[Frontiers in Bioscience 8, d20-31, January 1, 2003]

THE T BOX AND S BOX TRANSCRIPTION TERMINATION CONTROL SYSTEMS

Frank J. Grundy and Tina M. Henkin

Department of Microbiology, The Ohio State University, 484 W. 12th Avenue, Columbus, OH 43210 USA

TABLE OF CONTENTS

1. ABSTRACT

The T box and S box transcription termination control systems are widely used for control of gene expression in Gram-positive bacteria, but are rare in Gram-negative organisms. Both of these systems can be recognized in genomic data because of high conservation of primary sequence and structural elements. The T box system regulates a variety of amino acid-related genes, while the S box system is dedicated to genes involved in methionine metabolism. While both systems involve gene regulation at the level of premature termination of transcription, the molecular mechanisms employed are very different. In the T box system, expression is induced by stabilization of an antiterminator structure in the leader by interaction with the cognate uncharged tRNA; this prevents formation of the competing terminator helix, allowing synthesis of the full-length mRNA. Disruption of conserved leader features results in loss of readthrough. In the S box system, the antiterminator form of the leader is the more stable form. A competing anti-antiterminator must be stabilized by an unknown factor during growth in methionine to prevent formation of the antiterminator, thereby allowing formation of the terminator helix. Disruption of conserved leader elements results in constitutive expression.

2. INTRODUCTION

Control of transcription at the level of premature termination has emerged as a common strategy for gene regulation in prokaryotes. A variety of mechanisms for

modulating transcription termination have been uncovered. These include modification of the transcription machinery, effects on transcription from translation of the leader RNA, and interactions of the nascent transcript with RNA binding proteins or other RNAs (1, 2). In a number of cases, the mechanisms found are not randomly distributed throughout the bacterial world, but instead are clustered in certain groups. For example, transcriptional attenuation mediated by leader peptide translation is common in Gram-negative bacteria, but is rare in Gram-positive organisms, while the T box tRNA-dependent mechanism is common in Gram-positive organisms but is very rare in Gram-negative organisms. Similarly, systems like *Bacillus subtilis* TRAP, where an RNA binding protein promotes transcription termination, appears to be restricted to *Bacillus* species (3, 4). While the basis for the selective distribution of these regulatory systems is not clear, analysis of a variety of evolutionarily distant organisms provides new insights into how gene regulation at this level can be accomplished. This article will focus on the T box and S box transcription termination control systems, both of which are widely distributed in Gram-positive bacteria but are found only rarely in Gram-negative bacteria.

3. THE T BOX SYSTEM

The T box transcription termination control system was first uncovered by analysis of the *B. subtilis tyrS* gene, encoding tyrosyl-tRNA synthetase. Features of the *tyrS* leader, including an intrinsic transcriptional terminator and the T box itself, a conserved 14 nt element for which the system is named, were found upstream of multiple aminoacyl-tRNA synthetase (AARS) genes in *Bacillus* species, as well as in the *ilv-leu* biosynthesis operon (5). The conservation of the leader arrangement provided the first indication that this set of genes might be regulated by a common mechanism (5, 6). The fact that each gene in the group appeared to respond independently to limitation for the cognate amino acid, rather than to general amino acid starvation, indicated that there must be amino acid specificity in the regulatory response.

3.1. A conserved pattern in T box family leader RNAs

Comparative analysis of the initial set of 10 leaders revealed a complex pattern of highly conserved primary sequence and secondary structural elements, preceding the T box sequence and the intrinsic transcriptional terminator. The major structural elements were designated Stem I, Stem II, Stem IIA/B, and Stem III (7, 8). An antiterminator structure, which is formed by pairing of a portion of the T box sequence with residues on the 5' side of the terminator helix, was also identified (7); the terminator and antiterminator structures are mutually exclusive. The secondary structure model of the *B. subtilis tyrS* leader is shown in Figure 1. The terminator helix of each leader is predicted to be significantly more stable than the antiterminator structure and represents the default state of the leader RNA. Synthesis of an mRNA including the downstream coding region requires stabilization of the antiterminator, which prevents terminator formation.

3.1.1. A single codon provides the basis for amino acid-specific regulation

The similarities of the leader structures of multiple T box genes supported the idea that regulation operated by a common mechanism. The solution to the issue of amino acid specificity was provided by the recognition that each leader, in a specific position within an internal loop of Stem I, contains a triplet sequence that represents a codon for the amino acid matching the amino acid class of the downstream coding region, so that all tyrosyl genes contain a UAC tyrosine codon, leucine genes contain a CUC leucine codon, etc. (7). A single nucleotide substitution of the UAC tyrosine codon of the *tyrS* leader to a UUC phenylalanine codon resulted in loss of induction in response to limitation for tyrosine, and gain of a response to limitation for phenylalanine. This demonstrated that the codon, designated the "specifier sequence," is a major determinant of the specificity of the amino acid response. This type of analysis was extended to many other substitutions in *tyrS* and other model leaders, and the general result is that alteration of the specifier sequence can direct a switch in the amino acid specificity in some but not all cases, and that regulation is generally not as efficient, indicating that there may be additional determinants of the amino acid response (9-13).

3.1.2. The codon functions in the absence of translation

The identification of a codon as the major determinant of amino acid specificity suggested the obvious model that the leader region is translated, and that the efficiency of translation of the codon controls the termination/antitermination decision in a manner analogous to the *E. coli trp* attenuation system (14). Comparative analysis of the leaders did not reveal likely signals for translation initiation, and there was no conservation of the placement of open reading frames containing the specifier sequence. A simple test of this idea involved introduction of an extra nucleotide into the *tyrS* leader, immediately upstream of the UAC tyrosine codon. This insertion was predicted to result in a shift in the reading frame, if the leader was translated, and loss of induction in response to tyrosine limitation. Regulation was unaffected, indicating that the specifier sequence was unlikely to act at the level of translation (7). This conclusion was further supported by direct tests for translational activity of the *ilv-leu* leader (15).

