ÉTUDE DE LA RELATION STRUCTURE-FONCTION D'UN RIBORÉGULATEUR LYSINE

par

Simon Blouin

thèse présentée au Département de biologie en vue
de l'obtention du grade de docteur ès sciences (Ph.D.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, avril 2011

Your file Votre référence
ISBN: 978-0-494-75052-0
Our file Notre référence
ISBN: 978-0-494-75052-0

Le 28 avril 2011

le jury a accepté la thèse de Monsieur Simon Blouin
dans sa version finale.

Membres du jury

Professeur Daniel Lafontaine
Directeur de recherche
Département de biologie

Professeur Éric Massé
Membre
Département de biochimie

Monsieur Serguei Chteinberg
Membre externe
Université de Montréal

Professeur Kamal Bouarab
Président rapporteur
Département de biologie

SOMMAIRE

Une imposante portion du génome d'un organisme ne sera jamais exprimée sous forme de protéines. Cet acide désoxyribonucléique (ADN) non codant se trouve cependant à être, en bonne partie, transcrit en acide ribonucléique (ARN). Depuis plusieurs années, les études menées sur ces ARN non codants (ARNnc) démontrent qu'ils seraient impliqués dans différents modes de régulation génétique. Les riborégulateurs sont un exemple de telles voies de régulation nouvellement élucidées. La nouveauté du mécanisme réside dans le fait qu'en principe, ces ARNnc modulent l'expression génétique sans avoir recours aux protéines. C'est la liaison directe d'un métabolite à l'ARN messager (ARNm) qui, par l'induction de changements allostériques, influence l'expression du gène situé en aval du riborégulateur. La bactérie *Bacillus subtilis* possède une grande variété de riborégulateurs qui régulent l'expression de plus de 2% de ses gènes. Le présent travail porte sur le riborégulateur lysine *lysC* de *B. subtilis* et a comme objectif de comprendre l'influence de la structure du riborégulateur sur sa fonction de régulation de la transcription.

Dans un premier temps, mes travaux ont démontré la présence et l'importance fonctionnelle d'une interaction boucle-boucle entre les tiges P1 et P2, et d'un motif de type *kink turn* (*K-turn*) dans la tige P2. L'effet de ces motifs structuraux a été évalué par des essais de transcription, où le riborégulateur est inactivé lorsque des mutations sont introduites dans ces motifs. Des essais de fluorescence montrent aussi les répercussions de la formation de cette interaction sur la conformation du site de liaison. Finalement, nos résultats démontrent que l'interaction boucle-boucle dépend de la présence du motif *K-turn*, d'où l'importance de ce dernier.

Dans une seconde publication, nous avons démontré qu'il y avait un effet synergique entre le repliement de l'interaction boucle-boucle et celui d'une deuxième

interaction tertiaire de type hélice-boucle, entre la tige P2 et la boucle de la tige P4. L'influence de cette dernière sur l'organisation du site de liaison et sur la terminaison prématurée de la transcription a aussi été établie, entre autres avec l'utilisation d'un vecteur FRET impliquant les tiges P1 et P5 qui s'est avéré très informatif sur le repliement du coeur. Cet article fait également la démonstration de la contribution des autres éléments du riborégulateur qui ne sont pas impliqués dans des interactions tertiaires, notamment la tige P5 de l'aptamère et une tige nommée P6 située dans la plate-forme d'expression, dont les longueurs sont variables d'une séquence à l'autre.

Les résultats obtenus au cours de mon doctorat nous ont aidé à mieux comprendre comment les changements de structure du riborégulateur lysine affectent sa capacité à réguler l'expression d'un gène. De telles connaissances pourraient s'avérer précieuses pour une éventuelle utilisation des riborégulateurs comme outils biologiques ou comme cibles pour de nouveaux antibiotiques. J'ai d'ailleurs participé à l'écriture d'une revue qui traite des mécanismes de régulation des riborégulateurs et leur potentiel comme outils biologiques. Cette revue a été publiée dans le journal *ChemBioChem* et est jointe en annexe.

REMERCIEMENTS

Je tiens avant tout à remercier le Pr Daniel Lafontaine, sans qui tout ce travail n'aurait pas été possible. Je voudrais le remercier pour la confiance qu'il a manifestée envers moi et les opportunités qu'il m'a offertes. Je voudrais aussi remercier les Prs Marco Di Fruscio, Kamal Bouarab et Éric Massé pour avoir agit en tant que conseillers pendant une partie ou la totalité de mes études graduées. Merci au Pr Serguei Chteinberg de faire partie des quelques privilégiés qui, par obligation professionnelle, consacreront de leur précieux temps à lire cette thèse. Merci aux membres présents et passés du laboratoire qui ont été ma vie au quotidien des dernières années et dont plusieurs ont été beaucoup plus que des collègues; Jeff, mon éternel Ribo-co-chambreur, pour nous avoir tout appris du lab, même les potins, et pour son humour désopilant; Benny pour sa franchise entraînante; Renaud, notre incarnation du gros bon sens, pour ses discussions aussi éclairées qu'éclairantes; Jéjé pour son honnêteté sans limite, sa générosité et son hospitalité; Anne-Marie pour le FRET et les bons moments; Audrey pour avoir mis un peu d'ordre dans ce lab de gars; et Pat pour les débats émotifs.

Merci aux membres du laboratoire Blouin (pas le mien), avec qui nous avons toujours eu une relation particulière, autant au travail qu'autour de sources houblonnées.

Merci au FQRNT et au CRSNG qui m'ont permis de manger décemment et de payer les factures sans trop m'endetter pendant mes études graduées.

Merci à Marie et Jacques, mes parents, qui m'ont toujours laissé aller dans mes projets les plus fous et sans qui je ne me serais jamais rendu là.

Merci très spécial à Gab pour m'avoir soutenu, motivé et surtout enduré pendant toutes ces années qu'elle a su embellir et faire paraître moins longues!

«I feel a very unusual sensation - if it is not indigestion, I think it must be gratitude»

- Benjamin Disraeli (1804-1881)

TABLE DES MATIÈRES

LISTE DES ABRÉVIATIONS

2AP:	2-aminopurine
ADN:	Acide désoxyribonucléique
AEC:	*S*-(2-aminoéthyl)-L-cystéine
ARN:	Acide ribonucléique
ARNm:	ARN messager
ARNnc:	ARN non codant
ARNP:	ARN polymérase
ARNr:	ARN ribosomal
ARNt:	ARN de transfert
ATPase:	Adénosine triphospatase
di-GMPc:	Di-guanosine monophosphate cyclique
FRET:	*Fluorescene resonance energy transfer*
GlcN6P:	Glucosamine-6-phosphate
K-turn:	*Kink-turn*
Mfd:	*Mutation frequency decline*
miARN:	micro ARN
NTP:	Nucléotide triphosphate
Ops:	*Operon polarity supperssor*
PSTV:	Viroïde des tubercules de pomme de terre en fuseau (*Potato Spindle Tuber Viroid*)
RBS:	*Ribosome binding site*
ρ:	Rho
RNase:	Ribonucléase
rut:	*Rho-utilization site*
SAH:	*S*-adénosylhomocystéine
SAM:	*S*-adénosylméthyonine

SD:	Shine-Dalgarno
SELEX:	*Systematic evolution of ligands by exponential enrichment*
SHAPE:	S*elective 2'-hydroxyl acylation analyzed by primer extension*
siARN:	Petit ARN interférant (*Small Interfering*)
σ:	Sigma
smFRET:	*Single molecule* FRET
snARN:	Petit ARN nucléaire (*Small Nuclear*)
snoARN:	Petit ARN nucléolaire (*Small Nucleolar*)
TPP:	Thiamine pyrophosphate
TRAP:	*trp RNA-binding attenuation protein*
UTR:	*Untranslated region*
VIH:	Virus de l'immunodéficience humaine

LISTE DES TABLEAUX

CHAPITRE 1

CHAPITRE 2

LISTE DES FIGURES

INTRODUCTION

CONCLUSION

CHAPITRE 1

CHAPITRE 2

INTRODUCTION

1. Régulation par l'ARN

Une partie importante du génome des organismes est constituée de séquences non codantes, c'est-à-dire qui ne sont pas traduites en protéines. Cette portion non codante du génome augmente en importance avec le niveau de complexité de l'organisme. Par exemple, chez les procaryotes, on estime qu'environ 20% du génome est occupé par des séquences non codantes qui sont transcrites en ARNm, sans être traduites en protéines (Mattick et Makunin, 2006). À l'autre extrême, on évalue à 98,5% la portion du génome humain qui ne code pas pour des protéines (Mattick, 2004). On estime cependant que 40% de ce même génome serait transcrit en ARN (Matera *et al.*, 2007), ce qui laisse sous-entendre que l'ARNnc pourrait avoir un rôle à jouer, autre que celui d'être traduit en protéine

Une grande variété d'ARNnc est aujourd'hui répertoriée. Les premiers exemples connus sont les ARN de transfert (ARNt), qui apportent les acides aminés au ribosome pendant la traduction, ainsi que les ARN ribosomaux (ARNr), assurent l'activité peptidyl transférase du ribosome. On peut aussi mentionner les petits ARN nucléaires (snARN) et nucléolaires (snoARN), respectivement impliqués dans l'épissage et les modifications de l'ARNr (Mattick et Makunin, 2006). Les régions 5′ et 3′ non traduites (5′-UTR et 3′-UTR) des ARNm sont aussi considérées comme des ARNnc (Winkler, 2005b).

Plusieurs classes de petits ARNnc ayant un rôle de régulation ont récemment été découvertes. Principalement observés chez les eucaryotes, les exemples les mieux connus sont les petits ARN d'interférences (siARN) et les microARN (miARN) qui sont des éléments qui régulent en *trans*, c'est-à-dire qui agissent sur une autre

molécule. Ces petites séquences d'ARN se lient aux ARNm pour en induire la dégradation ou pour en inhiber la traduction (Rana, 2007).

Bien qu'ils semblent présents en moins grand nombre, des petits ARN de régulation ont aussi été identifiés chez les procaryotes. On estime qu'il y en aurait près d'une centaine chez *Escherichia Coli* (Gottesman et Storz, 2010), dont au moins une vingtaine interagiraient avec la protéine chaperone Hfq pour favoriser l'interaction spécifique avec un ARN (Gottesman et Storz, 2010; Zhang *et al.*, 2003).

Certains ARN agissant en *cis*, c'est-à-dire sur la même molécule d'ARN, ont aussi été identifiés comme éléments régulateurs. Cette régulation survient à la suite de la liaison d'une protéine, d'un autre ARN ou d'un métabolite (Winkler, 2005a). Par exemple, la biosynthèse du tryptophane est régulée par la liaison de la protéine TRAP (*Trp RNA-binding Attenuation Protein*) qui, en liant l'ARNm, inhibe la transcription et la traduction (Babitzke, 2004). La régulation par la liaison ARN-ARN a beaucoup été étudiée avec le mécanisme de la « *T-box* » qui détecte une carence en acide aminé. La reconnaissance d'un ARNt non chargé de son acide aminé active alors la transcription de gènes impliqués dans cette synthèse (Grundy et Henkin, 2003). Des ARN régulateurs, qui sont quant à eux impliqués dans la régulation génétique dirigée par une liaison ARN-métabolite, ont récemment été observés (Winkler et Breaker, 2003). Ces éléments, ici appelés riborégulateurs (du terme original *riboswitches*), font aujourd'hui l'objet de plusieurs études afin de nous aider à comprendre les mécanismes impliqués dans ce mode de régulation génétique.

2. Riborégulateurs

2.1 Historique

Dans les années 1990, la diversité des structures d'ARN et l'émergence des fonctions lui étant associé ont amené les chercheurs à développer de nouveaux outils. La méthode d'évolution artificielle SELEX (*Systematic Evolution of Ligands by Exponential Enrichment*) (Tuerk et Gold, 1990) a permis de développer des aptamères synthétiques, capables de lier spécifiquement de petits métabolites (Lauhon et Szostak, 1995). Il a même été rapporté que de tels aptamères artificiels pouvaient être placés en amont d'un gène pour en contrôler l'expression (Werstuck et Green, 1998).

Des liaisons ARN-métabolite aussi spécifiques n'étaient cependant pas encore observées *in vivo*, mais on commençait à considérer que des structures conservées en amont de certains gènes pourraient être impliquées dans leur régulation (Gelfand *et al.*, 1999; Miranda-Ríos *et al.*, 2001; Wei *et al.*, 1992). Des études menées sur le gène *btuB* de *E. coli* et sur l'opéron *cob* de *Salmonella typhimurium* ont démontré que des mutations situées dans les régions 5'-UTR correspondantes affectaient la régulation de la traduction, mais aucun facteur protéique impliqué dans ces mécanismes n'avait été identifié (Franklund et Kadner, 1997; Lundrigan *et al.*, 1991; Nou et Kadner, 1998; Nou et Kadner, 2000; Ravnum et Andersson, 1997; Richter-Dahlfors *et al.*, 1994). Devant l'absence de protéine responsable de la régulation, certains laboratoires sont arrivés à la conclusion que l'expression de certains gènes était régulée négativement par la présence d'un métabolite cellulaire qui induisait des changements structuraux au niveau de l'ARN (Grundy *et al.*, 2003; McDaniel *et al.*, 2003; Mironov *et al.*, 2002; Nou et Kadner, 2000). Il a par exemple été démontré en 2002 que le coenzyme B_{12} pouvait lier directement l'ARNm en amont du gène *btuB* pour induire un changement allostérique de l'ARNm et réguler la traduction, et

ce, sans l'aide de protéine (Nahvi *et al.,* 2002). Depuis, de nombreuses études ont démontré que ces structures d'ARNm sont conservées à travers les différentes espèces bactériennes et sont localisées en amont de plusieurs gènes impliqués dans le métabolisme de leur ligand respectif (Barrick *et al.,* 2004; Mandal *et al.,* 2003; Nudler et Mironov, 2004; Regulski *et al.,* 2008; Tucker et Breaker, 2005; Weinberg *et al.,* 2007; Weinberg *et al.,* 2010; Winkler et Breaker, 2003).

2.2 Définition

Le groupe de Breaker a été le premier à employer le terme *riboswitch* pour définir une portion structurée de l'ARNm non codant qui lie un métabolite pour réguler l'expression d'un gène, sans que des protéines soient impliquées (Nahvi *et al.,* 2002). Bien que certains groupes incluent parfois des éléments d'ARN thermosensibles (Chowdhury *et al.,* 2006; Johansson *et al.,* 2002), les «*T-box*» (Green *et al.,* 2010; Henkin, 2008; Smith *et al.,* 2010a) et les éléments d'ARN qui répondent aux changements de pH (Nechooshtan *et al.,* 2009) dans leur définition de riborégulateur à cause de leur capacité à réguler l'expression des gènes en fonction des changements physiologiques de la cellule, nous nous en tiendrons ici aux ARN capables de lier des métabolites non protéiques.

Les riborégulateurs sont constitués de deux domaines distincts: l'aptamère et la plate-forme d'expression (Figure 1). L'aptamère représente le domaine de liaison du métabolite et est responsable de la spécificité envers ce ligand. Les nucléotides qui sont directement impliqués dans la liaison du ligand sont généralement très conservés d'une espèce à l'autre (Serganov, 2009). La plate-forme d'expression est la deuxième région fonctionnelle d'un riborégulateur. C'est l'organisation de cette dernière qui influence l'expression du gène situé en aval. Ces deux portions distinctes d'un riborégulateur sont par contre intimement reliées et se chevauchent

de manière à pouvoir adopter deux conformations mutuellement exclusives. Les changements de conformation de l'aptamère, induits par la liaison du ligand, affectent ainsi directement la structure de la plate-forme d'expression. Cette dernière peut donc adopter une structure «permissive» ou «non permissive». Généralement, en absence de ligand, le riborégulateur est sous sa forme «permissive» qui permet l'expression du gène. Quand le ligand lie l'aptamère, ce dernier est stabilisé dans sa forme «non permissive» et la plate-forme d'expression adopte plutôt une structure qui empêche l'expression du gène en aval. Seuls les riborégulateurs adénine, glycine et SAH (*S*-adénosylhomocystéine) semblent agir comme activateurs en présence de leur ligand (Mandal et Breaker, 2004; Mandal *et al.*, 2004; Wang *et al.*, 2008).

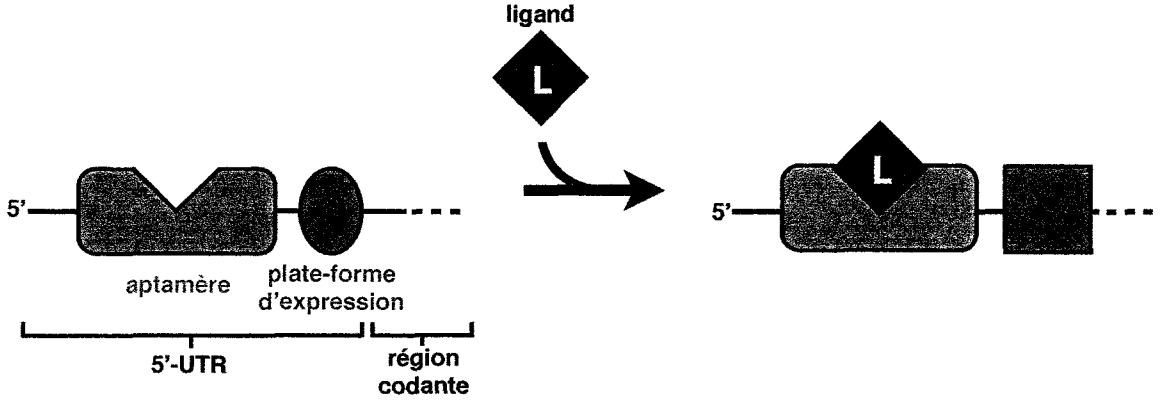


Figure 1. Anatomie d'un riborégulateur. L'aptamère et la plate-forme d'expression du riborégulateur sont représentés en vert et en bleu, respectivement. La liaison du ligand (L) à l'aptamère induit un changement de conformation de la plate-forme d'expression.

2.3 Familles

Depuis les premières observations faites sur les riborégulateurs flavine mononucléotide (FMN) et thiamine pyrophosphate (TPP) (Mironov *et al.,* 2002; Nahvi *et al.,* 2002; Winkler *et al.,* 2002a; Winkler *et al.,* 2002b), une variété grandissante de tels éléments de régulation a été observée dans les trois royaumes du vivant (Tucker et Breaker, 2005) et ils sont habituellement nommés et classés en fonction du ligand qu'ils reconnaissent (Figures 2 et 3). À ce jour, des riborégulateurs naturels sont connus pour lier l'adénine (Mandal et Breaker, 2004), la 7-aminométhyl-7-déazaguanine (preQ1) (Roth *et al.,* 2007), le coenzyme B_{12} (Nahvi *et al.,* 2002), la 2′-désoxyguanosine (Kim *et al.,* 2007), la glutamine (Ames et Breaker, 2011), la di-guanosine monophosphate cyclique (di-GMPc) (Sudarsan *et al.,* 2008), la glucosamine-6-phosphate (GlcN6P, nommé riborégulateur *glmS*) (Winkler *et al.,* 2004), la guanine (Mandal *et al.,* 2003), la glycine (Barrick *et al.,* 2004; Mandal *et al.,* 2004), la lysine (Grundy *et al.,* 2003; Sudarsan *et al.,* 2003b) , le magnésium (Cromie *et al.,* 2006), le molybdène (Moco) (Regulski *et al.,* 2008), la SAH (Wang *et al.,* 2008), la *S*-adénosylméthyonine (SAM) (Epshtein *et al.,* 2003; McDaniel *et al.,* 2003; Winkler *et al.,* 2003) et le tétrahydrofolate (Ames *et al.,* 2010). Bien que les riborégulateurs soient principalement présents chez les procaryotes, le riborégulateur TPP a aussi été identifié chez certains eucaryotes comme *Arabidopsis thaliana*, *Neurospora crassa* et *Aspergillus oryzae* (Kubodera *et al.,* 2003; Sudarsan *et al.,* 2003a).

Chacune des familles de riborégulateurs se distingue par la taille et la complexité de l'aptamère, qui sont généralement conservées au sein d'une même famille (Figure 3). Les structures de ces aptamères vont de la simple tige-boucle pouvant lier le ligand preQ1 (Roth *et al.,* 2007) aux architectures plus complexes, comme les jonctions à cinq voies des riborégulateurs lysine (Grundy *et al.,* 2003; Rodionov *et al.,* 2003; Sudarsan *et al.,* 2003b) et Moco (Regulski *et al.,* 2008). On observe même

des riborégulateurs à jonctions multiples comme ceux reconnaissant le coenzyme B_{12} (Mandal *et al.*, 2003) ou le TPP (Winkler *et al.*, 2002b).

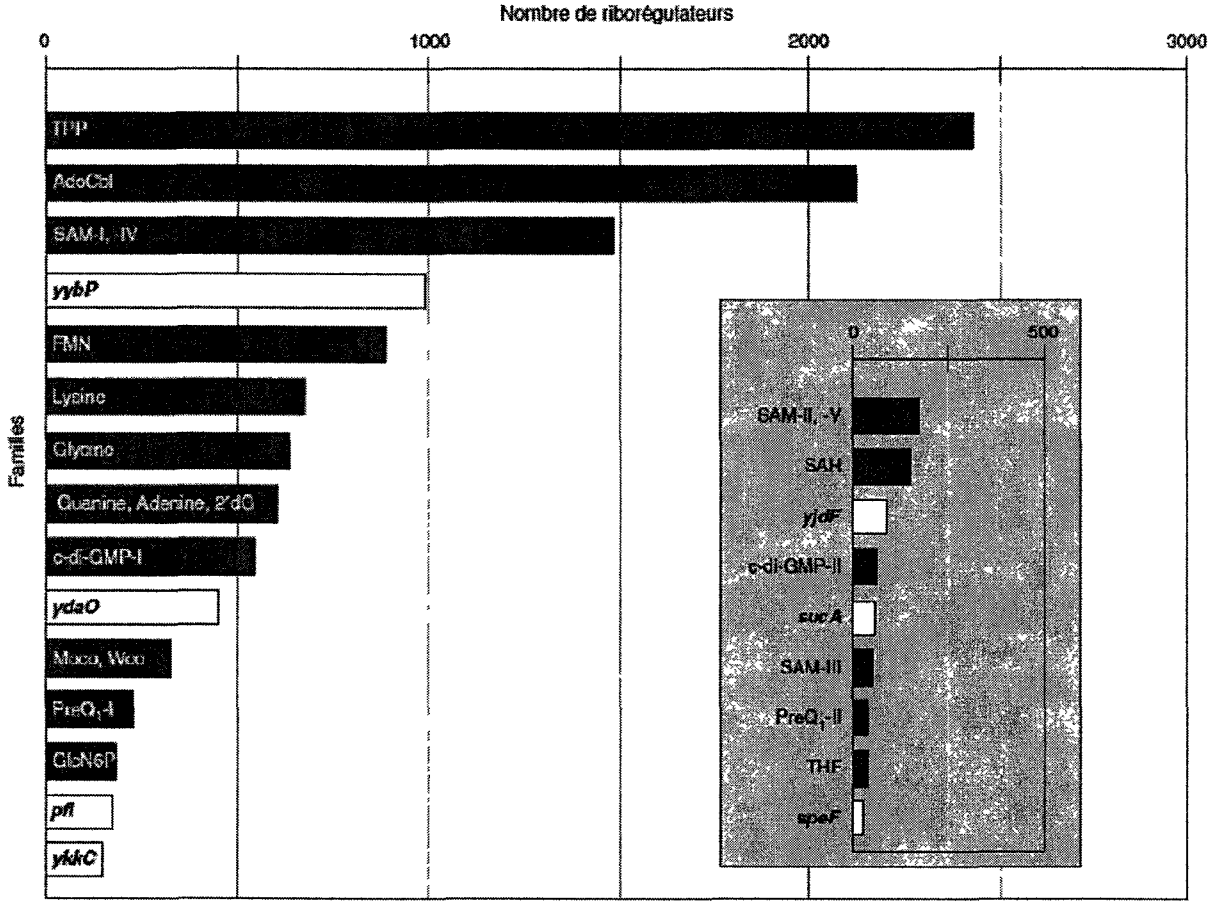


Figure 2. Représentation des différents riborégulateurs. Les riborégulateurs prédits pour lesquels il n'y a pas encore d'évidence expérimentale sont représentés par des rectangles blancs. Adapté de (Breaker, 2010).

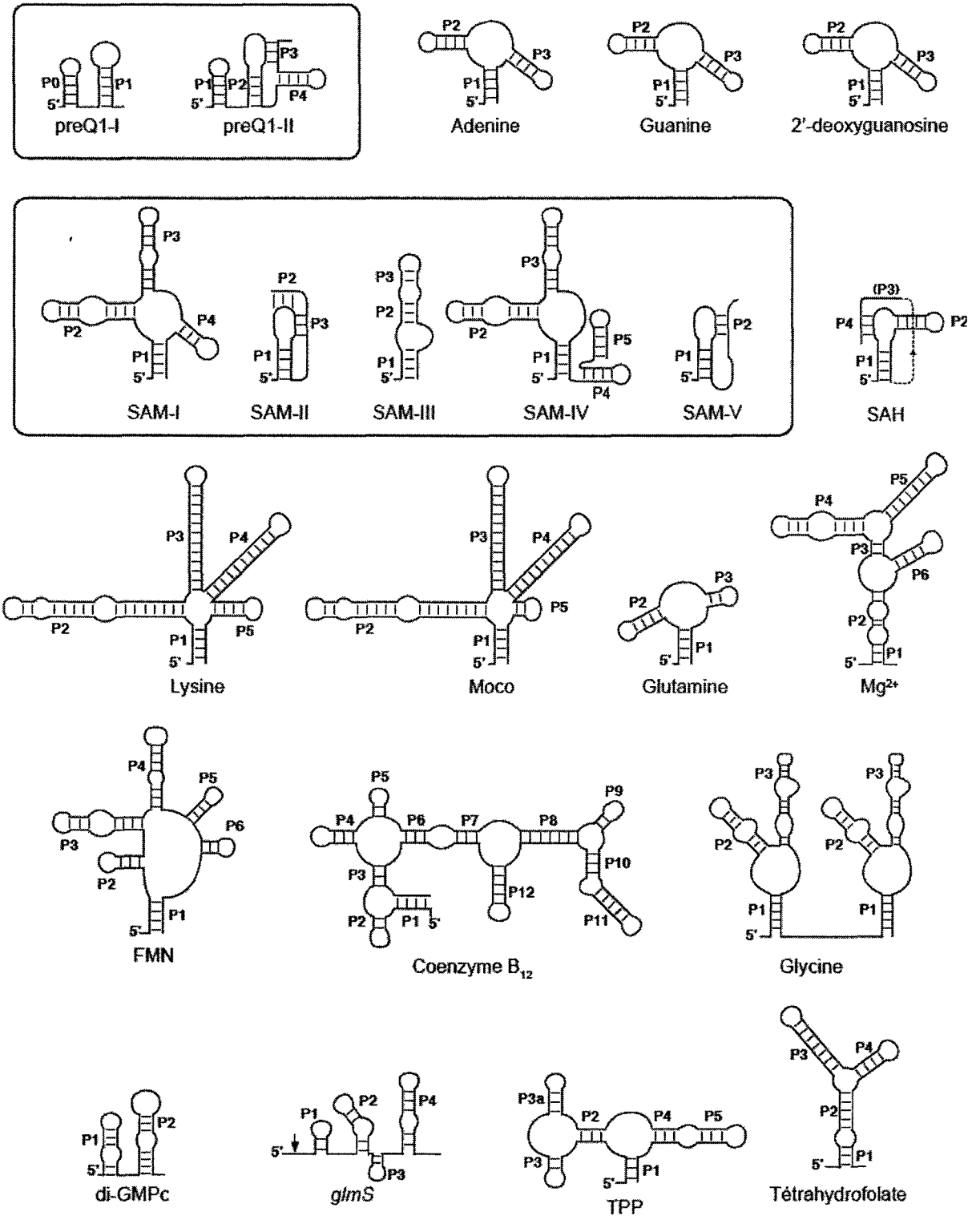

P0
P1
5'
preQ1-I
P3
P1
P2
P4
5'
preQ1-II
P2
P3
P1
5'
Adenine
Guanine
2'-deoxyguanosine
P4
SAM-I
SAM-II
SAM-III
SAM-IV
SAM-V
(P3)
SAH
P5
P6
Lysine
Moco
Glutamine
Mg2+
FMN
P7
P8
P9
P10
P11
P12
Coenzyme B12
Glycine
di-GMPc
glmS
P3a
TPP
Tétrahydrofolate

Figure 3. Familles de riborégulateurs. Les différentes familles de riborégulateurs répertoriées à ce jour sont identifiées et représentées par leur structure secondaire. L'extrémité 5′ de l'ARN et les tiges (P) des riborégulateurs sont indiquées. Les encadrés regroupent les sous-familles de riborégulateurs preQ1 et SAM qui lient le même ligand. Le site de clivage du riborégulateur-ribozyme *glmS* est indiqué par une flèche. Adapté de (Blouin *et al.*, 2009).

Chaque riborégulateur lie son ligand respectif avec une grande spécificité, c'est-à-dire qu'il discrimine aisément son ligand d'une molécule analogue. Par exemple, le riborégulateur lysine est aussi capable de lier la *S*-(2-aminoéthyl)-L-cystéine (AEC), un composé analogue à la L-lysine, mais avec une affinité trente fois moindre (Sudarsan *et al.*, 2003b). Cette spécificité est bien illustrée par les trois familles de riborégulateurs purines liant respectivement l'adénine, la guanine et la 2′-désoxyguanosine. Ces trois riborégulateurs sont constitués de jonctions à trois voies et présentent des structures secondaires et tertiaires très similaires (Edwards et Batey, 2009; Kim *et al.*, 2007). Des changements mineurs dans la séquence du site de liaison de l'aptamère permettent ainsi aux trois riborégulateurs, qui paraissent identiques au niveau de leur structure globale, à réagir à la présence de différents ligands.

À l'opposé des riborégulateurs purines, certaines familles utilisent différentes conformations d'ARN qui arrivent à reconnaître le même ligand. Les riborégulateurs preQ1 et SAM peuvent ainsi être répertoriés en sous-familles, puisqu'on retrouve respectivement deux (preQ1-I et preQ1-II) et cinq (SAM-I, SAM-II, SAM-III/S_{MK}, SAM-IV et SAM-V) classes de ces riborégulateurs (Corbino *et al.*, 2005; Fuchs *et al.*, 2006; Poiata *et al.*, 2009; Roth *et al.*, 2007; Weinberg *et al.*, 2008) qui, malgré leurs divergences de structure et de liaison avec le ligand, conservent une grande spécificité envers ce dernier (Blouin *et al.*, 2009).

2.4 Motifs structuraux

C'est avant tout la structure d'un riborégulateur qui dirige sa fonction de régulation. Tel que mentionné dans la section précédente, chaque famille de riborégulateurs se distingue généralement par sa structure secondaire qui lui est caractéristique. À l'intérieur d'une même famille, les alignements de séquences et les analyses de covariation révèlent que certains nucléotides de ces structures sont hautement conservés à travers les différentes espèces. Cette conservation est en partie expliquée par le besoin de maintenir les interactions avec le ligand et est généralement observable dans le coeur de l'aptamère qui sert de site de liaison. D'autres portions des riborégulateurs sont cependant aussi conservées à cause de leur implication dans le maintient de la structure tertiaire, qui est elle-même requise pour le parfait positionnement des nucléotides du coeur. L'ARN utilise ainsi une variété de motifs structuraux qui sont impliqués dans le maintient de la structure tertiaire des riborégulateurs.

2.4.1 Empilement d'hélices

Les riborégulateurs sont souvent organisés en jonctions qui relient plusieurs tiges ou hélices d'ARN. L'arrangement structural le plus souvent observé est l'empilement de ces hélices (*helical stacking*), où deux tiges d'ARN s'empilent l'une sur de l'autre dans un même axe pour donner l'apparence d'une seule longue hélice d'ARN double-brin. Ce type d'arrangement est entre autres facilement observable dans la plupart de jonctions d'ARN à trois voies comme le ribozyme à tête de marteau (*hammerhead*) (Martick et Scott, 2006). Les riborégulateurs liant les purines se replient eux aussi selon cette forme en «Y» où deux des trois tiges s'alignent (Batey *et al.*, 2004; Serganov *et al.*, 2004). Ces empilements sont aussi couramment observés dans des structures plus complexes comme les introns du groupe I

(Adams *et al.,* 2004) et les riborégulateurs SAM-I (Montange et Batey, 2006), où quatre hélices s'empilent deux par deux, pour former deux arrangements parallèles en forme de «H». Le riborégulateur lysine est un autre exemple qui démontre la facilité avec laquelle des tiges d'ARN double-brin s'empilent, où la jonction à cinq voies s'organise en forme de «X» avec deux éléments d'empilement (tiges P1-P2 et tiges P4-P5) à la manière des jonctions à quatre voies, avec en plus une cinquième tige (P3) positionnée grâce à d'autres motifs (Garst *et al.,* 2008; Serganov *et al.,* 2008).

2.4.2 Pseudonoeud

Un pseudonoeud se forme entre la boucle terminale d'une tige-boucle et une autre région simple-brin de l'ARN. Il a été démontré que ces structures peuvent être impliquées dans la régulation par la séquestration des sites d'initiation de la traduction ou par le recrutement de protéines (Babitzke *et al.,* 2009). L'importance d'un pseudonoeud a été démontrée dans l'organisation du coeur de l'aptamère du riborégulateur SAM-I (Heppell et Lafontaine, 2008). Il a aussi été démontré ou prédit que les riborégulateurs di-GMPc (Hengge, 2010), *glmS* (Wilkinson et Been, 2005), preQ1 (Klein *et al.,* 2009; Rieder *et al.,* 2009), SAH (Edwards *et al.,* 2010), SAM-II, SAM-IV (Weinberg *et al.,* 2008) et SAM-V (Poiata *et al.,* 2009) requièrent la formation d'un pseudonoeud pour leur activité de régulation ou pour le maintient de leur structure tertiaire.

2.4.3 Interaction boucle-boucle

Les interactions de type boucle-boucle peuvent être définies comme une variante des peudonoeuds. Elles se forment entre deux tiges-boucles lorsque les nucléotides

des boucles terminales sont complémentaires et peuvent former des paires de bases Watson-Crick. L'importance biologique des interactions boucle-boucle a pu être observée à maintes occasions, notamment dans le cas du virus de l'immunodéficience humaine (VIH) où la dimérisation de l'ARN viral est positivement influencée par la formation d'une interaction boucle-boucle (Haddrick *et al.*, 1996). La plupart des jonctions d'ARN à trois voies sont aussi stabilisées par une telle interaction. Tel que mentionné précédemment, elles adoptent une forme en «Y» où deux des trois tiges s'empilent. La structure est alors stabilisée par la formation de paires de bases entre l'une de ces tiges et la troisième (Figure 4) (de la Peña *et al.*, 2009). Chez les riborégulateurs, l'implication fonctionnelle de telles interactions a entre autres été démontrée dans les cas des riborégulateurs adénine (Lemay *et al.*, 2006) et lysine (Blouin et Lafontaine, 2007).

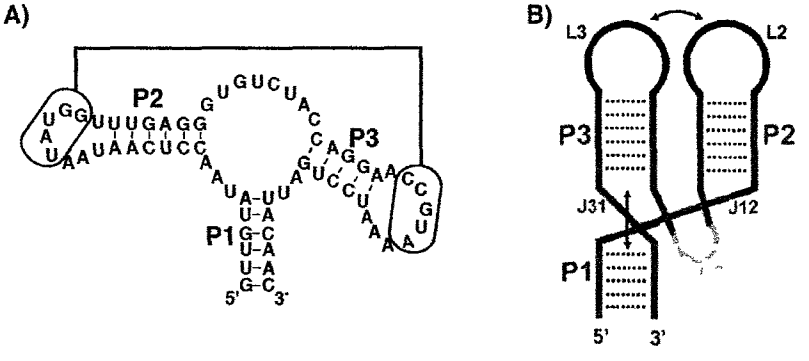


Figure 4. Interaction boucle-boucle (exemple d'une jonction à trois voies). (A) Structure secondaire du riborégulateur adénine. Les tiges et les deux séquences qui interagissent ensembles sont indiquées. Adapté de (Lemay *et al.*, 2006). **(B)** Schéma de la structure tertiaire où les tiges P1 et P3 s'empilent l'une sur l'autre et les boucles L2 et L3 forment une interaction boucle-boucle. Tiré de (de la Peña *et al.*, 2009).

2.4.4 *Kink-turn*

Les motifs *K-turns* sont retrouvés dans des hélices d'ARN double-brin et sont caractérisés par une boucle interne de trois nucléotides suivie de deux paires de bases G•A (Figure 5). Ils ont d'abord été identifiés et définis dans le ribosome (Klein *et al.,* 2001), mais on a rapidement dénombré plusieurs exemples de telles structures dans d'autres ARN (Dennis et Omer, 2005), comme les «*T-box*» et les riborégulateurs SAM-I, où on les avait nommés «motifs G•A» (Winkler *et al.,* 2001). Bien que l'on sache aujourd'hui qu'ils sont omniprésents dans l'ARN, la plupart des *K-turns* connus cadrent toujours dans le consensus établi en 2001 (Antonioli *et al.,* 2010; Klein *et al.,* 2001). Les *K-turns* ont la particularité d'induire un angle prononcé dans la double hélice d'ARN, et plusieurs d'entre eux servent de site de liaison pour des protéines (Klein *et al.,* 2001). Cette liaison protéique induirait une stabilisation de la structure repliée (Turner *et al.,* 2005). Dans les riborégulateurs lysine et SAM-I, la présence et l'importance de tels motifs pour permettre la formation d'interactions boucle-boucle et pseudo-noeud a été établie (Blouin et Lafontaine, 2007; Heppell et Lafontaine, 2008).

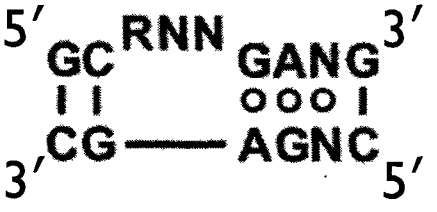


Figure 5. Structure secondaire consensus d'un motif *K-turn*. Les traits verticaux et les cercles représentent des paires de bases Watson-Crick et non canoniques, respectivement. Le trait horizontal représente la continuité de la séquence. Adapté de (Blouin et Lafontaine, 2007).

