (Genome Scanner for Terminators) algorithm first segregates the coding, upstream, and downstream regions based on the feature table entries of the genome sequence. Next it searches for palindromic sequences downstream of each gene (−20 to +270 nucleotides of the stop codon) without entering adjacent coding regions. The search is initiated at a G/C-rich (>50%) tetranucleotide, and a reverse complementary match is sought within the next 70 nucleotides. This defines the base of the stem, and the match is extended inward. Once a mismatch is encountered, all possible structures are computed allowing for different combinations of mismatches and gaps. The ΔG of formation of each of these structures was computed using the parameters from Turner *et al*. (22) and Jaeger *et al*. (23). Among all these structures, the one with the lowest ΔG was retained. Then the program moved to the next G/C-rich tetranucleotide and reinitiated the search.

The final set of structures used a minimal ΔG filter based on the G/C content of the bacteria. The species-specific ΔG_{cutoff} was set at $-0.230 \times (\%G_C) + 3.44$ based on two premises. Firstly, the basal ΔG of the

* The research was supported by grants from the Indian Council of Medical Research, Government of India. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at *http://www.jbc.org*) contains a list of terminator sequences used to optimize the algorithm.

§ Recipient of an Indian Academy of Sciences fellowship.

‖ To whom correspondence should be addressed. Tel.: 91-80-360-0668; Fax: 91-80-360-2697; E-mail: vraj@mcbl.iisc.ernet.in.

downstream region (20) is strongly correlated with the G/C content of the genome. Secondly, the algorithm should identify preferentially structures in the downstream rather than in the upstream region. The optimized cutoff value for ΔG was derived by iteratively weighing down $\Delta G_{downstream}$ so as to maximize the likelihood of identifying only downstream and not upstream structures. Under these constraints, the algorithm detects 10-fold or more structures in the downstream region compared with the upstream region. Furthermore, the distribution with respect to the stop codon shows a characteristic peak indicating a non-randomness in the distribution of structures. To further substantiate the statistical significance of this peak, a *t* test was performed to compare the average around the peak and that of a region that shows a background level of occurrence of structures −190–200 nucleotides downstream of the stop codon. The *p* value of the *t* tests for each genome is listed in Table I.

With the final parameters, the algorithm identified more than 90% of all experimentally shown terminators in different bacteria (29–32). This corresponds to a false negative rate of <10% and a false positive rate that varies between 5 and 10% in different genomes. All the putative terminators are classified based on the presence of a U trail as well as the position of adjacent structures (described below). The distribution of the structures is also analyzed and represented graphically. In the case of genes that are followed by multiple structures, the best candidate is identified again based on the lowest ΔG value. The program is available on request from the authors.

The whole genome sequences used for the present analysis are as follows, *Bacillus subtilis* (AL009126), *E. coli* K12 (U00096), *Mycobacterium leprae* (AL450380), *M. tuberculosis* H37Rv (AL123456), *Neisseria meningitidis* MC58 (AE002098), and *Xylella fastidiosa* (AE003849). The accession numbers of the GenBank™ entries are denoted in parentheses.

Bacterial Strains and Plasmids—E. coli strain DH10B was used for all cloning experiments, and *M. smegmatis* mc^2155 was used as the mycobacterial host for all *in vivo* assays for termination. The *M. tuberculosis* strain H37Ra was used to isolate genomic DNA. The *E. coli* cells were grown in Luria-Bertani medium whereas the *M. smegmatis* cells were grown as described in Ref. 24. Kanamycin was added at 35 μg/ml where appropriate.

The termination selection series of vectors (pTER) was generated by cloning various promoter fragments into the *E. coli*-mycobacterial promoter selection shuttle vector pSD7 (25). All of them contain different fragments encompassing the *M. smegmatis gyr* promoter and retain a unique *Bam*HI site only downstream of the promoter. pTER1 harbors a 257-bp[1] fragment whereas pTER5 harbors a 317-bp fragment. pTER7 harbors a 2.5-kb fragment, which includes the 5′ half of the *gyrB* gene (26). An internal *Bsr*GI site was filled to disrupt the reading frame and cause premature translation termination in pTER7Δ.

