Genetic Control by Metabolite-Binding Riboswitches

Wade C. Winkler[a] and Ronald R. Breaker*[a]

1. Introduction

Modern organisms must coordinate the expression of many hundreds of genes in response to metabolic demands and environmental changes. Each gene product must be regulated temporally, quantitatively, and often spatially. Additionally, genetic control processes must be dynamic, rapid, and selectively responsive to the specific conditions undergoing change. Therefore, organisms require sentries of genetic regulatory factors that continuously quantify a multitude of signals. Upon measurement of a particular signal, which may be one of many possible biochemical or physical cues, these regulatory factors must modulate expression of a specific subset of the organism's genes.

It has generally been assumed that proteins are the obligate sensors of these signals because protein is a proven medium for forming highly responsive sensors. However, recent findings demonstrate that mRNAs also are capable of acting as direct sensors of chemical[1–7] and physical conditions[8] for the purpose of genetic control. Classes of mRNA domains, collectively referred to as 'riboswitches', serve as RNA genetic control elements that sense the concentrations of specific metabolites by directly binding the target compound. Known riboswitches are responsible for sensing metabolites that are critical for fundamental biochemical processes including coenzyme B_{12},[1] thiamine pyrophosphate (TPP),[2] flavin mononucleotide (FMN),[3, 5] *S*-adenosylmethionine (SAM),[6, 7, 9] lysine,[10] guanine,[4] and adenine.[4, 11] Upon interaction with the appropriate small-molecule ligand, riboswitch mRNAs undergo a structural reorganization that results in the modulation of genes that they encode. To date, all riboswitches that have been examined in detail cause genetic repression upon binding their target ligand, although riboswitches that activate gene expression upon ligand binding are certainly possible.

In each instance, riboswitch domains have been subjected to a battery of biochemical and genetic analyses in order to convincingly demonstrate that direct interaction of small organic metabolites with mRNA receptors leads to a corresponding alteration in genetic expression. This review provides a brief summary of these efforts and of some of the general characteristics that are exhibited by riboswitches. Although these findings represent the initial steps in elucidating the principles that underlie RNA-based detection of small molecules, the prospects for riboswitches as genetic tools and as possible targets for development of antimicrobials are already beginning to emerge.

2. General Organization of Riboswitch RNAs

Bacterial riboswitch RNAs are genetic control elements that are located primarily within the 5′-untranslated region (5′-UTR) of the main coding region of a particular mRNA (Figure 1). Structural-probing studies (discussed further below) reveal that

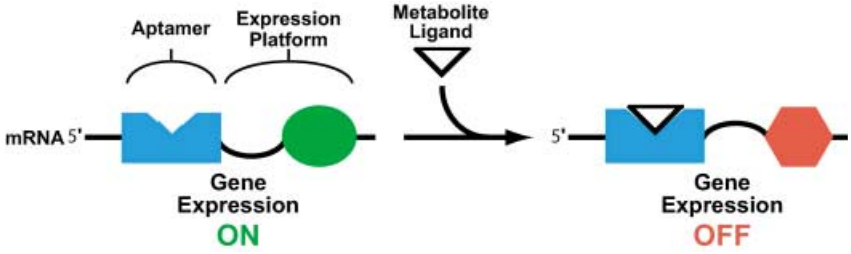


Figure 1. *General organization of riboswitch RNAs in bacterial mRNAs. Binding of the target ligand to the aptamer domain stabilizes an altered conformation of the expression platform that results in a change in gene expression for downstream gene(s).*

riboswitch elements are generally composed of two domains: a natural aptamer[12, 13] that serves as the ligand-binding domain, and an 'expression platform' that interfaces with RNA elements that are involved in gene expression (for example, Shine–Dalgarno (SD) elements, transcription terminator stems). These conclusions are drawn from the observation that aptamer domains synthesized in vitro bind the appropriate ligand in the absence of the expression platform.[2–4, 7] Moreover, structural-probing investigations suggest that, when it is examined independently, the aptamer domain of most riboswitches adopts a particular secondary- and tertiary-structure fold that is essentially identical to the aptamer structure when it is examined in the context of the entire 5′-leader RNA. This implies that, in many cases, the aptamer domain is a modular unit that folds independently of the expression platform.

Ultimately, the ligand-bound or -unbound status of the aptamer domain is interpreted through the expression platform, which is responsible for exerting an influence upon gene expression. The view of riboswitch domains as modular elements is further supported by the fact that aptamer domains are highly conserved amongst various organisms (and even between

[a] *Prof. R. R. Breaker, Dr. W. C. Winkler*
Department of Molecular, Cellular, and Developmental Biology
Yale University
Kline Biology Tower 506
266 Whitney Avenue
New Haven, CT 06520 (USA)
Fax: (+1) 203-432-6604
E-mail: ronald.breaker@yale.edu

  DOI: 10.1002/cbic.200300685

kingdoms, as is observed for the TPP riboswitch),[14] whereas the expression platform varies in sequence, structure, and in the mechanism by which expression of the appended open reading frame is controlled. For example, ligand binding to the TPP riboswitch of the *tenA* mRNA of *Bacillus subtilis* causes transcription termination.[5] This expression platform is distinct in sequence and structure compared to the expression platform of the TPP riboswitch in the *thiM* mRNA from *Escherichia coli*, wherein TPP binding causes inhibition of translation by an SD blocking mechanism.[2] The TPP aptamer domain is easily recognizable and of near identical functional character between these two transcriptional units, but the genetic control mechanisms and the expression platforms are very different.

