

The Number of G Residues in the *Bacillus subtilis* *pyrG* Initially Transcribed Region Governs Reiterative Transcription-Mediated Regulation[∇]

Alexander K. W. Elsholz,[†] Casper Møller Jørgensen, and Robert L. Switzer*

Department of Biochemistry, University of Illinois, Urbana, Illinois 61801

Received 16 October 2006/Accepted 1 December 2006

Derepression of *pyrG* expression in *Bacillus subtilis* involves CTP-sensitive reiterative transcription, which introduces up to 11 extra G residues at the 5' ends of *pyrG* transcripts. Insertion of three or more additional Gs at the 5' end of the *pyrG* initially transcribed region abolished reiterative transcription and caused constitutive expression.

Expression of the *pyrG* operon, which encodes CTP synthetase, is regulated in *Bacillus subtilis* and several other low-GC gram-positive bacteria by the intracellular CTP level (5, 8). CTP exerts its repressive effect by a novel mechanism that involves transcription attenuation at an intrinsic terminator located within an untranslated leader region. Two short segments of conserved nucleotide sequence are required for regulation, namely, a *pyrG* initially transcribed region (ITR) with the sequence 5'-GGGC (nontranscribed strand) and a C-plus-U-rich sequence (GCUCCCUUUC in *B. subtilis*) located in the 5' strand of the attenuator-specified terminator hairpin (Fig. 1A). Evidence that supports the mechanism for *pyrG* regulation illustrated in Fig. 1A has been presented (8–10). Low intracellular CTP concentrations cause transcription pausing at position +4 (normally a C) of the nascent *pyrG* transcript, which allows reiterative transcription (also called transcriptional slippage) to occur and introduce from 1 to 11 extra G residues before normal elongation resumes. When the intracellular CTP level is high, insertion of a C residue at position +4 proceeds readily without pausing, which precludes reiterative transcription. The extra G residues added to the 5' ends of *pyrG* transcripts under CTP limitation are proposed to act as antiterminator sequences; when present, they base pair with the C-plus-U-rich 5' strand of the terminator hairpin and preclude terminator formation (Fig. 1A). Thus, conditional reiterative transcription and the insertion of extra G residues result in CTP-sensitive antitermination and CTP-regulated transcription of the downstream *pyrG* gene. It is likely that this mechanism serves to regulate *pyrG* expression in several genera of low-GC gram-positive bacteria (10).

A series of *pyrG* transcripts with from 1 to 11 extra G residues is formed by reiterative transcription at the wild-type *pyrG* ITR in pyrimidine-starved cells (10). However, mutational insertion of four extra G residues at the start of the *pyrG* ITR (specifying the transcript sequence 5'-GGGGGGGC) was

sufficient to cause constitutive *pyrG* expression (10). In this study we sought to answer two related questions concerning the function of the extra 5' G residues. First, given that the addition of extra G residues permits formation of a progressively more stable antiterminator structure as more Gs are added, what number of G residues at the 5' end of *pyrG* transcripts is sufficient to cause fully constitutive expression? Second, what number of G residues in the *pyrG* ITR is optimal for reiterative transcription in pyrimidine-starved cells? In other cases where reiterative transcription has been studied, a homopolymeric run of at least three template residues is required for transcriptional slippage (1, 3, 4, 11). Our previous studies (10) confirmed that at least three G residues in the template *pyrG* ITR were required for reiterative transcription. However, reiterative transcription requires melting of the DNA template and nascent transcript within the initiation complex, so it would be expected that the greater stability of longer (rG)_n · (dC)_n duplexes would suppress melting and reduce reiterative transcription.

Both of these questions were addressed by construction of mutant *pyrG* promoter regions in which the number of G residues in the ITR was systematically increased from 3 (wild type) to 13 (Fig. 1B). Reiterative transcription in vivo with templates bearing several of the G_n insertions was characterized by primer extension mapping of the 5' ends of *pyrG* transcripts to determine their lengths and relative amounts, as previously described by Meng et al. (10) (Fig. 2). Derivatives of the high-copy plasmid pEB112 (6) containing the wild-type *pyrG* promoter-leader region (nucleotides –49 to +81) and mutants of this region in which one, two, three, and eight G residues were inserted into the ITR were constructed as described by Meng et al. (10) and used to transform *B. subtilis* HC-11 (*pyrB*::Spc^r). The pyrimidine auxotrophic *pyrB* host was used so that the cells could be starved for pyrimidines by growth on orotate, which is a poor pyrimidine source for *B. subtilis* (8, 9). Growth of the strain with cytidine, on the other hand, represses *pyrG* expression. High-copy plasmids containing the *pyrG* regions were used to produce transcripts at levels much higher than the background levels of transcripts produced from the wild-type chromosomal *pyrG* operon also present in strain HC-11 (which were barely detectable under the conditions used). The transformed cells were grown in minimal medium containing cytidine or orotate, i.e., repressing

* Corresponding author. Mailing address: Department of Biochemistry, University of Illinois, 600 South Mathews, Urbana, IL 61801. Phone: (217) 333-3940. Fax: (217) 244-5858. E-mail: rswitzer@uiuc.edu.