The putative terminators downstream of *tuf* and *Rv1324* were PCR-amplified from *M. tuberculosis* genomic DNA using primers 5′-ACCAGGATCCTCAAGTAGGTCTAC-3′ and 5′-CGGAGGATCCATGTCAGCGTAG-3′ and 5′-CGGCGGATCCTCGCCAACGCG-3′ and 5′-GAACGGATCCCCCGGGTTGTCGTAG-3′, respectively. The putative terminator downstream of the *M. smegmatis gyrA* gene was amplified from the pMN1Bg (26) clone using primers 5′-CCGAGATCTACGCGAGCGAGTTG-3′ and 5′-GCGGGATCCCCCGGGGCGCGTCGG-3′. All PCR products were cloned at the *Bam*HI site in the termination vector after digestion with *Bam*HI alone or with *Bgl*II as required.

Analysis of Termination in Vivo—M. smegmatis cells harboring various constructs were grown to midlog phase (1 A_{600}), harvested, washed, and resuspended in 100 mM Tris-HCl (pH 8.0). Cells were disrupted by sonication, and the extracts were recovered by centrifugation. An appropriate dilution was assessed for specific CAT activity as described before (27). All results were normalized to the activity of the equivalent promoter construct.

Analysis of Termination in Vitro—M. smegmatis RNA polymerase holoenzyme was purified as described before (28) with certain modifications. Briefly, the holoenzyme was enriched by polyethyleneimine and ammonium sulfate precipitation and purified through Superdex 200 followed by a DNA-cellulose column. Fractions were assessed for their ability to bind to a fragment encompassing the *M. smegmatis gyr* promoter.

Fragments containing the promoter with putative terminator regions were PCR-amplified using an appropriate primer downstream of the terminator with a vector-specific forward primer (25). The gel-purified fragments were used as templates for runoff transcription assays. RNA polymerase was incubated with 1 μg of template DNA for 10 min at 4 °C in 50 mM Tris-HCl (pH 8.0), 3 mM magnesium acetate, 100 μM EDTA, 100 μM dithiothreitol, 50 mM potassium chloride, 50 mg/ml bovine serum albumin, and 5% glycerol. The reaction was started by adding NTPs (final concentrations of 100 μM ATP, CTP, and GTP, 0.4 μM UTP, and 1 μCi of [α-^{32}P]UTP) and shifting to 37 °C. After 1 min, the reaction was supplemented with UTP (final concentration of 100 μM) and heparin (final concentration of 150 μg/ml). Reactions were stopped by the addition of equal volumes of formamide containing 0.025% bromphenol blue and 0.025% xylene cyanol and resolved on an 8% denaturing polyacrylamide gel. The reactions were visualized and quantitated by phosphorimaging (Fujifilm). The termination efficiency (TE) was calculated as follows.

For single structures, TE = 100 × TP/(RO + TP) where TP is the amount of terminated product and RO is the amount of runoff transcript. For total termination efficiency of the tandem structures, TE = 100 × (TP_1 + TP_2)/(RO + TP_1 + TP_2) where TP_1 and TP_2 are the amounts of product terminated downstream of the first and second structures, respectively.

For the first structure present in tandem, TE = 100 × TP_1/(RO + TP_1 + TP_2). For the second structure present in tandem, TE = 100 × TP_2/(RO + TP_1 + TP_2).

RESULTS

Algorithm—To delineate the elements involved in intrinsic transcription termination conserved in all bacteria, we have developed the GeSTer algorithm that identifies and classifies structures based on the trailing nucleotides and the position of adjacent structures. The algorithm identifies hairpin structures using the following parameters, a stem length ranging from 4 to 20 nucleotides with a loop of 3 to 10 nucleotides with a maximum of 3 unpaired nucleotides in the form of gaps or mismatches. These parameters are based on the qualitative assessment of all known terminators from different bacteria (29–32). The list of terminators compiled from previous literature has been provided as Supplemental Material. −20 to +270 nucleotides around the stop codon for each gene were searched, without entering adjacent coding regions. In the case of overlapping structures, the one with the lower ΔG was retained. Finally, structures were filtered using a minimal ΔG requirement based on the G/C content of the genome. With these parameters, the algorithm identified more than 90% of all experimentally shown terminators in different bacteria (29–32, Supplemental Material). Structures were classified as follows: (*a*) *E. coli* type/L-shaped, those with >3 U residues present in the 10 nucleotides trailing the structure; (b) *Mycobacterium* type/I-shaped, those with 3 or fewer U residues in the trail; (*c*) V-shaped, structures that are immediately followed (or preceded) by another structure; (*d*) Tandem/U-shaped, multiple structures that are present downstream of a single gene; and (*e*) Convergent/X-shaped, structures present between adjacent convergently oriented genes. It should be noted that all structures, other than the L-shaped ones, are symmetric and could potentially work in either orientation.