Aptamer domains for riboswitch RNAs typically range from $\approx$70 – 170 nucleotides in length (Scheme 1). These large sizes were somewhat unexpected given that in vitro evolution experiments identified a wide variety of small-molecule-binding aptamers, which are considerably shorter in length and structural intricacy (Scheme 1).[12, 13, 15] Although the reasons for the substantial increase in complexity and information content of the natural aptamer sequences relative to artificial aptamers remain to be proven, this complexity is most likely required to form RNA receptors that function with high affinity and selectivity. Apparent dissociation constants (K_D) for the ligand – riboswitch complexes range from low nanomolar to low micromolar values. It is also worth noting that some aptamer domains, when isolated from the appended expression platform, exhibit improved affinity for the target ligand over that of the intact riboswitch (typically $\approx$10- to 100-fold improvement).[2, 7] Presumably, there is an energetic cost in sampling the multiple distinct RNA conformations required by a fully intact riboswitch RNA, which is reflected by a loss in ligand affinity. Since the aptamer domain must serve as a molecular switch, this might also add to the functional demands on natural aptamers and help explain their more sophisticated structures.

3. Riboswitch Regulation of Transcription Termination in Bacteria

Bacteria primarily make use of two methods for termination of transcription. Certain genes incorporate a termination signal that is dependent upon the Rho protein,[16] while others make use of Rho-independent terminators (intrinsic terminators) to destabilize the transcription elongation complex.[17–19] The latter RNA elements are composed of a GC-rich stem-loop followed by a stretch of 6 – 9 uridyl residues. Intrinsic terminators are widespread throughout bacterial genomes[20] and are typically located at the 3′-termini of genes or operons. Interestingly, an increasing number of examples of intrinsic terminators are being identified within 5′-UTRs.[21]

Amongst the wide variety of genetic regulatory strategies employed by bacteria there is a growing class of examples wherein RNA polymerase responds to a termination signal within the 5′-UTR in a regulated fashion.[21, 22] Under certain conditions the RNA-polymerase complex is directed by external signals either to perceive or to ignore the termination signal. Presumably, one of at least two mutually exclusive mRNA conformations results in the formation or disruption of the RNA structure that signals transcription termination. A trans-acting (intermolecularly acting) factor, which in some instances is an RNA[23, 24] and in others is a protein,[25] is generally required for receiving a particular intracellular signal and subsequently stabilizing one of the RNA conformations. In contrast, riboswitches offer a direct link between RNA structure modulation and the metabolite signals that need to be interpreted by the genetic control machinery. A brief overview of the FMN riboswitch from a *B. subtilis* mRNA is provided below to illustrate this mechanism.

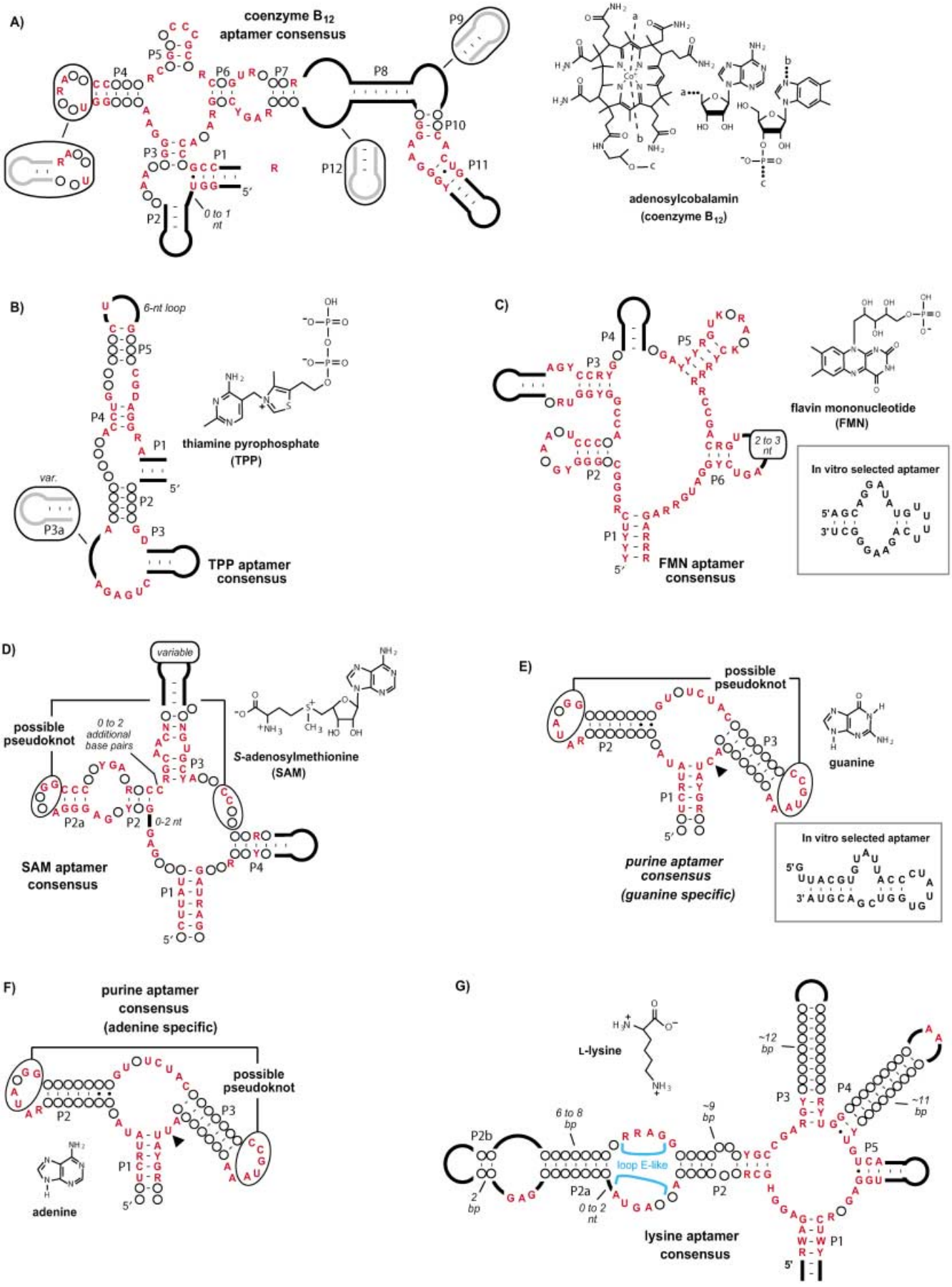
3.1. A natural aptamer for FMN

A highly conserved RNA domain, referred to as the *RFN* element, was identified in bacterial genes involved in the biosynthesis and transport of riboflavin and FMN.[26, 27] This element is required for genetic manipulation of the *ribDEAHT* operon (hereafter, *ribD*) of *B. subtilis*, as mutations resulted in a loss of FMN-mediated regulation.[28, 29] These data led to the proposal that either a protein-based FMN sensor[30] or FMN itself[26, 27] interacts with the *RFN* element in order to repress *ribD* gene expression. Although RNA sequences that specifically bind FMN have been identified through directed-evolution experimentation,[31–33] they exhibit no obvious resemblances to the *RFN* element (Scheme 1), a fact suggesting that FMN recognition by the *RFN* element would be achieved by a mechanism distinctive to that of the engineered aptamers.