2.4.5 Boucle E (*E-loop*)

Les motifs en boucle E (parfois nommés *S-turn*) ont été observés pour la première fois dans le viroïde des tubercules de pomme de terre en fuseau (PSTV) (Branch *et al.,* 1985). Ils ont depuis été répertoriés dans un bon nombre d'ARN, comme dans certains ribozymes, ARNr et dans une sous-famille de ribonucléase (RNase) P (Leontis et Westhof, 1998). Ce motif est représenté par une paire de bases A•A suivie d'un G non apparié, d'une paire A•U et d'une paire A•G (Figure 6). Comme pour les *K-turns*, certaines protéines ribosomales les utilisent comme site de liaison (Allison *et al.,* 1991). Le motif en boucle E est apparenté au motif sarcine-ricine, un motif hautement conservé dans l'ARNr (Szewczak et Moore, 1995) et défini par une série de paires de bases non canonique et d'un nucléotide non apparié. Ce type de motif semble être conservé dans les riborégulateurs glutamine (Ames et Breaker, 2011) et dans la tige P2 des riborégulateurs lysine (Garst *et al.,* 2008; Grundy *et al.,* 2003; Serganov *et al.,* 2008; Sudarsan *et al.,* 2003b).

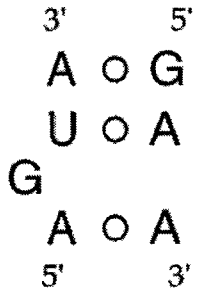


Figure 6. Arrangement typique d'une boucle E. Tiré de (Leontis et Westhof, 1998).

2.4.6 Boucle T (*T-loop*)

Une boucle T est un arrangement particulier de certaines pentaboucles des tiges d'ARN caractérisé par l'empilement d'une paire de bases non canonique entre le premier et le dernier nucléotide de la boucle. Elles ont d'abord été observées dans

la structure de l'ARNt, mais la récurrence du motif a aussi été confirmée avec la résolution de la structure du ribosome (Ban *et al.*, 2000; Nagaswamy et Fox, 2002). La présence de ce type de boucle favorise les interactions ARN-ARN et ARN-protéine (Nagaswamy et Fox, 2002). Leur présence a entre autres été rapportée dans les riborégulateurs B_{12}, FMN (Barrick et Breaker, 2007) et TPP (Baird *et al.*, 2010).

2.4.7 Tétraboucles et récepteurs

Les jonctions d'ARN sont constituées de plusieurs tiges-boucles réunies en un centre (la jonction). Les boucles terminales de ces tiges sont souvent organisées en tétraboucles (Woese *et al.*, 1990), c'est-à-dire que la double-hélice est coiffée d'une boucle de quatre nucléotides. Les tétraboucles les plus étudiées sont celles qui adoptent le consensus GNRA (Jaeger *et al.*, 1994) et elles sont fréquemment impliquées dans des interactions tertiaires par une liaison avec le sillon mineur d'une double-hélice distante. Une telle interaction composée d'une tétraboucle GNRA et d'un récepteur a été prédite chez le riborégulateur Moco (Roth et Breaker, 2009) et un arrangement similaire est présent dans le riborégulateur *glmS* (Klein et Ferré-D'Amaré, 2006). Un exemple bien documenté est le domaine P4-P6 de l'intron du groupe I de *Tetrahymena thermophila* (Cate *et al.*, 1996a; Cate *et al.*, 1996b), où une tétraboucle GAAA interagit avec un sillon mineur de l'ARN. Ce motif est aussi nommé «plate-forme adénosine» ou motif «A-mineur» (Nissen *et al.*, 2001) à cause de l'implication de deux adénosines de la tétraboucle dans l'interaction. Une interaction très similaire à celle du domaine P4-P6 est observée dans le riborégulateur lysine, où la boucle de la tige P4 interagit avec la tige P2. Seule différence, la tige P4 est coiffée d'une pentaboucle (GAAUA). L'uridine est cependant exposée vers l'extérieur de la structure et le reste de la séquence est organisée à la manière d'une tétraboucle qui interagit avec la tige P2 via les deux

premières adénosines (Garst *et al.,* 2008).

2.5 Mécanismes d'action

Bien que les riborégulateurs procaryotes soient principalement utilisés pour réguler les mécanismes de transcription et de traduction, il existe quelques cas où le changement de structure de l'ARN mène à un autre processus biologique qui modifie l'expression d'un gène. Le fait d'avoir des aptamères d'une même famille qui sont très conservés et des plates-formes d'expression variables permet cette diversité de mécanismes d'action en réponse à un même ligand, comme nous le verrons dans cette section.

2.5.1 Terminaison de la transcription

Les terminateurs de la transcription indépendants de Rho (ρ), aussi nommés terminateurs intrinsèques (voir section 4.4.2), sont très répandus dans les génomes procaryotes (Lillo *et al.,* 2002). Ils sont entres autres retrouvés en aval d'opérons pour terminer la transcription. Ces terminateurs intrinsèques sont composés d'une tige-boucle riche en paires de bases G-C, suivie d'une séquence de cinq à neuf uridines (Gusarov et Nudler, 1999; Schwartz *et al.,* 2003; Yarnell et Roberts, 1999). De telles structures déstabilisent le complexe d'élongation de la transcription, provocant ainsi la dissociation de l'ARN polymérase et la terminaison de la transcription. La formation de terminateurs intrinsèques est aussi l'un des moyens les plus répandus des plates-formes d'expression des riborégulateurs pour moduler l'expression de gènes. Le riborégulateur alterne donc entre deux conformations mutuellement exclusives qui dépendent de la présence ou non du ligand. La plate-forme d'expression peut ainsi adopter une structure antiterminatrice, qui

compétitionne directement avec la formation du terminateur, ou stabiliser l'aptamère pour empêcher cet antiterminateur de se former. (figure 7). Dans ce dernier cas, il y a formation du terminateur et terminaison prématurée de la transcription, puisque le complexe transcriptionnel est arrêté avant qu'il n'y ait transcription de la région codante du gène en aval.

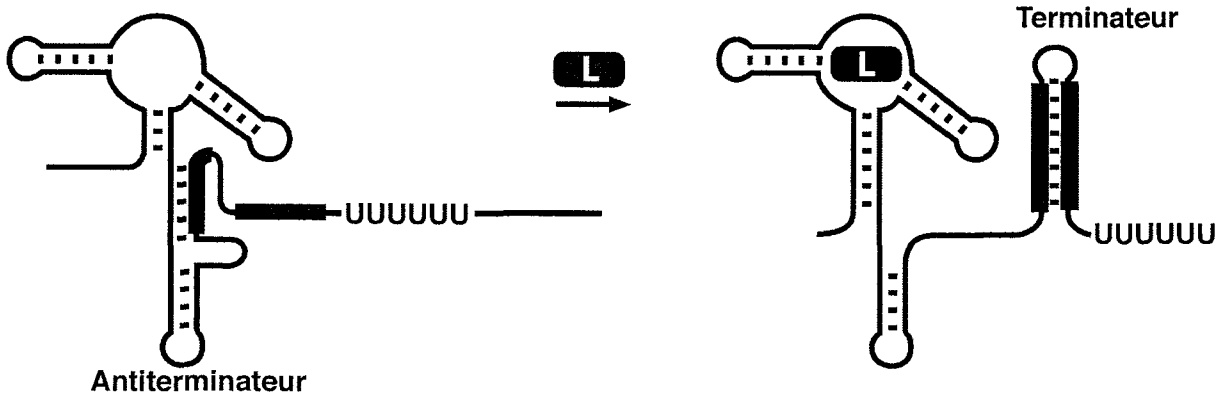


Figure 7. Terminaison prématurée de la transcription par un riborégulateur. La liaison du ligand (L) au riborégulateur empêche l'antiterminateur de se former pour permettre la formation d'un terminateur intrinsèque. Ce dernier est constitué d'une tige-boucle suivie d'une séquence de plusieurs uridines.

2.5.2 Initiation de la traduction

La liaison d'un ligand à un riborégulateur peut aussi moduler l'initiation de la traduction d'une protéine. D'un point de vue structural, le mécanisme est très similaire à celui de la terminaison de la transcription, c'est-à-dire que la plate-forme d'expression peut adopter deux conformations mutuellement exclusives, dépendamment de la liaison ou non du ligand à l'aptamère. Lorsqu'un tel

riborégulateur se retrouve dans sa forme «non permissive», une tige-boucle se forme pour séquestrer la séquence Shine-Dalgarno (SD, ou RBS pour *Ribosome Binding Site*) et/ou le codon d'initiation de la traduction (figure 8). Les riborégulateurs traductionnels sont ainsi localisés beaucoup plus près de la région codante du gène correspondant et ils modulent le recrutement du ribosome à l'ARNm. À l'opposé d'un contrôle transcriptionnel, l'ARNm d'un gène contrôlé par un riborégulateur traductionnel peut ainsi être transcrit au complet. Cette différence notable a récemment été soulevée dans le cas des riborégulateurs adénine, où un riborégulateur contrôlant la transcription du gène *pbuE* chez *B. subtilis* dépend d'une liaison co-transcriptionnelle entre le ligand et l'ARN. À l'inverse, le riborégulateur adénine qui régule la traduction du gène *add* de *Vibrio vulnificus* peut être complètement transcrit avant de lier l'adénine et de moduler l'initiation de la traduction (Lemay *et al.*, 2011).

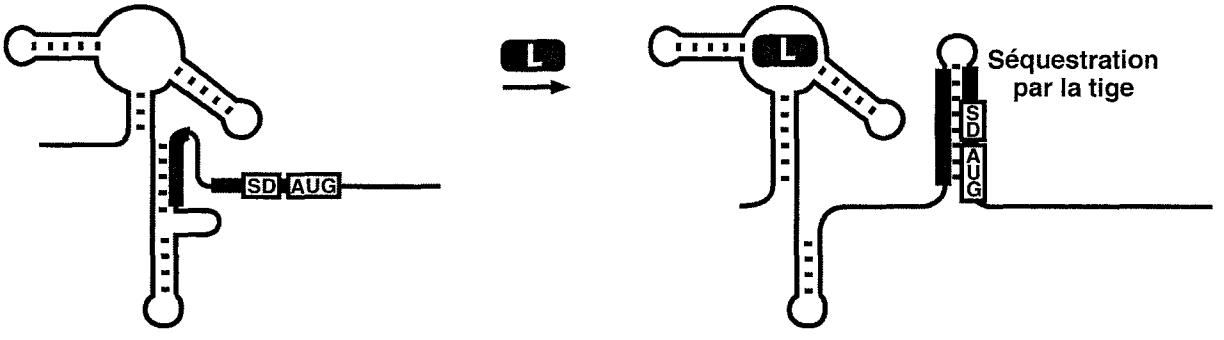


Figure 8. Contrôle de l'initiation de la traduction par un riborégulateur. La liaison du ligand (L) permet la formation d'une tige qui séquestre la séquence SD et le codon d'initiation de la traduction AUG.

Bien que cette règle ne soit pas impérative, il semble y avoir une tendance pour que les bactéries à Gram positif possèdent principalement des riborégulateurs qui régulent la terminaison de la transcription, alors que les bactéries à Gram négatif auraient plutôt des riborégulateurs contrôlant l'initiation de la traduction (Nudler et

Mironov, 2004; Vitreschak *et al.*, 2004; Winkler, 2005a; Winkler et Breaker, 2005). Cette propension pourrait s'expliquer par la proportion plus importante de polycistrons chez les bactéries à Gram positif (Nudler et Mironov, 2004), qui auraient ainsi un avantage énergétique à induire une terminaison prématurée de la transcription d'un ARNm.

2.5.3 Épissage alternatif et stabilité de l'ARNm

Comme la plupart des riborégulateurs connus, le riborégulateur TPP peut moduler l'expression de gènes en contrôlant les mécanismes de transcription ou de traduction chez les procaryotes. Il représente néanmoins un cas particulier, puisqu'il est à ce jour le seul riborégulateur à être aussi retrouvé chez les eucaryotes, où il joue un rôle dans le contrôle de l'épissage alternatif et de la stabilité de l'ARNm. Dans le premier cas, le riborégulateur contrôle l'accessibilité à deux différents sites d'épissage, où un seul de ces sites est utilisé en fonction de la présence ou non de TPP (Cheah *et al.*, 2007). Il a aussi été proposé qu'un élément riborégulateur liant l'arginine aurait un effet similaire sur l'épissage du gène *agaA* de *Aspergillus nidulans* (Borsuk *et al.*, 1999; Borsuk *et al.*, 2007). Le second exemple implique aussi le contrôle d'un événement d'épissage qui fait varier la longueur de l'extrémité 3′ d'un transcrit. En absence de ligand, il n'y a pas d'épissage et il en résulte un transcrit avec une extrémité 3′-UTR plus courte qui favorise l'expression du gène. À l'inverse, lorsqu'il y a liaison entre le TPP et le riborégulateur, le site d'épissage est accessible et le transcrit résultant possède une longue extrémité 3′-UTR qui favorise sa dégradation (Wachter *et al.*, 2007).

2.5.4 Dégradation de l'ARNm

Le riborégulateur liant la GlcN6P se distingue en étant à la fois un riborégulateur et un ribozyme, c'est-à-dire que cet ARN possède une activité catalytique, qui est dépendante de la présence de son ligand (Hampel et Tinsley, 2006; Winkler *et al.*, 2004). Son mode de régulation est ainsi différent parce que plutôt que d'induire un changement de conformation par sa liaison, la GlcN6P agit comme un co-facteur en participant directement à la réaction de coupure de l'ARNm (Cochrane *et al.*, 2007; Hampel et Tinsley, 2006; Klein et Ferré-D'Amaré, 2006; McCarthy *et al.*, 2005; Tinsley *et al.*, 2007). En présence d'une concentration suffisante de GlcN6P, il y a coupure dans la partie 5′-UTR de l'ARNm de la GlcN6P synthase (Winkler *et al.*, 2004). Il en résulte un transcrit avec une extrémité 5′-OH, qui est un substrat pour la RNase J1. Il y a ainsi une dégradation de l'ARNm et une diminution de la production de la protéine (Collins *et al.*, 2007).

2.5.5 Action en *trans*

Une des propriétés des riborégulateurs réside dans leur action en *cis*, c'est-à-dire qu'ils font partie de l'ARNm qu'ils régulent. Il a cependant été récemment rapporté qu'un riborégulateur SAM, retrouvé chez *Listeria monocytogenes*, aurait aussi un effet en *trans* en ciblant la régulation un autre gène (Loh *et al.*, 2010). Dans ce contexte, un riborégulateur SAM est situé en amont des gènes *lmo2419-lmo2418-lmo2417*. La liaison du ligand induit une terminaison de la transcription et seul le riborégulateur est transcrit. Ce court transcrit agit alors comme un petit ARN de régulation et peut s'hybrider à la portion 5′-UTR du gène *prfA* pour diminuer l'expression du facteur de virulence PrfA. Ce gène *prfA* est déjà contrôlé par un élément thermosensible situé dans sa région 5′-UTR qui permet au facteur de virulence d'être exprimé dans l'hôte (37°C), alors qu'il est habituellement réprimé à

des températures inférieures à 30°C (Johansson *et al.,* 2002). En plus de ce premier niveau de régulation, l'aptamère SAM peut s'hybrider au 5′-UTR de *prfA* à 37°C pour en empêcher l'expression. *L. monocytogenes* pourrait ainsi moduler l'expression de facteurs de virulence en fonction de la disponibilité des nutriments (Loh *et al.,* 2010).

2.5.6 Complexité des mécanismes

Une classe distincte de riborégulateurs disposés en tandem se distingue des mécanismes classiques (Figure 9). Le riborégulateur glycine s'apparente à ces exemples parce qu'il est souvent constitué de deux aptamères placés en tandem, en amont d'une seule plate-forme d'expression. Ces deux aptamères lient la glycine de façon coopérative, c'est-à-dire que la liaison d'une molécule de glycine au premier aptamère facilite la liaison d'un deuxième ligand au second aptamère (Mandal *et al.,* 2004). Un autre arrangement particulier de riborégulateur agissant en tandem a été identifié chez *Bacillus clausii.* Ce dernier est composé de deux aptamères qui lient respectivement la SAM et le coenzyme B_{12} (Sudarsan *et al.,* 2006). Contrairement au riborégulateur glycine, les deux aptamères sont suivis d'une tige terminatrice et les deux ligands peuvent inhiber la transcription de façon indépendante. Finalement, un autre exemple de riborégulateurs en tandem a récemment été observé chez *Bacillus anthracis*, où deux aptamères pouvant lier le même ligand (TPP) se suivent et sont tous les deux associés à une plate-forme d'expression (Welz et Breaker, 2007). Il a même été récemment observé que les riborégulateurs glutamine peuvent adopter des arrangements de un, deux ou trois aptamères (Ames et Breaker, 2011). Ces différents arrangements illustrent la flexibilité de la réponse possible par les riborégulateurs en fonction des conditions physiologiques.

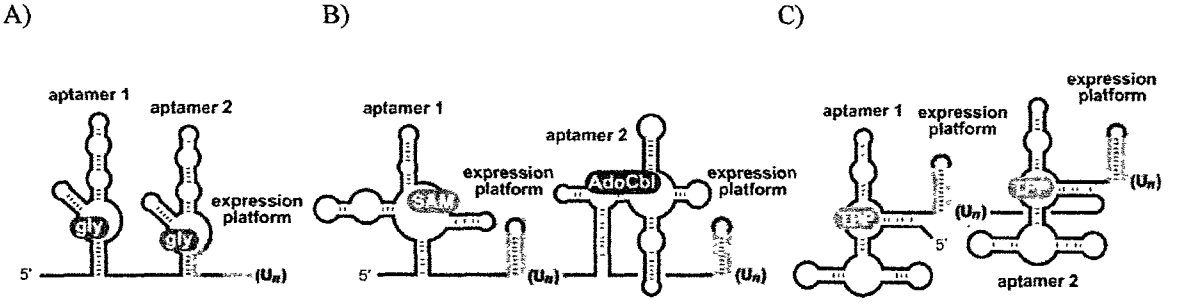


Figure 9. Exemples de riborégulateurs en tandem. (**A**) Riborégulateur glycine où 2 aptamères sont suivis d'une unique plate-forme d'expression. (**B**) Aptamères SAM et B_{12} (AdoCbl) de *B. clausii*, chacun suivi de sa plate-forme d'expression. (**C**) Double riborégulateur de *B. anthracis* formé de 2 aptamères TPP et de 2 plates-formes d'expression. Adapté de (Welz et Breaker, 2007).

3. Riborégulateur lysine

3.1 Historique et découverte

Chez *B. subtilis*, il est connu depuis les années 1970 que l'acide aminé lysine exerce un contrôle négatif sur l'expression de gènes impliqués dans sa propre synthèse (Grandgenett et Stahly, 1971; Rosner, 1975). Dans les années 1990, certaines mutations ont été identifiées dans la région 5′-UTR du gène *lysC* comme étant responsables d'une expression constitutive de l'enzyme aspatokinase II , impliquée dans la première étape de cette voie de synthèse de la lysine (Figure 10) (Lu *et al.,* 1991; von Hippel et Yager, 1992). Comparativement à ce qui était observé avec la souche de type sauvage, une fusion de la région 5′-UTR de ces mutants avec le gène *lacZ* ne présentait pas de diminution de l'expression de la β-galactosidase en présence d'un excès de l'acide aminé lysine. Déjà, on savait que la

lysine, le produit final de cette voie de biosynthèse, influençait négativement sa production et celle des produits intermédiaires de la voie de biosynthèse de l'aspartate, et que cette régulation impliquait la séquence en amont du gène *lysC* (Lu *et al*, 1991). La région 5′-UTR du même gène chez *E. coli* possède une certaine homologie de séquence avec celle de *B. subtilis* et des études menées en parallèles chez ces deux espèces arrivent aux mêmes conclusions (Patte *et al.*, 1998).

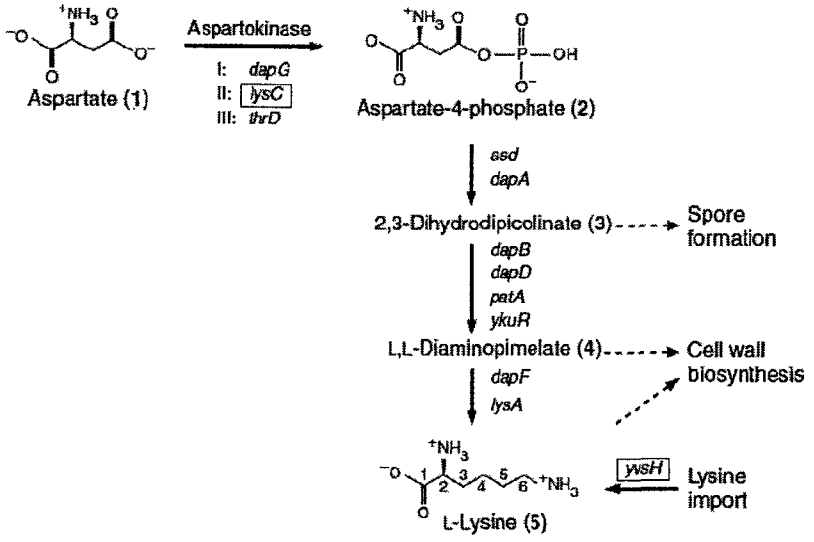


Figure 10. Synthèse de la lysine chez *B. subtilis*. Les gènes codant pour les enzymes impliquées sont indiqués. Les gènes encadrés (*lysC* et *yvsH*) sont sous le contrôle d'un riborégulateur lysine situé dans la région 5′-UTR de l'ARNm. Adapté de (Blount *et al.*, 2007).

En 1996, on s'aperçoit que l'expression du gène *lysC* chez *B. subtilis* est affectée par la lysine au niveau de sa transcription (Kochhar et Paulus, 1996). Les détails de ce mécanisme de rétrocontrôle sont encore mal connus, d'autant plus qu'aucun facteur protéique n'a été identifié comme étant impliqué dans cette régulation génétique (Liao et Hseu, 1998). Avec la découverte des riborégulateurs, certains groupes de recherche se sont intéressés de plus près aux séquences d'ARNm situées en amont de gènes impliqués dans la biosynthèse ou le transport de l'acide

aminé lysine (Grundy *et al.,* 2003; Rodionov *et al.,* 2003; Sudarsan *et al.,* 2003b). Les analyses bio-informatiques ont identifié chez plusieurs espèces des régions en amont de tels gènes ayant des structures secondaires similaires. Les études *in vitro* menées sur l'élément LYS (aussi nommé *L-box*) de *B. subtilis* indiquent que la structure de cette région de l'ARNm de *lysC* est réorganisée suite à une liaison spécifique de la lysine, et que cette réorganisation induit la terminaison de la transcription (Grundy *et al.,* 2003; Sudarsan *et al.,* 2003b). Ces études *in vitro*, faites en absence de protéine, venaient démontrer que le gène *lysC* de *B. subtilis* est placé sous le contrôle transcriptionnel d'un riborégulateur.

3.2 Implication du riborégulateur lysine

La voie de synthèse de la lysine est d'une importance capitale pour les bactéries. C'est en effet cette voie qui permet la synthèse d'acide diaminopimélique et de lysine, deux composés nécessaires aux parois bactériennes (Grundy *et al.,* 2003). Comme la régulation dépendante de la lysine était déjà observée pour les gènes *lysC* de *B. subtilis* et *E. coli*, des riborégulateurs lysine ont d'abord été recherchés et identifiés en amont de différents gènes impliqués dans cette voie de synthèse chez d'autres bactéries (Grundy *et al.,* 2003; Rodionov *et al.,* 2003; Sudarsan *et al.,* 2003b). Ils ont aussi été identifiés devant des gènes codant pour des transporteurs de lysine, comme *yvsH*, *lysT* et *lysW* (Jorth et Whiteley, 2010; Rodionov *et al.,* 2003). Il a été proposé qu'une troisième classe de gènes, impliqués dans l'utilisation de la lysine, pourraient être régulés par des riborégulateurs lysine. Il y aurait ainsi des riborégulateurs lysine qui contrôleraient un polycistron impliqué dans le catabolisme de la lysine chez *Thermoanaerobacter tengcongensis*, et le gène *FN1869* de *Fusobacterium nucleatum* qui coderait pour un exportateur de lysine (Rodionov *et al.,* 2003). Dans cette dernière classe, il y aurait augmentation de l'expression des gènes en présence de lysine. Si ces prédictions s'avèrent

véridiques, le riborégulateur lysine pourrait agir à la fois d'élément répresseur et activateur de l'expression des gènes. Les différentes plates-formes d'expression observées dans les séquences de riborégulateurs lysine sont soit des terminateurs intrinsèques de la transcription, soit des tiges qui séquestrent les signaux d'initiation de la traduction. Les deux riborégulateurs lysine *lysC* de *B. subtilis* et *E. coli*, régulent ainsi l'expression du gène au niveau de la transcription et de la traduction, respectivement.

Chez *B. subtilis*, deux riborégulateurs lysine ont été identifiés. Tel que mentionné précédemment, un premier a été identifié dans la région 5'-UTR du gène *lysC*. Comme il est impliqué dans la biosynthèse de la lysine, cette dernière inhibe sa propre synthèse en se liant au riborégulateur pour empêcher la transcription de ce gène (Grundy *et al.*, 2003; Sudarsan *et al.*, 2003b). Il a été prédit qu'un deuxième riborégulateur lysine se retrouve en amont du gène *yvsH*, codant pour un importateur de l'acide aminé lysine. L'inhibition de la transcription de ce gène par la lysine limite l'expression de cet importateur de manière à diminuer la prise de l'acide aminé dans le milieu externe par la bactérie.

3.3 Structure du riborégulateur

L'aptamère du riborégulateur lysine est formé d'une jonction à cinq voies, c'est-à-dire cinq hélices d'ARN (tiges P1 à P5) organisées autour d'un cœur (Figure 11). La région du cœur de l'aptamère est hautement conservée et il a été démontré qu'une réorganisation structurale de cette région a lieu en présence de L-lysine (Sudarsan *et al.*, 2003b). La structure tertiaire de l'aptamère (Garst *et al.*, 2008; Serganov *et al.*, 2008) se comporte comme une jonction à quatre voies, en adoptant une forme en «X» induite par l'empilements des tiges P1-P2 et P4-P5. La tige P3 est maintenue en place grâce une interaction boucle-boucle avec la tige P2 (Blouin et Lafontaine,

2007). Cette interaction est rendue possible par le repliement de la tige P2 qui se recourbe en deux points. Un de ces deux points d'inflexion correspond à un motif *K-turn* présent chez plusieurs riborégulateurs lysine et nécessaire pour la formation de l'interaction boucle-boucle (Blouin et Lafontaine, 2007). La tige P2 interagit aussi avec la boucle terminale de la tige P4, où deux adénosines de la boucle interagissent avec le sillon mineur de la tige P2 à la manière d'une interaction tétraboucle-récepteur, parfois nommée interaction hélice-boucle (Blouin *et al.,* 2010; Garst *et al.,* 2008; Serganov *et al.,* 2008). Au niveau de la tige P2, cette interaction est adjacente à un motif sarcine-ricine (boucle E) conservé dans les séquences de riborégulateurs lysine (Grundy *et al.,* 2003; Sudarsan *et al.,* 2003b). La tige P1, impliquée dans la transition entre les deux structures possibles, varie légèrement d'un riborégulateur à un autre au niveau de sa longueur, mais aussi au niveau de sa composition en paires de bases Watson-Crick. C'est cependant la tige P5 qui est la plus variable, en termes de longueur et de séquence (Rodionov *et al.,* 2003). Au niveau de la structure, cette tige P5 s'empile dans l'axe de P4 pour se retrouver juxtaposée à P1 dans la structure native du riborégulateur (Blouin *et al.,* 2010; Garst *et al.,* 2008; Serganov *et al.,* 2008).

Les plates-formes d'expression des riborégulateurs lysine oscillent entre les conformations «permissive» et «non-permissive» mutuellement exclusives. Ces deux conformations permettent ou non la transcription ou la traduction, en fonction du contexte. Dans le premier cas, la tige P1 est stabilisée par la présence du ligand pour permettre au terminateur de se former, alors que la partie 3′ de P1 et la partie 5′ du terminateur s'hybrident ensemble pour former un antiterminateur en absence de ligand (Figure 11A). Dans le cas d'un contrôle traductionnel, les changements structuraux sont très similaires, mais ils impliquent une tige séquestrant le RBS. Dans certaines séquences de riborégulateurs lysine, une autre tige-boucle (nommée P6) est retrouvée entre la tige P1 et le terminateur. Dans le cas du riborégulateur *lysC* de *B. subtilis*, la présence de cette tige favorise la formation de l'antiterminateur

(forme «permissive») (Blouin *et al.*, 2010).

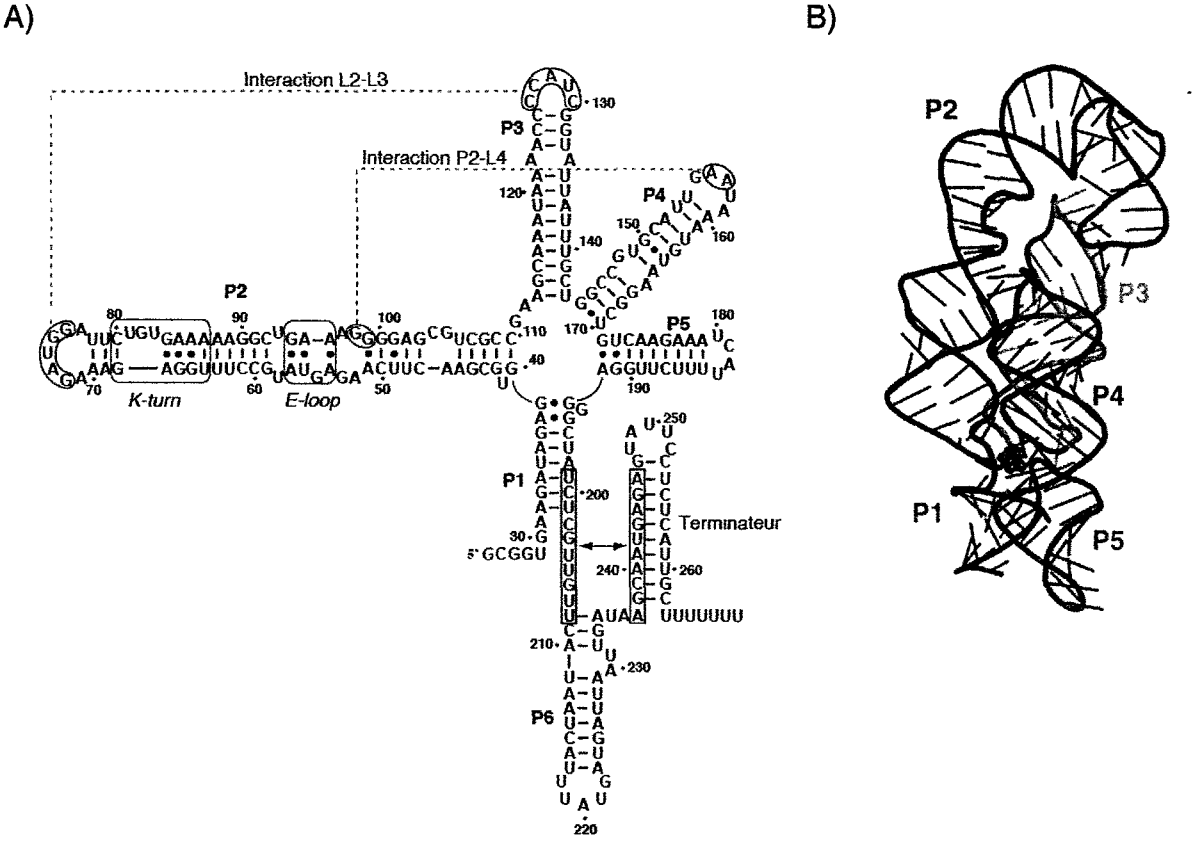


Figure 11. Structures secondaire et tertiaire du riborégulateur lysine. (A) Séquence et structure secondaire du riborégulateur lysine lysC de *B. subtilis.* Les motifs structuraux, les interactions tertiaires, les tiges P1 à P5 de l'aptamère ainsi que la tige P6 et le terminateur intrinsèque de la plate-forme d'expression sont indiqués. Lorsqu'hybridées, les 2 régions encadrées forment l'antiterminateur. Adapté de (Blouin *et al.*, 2010). **(B)** Structure tertiaire de l'aptamère du riborégulateur lysine de *Thermotoga maritima* (tiges P1 à P5). PDB 3D0U.

4. Transcription chez les procaryotes

La transcription constitue la première étape du processus de l'expression des gènes. Elle consiste à générer un brin d'ARN complémentaire à une matrice d'ADN. Ce mécanisme vital et ubiquitaire est accompli par des polymérases à ARN (ARNP) dépendantes de l'ADN et est soumis à une stricte régulation. La transcription est un mécanisme cyclique qui peut être décrit en trois grandes étapes: l'initiation, l'élongation et la terminaison.

4.1 La polymérase à ARN

Contrairement aux eucaryotes, une seule ARNP est retrouvée chez les procaryotes. L'enzyme est conservée dans l'évolution, tant au niveau de sa structure que de sa fonction (Darst, 2001), et est constituée de cinq sous-unités ($\alpha_2\beta\beta'\omega$) auxquelles se lie un facteur sigma (σ) pour la reconnaissance du promoteur. Les deux sous-unités α forment un homodimère impliqué dans la reconnaissance du promoteur et dans l'assemblage avec les sous-unités β et β', qui forment le centre catalytique. La sous-unité ω ne semble avoir qu'un rôle de chaperonne dans le repliement du domaine β' (Browning et Busby, 2004). L'ARNP est responsable de la polymérisation de l'ARN par l'extrémité 3'-OH du transcrit. Le mécanisme consiste à laisser entrer un nucléotide triphosphate (NTP) au site appelé «site A» de la polymérase, l'incorporer au transcrit via l'activité nucléotidyltransférase de l'enzyme en libérant un pyrophosphate, puis effectuer une translocation pour libérer à nouveau le site A (Nudler, 2009). L'efficacité et la spécificité de cette incorporation des NTPs sont assurées par deux éléments mobiles (nommées pont F (*bridge-helix*) et boucle G (*trigger-loop*)) situés dans la sous-unité β' (Borukhov et Nudler, 2008).

4.2 Initiation de la transcription

Pour qu'un gène soit transcrit par l'ARNP, celle-ci doit d'abord reconnaître son promoteur. Cette reconnaissance de la séquence promotrice est faite par un facteur σ (Burgess *et al.*, 1969) qui se lie aux cinq sous-unités de l'ARNP pour former l'holoenzyme. Un organisme procaryote possède généralement plusieurs facteurs σ qui sont exprimés en fonction des conditions ou du stade de croissance. La plupart font partie de la famille des facteurs σ^{70}, exprimés pour réguler le métabolisme basal des bactéries. L'utilisation de plusieurs facteurs σ capables de reconnaître des régions promotrices distinctes représente en fait le premier niveau de régulation de la transcription (Browning et Busby, 2004; Wösten, 1998).

La reconnaissance du promoteur par le facteur σ se fait par des séquences de l'ADN situées à environ 10 et 35 nucléotides en amont du site d'initiation de la transcription (+1), nommées respectivement boîtes -10 et -35 (Browning et Busby, 2004), qui sont principalement reconnues par les domaines 2.4 et 4.2 du facteur σ (Figure 12) (Paget et Helmann, 2003).

Cette reconnaissance du promoteur peut aussi être influencée, positivement ou négativement, par des facteurs de transcription. Dans le premier cas d'une activation de la transcription, l'activateur peut recruter l'ARNP en interagissant avec l'extrémité carboxy-terminale des sous-unités α (activation de classe I), ou avec le facteur σ (activation de classe II). Dans le cas d'une répression, un facteur de transcription peut agir par encombrement stérique en empêchant l'ARNP ou un activateur de se lier au promoteur. Finalement, des facteurs de transcription peuvent agir de manière positive ou négative par l'induction de changements structuraux qui modulent l'accessibilité des boîtes -10 et -35 (Browning et Busby, 2004).

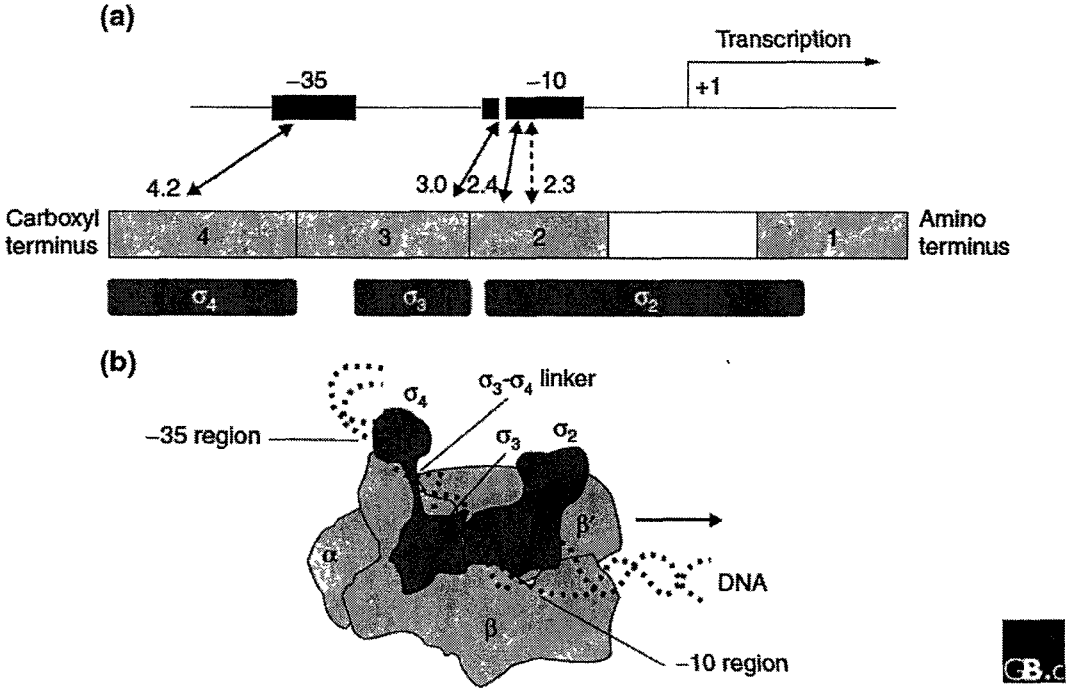


Figure 12. Interaction entre le facteur de σ^{70} *E. coli* et la région promotrice. (A) Les boîtes -10 et -35 du promoteur (haut) et les 4 domaines du facteur σ (bas) sont indiqués. **(B)** Holoenzyme ARNP où le facteur σ est en gris foncé. Les sous-unités de l'ARNP et les domaines du facteur σ sont indiqués. Tiré de (Paget et Helmann, 2003).

À la suite de cette liaison ADN-ARNP, le domaine 2.3 du facteur σ provoque la dissociation des deux brins d'ADN à proximité du site d'initiation de la transcription pour former un complexe «ouvert» (Helmann et Chamberlin, 1988; Juang et Helmann, 1994). La transcription d'un ARN est généralement précédée par la transcription de plusieurs petits transcrits dits «avortés» dont la longueur varie entre deux et quinze nucléotides avant que l'ARNP arrive à s'échapper du promoteur pour entrer dans la phase d'élongation (Borukhov et Nudler, 2008; Goldman *et al.*, 2009; Hsu *et al.*, 2003; Reppas *et al.*, 2006). L'efficacité de cette transition est influencée par la séquence promotrice et par le début de la séquence transcrite (Reppas *et al.*,

2006).