3.1.3. The effector is uncharged tRNA

The discovery that a single codon is the major determinant of amino acid specificity suggested that the most likely effector in the system is tRNA. Since the T box genes respond to amino acid limitation, the charging ratio of the cognate tRNA was a logical physiological signal. The role of tRNA as a regulator was confirmed by introduction of a nonsense codon at the position of the UAC specifier sequence in the *tyrS* leader. This mutation resulted in very low, uninducible expression in a wild-type strain (lacking a nonsense suppressor tRNA), indicating that expression is dependent on the presence in the cell of a tRNA that matches the specifier sequence (7). Since expression of these genes is induced in response to amino acid limitation, the requirement for tRNA for expression

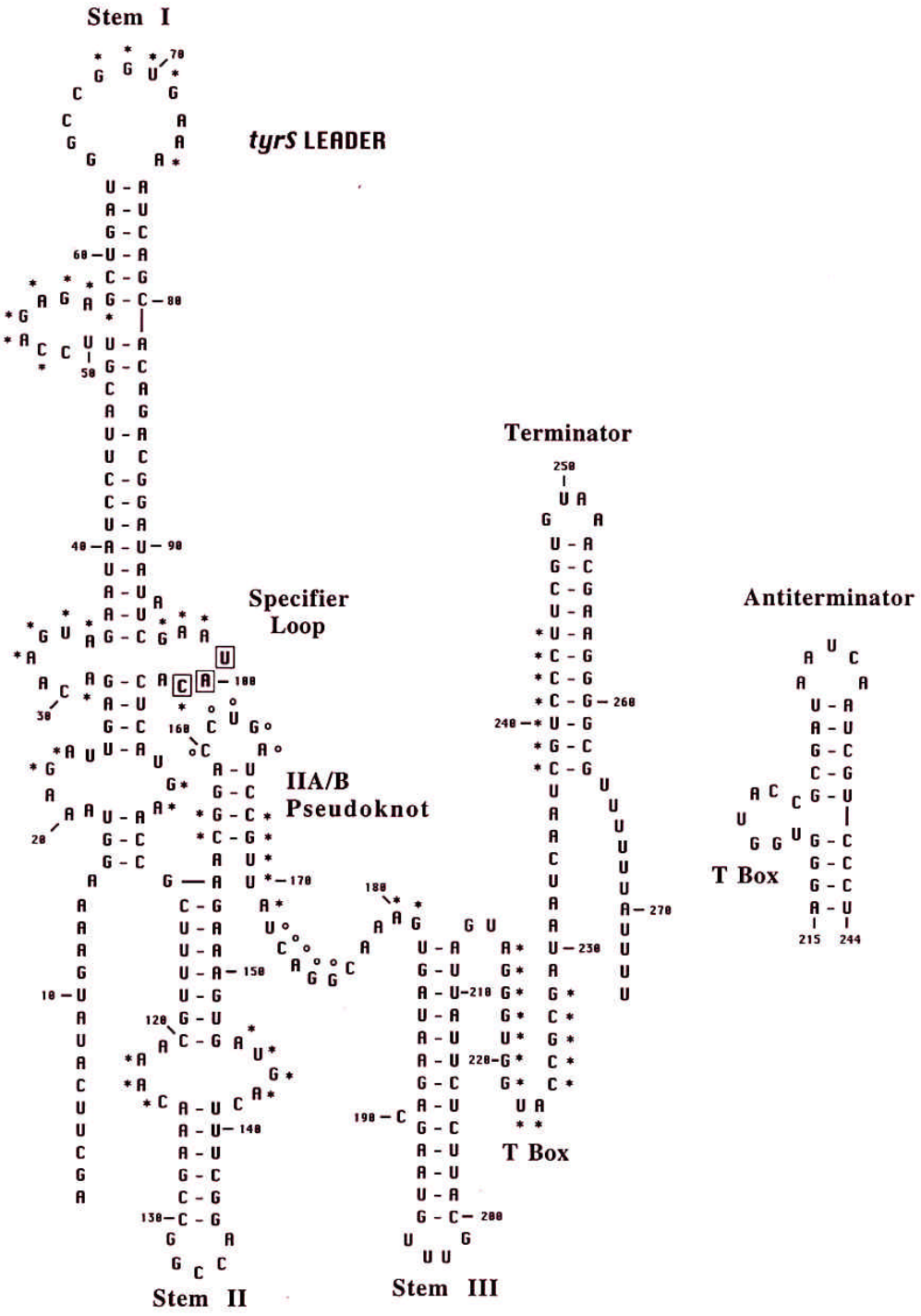

Figure 1. Structural model of the *B. subtilis tyrS* leader RNA. The leader is shown from the transcription start-point through the end of the transcriptional terminator; the TyrRS coding region begins further downstream. The leader is shown in the terminator form; the antiterminator form is shown to the right. The major structural features, Stem I, Stem II, the IIA/B pseudoknot, and Stem III are labeled. Asterisks indicate primary sequence elements conserved in most leaders; the T box is the most highly conserved element, and is labeled in both the terminator and antiterminator forms to highlight the alternate base-pairing arrangement. The UAC tyrosine codon, which serves as the specifier sequence, is boxed within the specifier loop. The IIA/B region is shown in the IIA form; the alternate pairing to form the IIB helix (or pseudoknot) is marked with small circles.

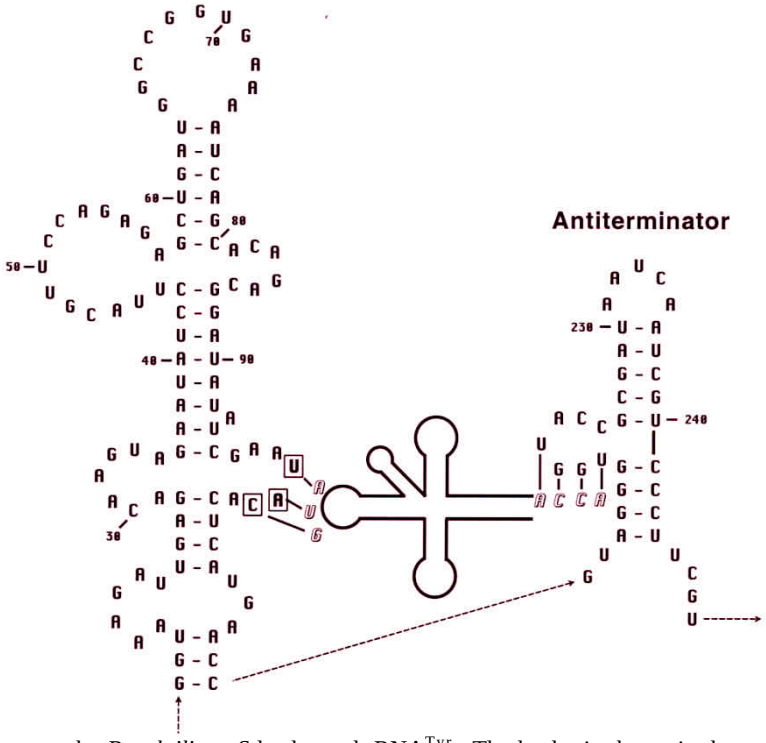

Figure 2. Interactions between the *B. subtilis tyrS* leader and $tRNA^{Tyr}$. The leader is shown in the antiterminator form (which interacts with the tRNA); the central portion of the leader (Stems II, IIA/B and III) are omitted for clarity. The known base-pairing interactions between the specifier sequence (boxed) and the anticodon of the tRNA (open letters), and between the acceptor end of the tRNA (open letters) and the antiterminator bulge, are shown.

suggested that uncharged tRNA might act as an inducer for the system. Introduction of a matching nonsense suppressor tRNA could partially overcome this effect, providing clear evidence that expression is dependent on codon-anticodon pairing between the specifier sequence and a matching tRNA (7).