3.1.1. Structural probing reveals FMN-mediated RNA folding modulation

Each internucleotide linkage in an RNA polymer is susceptible to spontaneous hydrolysis by an S_N2-like mechanism, wherein the 2′-oxygen atom attacks the adjacent phosphorus center, thereby leading to chain cleavage. For optimal speed, this reaction requires a 180° orientation between the attacking nucleophile, the phosphorus center, and the 5′-oxygen leaving group (in-line conformation).[34, 35] Nucleotides that are base paired, or otherwise structurally constrained, are typically incapable of adopting this configuration and therefore display low rates of spontaneous cleavage. In contrast, nucleotides that are structurally unrestrained exhibit much higher rates of spontaneous cleavage. These observations have been exploited in a structural-probing method, referred to as 'in-line probing', which establishes the relative rates of spontaneous cleavage for a given RNA polymer and correlates this with secondary- and tertiary-structure models.[34]

To assess whether the RFN element of *ribD* was responsive to FMN, a fragment of the corresponding 5′-UTR was 5′-^{32}P labeled and incubated in the absence and presence of FMN, and the resulting fragments were analyzed by polyacrylamide gel electrophoresis (PAGE). Interestingly, patterns differ between reactions that are incubated either with or without FMN, a fact signifying that there is a structural rearrangement of the RNA upon FMN binding to *ribD*.[3] The spontaneous cleavages of certain nucleotide positions located within interhelical regions of



◀ ***Scheme 1.*** *The known riboswitches. Consensus sequences and secondary structure models were derived by phylogenetic and biochemical analyses. Nucleotides in red are conserved in greater than 90% of the representative sequences, open circles identify nucleotide positions of variable sequence, and lines identify elements that are variable in sequence and length. Aptamer models were derived from literature citations as follows: A) coenzyme* B_{12}*,*[1, 40] *B) TPP,*[2, 14, 41] *C) FMN,*[3, 27] *D) SAM,*[7, 37] *E) guanine,*[39] *F) adenine,*[39, 48] *and G) lysine.*[10] *. Flavin- and guanine-binding aptamers isolated through in vitro selection experimentation*[31, 52] *are shown as insets in order to illustrate the structural complexity of natural aptamers versus their engineered counterparts. Letters R and Y represent purine and pyrimidine bases, respectively; K designates G or U; W designates A or U; H designates A, C, or U; D designates G, A, or U; N represents any of the four bases.*

the *RFN* element become significantly reduced in the presence of FMN; this suggests that these nucleotides are involved in forming an FMN–RNA complex, which forces structural constraints upon the RNA (Scheme 2). It is this type of structural modulation that can be harnessed by the expression platform for allosteric modulation of gene expression.

Additional evidence for direct binding of FMN by the *ribD RFN* element was generated by enzymatic probing. Oligonucleotides predicted to anneal with the *RFN* element were added to *ribD* transcripts in the presence and absence of FMN, and the resulting mixtures was digested with RNase H (which specifically cleaves RNA:DNA heteroduplexes) and analyzed by PAGE.[5] A significant portion of transcripts bind certain oligonucleotides in the absence of FMN, but not in the presence of FMN, a fact indicating that FMN stabilizes a structural rearrangement of *ribD* transcripts that in turn prevents annealing of the oligonucleotide.

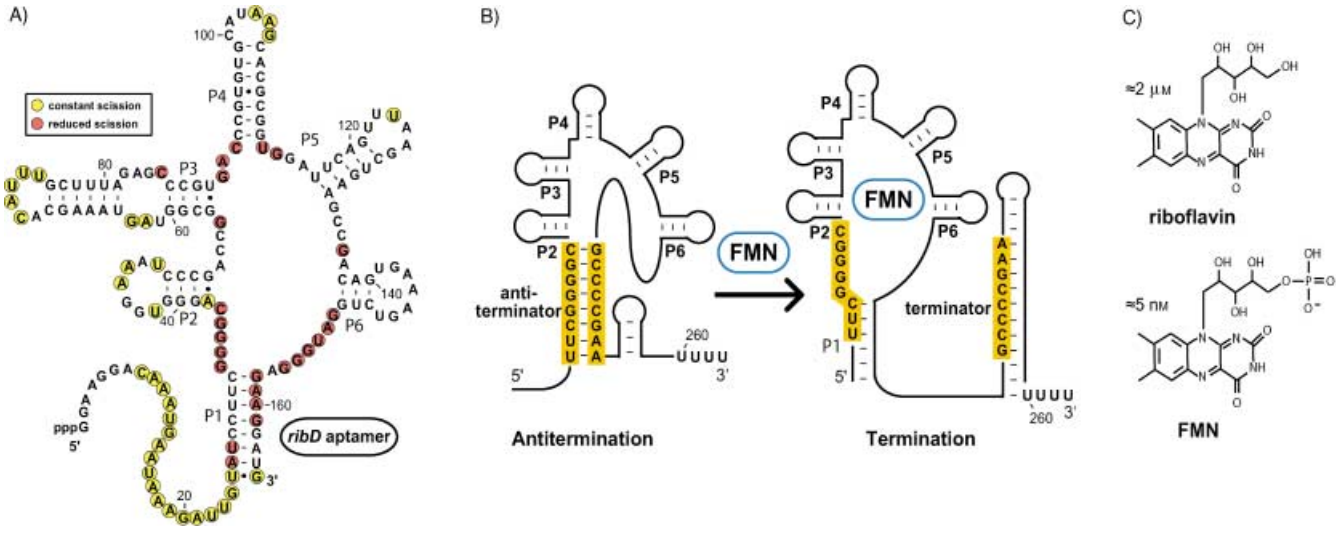
3.1.2. *Affinity and specificity of the FMN–ribD* complex