Whole Genome Analysis—When structures identified by the program were compiled, we found that there was a preponderance of hairpin potential sequences within 50 nucleotides downstream of the stop codon in bacterial genomes (Table I, Fig. 1), a characteristic one would expect of transcription terminators. In a few species either the L-shaped (*B. subtilis* and *N. meningitidis* in Table I) or I-shaped structures (*M. tuberculosis* and *M. leprae* in Table I) predominate. However, in many cases, these two classes constitute significant fractions of the structures identified (*E. coli* and *X. fastidiosa* in Table I).

A detailed analysis of *M. tuberculosis* and *E. coli* genomes revealed many interesting features. Firstly, the algorithm detects putative terminators downstream of 20–40% of genes. This is probably due to the operonic arrangement of many

[1] The abbreviations used are: bp, base pair(s); kb, kilobase(s); PCR, polymerase chain reaction; CAT, chloramphenicol acetyltransferase; TE, termination efficiency.

TABLE I
Representative whole genome analysis of bacterial sequences

Species	Genome	Genes	All	L	I	U	V	X	Peak	*p* value[a]
B. subtilis	4214814	4218	1608	1422	186	104	0	153	+21	0.0023
E. coli	4639221	4397	1734	918	816	256	13	139	+21	0.0004
M. leprae	3268203	1653	342	67	275	20	1	2	+32	0.0011
M. tuberculosis	4411529	3970	890	83	807	87	5	31	+37	0.0002
N. meningitidis	2272351	2096	885	621	264	83	2	67	+33	0.0002
X. fastidiosa	2679306	2821	481	198	283	31	3	23	+25	0.0039

[a] *t* test performed as described under "Experimental Procedures."

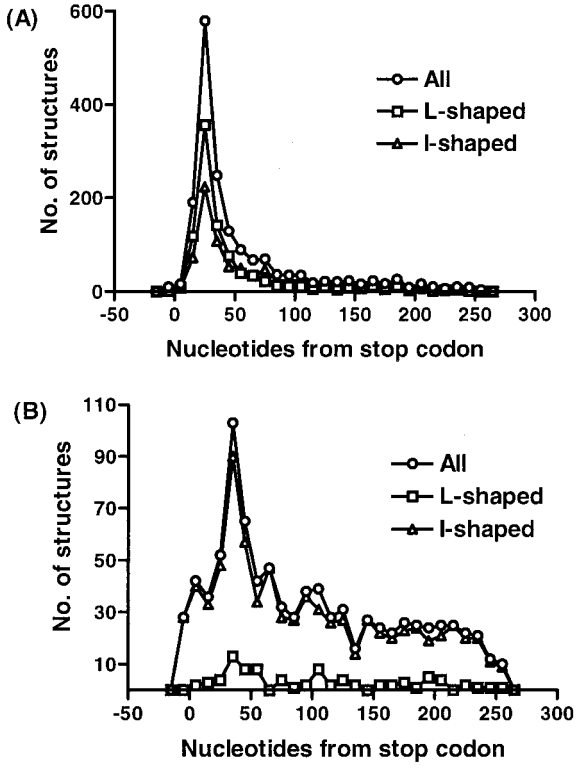

FIG. 1. **Distribution of terminators in bacterial genomes.** The distribution of all classes of terminators with the stop codon in *E. coli* (*A*) and *M. tuberculosis* (*B*). The distribution of L- and I-shaped terminators is also shown.

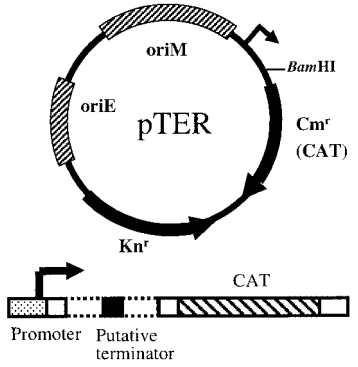

FIG. 2. **The termination selection vector series (pTER).** These vectors are *E. coli*-mycobacteria shuttle vectors and have a unique *Bam*HI site between the promoter and the CAT reporter gene. The origins of replication for mycobacteria (oriM) and *E. coli* (oriE) are indicated. The *M. smegmatis gyr* promoter is shown as an *arrow*. Upstream of the reporter system, there are stop codons in all frames.