Further support for the model that tRNA is the effector in the system was provided by isolation of mutations that result in overexpression of the *B. subtilis ilv-leu* operon (16). Several mutations were identified in the gene encoding $tRNA^{Leu}_{GAG}$ (which matches the CUC specifier sequence in the *ilv-leu* leader). These mutations were predicted to result in reduced charging of $tRNA^{Leu}_{GAG}$ in the absence of limitation for leucine, consistent with the idea that uncharged tRNA serves as the inducer. Zahler and coworkers also identified mutations in the *leuS* gene, encoding leucyl-tRNA synthetase, that similarly resulted in constitutive *ilv-leu* expression, presumably because of a decrease in the charging ratio of $tRNA^{Leu}$ (17).

A direct demonstration that the response to amino acid limitation is mediated by uncharged tRNA was obtained by experiments in which variants of $tRNA^{Tyr}$ were expressed in *B. subtilis* under the control of an inducible promoter. Expression of an unchargable variant of $tRNA^{Tyr}$ efficiently induced expression of a *tyrS-lacZ* transcriptional fusion during growth in rich medium (18). This indicates that the requirement for amino acid limitation acts by increasing the relative level of uncharged tRNA, since it can be bypassed by providing uncharged tRNA.

Measurements of tRNA charging ratios *in vivo* under inducing conditions by Northern analysis have provided further evidence that induction of T box genes occurs under conditions where the charging ratio of the cognate tRNA has been reduced (13, 19; M. T. Haldeman, F. G. and T. H., unpublished results). Although these studies are not highly quantitative, they indicate that a fairly modest reduction in the charging ratio is sufficient to promote efficient antitermination, suggesting that the mechanism is quite sensitive.

3.1.4. How does uncharged tRNA promote antitermination?

A second region of possible base-pairing between the tRNA and the leader was recognized by comparative analyses. A highly conserved bulge in the antiterminator contains the sequence UGGN, where the UGG triplet is 100% conserved and the N is variable. This region is complementary to the NCCA sequence comprising the 3' end of tRNA, where the N is variable (but is often a key identity determinant for recognition by the cognate AARS). The variable positions of the antiterminators and tRNAs were found to covary, and mutational analysis demonstrated that pairing at the N positions is required for efficient antitermination (18). This provided genetic evidence for a second interaction between the tRNA and the leaders (Figure 2).

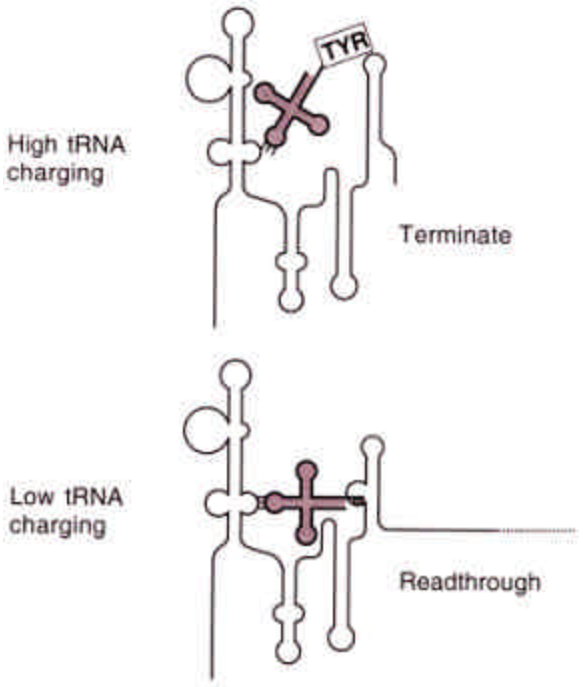

Figure 3. Model for *B. subtilis tyrS* antitermination control. When the charging ratio of $tRNA^{Tyr}$ is high, charged tRNA interacts with the specifier sequence, but is unable to interact with the antiterminator. The terminator helix forms, and transcription terminates, blocking expression of the downstream coding region. When the charging ratio of $tRNA^{Tyr}$ is low, uncharged tRNA interacts with the specifier sequence and the antiterminator bulge, stabilizing the antiterminator and preventing formation of the competing terminator helix. Transcription continues past the leader region terminator, and the full-length transcript encoding TyrRS is synthesized. Other genes in the T box family are regulated by an analogous mechanism, with a specific response to the charging ratio of the cognate tRNA.

The 3' A of the tRNA is the position of amino acid attachment in charged tRNA, and the pairing of this residue to the first U of the antiterminator bulge was predicted to abut the adjoining helical domain of the antiterminator. This pairing was therefore proposed to discriminate between charged and uncharged tRNA, with the amino acid of charged tRNA preventing a productive interaction with the antiterminator (18). Charging of the tRNA may also result in a conformational change in the tRNA, and results in binding by EF-Tu, either of which could interfere with antitermination activity.

3.1.5. Does charged tRNA play a role?

Several lines of evidence suggest that charged tRNA may also influence T box antitermination. Induction of *B. subtilis tyrS* expression by an unchargable tRNA is partially blocked if the cell contains a chargable tRNA matching the specifier sequence of the leader (18); similar results were observed for the *Lactococcus lactis trp* system (13). In addition, the constitutive expression of the *ilv-leu* operon conferred by mutations in $tRNA^{Leu}_{GAG}$ that reduced tRNA charging was lost if a second copy of the wild-type tRNA was introduced (16). Since the key difference between charged and uncharged tRNA is likely to be exerted at the acceptor end of the tRNA, a possible explanation for these results is that either charged or uncharged tRNA can make the initial interaction at the specifier sequence, but only uncharged tRNA can make the second productive interaction at the antiterminator. In this case, charged tRNA could interfere with antitermination by uncharged tRNA by competing for access to the specifier sequence, so that the system responds to the charging ratio of the matching tRNA rather than to the absolute levels of uncharged tRNA. The response to the charging ratio of the tRNA is also consistent with data showing that overexpression of certain AARSs results in repression (9, 12, 20); this effect is likely to be mediated at the level of tRNA charging, rather than by a direct interaction of the AARS with the leader.