If the RFN element serves as an aptamer for FMN, it should exhibit characteristics of a saturable receptor that has some ability to discriminate against related ligands. To obtain apparent K_D values for FMN, in-line probing assays were repeated with trace amounts of *ribD* RNA and increasing concentrations of FMN.[3] The ligand concentration that correlates with half-maximal modulation of RNA structure should reflect the apparent K_D value. These experiments indicate that the *ribD* RNA contains a saturable ligand binding site that exhibits an apparent K_D value of ≈ 5 nM. Furthermore, the RNA discriminates against riboflavin (the dephosphorylated form of FMN) by approximately three orders of magnitude. This exceptional ligand specificity of the *ribD* mRNA is surprising since the aptamer must generate a binding pocket for FMN that makes productive interactions with a phosphate group. An interesting aspect of future structure analyses will be to determine how RNA, a highly anionic polymer, is capable of preferentially recognizing the negatively charged phosphate group of FMN.

3.2. FMN-induced transcription termination

3.2.1. *In vitro transcription termination mediated by an FMN riboswitch*

The relative amounts of the major transcription products for the *ribD* leader region were examined by in vitro transcription with T7 RNA polymerase[3] or *Bacillus subtilis* RNA polymerase.[5] The *ribD* leader region contains a typical intrinsic terminator just upstream of the *ribD* coding region. Interestingly, transcripts that terminated at the intrinsic terminator are specifically induced by FMN, in the absence of additional protein factors. Furthermore, mutations in the *RFN* element abrogate this phenomenon.[5, 36] The left half of the terminator sequence is proposed to form alternative base-pairing interactions with a portion of the *RFN* element, thereby forming an antiterminator element. Sequence



Scheme 2. *Regulation of the* B. subtilis ribD *mRNA by FMN. A) In-line probing reveals ligand-dependent structure modulation with FMN: internucleotide linkages identified with red circles exhibit decreased amounts of spontaneous cleavage when* ribD *is incubated in the presence of FMN (indicating an increase in structural stabilization for these nucleotides) relative to incubation in the absence of FMN. Yellow circles identify linkages that exhibit consistently high levels of scission, which indicates that they are not modulated by presence of FMN. B) Model for the mechanism of* ribD *regulation. The* ribD *mRNA adopts an antitermination conformation in the absence of FMN. Increased levels of FMN stabilize an* RFN–*FMN complex that permits formation of the terminator structure. C) Chemical structure and apparent dissociation constants for riboflavin and FMN.*

alterations of the intrinsic terminator eliminate FMN-induced termination while alterations in the antiterminator result in constitutive termination.[5, 36] Taken together, these observations are consistent with a mechanistic model wherein FMN directly interacts with *ribD* transcripts during conditions of excess FMN. Complex formation subsequently induces transcription termination within the 5′-UTR (Scheme 2), which precludes gene expression by preventing the downstream coding regions from being transcribed. During conditions of limited FMN, an antiterminator structure is formed within the *ribD* nascent transcript, which allows for synthesis of the complete mRNA.

3.2.2. FMN-mediated control of transcription termination in vivo

The molecular details of riboswitch-mediated transcription termination are likely to be more complex than this rather simplistic model implies. For example, given that the 'decision' to form the terminator or antiterminator conformation occurs only once during transcription, the regulatory mechanism is likely to rely on precise transcriptional kinetics as well as the appropriate RNA folding pathways. Moreover, the kinetics of FMN interacting with the RNA receptor are presumed to be a critical factor. Although the affinity that the RNA has for FMN is exceptionally strong compared to engineered aptamers, it is possible that the kinetics of ligand association might be the more important determinant of genetic regulation. Indeed, all of these parameters are likely to conspire together in order to exert appropriate control over the intrinsic terminator. If true, then the kinetic parameters determined for FMN binding and for FMN-induced transcription termination in vitro might be an imperfect reflection of the function of riboswitches in vivo. Specifically, the impact of transcription speed and of the various cellular conditions might lead to significant differences in the performance characteristics of riboswitches.

3.3. Control of transcription termination by other riboswitches

Intrinsic terminators can be identified by computer-assisted search algorithms.[20] By using such bioinformatic analyses, it is possible to identify a subset of riboswitch RNAs that are predicted to contain an intrinsic terminator and an alternate antiterminator structural element.[27, 37–41] Therefore, the results described above for the FMN riboswitch may be indicative of the mechanisms used by many other riboswitch RNAs. Indeed, SAM- and TPP-dependent riboswitches have been demonstrated to exert control over termination by formation of mutually exclusive intrinsic terminator and antiterminator structures.[5–7] Furthermore, mutations that disrupt and subsequently restore helices within the SAM riboswitch aptamer result in loss and restoration, respectively, of SAM binding. Concurrently, these mutations also result in disruption or restoration of SAM-induced transcription termination in accordance with ligand-binding function.[7] It is also possible, and perhaps even likely, that some riboswitches will exert control over transcription termination signals that differ appreciably from classical intrinsic terminators.

4. Riboswitch Regulation of Translation Initiation in Bacteria

An alternative mechanism of genetic control by riboswitches is the modulation of translation initiation. Unlike transcription termination, the entire mRNA could be synthesized by RNA polymerase, but expression would be prevented by the riboswitch until the metabolite concentration reached a certain level. In most instances, we observed that riboswitches prevent translation initiation in the presence of high concentrations of target metabolite. However, it is certainly possible that allosteric modulation of riboswitch structures could lead to translation activation. The regulatory mechanism of translation control is briefly described below for a TPP riboswitch from *E. coli*.