[†] Present address: Institute for Microbiology, Ernst-Moritz-Arndt-University, Friedrich-Ludwig-Jahn-Strasse 15, D-17487 Greifswald, Germany.

[∇] Published ahead of print on 8 December 2006.

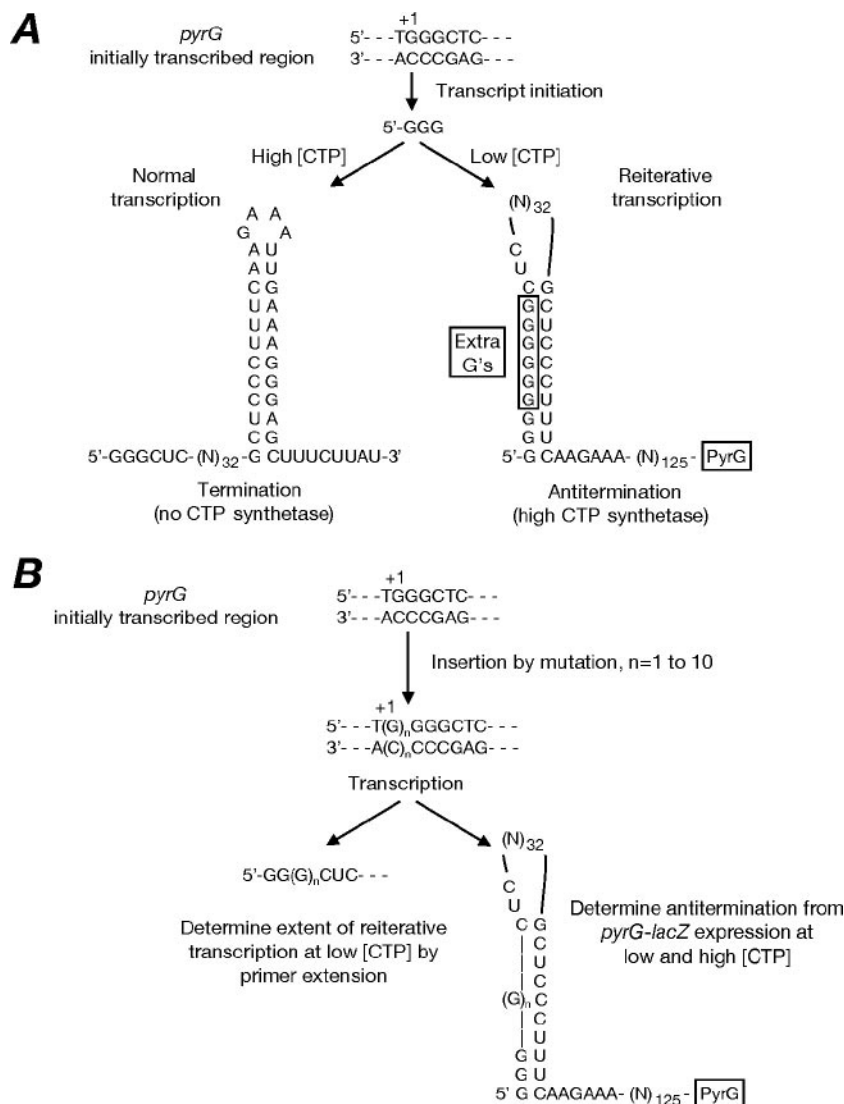


FIG. 1. (A) Proposed mechanism for the regulation of *B. subtilis pyrG* expression by CTP-sensitive reiterative transcription and antitermination (from reference 10). The 5' sequence of transcripts specified by the *pyrG* ITR depends on the intracellular CTP concentration. If the CTP level is high (left), a C residue is rapidly inserted at position +4 and the transcript encoded by the DNA is synthesized. This sequence permits formation of the leader terminator and reduced *pyrG* expression. If the CTP level is low (right), transcription pauses at +4, leading to reiterative transcription of a variable number of G residues. Transcripts bearing extra G residues are able to fold into an antiterminator hairpin, which suppresses termination and results in increased transcription of the downstream *pyrG* coding region. The nucleotides denoted by (N)₃₂ apparently play no role in regulation, because *pyrG-lacZ* fusions from which they were deleted retained normal regulation (9). (B) Rationale for experiments reported in this work. Mutational insertion of 1 to 10 additional G residues at +1 in the *pyrG* ITR leads to formation of transcripts with a known number of extra 5' G residues. The effects of such mutations on reiterative transcription were determined by primer extension mapping (left). The effects of such mutations on *pyrG* expression were assessed with *pyrG-lacZ* fusions grown under repressing and derepressing conditions (right). Note that numerous antiterminator structures can form, depending on the number of G residues inserted and the exact base pairing.