4.3 Élongation de la transcription

La liaison du facteur σ à l'ARNP est généralement transitoire et le facteur est relâché lorsque la polymérase passe en phase d'élongation (Reppas *et al.,* 2006). Lors de cette phase de la transcription, le complexe d'élongation ADN/ARNP/ARN est très stable et est influencé par divers mécanismes et facteurs de transcription, dont les principaux sont présentés dans cette section.

4.3.1 Pauses de l'ARNP

La vitesse d'élongation de la transcription par l'ARNP n'est pas constante. Cette vitesse d'élongation est entre autres modulée par des pauses de l'enzyme auxquelles on attribue différents rôles. Leur importance a notamment été démontrée dans certaines structures d'ARN, comme la RNase P, où des pauses de l'ARNP permettent le bon repliement du ribozyme (Pan *et al.,* 1999), et dans le cas de certains riborégulateurs, où une pause à un site précis ralentit la transcription pour permettre la liaison du ligand avant la transcription subséquente de structures alternatives (Lemay *et al.,* 2011; Wickiser *et al.,* 2005a; Wickiser *et al.,* 2005b). Ces pauses peuvent aussi permettre le recrutement de facteurs de transcription et constituent la première étape à un arrêt de la transcription. Finalement, comme les mécanismes de transcription et de traduction sont couplés chez les procaryotes, des pauses de l'ARNP lui permettent d'être synchrone avec la vitesse de traduction du ribosome. Cette synchronisation permet d'éviter que de longues portions (80 à 100 nucléotides) d'ARNm nouvellement transcrites soient dépourvues de ribosomes et puissent être liées par ρ qui induit alors la terminaison de la transcription (voir

section 4.4.1). La grande majorité des pauses de l'ARNP peuvent être répertoriées en 2 principaux types: les pauses de classe I, induites par la formation d'une tige-boucle, et les pauses de classe II qui sont causées par un mécanisme de retour en arrière (*backtracking*) de l'ARNP.

Les exemples les mieux caractérisés des pauses de classe I ont été identifiés dans les régions régulatrices des opérons histidine (*his*) et tryptophane (*trp*) (Artsimovitch et Landick, 2000; Chan *et al.*, 1997; Winkler et Yanofsky, 1981). Quant aux pauses de classe II, le cas le plus étudié est celui des sites *ops* (*Operon Polarity Supperssor*, voir section 4.3.2.2) où l'ARNP fait une pause pour permettre le recrutement du facteur de transcription RfaH.

Des pauses «ubiquitaires» de l'ARNP, qui ne correspondent à aucun de ces 2 types, sont cependant aussi observées (Kireeva et Kashlev, 2009; Neuman *et al.*, 2003). Le mécanisme général des pauses de l'ARNP tend de plus en plus à être expliqué par un changement allostérique du site actif de l'enzyme qui ralentit ou empêche l'incorporation d'un autre nucléotide (Landick, 2009). D'autres facteurs, comme la présence d'une tige-boucle ou l'action de retour en arrière de l'ARNP, ne feraient qu'augmenter la durée de telles pauses.

4.3.2 Principaux facteurs de transcription

4.3.2.1 NusA et NusG

NusA et NusG sont deux importants facteurs généraux d'élongation de la transcription qui interagissent directement avec l'ARNP (Li *et al.*, 1992; Schmidt et Chamberlin, 1984). Leur présence est ubiquitaire chez les bactéries, où ils semblent avoir des actions opposées.

NusA s'associe rapidement à l'ARNP lorsque le facteur σ est relâché et qu'elle transite en phase d'élongation (Mooney *et al.,* 2009). On attribue principalement à NusA un effet de ralentissement sur le taux de transcription de l'ARNP (Burns *et al.,* 1998; Lemay *et al.,* 2011), notamment en augmentant le temps des pauses à certains sites de classe I. Ce ralentissement peut entre autres résulter en un arrêt de la transcription aux terminateurs intrinsèques. De plus, NusA diminue l'efficacité de la terminaison de la transcription dépendante de ρ (Burns *et al.,* 1998) en réduisant l'activité adénosine triphosphatase (ATPase) de ρ (Schmidt et Chamberlin, 1984).

Contrairement à NusA, NusG agit positivement sur la transcription en accélérant la vitesse de l'ARNP (Bar-Nahum *et al.,* 2005) et en favorisant la «fuite» de l'ARNP d'un site de pause. Par contre, bien qu'il soit un activateur de la transcription, NusG peut interagir avec le facteur ρ par l'intermédiaire de son domaine carboxy-terminal (Li *et al.;* 1993). Cette interaction entre ρ et NusG favoriserait l'arrêt de la transcription à certains sites de terminaison dépendants de ρ (Burns et Richardson, 1995; Sullivan et Gottesman, 1992). L'association de NusG avec ρ serait aussi responsable du mécanisme d'antiterminaison de la transcription d'ARNr (Li *et al.,* 1993; Torres *et al.,* 2004). NusG est un facteur de transcription essentiel chez *E. coli*, mais pas chez *B. subtilis* où il augmente le temps de pause à certains sites spécifiques lors de la transcription de l'opéron *trpEDCFBA* (Yakhnin *et al.,* 2008). On attribue donc au facteur NusG un large éventail d'actions qui sont parfois fonctionnellement opposées.

4.3.2.2 RfaH

Le facteur de transcription RfaH est un paralogue de NusG. En ce sens, il influence positivement l'élongation de la transcription, diminue le taux de terminaison à certains sites indépendants de ρ et le temps des pauses de l'ARNP (Belogurov *et*

al., 2009). Mais à l'inverse de NusG, RfaH n'est pas un facteur général de la transcription essentiel à la bactérie. RfaH est spécifiquement recruté à certains sites de pause nommés suppresseurs de la polarité de l'opéron (sites *ops*) (Belogurov *et al.,* 2009) et est donc impliqué dans la transcription de certains opérons seulement, principalement ceux acquis par transferts horizontaux (Bailey *et al.,* 1997; Belogurov *et al.,* 2009). Les facteurs NusG et RfaH compétitionnent pour se lier à l'ARNP au même site. Cette compétition explique les différences observées dans la nature des gènes dont chacun contrôle l'élongation de la transcription (Belogurov *et al.,* 2009).

4.3.2.3 GreA et GreB

Les facteurs de transcriptions GreA et GreB favorisent l'activité endonucléolytique de l'ARNP lors de son mécanisme de retour en arrière. GreA et GreB sont respectivement responsables du clivage de petits (di et trinucléotides) et de longs fragments d'ARN libérés par cette marche arrière (Laptenko *et al.,* 2003; Shaevitz *et al.,* 2003). Ils favorisent ainsi l'élongation de la transcription en libérant plus rapidement l'ARNP des sites de pause induits par un retour en arrière de l'enzyme et en favorisant le passage de l'ARNP de la phase d'initiation vers la phase d'élongation. En induisant l'activité nucléolytique de l'ARNP, GreA et GreB sont aussi responsables du mécanisme de correction de l'enzyme qui permet le retrait des mauvaises incorporations nucléotidiques (Laptenko *et al.,* 2003).

4.3.2.4 Facteur σ

Bien que le facteur σ soit principalement impliqué dans la reconnaissance du promoteur pour l'initiation de la transcription, il semble y avoir des cas où il interagit avec le complexe d'élongation (Bar-Nahum et Nudler, 2001; Ring *et al.,* 1996). De

façon similaire à la protéine RfaH, le facteur σ peut induire des pauses de l'ARNP à des séquences spécifiques (Herbert *et al.*, 2010; Landick, 2006; Mooney et Landick, 2003). De plus, dans les rares occasions où le facteur σ reste associé au complexe d'élongation, il faciliterait la ré-initiation d'un nouveau cycle de transcription par l'ARNP (Bar-Nahum et Nudler, 2001).

4.4 Terminaison de la transcription

L'étape de terminaison de la transcription sert à dissocier le complexe ADN/ARNP/ ARN. Cette terminaison peut survenir à la fin d'un opéron pour recycler l'ARNP, au début d'un gène ou d'un opéron pour en contrôler l'expression (on parle ici d'atténuation), ou encore pour limiter la transcription d'un ARN qui n'est plus traduit (Nudler et Gottesman, 2002). Trois principaux mécanismes de terminaison de la transcription ont été caractérisés: la terminaison dépendante de ρ, la terminaison intrinsèque, ou indépendante de ρ, et la terminaison dépendante de la protéine Mfd (*Mutation Frequency Decline*).

4.4.1 Terminaison dépendante de ρ

Ce type de terminaison de la transcription implique le facteur ρ, une protéine hexamérique qui possède des activités ATPase et hélicase. Rho se lie à l'ARNm à des sites d'utilisation de ρ (*rut,* pour *Rho-utilization site*) qui sont riches en cytosines et dépourvues de structures stables et de ribosomes (Ciampi, 2006). Bien qu'on s'entende encore mal sur le mécanisme d'action précis de ρ (Figure 13), son mécanisme général peut se résumer en la liaison à l'ARN, sa translocation sur ce dernier en direction 5′-3′, puis en la dissociation du complexe d'élongation (Epshtein *et al.*, 2010; Roberts *et al.*, 2008).

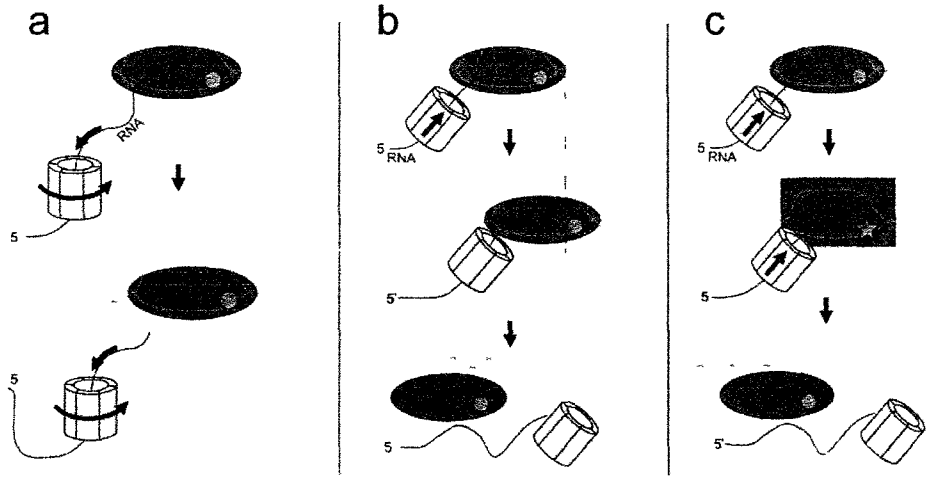


Figure 13. Modèles proposés pour la terminaison dépendante de ρ. (A) Rho (bleu) tire l'ARN (rouge) hors de l'ARNP (violet) pour défaire le complexe. **(B)** Rho se déplace sur l'ARN jusqu'à l'ARNP et la déloge. **(C)** Rho induit un changement allostérique dans le site actif (vert) de l'ARNP pour déstabiliser le complexe. Tiré de (Epshtein *et al*, 2010).

Même si ρ est principalement caractérisée comme un facteur de terminaison de la transcription, des groupes ont démontré que l'association de ρ avec l'ARNP, par l'intermédiaire des facteurs de transcription NusA et NusG, est beaucoup plus abondante qu'on le croyait, et ce, tout au long de l'élongation de la transcription (Epshtein *et al.,* 2010; Mooney *et al.,* 2009; Peters *et al.,* 2009). Bien qu'il soit impliqué dans la terminaison de la transcription de façon globale, ρ semble particulièrement important pour limiter la transcription de gènes acquis par transfert horizontal (Cardinale *et al.,* 2008).

4.4.2 Terminaison indépendante de ρ (intrinsèque)

À l'inverse de la terminaison de la transcription dépendante du facteur ρ, la terminaison intrinsèque ne requiert pas de facteur protéique. Cette terminaison indépendante de ρ est seulement contrôlée par des événements cinétiques et allostériques dictés par la séquence de l'ADN et la structure de l'ARNm. Ce type de terminaison survient à des séquences qui forment une tige-boucle riche en paires de bases G-C suivie d'une séquence d'environ sept à neuf uridines (Nudler et Gottesman, 2002). Cette séquence d'uridines induit d'abord une pause de l'ARNP qui permet la formation des conformations appropriées de la tige-boucle et des structures en amont nécessaires à la régulation (Larson *et al.*, 2008). La formation de la tige-boucle en amont de la séquence de poly-uridines déstabilise ensuite le complexe ADN/ARNP/ARN, dont la stabilité est déjà affaiblie par les multiples interactions A-U entre l'ADN et l'ARN (Gusarov et Nudler, 1999; Larson *et al.*, 2008) (Figure 14). La formation et l'efficacité des terminateurs intrinsèques sont fortement influencées par la séquence en amont qui peut entrer en compétition avec la formation de la tige-boucle terminatrice. Ce phénomène est notamment observé dans les cas de régulation de l'opéron *trp* et du contrôle transcriptionnel par les riborégulateurs.

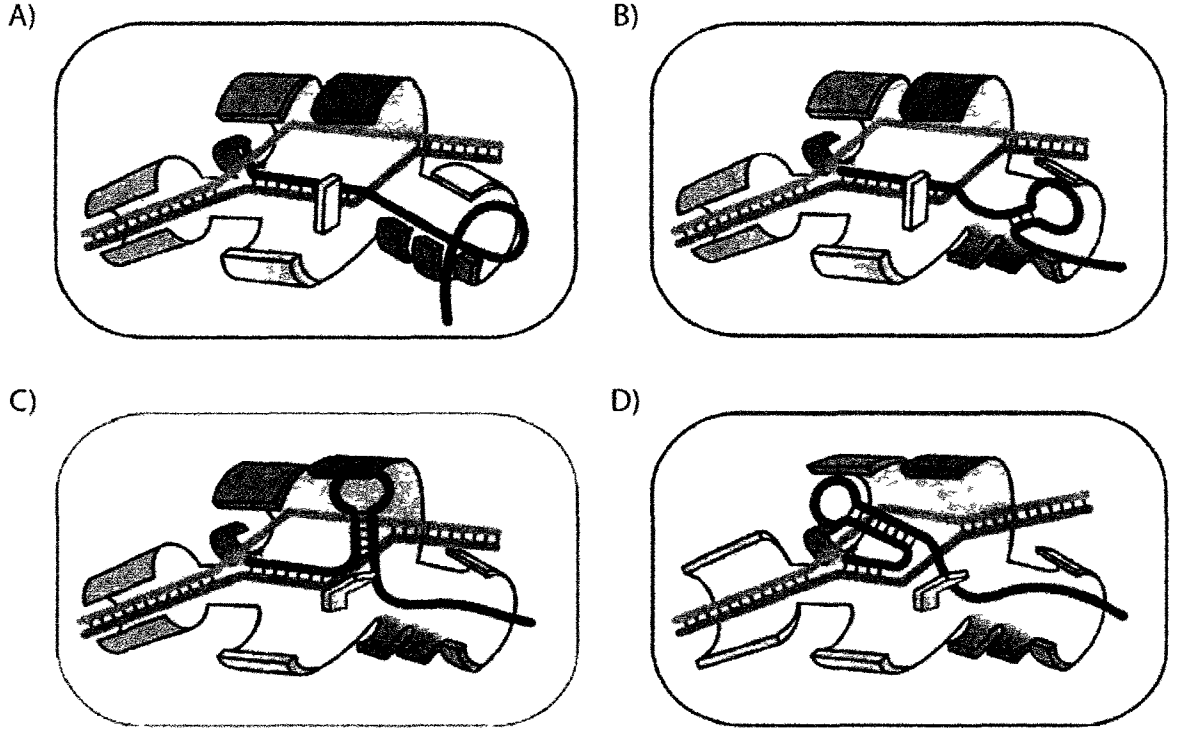


Figure 14. Représentation schématique de la terminaison intrinsèque de la transcription. **(A)** L'ARNP fait une pause après avoir transcrit la séquence d'urdines de l'ARN (brin rouge). **(B)** Formation de la tige-boucle en amont de la séquence d'uridines. **(C)** Intrusion de la tige-boucle dans l'ARNP. **(D)** La tige-boucle se forme complètement, déstabilise l'hybride ADN/ARN (vert/rouge) et induit un changement de conformation du site actif de l'ARNP. Adapté de (Epshtein *et al.*, 2007).

4.4.3 Terminaison dépendante de Mfd

Le facteur protéique Mfd a d'abord été identifié dans le mécanisme de réparation suite à des dommages causés par les rayons ultraviolets (Witkin, 1994). Mfd agit comme un agent de terminaison de la transcription en reconnaissant les ARNP bloquées, notamment suite à des lésions dans l'ADN, puis en induisant la

dissociation du complexe d'élongation (Roberts *et al.*, 2008). Il serait aussi le principal responsable du couplage entre la transcription et la réparation de l'ADN en permettant le recrutement de protéines impliquées dans le mécanisme de réparation, comme UvrA (Park *et al.*, 2002).

4.5 Transcription de type *single round*

L'utilisation d'un système minimal de transcription est requis pour pouvoir disséquer le mécanisme de terminaison intrinsèque et quantifier l'efficacité d'une séquence terminatrice. La réaction de transcription *in vitro* de type *single round* se fait à partir d'une séquence d'ADN qui contient un promoteur, la région régulatrice avec le terminateur (comme un riborégulateur) et une séquence ajoutée en aval du terminateur qui permet de différencier par leur longueur les deux principaux produits de transcription obtenus (Artsimovitch et Henkin, 2009). À partir de la position d'initiation de la transcription (+1), la matrice d'ADN requiert une séquence d'environ 15 à 20 paires de bases où seulement trois des quatre nucléotides sont représentés, de manière à bloquer l'élongation en absence du quatrième nucléotide dans la réaction. On arrive ainsi à initier la transcription, puis bloquer et synchroniser tous les complexes (Mironov *et al.*, 2009; Nudler et Gusarov, 2003; Nudler *et al.*, 2003). L'élongation de la transcription se poursuit par l'ajout de tous les nucléotides, en plus de rifampicine ou d'héparine pour prévenir la ré-initiation d'un cycle de transcription par l'ARNP en s'y liant (Tavormina *et al.*, 1996; Walter *et al.*, 1967). L'utilisation d'un cycle unique de transcription prévient ainsi l'obtention de produits intermédiaires pour mieux quantifier la différence entre les deux principaux produits, c'est-à-dire le transcrit terminé et le transcrit pleine longueur. Les conditions expérimentales sont spécifiquement modifiées par l'ajout d'un ligand ou d'un facteur protéique dont on veut observer l'effet sur la terminaison ou l'antiterminaison de la transcription. Ce type de transcription *in vitro* a été largement utilisé pour étudier les

pauses de l'ARNP (Lemay *et al.,* 2011; Sevostyanova *et al.,* 2008), l'effet de facteurs de transcription (Burns *et al.,* 1998; Herbert *et al.,* 2010; Sevostyanova et Artsimovitch, 2010; Yakhnin *et al.,* 2008) et les mécanismes de terminaison prématurée de la transcription par les «*T-box*» (Grundy et Henkin, 2004; Grundy *et al.,* 2002) et les riborégulateurs (Blouin *et al.,* 2010; Blouin et Lafontaine, 2007; Epshtein *et al.,* 2003; Heppell *et al.,* 2011; Mironov *et al.,* 2009; Wickiser *et al.,* 2005b; Winkler *et al.,* 2002a).

5. Objectifs du projet de recherche

Trois publications portant sur le riborégulateur lysine ont vu le jour dans le courant de l'année 2003 (Grundy *et al.,* 2003; Rodionov *et al.,* 2003; Sudarsan *et al.,* 2003b). Suite à la récente découverte des riborégulateurs (Nahvi *et al.,* 2002), ces publications sur le riborégulateur lysine faisaient entre autres état de la conservation de la structure secondaire de l'élément régulateur par des analyses phylogénétiques, tout en proposant une structure secondaire qui suggérait l'interaction entre les boucles terminales des tiges P2 et P3. Les groupes de Breaker et Henkin avaient aussi démontré que la lysine induisait une terminaison prématurée de la transcription du gène *lysC* chez *B. subtilis*, tout en induisant des changements structuraux dans la région 5'-UTR de cet ARNm (Grundy *et al.,* 2003; Sudarsan *et al.,* 2003b). Il y avait cependant très peu d'évidences expérimentales pour confirmer la structure prédite, et encore moins pour établir la relation entre celle-ci et la fonction régulatrice du riborégulateur.

Nous avions ainsi émis comme hypothèse de départ que la fonction de régulation de la transcription du riborégulateur lysine *lysC* de *B. subtilis* dépendait directement de sa structure. La première étape du projet a consisté à vérifier l'existence d'un *K-turn* et d'une interaction boucle-boucle, deux motifs dont la présence dans le

riborégulateur avait été prédite par les études phylogénétiques, en plus d'en évaluer l'importance fonctionnelle par des essais de transcription *in vitro* de type *single round*. Ces résultats ont fait l'objet d'une publication (Blouin et Lafontaine, 2007) présentée dans le chapitre 1.

Dans un deuxième chapitre (Blouin *et al.,* 2010), des essais de SHAPE (*Selective 2'-Hydroxyl Acylation analyzed by Primer Extension*), de fluorescence et d'électrophorèses en gels natifs nous ont permis d'observer l'impact des différents motifs et interactions sur la structure générale de l'aptamère et sur la conformation du site de liaison. Nous avons aussi confirmé l'importance du maintient de ces structures en *single round*, en plus de mettre en évidence l'effet de structures distantes localisées dans la plate-forme d'expression. Fait intéressant, cette publication nous a aussi permis de démontrer expérimentalement l'absence de liaison du ligand à un mutant G39C, préalablement identifié comme résistant à certains antibiotiques (Lu *et al.,* 1991; Patte *et al.,* 1998; Vold *et al.,* 1975) et déjà utilisé comme contrôle négatif (Blouin et Lafontaine, 2007; Sudarsan *et al.,* 2003b). La mise au point de techniques comme les transcriptions *single round* et le SHAPE a notamment contribué à exposer la forte relation existant entre la structure et la fonction des riborégulateurs de différentes familles et elles se sont avérées très profitables pour d'autres publications auxquelles j'ai participé, mais qui ne seront pas présentées dans cette thèse (Heppell *et al.,* 2011; Lemay *et al.,* 2011).

CHAPITRE 1

L'APTAMÈRE DU RIBORÉGULATEUR LYSINE CONTIENT UNE INTERACTION BOUCLE-BOUCLE ET UN MOTIF *K-TURN* NÉCESSAIRES À LA RÉGULATION GÉNÉTIQUE PAR LE RIBORÉGULATEUR

1.1 Présentation de l'article et contribution

L'influence de la structure de l'ARN est cruciale dans la régulation par les riborégulateurs. Cependant, les premières évidences menant à l'identification d'une famille de riborégulateurs sont presque toujours fournies par des analyses de bio-informatique. Il y a néanmoins une réelle nécessité de recourir à des méthodes biochimiques plus expérimentales pour vérifier de telles prédictions, en étudier la fonction, ou pour approfondir et faire ressortir des différences entre plusieurs riborégulateurs d'une même famille. C'est principalement dans cette optique que s'inscrit ce premier article.

Comme le motif *K-turn* prédit dans le riborégulateur lysine différait du consensus établi en 2001 (Klein *et al.,* 2001), nous avons vérifié par gels natifs qu'il induisait un angle prononcé dans la double-hélice d'ARN et qu'il pouvait lier la protéine ribosomale L7Ae, deux principales caractéristiques des *K-turns.* Ces résultats, qui représentent la première démonstration de la présence de ce motif dans les riborégulateurs, nous ont amené à redéfinir un consensus à partir de *K-turns* semblables retrouvés dans d'autres séquences de riborégulateurs lysine. Des essais de RNase T1 ont aussi démontré que l'interaction boucle-boucle entre les tiges P2 et P3 pouvait se former en présence d'ions magnésium et que la présence du motif *K-turn* est nécessaire à la formation de cette interaction. Ces observations

ont été corroborées par les essais de transcriptions *single round*, qui ont aussi mis en évidence l'importance fonctionnelle de tels motifs. Finalement, l'utilisation d'une 2-aminopurine (2AP) incorporée à la séquence de l'aptamère a permis d'observer une déficience dans le repliement du coeur pour des mutants de ces motifs, faisant ainsi ressortir l'importante relation qui existe entre la structure et la fonction du riborégulateur.

J'ai réalisé toutes les expériences de cet article et participé à la rédaction du manuscrit avec le Pr Daniel Lafontaine.

1.2 Manuscrit

Blouin, S. et Lafontaine, D.A. (2007). A loop-loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. RNA *13*, 1256-1267.

A loop-loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control

Simon Blouin and Daniel A. Lafontaine*

Département de biologie, Faculté des sciences,
Université de Sherbrooke,
Sherbrooke, Québec,
Canada, J1K 2R1

Running title: A loop-loop interaction in the lysine aptamer

Keywords: Riboswitch; loop-loop interaction; kink-turn; 2-aminopurine fluorescence

*To whom correspondence should be addressed

Contact information:
Tel. 819-821-8000, ext: 65011,
Fax. 819-821-8049,
email: Daniel.Lafontaine@USherbrooke.ca

ABSTRACT

The lysine riboswitch is associated to the *lysC* gene in *Bacillus subtilis* and the binding of lysine modulates the RNA structure to allow the formation of an intrinsic terminator presumably involved in transcription attenuation. The complex secondary structure of the lysine riboswitch aptamer is organized around a five-way junction that undergoes structural changes upon ligand binding. Using single-round transcription assays, we show that a loop-loop interaction is important for lysine-induced termination of transcription. Moreover, upon close inspection of the secondary structure, we find that an unconventional kink-turn motif is present in one of the stems participating in the loop-loop interaction. We show that the K-turn adopts a pronounced kink and that it binds the K-turn-binding protein L7Ae of *Archaeoglobus fulgidus* in the low nanomolar range. The functional importance of this K-turn motif is revealed from single-round transcription assays which show its importance for efficient transcription termination. This motif is essential for the loop-loop interaction, and consequently, for lysine binding. Taken together, our results depict for the first time the importance of a K-turn-dependent loop-loop interaction for the transcription regulation of a lysine riboswitch.

INTRODUCTION

Riboswitches are highly structured genetic control elements that are located in the 5′-untranslated region (5′-UTR) of several messenger RNAs of prokaryotes and also in eukaryotes (Kubodera et al., 2003; Lai, 2003; Sudarsan et al., 2003a; Mandal & Breaker, 2004). Riboswitches are composed of two modular domains, an aptamer and an expression platform, which respectively senses a specific metabolite and controls gene expression. In all cases studied, riboswitches perform their regulation via ligand-dependent structural changes, and because they can operate *in vitro* in

absence of protein cofactor, it has been proposed that they can regulate gene expression without the assistance of any protein (reviewed in (Grundy & Henkin, 2004; Nudler & Mironov, 2004; Winkler, 2005)), making them potential remnants of ancient gene regulatory systems. A large number of different riboswitches has been described (reviewed in (Breaker, 2006)), and a recent bioinformatic study has provided evidence that additional riboswitches still await to be discovered given that numerous additional RNA motifs sharing characteristics with riboswitches were identified (Barrick et al., 2004).

In most bacteria, the amino acid lysine derives from aspartate via the diaminopimelate (DAP) pathway (Patte, 1994). This pathway is of special interest for antibacterial drugs given that the absence of DAP in mammalian cells allows for the use of the DAP biosynthetic genes as bacteria-specific drug targets (Hutton et al., 2003). The first two stages of the DAP pathway are common for the biosynthesis of amino acids of the aspartate family, namely lysine, threonine and methionine, and are catalyzed by aspartokinase and aspartate semialdehyde dehydrogenase. *Bacillus subtilis* has three aspartokinases: DAP-inhibited DapG, lysine-inhibited aspartokinase-II LysC, and aspartokinase-III, that is inhibited by simultaneous addition of threonine and lysine and encoded by the *yclM* gene. Although most genes of lysine biosynthesis are repressed by lysine, little is currently known about their regulation mechanisms (Patte, 1994; Kochhar & Paulus, 1996; Patte et al., 1998; Mader et al., 2002). Mutations in leader regions of *lysC* genes *E. coli* and *B. subtilis* have been shown to release repression of gene expression by lysine (Lu et al., 1992; Patte et al., 1998). Based on the identification of a short transcript corresponding to the *B. subtilis lysC* leader region, it has been proposed that lysine controls transcription and translation by a mechanism in which no protein factor serves as the genetic regulator (Kochhar & Paulus, 1996; Liao & Hseu, 1998; Rodionov et al., 2003). Instead, a group of conserved elements was identified upstream of several lysine biosynthesis genes and was shown to be important for

lysine-dependent transcription termination *in vitro* (Grundy et al., 2003; Sudarsan et al., 2003b). Through phylogenetic analyses, this motif was designated as the L box or the LYS element and was found to reside upstream of additional lysine related genes (Rodionov et al., 2004). The binding of lysine to the *lysC* RNA leader region was shown to cause structural changes consistent with the hypothesis that lysine binding promotes transcription attenuation by preventing the formation of an antiterminator stem through a riboswitch mechanism (Grundy et al., 2003; Sudarsan et al., 2003b).

The secondary structure of the L box aptamer domain of the *B. subtilis* lysine riboswitch (Fig. 1) is based on conserved primary sequence elements (Grundy et al., 2003; Rodionov et al., 2003; Sudarsan et al., 2003b). The core of the L box is organized around a five-way junction exhibiting a 2HS1HS4HS4HS3 arrangement (using the IUB nomenclature of junctions (Lilley et al., 1996)) and it has been shown by in-line probing assays to be reorganized upon lysine binding (Sudarsan et al., 2003b). The binding of lysine stabilizes the P1 stem which can be considered as an anti-antiterminator given that its formation prevents the adoption of an antiterminator stem (Grundy et al., 2003; Rodionov et al., 2003; Sudarsan et al., 2003b). Helices P2-P5 are most probably involved in the formation of the correct structure of the aptamer and a loop E-like motif or a S turn motif has been speculated to reside in the middle region of the P2 helix (Grundy et al., 2003; Sudarsan et al., 2003b). In addition, covariation analysis has suggested that a loop-loop interaction occurs between the P2 and P3 stem-loops (Fig. 1) (Grundy et al., 2003; Rodionov et al., 2003). Similar long-range interactions are found in other riboswitch systems such as the adenine riboswitch where we have shown that a loop-loop interaction is very important for ligand binding (Lemay et al., 2006). An additional internal loop in the P2 helix that matches the GA motif found in both the T box and S box leader RNAs (Winkler et al., 2001) has been predicted to correspond to the K-turn motif (Klein et al., 2001) which introduces a severe kink in the RNA backbone (Grundy et al., 2003).

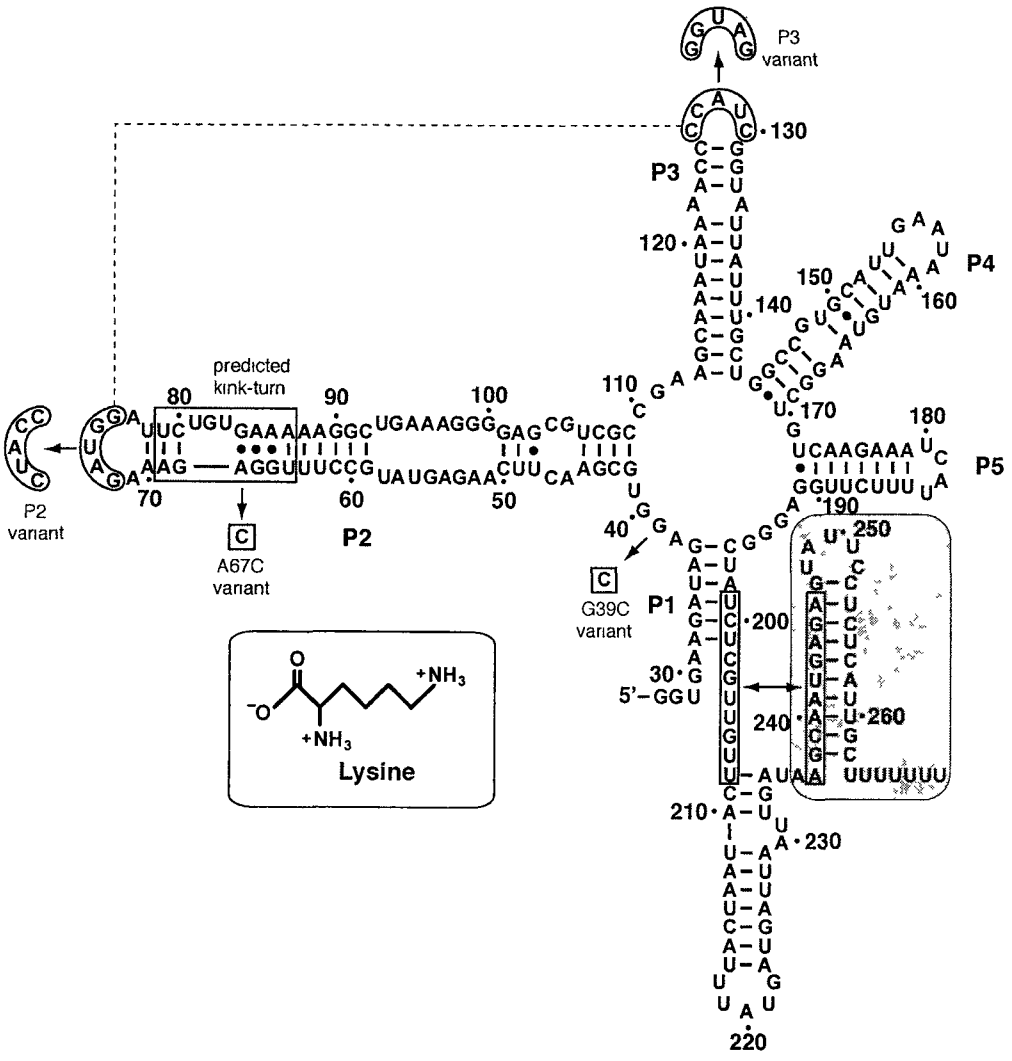


FIGURE 1. Sequence and secondary structure of the lysC lysine riboswitch of *Bacillus subtilis* (Epshtein et al., 2003; Grundy et al., 2003; Sudarsan et al., 2003a). The represented secondary structure is shown with the transcription terminator structure (shaded region). Regions involved in the formation of the antiterminator helix are boxed and indicated by an arrow. The loop-loop interaction is indicated by a dotted line. The K-turn is shown in the P2 helical domain in a box. The mutations studied in this work are represented by boxed sequences. The chemical structure of lysine is shown in the insert.

Here, a mutational analysis shows that a loop-loop interaction between the stem-loops P2 and P3 is important for the biological function of the lysine riboswitch. Upon examination of the P2 helical domain, we identify an unconventional kink-turn motif and show its importance for lysine-induced termination of transcription, most probably by allowing formation of the loop-loop interaction. By introducing the fluorescent adenine structural analog 2-aminopurine (2AP) in the core of the aptamer, we observe an important change induced by magnesium ions which is dependent upon the correct formation of the loop-loop interaction, indicating that distal structural elements are important for the core folding of the lysine aptamer. Our results provide the first experimental evidence that a loop-loop interaction in the *B. subtilis* lysine riboswitch has a major role in the riboswitch biological function, and strongly suggest that an unconventional K-turn motif is very important for the adoption of the loop-loop interaction.

RESULTS

Lysine promotes premature termination of transcription

We first analyzed transcription attenuation by the lysine riboswitch using single-round *in vitro* transcription assays. This technique has been previously used to study a variety of riboswitches (Epshtein et al., 2003; McDaniel et al., 2003; Mandal et al., 2004; McDaniel et al., 2005; Sudarsan et al., 2006), and particularly the one responding to lysine (Grundy et al., 2003; Sudarsan et al., 2003b; Blount et al., 2007). A DNA template containing the *B. subtilis glyQS* promoter upstream the lysine riboswitch and a 95 nucleotides sequence downstream the terminator stem was generated for our studies (Fig. 2A). The conformation of the riboswitch in absence of ligand should prevent the formation of the terminator stem, allowing the RNA polymerase to transcribe a complete mRNA (readthrough product). However, lysine

binding to the aptamer is expected to stabilize the anti-antiterminator P1 stem and to allow the terminator stem to be formed, ultimately leading to premature transcription termination (terminated product).

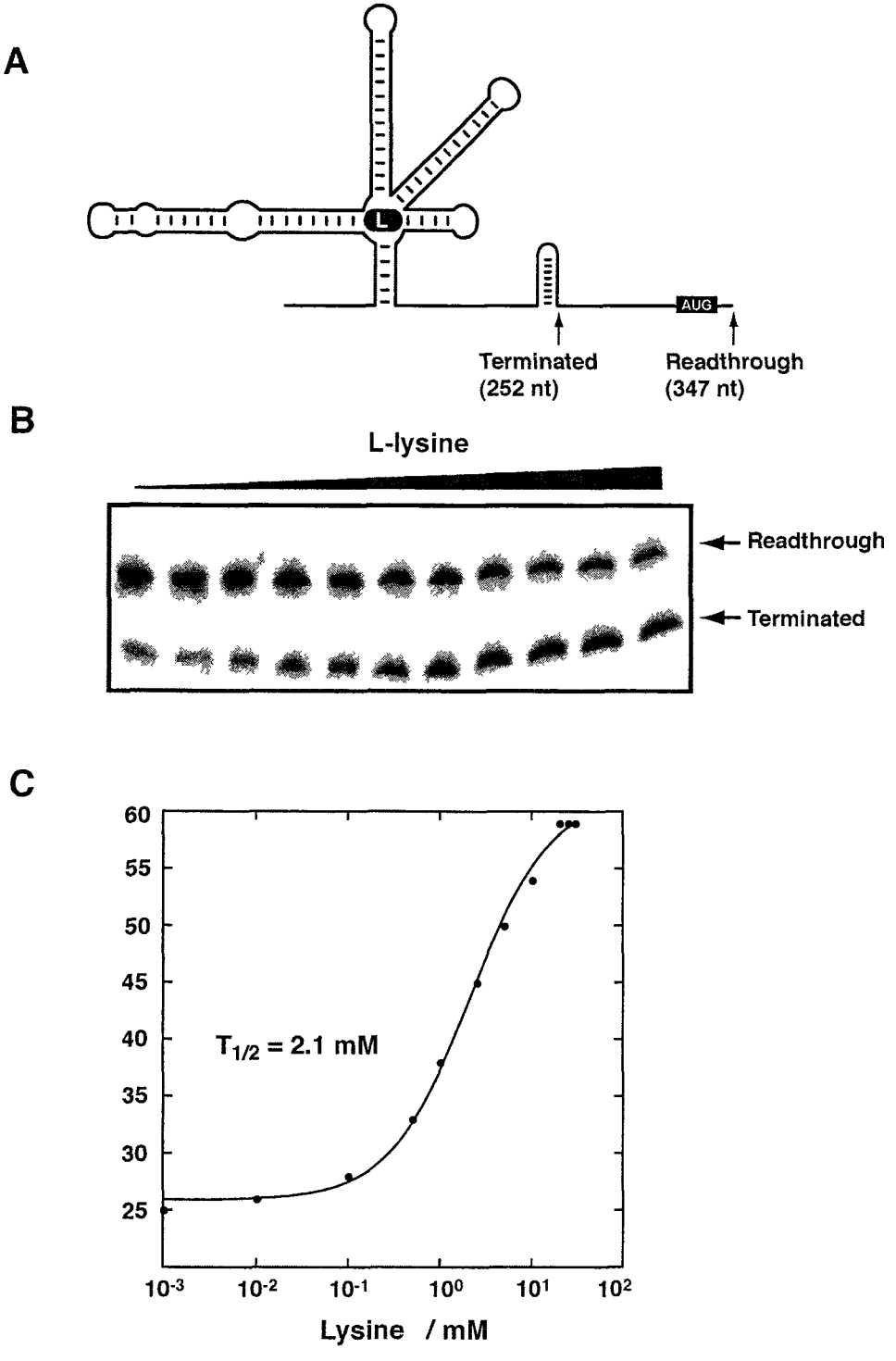


FIGURE 2. Lysine-induced riboswitch transcription termination. **A:** Schematic showing constructs used in these experiments. The terminated (252 nt) and readthrough (347 nt) transcription products are indicated by arrows. Lysine is indicated by a rounded rectangle. While the premature termination product is expected to occur in presence of ligand, transcription readthrough is expected in its absence. **B:** Single-round transcriptions performed in presence of 1 μM, 10 μM, 100 μM, 500 μM, 1 mM, 2.5 mM, 5 mM, 10 mM, 20 mM, 25 mM and 30 mM L-lysine, and resolved on 5% denaturing polyacrylamide gels. Readthrough and terminated products are indicated on the right. **C:** Percentage of termination (%T) is plotted as a function of lysine concentration. The line shows a two-state model from which a $T_{1/2}$ value = 2.1 mM lysine was calculated.