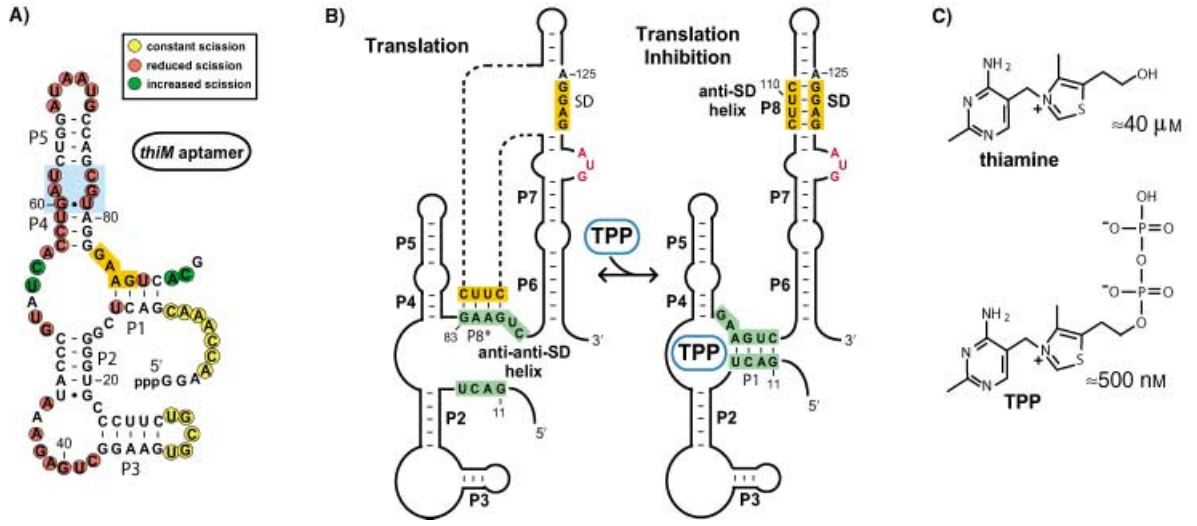
4.1. A natural aptamer for TPP

A conserved RNA element, referred to as the *thi* box, was identified within 5′-UTRs of mRNAs that are responsible for thiamine biosynthesis and transport.[41, 42] Genetic experiments confirmed that this structural element is required for thiamine-dependent regulation of *Rhizobium meliloti* thiamine biosynthesis genes,[42] yet no regulatory factor had been identified through classical genetic experimentation. Therefore, it was possible that the *thi* box might serve as a portion of a riboswitch that responds to thiamine or its derivatives.

In *E. coli*, thiamine biosynthesis and transport genes are primarily located within three operons and four single genes,[43] wherein each operon is preceded by a *thi* element. To begin to assess the regulatory properties of these sequences, the leader regions for the *thiMD* and *thiCEFSGH* operons were utilized to construct transcriptional and translational fusions to a *lacZ* reporter gene.[2] Addition of exogenous thiamine results in repression of the *lacZ* reporter gene in *E. coli*. Results from these data demonstrate that the *thiM* gene is regulated primarily at the level of translation while the *thiC* leader region confers both transcriptional and translational regulation to the *lacZ* reporter.

4.1.1. Direct binding of thiamine pyrophosphate by E. coli *mRNAs*

As described above for the FMN aptamer, direct binding of TPP to the *thiM* and *thiC* leaders was demonstrated by in-line probing assays.[2] The addition of thiamine, thiamine monophosphate (TP), or TPP leads to structural rearrangement of the *thiM* RNA, particularly in the region encompassing the *thi* element (Scheme 3). Significantly, TPP, which is typically the bioactive form of thiamine, exhibits the best affinity of the ligands, with an apparent K_D value of 500 nM, while TP and thiamine exhibit apparent K_D values of 3 and 40 μM, respectively. In-line probing assays of RNAs resembling the *thiC* leader region reveal even more dramatic discrimination between thiamine and its phosphorylated forms, with a difference of more than 1,000-fold exhibited between binding of thiamine and TPP. These data are consistent with genetic experiments that suggested that TPP synthesis is required for regulation.[44, 45] Also, this system



Scheme 3. *Regulation of the* E. coli thiM *mRNA by TPP. A) In-line probing reveals ligand-dependent structure modulation with TPP: internucleotide linkages identified with red circles exhibit decreased amounts of spontaneous cleavage when* thiM *is incubated in the presence of TPP compared to incubation in the absence of ligand. In contrast, linkages identified with green circles exhibit increased amounts of cleavage when* thiM *is incubated with TPP compared to incubation in the absence of ligand. The blue-shaded box indicates the pyrophosphate-recognition region (as described in the text). B) Model for the mechanism of* thiM *regulation. In the absence of TPP, the anti-SD sequence interacts with part of the aptamer domain to form anti-anti-SD. As TPP is increased, aptamer – TPP complexes are formed and the anti-SD favors pairing with the SD. C) Chemical structure and apparent dissociation constants for thiamine and TPP.*

provides another example of a natural RNA aptamer that makes productive contacts to phosphate groups.

4.1.2. Confirmation of TPP binding by equilibrium dialysis

RNAs resembling the *thiM* leader region were synthesized and placed into one side of a two-chamber equilibrium dialysis apparatus, in which the compartments are separated by a dialysis membrane with a molecular-weight cut off of 5000 daltons. ^{3}H-thiamine was preferentially retained within the *thiM*-containing chamber when allowed to equilibrate between chambers.[2] This effect could be eliminated by providing excess unlabeled thiamine, but could not be reversed when supplemented with oxythiamine, a close chemical analogue of thiamine. Additionally, a mutated version of *thiM* was unable to shift ^{3}H-thiamine into the RNA-containing chamber. Together, these data are indicative of the formation of stable *thiM* – thiamine complexes, wherein the sequence of the RNA and the chemical form of the ligand are critical for maximal binding affinity.