and derepressing conditions, respectively, and RNA was harvested from the cells as previously described (10). A single-stranded DNA (primer C) complementary to nucleotides +15 to +34 of the *pyrG* transcript was chemically synthesized and 5' end labeled with [γ -³³P]ATP and T4 polynucleotide kinase. The labeled DNA was used as the primer for avian myeloblastosis virus reverse transcriptase, and the primer-extended products were analyzed by electrophoresis as described previously (10). Transcripts specified by the wild-type *pyrG* plasmid and harvested from cells grown with excess cytidine mapped to the +1 position of the ITR, as previously reported (8, 10). In

this assay transcripts that were one to four nucleotides longer were also detected in cells grown with excess cytidine (Fig. 2, WT +C). These longer transcripts may indicate that limited reiterative transcription occurred even under repressing conditions, but we believe that they are largely an artifact of using templates borne on multicopy plasmids, because they were much less prominent when transcripts from the chromosomal *pyrG* operon were analyzed (8, 10). Extensive reiterative transcription occurred in pyrimidine-starved cells with the wild-type template, as shown by the formation of a ladder of primer-extended products, corresponding to transcripts with extensions up

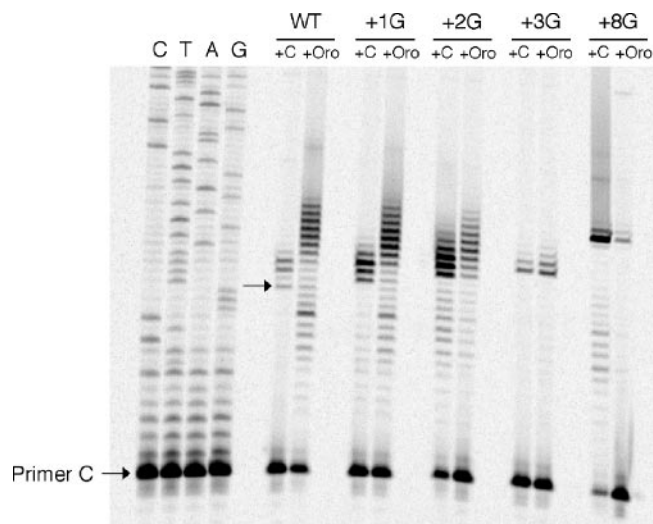


FIG. 2. Primer extension mapping of *pyrG* transcripts extracted from the pyrimidine-requiring *B. subtilis* strain HC-11 bearing pEB112-derived plasmids that contained the promoter-leader region of wild-type *pyrG* (WT) or the corresponding region with (G)_n insertion mutations (+1G, +2G, +3G, and +8G). Cells were grown on minimal medium (8) at 37°C with excess cytidine (+C) or orotate (+Oro), a poor pyrimidine source, and were harvested in mid-exponential phase as previously described (8). The sequencing ladder is labeled to correspond to the sequence of the nontranscribed strand. The arrow denotes G +1, the previously identified (10) transcription initiation site encoded by the *pyrG* gene. The position of 5'-end-labeled primer C (10), used both for sequencing and primer extension, is also shown.

to 11 nucleotides long. We have previously demonstrated that these extensions are formed by the addition of extra 5' G residues (10). The most abundant poly(G) extensions ranged from 4 to 10 nucleotides long (Fig. 2, WT +Oro).