genes. In addition, some of the other genes may rely on the Rho protein for transcription termination. In agreement with this, a Rho homologue has been identified in the *M. tuberculosis* genome as well. Secondly, there is dramatic concentration of structures about 21 nucleotides downstream of the stop codon in *E. coli* with relatively few structures present in the rest of the downstream region. In *M. tuberculosis*, on the other hand, the structures peak 37 nucleotides downstream of the stop codon with a significant fraction spread throughout the downstream region. Furthermore, as discussed above, *E. coli* shows marginal preference in using L-shaped structures over I-shaped structures whereas *M. tuberculosis* almost exclusively uses I-shaped structures. However, it is noteworthy that irrespective of their frequency of occurrence, the non-L-shaped structures are concentrated at the same position as the L-shaped structure (Fig. 1) indicating that they serve a similar purpose, *i.e.* of transcription termination. Thirdly, a significant portion of the genes employ multiple structures, either V- or U-type, for bringing about termination. Of these the tandem structures are more frequent. For instance, 10% of the putative terminators in *M. tuberculosis* and 15% in *E. coli* are of the U-type. Lastly, few convergently oriented genes use a single structure present in the shared downstream region between them.

Structures without a U Trail Are Efficient Terminators in M. smegmatis—Although L-shaped structures function in many species including *E. coli*, the V-shaped structures have been identified previously at least in *Streptomyces* (31). In addition, X-shaped structures function both in *E. coli* (33) and *Streptococcus* (32). Therefore, we decided to test the ability of structures that lack an obvious U trail to bring about transcription termination in mycobacteria. Toward this end, we constructed a mycobacteria-specific termination selection vector (pTER5; Fig. 2) by cloning the *M. smegmatis* gyrase promoter (27) upstream of a CAT reporter gene. A fragment cloned between the promoter and the reporter gene would reduce transcriptional read-through if it were a terminator, thereby leading to chloramphenicol sensitivity and a quantitative decrease in specific CAT activity.

Representative I-shaped terminators were PCR-amplified from the *M. tuberculosis* genome and cloned into the termination vector (see "Experimental Procedures"). The terminator downstream of *tuf* gene harbors an AU-rich trail. When present upstream of the CAT gene, it reduces transcription read-through by ~80% (Fig. 3, *A* and *D*) indicating that a classical U trail as defined in *E. coli* is not essential for transcription termination. Surprisingly, the terminator showed comparable efficiency in the reverse orientation that lacks an appreciable AU-rich trail. To substantiate this observation, we tested the terminator present downstream of *Rv1324* for its ability to bring about transcription termination bidirectionally. This structure is flanked on both sides by G/C-rich stretches. In agreement with the above results, this terminator also functions with comparable efficiency in both orientations (Fig. 3, *B* and *D*). This clearly demonstrates that the U trail is not essential for the functioning of the terminator in mycobacteria. To analyze the termination efficiency of U-shaped structures, we used the putative terminator present downstream of the *gyrA* gene in *M. smegmatis* (Fig. 3*C*). The individual structures here are weaker than the structures tested so far; however, in tan-