3.2. The basic model for T box antitermination

Multiple amino acid-related genes with the conserved features of the T box system can be found in a single genome; there are 19 T box genes in *Bacillus subtilis* (21). Each gene is regulated independently in response to the charging ratio of the cognate tRNA, so that uncharged tRNA serves as the signal for increased readthrough. An example of how these systems function is illustrated with *B. subtilis tyrS* and $tRNA^{Tyr}$ (Figure 3). The tRNA-leader RNA interaction occurs in the absence of translation, so that this system represents a novel example of codon-anticodon pairing.

3.3. Leader RNA requirements

Leader RNAs in the T box family contain a complicated array of conserved features (Figure 1). Most of the conservation is at the level of secondary structure, as demonstrated by covariation (7; unpublished results). The basic structural pattern predicted by comparative analysis has been confirmed by chemical and enzymatic probing of the *B. subtilis thrS* leader RNA *in vivo* and *in vitro* (22). Mutations that disrupt the conserved elements in the context of the *tyrS* leader generally result in a decrease or abolition of tRNA-dependent antitermination, indicating that these elements are important for function (8, 23; unpublished results); similar results have been obtained for other T box family genes (9, 15, 24, 25). Certain leader features are absent from or modified in individual leaders or groups of leaders, but the majority of the leaders identified to date fit the basic pattern represented by the *B. subtilis tyrS* leader (unpublished results). The role of most of these conserved elements is not clear. Obvious possibilities include roles in tRNA binding, proper folding of the leader RNA, binding of additional factors that may be required for leader RNA-tRNA recognition, or interactions with the transcriptional machinery.

The most highly conserved leader sequence element is the T box itself, which forms the 5' side of the antiterminator and includes four nucleotides (UGGN) that pair with the acceptor end of the tRNA (NCCA). The distal three nucleotides (ACC) in the antiterminator bulge are also highly conserved; the A is 100% conserved, while the C's

vary occasionally (26). Mutations at any of these positions drastically affect antitermination, indicating that these residues are important for function. A small RNA modeled on the antiterminator region folds as predicted and is capable of sequence-specific binding to either full-length tRNA or an acceptor end mini-helix (27). Mutation of the bulge region disrupts tRNA binding *in vitro*, consistent with the mutational analyses *in vivo.* Certain residues in the antiterminator helices can be altered without a major effect on function, despite high conservation, while residues in the helix above the bulge are highly sensitive to mutation, suggesting that they play an important role in formation of the proper structure (26).

3.4. tRNA requirements

The ability of tRNAs other than $tRNA^{Tyr}$ to interact with the *tyrS* leader was investigated by alteration of the known leader RNA determinants for tRNA interaction (the specifier sequence and the antiterminator variable position) to match new tRNA types, and testing for induction in response to the new amino acid specficity (7, 11). This approach has been applied to a number of leaders, and the general result is that while the tRNA specificity can be switched in certain cases, other tRNAs may interact inefficiently with a given leader, or not at all (9-13). These results suggest that there are additional elements involved in the tRNA-leader interaction.

Mutational analysis was used to examine the sequence and structural requirements of $tRNA^{Tyr}$ for *tyrS* antitermination, using a system in which the function of the tRNA in antitermination could be examined independent of other requirements for tRNA synthesis and activity in the cell (28). Mutations in the helical regions of $tRNA^{Tyr}$ could be introduced as long as base-pairing was maintained. In addition, the long variable arm of $tRNA^{Tyr}$ could be replaced by either a short variable arm or a long helical insertion. However, alterations that affect the structure of the tRNA or the overall tertiary folding of the tRNA disrupted antitermination activity, suggesting that the leader-tRNA interaction is dependent on the entire tRNA tertiary structure. An interaction between the D and T arms of the tRNA with the leader was suggested to be required for *L. lactis trp* gene expression (29), but the proposed base-pairing is not supported by covariation analysis in other leaders or mutational studies.

3.5. RNA processing

Processing of the readthrough transcript has been reported for several genes in the T box family (13, 30, 31). The processing event for the *thrS* leader is RNase E-dependent in *E. coli*, but the enzyme responsible for this activity in *B. subtilis* has not yet been identified (32). The cleavage event has been mapped to the antiterminator for some genes, while for the *gltX-cysES* transcript cleavage occurs further downstream. The processing event does not appear to be dependent on the tRNA, but deletion of regions downstream of the processing site blocks the reaction for *thrS*, suggesting that there are elements in this region that are necessary for recognition by the cleavage enzyme (32). Processing is predicted to stabilize the readthrough transcript, amplifying the antitermination effect by increasing the functional life-time of the transcript.

3.6. Distribution of the T box system

The T box system can be identified in genomic sequences by searching for the T box sequence itself or by examination of regions upstream of genes known to be regulated by this mechanism in other organisms. T box genes have been most commonly found in low G+C Gram-positive bacteria, but are found rarely in high G+C Gram-positive species, in *Deinococcus radiodurans*, and in *Geobacter sulfurreducens*, a member of the delta-proteobacteria (unpublished results). The only group of Gram-positive organisms in which this system has not yet been found is the *Mycoplasma* group. The number of T box leaders in an individual organism varies from a single gene (in the high G+C group) to as many as 38 (in *Bacillus anthracis*). Several T box genes in organisms other than *Bacillus* have been characterized, and these in general exhibit the predicted features of the regulatory system, with a few interesting variations (13, 24, 25, 33).

Leaders exhibiting the characteristic features of the T box family have been found upstream of AARS genes, amino acid biosynthesis genes and genes likely to encode amino acid transporters. In addition, regulation by this system has been demonstrated for the *B. subtilis yczA* gene, which encodes a regulator of the activity of the TRAP attenuation protein, which in turn regulates *trp* biosynthesis gene expression (19, 34). All of the T box genes are in some way related to amino acid metabolism, consistent with a response to tRNA charging as the physiological signal. Most of the amino acids are represented in this family, with the notable exceptions of glutamyl and lysyl genes; glutaminyl genes are rare (unpublished results). The basis for this is unknown, but may reflect the ability of the corresponding tRNAs to engage in the leader-tRNA interaction. The distribution of amino acid class also varies in different groups of organisms; for example, genes involved in methionine metabolism in *Streptococcus* and *Enterococcus* are regulated by the T box system, while these genes in *Bacillus* or *Clostridium* sp. are instead regulated by the S box transcription termination control system.