When the transcripts specified by plasmids carrying *pyrG* ITRs containing one, two, three, or eight G insertions were mapped, the shortest transcripts from cells grown under repressing conditions were longer than the wild-type transcripts by one, two, three, and eight nucleotides, respectively, as expected. As with the wild-type template, some marginally extended transcripts were formed in repressed cells when the template contained insertions of one or two G residues, but these transcripts were much less abundant with the templates containing three and eight G insertions. This observation indicates that an ITR sequence of longer than five G residues suppresses reiterative transcription. This conclusion is reinforced by the mapping of transcripts from the same strains grown under derepressing conditions. A ladder of longer transcripts similar to that seen with the wild-type template was formed in starved cells containing the *pyrG* template with a single G insertion. When the template contained two inserted G residues, the ladder of transcripts also indicated the occurrence of reiterative transcription in pyrimidine-starved cells, but the ladder was somewhat shorter than that found with the wild-type template, and only a few of the transcripts were longer than those observed with repressed cells. When the *pyrG* template contained three or eight inserted G residues, little reiterative transcription occurred under any condition.

It was demonstrated previously that reiterative transcription

at the *pyrG* ITR requires a run of three G residues at the site of transcription initiation, followed by a C or T residue (10). Here we have demonstrated that a run of four or five G residues also permits reiterative transcription, but a run of six or more G residues suppresses it. As noted above, reiterative transcription requires melting of the hybrid formed by the DNA template and the nascent transcript, so the greater stability of the (rG)_n · (dC)_n duplex when n is greater than 5 may preclude reiterative transcription by suppressing melting.

The effect of various G insertions on expression of corresponding *pyrG'*-*lacZ* fusions integrated into the *B. subtilis* chromosome was also determined. Derivatives of the pDH32 integration plasmid (2) that contained *pyrG'*-*lacZ* fusions bearing either the wild-type *pyrG* promoter-leader region (nucleotides -49 to +81), which specifies transcripts initiating with 5'-GGGCUC, or mutated *pyrG* promoter-leader regions that specified transcripts initiating with 5'-GGG(G)_nCUC (where n equals 1, 2, 3, 4, 6, 8, and 10) were constructed using previously published procedures (8, 9). DNA sequencing confirmed the identities of the *pyrG'* regions of the plasmids. The plasmids were integrated into the *amyE* locus of the chromosome of *B. subtilis* strain HC-11 (*pyrB*::*Spc*^r) as described previously (8). (Note that two of these strains, QM401 and QM425, were prepared previously [8, 10].) The integrant strains were grown on minimal medium containing 200 μg of cytidine per ml (repressing conditions) and on the same medium containing 100 μg of orotate per ml (derepressing conditions), and β-galactosidase assays were used to assess *pyrG* expression under both conditions using published methods (8). As seen in previous studies (8, 9), expression of the wild-type *pyrG'*-*lacZ* fusion was derepressed 14-fold by growth under pyrimidine-limiting conditions, relative to growth on excess cytidine (Table 1). Insertion of even a single G residue in the *pyrG* ITR led to substantial derepression of expression. Insertion of two or more G residues resulted in essentially constitutive *pyrG* expression (Table 1). (The residual 1.8- to 2.5-fold-higher level of expression in pyrimidine-starved cells containing three or more inserted G residues was previously observed in a *pyrG'*-*lacZ* fusion in which the attenuator terminator was deleted [8] and results from a phenomenon unrelated to transcription attenuation.) The highest level of *pyrG* expression under both re-

TABLE 1. Effects of G insertion mutations on expression of *pyrG'*-*lacZ* fusions in repressed and pyrimidine-starved *B. subtilis* HC-11 cells

Strain (insertion)	β-Galactosidase activity (Miller units) ^a with:		Fold regulation
	Cytidine	Orotate	
QM401 (none [wild type])	41 ± 4	574 ± 34	14
AE101 (+1G)	220 ± 25	1,240 ± 82	5.6
AE102 (+2G)	415 ± 30	900 ± 60	2.2
AE103 (+3G)	500 ± 29	910 ± 48	1.8
QM425 (+4G)	800 ± 27	1,800 ± 115	2.2
AE104 (+6G)	645 ± 45	1,610 ± 103	2.5
AE105 (+8G)	560 ± 63	1,130 ± 66	2
AE106 (+10G)	480 ± 54	870 ± 35	1.8

^a The data are the combined results of duplicate assays from three replicate experiments (mean ± standard deviation).

pressing and derepressing conditions was observed with the *pyrG'*-*lacZ* fusion containing four inserted G residues.