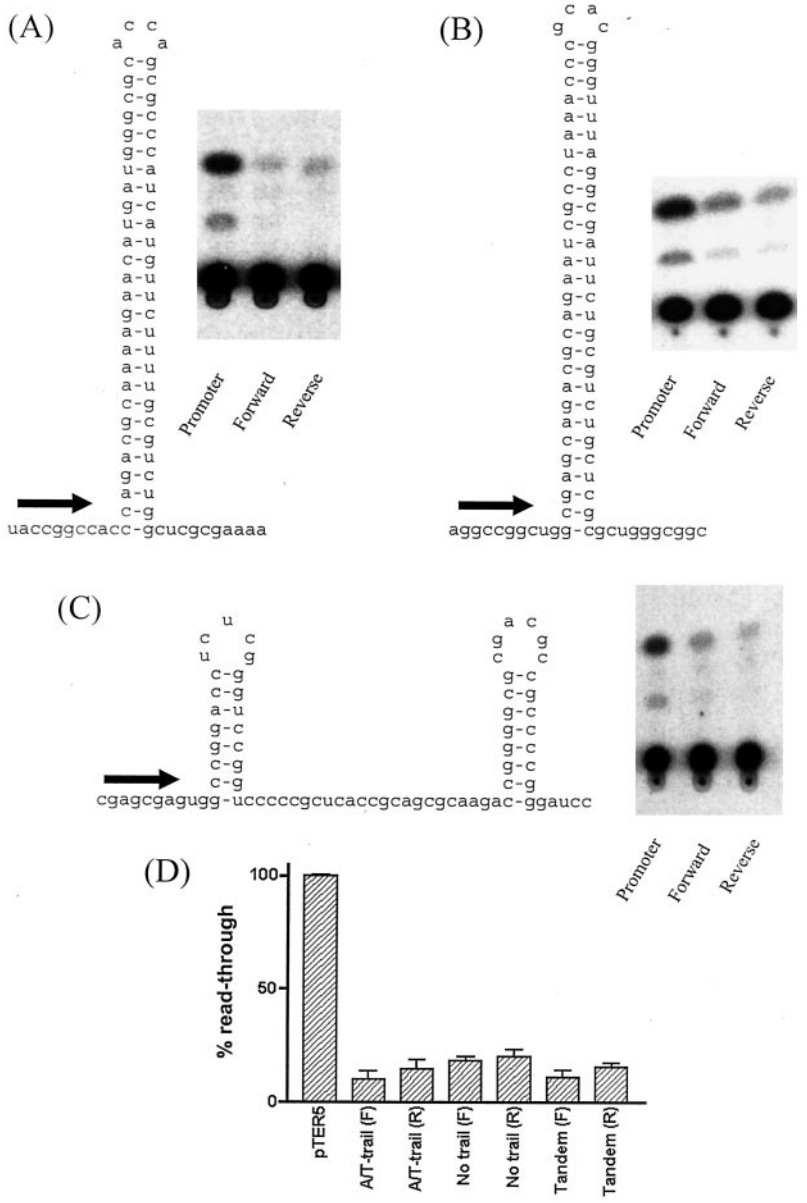

FIG. 3. **Different classes of terminators in mycobacteria.** Representative terminators with (*A*) or without (*B*) an A/U trail or present in tandem (*C*) identified by the algorithm are shown along with a representative CAT activity indicating a decrease in read-through transcription. All terminators were tested in pTER5 (see "Experimental Procedures" and Fig. 2). *D*, summary of the analysis of read-through transcription. Each value is an average of at least three independent experiments.

dem they show a similar termination efficiency in both orientations (Fig. 3, *C* and *D*). Thus, both I- and U-shaped structures function with high efficacy *in vivo*.

Terminators Work Only in the Untranslated Region—In the experiments described so far, the terminators were cloned more than 50 nucleotides downstream of the promoter in the 5′-untranslated region. When such a structure was moved closer to the promoter (27 nucleotides downstream), there was no detectable effect on termination efficiency (compare pTER1 and pTER5 in Fig. 4). However, when cloned within the coding region (1.1 kb downstream of the promoter), the structure had no detectable effect on transcription read-through in either orientation (compare pTER7 and pTER5 in Fig. 4). On the other hand, in the same construct, when translation was moved out of frame, leading to a premature stop codon, the structure brings about termination with efficiency comparable with the 5′-untranslated region context (compare pTER7Δ and pTER7 in Fig. 4). Thus, terminators appear to be effective only in the non-coding region. The close coupling of transcription and translation in bacteria probably prevents the extrusion of these structures in the RNA in the coding region.

Terminators Work Efficiently in Vitro—To ensure that the structures were genuine intrinsic transcription terminators, we analyzed their ability to bring about termination *in vitro* using purified RNA polymerase from *M. smegmatis*. The templates containing the promoter and the various terminators in either orientation were generated as described under "Experimental Procedures." Fig. 5 shows results of a representative *in vitro* transcription termination assay. The majority of the transcript appears to terminate a few nucleotides downstream of the structure in every case. The termination efficiency of these structures is comparable with those obtained *in vivo* (Table II). Furthermore, in agreement with the results obtained in the *in vivo* experiments, all three structures work bidirectionally (compare "Forward" and "Reverse" in Table II). Of particular interest is the tandem terminator in which transcription terminates downstream of each structure. Notably, the first structure encountered by the polymerase (t_1 in the forward orientation and t_2 in the reverse orientation) works at low efficiency (~45%, Table II). On the other hand, the same structure when encountered second (t_2 in the forward orientation and t_1 in the reverse orientation) shows appreciably higher termination ef-