Single-round *in vitro* transcription reactions were carried out by using *E. coli* RNAP in presence of increasing concentrations of lysine. In presence of 1 μM ligand, low transcription termination is observed with ~25% of terminated products (Fig. 2B). However, the addition of L-lysine leads to an increase of transcription termination up to a value of ~60% in presence of 10 mM lysine (Fig. 2B). The ratio of terminated product (%T) was calculated for each reaction and plotted as a function of ligand concentration (Fig. 2C). Resulting data were well fitted by a simple two-state model (see Materials and Methods) and the lysine concentration required to obtain half of the change in transcription termination (defined as $T_{1/2}$) is 2.1 ± 0.2 mM. Similar results were observed in a recent study where a value of $T_{1/2}$ ~5 mM was obtained (Blount et al., 2007). Given that $T_{1/2}$ values are highly dependent on experimental protocols, the small difference between our value and the one reported most probably results from variations in transcription conditions (e.g., ribonucleotide concentrations, S Blouin and DA Lafontaine, personal observation).

The loop-loop interaction is important for transcription termination

It has been previously proposed that a loop-loop interaction takes place between loops P2 and P3 (Fig. 1;(Grundy et al., 2003)). To determine whether the loop-loop motif is important for ligand binding, single-round *in vitro* transcriptions were carried out by disrupting the putative interaction (Fig. 3). To assess the relative importance of the loop-loop interaction, a single-point mutation (G39C), which has previously been shown to cause *lysC* derepression (Vold et al., 1975; Lu et al., 1992; Sudarsan et al., 2003b), was used in our transcription assay as a negative control. As expected, upon introduction of the G39C mutation, the lysine-induced transcription attenuation mechanism was severely affected, yielding only ~20% of termination in the absence and presence of lysine (Fig. 3). Similar results have been previously obtained by Breaker and co-workers (Sudarsan et al., 2003b), indicating the importance of G39 in the riboswitch regulation control.

Two riboswitch variants were made in which each of the P2 or P3 loop nucleotides involved in the proposed loop-loop interaction was changed for its respective counterpart (P2 and P3 variants, respectively, Fig. 3). In absence of lysine, both riboswitch variants yielded a reduced transcription termination efficiency (~14%). Addition of lysine to transcription reactions resulted in termination efficiencies of ~30% and ~26% for the P2 and P3 variants, respectively. In contrast to what is observed for the G39C variant, the presence of lysine results in an increase of transcription termination, albeit to a reduced level compared to the natural sequence. However, when both P2 and P3 loop mutations are introduced simultaneously, the lysine-induced transcription termination activity is restored near to the wild-type level (Fig. 3). The observed restoration of the riboswitch activity suggests that the loop-loop interaction is required for efficient transcriptional control, and that the interaction takes place provided that base pairing is maintained.

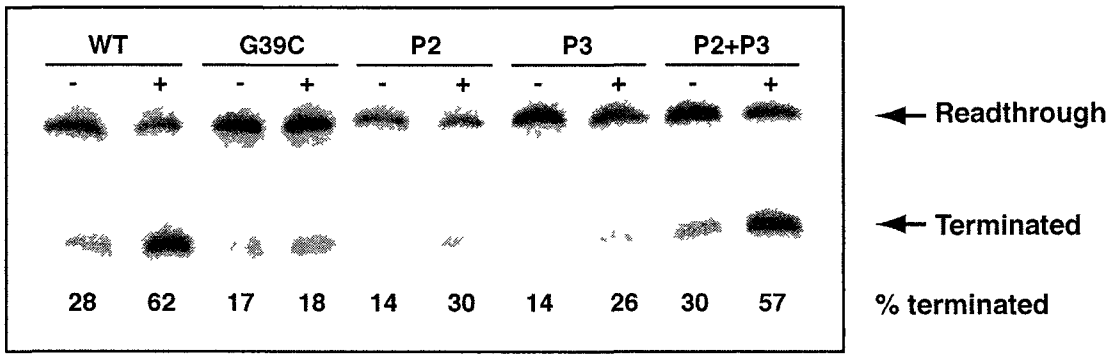


FIGURE 3. A loop-loop interaction is important for transcription termination. Single round transcriptions were carried out in absence (-) and in presence (+) of 5 mM lysine for each riboswitch variant. Readthrough and terminated products were separated by gel electrophoresis and are indicated on the right. Reactions were performed for the natural sequence (WT), the G39C core variant, P2 and P3 loop-loop-deficient variants, and for the P2+P3 variant. The percentage of terminated product for each reaction is indicated below the gel.

An unconventional K-turn motif is present in the L box domain

Upon visual inspection of the secondary structure of the *B. subtilis lysC* aptamer domain, it can be observed that a region of the P2 arm may fold as a K-turn motif (Fig. 1). The K-turn structure is known to introduce a severe bend into the helical axis where it subtends an included angle of 60° (Klein et al., 2001). This motif is characterized by the presence of two tandem G•A pairs, an asymmetric loop and a nucleotide completely exposed to the solvent (Klein et al., 2001). The K-turn is organized around a central coaxial base stack that makes an important hydrophobic core for the structure of the motif. The overall structure of the K-turn is highly conserved and occurs in several locations in the 23S rRNA *Haloarcula marismortui*

(Ban et al., 2000) and in the 16S rRNA of *Thermus thermophilus* (Schluenzen et al., 2000). K-turn motifs are very often bound by protein factors making them important recognition elements. The *B. subtilis lysC* and the consensus K-turn motifs shown in Figure 4A indicate that the *lysC* K-turn sequence does not match the consensus sequence. For instance, the single-stranded sequence of *lysC* (UGU) does not correspond to the consensus (RNN), and some conserved base pairs also differ. Because of the observed differences, we decided to investigate if the putative *lysC* K-turn motif adopts a structure similar to conventional K-turn elements.

The highly kinked conformation of the K-turn has previously been studied using gel electrophoresis by Lilley and co-workers where it was shown that K-turn containing molecules exhibit severely retarded electrophoretic mobility (Goody et al., 2004). The gel electrophoresis technique is very sensitive to study global conformational changes in nucleic acids (Marini et al., 1982; Diekmann & Wang, 1985; Bhattacharyya & Lilley, 1989; Hsieh & Griffith, 1989; Bhattacharyya et al., 1990; Tang & Draper, 1990; Luebke & Tinoco, 1996; Lafontaine et al., 2001, 2002a; Goody et al., 2004; Lemay et al., 2006; Lemay & Lafontaine, 2007), and provides a simple yet powerful means to analyze the K-turn motif in solution. To analyze the extent of bending, a set of comparison duplexes was constructed by placing a series of oligoadenine bulges in the same sequence context, to generate RNA duplexes having the same length as the K-turn-containing molecules. Such oligoadenine bulges cause axial bending, the magnitude of which depends on the number of unpaired bases in the bulge (Bhattacharyya et al., 1990; Tang & Draper, 1990; Gohlke et al., 1994; Lilley, 1995; Zacharias & Hagerman, 1995; Luebke & Tinoco, 1996). The relative mobilities of radioactively labeled RNA species were compared in a native polyacrylamide gel and electrophoresed in presence of 10 mM magnesium ions. The series of oligoadenine bulged species exhibits the expected progressive reduction of mobility with bulge size, which is known to result from an axial kinking that becomes greater as the number of unpaired bases increases (Fig. 4B) (Gohlke

et al., 1994; Zacharias & Hagerman, 1995). The *lysC* K-turn species, when introduced in a corresponding RNA duplex, displays a markedly retarded mobility which is similar to an RNA species having a five adenine bulge (Fig. 4B). This is in agreement with crystal structures of RNA molecules containing conventional K-turn structures (Ban et al., 2000; Schluenzen et al., 2000; Klein et al., 2001) and with a previous gel electrophoresis analysis performed on the Kt-7 K-turn motif where a highly kinked conformation was deduced (Goody et al., 2004). According to transient electric birefringence (TEB) experiments, an RNA species having a five adenine bulge displays a bend angle of ~ 70 ± 3 degrees in presence of 2 mM magnesium (Zacharias & Hagerman, 1995). This value is very close to the observed subtended angle in crystal structures (~60°) (Klein et al., 2001) and is consistent to the *lysC* K-turn motif adopting a structure very similar to a conventional K-turn motif. The effect of sequence changes in the G•A mismatches on the structure of the K-turn has been previously studied whereby the replacement of A67 by a cytosine highly perturbs the formation of the helical bend (Goody et al., 2004). Introducing a similar mutation in the context of the lysC K-turn (Kt' variant) clearly affected the migration pattern (Fig. 4B). The loss of the retarded migration is consistent with the inability of the mutated K-turn motif to adopt a highly kinked structure, as found in the Kt-7 motif (Goody et al., 2004). Thus, we conclude that a K-turn motif is present in the *lysC* riboswitch aptamer and that its folding properties are very similar to a conventional K-turn motif.

Despite variations in their sequences, K-turn structures found in rRNA all possess a similar three-dimensional structure where the average backbone r.m.s.d. is 1.7 Å (Klein et al., 2001). Most of these K-turn sequences make significant interactions with ribosomal proteins of unrelated structures (Ban et al., 2000; Klein et al., 2001). To further characterize the *lysC* K-turn motif, we studied its protein binding properties by using the L7Ae protein, which has previously been shown to bind the Kt-7 K-turn motif (Turner et al., 2005). We examined the binding of L7Ae to the complete riboswitch aptamer domain using gel electrophoretic retardation analysis (Fig. 4C).

For this purpose we prepared a radioactively ^{32}P-labeled *lysC* aptamer molecule by *in vitro* transcription and incubated it with increasing concentrations of L7Ae. Our analysis by electrophoresis shows that two species are apparent and correspond to the free and to the L7Ae-aptamer complex. The radioactive bands were quantified and analyzed, and the data fitted very well to a simple model (see Materials and Methods) that assumes the non-cooperative binding of L7Ae protein, giving a dissociation constant (K_d) of 8.7 ± 2.5 nM. This value is in good agreement with recent studies performed on the box C/D snoRNA and the Kt-7 kink-turn motif, for which K_d values near the low nM range were found (Kuhn et al., 2002; Turner et al., 2005). The formation of the complex was also found to be specific given that the A67C mutation completely abolishes the formation of a complex with L7Ae (data not shown).

We used the WebLogo application (Crooks et al., 2004) to generate a graphical representation of the K-turn motif found in various L box aptamers obtained through the *Rfam* database (Griffiths-Jones et al., 2003). Among the 98 available sequences, several of them did not exhibit a recognizable K-turn motif, leading us to establish specific criteria that are found in most K-turn sequences. Specifically, flanking Watson-Crick base pairs, two tandem G•A pairs and a stretch of three unpaired nucleotides (NNN) were used as selection criteria. Moreover, to avoid redundancy between variants, sequences that displayed more than 95% of homology over the entire aptamer sequence were included only once, thus yielding a total of 29 selected sequences. Finally, three additional sequences matching our selection criteria were found in the literature (Grundy et al., 2003) bringing the total number of sequences to 32. Using these sequences, a graphical representation of the K-turn motif was generated which was arranged to match the known secondary structure consensus (Fig. 4D). From this analysis, we find that the identity of flanking Watson-Crick base pairs (base pairs 64-87 and 69-79) is not conserved as long as base pairing is preserved. However, tandem GA pairs (pairs 66-85 and 67-84) are

conserved to a high degree which is expected given their important role in the formation of the K-turn architecture (Moore et al., 2004). The most represented three nucleotides single-stranded sequence corresponds to GAU, which is different of the sequence of the *B. subtilis lysC* K-turn motif, but similar to the established K-turn consensus motif (Fig. 4D). However, according to this representation, K-turn motifs can be adopted from several other sequences that do not necessarily match the K-turn consensus. Taken together, our results indicate that K-turn sequences found in lysine aptamers may differ from the originally proposed K-turn motif and thus suggest that the K-turn element can be adopted by a diverse array of RNA sequences.

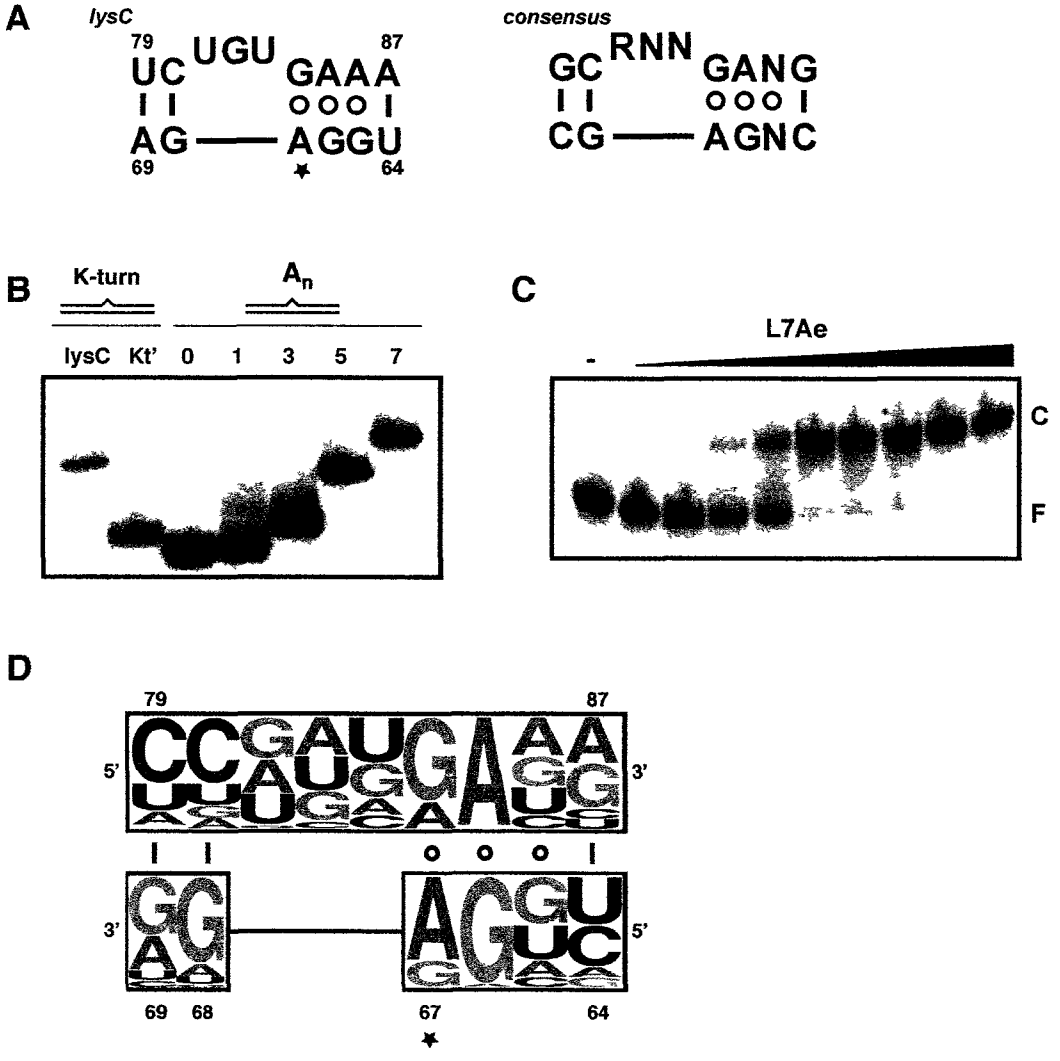


FIGURE 4. A K-turn motif is present in the *lysC* aptamer. **A:** The sequence and predicted secondary structure of *B. subitlis lysC* K-turn motif (left). The secondary structure has been written to match the established K-turn consensus (right) (Klein et al., 2001). Watson-Crick and purine base pairs are shown by vertical lines and circles, respectively. The star shows the nucleotide A67 known to be important for the adoption of the K-turn motif (Goody et al., 2004). R denotes a purine consensus. **B:** Electrophoretic migration of the *lysC* K-turn motif in presence of magnesium ions. Comparison of electrophoretic mobility of RNA duplexes containing a centrally located *lysC* K-turn motif (K-turn) or variants (A_n) with duplexes kinked by virtue of a central oligoadenine bulge. Note that the *lysC* K-turn sequence migrates similarly to the five adenine bulge species. Introduction of an A67C mutation (Kt' variant) completely abolishes the K-turn structure, as previously observed for the Kt-7 variant (Goody et al., 2004). **C:** Gel mobility shift assay of lysine aptamer as a function of the K-turn binding L7Ae protein. The aptamer was incubated in presence of increasing concentrations of L7Ae, and the complexes analyzed by non-denaturing polyacrylamide gel electrophoresis. The aptamer was incubated in absence (-) and in presence of 0.8 nM, 2 nM, 5 nM, 6.7 nM, 10 nM, 12.5 nM, 16 nM, 50 nM and 100 nM L7Ae. The free (F) and L7Ae-complexed aptamer (C) are indicated on the right. **D:** Graphical representation of the K-turn motif obtained from 32 lysine aptamer sequences. The logo has been arranged to match the secondary structure of the K-turn motif. A proportional representation for each residues is shown. Watson-Crick and purine base pairs are shown by vertical lines and circles, respectively. A star indicates the location of A67. Nucleotide positions are indicated below each representation.

Next, we studied the formation of the loop-loop interaction using the RNase T1 endonuclease which is specific to single-stranded guanines. Using the RNase T1 assay, we have previously shown with the adenine riboswitch aptamer that a magnesium-dependent loop-loop interaction can form in absence of ligand (Lemay et al., 2006). We expected that the presence of magnesium ions would induce the formation of the loop-loop interaction, and upon its formation, that guanines located in the P2 loop (positions 72, 75 and 76) would be protected from RNase T1 digestion.

In the context of the natural sequence (WT), RNase T1 cleavage is observed in absence of magnesium ions at expected positions within the P2 loop region (G72, G75 and G76, Fig. 5A). In presence of 10 mM magnesium ions, strong protections are observed for all three positions indicating the adoption of the loop-loop interaction. However, in the context of the Kt' variant, which prevents the adoption of the K-turn structure, the magnesium-dependent protections are no longer observed indicating the inability of the loop-loop interaction to form when the K-turn motif is not properly folded (Fig. 5A).

The functional importance of the K-turn motif was assessed by using a riboswitch variant carrying the Kt' mutation using single-round transcription assays (Fig. 5B). We observe that the presence of the Kt' mutation significantly disrupts the lysine-induced riboswitch regulation control, as shown by lower transcription termination efficiencies obtained in absence and in presence of lysine (~13% and ~31%, respectively). This result is very similar to those obtained for the P2 and P3 riboswitch variants (Fig. 3), suggesting that the formation of the K-turn motif is important for the formation of the P2-P3 loop-loop interaction. Thus, as previously found for ribosomal (Klein et al., 2001) and small non-coding RNAs (Dennis & Omer, 2005), K-turns also mediate important RNA tertiary structure interactions which are relevant for the biological function of riboswitches.

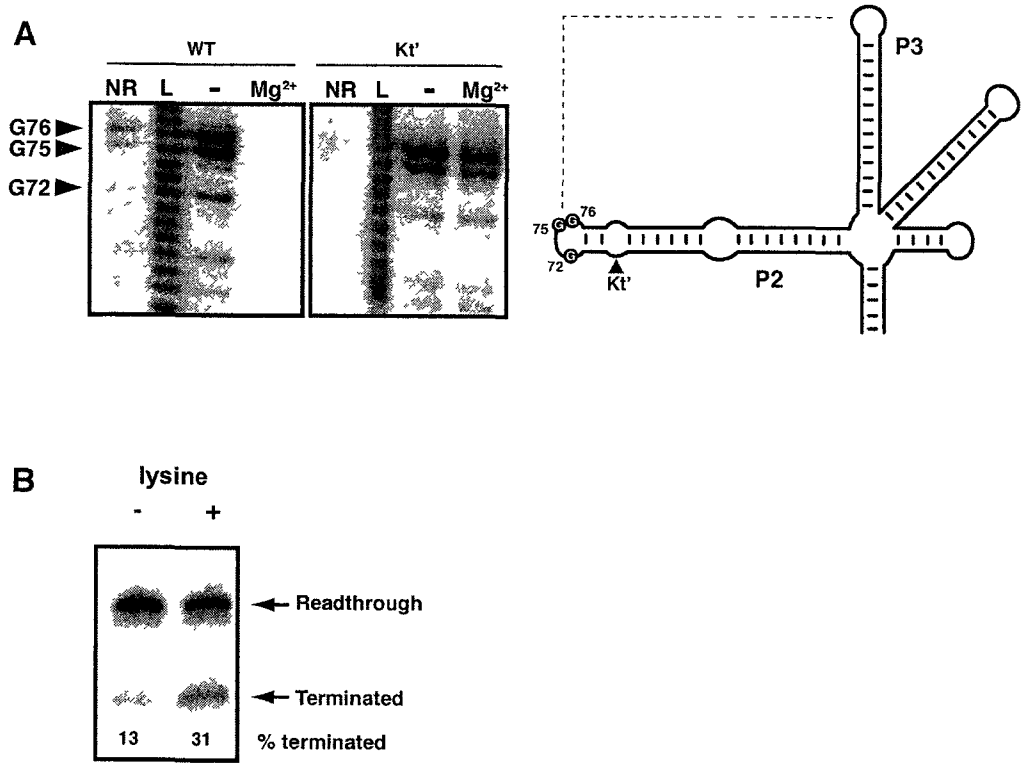


FIGURE 5. The kink-turn motif is important for loop-loop formation and transcription control. **A:** Partial digestions of end-labeled wild-type and Kt' variant aptamers by RNase T1 show that G72, G75 and G76 present in the P2 helical domain are protected from cleavage by the addition of magnesium for the wild type molecule (left panel). However, no protection is observed when the Kt' mutant is introduced in the aptamer (right panel). Lane NR contains unreacted RNA, while lane L contains RNA subjected to partial alkaline digestion. Positions of strong cleavage are indicated on the left. A schematic is shown on the right representing the three observed guanines and the loop-loop interaction. **B:** Single round transcriptions of the Kt' mutant performed in the absence (-) and in the presence (+) of 5 mM lysine. Readthrough and terminated products were separated by gel electrophoresis and are indicated on the right. The percentage of terminated product is indicated below the gel for each reaction.

Ion-induced folding of the lysine aptamer domain

Our single-round transcription data indicate that adoption of the loop-loop interaction, via the correct formation of the K-turn structure, is very important for the lysine riboswitch regulation. Since in-line probing has shown a reorganization of the folding of the core domain upon ligand binding (Sudarsan et al., 2003b), we speculated that formation of this tertiary interaction should be important for the folding of the lysine riboswitch core domain. We took advantage of the non-conserved nature of the core position 194 to introduce the fluorescent local reporter 2-aminopurine (2AP) to make a spectroscopic study of conformational changes induced by magnesium ions binding. Fluorescence of 2AP is strongly quenched by the stacking of adjacent nucleobases, so changes in the intensity of fluorescence are indicative of changes in the local environment of the fluorophore (Ward et al., 1969; Xu et al., 1994; Stivers, 1998; Jean & Hall, 2001; Lafontaine et al., 2002b). 2AP fluorescence emission spectra were recorded in the presence of 90 mM Tris-borate and 100 mM KCl, and upon excitation of 2AP, a relatively low fluorescence was detected, indicating that 2AP is not highly exposed to the solvent. However, upon addition of magnesium ions, an increase of the fluorescence intensity is observed indicating that the 2AP local structure is altered, most probably resulting from the exposition of 2AP to the solvent (Fig. 6, insert). Fluorescence changes calculated for each magnesium ions concentration allow to estimate apparent magnesium-binding parameters assuming a simple two-state transition model (see Materials and Methods). Good fits were obtained with this model, giving an apparent affinity for magnesium ions $[Mg^{2+}]_{1/2}$ of 1.68 mM and a Hill coefficient n = 1.6 ± 0.1, suggesting a cooperative character for the transition of the 2AP fluorescence change (Table 1).

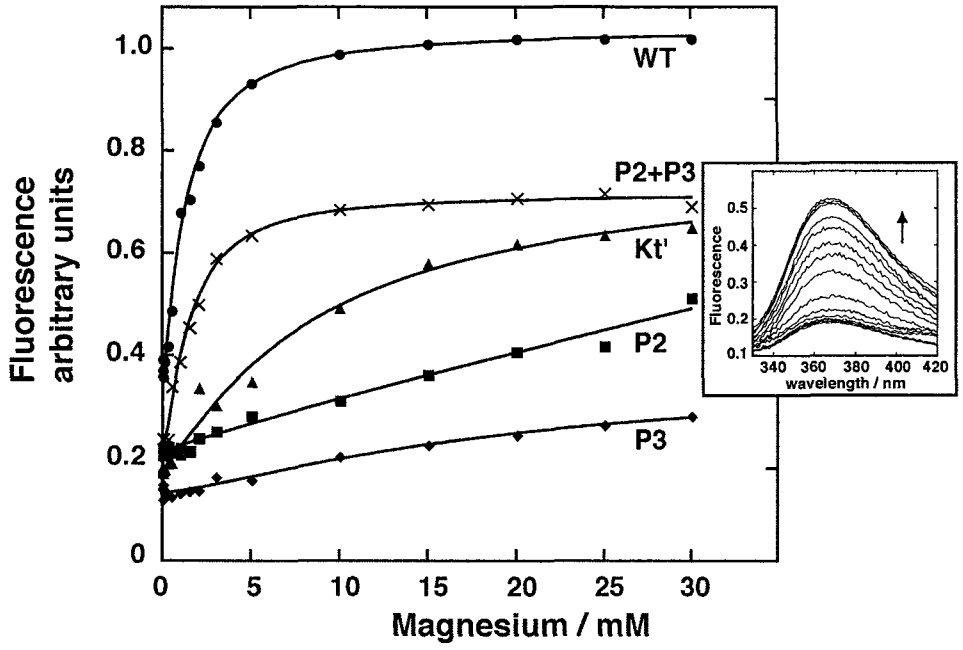


FIGURE 6. Formation of the loop-loop interaction is important for the reorganization of the aptamer core region. The insert shows fluorescence emission spectra (λex = 300 nm) from 330 nm to 420 nm for each magnesium ions concentration. An increase of fluorescence is observed when the fluorescent lysine aptamer is incubated by the addition of magnesium ions. The normalized 2AP fluorescence intensity is plotted as a function of magnesium ions for the natural sequence (WT), the loop-loop deficient variants P2 (squares) and P3 (diamonds), the doubly substituted aptamer P2+P3 (crosses) and the K-turn-compromised variant (triangles). The experimental data were fitted (line) by regression to a simple two-state model where the binding of metal ions to the aptamer induces a structural change.

TABLE 1. Magnesium ion titrations on 2AP-lysine aptamers.

Variant	$[Mg^{2+}]_{1/2}$ (mM)	Hill coefficient (*n*)
WT	1.68	1.6 ± 0.1
P2[a]	23.6	1.1 ± 0.3
P3[b]	38.1	0.98 ± 0.23
P2+P3[c]	1.54	1.4 ± 0.1
Kt'[d]	9.70	1.1 ± 0.1

TABLE 1. Half of apparent concentrations for the magnesium ion-induced process ($[Mg^{2+}]_{1/2}$) were measured under standard conditions. Asymptotic standard deviations on the fits indicate that errors in $[Mg^{2+}]_{1/2}$ are generally less than 10% of values. [a]Nucleotides G72 to G76 replaced with their Watson Crick complement. [b]Nucleotides C126 to C130 replaced with their Watson-Crick complement. [c]Nucleotides G72 to G76 relocated to positions 126 to 130, and nucleotides C126 to C130 relocated to positions 72 to 76. [d]Nucleotide A67 replaced with a cytosine.

Formation of the loop-loop interaction was also studied using 2AP fluorescence. Aptamer variants were made in which either the P2 or the P3 loop region was exchanged for its counterpart and resulting constructs were monitored using 2AP fluorescence (Fig. 6). Upon magnesium ions titration, it is immediately apparent that the folding of the aptamer domain is severely impaired, giving $[Mg^{2+}]_{1/2}$ values of 23.6 mM and 38.1 mM for the P2 and P3 variants, respectively (Table 1). Hill coefficient values were found to decrease slightly ($n \sim 1$), but were not considered as an indicator of perturbed folding given the small magnitude change. We also tested the P2+P3 variant and values of $[Mg^{2+}]_{1/2}$ of 1.54 mM and n = 1.4 ± 0.1 were obtained, which is very close to values obtained for the natural sequence (Table 1). In addition, given that the correct folding of the K-turn motif is required for loop-loop formation (Fig. 5A), we examined the effect of the disruption of the K-turn structure

on the core folding. Upon introduction of the Kt' mutation, the folding of the core is found to be perturbed, albeit to a lower extent compared to loop-loop variants, with values of $[Mg^{2+}]_{1/2}$ = 9.7 mM and n = 1.1 ± 0.1. Taken together, our 2AP fluorescence results suggest that a structural reorganization is occurring in the core region of the lysine aptamer, and that this reorganization requires the formation of the kink-turn dependent loop-loop interaction.

DISCUSSION

The study of riboswitches is of particular interest because RNA-ligand interactions are potentially very important in the elaboration of novel drug targets that could disrupt the expression of essential bacterial genes. Some examples clearly show that compounds competing with natural ligands such as lysine (Blount et al., 2007) and thiamine pyrophosphate (Sudarsan et al., 2005) alter the expression of essential genes and affect viability of pathogenic bacteria. The lysine riboswitch has been previously shown to regulate the expression of genes involved in the lysine biosynthesis pathway or transport (Grundy et al., 2003; Rodionov et al., 2003; Sudarsan et al., 2003b). Previous works also showed that lysine induces premature termination of transcription of the *B. subtilis lysC* gene by a riboswitch regulation mechanism. In this study, we performed a structural characterization of the lysine riboswitch aptamer domain and identified a novel tertiary interaction that is essential for the riboswitch activity.

While the existence of a loop-loop interaction in the lysine aptamer was previously proposed (Grundy et al., 2003; Rodionov et al., 2003), no direct experimental evidence was provided to unambiguously confirm its presence and its functional importance. Our single-round transcription data clearly show that the *lysC* lysine riboswitch relies on the formation of a loop-loop interaction to efficiently regulate

transcription in presence of lysine (Fig. 3). Interestingly, loop-loop-deficient mutants (P2 and P3) still retain a certain capacity to respond to lysine during transcription. These results suggest that although riboswitch regulation is affected by loop-loop disruption, transcription regulation still occur in a lysine dependent manner, indicating that lysine can bind to the aptamer domain. Accordingly, in-line probing results of the lysine aptamer have shown that the core region of the aptamer is reorganized upon ligand binding; the lysine binding site is thus very likely to reside within the core region of the aptamer domain, and mutations performed peripherally of the aptamer are not likely to directly alter the ligand binding site (Sudarsan et al., 2003b). However, when the G39C mutation is introduced in the core of the aptamer, the riboswitch is unable to perform transcription regulation in presence of lysine (Fig. 3), suggesting that G39 could be directly involved in the formation of the lysine binding site. This is supported by in-line probing results which have shown that the region J1/2, and therefore G39, is rearranged upon ligand binding (Sudarsan et al., 2003b). Thus, together with in-line probing data, our results suggest that the loop-loop interaction is required for the proper folding of the aptamer domain, but not for ligand binding.

Our 2AP fluorescence data show that the loop-loop formation occurs in absence of ligand and influences the folding of the core (Fig. 6). The detailed analysis of the magnesium-induced folding process of the *lysC* riboswitch has revealed that the folding of the core domain is dependent on the correct formation of the loop-loop interaction, and that disruption of the loop-loop interaction leads to a marked decrease in the ability of the core to fold correctly. Because 2AP fluorescence experiments are performed in absence of lysine, it suggests that the loop-loop interaction participates in the folding of the core in absence of ligand. This rearrangement could potentially be very important for the organization of the core region prior to ligand binding. Similarly, folding effects of a loop-loop interaction have also previously been characterized in the hammerhead ribozyme where it was found

that the magnesium ions concentration required for effective ribozyme cleavage was decreased (Khvorova et al., 2003; Penedo et al., 2004; Martick & Scott, 2006). This indicates that these peripheral folding elements (i.e. loop-loop interaction) may be important as auxiliary folding enhancers and that given their high representation in riboswitch secondary structures, may have important *in vivo* implications as found for the hammerhead ribozyme (Khvorova et al., 2003). In the present case, it is very likely that riboswitches bearing compromised loop-loop structures exhibit reduced transcription termination because embedded mutant aptamers are only partially folded, and are thus precluded to perform efficient ligand binding under the conditions used for transcription assay.

The presence of an unconventional kink-turn structure is observed in the P2 helical domain of the aptamer (Fig. 1). Our native gel electrophoresis together with single-round transcription data indicate that the *lysC* K-turn element introduces a sharp kink in an RNA helix very similar to an archetypal molecule (Klein et al., 2001; Goody et al., 2004), and that its correct folding is required for efficient riboswitch regulation (Figs 4 and 5B). Because it is located proximally to P2 the stem-loop, which is involved in the formation of the loop-loop interaction, it is very likely that correct folding of the K-turn is of primary importance for the formation of the loop-loop structure. Indeed, RNase T1 probing and 2AP fluorescence assays indicate that disruption of the K-turn inhibits formation of the loop-loop interaction and the magnesium-induced reorganization of the core domain of the aptamer (Figs 5A and 6). Because the *lysC* kink-turn does not fit the K-turn consensus sequence, the fact that it has similar folding properties compared to an archetypal motif suggests that K-turns are more widespread than previously envisioned (Klein et al., 2001), as described recently (Dennis & Omer, 2005). This is supported by a recent study in which it was observed that K-turn motifs exhibiting similar characteristics to the *lysC* variant are also found numerous times in ribosomal K-turns (Lescoute et al., 2005). Furthermore, additional K-turns located in box C/D and H/ACA of snoRNP are

essential for methylation and pseudouridylation, where protein binding to the K-turn stabilizes the RNA kinked-structure during the formation of the RNP complex (Henras et al., 2004; Dennis & Omer, 2005; Turner et al., 2005).

Additional examples of kink-turn containing RNA molecules can be found in alternative biological contexts. For instance, the crystal structure of the the SAM-responding riboswitch (Epshtein et al., 2003; McDaniel et al., 2003; Winkler et al., 2003) contains a kink-turn motif that is most likely important for the formation of a long-range pseudoknot structure (Montange & Batey, 2006), both of which being important for the SAM riboswitch regulation (Winkler et al., 2001; McDaniel et al., 2005). In addition, in the 23S ribosomal RNA, the Kt-7 kink-turn is located in the 5-6-7 junction (Ban et al., 2000; Klein et al., 2001). Interestingly, the 5-6-7 junction also exhibits a loop-loop interaction between two of its stem-loops, an arrangement very similar to what is found in the *lysC* aptamer. However, in the case of the ribosomal junction, the crystal structure shows that ribosomal proteins L24 and L29 interact with Kt-7, suggesting that their presence could be important for the correct folding of the K-turn motif, and consequently for the loop-loop structure (Ban et al., 2000). Thus, because K-turn motifs serve as protein anchoring points in the building of complex architectures, the *lysC* K-turn could also be involved in protein interactions, as observed with L7Ae in our native gel assays (Fig. 4C). Although the *in vivo* relevance of this result needs to be experimentally addressed, it suggests that, even if such RNA elements can work *in vitro* without any protein factor, the *in vivo* mechanisms of some K-turn containing riboswitches may be different by employing additional factors.

ACKNOWLEDGMENTS

We thank members of the Lafontaine laboratory and Drs Carlos Penedo and Alain Lavigueur for helpful discussions. We also thank DMJ Lilley for generous gift of the L7Ae protein. This work was supported by the National Sciences and Engineering Research Council of Canada (NSERC). D.A.L. is a Chercheur-Boursier Junior I of the Fonds de la Recherche en Santé du Québec.

MATERIALS AND METHODS

Synthesis of RNA molecules

RNA molecules were transcribed from PCR DNA template using the T7 RNA polymerase (Milligan et al 1987). All transcribed RNA species begin with a 5′-GCG sequence to minimize the 5′-heterogeneity of the RNA population (Pleiss et al., 1998). RNA molecules containing 2-aminopurine (2AP) were purchased from Dharmacon (Boulder), deprotected and purified as described previously (Wilson & Lilley, 2002). The internally-labeled 2AP aptamer was assembled from oligonucleotides with the following sequences (all written 5′ to 3′):

5′ strand:

GCGGUGAAGAUAGAGGUGCGAACUUCAAGAGUAUGCCUUUGGAGAAAGAUG GAUUCUGUGAAAAAGGCUGAAAGGGGAGCGUCGCCGAAGCAAAUAAAACCCC AUCGGUAUUAUUUGCUGGCCGUGCAUUGAAUAAAUGUAAGGCUGUCAAGAA GCAUC

3′ strand:

5′-PO4-AUUGCUUCUUGGAG(2AP)GCUAUCUUCACC. P2 and P3 variant sequences in the 5′ strand shown underlined represent the regions that were changed for their Watson-Crick counterpart for the study of the loop-loop interaction. Purified 5′ and 3′ strands RNA strands were annealed by heating a mixture (molar

ratio 1.1:1) to 65°C in 10 mM Hepes, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, and slowly cooled to 37°C. T4 RNA ligase (New England BioLabs) was then added to the reaction and samples were incubated at 37°C for 3 h. Full-length ligated RNA molecules were purified by denaturing polyacrylamide gel electrophoresis, electro-eluted and dissolved in water.