Interpretation of the effects of the G insertions on *pyrG'*-*lacZ* expression is complex, because the insertions have two consequences that are expected to have opposite effects on *pyrG* expression. Reiterative transcription is necessary to generate the 5' poly(G) tails on *pyrG* transcripts that act as the antiterminator and suppress the downstream attenuator. From this consideration alone, one would predict that the fusions in which reiterative transcription occurred, i.e., those with one and two G insertions (Fig. 2, +1G and +2G), would behave more or less like the wild type and that those containing three or more G insertions, in which reiterative transcription was suppressed (Fig. 2, +3G and +8G), would fail to antiterminate and would be repressed under all growth conditions. In fact, all of the G insertion mutants were expressed at elevated levels; that is, they failed to terminate normally under repressing conditions. The *pyrG'*-*lacZ* fusion containing a single G insertion showed some residual regulation, which presumably resulted from its ability to carry out reiterative transcription, but was expressed at an elevated level under repressing conditions, compared to the wild-type fusion. Insertions of two or more G residues led to constitutive expression, whether reiterative transcription occurred or not. Examination of the patterns of transcript formation in these insertion strains (Fig. 2) leads us to suggest that derepression results from formation of *pyrG* transcripts containing six or more G residues at their 5' ends, whether these result from reiterative transcription (wild type and one and two G insertions) or are encoded in the *pyrG* template (three or more G insertions). While a 5'-GGGGGG segment of the transcript is apparently sufficient to serve as an antiterminator by base pairing with the downstream GCUCC CUUUC segment of the attenuator hairpin, the highest level of expression was obtained with the four-G insertion mutant (Table 1), corresponding to a 5' poly(G) tract of seven residues. The most abundant poly(G) tracts formed by reiterative transcription in pyrimidine-starved cells with the wild-type *pyrG* template contained 7 to 13 G residues (Fig. 2, WT +Oro), which are clearly sufficient for optimal antitermination.

It is evident that the *pyrG* DNA has been selected to contain the optimal number of rG-dC base pairs specifying the 5' terminus of its transcripts; that number is three. A *pyrG* template specifying two G residues in the transcript, i.e., one in which one of the three wild-type G residues is mutated to A, fails to carry out significant reiterative transcription (10), cannot antiterminate, and is expressed at very low levels (9). A *pyrG* template specifying four G residues is capable of reiterative transcription but antiterminates to a significant degree even under repressing conditions, leading to suboptimal regulation. Templates specifying five or more G residues are progressively more defective in reiterative transcription and antiterminate under all conditions.

Our studies identified that a 5' poly(G) sequence containing six G residues is highly effective in antitermination of *pyrG* expression, and a 5' poly(G) sequence of seven to nine G residues is optimal for antitermination. However, the stability of the putatively optimal antiterminator stem-loop structure with (G)₆C base paired to the terminator sequence GCUCC CUUUC, as calculated using Mfold (12), is approximately equal to the calculated stability of the *B. subtilis pyrG* leader

terminator (attenuator) stem-loop, whereas antiterminator structures with fewer G residues are progressively weaker by about 0.5 kcal/mol per G residue. (Numerous possible structures are predicted by Mfold, depending on the exact sequences folded and constraints used, so such computations should be interpreted cautiously.) We suggest that the competition in vivo between antiterminator and terminator conformations of *pyrG* leader RNA does not occur as a simple thermodynamic equilibrium between the two folded forms. If it did, termination would be expected to equal or exceed antitermination even under derepressing conditions. Rather, we predict that the system is kinetically controlled. We propose that during pyrimidine starvation reiterative transcription leads to formation of 5' poly(G) tracts on the transcripts that are able to form the antiterminator structure before the terminator is fully transcribed and emergent from RNA polymerase. By the time the terminator is able to form and compete with the antiterminator structure, transcription has proceeded through the terminator sequence, and RNA polymerase is committed to transcription of the entire downstream *pyrG* coding region. Only in the absence of the upstream 5' poly(G) tracts and the consequent absence of interfering (albeit transient) antiterminator formation is the *pyrG* leader terminator able to form quickly enough to terminate within the leader region.