4. THE S BOX SYSTEM

The identification of T box genes in *Enterococcus* and *Streptococcus* sp. with methionyl specificity led us to notice that the corresponding genes in *B. subtilis* did not appear to be members of the T box regulon. These genes contain leader regions 200 nt in length which include elements resembling intrinsic transcriptional terminators. However, the leaders lack the sequence and structural elements characteristic of T box family members and are instead closely related to each other. Eleven transcriptional units with this new set of features were identified in the *B. subtilis* genome (35). Several of these genes were predicted to encode proteins involved in methionine metabolism, while the remainder were genes of unknown function. The genes in this group

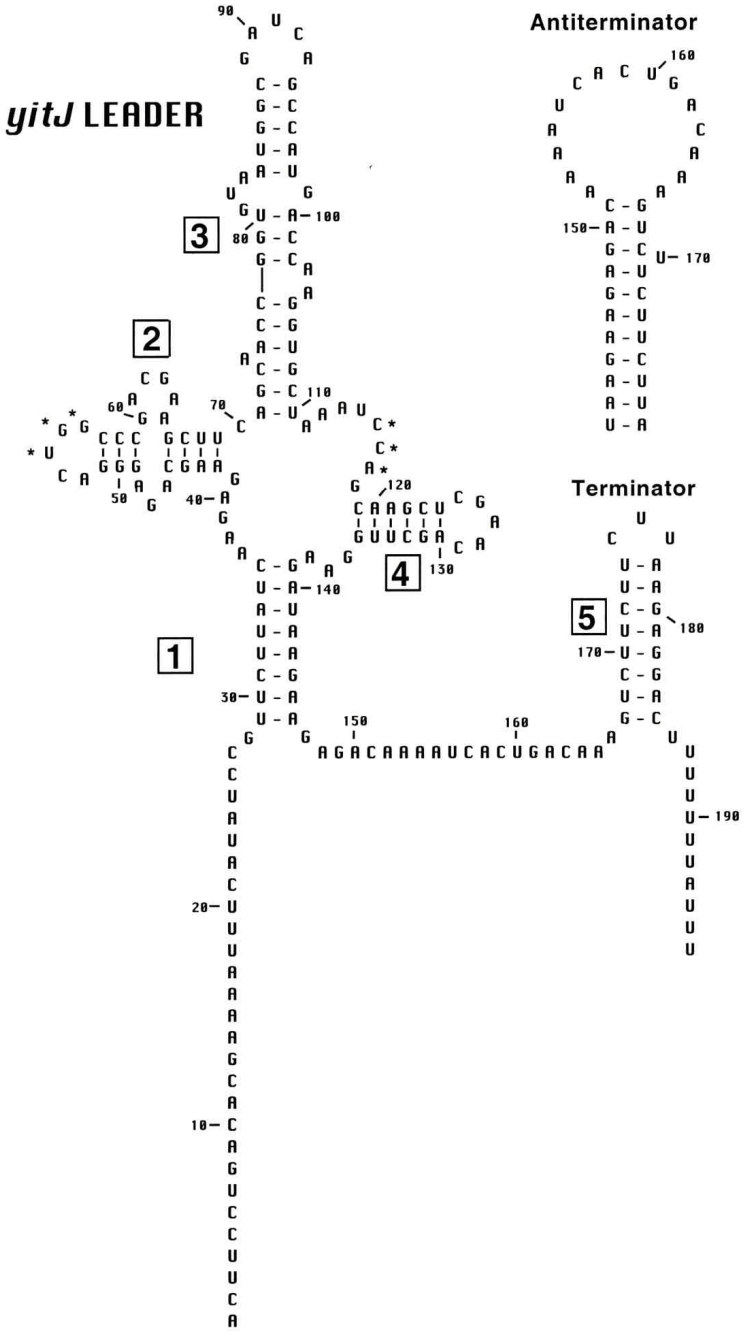

Figure 4. Structural model of the *B. subtilis yitJ* leader RNA. The leader is shown from the transcription start-point through the leader region terminator; the coding region is located further downstream. The leader is shown in the terminator form; the antiterminator, derived from pairing of sequences on the 5' side of the terminator (helix 5) with sequences on the 3' side of helix 1, is shown above the terminator. The core conserved region of S box leaders includes helices 1, 2, 3 and 4. The upper region of helix 3 is variable. The asterisks indicate residues that covary, suggesting that they may form a tertiary interaction.

were designated the S box regulon, since they appeared to be involved in sulfur metabolism.

4.1. S box leader requirements

Comparative analysis of the leaders in this set revealed that they could all form an antiterminator structure capable of competing with formation of the terminator helix (35; Figure 4 illustrates one example). The 5' side of the antiterminator could also participate in formation of another alternate structure, designated an anti-antiterminator, since formation of this structure competes with formation of the antiterminator. This led to a basic

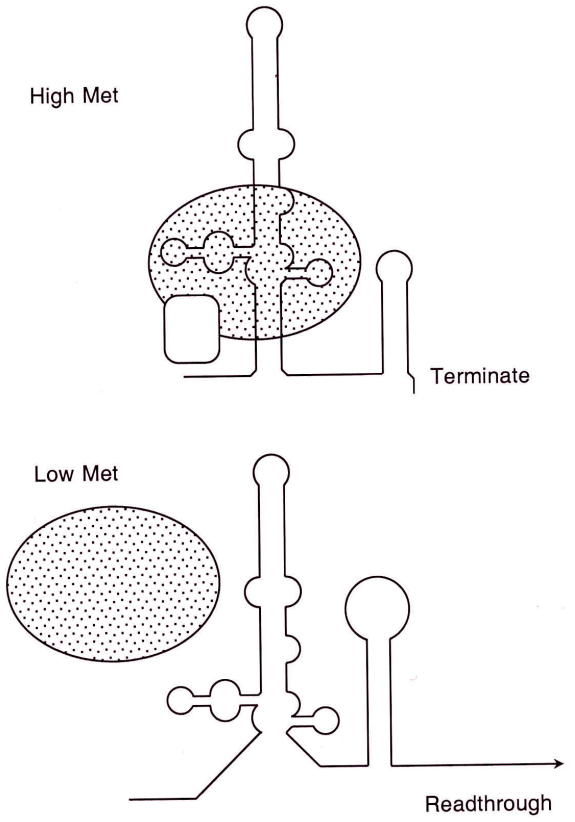

Figure 5. Model for S box antitermination control. When methionine (box) is abundant, an unknown factor binds to and stabilizes the anti-antiterminator form of the leader, preventing formation of the competing antiterminator, and allowing formation of the terminator helix. When methionine levels are low, the anti-antiterminator is destabilized, permitting formation of the antiterminator, which sequesters sequences required for formation of the terminator helix. Transcription continues past the terminator region, and the downstream genes are expressed. All S box genes are regulated in concert in response to methionine abundance.

model in which formation of the antiterminator, and readthrough, occurs when methionine is limiting; high levels of methionine were proposed to result in stabilization of the anti-antiterminator structure, which prevents formation of the competing antiterminator, thereby allowing formation of the terminator helix and premature termination of transcription (35). The anti-antiterminator region is postulated to form a complex cruciform structure with high conservation of both primary sequence and secondary structure.