The electrophoretic retardation of K-turn and oligoadenine bulged species was performed using 120 base paired RNA duplexes. The RNA strand containing the three nucleotide bulge of the duplex containing the lysC K-turn motif has the following sequence:

GCGACACACGGACAAGCAAUGUGUGACGUAAUAUGGAGGGCGGACAAGCAA
UGUGUGUUCUGUGAAAAAUGUACUGUUCAACUUGACAGUCCUCGCAUGUAC
UGUUCAACUUAAGACCCACGCUUC

where the K-turn motif is underlined. The sequences of the A67C mutation and oligoadenine bulged species are derived from this sequence.

In vitro transcription assays

DNA templates for *in vitro* transcriptions were obtained by PCR with a combination of oligonucleotides corresponding to the *Bacillus subtilis glyQS* promoter fused to the *B. subtilis* lysine riboswitch and a 95 nucleotides sequence downstream of the expression platform. PCR products were cleaned with the Qiagen QIAquick gel extraction kit. The promoter region ends by an adenine and is fused to the C17 position of the riboswitch sequence, allowing the transcription to be initiated by the ApC dinucleotide. In the terminated species, RNAP stops at position 269. The readthrough species allows the RNAP to stop 95 nucleotides downstream of that position (i.e., 49 nucleotides after the AUG).

Reactions were initiated with 300 fmol of DNA template in a buffer containing 150 μM ApC, 2,5 μM rATP and rGTP, 0,75 μM rUTP, 2 μCi [α-^{32}P]-UTP, 50 mM Tris-HCl (pH 8.0 at 23°C), 20 mM $MgCl_2$, 0,1 mM EDTA, 1 mM DTT, 10% glycerol, 0,25 U *E. coli* RNAP (Epicentre), and L-lysine when required. This initiation mixture was incubated 15 min at 37°C. Reactions were completed by the addition of each rNTP to a final concentration of 65 μM and 0.45 mg/ml heparin to avoid re-initiation of transcription. Reactions were incubated at 37°C for an additional 15 min and stopped with an equal volume of formamide. Transcription products were analyzed using a 5% denaturing PAGE and visualized on a Phosphor Imager screen. The lysine-induced transcription termination can be described by the equation,

$$\alpha = T_I + T_M \bullet ([\text{lysine}] / (T_{1/2} + [\text{lysine}]))$$

where α is the percentage terminated, T_I and T_M are the initial and maximal changes in transcription termination, and $T_{1/2}$ is the lysine concentration required to obtain half of the change. The equation assumes a simple 1:1 stoichiometry between the aptamer and lysine as expected from in-line probing results (Sudarsan et al., 2003b). The reported errors are the standard uncertainties of the data from the best-fit theoretical curves. The standard uncertainty of the measurement is thus assumed to be approximated by the standard deviation of the points from the fitted curve (Flannery et al., 1992; Rist & Marino, 2001).

Fluorescence spectroscopy

Fluorescence spectroscopy was performed on a Quanta Master fluorometer. All data were collected at 10°C in 90 mM Tris-borate, pH 8.3 and 100 mM KCl. Excitation for 2-aminopurine fluorescence was done at 300 nm to obtain a good separation between the Raman peak and 2AP fluorescence signal. Spectra were corrected for background, and intensities were determined by integrating data collected over the range 330-450 nm.

Relative 2AP fluorescence was determined by monitoring the fluorescent emission of a fixed concentration of internally-labeled 2AP lysine aptamer (100 nM) and titrating the concentration of magnesium ions. Fluorescence data were fitted to a simple two-state model that assumes an all-or-none conformational transition induced by the binding of magnesium ions, with a Hill coefficient n and an apparent association constant K_A. The proportion of folded aptamer (α) is given by:

$$\alpha = K_A \bullet [Mg^{2+}]^n / (1 + K_A \bullet [Mg^{2+}]^n)$$

As the parameters K_A and n co-vary, we present the $[Mg^{2+}]^{1/2} = (1/K_A)^{1/n}$, which gives a robust estimate of the affinity for magnesium ions (Lafontaine et al., 2001, 2002a; Lemay et al., 2006).

Gel electrophoresis under non-denaturing conditions

Radioactively [5′-^{32}P]-labeled RNA (final concentration < 1nM) in 90 mM Tris and 89 mM boric acid, pH 8.5 (TB buffer) and 10 mM $MgCl_2$ was electrophoresed in 10% acrylamide:bisacrylamide (29:1) gels in TB with 10 mM $MgCl_2$ at 4°C at 150 V for 48 h. The running buffer was circulated during electrophoresis. Gels were exposed to Phosphor Imager screens.

Gel electrophoresis under non-denaturing conditions in presence of L7Ae

RNA samples were prepared in 90 mM Tris-borate, pH 8.3, 1 mM EDTA (1X TBE), 25 mM NaCl and 5% glycerol. The concentration of L7Ae is indicated in the text. Mixtures were incubated on ice for 1 h before electrophoresis on 6.6% acrylamide:bisacrylamide (29:1) gels in 1X TBE with 25 mM NaCl at 4°C at 250V for 5 h. The running buffer was circulated during electrophoresis. Gels were exposed to Phosphor Imager screens. The formation of L7Ae-aptamer complex was analyzed using a simple non-covalent binding model:

$$\alpha = [\text{aptamer}] / (K_{Dapp} + [\text{aptamer}])$$

where α is the fraction bound and K_{Dapp} is the apparent dissociation constant. The equation assumes a simple 1:1 stoichiometry between the aptamer and L7Ae as it is expected from the crystal structure (Moore et al., 2004).

Partial RNase T1 cleavage

[5′-^{32}P] RNA was incubated in a 50 mM Tris-HCl (pH 8.3), 100 mM KCl in the absence or presence of 10 mM $MgCl_2$ for 5 minutes at 37°C. RNase T1 (0.2 U) was then added and allowed to react for 2 minutes at 37°C. Reactions were stopped by the addition of a volume of 95% formamide. Products were separated on denaturing polyacrylamide gels, which were dried and exposed to Phosphor Imager screen.

REFERENCES

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 289 : 905-920.

Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I, Wickiser JK, Breaker RR. 2004. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc Natl Acad Sci U S A 101 :6421-6426.

Bhattacharyya A, Lilley DM. 1989. The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). Nucleic Acids Res 17 :6821-6840.

Bhattacharyya A, Murchie AI, Lilley DM. 1990. RNA bulges and the helical periodicity of double-stranded RNA. Nature 343 :484-487.

Blount KF, Wang JX, Lim J, Sudarsan N, Breaker RR. 2007. Antibacterial lysine analogs that target lysine riboswitches. Nat Chem Biol 3 :44-49.

Breaker RR. 2006. Riboswitches and the RNA World. In: R.F. Gesteland TRCaJFA, ed. The RNA World: Third Edition . New York: Cold Spring Harbor Laboratory Press. pp 89-108.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14 :1188-1190.

Dennis PP, Omer A. 2005. Small non-coding RNAs in Archaea. Curr Opin Microbiol 8 :685-694.

Diekmann S, Wang JC. 1985. On the sequence determinants and flexibility of the kinetoplast DNA fragment with abnormal gel electrophoretic mobilities. J Mol Biol 186 :1-11.

Epshtein V, Mironov AS, Nudler E. 2003. The riboswitch-mediated control of sulfur metabolism in bacteria. Proc Natl Acad Sci U S A 100 :5052-5056.

Flannery BP, Teukolsky SA, Vetterling WT. 1992. Numerical Recipes in Fortran, 2nd Edn. Cambridge: Cambridge University Press, UK.

Gohlke C, Murchie AI, Lilley DM, Clegg RM. 1994. Kinking of DNA and RNA helices by bulged nucleotides observed by fluorescence resonance energy transfer. Proc Natl Acad Sci U S A 91 :11660-11664.

Goody TA, Melcher SE, Norman DG, Lilley DM. 2004. The kink-turn motif in RNA is dimorphic, and metal ion-dependent. Rna 10 :254-264.

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. Nucleic Acids Res 31 :439-441.

Grundy FJ, Henkin TM. 2004. Regulation of gene expression by effectors that bind to RNA. Curr Opin Microbiol 7 :126-131.

Grundy FJ, Lehman SC, Henkin TM. 2003. The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. Proc Natl Acad Sci U S A 100 :12057-12062.

Henras AK, Dez C, Henry Y. 2004. RNA structure and function in C/D and H/ACA s(no)RNPs. Curr Opin Struct Biol 14 :335-343.

Hsieh CH, Griffith JD. 1989. Deletions of bases in one strand of duplex DNA, in contrast to single-base mismatches, produce highly kinked molecules: possible relevance to the folding of single-stranded nucleic acids. Proc Natl Acad Sci U S A 86 :4833-4837.

Hutton CA, Southwood TJ, Turner JJ. 2003. Inhibitors of lysine biosynthesis as antibacterial agents. Mini Rev Med Chem 3 :115-127.

Jean JM, Hall KB. 2001. 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking. Proc Natl Acad Sci U S A 98 :37-41.

Khvorova A, Lescoute A, Westhof E, Jayasena SD. 2003. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Biol 10 :708-712.

Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001. The kink-turn: a new RNA secondary structure motif. Embo J 20 :4214-4221.

Kochhar S, Paulus H. 1996. Lysine-induced premature transcription termination in the lysC operon of Bacillus subtilis. Microbiology 142 (Pt 7) :1635-1639.

Kubodera T, Watanabe M, Yoshiuchi K, Yamashita N, Nishimura A, Nakai S, Gomi K, Hanamoto H. 2003. Thiamine-regulated gene expression of Aspergillus oryzae thiA requires splicing of the intron containing a riboswitch-like domain in the 5′-UTR. FEBS Lett 555 :516-520.

Kuhn JF, Tran EJ, Maxwell ES. 2002. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. Nucleic Acids Res 30 :931-941.

Lafontaine DA, Norman DG, Lilley DM. 2001. Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction. Embo J 20 :1415-1424.

Lafontaine DA, Norman DG, Lilley DM. 2002a. The global structure of the VS ribozyme. Embo J 21 :2461-2471.

Lafontaine DA, Wilson TJ, Zhao ZY, Lilley DM. 2002b. Functional group requirements in the probable active site of the VS ribozyme. J Mol Biol 323 : 23-34.

Lai EC. 2003. RNA sensors and riboswitches: self-regulating messages. Curr Biol 13 :R285-291.

Lemay JF, Lafontaine DA. 2007. Core requirements of the adenine riboswitch aptamer for ligand binding. RNA 13 :339-350.

Lemay JF, Penedo JC, Tremblay R, Lilley DM, Lafontaine DA. 2006. Folding of the adenine riboswitch. Chem Biol 13 :857-868.

Lescoute A, Leontis NB, Massire C, Westhof E. 2005. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. Nucleic Acids Res 33 : 2395-2409.

Liao HH, Hseu TH. 1998. Analysis of the regulatory region of the lysC gene of Escherichia coli. FEMS Microbiol Lett 168 :31-36.

Lilley DM. 1995. Kinking of DNA and RNA by base bulges. Proc Natl Acad Sci U S A 92 :7140-7142.

Lilley DMJ, Clegg RM, Diekmann S, Seeman NC, von Kitzing E, Hagerman P. 1996. Nomenclature Committee of the International Union of Biochemistry: a nomenclature of junctions and branchpoints in nucleic acids. Recommendations 1994. J Mol Biol 255 :554-555.

Lu Y, Shevtchenko TN, Paulus H. 1992. Fine-structure mapping of cis-acting control sites in the lysC operon of Bacillus subtilis. FEMS Microbiol Lett 71 :23-27.

Luebke KJ, Tinoco I, Jr. 1996. Sequence effects on RNA bulge-induced helix bending and a conserved five-nucleotide bulge from the group I introns. Biochemistry 35 :11677-11684.

Mader U, Homuth G, Scharf C, Buttner K, Bode R, Hecker M. 2002. Transcriptome and proteome analysis of Bacillus subtilis gene expression modulated by amino acid availability. J Bacteriol 184 :4288-4295.

Mandal M, Breaker RR. 2004. Gene regulation by riboswitches. Nat Rev Mol Cell Biol 5 :451-463.

Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR. 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science 306 :275-279.

Marini JC, Levene SD, Crothers DM, Englund PT. 1982. Bent helical structure in kinetoplast DNA. Proc Natl Acad Sci U S A 79 :7664-7668.

Martick M, Scott WG. 2006. Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell 126 :309-320.

McDaniel BA, Grundy FJ, Artsimovitch I, Henkin TM. 2003. Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. Proc Natl Acad Sci U S A 100 :3083-3088.

McDaniel BA, Grundy FJ, Henkin TM. 2005. A tertiary structural element in S box leader RNAs is required for S-adenosylmethionine-directed transcription termination. Mol Microbiol 57 :1008-1021.

Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res 15 :8783-8798.

Montange RK, Batey RT. 2006. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. Nature 441 :1172-1175.

Moore T, Zhang Y, Fenley MO, Li H. 2004. Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. Structure 12 :807-818.

Nudler E, Mironov AS. 2004. The riboswitch control of bacterial metabolism. Trends Biochem Sci 29 :11-17.

Patte JC. 1994. Biosynthesis of threonine and lysine . Washington D.C.: American Society for Microbiology.

Patte JC, Akrim M, Mejean V. 1998. The leader sequence of the Escherichia coli lysC gene is involved in the regulation of LysC synthesis. FEMS Microbiol Lett 169 :165-170.

Penedo JC, Wilson TJ, Jayasena SD, Khvorova A, Lilley DM. 2004. Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. Rna 10 :880-888.

Pleiss JA, Derrick ML, Uhlenbeck OC. 1998. T7 RNA polymerase produces 5′ end heterogeneity during in vitro transcription from certain templates. Rna 4 : 1313-1317.

Rist M, Marino J. 2001. Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence. Nucleic Acids Res 29 :2401-2408.

Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. 2003. Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? Nucleic Acids Res 31 :6748-6757.

Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. 2004. Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory systems. Nucleic Acids Res 32 :3340-3353.

Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A. 2000. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell 102 : 615-623.

Stivers JT. 1998. 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. Nucleic Acids Res 26 :3837-3844.

Sudarsan N, Barrick JE, Breaker RR. 2003a. Metabolite-binding RNA domains are present in the genes of eukaryotes. Rna 9 :644-647.

Sudarsan N, Cohen-Chalamish S, Nakamura S, Emilsson GM, Breaker RR. 2005. Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. Chem Biol 12 :1325-1335.

Sudarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, Breaker RR. 2006. Tandem riboswitch architectures exhibit complex gene control functions. Science 314 :300-304.

Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. 2003b. An mRNA structure in bacteria that controls gene expression by binding lysine. Genes Dev 17 :2688-2697.

Tang RS, Draper DE. 1990. Bulge loops used to measure the helical twist of RNA in solution. Biochemistry 29 :5232-5237.

Turner B, Melcher SE, Wilson TJ, Norman DG, Lilley DM. 2005. Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. Rna 11 :1192-1200.

Vold B, Szulmajster J, Carbone A. 1975. Regulation of dihydrodipicolinate synthase and aspartate kinase in Bacillus subtilis. J Bacteriol 121 :970-974.

Ward DC, Reich E, Stryer L. 1969. Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. J Biol Chem 244 :1228-1237.

Wilson TJ, Lilley DM. 2002. Metal ion binding and the folding of the hairpin ribozyme. Rna 8 :587-600.

Winkler WC. 2005. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. Curr Opin Chem Biol 9 :594-602.

Winkler WC, Grundy FJ, Murphy BA, Henkin TM. 2001. The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. Rna 7 :1165-1172.

Winkler WC, Nahvi A, Sudarsan N, Barrick JE, Breaker RR. 2003. An mRNA structure that controls gene expression by binding S-adenosylmethionine. Nat Struct Biol 10 :701-707.

Xu D, Evans KO, Nordlund TM. 1994. Melting and premelting transitions of an oligomer measured by DNA base fluorescence and absorption. Biochemistry 33 :9592-9599.

Zacharias M, Hagerman PJ. 1995. Bulge-induced bends in RNA: quantification by transient electric birefringence. J Mol Biol 247 :486-500.

CHAPITRE 2

IMPORTANCE DES ÉLÉMENTS PÉRIPHÉRIQUES DANS LA RÉGULATION DE LA TRANSCRIPTION PAR LE RIBORÉGULATEUR LYSINE

2.1 Présentation de l'article et contribution

Les éléments essentiels des riborégulateurs sont généralement conservés parmi les séquences d'une même famille, que ce soit au niveau de la séquence, ou de la structure où on peut noter une covariation dans les éléments comme les hélices ou les pseudonoeuds. Dans le cas du riborégulateur lysine, le coeur de l'aptamère représente la région la mieux conservée puisqu'il constitue le site de liaison du ligand. Les régions impliquées dans le *K-turn* et la boucle E sont aussi passablement conservées, ce qui a permis la prédiction de tels motifs (Grundy *et al.,* 2003; Rodionov *et al.,* 2003; Sudarsan *et al.,* 2003b). Les deux adénines de la boucle terminale de P4 sont aussi conservées dans 100% des séquences identifiées, ce qui laissait pressentir son implication dans une interaction tertiaire quelconque, même avant la publication de la structure native d'un riborégulateur lysine (Andrew D Garst *et al.,* 2008; Serganov *et al.,* 2008).

Dans cette deuxième publication, nous avons entre autres démontré l'importance distincte des deux interactions tertiaires sur la structure globale de l'aptamère par des essais en gels natifs, en plus de démontrer avec la technique de SHAPE qu'il existe une synergie de repliement entre ces deux interactions. Comme pour le premier article, nous avons utilisé la construction contenant une 2AP interne pour observer les changements de conformation du coeur de l'aptamère. Ces essais de fluorescence nous ont permis de conclure que l'interaction P2-L4 influence

directement le repliement du site de liaison, d'où son importance au niveau de la fonction du riborégulateur. Cet article soulève aussi le rôle et l'importance des éléments moins conservés, comme la très variable tige P5 et la tige identifiée comme P6 localisée dans la plate-forme d'expression.

Les électrophorèses en conditions natives ont été réalisées par Raja Chinnappan, qui a aussi rédigé cette portion des procédures expérimentales. J'ai réalisé toutes les autres expériences en plus de rédiger le manuscrit et de faire les figures avec le Pr Daniel Lafontaine.

2.2 Manuscrit

Blouin, S., Chinnappan, R. et Lafontaine, D.A. (2010). Folding of the lysine riboswitch: importance of peripheral elements for trancriptional regulation. Nucleic Acids Research. [publication électronique avant impression]

Folding of the lysine riboswitch: importance of peripheral elements for transcriptional regulation

Simon Blouin, Raja Chinnappan and Daniel A. Lafontaine

Groupe ARN/RNA Group, Département de biologie, Faculté des sciences, Université de Sherbrooke, Sherbrooke, Québec, Canada, J1K 2R1.

Keywords: Riboswitch, ligand-induced conformational changes, RNA folding, lysine
Running title: Importance of lysine riboswitch peripheral elements.

Address correspondence to:
Daniel Lafontaine
2500, blvd Université
Sherbrooke, Québec, Canada, J1K 2R1
E-mail: Daniel.Lafontaine@USherbrooke.ca

ABSTRACT

The *Bacillus subtilis lysC* lysine riboswitch modulates its own gene expression upon lysine binding through a transcription attenuation mechanism. The riboswitch aptamer is organized around a single five-way junction that provides the scaffold for two long-range tertiary interactions (loop L2-loop L3 and helix P2-loop L4)—all of this for the creation of a specific lysine binding site. We have determined that the interaction P2-L4 is particularly important for the organization of the ligand binding site and for the riboswitch transcription attenuation control. Moreover, we have observed that a folding synergy between L2-L3 and P2-L4 allows both interactions to fold at lower magnesium ion concentrations. The P2-L4 interaction is also critical for the close juxtaposition involving stems P1 and P5. This is facilitated by the presence of lysine suggesting an active role of the ligand in the folding transition. We also show that a previously uncharacterized stem-loop located in the expression platform is highly important for the riboswitch activity. Thus, folding elements located in the aptamer and the expression platform both influence the lysine riboswitch gene regulation.

INTRODUCTION

Riboswitches are genetic control elements located in untranslated regions of mRNA that positively or negatively modulate mRNA cellular levels, mostly through conformational changes. Riboswitches have been found in all kingdoms of life and shown to regulate gene expression by transcription attenuation, translation initiation, ribozyme-mediated mRNA processing and mRNA splicing (1). These RNA regulatory elements directly bind cellular metabolites and alter the expression of downstream genes that are usually associated with the biosynthesis or transport of target metabolites (2-4). The riboswitch-driven genetic control can also be initiated through

a number of cellular factors, such as magnesium ions (5), temperature (6) and uncharged tRNA (7).

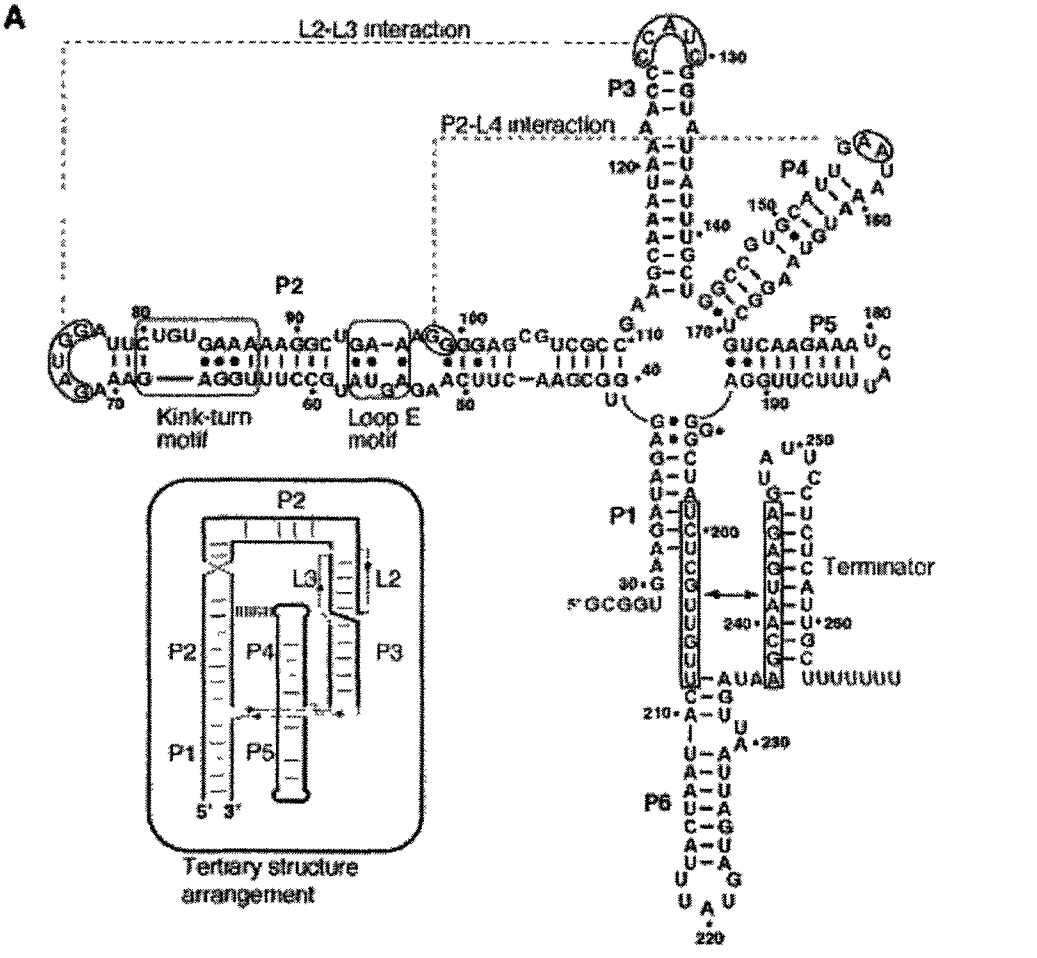
Riboswitches are composed of two modular domains: an aptamer and an expression platform. The most conserved region of riboswitches is the aptamer, which is a receptor that specifically binds a target metabolite. However, the expression platform varies widely in sequence and structure and typically affects gene expression through secondary structure reorganization. Switching between gene activation and repression is accomplished by the ligand-induced riboswitch reorganization, which adopts mutually exclusive secondary structures (8,9). Recently, several bioinformatic studies have provided strong support for putative additional RNA regulatory elements that still await characterization (10-15).

In the bacterium *Bacillus subtilis*, two riboswitches are involved in the regulation of the lysine cellular level by controlling the expression of an aspartokinase II (*lysC*) and a lysine transporter (*yvsH*) (16-18). Lysine binding to the *lysC* riboswitch was shown to cause structural changes in the aptamer domain and to downregulate gene expression by transcription attenuation (16,17). The lysine riboswitch secondary structure is based on conserved elements previously shown to be reorganized upon lysine binding (Figure 1A)—leading to the anti-antiterminator (P1) stem stabilization and transcription attenuation (16,17,19,20). Recently, lysine riboswitch crystal structures have shown that the riboswitch aptamer is organized around a five-way junction (19,20) with helical stacking occurring between stems P1 and P2 and between stems P4 and P5 (Figure 1A). Both stacks are arranged in an X conformation commonly observed for RNA four-way junctions, where a close juxtaposition is observed between P1 and P5 for the lysine aptamer. The global fold of the riboswitch is defined by the presence of long-range tertiary interactions occurring between loops L2-L3 (L2-L3) and helix P2-loop L4 (P2-L4). The L2-L3 loop-loop interaction was previously shown to be dependent of the kink-turn motif

located in the P2 stem and found to be important for the riboswitch folding and transcription regulation (21). The helix-loop interaction P2-L4 was recently established from crystallographic studies and involves invariant residues of the stem P2 and the loop L4 (Figure 1A). The resulting complex architecture of the riboswitch creates a tight binding pocket that specifically recognizes both charged ends of the bound lysine, thus ensuring high discrimination against lysine structural analogs (19,20). However, due to the presence of a small cavity in the binding site, the riboswitch can accommodate the binding of antimicrobial lysine analogs carrying modifications at position C4 such as *S*-(2-aminoethyl)-L-cysteine (AEC) and L-4-oxalysine (19). Resistance to such compounds is conferred through mutations in the lysine riboswitch aptamer that disrupt critical elements in the core or tertiary interactions (22-24).

We have previously shown that the long-range interaction L2-L3 is important for the transcription regulation of the *lysC* lysine riboswitch (21). Upon the folding of a kink-turn motif located in the P2 stem, loops L2 and L3 interact to reorganize the riboswitch core and to allow premature transcription termination. Because nucleotides located in P2 and P4 directly interact with lysine (19,20), we hypothesized that the P2-L4 interaction would be of paramount importance for the lysine riboswitch function. Thus, to understand how the lysine riboswitch makes use of peripheral elements such as L2-L3 and P2-L4, a series of biophysical and biochemical assays were performed to study the role of these elements on the riboswitch folding and transcriptional activity. Our results show that the folding of either L2-L3 or P2-L4 is optimal when both interactions are functional, as observed from the Mg^{2+} concentration required to fold the aptamer. As reported for L2-L3 (21), we also find that the presence of P2-L4 is very important for the organization of the binding site and for the riboswitch activity. The P2-L4 interaction was also determined to be crucial for a folding transition occurring between stems P1 and P5, as observed in the crystal structure. Interestingly, the P5 stem was found to

functionally replace the P1 stem as the anti-antiterminator, suggesting that various peripheral elements may be used to regulate riboswitch activity. Lastly, we establish that the previously uncharacterized helix P6 found in the expression platform is very important for the adoption of the ON state, indicating that folding elements located outside the aptamer region may also have a strong influence on riboswitch activity.



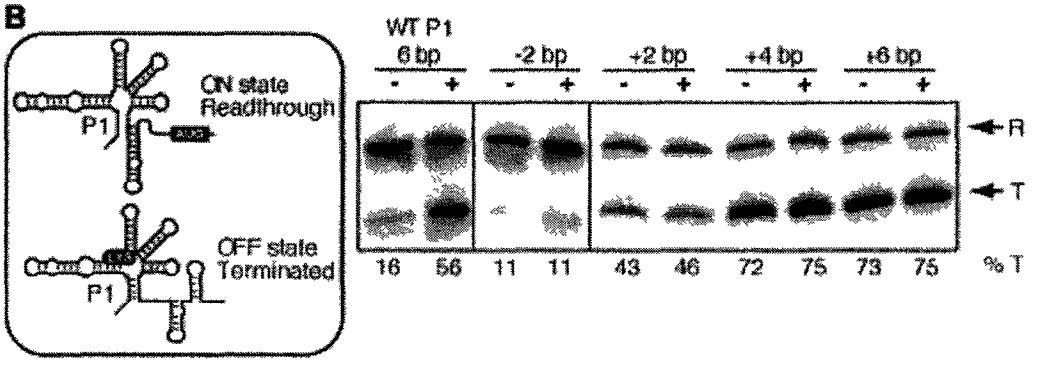


Figure 1. Secondary structure and activity of the lysine riboswitch as a function of lysine. (**A**) Secondary structure representing the lysine riboswitch. Regions involved in the formation of the kink-turn, loop E and terminator are indicated (16-21). Loop-loop L2-L3 and helix-loop P2-L4 interactions are shown by dotted lines and regions involved in the formation of the antiterminator are boxed and indicated by an arrow. Position 194 where a 2AP is introduced is shown by a star. The tertiary structure arrangement is shown in the inset, where discontinuities in the single-stranded regions are represented by dashed lines. The P2-L4 interaction is represented by vertical lines. (**B**) Single-round *in vitro* transcriptions performed as a function of P1 length in absence (-) or in presence (+) of 5 mM lysine. Variations in the total number of base pairs in P1 are indicated. Readthrough and prematurely terminated transcripts are indicated on the right and the percentage of termination is indicated below each reaction lane. Both structures representing the ON and OFF states are shown in the inset where the P1 stem is stabilized in presence of lysine.

MATERIALS AND METHODS

Synthesis of RNA molecules

RNA molecules were transcribed from double-stranded DNA template using T7 RNA polymerase (25). Templates for transcription were made by PCR from synthetic DNA oligonucleotides (Sigma Genosys, Canada). All transcribed RNA species begin with a 5′-GCG sequence to minimize the 5′-heterogeneity of the RNA population (26). RNA molecules containing 2-aminopurine (2AP) were bought from Integrated DNA Technologies (IDT) and purified as described previously (27). The internally-labeled 2AP aptamers were assembled by ligation from oligonucleotides of the following sequences (all written 5′ to 3′):

5′ transcribed strand:

GCGGUGAAGAUAGAGGUGCGAACUUCAAGAGUAUGCCUUUGGAGAAAGAUG
GAUUCUGUGAAAAAGGCUGAAAGGGGAGCGUCGCCGAAGCAAAUAAAACCCC
AUCGGUAUUAUUUGCUGGCCGUGCAUUGAAUAAAUGUAAGGCUGUCAAGAA
GCAUC

3′ strand:

5′-PO_4-AUUGCUUCUUGGAG(2AP)GCUA UCUUCACC.

Mutations in the 5′ transcribed strand were also prepared with different oligonucleotides carrying sequence changes indicated in the text. Reconstituted 2AP-containing riboswitches were prepared as previously described (21). Dual-labeled riboswitches used in FRET assays were prepared by annealing a 5′ transcribed strand:

GCGGUGAAGAUAGAGGUGCGAACUUCAAGAGUAUGCCUUUGGAGAAAGAUG
GAUUCUGUGAAAAAGGCUGAAAGGGGAGCGUCGCCGAAGCAAAUAAAACCCC
AUCGGUAUUAUUUGCUGGCCGUGCAUUGAAUAAAUGUAAGGCUGUCAAGAA
GCAUCGG

and a 3′ synthetic strand:

5′-Cy3-CCGAUGCUUCUUGGAGGGCUAUCUU (5-N-U)CACC

RNA were bought from IDT and purified as described previously (27). The 3′ strand was internally labelled with fluorescein (Invitrogen) at the 5′-amino-allyl uridine nucleotide, according to the manufacturer's protocol. Purified 5′ and 3′ strands were annealed by heating a mixture (molar ratio 1.1:1) to 65°C in 10 mM Hepes, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, and slowly cooled to 30°C. The complex was purified on 8% non-denaturing acrylamide:bisacrylamide (29:1) gel, electro-eluted, precipitated in ethanol and dissolved in water.

Single-round in vitro transcription assays

DNA templates for single-round *in vitro* transcriptions were prepared by PCR with a combination of oligonucleotides corresponding to *Bacillus subtilis glyQS* promoter

fused to *B. subtilis* lysine riboswitch and a 95 nucleotides sequence downstream of the expression platform (21). The promoter region ends by an adenine and is fused to the C17 position of the riboswitch sequence, allowing the transcription to be initiated by the ApC dinucleotide. In the terminated species, the RNA polymerase (RNAP) stops at position 269. The readthrough species allows the RNAP to continue 95 nucleotides further (i.e., 49 nucleotides after the AUG). Reactions were performed and analyzed as previously described (21). Experiments have been performed at least three times and all exhibited very similar uncertainties (< 6%).

2AP fluorescence spectroscopy

Fluorescence spectroscopy was performed on a Quanta Master fluorometer. All data were collected at 10°C in 90 mM Tris-borate, pH 8.3 and 100 mM KCl and analyzed as performed previously (21). Briefly, excitation for 2AP fluorescence was done at 300 nm to obtain a good separation between the Raman peak and the 2AP fluorescence signal. The fluorescence data are fitted to a simple two-state model that assumes an all-or-none conformational transition induced by the binding of magnesium ions, with a Hill coefficient *n* and an apparent association constant K_A. The proportion of folded aptamer (α) is given by:

$$\alpha = K_A \bullet [Mg^{2+}]^n / (1 + K_A \bullet [Mg^{2+}]^n)$$

As the parameters K_A and *n* co-vary, we present the $[Mg^{2+}]_{1/2} = (1/K_A)^{1/n}$, which gives a robust estimate for the affinity of magnesium ions (27,30). Experiments have been performed at least three times and all exhibited very similar uncertainties (< 5%).

FRET analysis

Fluorescence spectroscopy was performed on a Quanta Master fluorometer and spectra were corrected for lamp fluctuations and instrument variations as described (27). Polarization artifacts were avoided by setting excitation and emission polarizers

crossed at 54.75°. Values of E_{FRET} were determined using the acceptor normalization method (31,32). All data were collected at 4°C in 90 mM Tris-borate, pH 8.3 and 100 mM KCl. Fluorescein and cyanine 3 fluorophores were excited at 490 nm and 547 nm, respectively. Data were collected using 10 pmol of fluorescent dual-labelled riboswitch and magnesium ions were titrated over the range indicated. Experiments have been performed at least three times and all exhibited very similar uncertainties (< 5%).

Native gel electrophoresis

[5′-^{32}P] labeled lysine riboswitch aptamers were incubated at 70°C for 3 min and allowed to slowly cool to room temperature to ensure homogenous folding. Samples were then electrophoresed in 90 mM Tris-borate, pH 8.3 and 1 mM $MgCl_2$ in 8% acrylamide:bisacrylamide (29:1) gel in Tris-borate buffer containing 1 mM $MgCl_2$ at room temperature at 150 V for 18 h with the running buffer circulated during the electrophoresis. The gels were exposed to PhosphorImager screens and imaged.

SHAPE analysis

Wild-type and mutant riboswitch aptamers were prepared as previously described, with the addition of the sequence GGAGCUAAAGAGAAGCGGAAG at the 3′ end to allow binding of a complementary DNA oligonucleotide primer. Each reaction was prepared using 1 pmol of purified aptamer resuspended in 2 volumes of 0.5X TE buffer, in which was added 1 volume of 3.3X folding buffer containing 333 mM K-HEPES, pH 8.0, 333 mM NaCl and the desired concentration of $MgCl_2$. Samples were heated to 65°C and allowed to slowly cool to 30°C before being pre-incubated 10 minutes at 37°C. *N*-methylisatoic anhydride (Invitrogen) dissolved in DMSO was then added and allowed to react for 80 minutes at 37°C. Modified RNA were ethanol precipitated, washed with 70% ethanol and resuspended in 0.5X TE buffer. For SHAPE analysis in the presence of ligand, L-lysine (Sigma) was added to a final concentration of 5 mM, and the final 100 mM NaCl concentration was replaced by 75

mM NaCl and 25 mM KCl. Reverse transcription reactions were performed as previously described (33) and reverse transcribed products were separated on 5% denaturing polyacrylamide gel. The gels were exposed to PhosphorImager screens and imaged. Individual band intensities in Figure 3B were integrated using SAFA (34).

RESULTS

The P1 stem stability is crucial for lysine riboswitch transcription regulation

With the exception of the glmS ribozyme *(35)*, all reported riboswitches control gene expression by modulating the formation of a helical domain, the P1 stem, which is of central importance for the folding of the expression platform and riboswitch activity. In a biological context, the P1 stem of the lysine riboswitch is not folded in absence of lysine, which allows gene transcription to occur (ON state, Figure 1B, inset). However, upon lysine binding, the P1 stem is stabilized and premature transcription termination is favored (OFF state). Examination of crystal structures suggests that the lysine riboswitch uses complex folding elements to recognize lysine and to achieve ligand-dependent P1 stem stabilization. To investigate whether P1 stem stability is sufficient for transcription attenuation or whether lysine-induced riboswitch core reorganization is required, we varied the stability of the P1 stem of the *B. subtilis lysC* lysine riboswitch and studied its effect on premature transcription termination.

Single-round *in vitro* transcriptions were carried out using a DNA template containing the *B. subtilis glyQS* promoter upstream of the lysine riboswitch and a 95 nucleotide sequence downstream of the terminator. When performing transcription reactions using *E. coli* RNAP in absence of lysine, a low transcription termination efficiency of

16% was observed (Figure 1B). However, when 5 mM lysine was present in the transcription reaction, 56% premature termination was obtained indicating that ligand binding to the lysine riboswitch modulates transcription attenuation *in vitro*, as previously reported (16,17,21,22). To investigate the influence of the P1 stem on riboswitch activity, we increased P1 length by 2 bp and observed a disrupted riboswitch regulation regardless of the presence of the ligand (≥43%; Figure 1B). Further elongation of the P1 stem by 4 bp and 6 bp yielded similar high termination efficiencies (≥72%) suggesting that the riboswitch is locked in the OFF state by the high stability of P1 (Figure 1B). In contrast, when 2 bp were removed from this stem (-2 bp, Figure 1B), a reduced termination efficiency of only 11% was obtained independently of lysine's presence. Together, these results suggest that the formation of the RNA-ligand complex is not required to yield premature transcription termination. Indeed, the modulation of the P1 stem stability is sufficient to lock the riboswitch either in the ON or OFF state resulting in significant differences in basal expression levels.

The P2-L4 interaction is important for *lysC* riboswitch activity

Our laboratory has previously shown that stem-loops P2 and P3 are involved in a loop-loop interaction that is important for the riboswitch structure (21). An additional tertiary interaction involving conserved elements from P2 and L4 has recently been discovered (19,20), which is of particular interest since both P2 and P4 stem-loops directly interact with lysine. Thus, to determine the importance of the P2-L4 interaction for riboswitch function, a series of riboswitch constructs were analyzed using single-round *in vitro* transcription.