Reiterative transcription occurs when nucleotides are added repetitively to the 3' end of a nascent transcript after that transcript slips toward the 5' end relative to the DNA template strand. Slippage of the transcript generally occurs when RNA polymerase pauses because it is unable to add the next encoded nucleotide. The slippage site is a homopolymeric sequence, at least three nucleotides in length, on the template that specifies a complementary sequence in the transcript (GGG in *pyrG*), which undergoes elongation to longer homopolymeric segments (4). Very few systematic studies of the dependence of reiterative transcription on the length of the homopolymeric slippage site in the template are available, and most have studied reiterative transcription in vitro, not in vivo, as in this work. We have identified only one other example of reiterative transcription yielding poly(G) addition at the 5' end of transcripts, namely, that catalyzed by T7 RNA polymerase with GTP only and a template specifying 5'-GGGN transcripts (7). In this case a template specifying two Gs or one G sharply reduced or abolished reiterative transcription, respectively. Templates specifying a sequence of four or more Gs were not studied. Reiterative transcription has been observed more frequently when the template specifies a tract of U or A residues at the 5' end of the transcript (4). In several such cases the effects of substitution mutations in the homopolymeric segments of the template documented that a run of at least three T or A residues was required (1, 3, 11). However, only in the studies by Cheng et al. (1) of reiterative transcription and regulation of the *Escherichia coli upp* promoter was the effect of length of the homopolymeric segment of T residues in the template systematically examined. These workers found that a minimum of three T residues in the template was required for reiterative transcription of U residues in the *upp* transcript. Stepwise variation in the number of template Ts from three to eight did not alter reiterative transcription significantly in vitro but appeared to increase it in vivo, as judged from the regu-

lation of corresponding *upp-lacZ* fusions. This length dependency is quite different from that observed in our studies with reiterative transcription of G residues with the *pyrG* template. From these differences it seems likely that the detailed dependence of reiterative transcription on template sequence will differ from system to system. It should be noted, however, that the mechanism of regulation of *upp* expression by reiterative transcription differs markedly from the mechanism of *pyrG* regulation. In the *upp* system reiterative transcription of U residues at high UTP levels leads to abortive initiation and reduced expression, whereas reiterative transcription at the *pyrG* ITR occurs at low CTP levels and leads to antitermination and elevated expression.

We gratefully acknowledge Qi Meng for providing *B. subtilis* strains QM401 and QM425 and developing the methods used in this work and Charles L. Turnbough, Jr., for valuable comments on the manuscript.

This research was supported by Public Health Service grant GM47112 from the National Institute of General Medical Studies.

REFERENCES

1. Cheng, Y., S. M. Dylla, and C. L. Turnbough, Jr. 2001. A long T-A tract in the *upp* initially transcribed region is required for regulation of *upp* expression by UTP-dependent reiterative transcription in *Escherichia coli*. *J. Bacteriol.* **183**:221–228.
2. Grandoni, J. A., S. B. Fulmer, V. Brizzio, S. A. Zahler, and J. M. Calvo. 1993. Regions of the *Bacillus subtilis* *ilv-leu* operon involved in regulation by leucine. *J. Bacteriol.* **175**:7581–7593.
3. Guo, H. C., and J. W. Roberts. 1990. Heterogeneous initiation due to slippage at the bacteriophage 82 late promoter in vitro. *Biochemistry* **27**:10702–10709.
4. Jacques, J.-P., and D. Kolakofsky. 1991. Pseudo-templated transcription in prokaryotic and eukaryotic organisms. *Genes Dev.* **5**:707–713.
5. Jørgensen, C. M., K. Hammer, and J. Martinussen. 2003. CTP limitation increases expression of CTP synthase in *Lactococcus lactis*. *J. Bacteriol.* **185**:6562–6574.
6. Leonhardt, H., and J. Alonso. 1988. Construction of a shuttle vector for inducible gene expression in *E. coli* and *B. subtilis*. *J. Gen. Microbiol.* **134**:605–609.
7. Martin, C. T., D. K. Muller, and J. E. Coleman. 1988. Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* **27**:3966–3974.
8. Meng, Q., and R. L. Switzer. 2001. Regulation of transcription of the *Bacillus subtilis* *pyrG* gene encoding cytidine triphosphate synthetase. *J. Bacteriol.* **183**:5513–5522.
9. Meng, Q., and R. L. Switzer. 2002. *cis*-acting sequences of *Bacillus subtilis* *pyrG* mRNA essential for regulation by antitermination. *J. Bacteriol.* **184**:6734–6738.
10. Meng, Q., C. L. Turnbough, Jr., and R. L. Switzer. 2004. Attenuation control of *pyrG* expression in *Bacillus subtilis* is mediated by CTP-sensitive reiterative transcription. *Proc. Natl. Acad. Sci. USA* **101**:10943–10948.
11. Xiong, X. F., and W. S. Resnikoff. 1993. Transcriptional slippage during the transcription initiation process at a mutant *lac* promoter in vivo. *J. Mol. Biol.* **231**:569–580.
12. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**:3406–3415.