Mutational analysis confirmed the basic predictions of the model (35). Mutations that disrupted the terminator or anti-antiterminator regions resulted in high level, constitutive expression, consistent with a negative role in expression, while mutations that disrupted the antiterminator resulted in low level, uninducible expression, consistent with requirement of this element for transcriptional readthrough. In addition, mutations within conserved elements of the anti-antiterminator domain that did not directly compete with the antiterminator also resulted in constitutive expression, suggesting that these elements are required for repression during growth in methionine (23, 35). These results are consistent with a model in which the conserved elements in the 5' region of the S box leaders serve as a binding site for a negative-acting regulatory factor, since disruption of these elements results in constitutive expression (Figure 5).

4.2. How does the S box system respond to methionine levels?

The nature of the regulatory factor, and the molecular mechanism for the response to methionine availability, remain to be identified. One approach to identification of the regulatory factor involved isolation of mutants in which expression of a *yitJ-lacZ* fusion was constitutive. One mutant that acts in *trans* has been identified (B. A. Murphy, F. J. G. and T. M. H., unpublished); this mutant was shown to affect multiple S box genes, and the increase in expression was dependent on a functional antiterminator, consistent with the phenotype expected if a general repressor of antitermination has been disrupted. It will be of great interest to identify the genetic lesion in this strain.

The only information concerning the physiological signal to which the S box regulon responds is the observation that transcriptional fusions exhibit very low expression during growth in the presence of methionine, and expression is strongly induced when a methionine auxotroph is grown in the absence of methionine. Expression is also induced when fusions are introduced into a methionine prototroph and cells are switched from growth medium containing methionine to medium from which methionine has been omitted; similarly, expression is quickly turned off when methionine is added to cultures grown in the absence of methionine. The lower magnitude of the response to methionine in a prototroph is presumed to be due to intracellular production of methionine. The response to methionine availability occurs at the level of transcriptional termination, since fusions in which the terminator is absent exhibit high level expression regardless of the presence of methionine (35).

It is not clear at this time whether methionine is the true molecular effector, or whether some other physiological parameter changes when methionine levels change. For example, regulation of methionine biosynthesis genes in *E. coli* responds in large part to levels of S-adenosylmethionine (SAM), which drop when cells are limited for methionine (36). Since the *metK* (SAM synthetase) gene is in the S box regulon in a number of species, the *metK* gene would be subject to autoregulation if SAM is the effector.

4.3. Distribution of the S box system

Genes with leader regions fitting the pattern of the S box leaders have been found primarily in low G+C Gram-positive species; three genes have also been

identified in the Gram-negative organism *Geobacter sulfurreducens*, in which the T box system has also been found (unpublished results). This system appears to be absent in the high G+C Gram-positive branch, and most methionine genes in *Streptococcus* and *Enterococcus* sp. are instead regulated by the T box system. Nearly all of the genes identified for which functions can be reliably predicted have some role in methionine metabolism (37). The *cysH* operon of *B. subtilis*, which contains an S box leader motif, appears instead to be regulated at the level of transcription initiation in response to cysteine availability (38).

5. PREDICTIONS OF GENE FUNCTION FROM REGULATORY SIGNALS

The rapid accumulation of microbial genomic sequence data has provided a wonderful resource for characterization of the distribution and organization of genes in a variety of species. A major limitation, however, is in the ability to identify the probable physiological role of genes of unknown function. The identification of sequence elements diagnostic of regulation in response to a particular physiological signal provides a new tool for making predictions about the probable function of unknown genes.

5.1. Identification of putative T box and S box genes

Regulation of a gene by either the T box or S box system can be predicted on the basis of identification of conserved primary sequence and secondary structural elements in regions upstream of coding sequences. These elements can be recognized by searching genomic data of uncharacterized organisms for the most highly conserved elements, followed by inspection of the surrounding regions for the remaining conserved features, and analysis of the downstream coding regions for similarity to genes of known function in other organisms. An alternative approach involves identification of genes predicted to be involved in a particular function or pathway, and analysis of the upstream DNA for the patterns characteristic of T box or S box leaders. The first approach is biased toward genes with the highest level of similarity to the conserved pattern, but allows identification of new genes and gene families. The second approach is biased toward genes that have been previously identified, but allows greater divergence in conserved sequence elements. Both approaches have been applied to the identification of new T box and S box genes in a variety of genomes, and have yielded large data sets for both sets of leaders (unpublished results). No difference in the level of conservation of the diagnostic sequence elements has been observed in comparison of leaders obtained from the two search schemes, indicating that the patterns of sequence conservation observed are independent of the search criteria and reflect true conservation.

5.2. T box leaders provide insight into amino acid specificity

For the T box system, the identification of the specifier sequence within the Stem I region provides a clear indication of the probable regulatory pattern of the gene. For example, the presence of a UAC tyrosine codon at this position in the *B. subtilis tyrS* leader is the key determinant for regulation of the expression of this gene in response to charging of $tRNA^{Tyr}$ (7). The high conservation of AARS genes generally allows prediction of gene function with high confidence; the identity of the specifier sequence in this case provides confirmation of the amino acid assignment. A number of putative amino acid biosynthesis genes have also been found that are preceded by T box leaders (unpublished results); most of these genes are relatively easy to assign, with the specifier sequence providing an extra level of confidence. The third major class of genes with T box leaders exhibit similarity to various classes of transporter proteins. In this case, the amino acid specificity predicted by the specifier sequence provides key information about the probable physiological conditions to which expression of the gene responds, and allows a clear prediction of the type of compound likely to serve as the substrate of the regulated transporter. For example, the *yvbW* gene of *B. subtilis* exhibits high similarity to amino acid transporters of a variety of specificities; the presence of a CUC leucine specifier sequence predicted that expression would be induced by growth under limiting leucine conditions, and this was confirmed by testing expression of a *yvbW-lacZ* transcriptional fusion (S. Rollins and T. H., unpublished results). This type of analysis therefore provides a new approach to predicting the function of putative transporter genes, based on the predicted (and easily testable) patterns of gene regulation.