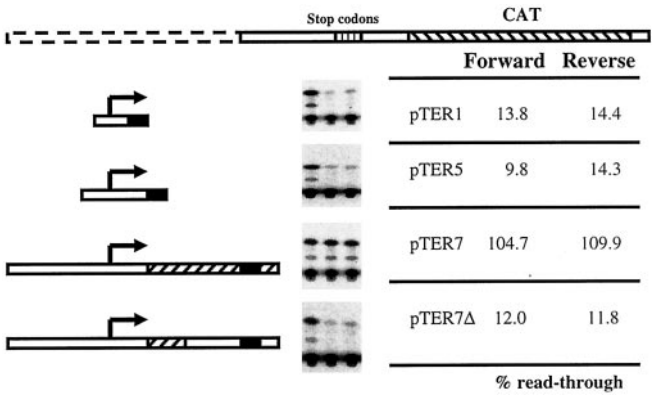

FIG. 4. **Effect of distance from the promoter and translation on termination efficiency.** The terminator from Fig. 2*A* was placed at different distances from the promoter, −27 bp (pTER1), 77 bp (pTER5), and 1.1 kb (pTER7) downstream of the transcription start site. Representative CAT assays and the means obtained from at least three independent experiments is shown. The promoter (*arrow*), terminator (*filled box*), and translated regions (*hatched box*) are indicated.

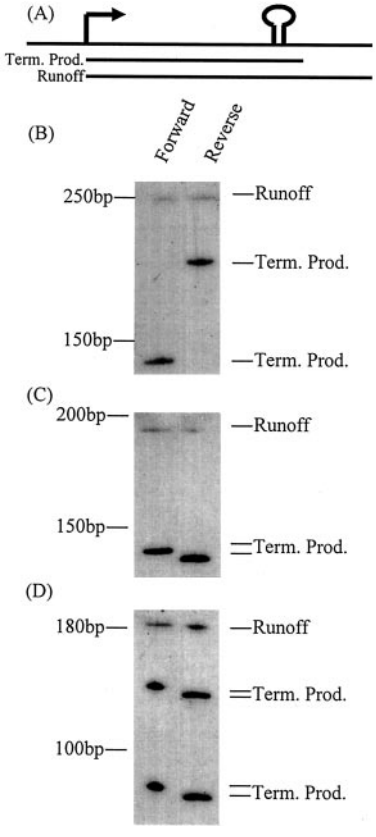

FIG. 5. ***In vitro* termination assay.** *A*, schematic representation for the assay. Runoff transcription assays were performed with constructs harboring the terminators shown in Fig. 2, in either orientation. The terminators used in the assay are as follows, *tuf* terminator (*B*), *Rv1324* terminator (*C*), and the *gyrA* terminator (*D*). The positions of full-length runoff and terminated (*Term. Prod.*) products are indicated. A sequencing ladder was used as molecular mass marker.

ficiency (~70%, Table II), probably due to a slowing down of the polymerase at the first structure. Together the two structures in tandem show an efficiency comparable with the other individual structures that have a longer stem. Thus, in conclusion, structures with long stems (>27 bp) function alone efficiently as terminators whereas structures with shorter stems (~8 bp) individually form inefficient terminators. However, the cell recruits these shorter structures in tandem where two of them together now constitute a single efficient terminator. This is especially important because the U-shaped structures constitute 10% of the structures in *M. tuberculosis*.

TABLE II
Termination efficiency of various structures in vitro

Terminator	Termination efficiency ± S.D.[a]	
	Forward	Reverse
	%	
AT trail	84.4 ± 3.5	82.4 ± 5.3
No trail	82.9 ± 4.1	92.0 ± 3.8
Tandem (total)	82.6 ± 4.3	84.0 ± 3.5
Tandem (t_1)	43.2 ± 3.9	71.3 ± 3.7[b]
Tandem (t_2)	69.3 ± 3.3[b]	44.5 ± 3.0

[a] Calculated as described under "Experimental Procedures."
[b] Efficiency of the second structure is calculated as a percentage of read-through from the first structure.

DISCUSSION

Intrinsic terminators represent an extremely economical mechanism of transcription termination. Earlier attempts to identify intrinsic terminators have, in general, had limited success in bacterial species other than *E. coli* (21, 30, 34, 35). This is probably because they fail to take into account the possibility that secondary structure alone could work as a terminator. As a result, they identify only the L (and possibly the X) subsets of the terminators identified by the present algorithm. The only other theoretical analysis of the distribution of secondary structures in the non-coding region similarly failed to detect a concentration of structures downstream of the stop codon in the majority of genomes (20). This is probably because of the 60-base window (moved in steps of 10 bases) employed in the study. Such a rigid window and large step size would lead to blunting of peaks, which reduces the resolution of their analysis. Therefore, in many organisms, including *M. tuberculosis*, the modest concentration of structures is no longer statistically distinguishable from the background ΔG. In contrast, the present algorithm varies the window size dynamically to specifically identify individual stem-loop structures, thereby improving the sensitivity and accuracy of the prediction.