4.2. Binding of thiamine derivatives correlates with structural modulation

Close inspection of in-line probing data for *thiM* reveals two surprising patterns of structural modulation. First, the relative rates of spontaneous fragmentation between reactions containing either thiamine or TPP differ within an internal loop of the *thi* element (Scheme 3). Nucleotides in this region adopt an increase in structural order in the presence of TPP but not with thiamine, a fact implying that this region is somehow involved in the formation of a pyrophosphate-recognition pocket. Secondly, the region of the SD sequence is the only portion outside of the *thi* element that becomes structurally modulated in the presence of TPP.

Specifically, the SD sequence exhibits a significant decrease in spontaneous cleavage relative to reactions lacking TPP; this suggests that the SD is converted into a more structurally constrained form upon binding of TPP. This idea is consistent with a mechanism (Scheme 3) whereby in the absence of TPP the SD has a significant degree of single-stranded character and is accessible for translation initiation. An anti-SD sequence is proposed to interact with an anti-anti-SD sequence within the TPP aptamer under these conditions. In contrast, during conditions of excess TPP, a TPP – RNA complex is formed that disrupts the base pairing of the anti-SD sequence. The anti-SD sequence is then free to interact directly with the SD, decrease the single-stranded character of the region, and subsequently decrease the efficiency of translation initiation. Preliminary site-directed mutagenesis of the *thiM* mRNA supports this overall model.[2] Specifically, mutations that disrupt TPP binding also disrupt regulation of translation for *thiM* – *lacZ* fusions, while mutations that alter the anti-SD sequence affect regulation but do not affect TPP binding. Thus, binding of thiamine correlates with both the structural accessibility of the SD and the translation efficiency in vivo.

4.3. Control of translation initiation by other riboswitches

Bioinformatics analyses have suggested that molecular mechanisms similar to that of *thiM* also might be recurrent amongst riboswitch RNAs. Specifically, anti-SD and anti-anti-SD structures have been proposed for several riboswitch classes, including FMN,[27] lysine,[10] TPP,[41] coenzyme B_{12},[1, 40] and SAM.[7] In general, riboswitches from Gram-negative organisms seem to favor expression platforms that exert control over translation, while

riboswitches from Gram-positive bacteria appear to predominately use expression platforms that control transcription termination. The latter might reflect a greater reliance upon multigene transcriptional units in Gram-positive organisms, which could be more efficient to preclude transcription of long operons when the gene products are unnecessary.

Biochemical evidence for riboswitch-mediated control over translation initiation has also been obtained for FMN and coenzyme B_{12} riboswitches.[1, 3] FMN binding to a riboswitch that regulates the *B. subtilis ypaA* gene results in alteration of the SD structural context, in a similar manner to what was observed for *thiM*.[3] Interestingly, this genetic control element has also been proposed to regulate *ypaA* transcription,[46] although the leader region does not contain an obvious intrinsic terminator structure. Binding of coenzyme B_{12} to the *E. coli btuB* riboswitch has also been demonstrated to correlate with regulation of translation in vivo. However, this RNA differs from the *thiM* paradigm as the mechanism for control over translation efficiency, which has not yet been elucidated, appears to be derived from something other than the formation of alternate, mutually exclusive stem-loop structures (anti-SD and anti-anti-SD).

Preliminary data also indicate that certain riboswitch RNAs exert control over transcription and translation by using the same RNA sequence.[47] For this class of riboswitches, the SD sequence is contained within an intrinsic terminator. Therefore, the formation of the terminator structure also enacts formation of a SD-sequestering structure. In total, all of these observations suggest that the *thiM* and *ribD* riboswitches represent useful paradigms for riboswitch-mediated control of translation and transcription, respectively. However, there is likely to be a greater variety of molecular mechanisms utilized by riboswitch RNAs for control of gene expression. Indeed, TPP riboswitches that must be employing different mechanisms of control have been identified in several plant and fungal species.[14] The placement of these RNAs near splice sites in some instances and in the 3′-UTR in others is suggestive of TPP-responsive control over splicing and mRNA stability or expression, respectively.

5. Early Origins?

The FMN, TPP, lysine, and coenzyme B_{12} riboswitch RNAs are widespread among evolutionarily distant microorganisms, a fact that implies there is an ancient origin for these RNA genetic elements.[10, 27, 40, 41] SAM, guanine, and adenine riboswitches are also represented in numerous different genera, although they appear to be primarily limited to Gram-positive bacteria, with a few Gram-negative bacteria as exceptions.[7, 39, 48] In all instances, the structural and sequence conservation of riboswitch classes is limited to the aptamer domain (Scheme 1). This is not unexpected given that the aptamer RNA must preserve its capability to bind the target chemical, which has not been significantly modified through evolution. In contrast, there is considerable sequence and structural diversity between expression platforms, even between riboswitches of the same class and within the same organism. Together, these data hint that the ligand-binding properties of riboswitch aptamer domains have been maintained throughout expansive evolutionary timescales.

Furthermore, the ligands for riboswitch RNAs have been proposed to be functional relics from a hypothetical RNA-based world, in which RNA polymers provided all the necessary catalytic and genomic functions of the earliest organisms.[49, 50] Therefore it is tempting to speculate that, as cofactor-binding RNAs, the aptamer domains from riboswitches may have been useful in the context of an RNA-based world for some of the earliest forms of genetic control, for allosteric modulation of ribozymes, or as part of ribozymes that utilized the ligands as essential cofactors.

6. Riboswitches as Drug Targets and Genetic Tools

Riboswitches are utilized for control of numerous genes involved in the biosynthesis and transport of prokaryotic enzymatic cofactors. At least 69 genes, which represents nearly 2% of the total genomic content of *B. subtilis*, are under the control of riboswitch RNAs (Table 1); this exemplifies the extensive use of riboswitch RNAs for genetic control in prokaryotes.[39] Many riboswitch-mediated genes are expected to be essential under most growth conditions. Interference with riboswitch function is then predicted to result in dramatic destabilization of vital metabolic pathways and, perhaps, cessation of growth. Therefore, it seems likely that compounds that closely resemble the target metabolites will bind to riboswitch RNAs and cause a decrease in gene expression. If this analogue-induced disruption of gene expression is sufficient, then such compounds might be candidates for antimicrobial applications.