An interesting example of using the T box system to investigate gene function is provided by the *B. subtilis yczA* gene, which contains a T box leader with a tryptophan specifier sequence and for which regulation in response to accumulation of uncharged $tRNA^{Trp}$ has been demonstrated (19). The product of this gene, now designated AT, was demonstrated to play a central role in regulation of the activity of the TRAP protein (34), which in turn represses the expression of tryptophan biosynthesis genes at the levels of transcription termination and translation initiation (4). The identification of AT, and its response to $tRNA^{Trp}$ charging, provides an explanation for the observation that a mutation in tryptophanyl-tRNA synthetase results in derepression of tryptophan biosynthesis genes (39), a result not previously accounted for in the TRAP regulatory system. The presence of the tryptophanyl T box leader upstream of *yczA* was the key clue to the relationship between this gene and the tryptophan pathway, and to the identification of this protein as a regulator of TRAP function.

5.3. S box leaders signal a role in methionine metabolism

The majority of genes originally identified in the *B. subtilis* genome with leaders containing the set of conserved features used to define the S box family were annotated as genes of unknown function (35). Since a few of the genes in the group were known to be involved in methionine metabolism (e.g., *metC, metK*), we tested for regulation in response to methionine limitation and showed that expression of several of these genes, including genes of

unknown function, was controlled by methionine availability (35, 40; unpublished results). We therefore attempted to predict likely functions for some of these genes in the known pathways of methionine biosynthesis (37, 40). We also demonstrated that insertional inactivation of the *yitJ* gene resulted in methionine auxotrophy, consistent with the prediction that it encodes a novel form of methylene tetrahydrafolate reductase (37). Similarly, we predicted that the *yjcIJ* genes were likely to encode cystathionine gamma-synthase and beta-lyase, respectively; this has now been confirmed experimentally (41). We also predicted that the *ykrTS* and *ykrWXYZ* operons were likely to be involved in recycling of methylthioadenosine, a byproduct of polyamine biosynthesis, to methionine, and this has now been confirmed (40, 42). These analyses demonstrate the predictive power of identification of key conserved elements of regulatory systems.

6. PERSPECTIVES

The T box and S box transcription termination control systems represent two different mechanisms for control of expression of amino acid-related genes. Both systems are broadly used in Gram-positive bacteria, although the two systems are differentially distributed within this group. The specific utilization of the cognate tRNA as the effector allows the T box mechanism to be used for genes of different amino acid classes within the same organism. In contrast, the S box system is dedicated to genes involved in methionine metabolism. It is not clear why a methionine-specific system has been developed, since in some organisms the T box system is also used for methionine genes, and the S box system is absent. The two systems represent different ways in which to control gene expression at the level of premature termination of transcription, providing alternative mechanisms to solving the same regulatory problem. The rarity of these systems in Gram-negative organisms may reflect evolutionary history, or may be due to differences in transcription elongation at the molecular level that make different systems better adapted to different types of regulatory controls.

7. ACKNOWLEDGMENT

Work in our laboratory is supported by National Institutes of Health grants R01-GM47823 and R01-GM63615.

8. REFERENCES

1. Henkin T M: Transcription termination control in bacteria. *Curr Opin Microbiol* 3, 149-153 (2000)

2. Henkin T M: Transcripton termination control in bacteria. In: Signals, switches, regulons and cascades: control of bacterial gene expression. Eds: Hodgson D A, Thomas C M, Cambridge University Press, United Kingdom, p. 169-181 (2002)

3. Yanofsky C: Transcription attenuation: once viewed as a novel regulatory strategy. *J Bacteriol* 182, 1-8 (2000)

4. Babitzke P, P Gollnick: Posttranscription initiation control of tryptophan metabolism in *Bacillus subtilis* by the *trp* RNA-binding attenuation protein (TRAP), anti-TRAP, and RNA structure. *J Bacteriol* 183, 5795-5802 (2001)

5. Henkin T M, B L Glass, F J Grundy: Analysis of the *Bacillus subtilis tyrS* gene: conservation of a regulatory sequence in multiple tRNA synthetase genes. *J Bacteriol* 174, 1299-1306 (1992)

6. Putzer H, N Gendron, M Grunberg-Manago: Coordinate expression of the two threonyl-tRNA synthetase genes in *Bacillus subtilis*: control by transcriptional antitermination involving a conserved regulatory sequence. *EMBO J* 11, 3117-3127 (1992)

7. Grundy F J, T M Henkin: tRNA as a positive regulator of transcription antitermination in *B. subtilis*. *Cell* 74, 475-482 (1993)

8. Rollins S M, F J Grundy, T M Henkin: Analysis of *cis*-acting sequence and structural elements required for antitermination of the *Bacillus subtilis tyrS* gene. *Mol Microbiol* 25, 411-421 (1997)

9. Putzer H, S Laalami, A A Brakhage, C Condon, M Grunberg-Manago: Aminoacyl-tRNA synthetase gene regulation in *Bacillus subtilis*: induction, repression, and growth-rate regulation. *Mol Microbiol* 16, 709-718 (1995)

10. Marta P T, R D Ladner, J A Grandoni: A CUC triplet confers leucine-dependent regulation of the *Bacillus subtilis ilv-leu* operon. *J Bacteriol* 178, 2150-2153 (1996)

11. Grundy F J, S E Hodil, S M Rollins, T M Henkin: Specificity of tRNA-mRNA interactions in *Bacillus subtilis tyrS* antitermination. *J Bacteriol* 179, 2587-2594 (1997)

12. Luo D, J Leautey, M Grunberg-Manago, H. Putzer: Structure and regulation of expression of the *Bacillus subtilis* valyl-tRNA synthetase gene. *J Bacteriol* 179, 2472-2478 (1997)

13. van de Guchte M, S D Ehrlich, A Chopin: $tRNA^{Trp}$ as a key element of antitermination in the *Lactococcus lactis trp* operon. *Mol Microbiol* 29, 61-74 (1998)

14. Landick R, C L Turnbough Jr., C Yanofsky: Transcription attenuation. In *Escherichia coli* and *Salmonella:* cellular and molecular biology. Ed: Neidhardt F C, ASM Press, Washington DC, p. 1263-1286 (1996)

15. Grandoni J A, S B Fulmer, S B Brizzio, S A Zahler, J M Calvo: Regions of the *Bacillus subtilis ilv-leu* operon involved in regulation by leucine. *J Bacteriol* 175, 7581-7593 (1993)

16. Garrity D B, S A Zahler: Mutations in the gene for a tRNA that functions as a regulator of a transcriptional attenuator in *Bacillus subtilis*. *Genetics* 137, 627-636 (1994)