We used the *M. tuberculosis* genome as a test for our algorithm. Surprisingly, although there were many secondary structure potential sequences present downstream of genes, most of these were devoid of a trail of U (Table I). Notably, irrespective of whether the structures are followed by a U trail or not, they are concentrated approximately within 50 nucleotides downstream of the stop codon (Fig. 1). Thus, both classes of structures appear to have evolved for a conserved function in transcription termination. Interestingly, even in *E. coli*, a significant fraction of the structures lack a discernible U trail. In agreement with this, we have experimentally shown that the U trail does not play a primary role in transcription termination both inside the mycobacterial cell (Fig. 3) and *in vitro* with purified mycobacterial RNA polymerase (Fig. 4). On the other hand, the significance of the trail sequence in *E. coli* is not completely clear. In different systems, the U trail has been shown to be either essential (10, 17), unnecessary (36), or necessary only in the absence of appropriate elements downstream of the termination site (19).

The efficiency of termination is believed to be determined by the opposing influences of the rates of elongation and release (1, 37). Recently, a paused form of the polymerase that reacts slowly with the nucleotides has been proposed as an intermediate prior to the actual step of release (38, 39). Revised models based on single molecule experiments evoke kinetic competition between elongation rates and the largely irreversible formation of the paused complex rather than the actual step of release (39). Most bacterial coding sequences have evolved to

favor the former rather than latter reaction. Terminators represent sequences, which specifically alter one or both of these reactions leading to transcription termination.

Support for the above models came from the analysis of mutationally altered polymerases that have a lower elongation rate and show a concomitant increase in termination (40). Similar results are obtained using a wild-type polymerase in the presence of limiting concentrations of nucleotides (7, 40). Furthermore, recent work shows that the primary role of the U trail may be to decrease the rate of elongation (38) and thereby allow the hairpin to extrude and dislodge the nascent chain from the catalytic site. In *M. tuberculosis*, where the rate of RNA chain elongation is about 10-fold slower than *E. coli* (41), such a role for the U trail would be redundant. Therefore, an I-shaped structure, even without the stalling effect of the U trail, could work as efficiently as an L-shaped structure. Thus, in the framework of the kinetic competition model, a lower elongation rate would mean that the enhancement required in the rate of pausing/release to bring about termination would be correspondingly lower. An alternative explanation for the low representation of the L-shaped structures in *M. tuberculosis* could be the high G/C content of the organism. However, *M. leprae*, an organism closely related to *M. tuberculosis*, shows a similar preference for I-shaped structures although it has a considerably lower G/C content. In addition, we find no simple relationship between the G/C content of an organism and its preference for one or the other type of structure. On the other hand, our hypothesis predicts a correlation between the prevalence of the I-shaped structure with a lower rate of RNA chain elongation. In agreement with this prediction, when the *E. coli* RNA polymerase itself is made to move slowly in the presence of limiting amounts of nucleotides, it terminates efficiently even in the absence of a U-trail (7). Thus, our results substantiate the model of kinetic competition between the rates of elongation and termination (1, 39).

Another point of interest is that we found that the cells are protected against premature termination at structures within the coding region by the translating ribosomes (Fig. 4). This mechanism would not be operational in tRNA and rRNA genes. These two classes of highly transcribed genes are known to have extensive secondary structure in their RNA without the protective influence of translation. Therefore, a terminator structure within the coding region of such genes would be disastrous to the cell. Significantly, the algorithm does not identify putative terminators in the coding regions of these genes, implying that the identified structures are genuine terminators.

Acknowledgments—We thank Anil K. Tyagi for pSD7, Narasimha Prakash for suggestions and discussion, and M. Chatterji for technical assistance, discussion, and critical reading of the manuscript.