Adenines at positions A156 and A157 in L4 are very conserved (16-18) and interact with G98 and G99 of the P2 helical domain (Figure 1A). Deletion of A156

significantly reduced premature termination in absence as well as in presence of lysine, clearly indicating that the P2-L4 interaction is important for lysine-mediated riboswitch activity (Figure 2B). Moreover, when both adenines 156 and 157 were replaced with cytosines (L4 mutant), a similar reduction in premature termination was observed (Figure 2B), consistent with the conserved nature of these nucleotides. A similar result was obtained when introducing a G98A/G99A mutation in the P2 helical domain (P2 mutant; Figure 2B).

The importance of the P2-L4 interaction was also assessed by modulating the length of the P4 stem as we anticipated that it would preclude the P2-L4 interaction to take place. As expected, the riboswitch activity was severely compromised with a complete disruption in lysine-mediated transcription termination when the P4 stem was either reduced to 4 bp or elongated to 19 bp (Figures 2A and 2B). Taken together, these results are consistent with the formation of the P2-L4 tertiary interaction being functionally important for the *B. subtilis lysC* lysine riboswitch.

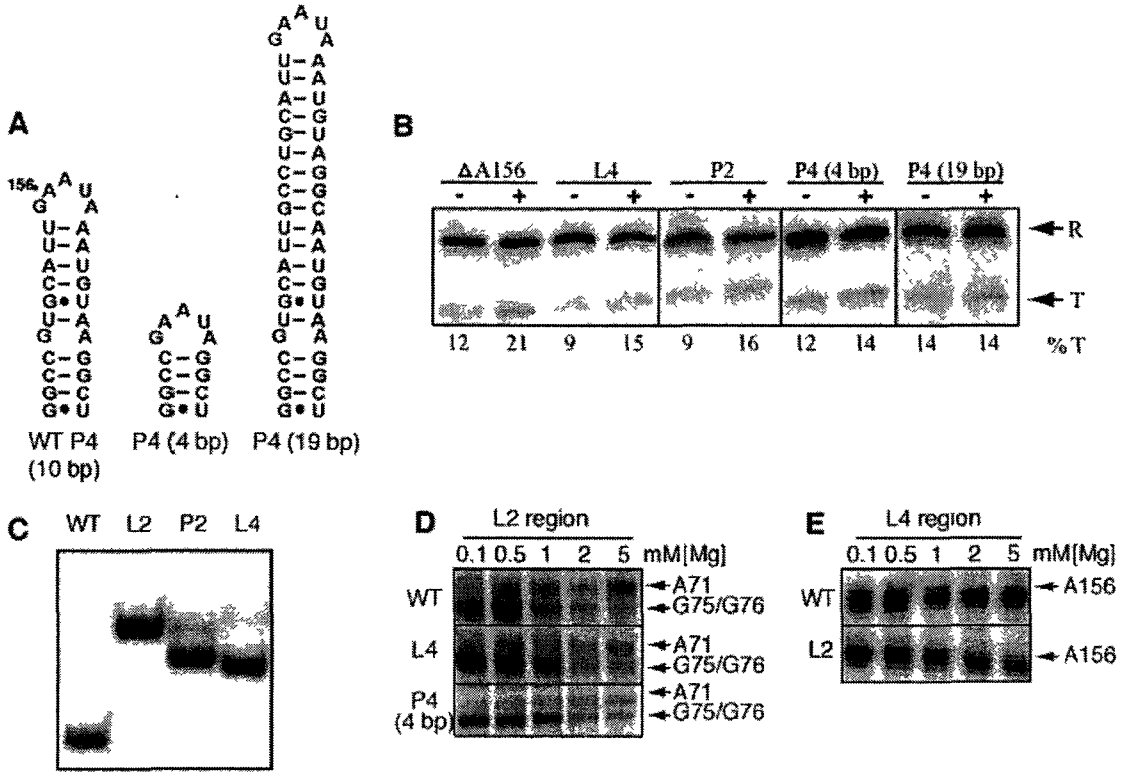


Figure 2. The P2-L4 interaction is important for the folding and activity of the riboswitch. (**A**) Secondary structure of the 4 bp and the 19 bp P4 stem mutants. (**B**) Single-round in *vitro* transcription assays made using selected P4 riboswitch mutants in absence (-) or in presence (+) of 5 mM lysine. Reactions were performed for the A156-deleted variant (ΔA156), A156C/A157C (L4), G98C/G99C (P2), P4 (4 bp) and P4 (19 bp). Readthrough (R) and prematurely terminated (T) transcripts are indicated on the right and the percentage of termination (%T) is indicated below each reaction lane. Vertical bars separate results obtained from different gels. (**C**) Non-denaturing gel electrophoresis of the wild-type (WT) and lysine aptamer mutants in the presence of 1 mM magnesium ions. The mutants L2 (G72C/A73U/U74A/G75C/G76C), P2 and L4 disrupt one of either L2-L3 or P2-L4 tertiary interactions, which results in a slower electrophoretic migration rate. (**D**) and (**E**) SHAPE modification of the lysine riboswitch performed at various magnesium ion concentrations. SHAPE data show that disruption of the P2-L4 interaction increases the magnesium requirement to induce the L2-L3 interaction (**D**) and that the disruption of the L2-L3 interaction increases the magnesium requirement to form the P2-L4 interaction (**E**). Arrows show SHAPE-reactive sites that are significantly modulated by magnesium ions.

Long-range interactions L2-L3 and P2-L4 define the global fold of the riboswitch

The tertiary structure of the lysine riboswitch comprises the long-range tertiary interactions L2-L3 and P2-L4 that are important for the transcriptional regulation (Figure 2B and (21)). To investigate how interactions L2-L3 and P2-L4 influence the

riboswitch global structure, we employed the native gel assay, a technique that has been used to study global conformational changes in nucleic acids (27,36-40). Given that a compact structure migrates faster than an unfolded one, which is generally more extended, we speculated that aptamers correctly folded should display a faster migration rate compared to those incorrectly folded.

Clear differences in the migration pattern in a native polyacrylamide gel with 1 mM magnesium ions were detected between wild-type and several mutants, suggesting a disruption in distinct tertiary structures (Figure 2C). A riboswitch aptamer having a disabled L2-L3 interaction (mutant L2) exhibited a slower migration when compared to the wild-type sequence (Figure 2C). A precedent for this migration behavior has been observed for the adenine riboswitch, where the disruption of the loop-loop interaction reduced the migration rate in native gels (27,38). Disruption of the P2-L4 interaction (P2 and L4 mutants) also reduced the aptamer migration rate, albeit not as much as the L2 mutant (Figure 2C). Thus, our results are consistent with both L2-L3 and P2-L4 tertiary interactions being important for the adoption of the global structure of the aptamer.

The L2-L3 and P2-L4 interactions are interdependently folding units

Although our native gel data suggest that both L2-L3 and P2-L4 tertiary interactions can be individually disrupted to alter the global riboswitch structure, we wanted to gain more quantitative information about the folding requirement of each interaction. We thus used selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) to obtain information about the folding of the lysine aptamer (33). This technique is particularly relevant to discriminate local nucleotide flexibility against constrained RNA regions, where 2'-hydroxyl groups in flexible regions are prone to react with electrophiles like *N*-methylisatoic anhydride (NMIA). SHAPE chemistry is therefore

an excellent tool to look closely at the formation of tertiary interactions in the lysine riboswitch, as previously reported (20). When subjected to NMIA reaction as a function of Mg^{2+} concentration, the lysine riboswitch aptamer exhibited clear changes in NMIA reactivity, especially in regions involved in L2-L3 and P2-L4 long-range interactions (Figure S1A). The relative reaction intensity was plotted as a function of Mg^{2+} to calculate the concentration required to obtain half of the change ($[Mg^{2+}]_{1/2}$, see Materials and Methods). This analysis was done for the L2, L3, P2 and L4 regions of the riboswitch that are involved in L2-L3 and P2-L4 tertiary interactions (Table 1). While L2 and L3 gave very similar results (0.82 mM and 0.85 mM respectively), P2 and L4 gave slightly different $[Mg^{2+}]_{1/2}$ values (1.11 and 0.59 mM, respectively). However, partial nuclease digestion of an isolated P4 stem-loop revealed that the L4 region undergoes an independent Mg^{2+}-induced structural rearrangement indicating that the $[Mg^{2+}]_{1/2}$ value obtained for L4 in the context of the riboswitch probably reflects multiple folding processes (S. Blouin and D.A. Lafontaine, unpublished results). Nevertheless, SHAPE results obtained for L2, L3 and P2 regions clearly suggest that both long-range tertiary interactions fold with magnesium ions in the low mM range.

We next studied the L2-L3 interaction as a function of Mg^{2+} concentration in the context of a riboswitch exhibiting a compromised P2-L4 interaction (Figure 2D). SHAPE experiments performed on both wild-type and L4 mutant molecules clearly indicated that the interaction L2-L3 required more magnesium ions to fold in absence of P2-L4. For instance, although the wild-type aptamer showed protection against NMIA reaction for positions G75/G76 in presence of 1 mM Mg^{2+}, the L4 mutant exhibited such protection only when using 2 mM Mg^{2+} (Figure 2D). Similarly, the position A71 did not react to NMIA at 5 mM Mg^{2+} in the context of the L4 mutant, as observed for the wild-type aptamer. Very similar results were obtained when using an aptamer mutant having the P4 stem reduced to 4 bp (P4 mutant, Figure 2D).

Table 1. SHAPE modification of the lysine riboswitch as a function of magnesium ions

Riboswitch region	$[Mg^{2+}]_{1/2}$ (mM)
L2[a]	0.82
P2[b]	1.11
L3[c]	0.85
L4[d]	0.59

Half of apparent concentrations for the magnesium-induced folding process ($[Mg^{2+}]_{1/2}$) were measured under standard conditions. Asymptotic standard deviations on the fits indicate that errors in ($[Mg^{2+}]_{1/2}$) are generally less than 10% of values.

[a]Nucleotides A71, G75 and G76 were monitored.

[b]Nucleotide G98 was monitored.

[c]Nucleotide A128 was monitored.

[d]Nucleotide A156 was monitored.

We also analyzed the P2-L4 interaction as a function of the L2-L3 tertiary interaction (Figure 2E). Similarly to our results for the L2-L3 interaction, the P2-L4 interaction also required more magnesium ions to fold when L2-L3 was disabled, as deduced from the absence of protection of A156 at 1 mM Mg^{2+}. Taken together, these results suggest that although both L2-L3 and P2-L4 interactions can fold independently of each other, they require less Mg^{2+} to form when both interactions are functional.

The G39C mutation inhibits binding without disrupting the riboswitch global folding

We and others previously characterized a riboswitch core mutation (G39C) known to cause *lysC* derepression and to severely perturb the lysine riboswitch transcriptional control *in vitro* (17,21,23,41). According to the recently crystalized riboswitch structures (19,20), G39 interacts with lysine via its 2'OH in the bound conformation. To characterize the influence of G39 on the folding of the riboswitch, we analyzed the influence of this core mutation on the lysine aptamer using native gel electrophoresis. In contrast to what we obtained when disrupting interactions L2-L3 or P2-L4, the introduction of a G39C mutation did not result in a retarded gel migration when compared to the wild-type molecule (Figure 3A), suggesting that both L2-L3 and P2-L4 interactions can form in this context.

We next studied the influence of G39C on the lysine-induced folding of the aptamer using SHAPE assays (Figure 3B). To establish if ligand binding can assist in aptamer folding, we used a Mg^{2+} concentration of 0.5 mM that does not allow formation of interactions L2-L3 and P2-L4 (Figures 2D and 2E). When the wild-type molecule was subjected to NMIA reaction using a concentration of 5 mM lysine, clear changes could be observed in regions involved in tertiary interactions L2-L3 and P2-L4 (Figure 3B). Additional reactions were observed in the aptamer core region that constitutes the lysine binding pocket, such as G111, A112, and G193 (19,20). Moreover, significant lysine-induced NMIA reactivity were obtained in the P5 stem, as previously reported (20). In contrast, the mutant G39C did not show significant changes in presence of lysine (Figure 3C), consistent with its negative effect on transcription regulation (17,21).

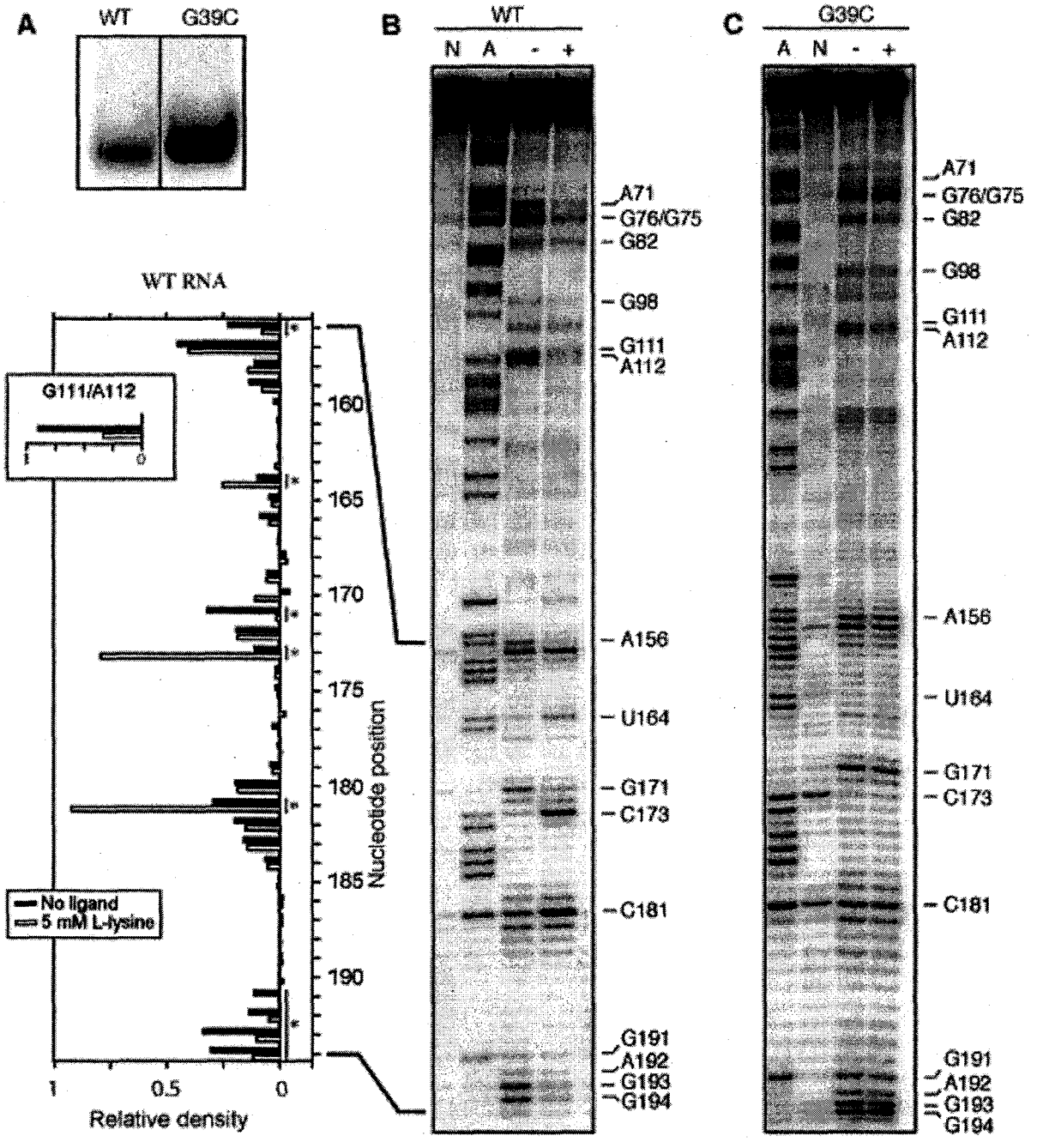

A
WT
G39C
WT RNA
G111/A112
No ligand
5 mM L-lysine
Relative density
Nucleotide position
B
WT
N A - +
A71
G76/G75
G82
G98
G111
A112
A156
U164
G171
C173
C181
G191
A192
G193
G194
C
G39C
A N - +
A71
G76/G75
G82
G98
G111
A112
A156
U164
G171
C173
C181
G191
A192
G193
G194

Figure 3. The G39C mutation inhibits binding without disrupting the global folding of the riboswitch. (**A**) Non-denaturing gel electrophoresis of the wild-type (WT) and G39C mutant in presence of 1 mM magnesium ions. The G39C mutant exhibits a migration similar to the wild-type, indicating that it does not perturb the global folding of the riboswitch. Note that both lanes were taken from the same gel but were manually juxtaposed to clearly indicate the co-migration of both aptamers. (**B**) SHAPE modification of the lysine riboswitch performed on the wild-type aptamer (WT). Reactions were performed at 0.5 mM $MgCl_2$, in absence (-) and in presence (+) of 5 mM L-lysine. N and A represent unreacted RNA and an adenine sequencing lane, respectively. Nucleotides of the wild-type aptamer exhibiting a lysine-induced modification are indicated on the right of the gel. The histogram on the left represents the relative density of bands corresponding to nucleotides 156 to 194, where filled and empty bars represent quantifications in absence and in presence of lysine, respectively. Positions showing significant difference in lysine-induced modification are marked by an asterisk. The inset represents a quantification of the reactivity of nucleotides G111 and A112. (**C**) SHAPE modification of the lysine riboswitch performed for the G39C aptamer mutant. Reactions were performed at 0.5 mM $MgCl_2$, in absence (-) and in presence (+) of 5 mM L-lysine. N and A represent unreacted RNA and an adenine sequencing lane, respectively.

The P2-L4 interaction organizes the riboswitch binding pocket

We have previously employed the fluorescent local reporter 2AP to show that the formation of the L2-L3 long-range tertiary interaction is important for the lysine riboswitch core folding (21). In these studies, 2AP was introduced at the non-

conserved position 194 (Figure 1A), which, according to the crystal structure of the *Thermotoga maritima* lysine riboswitch variant (19,20), is exposed to the solvent and makes stacking interaction with an extra bulged nucleotide found in the P3 helical domain (19,20). Given that the bulged nucleotide is not present in the *B. subtilis lysC* riboswitch, it is expected that G194 does not have significant interaction with the rest of the core of the riboswitch, as suggested by its non-conserved nature.

2AP fluorescence was used to characterize the influence of the P2-L4 interaction on the folding of the riboswitch core domain. The addition of Mg^{2+} to a 2AP-labeled wild-type lysine aptamer resulted in an increase in fluorescence (Figure 4), which is consistent with position 194 being exposed to the solvent in the crystal structure (19,20). Fluorescence changes were fitted using a simple two-state transition model that allowed us to estimate values for apparent Mg^{2+}-binding parameters. With this model, an apparent affinity for magnesium ions $[Mg^{2+}]_{1/2}$ of 1.9 mM and a Hill coefficient n = 1.6 ± 0.1 were obtained suggesting that the transition exhibits a cooperative character, which is consistent with our previous findings (21). The disruption of the P2-L4 interaction (L4 mutant) severely perturbed the 2AP folding transition, as opposed to the folding of the L2-L3 interaction (Figure 2E), suggesting that the P2-L4 interaction is very important for the reorganization of the riboswitch core (Figure 4). As a comparison, we also performed the same experiment using a riboswitch having a G39C mutation that strongly inhibits riboswitch activity but that does not perturb global folding (Figure 3A). As found for the L4 mutant, almost no change in 2AP fluorescence was observed as a function of Mg^{2+}, indicating that the local folding is also strongly perturbed. Thus, together with native gel data, our 2AP fluorescence results show that formation of a P2-L4 long-range interaction is very important for riboswitch global folding and core reorganization to ensure lysine-dependent riboswitch activity (Figure 2B).

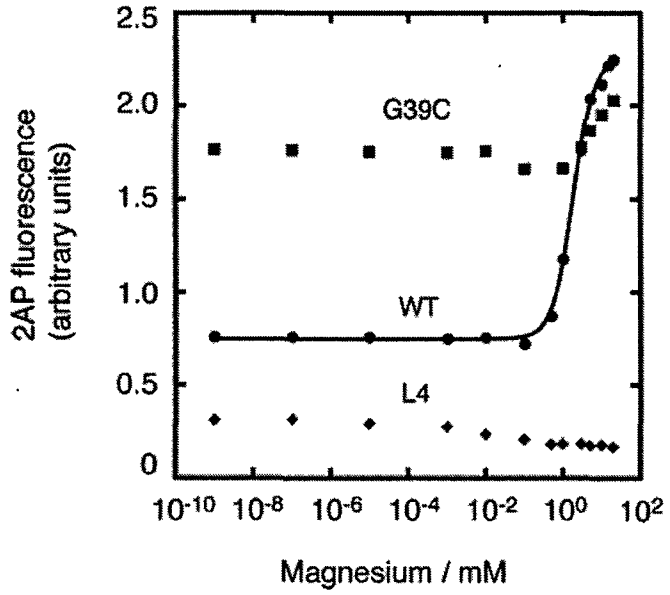


Figure 4. The formation of the P2-L4 interaction is important for riboswitch core folding. Normalized 2AP fluorescence intensity plotted as a function of magnesium ions for the WT (circles), the P2-L4-deficient L4 mutant (diamonds) and the G39C mutant (squares). The experimental data were fitted by regression to a simple two-state model where the binding of metal ions to the aptamer induces a structural change. Note that no significant change is observed when using either the L4 or G39C mutant.

The P5 stem is important for riboswitch activity

The P5 stem is the least conserved helical domain of the lysine riboswitch where only the three-first base pairs located in proximity to the junction are conserved (16-18,22). Crystal structures of the lysine riboswitch show that the P5 stem is juxtaposed to the 3′ strand of P1 (19,20). Interestingly, it has been previously reported that P5 is reactive to structural probing depending of the presence of lysine,

suggesting that P5 is involved in a riboswitch structural reorganization (17,19,20). Single-round transcription assays were thus used to determine the importance of the P5 stem for the *lysC* riboswitch activity. A riboswitch construct having the P5 stem-loop reduced to 5 bp exhibited poor premature termination efficiency regardless of lysine's presence (Figures 5A and 5B). However, a P5 stem increased to 16 bp yielded higher premature transcription termination in absence or presence of lysine (Figure 5B). These results show that the stability of the P5 stem correlates with the riboswitch transcription termination efficiency, which is reminiscent of what we observed for the P1 stem (Figure 1B). These results suggest that P5 is important for the riboswitch regulation mechanism.

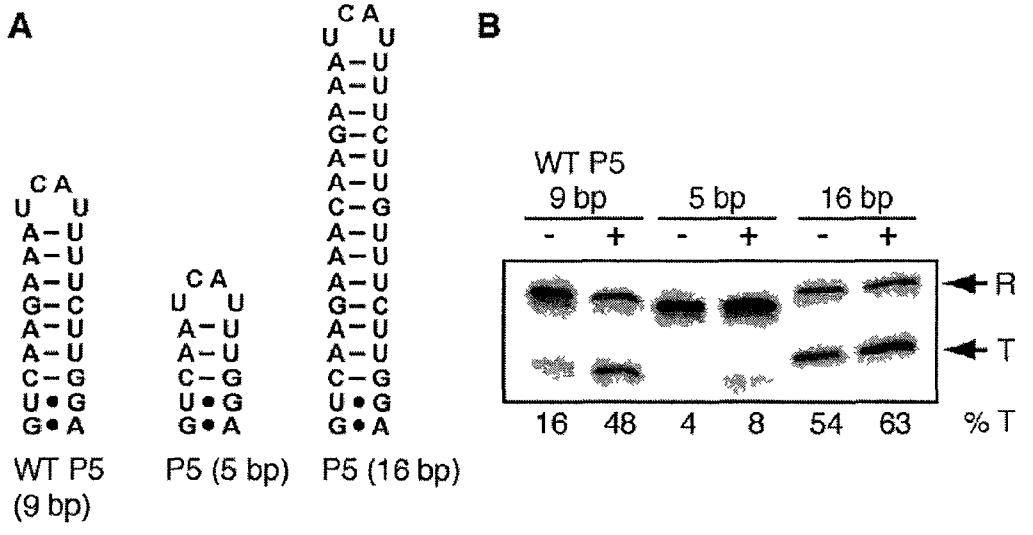


Figure 5. The P5 stem is important for lysine riboswitch activity. (**A**) Secondary structures of the 5 bp and the 16 bp P5 stem mutants. (**B**) Single-round *in vitro* transcription assays made using selected P5 riboswitch mutants in absence (-) or in presence (+) of 5 mM lysine. Reactions were performed for the wild-type riboswitch, the P5 (5 bp) and the P5 (16 bp) mutants. Readthrough (R) and prematurely terminated (T) transcripts are indicated on the right and the percentage of termination (%T) is indicated below each reaction lane.

A P1-P5 folding transition is observed by Fluorescence Resonance Energy Transfer

Although crystal structures generate invaluable information about folded state(s) of the aptamer domain, they do not provide details regarding the structural sampling that riboswitches may experience for efficient ligand binding. However, fluorescence resonance energy transfer (FRET) (36,42) is especially informative in the study of conformational changes of nucleic acids in solution and provides quantitative data on global structure and conformational transitions (27,43,44). In contrast to most riboswitches studied, the lysine riboswitch is especially intriguing given that it exhibits very similar structures regardless of the presence of Mg^{2+} or its cognate ligand (19,20).

We therefore introduced fluorophores in helices P1 and P5 and monitored the efficiency of FRET (E_{FRET}) between both helices (Figure 6A). In these experiments, the aptamer sequence was minimally altered to allow the introduction of fluorophores and assure the stability of the fluorescent construct. Since P1 and P5 are in close proximity in crystal structures, we predicted that the folded state would produce a large E_{FRET} value. As expected, addition of Mg^{2+} resulted in an E_{FRET} increase (Figure 6A), consistent with the location of stems P1 and P5 in the folded riboswitch structure. When the data are fitted to a simple two-state model that assumes an all-or-none conformational transition, we can calculate the magnesium ion concentration at which the transition is half complete. In absence of lysine, the data fitted values of $[Mg^{2+}]_{1/2}$ = 0.5 mM and n = 1.3 ± 0.1, suggesting that the transition does not occur in a cooperative manner (Table 2). When the experiment was repeated in presence of 5 mM lysine, a different transition was observed exhibiting values of $[Mg^{2+}]_{1/2}$ = 0.05 mM and n = 0.79 ± 0.10, suggesting that the presence of lysine decreases the Mg^{2+} requirement by 10-fold (Table 2). These results indicate that lysine binding is important for the close juxtaposition occurring

between P1 and P5, which is in agreement with structural probing data (17,19,20).

Table 2. Efficiency of FRET of the P1-P5 vector for various lysine riboswitch mutants

Variant	$[Mg^{2+}]_{1/2}$ (mM)	Hill coefficient (*n*)
WT	0.50	1.3 ± 0.1
WT + lysine	0.05	0.79 ± 0.10
L2[a]	0.07	0.75 ± 0.07
L4[b]	n.d.	n.d.
G39C	2.78	0.5 ± 0.2

Half of apparent concentrations for the magnesium-induced folding process ($[Mg^{2+}]_{1/2}$) were measured under standard conditions. Asymptotic standard deviations on the fits indicate that errors in ($[Mg^{2+}]_{1/2}$) are generally less than 10% of values. Values were not determined for the L4 mutant as no significant change was observed.

[a]Nucleotides G72-G76 replaced with their Watson-Crick complement.

[b]Nucleotides A156-A157 were replaced with cytidines.

To determine the influence of riboswitch long-range interactions on the P1-P5 FRET transition, we introduced a series of mutations in the lysine riboswitch and studied their effects on P1-P5 folding (Figure 6B). Interestingly, the three riboswitch mutants gave distinct folding profiles. For instance, while the disruption of the L2-L3 interaction (L2 mutant) slightly altered the P1-P5 folding, the introduction of a G39C mutation significantly reduced the ability of P1 and P5 to fold (Figure 6B and Table 2). However, the disruption of the P2-L4 interaction (L4 mutant) completely

abolished the folding transition, indicating that P2-L4 is crucial in the P1-P5 close juxtaposition.

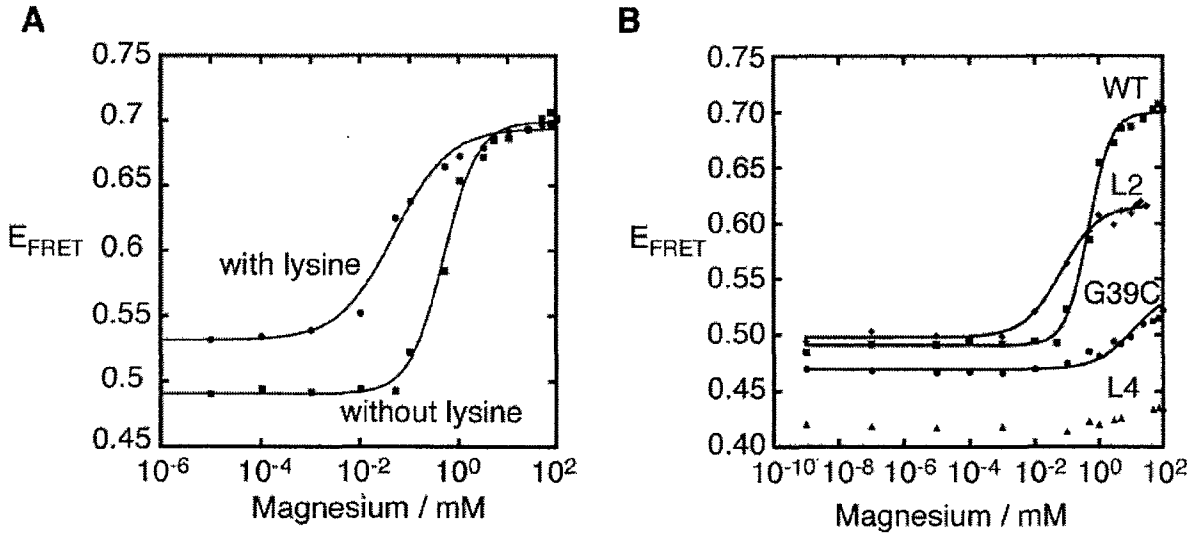


Figure 6. Lysine riboswitch folding monitored using a P1-P5 FRET vector. (**A**) Plot of the efficiency of FRET as a function of magnesium ions in the absence (squares) or the presence (circles) of 5 mM lysine. The FRET increase corresponds to a shortening of the distance between fluorophores attached to P1 and P5 stems. The data were fitted by regression to a simple two-state model in which the binding of Mg^{2+} to the aptamer induces a structural change. (**B**) Plot of the efficiency of FRET as a function of magnesium ion concentration for the WT molecule (squares), the L2 mutant (diamonds), the G39C mutant (circles) and the L4 mutant (triangles) in absence of lysine. Note that no significant change is observed when using the L4 riboswitch variant.

The P5 stem can control the riboswitch activity

Our single-round *in vitro* transcription data show that the transcription termination is proportional to P5 stem stability (Figure 5B), which is reminiscent of our results obtained for the P1 stem (Figure 1B). Thus, given that the riboswitch activity can be modulated by altering the stability of P5, we speculated that a riboswitch construct in which the expression platform is fused to the P5 stem could control transcription termination as a function of lysine (Figure 7A). When assessing such a construct using single-round *in vitro* transcription assays, a low premature transcription of 21% was obtained in absence of lysine (Figure 7B, left panel). However, the addition of lysine resulted in a significant increase in premature termination (41 %), consistent with the idea that the engineered riboswitch can modulate transcription.

Next, because the stabilization of P1 in the wild-type construct resulted in increased premature termination (Figure 1B), we speculated that the stabilization of P5 in this alternative construct should have the same effect. We thus increased the length of the P5 stem to 10 bp and used this construct in single-round transcription assays (Figure 7B, right panel). A marked increase in premature transcription termination (82%) was observed independently of lysine, indicating that the stabilization of P5 can efficiently terminate transcription prematurely. Taken together, these results support the idea that the expression platform can be relocated in different regions of the riboswitch and still allow for riboswitch activity.

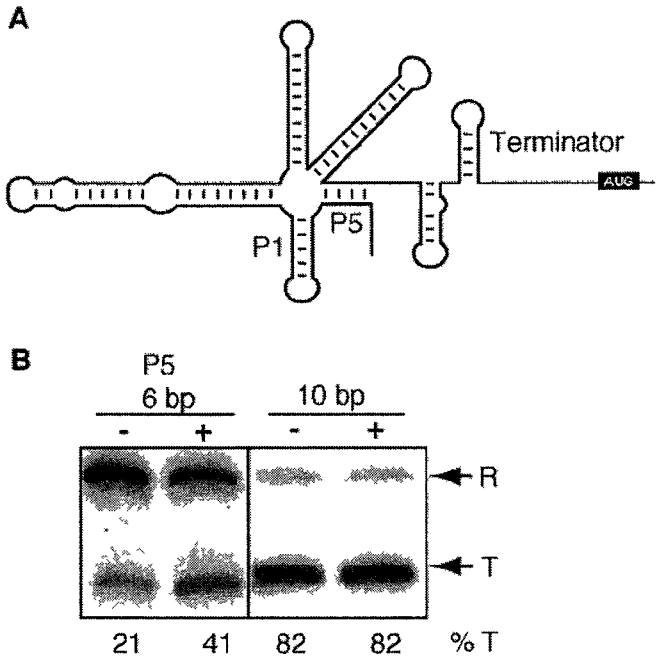


Figure 7. The P5 stem can be used as an anti-antiterminator. (**A**) Schematic of the lysine riboswitch in which the expression platform is fused to the P5 stem. (**B**) Single-round *in vitro* transcriptions were performed using constructs with a P5 stem having either 6 bp or 10 bp in the absence (-) and the presence (+) of 5 mM lysine. Readthrough (R) and prematurely terminated (T) transcripts are indicated on the right and the percentage of termination (%T) is indicated below each reaction lane.

The P6 stem positively influences antitermination

In this work, we have studied the importance of aptamer structural elements for the riboswitch transcription regulation. However, although the aptamer is crucial for the riboswitch function, the expression platform is also very important given that it is directly involved in genetic control (16,17,19-21). Upon examination of the *lysC* expression platform (Figure 1A), an additional stem-loop motif is found between P1

and the terminator, which we named P6. Based on the riboswitch secondary structure, the role of the P6 stem is not readily apparent as it could either be involved in the formation of the ON or OFF state. For example, P6 could be involved in a stacking interaction either with the antiterminator stem, which could stabilize the ON state, or with the terminator helix, which could favor the OFF state.

To determine if P6 has any influence on the riboswitch activity, various constructs differing in their P6 stability were engineered and used in single-round *in vitro* transcription assays (Figure 8A). When using a riboswitch in which three base pairs of the P6 stem were inhibited, thus making a 7 bp stem, a significant increase in termination was obtained both in absence (47%) and in presence of lysine (62%; Figure 8B). Higher termination efficiencies were also obtained when decreasing the P6 stem to 4 bp (≥67%) or 0 bp (≥77%). In contrast, when a more stable P6 helix (13 bp) was used, a marked reduction in termination efficiency was observed (≤14%; Figure 8B). Thus, our *in vitro* transcription data suggest that the formation of P6 stem is important for antitermination to take place, clearly suggesting a crucial role for P6 in the riboswitch function.

To assess to which extent the formation of P6 can inhibit the formation of the terminator structure, we first engineered two constructs in which the terminator stem was elongated to 12 bp or 14 bp. As expected, single-round transcription assays using riboswitches with either a 12 bp or a 14 bp terminator gave high termination efficiency (Figure 8C, left panel). When these experiments were repeated using a construct in which P6 was elongated to 13 bp, a significant decrease in transcription termination was observed when using a 12 bp terminator (Figure 8C, right panel), consistent with P6 promoting antitermination. However, no significant change was observed when the terminator was elongated to a 14 bp. Thus, these results indicate that the stability of both the P6 stem and terminator is important for the proper riboswitch regulation mechanism.

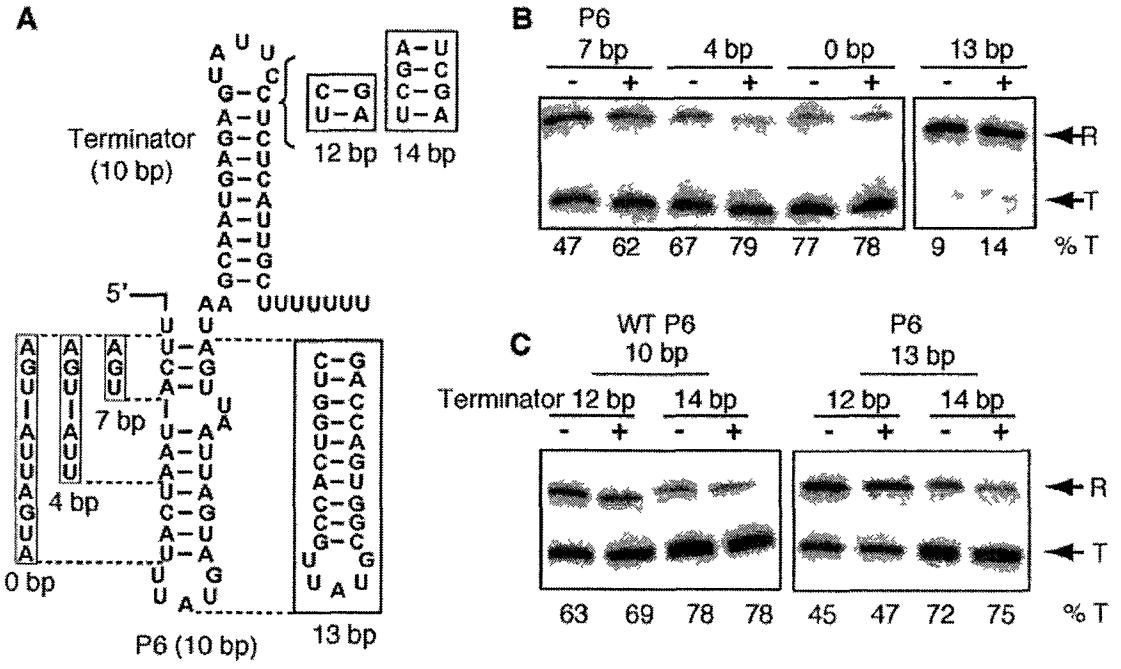


Figure 8. The structure of the P6 stem is important for transcription termination. (**A**) Secondary structure of P6 and terminator mutants. Mutated sequences for the 0 bp, 4 bp and 7 bp P6 mutants are boxed and only replace the corresponding 5′ side of P6 enclosed between dotted lines. The 13 bp P6 stem is also boxed and replaces the entire wild-type P6 stem enclosed between dotted lines. The location and the additional base pairs introduced in the terminator stem (12 bp and 14 bp mutants) are also shown. (**B**) Single-round *in vitro* transcription assays were made using selected P6 mutants in absence (-) or in presence (+) of 5 mM lysine. Reactions were performed for the 0 bp, 4 bp, 7 bp and the 13 bp P6 mutants. (**C**) Single-rounds *in vitro* transcriptions assays were made using selected terminator stem mutants in absence (-) or in presence (+) of 5 mM lysine. Reactions were performed for the 12 bp and 14 bp terminator stem mutants either in the context of the wild-type 10 bp (left panel) or the 13 bp (right panel) P6 stem. Readthrough (R) and prematurely terminated (T) transcripts are indicated on the right and the percentage of termination (%T) is indicated below each reaction lane.