There is clear precedence for the targeting of RNAs with small-molecule drugs,[51] the most obvious example being that of ribosomal RNA. Several other bacterial-specific RNAs have been explored as candidates for small-molecule drug interaction; however, the approach relies upon screening large chemical libraries for those chemicals that fortuitously interact with the RNA of interest, even though the RNA itself does not naturally form a binding pocket for small organic molecules. Riboswitch RNAs therefore may exhibit an advantage in antimicrobial development given that they serve as a receptor for small-molecule ligands, much like their protein receptor counterparts. The continued exploration of the molecular recognition of target ligands by riboswitch RNAs will ultimately reveal whether 'blind spots', that can be exploited for the development of antimicrobials, exist within the molecular recognition landscape.

In addition to their potential as targets for chemical inhibition, a detailed understanding of the mechanisms utilized by natural riboswitch RNAs may lead to the development of novel genetic control elements. Numerous aptamer RNA sequences have been identified that interact with a wide variety of small organic molecules.[15] It is reasonable to expect that engineered riboswitches could be generated that respond to nonbiological, or otherwise metabolically inert, compounds. The range of uses for such genetic control elements will then be limited only by the range of human imagination and the speed at which engineered aptamers and expression platforms could be built.

Further exploration of riboswitch mechanism, character, and distribution in biological systems will require a combination of

Table 1. *Distribution of known riboswitch classes in B. subtilis.*[a]

Ligand	Transcriptional unit	Predicted gene function(s)
lysine	*lysC*	aspartokinase II
FMN	*ypaA*	putative flavin transporter
	ribD-ribE-ribBA-ribH	riboflavin biosynthesis
coenzyme B_{12}	*yvrC-yvrB-yvrA-yvqK*	unknown; similar to iron transport proteins
TPP	*thiC*	biosynthesis of thiamine pyrimidine moiety
	tenA1-thiX1-thiY1-thiz1-thiE2-thiO-thiS-thiG-thiF-thiD	thiamine biosynthesis
	ykoF-ykoE-ykoD-ykoC	unknown
	yuaJ	unknown; putative thiamine transporter
	ylmB	similar to acetylornithine deacetylase
guanine	*yxjA*	similar to pyrimidine nucleoside transport
	xpt-pbuX	xanthine permease
	pbuG	hypoxanthine/guanine permease
	purE-purK-purB-purC-purS-purQ-purL-purF-purM-purN-purH-purD	purine biosynthesis
adenine	*ydhL*	unknown
SAM	*yitJ*	putative methylene tetrahydrafolate reductase
	metI-metC	methionine biosynthesis
	ykrT-ykrS	5′-methylthioadenosine recycling pathway
	ykrW-ykrX-ykrY-ykrZ	5′-methylthioadenosine recycling pathway
	cysH-cysP-sat-cysC-ylnD-ylnE-ylnF	cysteine biosynthesis
	yoaD-yoaC-yoaB	unknown
	metE	methionine synthase, B_{12}-independent
	metK	*S*-adenosylmethionine synthetase
	yusC-yusB-yusA	unknown ABC transporter
	yxjG	unknown
	yxjH	unknown

[a] Gene nomenclature is derived from the SubtiList database[53] except for *metI* and *metC*, which are recent designations.[54] Functional roles for *ypaA*,[55] *yuaJ*,[41] *ykrTS*,[56] and *ykrWXYZ*[56] have recently been proposed.

genomic, bioinformatics, and biochemical techniques. It is possible that the riboswitches observed to date are of ancient origin, and thus, their evolutionary distribution and functions might reflect this possibility. Further examination of the kinetic and molecular recognition properties will also reveal whether riboswitches are equal to the task of genetic control, as compared to protein factors, or whether we are observing the last vestiges of an ancient but imperfect form of genetic regulation machinery. Given the fundamental metabolic processes that riboswitches control in modern organisms, further studies hold considerable promise for revealing insights into how organisms orchestrate complex genetic networks.