REFERENCES

1. von Hippel, P. H. (1998) *Science* **281,** 660–665
2. Richardson, J. P., and Greenblatt, J. (1996) in *Escherichia coli* and *Salmonella*: *Cellular and Molecular Biology* (Neidhardt, F. C., ed) 2nd Ed., pp. 822–848, ASM Press, Washington D. C.
3. Platt T. (1994) *Mol. Microbiol.* **11,** 983–990
4. Das, A. (1993) *Annu. Rev. Biochem.* **62,** 893–930
5. Platt T. (1986) *Annu. Rev. Biochem.* **55,** 339–372
6. Yager, T. D., and von Hippel, P. H. (1987) in *Escherichia coli* and *Salmonella*: *Cellular and Molecular Biology* (Neidhardt, F. C., ed) 1st Ed., pp. 1241–1275, ASM Press, Washington D. C.
7. Yarnell, W. S., and Roberts, J. W. (1999) *Science* **284,** 611–615
8. Fisher, R., and Yanofsky, C. (1983) *J. Biol. Chem.* **258,** 9208–9212
9. Ryan, T., and Chamberlin, M. J. (1983) *J. Biol. Chem.* **258,** 4690–4693
10. Lynn, S. P., Kasper, L. M., and Gardner, J. F. (1988) *J. Biol. Chem.* **263,** 472–479
11. Yang, M. T., and Gardner, J. F. (1989) *J. Biol. Chem.* **264,** 2634–2639
12. Cheng, S. W., Lynch, E. C., Leason, K. R., Court, D. L., Shapiro, B. A., and Friedman, D. I. (1991) *Science* **254,** 1205–1207
13. Farnham, P. J., and Platt, T. (1981) *Nucleic Acids Res.* **9,** 563–577
14. Wang, D., Severinov, K., and Landick, R. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 8433–8438
15. Artsimovitch, I., and Landick, R. (1998) *Genes Dev.* **12,** 3110–3122
16. Arndt, K. M., and Chamberlin, M. J. (1990) *J. Mol. Biol.* **213,** 79–108
17. Wilson, K. S., and von Hippel, P. H. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 8793–8797
18. Martin, F. H., and Tinoco, I., Jr. (1980) *Nucleic Acids Res.* **8,** 2295–2299
19. Reynolds, R., and Chamberlin, M. J. (1992) *J. Mol. Biol.* **224,** 53–63
20. Washio, T., Sasayama, J., and Tomita, M. (1998) *Nucleic Acids Res.* **26,** 5456–5463
21. Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O., and Salzberg, S. L. (2000) *J. Mol. Biol.* **301,** 27–33
22. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17,** 167–192
23. Jaeger, J. A., Turner, D. H., and Zuker, M. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 7706–7710
24. Nagaraja, V., and Gopinathan, K. P. (1980) *Arch. Microbiol.* **124,** 249–254
25. Das Gupta, S. K., Bashyam, M. D., and Tyagi, A. K. (1993) *J. Bacteriol.* **175,** 5186–5192
26. Madhusudan, K., and Nagaraja, V. (1995) *Microbiology* **141,** 3029–3037
27. Unniraman, S., and Nagaraja, V. (1999) *Genes Cells* **4,** 697–706
28. Predich, M., Doukhan, L., Nair, G., and Smith, I. (1995) *Mol. Microbiol.* **15,** 355–366
29. Hartmann, R. K., and Erdmann, V. A. (1991) *Nucleic Acids Res.* **19,** 5957–5964
30. d'Aubenton-Carafa, Y., Brody, E., and Thermes, C. (1990) *J. Mol. Biol.* **216,** 835–858
31. Ingham, C. J., Hunter, I. S., and Smith, M. C. (1995) *Nucleic Acids Res.* **23,** 370–376
32. Steiner, K., and Malke, H. (1995) *Mol. Gen. Genet.* **246,** 374–380
33. Postle, K., and Good, R. F. (1985) *Cell* **41,** 577–585
34. Brendel, V., and Trifonov, E. N. (1984) *Nucleic Acids Res.* **12,** 4411–4427
35. Brendel, V., Hamm, G. H., and Trifonov, E. N. (1986) *J. Biomol. Struct. Dyn.* **3,** 705–723
36. Abe, H., and Aiba, H. (1996) *Biochimie (Paris)* **78,** 1035–1042
37. von Hippel, P. H., and Yager, T. D. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 2307–2311
38. Gusarov, I., and Nudler, E. (1999) *Mol. Cell* **3,** 495–504
39. Yin H, Artsimovitch I, Landick R, and Gelles J. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 13124–13129
40. McDowell, J. C., Roberts, J. W., Jin, D. J., and Gross, C. (1994) *Science* **266,** 822–825
41. Harshey, R. M., and Ramakrishnan, T. (1977) *J. Bacteriol.* **129,** 616–622