DISCUSSION

Riboswitches are exquisite RNA molecular sensors that are able to discriminate against closely similar cellular targets, which often involve the formation of high affinity RNA-ligand complexes. Although riboswitch ligand binding sites are constituted of highly conserved nucleotides, it is expected that less conserved peripheral elements are also important for the adoption of the native RNA structure. Perhaps one of the best example concerning the importance of peripheral elements lies in the hammerhead ribozyme. This ribozyme has been considered for several years as a simple three-way RNA junction in which the folding pathway was entirely driven by the magnesium ion-dependent junction folding (45). However, recent findings have shown that a critical loop-loop interaction is important to achieve catalysis at low magnesium ion concentrations (46-48), and, importantly, the comparison of hammerhead crystal structures obtained both in absence (49,50) and in presence (51) of the loop-loop interaction shows that a significant structural reorganization of the hammerhead core region is caused by the formation of the tertiary interaction. This is consistent with long-range tertiary interactions being important for the folding of core regions, which is of particular interest for riboswitches as they often employ core regions to form ligand binding sites (2).

The structure of the lysine riboswitch is defined by the presence of two long-range tertiary interactions that are proximally juxtaposed, and that are scaffolded through a five-way junction in which the lysine binding site is centrally located (Figure 1A). Recent crystal structures obtained in absence or in presence of lysine shown that tertiary interactions can be assembled independently of the ligand, suggesting that these interactions can fold the binding site at the junctional core, as we (21) and others (19,20) have shown. The work described here focuses on the role of peripheral elements on the riboswitch folding and lysine-dependent regulation control.

Importance of the P2-L4 interaction for riboswitch regulation

SHAPE data indicate that both L2-L3 and P2-L4 interactions require less magnesium ions to fold when both interactions are functional (Figures 2D and 2E), suggesting that the formation of each interaction can reduce the conformational space that the remaining one has to sample to fold. In a transcriptional context, the polarity of RNA polymerization suggests that L2-L3 forms first to subsequently assist in P2-L4 formation (Figure 1A). Importantly, the P2-L4 interaction is crucial for the riboswitch core folding as revealed by 2AP fluorescence assays where no fluorescence increase was observed when altering the L4 loop sequence (Figure 4). Because 2AP is located at position 194 and that 2AP fluorescence is very sensitive to its local environment, we expect the formation of base pairs A38•G195 and G39•G193, which are in proximity to the ligand binding site, to have a significant effect on 2AP fluorescence (Figure 9A). The marked decrease of fluorescence emission when P2-L4 is disrupted suggests that the latter is important for the correct folding of the riboswitch core region, and consequently for riboswitch activity (Figure 2B). Because core nucleotides G40 and G144 are part of stems P2 and P4, respectively, it is likely that the folding of the P2-L4 interaction is important for the correct positioning of these two nucleotides in the core region. Therefore, these results strongly suggest that P2-L4 is particularly important to organize the riboswitch ligand binding site to ensure proper lysine-dependent riboswitch transcription activity. Our data is in agreement with previously obtained spontaneous bacterial mutants that displayed a *lysC* constitutive phenotype (24), most of which mapped to the riboswitch helical domains. These mutations very likely inhibit both L2-L3 and P2-L4 tertiary interactions thus precluding riboswitch activity. However, even if mutations disabling the P2-L4 tertiary interaction have significant impacts on both global structure and riboswitch activity, such mutants still display limited lysine-induced effects on transcription termination (Figure 2B and (21)). In contrast, the binding mutant G39C does not have any impact on global folding of the RNA (Figure

3A), but sufficiently disrupts local folding of the riboswitch binding pocket (Figure 4) to completely inhibit ligand binding (Figure 3C and (21)). These results are in good agreement with previous reports that identified *lysC* riboswitch mutants at nucleotide position 39 that were resistant to the lysine analog AEC (23,40). One reason for the absence of a lysine-induced riboswitch regulation in the context of a G39C mutant could be that C39 interacts with G111, which would perturb the riboswitch core as observed by 2AP fluorescence. Moreover, such interaction could readily destabilize the P1 stem thus leading to the complete absence of response to lysine in transcription regulation.

The lysine riboswitch aptamer exhibits a P1-P5 folding transition

The P5 helix is one of the least conserved elements of the aptamer and according to sequence alignments, the length of P5 can vary from 4 bp to more than 15 bp (16,17,52). Although no obvious role for P5 can be deduced from alignments, crystal structures show that P5 stacks onto the P4 stem and lies in close proximity to P1, suggesting that P5 could be important for P1 stabilization and riboswitch regulation. Transcription assays show that the stability of P5 is important for riboswitch function (Figure 5B), and as deduced from FRET transitions (Figure 6A), the close juxtaposition between P1 and P5 is Mg^{2+}-dependent and occurs in the low mM range, similar to what we observed from SHAPE experiments for the folding of interactions L2-L3 and P2-L4 (Figure S1B). Interestingly, the presence of lysine decreases the Mg^{2+} requirement by 10-fold, suggesting the involvement of lysine in the riboswitch folding. According to the crystal structure, interactions G39•G193 and G171•A192 are part of P1 and P5, respectively, and are involved in the formation of a purine quartet that is located underneath the bound lysine (Figure 9B). The structural relationship between the formation of the quartet and P1-P5 folding transition will require additional efforts. For instance, it will be crucial to understand if

the P1-P5 transition observed by FRET is important for the creation of this quartet, the latter being highly likely required for lysine binding. SHAPE analysis also show that the presence of lysine facilitates the folding of the riboswitch (Figure 3B), suggesting that ligand binding stabilizes the folded state of the lysine riboswitch (Figures 9A and 9B).

Although the nature of the FRET transition could purely result from a translational movement of one helix, a mixed nature involving translation and helical rotation could also produce the observed FRET increase. For instance, Batey and coworkers observed that one of the main differences between lysine riboswitch crystals obtained in absence or presence of lysine lies in the rotational orientation of the 5′-side of the P1 helix (20). Nevertheless, our results are consistent with the P1-P5 FRET transition being highly dependent on the correct formation of the P2-L4 interaction, suggesting that P2-L4 assists the P1-P5 folding transition most probably through the P4-P5 stacking interaction.

The P5 stem can act as an anti-antiterminator to control riboswitch activity

Very little information is available on the relative importance of the expression platform for riboswitch regulation. Our transcription assays indicate that P1 length is optimal for the *lysC* riboswitch and that the riboswitch system is very tight since the addition or the removal of only 2 bp significantly disrupts the ligand-induced transcription attenuation, without completely preventing ligand binding (data not shown). The addition of 2 bp is predicted to result in a $\Delta\Delta G^{\circ}_{37}$ of -4.7 kcal/mol (53), a value which agrees well with those of -8.3 kcal/mol and -7.6 kcal/mol that were obtained for the binding to the adenine riboswitch of 2AP (54) and adenine (55), respectively. Interestingly, examination of a recently published lysine riboswitch sequence alignment shows that an important number of P1 stem sequences do not

only vary in their length (56), but also in their Watson-Crick base pair ratio as various mismatches are found in the P1 helical domain, suggesting that the stability of P1 stems may be used to fine-tune the gene expression regulation. Our observations about the importance of the P1 stem sequence for the riboswitch activity are consistent with previous findings (16).

In our transcription assays, we also observed that the stability of the P5 stem is important for the riboswitch function (Figure 5). In our constructs, if we consider the P5 helical domain independently of the rest of the riboswitch molecule, the use of a P5 stem consisting of 5 bp and of 16 bp correspond to a $\Delta\Delta G^{\circ}_{37}$ of 6.3 kcal/mol and -11.5 kcal/mol, respectively, compared to the wild-type molecule. Because of the correlation between P5 stabilities and transcription regulation, we speculated that P5 could replace P1 as the anti-antiterminator. Although lysine-dependent transcription attenuation is not as efficient as what we obtained in the context of the wild-type molecule (Figure 7B), a riboswitch activation of twofold observed in presence of lysine suggests that a re-engineered riboswitch harboring an expression platform fused to P5 still displays a lysine-induced stabilization that leads to a premature transcription termination. This suggests that the riboswitch activity can also be controlled by using P5 as an anti-antiterminator. Interestingly, nucleotides G195-U197 are found in close proximity to the P5 stem. Because these nucleotides are located in the 3′ strand of the P1 stem, it is possible that the presence of P5 may influence the stability of the P1 stem. A nearly identical configuration is present in the SAM-I riboswitch (57) where natural variations in P4 stem also influence ligand-induced gene regulation (Heppell B. *et al*, in preparation), suggesting that the use of non-conserved helical domains may be a common feature employed by riboswitches to stabilize the P1 stem in the ligand-bound conformation. The successful relocation of the expression platform to the P5 helical domain shows an additional level of riboswitch modularity, best exemplified by the TPP riboswitch that is found in various expression platform configurations to regulate different biological process, such as

transcription, translation, mRNA splicing and stability (1).

The P6 stem is important for the riboswitch regulatory activity

The formation of the stem-loop P6 was found to be important for the lysine riboswitch activity (Figure 8B). Although its position within the expression platform could not readily provide information about its functional role, the latter is consistent with the transcription polarity of the riboswitch where secondary structure elements should fold as they are transcribed. Thus, in the case of the lysine riboswitch, given that P6 and the antiterminator sequences are located proximally, it is likely that both elements are involved in the formation of a large antiterminator helical domain that is required for an efficient riboswitch regulation. Given that the P6 stem is not conserved among lysine riboswitches (18), it is probable that the sequence of P6, and also of the terminator, have evolved for efficient gene regulation in various cellular contexts, agreeing well with our results showing that the addition or the removal of two or three base pairs in either domain is clearly detrimental for riboswitch activity (Figure 8).

The high specificity that riboswitches display toward their cognate ligand is mainly associated to the proper organization of the core domain, which is well illustrated by the conservation of involved nucleotides. Our results highlight the importance of long-range interactions for the formation of the ligand binding site as these elements have to be precisely positioned to ensure efficient ligand binding and riboswitch activity. The importance of peripheral elements for riboswitches is likely to be found in other riboswitch families, as recently observed (58).

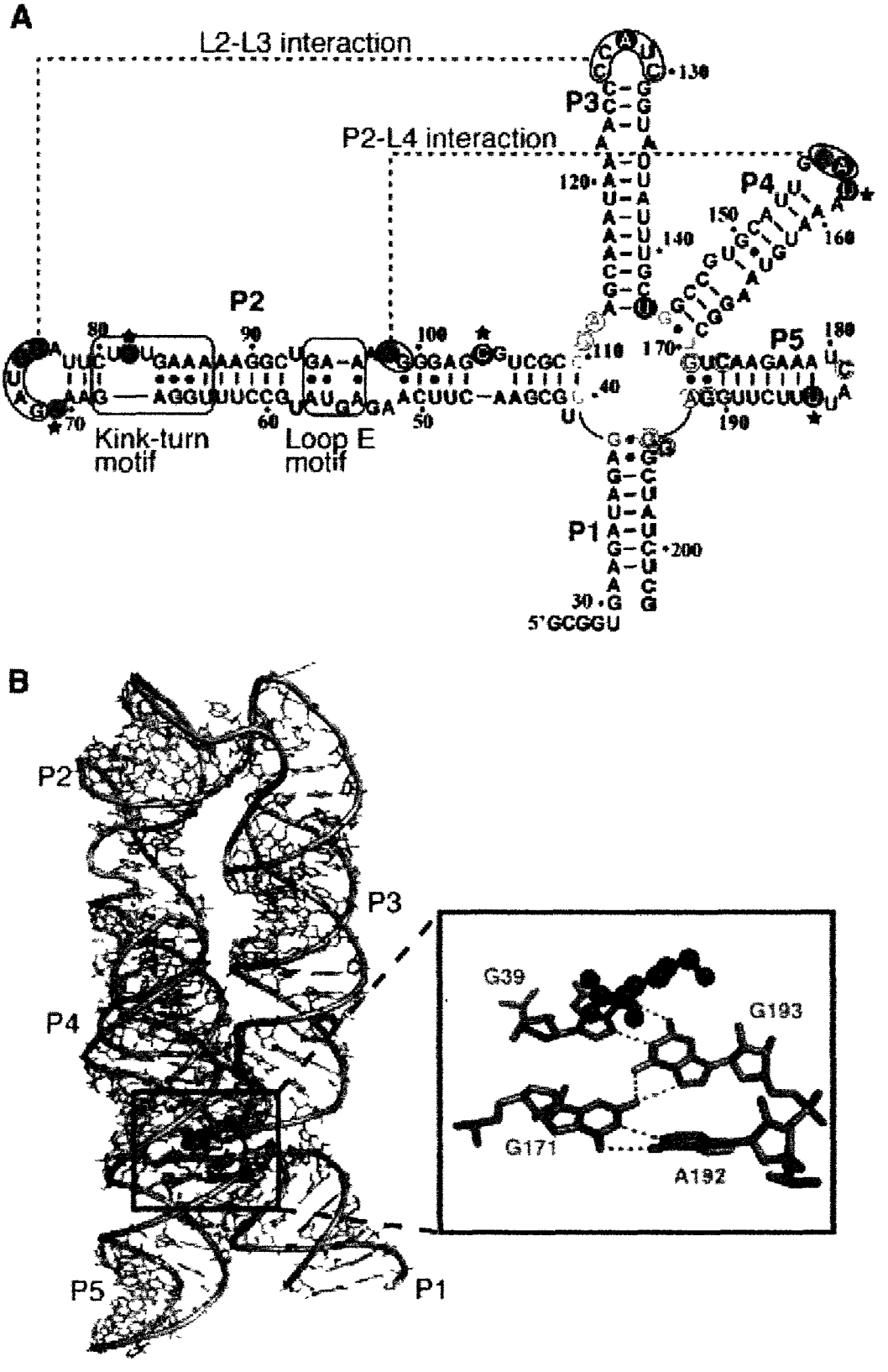

A
L2-L3 interaction
P2-L4 interaction
P3
P4
P5
P2
P1
Kink-turn motif
Loop E motif
5'GCGGU
B
P2
P3
P4
P5
P1
G39
G193
G171
A192

Figure 9. Folding of the lysine riboswitch as a function of magnesium ions and lysine. (**A**) Secondary structure of the lysine riboswitch summarizing NMIA reactivity obtained under different conditions. Regions involved in the formation of the kink-turn and loop E motif are boxed. L2-L3 and P2-L4 long-range tertiary interactions are indicated by dashed lines. Nucleotides reacting to NMIA upon increasing the concentration of magnesium ions or 5 mM L-lysine are indicated by black and empty circles, respectively. Nucleotides reacting in both conditions are indicated by gray circles. Nucleotides that are identified by a star show an increase in NMIA reaction. Nucleotides involved in the binding pocket and in the purine quartet are indicated in green and blue, respectively. (**B**) Representation of the lysine riboswitch crystal structure (19,20). Lysine is represented in red and nucleotides forming the purine quartet spanning stems P1 and P5 are indicated in blue. The inset shows the bottom junction layer where non-canonical base pairs G39•G193 and G171•A192 identified in blue are involved in the purine quartet. Note that numbering of the crystalized lysine riboswitch of *Thermotoga maritima* has been replaced by corresponding positions in the *lysC* lysine riboswitch of *Bacillus subtilis* (21).

REFERENCES

1. Serganov, A. and Patel, D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet*, **8**, 776-790.
2. Roth, A. and Breaker, R.R. (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem*, **78**, 305-334.

3. Edwards, T.E., Klein, D.J. and Ferre-D'Amare, A.R. (2007) Riboswitches: small-molecule recognition by gene regulatory RNAs. *Curr Opin Struct Biol*, **17**, 273-279.
4. Schwalbe, H., Buck, J., Furtig, B., Noeske, J. and Wohnert, J. (2007) Structures of RNA switches: insight into molecular recognition and tertiary structure. *Angew Chem Int Ed Engl*, **46**, 1212-1219.
5. Cromie, M.J., Shi, Y., Latifi, T. and Groisman, E.A. (2006) An RNA sensor for intracellular Mg(2+). *Cell*, **125**, 71-84.
6. Chowdhury, S., Maris, C., Allain, F.H. and Narberhaus, F. (2006) Molecular basis for temperature sensing by an RNA thermometer. *Embo J*, **25**, 2487-2497.
7. Grundy, F.J. and Henkin, T.M. (2004) Regulation of gene expression by effectors that bind to RNA. *Curr Opin Microbiol*, **7**, 126-131.
8. Tucker, B.J. and Breaker, R.R. (2005) Riboswitches as versatile gene control elements. *Curr Opin Struct Biol*, **15**, 342-348.
9. Gilbert, S.D. and Batey, R.T. (2006) Riboswitches: fold and function. *Chem Biol*, **13**, 805-807.
10. Yao, Z., Barrick, J., Weinberg, Z., Neph, S., Breaker, R., Tompa, M. and Ruzzo, W.L. (2007) A Computational Pipeline for High- Throughput Discovery of cis-Regulatory Noncoding RNA in Prokaryotes. *PLoS Comput Biol*, **3**, e126.
11. Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F., Sudarsan, N. *et al.* (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res*, **35**, 4809-4819.
12. Kazanov, M.D., Vitreschak, A.G. and Gelfand, M.S. (2007) Abundance and functional diversity of riboswitches in microbial communities. *BMC Genomics*, **8**, 347.
13. Freyhult, E., Moulton, V. and Clote, P. (2007) Boltzmann probability of RNA structural neighbors and riboswitch detection. *Bioinformatics*, **23**, 2054-2062.

14. Puerta-Fernandez, E., Barrick, J.E., Roth, A. and Breaker, R.R. (2006) Identification of a large noncoding RNA in extremophilic eubacteria. *Proc Natl Acad Sci U S A*, **103**, 19490-19495.
15. Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I. *et al.* (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A*, **101**, 6421-6426.
16. Grundy, F.J., Lehman, S.C. and Henkin, T.M. (2003) The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc Natl Acad Sci U S A*, **100**, 12057-12062.
17. Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S. and Breaker, R.R. (2003) An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev*, **17**, 2688-2697.
18. Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. and Gelfand, M.S. (2003) Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Res*, **31**, 6748-6757.
19. Serganov, A., Huang, L. and Patel, D.J. (2008) Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature*, **455**, 1263-1267.
20. Garst, A.D., Heroux, A., Rambo, R.P. and Batey, R.T. (2008) Crystal structure of the lysine riboswitch regulatory mRNA element. *J Biol Chem*, **283**, 22347-22351.
21. Blouin, S. and Lafontaine, D.A. (2007) A loop-loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA*, **13**, 1256-1267.
22. Blount, K.F., Wang, J.X., Lim, J., Sudarsan, N. and Breaker, R.R. (2007) Antibacterial lysine analogs that target lysine riboswitches. *Nat Chem Biol*, **3**, 44-49.

23. Lu, Y., Shevtchenko, T.N. and Paulus, H. (1992) Fine-structure mapping of cis-acting control sites in the lysC operon of Bacillus subtilis. *FEMS Microbiol Lett*, **71**, 23-27.
24. Patte, J.C., Akrim, M. and Mejean, V. (1998) The leader sequence of the Escherichia coli lysC gene is involved in the regulation of LysC synthesis. *FEMS Microbiol Lett*, **169**, 165-170.
25. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res*, **15**, 8783-8798.
26. Pleiss, J.A., Derrick, M.L. and Uhlenbeck, O.C. (1998) T7 RNA polymerase produces 5′ end heterogeneity during in vitro transcription from certain templates. *Rna*, **4**, 1313-1317.
27. Lemay, J.F., Penedo, J.C., Tremblay, R., Lilley, D.M. and Lafontaine, D.A. (2006) Folding of the adenine riboswitch. *Chem Biol*, **13**, 857-868.
28. Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1992) *Numerical Recipes in Fortran, 2nd Edn.* Cambridge University Press, UK, Cambridge.
29. Rist, M. and Marino, J. (2001) Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence. *Nucleic Acids Res*, **29**, 2401-2408.
30. Lafontaine, D.A., Wilson, T.J., Zhao, Z.Y. and Lilley, D.M. (2002) Functional group requirements in the probable active site of the VS ribozyme. *J Mol Biol*, **323**, 23-34.
31. Clegg, R.M., Murchie, A.I., Zechel, A., Carlberg, C., Diekmann, S. and Lilley, D.M. (1992) Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry*, **31**, 4846-4856.
32. Murchie, A.I., Clegg, R.M., von Kitzing, E., Duckett, D.R., Diekmann, S. and Lilley, D.M. (1989) Fluorescence energy transfer shows that the four-way DNA junction is a right-handed cross of antiparallel molecules. *Nature*, **341**, 763-766.

33. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. and Weeks, K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc*, **127**, 4223-4231.

34. Das, R., Laederach, A., Pearlman, S.M., Herschlag, D. and Altman, R.B. (2005) SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA*, **11**, 344-354.

35. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A. and Breaker, R.R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, **428**, 281-286.

36. Lilley, D.M. (1998) Folding of branched RNA species. *Biopolymers*, **48**, 101-112.

37. Silverman, S.K. and Cech, T.R. (1999) Energetics and cooperativity of tertiary hydrogen bonds in RNA structure. *Biochemistry*, **38**, 8691-8702.

38. Lemay, J.F. and Lafontaine, D.A. (2007) Core requirements of the adenine riboswitch aptamer for ligand binding. *RNA*, **13**, 339-350.

39. Heppell, B. and Lafontaine, D.A. (2008) Folding of the SAM aptamer is determined by the formation of a K-turndependent pseudoknot. *Biochemistry*, **47**, 1490-1499.

40. Yamauchi, T., Miyoshi, D., Kubodera, T., Nishimura, A., Nakai, S. and Sugimoto, N. (2005) Roles of Mg(2+) in TPP-dependent riboswitch. *FEBS letters*, **579**, 2583-2588.

41. Vold, B., Szulmajster, J. and Carbone, A. (1975) Regulation of dihydrodipicolinate synthase and aspartate kinase in Bacillus subtilis. *Journal of bacteriology*, **121**, 970-974.

42. Forster, T. (1948) Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann Phys*, **2**, 55-75.

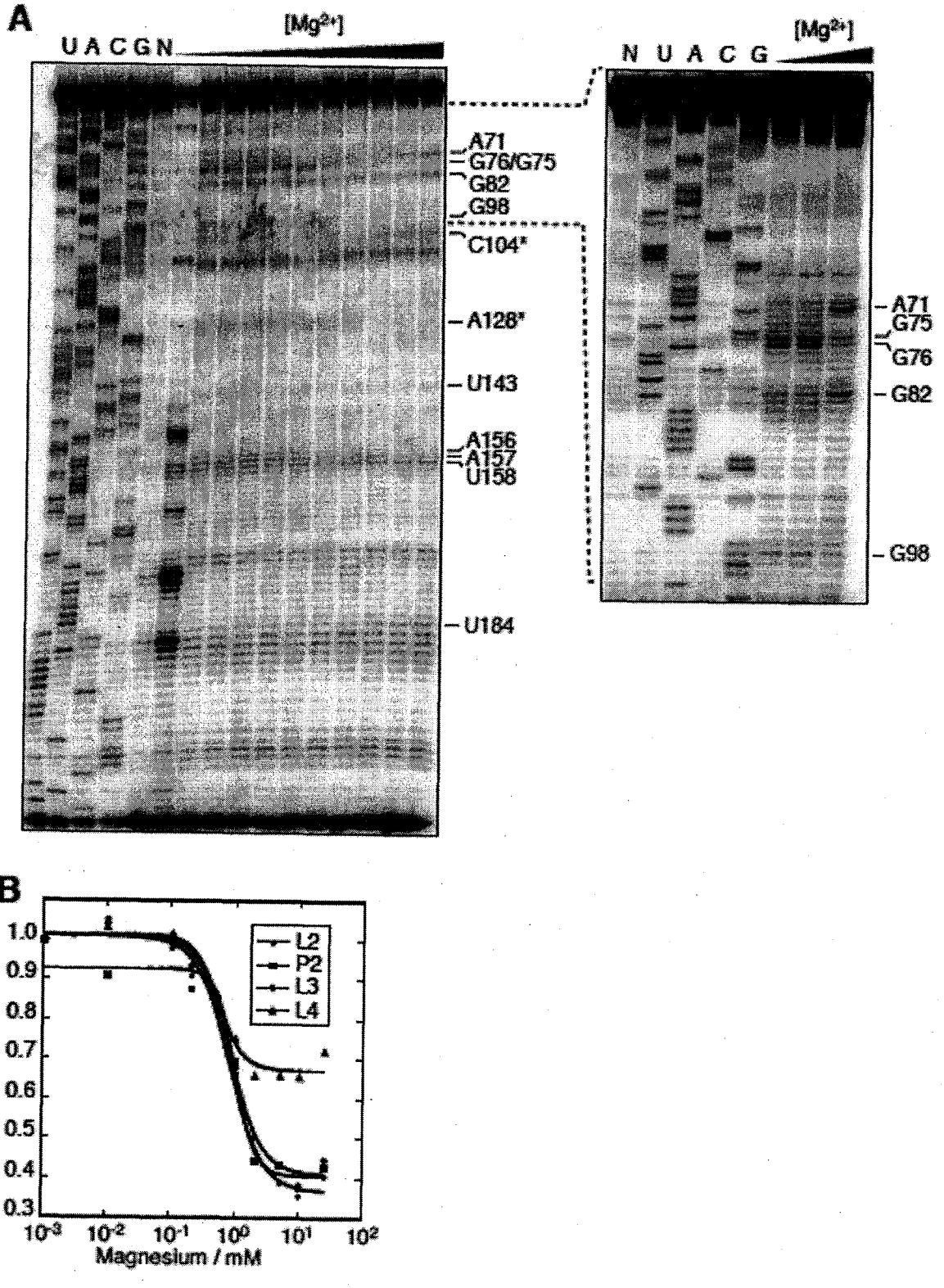
43. Lafontaine, D.A., Norman, D.G. and Lilley, D.M. (2002) The global structure of the VS ribozyme. *Embo J*, **21**, 2461-2471.

44. Walter, N.G., Hampel, K.J., Brown, K.M. and Burke, J.M. (1998) Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer. *Embo J*, **17**, 2378-2391.
45. Blount, K.F. and Uhlenbeck, O.C. (2005) The structure-function dilemma of the hammerhead ribozyme. *Annu Rev Biophys Biomol Struct*, **34**, 415-440.
46. Khvorova, A., Lescoute, A., Westhof, E. and Jayasena, S.D. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat Struct Biol*, **10**, 708-712.
47. De la Pena, M., Gago, S. and Flores, R. (2003) Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. *Embo J*, **22**, 5561-5570.
48. Penedo, J.C., Wilson, T.J., Jayasena, S.D., Khvorova, A. and Lilley, D.M. (2004) Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *Rna*, **10**, 880-888.
49. Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature*, **372**, 68-74.
50. Scott, W.G., Finch, J.T. and Klug, A. (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell*, **81**, 991-1002.
51. Martick, M. and Scott, W.G. (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, **126**, 309-320.
52. Vitreschak, A.G., Rodionov, D.A., Mironov, A.A. and Gelfand, M.S. (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet*, **20**, 44-50.
53. Xia, T., SantaLucia, J., Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C. and Turner, D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719-14735.

54. Wickiser, J.K., Cheah, M.T., Breaker, R.R. and Crothers, D.M. (2005) The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry*, **44**, 13404-13414.
55. Greenleaf, W.J., Frieda, K.L., Foster, D.A., Woodside, M.T. and Block, S.M. (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science (New York, N.Y*, **319**, 630-633.
56. Barrick, J.E. and Breaker, R.R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol*, **8**, R239.
57. Montange, R.K. and Batey, R.T. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, **441**, 1172-1175.
58. Buck, J., Noeske, J., Wohnert, J. and Schwalbe, H. Dissecting the influence of Mg2+ on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res*, **38**, 4143-4153.

ACKNOWLEDGEMENTS

We thank members of the Lafontaine laboratory for discussion, Simon Fournier for technical assistance, Dr. Alain Lavigueur for critical reading of the manuscript and the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support. S.B. holds a doctoral fellowship from NSERC and D.A.L. is a Canadian Institutes of Health Research (CIHR) New Investigator scholar as well as a Chercheur–boursier Junior 2 from the Fonds de la recherche en Santé du Québec (FRSQ).


A
U A C G N
[Mg2+]
N U A C G
[Mg2+]
A71
G76/G75
G82
G98
C104*
A128*
U143
A156
A157
U158
U184
A71
G75
G76
G82
G98
B
NMIA reactivity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
L2
P2
L3
L4
10-3
10-2
10-1
100
101
102
Magnesium / mM

Supplementary Figure 1. The folding of the lysine riboswitch as a function of magnesium ions. (**A**) SHAPE modification of the lysine riboswitch performed at 0 μM, 1 μM, 10 μM, 100 μM, 200 μM, 500 μM, 1 mM, 2 mM, 5 mM, 10 mM and 25 mM $MgCl_2$. Sequencing reactions are indicated for each nucleotides and positions where NMIA reaction was modified upon magnesium ion-inducing folding are indicated on the right. N represents a primer extension performed on unreacted RNA. The experiment performed in absence of magnesium ions shows multiple reaction sites most probably caused by NMIA overreacted with the RNA, as shown by the low fraction of full-length product. The gel on the right represents reverse transcribed products obtained using a primer hybridized on positions 105 to 126 to allow better resolution of reacting nucleotides located in the P2 stem. Reactions were performed at magnesium ion concentrations of 1 μM, 1 mM and 10 mM. Asterisks indicate regions that were assigned according to a previous study by the Batey group (20). (**B**) Relative intensity of reaction at positions involved in tertiary interactions plotted as a function of magnesium ion concentration for regions L2, L3, P2 and L4. The magnesium ion concentration required to reach half of the observed protection ($[Mg^{2+}]_{1/2}$) is indicated for each quantified position in Table 1.

CONCLUSION

Les travaux présentés dans cette thèse visaient à démontrer l'importance de la structure d'un riborégulateur sur sa fonction régulatrice, et comment ces deux éléments sont intimement liés. Plus spécifiquement, ces travaux ont été réalisés sur le riborégulateur lysine qui contrôle la transcription du gène *lysC* de *B. subtilis*, seul riborégulateur lysine pour lequel il existait des évidences expérimentales de son mode d'action au début de mon doctorat (Grundy *et al.,* 2003; Sudarsan *et al.,* 2003b).

Dans un premier temps, mes travaux ont permis de confirmer la présence du motif *K-turn* dans le riborégulateur lysine *lysC* de *B. subtilis*. Une telle démonstration est venue renforcer l'idée que les motifs *K-turn* sont beaucoup plus répandus et peuvent différer de ce que laissaient présager les premières observations (Dennis et Omer, 2005; Klein *et al.,* 2001; Lescoute *et al.,* 2005). Dans le cas précis du riborégulateur lysine, l'importance du motif a été clairement illustrée par les essais de RNase T1 qui démontrent son implication dans la formation de l'interaction boucle-boucle. De plus, les essais de transcription *in vitro* réalisés dans ce premier article soulignent l'importance des deux motifs (*K-turn* et interaction boucle-boucle) au niveau de la régulation de la transcription par le riborégulateur. Bien que les mutations introduites dans ces motifs structuraux affectent l'efficacité de terminaison de la transcription, la présence du ligand induit toujours une augmentation de cette terminaison, suggérant une liaison au moins partielle de la lysine à de tels riborégulateurs mutés. Par contre, avec une construction contenant une 2AP interne à la jonction des tiges P1 et P5 du riborégulateur (position 194), il a été possible d'observer des déficiences dans le repliement du coeur de l'aptamère avec l'introduction de ces mutations dans les éléments périphériques.

La suite des travaux a principalement permis de mieux comprendre l'étroite relation

existant entre la structure du riborégulateur et son implication dans l'atténuation de la transcription du gène *lysC*. Ainsi, l'effet des mutations introduites dans les interactions tertiaires a été observé sur gels natifs, et par l'augmentation de la concentration de magnésium requise pour le repliement de l'aptamère (expériences de SHAPE). De plus, bien que l'aptamère puisse adopter sa structure native même en absence de lysine (Garst *et al.,* 2008; Serganov *et al.,* 2008), j'ai démontré que la liaison de celle-ci stabilise l'ARN et facilite son repliement. Par contre, les expériences de SHAPE ont démontré les protections induites par ce repliement, en plus de celles induites par la lysine au site de liaison, n'étaient pas observables dans le cas d'un mutant G39C. Il s'agit de la première démonstration qu'une mutation ponctuelle dans le coeur de l'aptamère empêche complètement la liaison du ligand par des perturbations locales, observables par les expériences de 2AP, ce qui vient expliquer la résistance aux antibiotiques observée depuis longtemps pour certaines souches mutantes de *B. subtilis* (Lu *et al.,* 1991; Vold *et al.,* 1975). D'autre part, les données obtenues en lien avec la tige P4 et son interaction avec P2 démontrent l'importance de la conservation de la structure. Par exemple, la déstabilisation de cette interaction résulte directement en une désorganisation du site de liaison. Finalement, mes travaux ont permis de proposer un rôle pour les tiges variables P5 et P6, en plus de souligner l'importance de tels éléments malgré le faible niveau de conservation qu'ils peuvent démontrer.

En terminant, ce travail a permis de décortiquer l'influence des différents éléments structuraux d'une jonction d'ARN complexe, tout en les mettant en perspective avec sa fonction régulatrice. Les résultats font la lumière sur un mécanisme de résistance bactérienne à certains antibiotiques et sur l'influence de régions très variables entre les différentes séquences de riborégulateurs lysine. Ces dernières données soulignent entre autres l'importance de l'attention à porter à de telles variations dans la recherche ou la conception de nouveaux antibiotiques qui cibleraient précisément certaines bactéries (Blount *et al.,* 2007; Lee *et al.,* 2009; Jérôme Mulhbacher *et al.,*

2010). Finalement, les méthodes ici mises au point pour l'étude des riborégulateurs représentent des outils additionnels pour de tels travaux et ont notamment servi à d'autres recherches similaires du laboratoire (Heppell *et al.,* 2011; Lemay *et al.,* 2011).

DISCUSSION ET PERSPECTIVES

Bien que mes travaux aient contribué à une meilleure compréhension du mécanisme de régulation de la transcription du gène *lysC* chez *B. subtilis*, ce travail soulève d'autres questions, tant au niveau du riborégulateur lui-même qu'au niveau de la famille des riborégulateurs lysine. Même si la structure globale de l'aptamère est conservée et que son repliement semble être similaire d'une séquence à l'autre, des différences structurales, notamment au niveau de la plate-forme d'expression, et fonctionnelles nous empêchent de généraliser le mécanisme en entier. Certains de ces points attirent particulièrement mon attention et méritent d'en discuter plus longuement.

Mécanisme de liaison du ligand

L'importance des pauses de l'ARNP pendant la transcription a été démontrée dans le fonctionnement de certains riborégulateurs qui contrôlent la terminaison prématurée de la transcription (Lemay *et al.,* 2011; Wickiser *et al.,* 2005a; Wickiser *et al.,* 2005b). Ces pauses permettent au ligand de lier l'aptamère avant que la plate-forme d'expression ne soit transcrite. On dit de ces riborégulateurs qu'ils opèrent sous un régime cinétique qui dépend de la vitesse de transcription. Dans un tel cas, si la plate-forme d'expression est transcrite, le riborégulateur est «bloqué» dans une de ses deux conformations et la présence du ligand n'est plus capable d'induire un changement de conformation. La présence d'un site de pause est donc nécessaire pour que la liaison se fasse pendant la transcription, avant que l'ARNP n'ait transcrit la plate-forme. À l'inverse, d'autres riborégulateurs sont plutôt dirigés selon un mode thermodynamique, où le ligand peut se lier et induire un changement de structure réversible même si l'ARN est complètement transcrit. Nous avons récemment démontré la différence entre les deux types de liaison pour deux

riborégulateurs adénine (Lemay *et al.,* 2011). Un de ces riborégulateur contrôle la transcription et la liaison doit être co-transcriptionnelle, alors qu'un deuxième qui régule la traduction semble exclusivement dirigé par l'affinité du ligand. Dans ce deuxième cas, le changement de structure induit par le ligand est dit «réversible». Si on observe les travaux du groupe de Henkin sur les riborégulateurs SAM, les résultats indiquent une même tendance où les riborégulateurs «irréversibles» contrôlent la terminaison de la transcription, et à l'opposé, les riborégulateurs «réversibles» modulent l'initiation de la traduction (Smith *et al.,* 2010b; Tomsic *et al.,* 2008). Comme nous savons que le riborégulateur *lysC* de *B. subtilis* contrôle au niveau de la transcription, il serait intéressant de vérifier la présence de sites de pauses et leur implication potentielle dans le repliement et/ou la liaison du ligand. Si par exemple l'arrêt de l'ARNP à certaines positions s'avère influencer la régulation par le riborégulateur, il faudrait voir si, comme la longueur de P5 et la présence de P6, ces pauses varient d'une séquence à l'autre. Il est permis de penser qu'une telle régulation est dépendante du contexte, et qu'elle pourrait en partie expliquer les variations observées dans la régulation par les différents riborégulateurs d'une même famille, retrouvés dans un même génome (Tomsic *et al.,* 2008). Idéalement, l'identification d'un riborégulateur lysine qui contrôlerait exclusivement l'initiation de la traduction pourrait mener à d'intéressantes conclusions sur les différents modèles de liaison, tel que démontré pour les riborégulateurs adénine (Lemay *et al.,* 2011).

Implication de la tige P5

Une des principale conclusion tirée de la résolution de la structure d'un riborégulateur lysine (Garst *et al.,* 2008; Serganov *et al.,* 2008) est l'absence de changements structuraux majeurs induits par la présence du ligand. Seul un très léger effet de rotation de la tige P1 est observable entre les cristaux obtenus en absence et en présence de ligand (Garst *et al.,* 2008). Ces données suggèrent donc

que la lysine se lie à un riborégulateur déjà replié. Elles vont aussi dans le même sens que la plupart de mes résultats, qui démontrent le repliement de l'aptamère en absence de ligand. Par contre, mes essais de FRET démontre que, en solution, ce repliement de l'ARN est grandement facilité par la présence du ligand (voir chapitre 2, figure 6). Cette différence est possiblement camouflée par les conditions de cristallisation qui ne montrent que l'état replié le plus stable de l'ARN. De plus, des essais préliminaires de smFRET (*Single Molecule* FRET) montrent une conformation repliée ayant une valeur de FRET finale légèrement plus élevée en présence de lysine (résultats non publiés). Cette observation, non détectable dans mes essais de FRET «traditionnel», c'est-à-dire sur une population de molécules, pourrait entre autres être causé par une rotation de la tige P1, comme nous l'avons observé pour le riborégulateur SAM-I (Heppell *et al.,* 2011). Par contre, à l'inverse du riborégulateur SAM-I, des expériences de CGE (*Comparative Gel Electrophoresis*) ont confirmé l'effet de rapprochement des tiges P1 et P5 (résultats non publiés). Il pourrait ainsi y avoir une combinaison entre la juxtaposition des deux tiges et une rotation de P1, où le rapprochement est facilité par la présence du ligand et la rotation en est strictement dépendante.