Keywords: aptamers · gene expression · riboswitches · RNA · transcription

[1] A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, R. R. Breaker, *Chem. Biol.* **2002**, *9*, 1043 – 1049.
[2] W. Winkler, A. Nahvi, R. R. Breaker, *Nature* **2002**, *419*, 952 – 956.
[3] W. C. Winkler, S. Cohen-Chalamish, R. R. Breaker, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15 908 – 15 913.
[4] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, *Cell* **2003**, *113*, 577 – 586.
[5] A. S. Mironov, I. Gusarov, R. Rafikov, L. E. Lopez, K. Shatalin, R. A. Kreneva, D. A. Perumov, E. Nudler, *Cell* **2002**, *111*, 747 – 756.
[6] B. A. McDaniel, F. J. Grundy, I. Artsimovitch, T. M. Henkin, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3083 – 3088.
[7] W. C. Winkler, A. Nahvi, N. Sudarsan, J. E. Barrick, R. R. Breaker, *Nat. Struct. Biol.* **2003**, *10*, 701 – 707.
[8] J. Johansson, P. Mandin, A. Renzoni, C. Chiaruttini, M. Springer, P. Cossart, *Cell* **2002**, *110*, 551 – 561.
[9] V. Epshtein, A. S. Mironov, E. Nudler, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5052 – 5056.
[10] N. Sudarsan, J. K. Wickiser, R. R. Breaker, unpublished results.
[11] M. Mandal, R. R. Breaker, unpublished results.
[12] T. Hermann, D. J. Patel, *Science* **2000**, *287*, 820 – 825.
[13] L. Gold, B. Polisky, O. Uhlenbeck, M. Yarus, *Annu. Rev. Biochem.* **1995**, *64*, 763 – 797.
[14] N. Sudarsan, J. E. Barrick, R. R. Breaker, *RNA* **2003**, *9*, 644 – 647.
[15] M. Famulok, *Curr. Opin. Struct. Biol* **1999**, *9*, 324 – 329.
[16] J. P. Richardson, *Biochim. Biophys. Acta* **2002**, *1577*, 251 – 260.
[17] I. Gusarov, E. Nudler, *Mol. Cell* **1999**, *3*, 495 – 504.
[18] W. S. Yarnell, J. W. Roberts, *Science* **1999**, *284*, 611 – 615.
[19] A. Schwartz, A. R. Rahmouni, M. Boudvillain, *EMBO J* **2003**, *22*, 3385 – 3394.
[20] F. Lillo, S. Basile, R. N. Mantegna, *Bioinformatics* **2002**, *18*, 971 – 979.
[21] T. M. Henkin, *Curr. Opin. Microbiol.* **2000**, *3*, 149 – 153.
[22] E. Nudler, M. E. Gottesman, *Genes Cells* **2002**, *7*, 755 – 768.
[23] F. J. Grundy, W. C. Winkler, T. M. Henkin, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11121 – 11126.
[24] T. M. Henkin, C. Yanofsky, *Bioessays* **2002**, *24*, 700 – 707.
[25] J. Stulke, *Arch. Microbiol.* **2002**, *177*, 433 – 440.
[26] M. S. Gelfand, A. A. Mironov, J. Jomantas, Y. I. Kozlov, D. A. Perumov, *Trends Genet.* **1999**, *15*, 439 – 442.
[27] A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, M. S. Gelfand, *Nucleic Acids Res.* **2002**, *30*, 3141 – 3151.
[28] Y. V. Kil, V. N. Mironov, I. Gorishin, R. A. Kreneva, D. A. Perumov, *Mol. Gen. Genet.* **1992**, *233*, 483 – 486.
[29] V. N. Mironov, A. S. Kraev, M. L. Chikindas, B. K. Chernov, A. I. Stepanov, K. G. Skryabin, *Mol. Gen. Genet.* **1994**, *242*, 201 – 208.
[30] J. Perkins, J. Pero in *Bacillus subtilis and Its Closest Relatives: From Genes to Cells* (Eds.: A. L. Sonenshein, J. A. Hoch, R. Losick), American Society for Microbiology, Washington D.C., **2002**, pp. 271 – 286.
[31] P. Burgstaller, M. Famulok, *Angew. Chem.* **1994**, *106*, 1163 – 1166; *Angew. Chem. Int. Ed.* **1994**, *33*, 1084 – 1087.
[32] C. T. Lauhon, J. W. Szostak, *J. Am. Chem. Soc.* **1995**, *117*, 1246 – 1257.
[33] M. Roychowdhury-Saha, S. M. Lato, E. D. Shank, D. H. Burke, *Biochemistry* **2002**, *41*, 2492 – 2499.

[34] G. A. Soukup, R. R. Breaker, *RNA* **1999**, *5*, 1308 – 1325.
[35] V. Tereshko, S. T. Wallace, N. Usman, F. E. Wincott, M. Egli, *RNA* **2001**, *7*, 405 – 420.
[36] W. C. Winkler, S. Cohen-Chalamish, R. R. Breaker, unpublished results.
[37] F. J. Grundy, T. M. Henkin, *Mol. Microbiology* **1998**, *30*, 737 – 749.
[38] S. Kochhar, H. Paulus, *Microbiology* **1996**, *142*, 1635 – 1639.
[39] M. Mandel, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, *Cell* **2003**, *113*.
[40] A. Nahvi, J. E. Barrick, M. S. Ebert, R. R. Breaker, unpublished results.
[41] D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, M. S. Gelfand, *J. Biol. Chem.* **2002**, *277*, 48 949 – 48 959.
[42] J. Miranda-Rios, M. Navarro, M. Soberon, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9736 – 9741.
[43] T. P. Begley, D. M. Downs, S. E. Ealick, F. W. McLafferty, A. P. Van Loon, S. Taylor, N. Campobasso, H. J. Chiu, C. Kinsland, J. J. Reddick, J. Xi, *Arch. Microbiol.* **1999**, *171*, 293 – 300.
[44] E. Webb, F. Febres, D. M. Downs, *J. Bacteriol.* **1996**, *178*, 2533 – 2538.
[45] E. Webb, D. Downs, *J. Biol. Chem.* **1997**, *272*, 15 702 – 15 707.
[46] J. M. Lee, S. Zhang, S. Saha, S. Santa Anna, C. Jiang, J. Perkins, *J. Bacteriol.* **2001**, *183*, 7371 – 7380.
[47] W. C. Winkler, I. Puskarz, R. R. Breaker, unpublished results.
[48] M. Mandel, R. R. Breaker, unpublished results.
[49] H. B. White III, *J. Mol. Evol.* **1976**, *7*, 101 – 104.
[50] G. F. Joyce, *Nature* **2002**, *418*, 214 – 221.
[51] G. J. Zaman, P. J. Michiels, C. A. van Boeckel, *Drug Discov. Today* **2003**, *8*, 297 – 306.
[52] D. Kiga, Y. Futamura, K. Sakamoto, S. Yokoyama, *Nucleic Acids Res.* **1998**, *26*, 1755 – 1760.
[53] I. Moszer, L. M. Jones, S. Moreira, C. Fabry, A. Danchin, *Nucleic Acids Res.* **2002**, *30*, 62 – 65.
[54] S. Auger, W. H. Yuen, A. Danchin, I. Martin-Verstraete, *Microbiology* **2002**, *148*, 507 – 518.
[55] R. A. Kreneva, M. S. Gel'fand, A. A. Mironov, A. Iomantas Iu, I. Kozlov Iu, A. S. Mironov, D. A. Perumov, *Genetika* **2000**, *36*, 1166 – 1168.
[56] B. A. Murphy, F. J. Grundy, T. M. Henkin, *J. Bacteriol.* **2002**, *184*, 2314 – 2318.

Received: June 11, 2003 [M 685]