17. Van der Horn B P, S A Zahler: Cloning and nucleotide sequence of the leucyl-tRNA synthetase gene of *Bacillus subtilis*. *J Bacteriol* 174:3928-3935 (1992)

18. Grundy F J, S M Rollins, T M Henkin: Interaction between the acceptor end of tRNA and the T box stimulates antitermination in the *Bacillus subtilis tyrS* gene: a new role for the discriminator base. *J Bacteriol* 176, 4518-4526 (1994)

19. Sarsero J P, E Merino, C Yanofsky: A *Bacillus subtilis* operon containing genes of unknown function senses $tRNA^{Trp}$ charging and regulates expression of the genes of tryptophan biosynthesis. *Proc Natl Acad Sci USA* 97, 2656-2661 (2000)

20. Gagnon Y, R Breton, H Putzer, M Pelchat, M Grunberg-Manago, J Lapointe: Clustering and co-transcription of the *Bacillus subtilis* genes encoding the aminoacyl-tRNA synthetases specific for glutamate and for cysteine and the first enzyme for cysteine biosynthesis. *J Biol Chem* 269, 7473-7482 (1994)

21. Chopin A, V Biaudet, S D Ehrlich: Analysis of the *Bacillus subtilis* genome reveals nine new T-box leaders. *Mol Microbiol* 29, 662-664 (1998)

22. Luo D, C Condon, M Grunberg-Manago, H Putzer: *In vitro* and *in vivo* structure probing of the *thrS* leader in *Bacillus subtilis*. *Nucl Acids Res* 26, 5379-5387 (1998)

23. Winkler W C, F J Grundy, B A Murphy, T M Henkin: The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA* 7, 1165-1172 (2001)

24. Delorme C, S D Ehrlich, P Renault: Regulation of expression of the *Lactococcus lactis* histidine operon. *J Bacteriol* 181, 2026-2037 (1999)

25. Frenkiel H, J Bardowski, S D Ehrlich, A Chopin: Transcription of the *trp* operon in *Lactococcus lactis* is controlled by antitermination in the leader region. *Microbiology* 144, 2103-2111 (1998)

26. Grundy F J, T R Moir, M T Haldeman, T M Henkin: Sequence requirements for terminators and antiterminators in the T box transcription antitermination system: disparity between conservation and functional requirements. *Nucl Acids Res* 30, 1646-1655 (2002)

27. Gerdeman M S, T M Henkin, J V Hines: *In vitro* structure-function studies of the *Bacillus subtilis tyrS* antiterminator: evidence for factor-independent tRNA acceptor stem binding specificity. *Nucl Acids Res* 30, 1065-1072 (2002)

28. Grundy F J, J A Collins, S M Rollins, T M Henkin: tRNA determinants for transcription antitermination of the *Bacillus subtilis tyrS* gene. *RNA* 6, 1131-1141 (2000)

29. van de Guchte M, S D Ehrlich, A Chopin: Identity elements in tRNA-mediated transcription antitermination: implication of tRNA D- and T-arms in mRNA recognition. *Microbiology* 147, 1223-1233 (2001)

30. Condon C, H Putzer, M Grunberg-Manago: Processing of the leader mRNA plays a major role in the induction of *thrS* expression following threonine starvation in *Bacillus subtilis*. *Proc Natl Acad Sci USA* 93, 6992-6997 (1996)

31. Pelchat M, J Lapointe: *In vivo* and *in vitro* processing of the *Bacillus subtilis* transcript coding for glutamyl-tRNA synthetase, serine acetyltransferase, and cysteinyl-tRNA synthetase. *RNA* 5, 281-289 (1999)

32. Condon C, H Putzer, D Luo, M Grunberg-Manago: Processing of the *Bacillus subtilis thrS* leader is RNase E-dependent in *Escherichia coli*. *J Mol Biol* 268, 235-242 (1997)

33. Grundy F J, M T Haldeman, G M Hornblow, J M Ward, A F Chalker, T M Henkin: The *Staphylococcus aureus ileS* gene, encoding isoleucyl-tRNA synthetase, is a member of the T-box family. *J Bacteriol* 179, 3767-3772 (1997)

34. Valbuzzi A, C Yanofsky: Inhibition of the *B. subtilis* regulatory protein TRAP by the TRAP-inhibitory protein, AT. *Science* 293, 2057-2059 (2001)

35. Grundy F J, T M Henkin: The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol Microbiol* 30, 737-749 (1998)

36. Greene R C: Biosynthesis of methionine. In *Escherichia coli* and *Salmonella*: cellular and molecular biology. Ed: Neidhardt F C, ASM Press, Washington DC, p. 542-560 (1996)

37. Grundy F J, T M Henkin: Synthesis of serine, glycine, cysteine, and methionine. In *Bacillus subtilis* and its closest relatives: from genes to cells. Ed: Sonenshein A L, Hoch J A, Losick R, ASM Press, Washington DC, p. 245-254 (2002)

38. Mansilla M C, D Albanesi, D de Mendoza: Transcriptional control of the sulfur-regulated *cysH* operon encoding genes involved in L-cysteine biosynthesis in *Bacillus subtilis. J Bacteriol* 182, 5885-5892 (2000)

39. Steinberg W: Temperature-induced derepression of tryptophan biosynthesis in a tryptophanyl-transfer ribonucleic acid synthetase mutant of *Bacillus subtilis. J Bacteriol* 117, 1023-1034 (1974)

40. Murphy B A, F J Grundy, T M Henkin: Prediction of gene function in methylthioadenosine recycling from regulatory signals. *J Bacteriol* 184, 2314-2318 (2002)

41. Auger S, W H Yuen, A Danchin, I Martin-Verstraete: The *metIC* operon involved in methionine biosynthesis in *Bacillus subtilis* is controlled by transcription antitermination. *Microbiology* 148, 507-518 (2002)

42. Sekowska A, L Mulard, S Krogh, J K Tse, A Danchin: MtnK, methylthioribose kinase, is a starvation-induced protein in *Bacillus subtilis. BMC Microbiol* 1, 15 (2001)

Key Words: Microbiology, Bacteria, Antitermination, Attenuation, *Bacillus*; Gram-positive, Regulation, tRNA, Aminoacyl-tRNA Synthetase, Amino Acid Biosynthesis, Review

Send correspondence to: Dr Tina M. Henkin, Department of Microbiology, The Ohio State University, 484 W. 12th Avenue, Columbus, OH 43210 USA, Tel: 614-688-3831, Fax: 614-292-8120, E-mail: henkin.3@osu.edu