Cette tige P5 est particulièrement intéressante et intrigante à cause de sa nature non conservée (en termes de longueur). Comme pour la tige P4 du riborégulateur SAM-I (Heppell *et al.,* 2011), nous avons démontré que la stabilité de cette tige variable influence le processus de terminaison de la transcription. Par contre, la tige non conservée P4 de la famille SAM-I n'est pas essentielle au riborégulateur et est absente de certaines séquences (Heppell *et al.,* 2011; Weinberg *et al.,* 2008). Dans la famille des riborégulateurs lysine, cette variabilité naturelle semble moins extrême, puisque la tige P5 est présente dans toutes les séquences connues (Grundy *et al.,* 2003; Jorth et Whiteley, 2010; Rodionov *et al.,* 2003; Sudarsan *et al.,* 2003b). Cette présence essentielle peut sans doute s'expliquer avec la formation d'une plate-forme de quatre purines qui semblent soutenir la lysine au site de liaison

et qui implique les premières paires de bases de P1 et P5 (voir chapitre 2, figure 9) (Serganov *et al.,* 2008). La tige P4 semble aussi avoir un rôle essentiel par son interaction avec P2, probablement en induisant la stabilité de P5 sur laquelle elle s'empile. Une mutation dans cette interaction pourrait ainsi avoir comme effet de déstabiliser le positionnement de P5, d'où l'absence de repliement du coeur (voir chapitre 2, mutant L4 dans les figures 4 et 6). La diminution de l'efficacité dans la terminaison de la transcription du mutant L4 serait donc provoquée par l'absence de cette plate-forme stable. Si P5 et la plate-forme sont aussi importants pour la liaison, j'estime plutôt minces les chances qu'on puisse observer des séquences de riborégulateurs lysine qui ne renferment pas cette tige P5.

Tel que mentionné dans la section précédente, le riborégulateur *lysC* de *B. subtilis* régule au niveau de la transcription et est potentiellement dirigé selon un mode de liaison cinétique. On pourrait penser que la tige P5 induit une pause de l'ARNP, ce qui expliquerait l'augmentation de l'efficacité du riborégulateur lorsqu'on allonge cette tige (voir chapitre 2, figure 5). Par contre, vu l'implication des tiges P1 et P5 dans la liaison du ligand, il est peu probable qu'une telle pause soit nécessaire à la liaison de la lysine à un aptamère dont la tige P5 n'est pas complètement transcrite. Elle pourrait par contre permettre le bon positionnement de P4, pour son interaction avec P2, et ainsi induire la stabilité de l'empilement P4-P5.

Cible potentielle pour des antibiotiques

Comme les riborégulateurs contrôlent souvent des gènes essentiels chez les bactéries, leur découverte a suscité beaucoup d'espoir en leur utilisation comme cibles pour des antibiotiques (Blouin *et al.,* 2009; Blount et Breaker, 2006; Blount *et al.,* 2007; Kim *et al.,* 2009; Lee *et al.,* 2009; Mulhbacher *et al.,* 2010). Comme la voie de biosynthèse de la lysine est d'une importance capitale pour plusieurs bactéries,

les riborégulateurs lysine impliqués dans ce métabolisme représentent des cibles de choix pour ce type d'application. Il a d'ailleurs été démontré que l'analogue de lysine AEC était capable de lier le riborégulateur lysine et il a été proposé que cette liaison serait responsable de l'activité antibiotique de l'AEC (Grundy *et al.,* 2003; Sudarsan *et al.,* 2003b). Cette hypothèse était entre autres soutenue par l'identification de mutations à l'intérieur du riborégulateur qui étaient responsables d'une résistance à l'AEC (Blount *et al.,* 2007; Lu *et al.,* 1992; Patte *et al.,* 1998). De plus, l'obtention des structures cristallines du riborégulateur lysine montrent la présence d'une cavité au site de liaison qui semble permettre certaines modifications en position C4 du ligand (Garst *et al.,* 2008; Serganov *et al.,* 2008). D'autres composés analogues à la lysine, modifiés en cette position, pourraient donc être employés pour cibler le riborégulateur lysine de certaines souches bactériennes et inhiber l'expression de gènes en aval. Ce genre de recherches nécessitent par contre des études de type «cas par cas» qui demandent de vérifier la liaison et l'inhibition de l'expression pour le ligand et le riborégulateur précis qu'on veut cibler, en plus de tester le système pour l'apparition de résistances potentielles. Le groupe de Henkin a aussi soulevé la complexité de ce genre de recherche avec l'exemple précis du riborégulateur lysine de *E. coli* et du composé antibiotique AEC (Ataide *et al.,* 2007). Dans ce cas, il semblerait que la principale cible de l'AEC ne soit pas le riborégulateur, mais plutôt l'aminoacyl-ARNt-synthétase responsable de la liaison de la lysine à son ARNt. L'action antibiotique de l'AEC résiderait ainsi dans son incorporation dans les protéines en remplacement de la lysine. Les mutations observées dans la séquence du riborégulateur qui confèrent une résistance à l'AEC ne seraient qu'une réponse à ce mécanisme pour produire la lysine de façon constitutive. Cette surproduction de lysine suffirait à compétitionner avec l'AEC et contrer son effet. Par conséquent, même s'il existe un grand potentiel à l'utilisation de riborégulateurs comme nouvelles cibles d'antibiotiques, le développement de nouveaux composés demande une compréhension globale des mécanismes d'action de ces molécules et des

résistances qui peuvent se développer, qui va au-delà de la compréhension de l'interaction ligand-riborégulateur.

Implication de protéines

Même si on soupçonnait depuis longtemps l'implication de la région 5′-UTR dans la régulation de plusieurs gènes, la découverte des riborégulateurs a tardé à émerger parce qu'on cherchait principalement à identifier des facteurs protéiques impliqués dans de telles rétroactions par un ligand (Mandal *et al.,* 2003; Nou et Kadner, 2000; Winkler *et al.,* 2002b). Par exemple, le groupe de Paulus avait déjà proposé un mécanisme d'atténuation de la transcription du gène *lysC* de *B. subtilis*, différent du riborégulateur qui était alors méconnu, mais qui impliquait tout de même des structures alternatives dans la région 5′-UTR de l'ARNm de *lysC* (Figure 15) (Chen *et al.,* 1987).

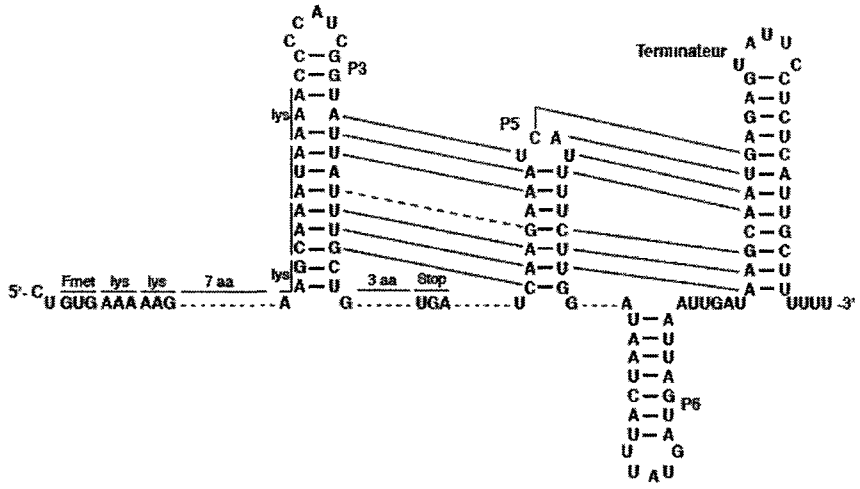


Figure 15. Mécanisme proposé pour l'atténuation de la transcription du gène *lysC* de *B. subtilis* (1987). Les codons de départ (Fmet), d'arrêt (Stop) et lysine (lys) du petit peptide sont indiqués. Les traits indiquent les paires de bases alternatives. Le nom des tiges correspondant au riborégulateur ont été ajoutées. Adapté de (Chen *et al.,* 1987).

Par contre, ce modèle proposé était comparable à la régulation de l'opéron tryptophane et impliquait la traduction d'un petit peptide atténuateur riche en acides aminés lysine. Selon ce modèle, le ralentissement du ribosome causé par une carence en acides aminés lysine permettrait la formation de l'antiterminateur. Ce n'est que quelques années plus tard que le même groupe a finalement mentionné n'avoir jamais réussi à démontré que cette potentielle région codante pouvait être traduite (Kochhar et Paulus, 1996).

C'est d'ailleurs dans cette absence de protéine que réside l'élégante simplicité et l'aspect nouveauté du mécanisme de régulation génétique par les riborégulateurs. Bien qu'ils fonctionnent ainsi dans un système minimal, il n'est toujours pas impossible que certains facteurs protéiques interagissent avec des riborégulateurs pour en influencer la régulation. La présence de motifs comme les boucles E et les *K-turns* dans certains riborégulateurs (Ames et Breaker, 2011; Blouin et Lafontaine, 2007; McDaniel *et al.*, 2005; Montange et Batey, 2006; Winkler *et al.*, 2001), en plus d'exemples qui démontrent l'implication de protéines ribosomales dans la stabilité d'autres ARN (Cho *et al.*, 2010), viennent renforcer cette idée. C'est d'ailleurs en partie par une liaison ARN-L7Ae, une protéine ribosomale, que nous avons mis en évidence la présence du *K-turn* dans le riborégulateur lysine *lysC* de *B. subtilis*. Malheureusement, en utilisant un système à deux aptamères (Figure 16) qui permet de récupérer le riborégulateur dans un extrait de protéines à partir d'une liaison ARN-streptavidine (Srisawat et Engelke, 2001), nous n'avons jamais réussi à identifier une protéine de *B. subtilis* pouvant lier ce riborégulateur. Compte tenu de la présence de l'interaction boucle-boucle entre les tiges P2 et P3 du riborégulateur lysine, il est fort probable que l'angle induit par le *K-turn* soit naturellement stabilisé, sans nécessiter la liaison d'une protéine comme il a été observé pour certains autres *K-turns* (Turner *et al.*, 2005). Mes essais ont cependant été limités au riborégulateur *lysC* de *B. subtilis*, alors que plusieurs autres séquences semblent renfermer un tel motif (Blouin et Lafontaine, 2007).

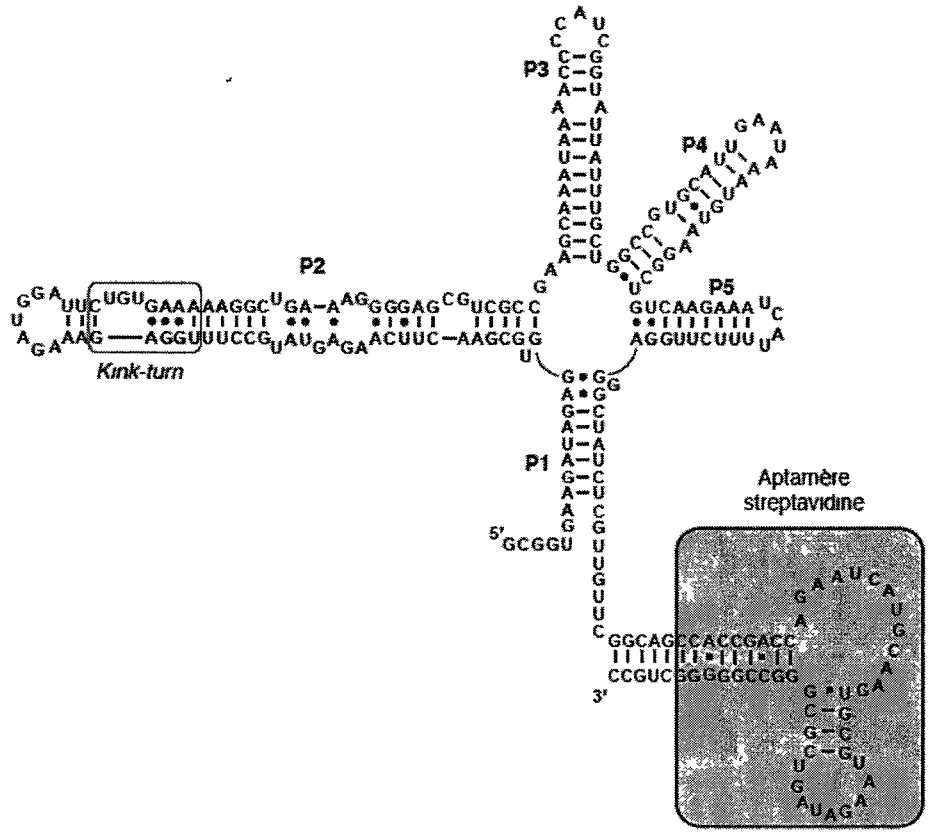


Figure 16. Structure d'un aptamère «récupérable». L'aptamère du riborégulateur lysine *lysC* de *B. subtilis* couplé à un aptamère pouvant lier la streptavidine (encadré ombragé). Le *K-turn* capable de lier certaines protéines est indiqué.

Je crois qu'il serait intéressant de tester ces autres riborégulateurs dans leur contexte naturel pour détecter d'éventuelles interactions protéine-riborégulateur. Il est aussi possible que cette interaction ne soit que transitoire ou qu'elle ne se produise seulement dans certaines conditions de croissance. On peut par exemple imaginer qu'en cas d'un arrêt de la transcription, le riborégulateur agit en *trans* comme un petit ARN avec la participation d'une protéine, un peu à la manière de Hfq. Nous savons qu'un système semblable est possible avec un riborégulateur SAM (Loh *et al.*, 2010) et que la présence du riborégulateur lysine est parfois

observable pour de longues périodes de temps (Jorth et Whiteley, 2010; Phan et Schumann, 2009), ce qui supporte l'hypothèse d'une telle utilisation du riborégulateur. Cette possibilité demeure toutefois sans plus d'évidences expérimentales, pour l'instant.

Mécanismes divergents

La plupart des riborégulateurs identifiés à ce jour régulent l'expression génétique en modulant les mécanismes de la transcription ou de la traduction. Avec une régulation en *cis*, dirigée par des changements structuraux qui ne requiert pas de facteur protéique, ces mécanismes sont d'une simplicité étonnante. Par contre, tel que mentionné en introduction, le spectre d'action des riborégulateurs s'est élargi avec leur implication dans l'épissage, la stabilité de l'ARN et l'action en *trans* d'un riborégulateur SAM. De plus, même si le fonctionnement strict du riborégulateur ne semble pas impliquer de protéine, ses changements structuraux peuvent aussi moduler des interactions protéiques. Par exemple, le riborégulateur lysine de *E. coli* permettrait l'action d'une RNase, qui dépend de la structure qu'adopte le riborégulateur (Caron *et al*, en préparation). De telles observations ouvrent la porte à de nouvelles perspectives de recherche sur les riborégulateurs. Maintenant que l'on commence à mieux comprendre leurs mécanismes de régulation, il devient intéressant d'élargir ces connaissances à l'implication plus globale qu'un riborégulateur peut avoir sur l'expression génétique, que ce soit en *cis*, en *trans* ou une combinaison qui résulte en une double action.

Sous-familles

De manière très globale, la taille d'un riborégulateur est souvent proportionnelle à la taille de son ligand. Par exemple, les riborégulateurs liant les purines sont parmi les plus petits aptamères connus (Garst *et al.,* 2010). À l'opposé, les riborégulateurs B_{12} sont compris dans des séquences beaucoup plus longues ont une structure plutôt complexe pour lier cet immense ligand (Nahvi *et al.,* 2002). Le riborégulateur lysine ne suit par contre pas cette logique en ayant un aptamère plutôt imposant (près de 200 nucléotides) qui reconnaît un ligand de petite taille. De plus, en observant la structure du riborégulateur liant la lysine (Garst *et al.,* 2008; Serganov *et al.,* 2008), on note que les tiges P2, P3 et P4, toutes trois impliquées dans les interactions tertiaires, forment une portion importante de l'aptamère, quoique très distante du site de liaison. Même si les interactions sont cruciales pour le maintient de la conformation du site de liaison (Blouin *et al.,* 2010), il est permis de se demander pourquoi la pression évolutive a favorisé le maintient d'aussi longues tiges qui, à première vue, semblent accessoires. On pourrait croire à une liaison protéique, notamment par le *K-turn* ou la boucle E ainsi exposés, mais si de telles liaisons existent, elles ne semblent pas prioritaires au mécanisme de riborégulation. On ne peut pas non plus omettre l'hypothèse que la séquence du riborégulateur agit en *trans* sur la régulation d'un autre gène, ce qui pourrait expliquer cette conservation de longueur de la séquence. Nous savons cependant que des structures très divergentes, comme les sous-familles de riborégulateurs SAM et preQ1, arrivent à lier un même ligand. Il serait ainsi intéressant d'identifier des sous-familles de riborégulateurs lysine, où de nouvelles structures arriveraient quand même à organiser une liaison avec la lysine.

«Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.»

- Sir Winston Churchill (1874-1965)

ANNEXE

Blouin, S., Mulhbacher, J., Penedo, J. C. et Lafontaine, D. A. (2009). Riboswitches: ancient and promising genetic regulators. Chembiochem *10*, 400-416.

Riboswitches: Ancient and promising genetic regulators

Simon Blouin,[a] Jérôme Mulhbacher,[a] J. Carlos Penedo,[b] and Daniel A. Lafontaine* [a]

[a] S. Blouin, Dr. J. Mulhbacher, Prof. D. A. Lafontaine
Dpartement de biologie, Université de Sherbrooke
2500, Boulevard Université, Sherbrooke, J1K 2R1 (Canada)
Fax: (+1) 819-821-8049
E-mail: daniel.lafontaine@usherbrooke.ca
[b] Dr. J. C. Penedo
University of St. Andrews, School of Physics and Astronomy
North Haugh, St. Andrews, KY16 9SS (UK)
Fax: (+44) 1334-463104

Abstract

Newly discovered metabolite-sensing riboswitches have revealed that cellular processes extensively make use of RNA structural modulation to regulate gene expression in response to subtle changes in metabolite concentrations. Riboswitches are involved in various regulation levels of gene expression such as transcription attenuation, translation initiation, mRNA splicing and mRNA processing. Riboswitches are found in the three kingdoms of life, and in various cases, are involved in the regulation of essential genes, making their regulation an essential part of the cell survival. Because riboswitches operate without the assistance of accessory proteins, they are believed to be remnants of an ancient time, where gene regulation was strictly based on RNA, from which are left numerous "living molecular fossils", as exemplified by ribozymes, and more spectacularly, by the ribosome. Due to their nature, riboswitches hold high expectations for the manipulation of gene expression and the detection of small metabolites, offering also an unprecedented potential for the discovery of novel classes of antimicrobial agents.

1. Introduction

Only twenty-five years ago, RNA was mainly considered as a cellular messenger carrying out genetic information from the genome to the ribosome for the production of proteins. Only protein enzymes were then considered to be truly "molecular effectors", which were involved in essential biological catalytic processes. The discovery of RNA enzymes, or ribozymes,[1, 2] changed the way that scientists perceived the role of RNA as a cellular component. Ribozymes, by playing both informational and catalytic roles, inspired the RNA World hypothesis, which was put forward to address the "chicken and egg" problem of how a translation system could emerge without proteins already in place. Thus, if RNA or some RNA-related

molecules could encode information and perform catalytic reactions, then protein-driven chemical reactions could have evolved from those molecules to yield the complexity observed in modern cellular processes. However, for such a primitive world to be viable, it would also require that a large collection of catalytic reactions would benefit from feedback regulatory loops as a way to regulate gene expression. These regulating mechanisms would need to be highly specific to prevent unwanted regulatory cofactors from perturbing/altering metabolic regulation by acting on unrelated biochemical pathways. Because RNA has only a handful of functional groups compared to proteins, it can be hard to imagine that RNA can build tertiary structures complex enough to form specific ligand binding sites. However, it was repeatedly shown using SELEX (Systematic Evolution of Ligands by EXponential enrichment) that RNA can be a 'good contortionist' at building complex architectures.[3] A large number of artificial RNA aptamers have been isolated responding to various small molecules such as theophylline, adenosine triphosphate, amino acids and organic dyes (recently reviewed[4]). By taking advantage of the fact that formation of the RNA-ligand complex increases the stability of an RNA structure compared to that of the RNA by itself, several gene regulation systems were engineered in which Shine-Dalgarno (SD) and initiation codon sequences were selectively sequestered depending on the presence of an exogenously added ligand which altered the structure of the mRNA upon binding.[4] Such results strongly suggested that the RNA World could have benefited from feedback regulatory mechanisms to regulate gene expression.

An additional evidence for an RNA-based primitive world came from recent work showing that naturally occurring RNA motifs can act as specific regulatory switches. Strikingly, it was found that these RNA switches were able to use feedback regulatory loops to monitor and control their own expression, apparently all of this in absence of any helper protein. These RNA elements, called riboswitches, are literally 'molecular switches' that can sense the presence of small cellular

metabolites and to fold into a different conformation to either promote or inhibit expression of the protein encoded downstream. Given that a number of excellent recent reviews have covered in great detail riboswitch molecular structures,[5-24] this review will mainly focus on riboswitch regulation mechanisms. We will also explore the growing potential of riboswitches to be used as antibacterial cellular targets, molecular biosensors and also as ligand inducible effectors. In a nutshell, this review will attempt to show the current knowledge about riboswitches and the vast possibilities that they offer as molecular tools.

2. Riboswitches as novel gene control elements

Riboswitches are located in untranslated regions of mRNA and are able to sense the concentration of a target cellular metabolite, which is almost always related to the gene product encoded by the riboswitch downstream sequence. To be able to sense various types of cellular changes, riboswitches use different molecular architectures that are highly adapted to monitor biological cues. The variety of riboswitch architectures spans from single temperature-responsive stem-loops to complex structures able to specifically bind large metabolites such as coenzyme B_{12}. In all cases studied, this cis-acting regulation mechanism enables a tight regulation control that allows the sensing RNA molecule to be directly regulated. This is a different strategy compared to RNAi-based trans-acting system where a short interfering RNA molecule acts as a guide to target complementary sequences within mRNAs, leading to the activation of the RNA-Induced Silencing Complex (RISC), and ultimately to mRNA silencing.[25] With the exception of the *glmS* ribozyme,[26-28] all reported riboswitch representatives control gene expression by modulating the formation of at least one helical domain. Surprisingly, only by performing such a minimal structure alteration, riboswitches are able to modulate a large array of diverse biological processes such as transcription attenuation, translation initiation,

mRNA splicing and mRNA processing.

The term 'riboswitch' was coined by Ronald Breaker in 2002 when he reported that mRNA encoding enzymes involved in vitamin B_1 and B_{12} biosynthesis in *Escherichia coli* could bind associated metabolites without helper proteins being involved.[29, 30] Similar findings were also obtained from other groups when they observed that ligand-dependent feedback regulation mechanisms could rely entirely on RNA structural changes.[31-34] However, before the riboswitch concept was generally accepted, several hints came from previous biochemical studies, which paved the way to a significant conceptual progress, and ultimately to the proposition that RNA structures can directly bind cellular metabolites to modulate gene expression.[35-37] For instance, based on mutational analyses, it was proposed that elements found in the 5′-UTR of mRNA involved in the synthesis of thiamine pyrophosphate (TPP), riboflavin and adenosylcobalamin were required for gene regulation of downstream gene products.[32, 38, 39] These studies provided a clear answer as to why several attempts to identify regulatory proteins involved in the gene regulation did not reveal expected protein cofactors. After the riboswitch concept was put forward, a 'mini revolution' took place among the RNA community, which led to the discovery of many distinct classes of RNA switches. Natural cis-acting riboswitches have been shown to respond to various metabolites such as adenine,[40] adenosylcobalamin (B_{12}),[30] flavin mononucleotide,[31, 41] guanine,[42] 2'-deoxyguanosine,[43] glucosamine-6-phosphate,[26] glycine,[44] lysine,[33, 45] molybdenum,[46] 7-aminomethyl-7-deazaguanine (preQ1),[47] S-adenosyl methionine (SAM),[34, 48-51], S-adenosylhomocysteine (SAH)[52] and TPP.[29, 31]

The sensing of target metabolites by riboswitches is performed by the aptamer domain, which is responsible for the high affinity and specificity of riboswitch-ligand complexes (Figure 1).[6] Where studied, it has been shown that the aptamer folds into a defined structure that produces specific contacts with most of the atomic

groups of the ligand ensuring a very high ligand binding specificity. Importantly, the formation of the ligand-aptamer complex has a direct influence on the folding of the downstream expression platform, which is used to control gene expression. The adaptive recognition of the binding site appears to be very high as a large spectrum of ligand binding riboswitches has been found until now.[6] The expression platform also shows an important level of plasticity since it can adopt a large array of secondary structures that can be used to regulate most of the important biological processes, and this either by upregulating or downregulating gene expression. The modularity of the aptamer and the expression platform makes riboswitches the 'Swiss army knife' of cellular regulation, and even more remarkably, all of this without the absolute requirement of protein cofactors. The genetic decision relying on the folding status of the riboswitch is dictated by the binding of the target metabolite to the aptamer domain. In most cases, when the concentration of the target metabolite is high in the cell, its binding to the riboswitch aptamer results in the repression of its synthesis (Figure 1). However, gene expression is stimulated by ligand binding in cases in which it is involved either in metabolite cellular transport,[40] or as an energy source.[44]

The paradigm of riboswitch regulation is well represented by the purine riboswitch class, which consists of guanine and adenine riboswitches that negatively and positively regulate gene expression upon ligand binding, respectively (Figure 1).[40, 42] In the case of the guanine riboswitch, a transcription terminator element is formed only in the presence of guanine, resulting in gene regulation repression,[42] which can vary according to the riboswitch representative.[53] For the adenine riboswitch, the presence of adenine is associated with the disruption of the transcription terminator thereby allowing gene expression to take place (Figure 1). Crystal structures show that despite their high structural similarities, both riboswitch aptamers display very high specificity toward their respective ligand,[40, 42] which is essentially achieved by the formation of a Watson-Crick base pair between the

ligand and a nucleobase located in the aptamer domain.[54, 55] The adenine riboswitch is also thought to modulate at the translational level,[55, 56] which is another common regulation mechanism used by bacterial riboswitches (see section 4.2). Translational riboswitch control has also been characterized in detail in the case of the ROSE (repressor of heat-shock gene expression) element, which controls translation initiation via thermosensing making the miniROSE element one of the simplest riboswitch structures (Figure 1).[57] This control mechanism is achieved by using only a simple helical domain that sequesters the Shine-Dalgarno (SD) sequence and the AUG start codon, both of which are exposed when the temperature is increased to allow expression of heat-shock gene products.[57] A similar thermosensible regulation mechanism is involved in infections caused by the food-borne pathogen *Listeria monocytogenes*.[58] It has been shown that invasion proteins and phospholipases involved in the infectious process are controlled by *prfA*, which is not expressed below 30°C, resulting in low expression of virulence. However, when *L. monocytogenes* enters an animal host, the temperature shifts to 37°C, resulting to *prfA* expression and production of proteins involved in the infectious process. Interestingly, a thermosensor RNA structural element has been shown to control *prfA* expression by masking the ribosomal binding region at 30°C but by allowing ribosome access at 37°C.[58]

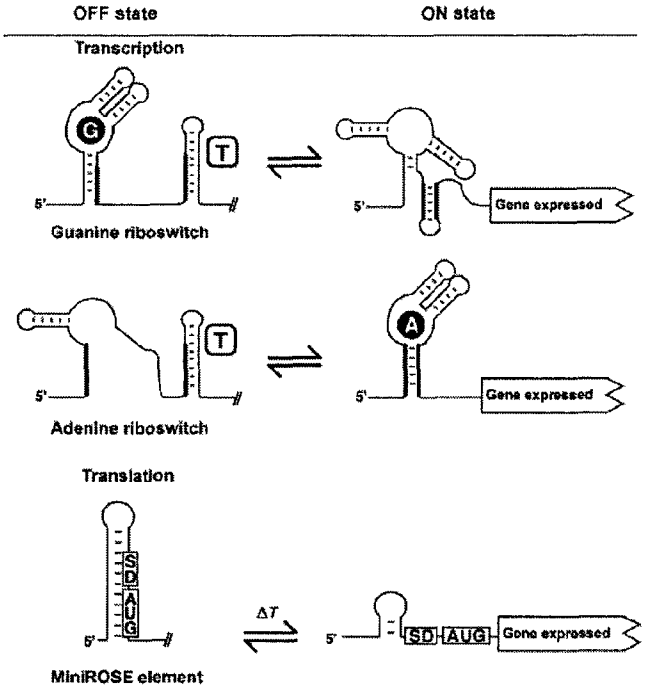


Figure 1. Riboswitch gene regulation mechanisms. The guanine riboswitch is shown as an example of a negatively-controlling riboswitch as premature transcription termination is favored in the presence of guanine. The adenine riboswitch positively controls gene expression since the presence of adenine is associated with transcription elongation. The mini-ROSE element positively modulates translation initiation when temperature is high enough. Guanine, adenine and transcription terminators are shown by G, A and T, respectively.

3. Riboswitch structural diversity

Riboswitches fold into diverse secondary structures, which are often organized around helical junctions of varying degrees of complexity. For instance, in the case of purine riboswitches, and most probably also for the recently discovered 2'-deoxyguanosine riboswitch (Figure 2), the aptamer folds around a three-way

junction displaying a loop-loop interaction, which is important for folding and activity.[40, 42] The junction also contains a helical stacking interaction occurring between stems P1 and P3, which is frequently found in three-way RNA junctions.[14] The hammerhead ribozyme shows a highly similar global tertiary fold as it also contains a loop-loop interaction as well as an equivalent helical stacking interaction. In the case of the hammerhead, the presence of the loop-loop interaction was only recently appreciated when it was shown that this distal interaction is very important for the ribozyme function at physiological conditions.[59-62] The presence of the loop-loop interaction has been shown to be also important for purine riboswitches.[54, 63-66] As opposed to the hammerhead ribozyme, a minimal adenine riboswitch aptamer lacking the loop-loop interaction does not readily fold into a similar global tertiary structure as compared to the wild-type (Lemay and Lafontaine, unpublished observations), most probably indicating that the aptamer core is less rigid than the hammerhead ribozyme and cannot by itself drive the riboswitch to the active conformation. A naturally-occurring variant of the purine riboswitch has recently been reported in only one bacterial species that exhibits selective and high-affinity binding of 2'-deoxyguanosine,[43] indicating that the ligand specificity of existing riboswitches can be reprogrammed without significantly altering the global aptamer architecture.

Riboswitches have found different molecular solutions to create ligand binding aptamers for both preQ1 and SAM (Figure 2). Whilst aptamers preQ1-I and preQ1-II exhibit relatively similar secondary structures, those specific to SAM display very different structures ranging from a simple stem-loop to complex three- or four-way junctions, suggesting that evolutionary distinct RNA aptamers may converge to the same function (Figure 2). Crystal structures of SAM-I, SAM-II and SAM-III aptamers show that although the ligand is bound in distinct ways,[67-69] high affinity and specificity is preserved in both aptamers, enabling a strong discrimination against most analogues of SAM, such as SAH.[70] For instance, to recognize the positively charged sulfur group as well as the activated methyl group, the SAM-II aptamer

makes use of carbonyl oxygen groups from uracils, a very similar strategy to the one employed by SAM-I, which employs two universally conserved A-U pairs to provide the carbonyl oxygen atoms. Interestingly, a particular SAH binding aptamer which shares structural similarities with the SAM-II aptamer[52] suggests a common ligand binding site for both ligands as SAM and SAH only differ by the presence of a single methyl group and a positive charge. The existence of several different aptamers for preQ1 and SAM is consistent with the idea that riboswitches exploit the structural versatility inherent to RNA molecules and suggests that other RNA binding aptamers sharing similar ligand binding strategies are bound to be discovered.

The lysine riboswitch is centered on a unique five-way junction that is most probably important for lysine binding.[33, 45, 71] A long-range loop-loop interaction occurring between stems P2 and P3 has been shown to be important for the aptamer global folding, ligand binding and riboswitch transcriptional control activity.[71] This study also found that the correct folding of a kink-turn element present in the P2 stem is crucial for the loop-loop interaction to take place. The simultaneous presence of a kink-turn element together with a long-range tertiary interaction is not unique to the lysine riboswitch since a similar architectural arrangement is also observed in the SAM-I aptamer. From the SAM aptamer crystal structure, it is clear that the presence of the K-turn is critical for the long-range pseudoknot interaction to form, which in turn is vital for the ligand binding activity of the aptamer.[72] In addition, in the 23S ribosomal RNA, a three-way junction also exhibits a loop–loop interaction in which one of the interacting arms contains a kink-turn motif.[73, 74] In the case of the ribosomal junction, the crystal structure shows that ribosomal proteins L24 and L29 interact with the contained kink-turn (Kt-7), suggesting that their presence could be important for the correct folding of the K-turn motif and, consequently, for the loop–loop structure.

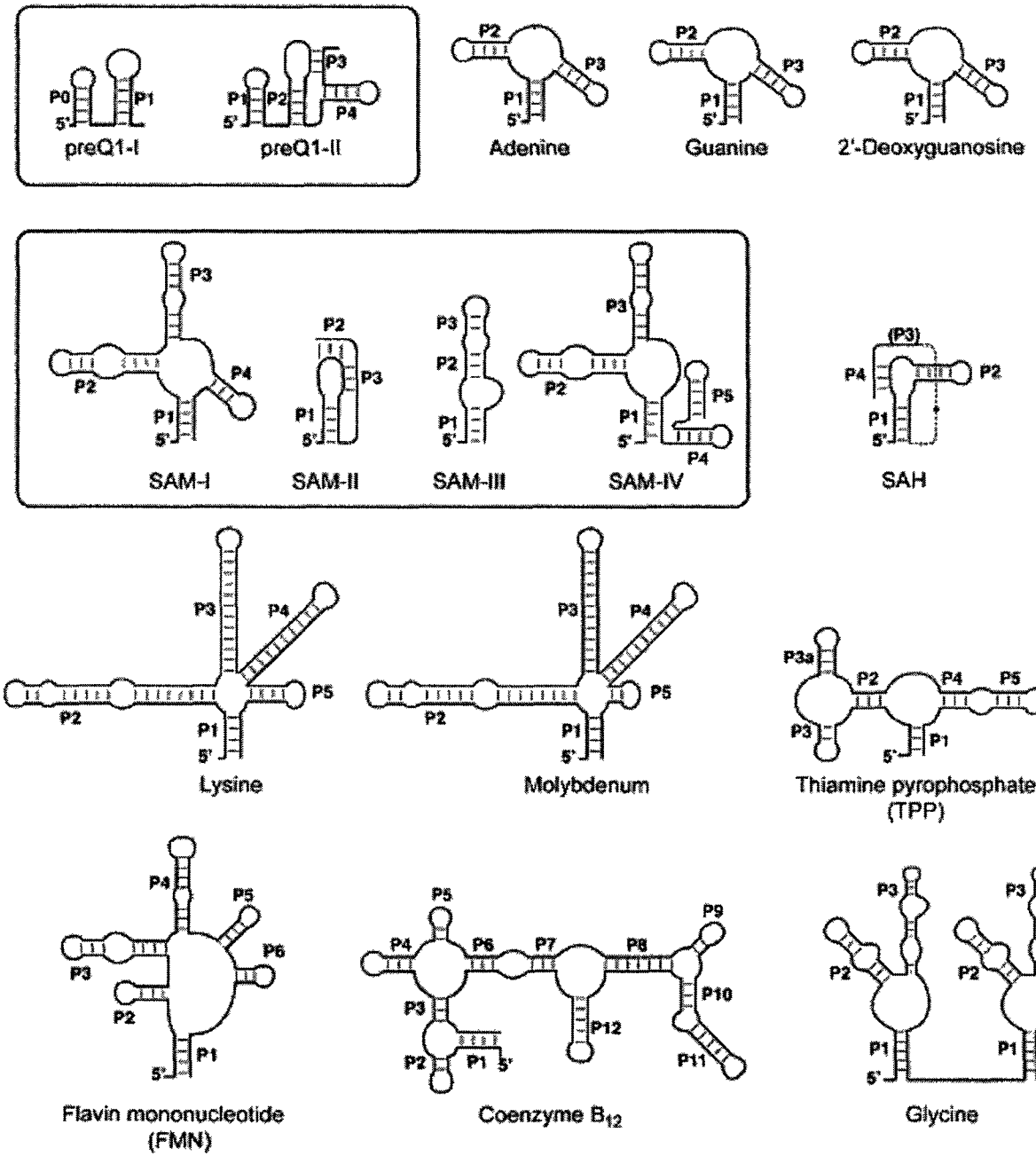

P0
P1
5'
preQ1-I
P3
P1
P2
P4
5'
preQ1-II
P2
P3
P1
5'
Adenine
Guanine
2'-Deoxyguanosine
P3
P4
P2
P1
5'
SAM-I
P2
P3
P1
5'
SAM-II
P3
P2
P1
5'
SAM-III
P5
SAM-IV
(P3)
SAH
P5
Lysine
Molybdenum
P3a
Thiamine pyrophosphate
(TPP)
P6
Flavin mononucleotide
(FMN)
P9
P10
P12
P11
Coenzyme B_{12}
Glycine

Figure 2. Secondary structures of metabolite-binding riboswitch aptamers. The sensed metabolite is indicated for each riboswitch class. Aptamers sensing an identical ligand are grouped in a rounded rectangle. An additional P3 stem-loop located between P2 and P4 may be present in some SAH natural representatives and is indicated between parentheses. A zero length-linker region is identified by a dashed line and the strand polarity is indicated by an arrow. Aptamer secondary structures should be considered only as artistic representations since known tertiary interactions are not represented and do not contain the exact number of base pairs. The *glmS* ribozyme is shown in Figure 3.

The glycine-specific riboswitch is unique among known riboswitches since it is the only one that employs cooperative binding to control gene expression.[44] The secondary structure of the aptamer encompasses two similar domains joined by a linker sequence (Figure 2). Because these two domains are interrelated, cooperative binding can be used to efficiently respond to ligand within a small range of concentrations. A recent study employing Nucleotide Analog Interference Mapping (NAIM) has identified two sites that are involved in cooperative tertiary interactions.[75] In this study, Kwon and Strobel have found that the minor groove of the P1 stem of the first aptamer as well as the major groove of the P3 stem of both aptamers are likely to be important for inter-domains docking. These findings suggest that the glycine riboswitch uses the P1 stem of the first aptamer not to modulate the expression platform as done in other riboswitches, but to contribute to tertiary interactions that are important for ligand binding cooperativity. This finding is a striking example that riboswitches can exhibit complex regulation characteristics that were previously thought to belong only to protein cofactors. As if the remarkable riboswitch versatility already discussed was not enough to control gene expression, recently discovered riboswitches exhibit a further degree of complexity.[44, 76, 77] For instance, molecular arrangements consisting of two independent TPP riboswitch

units have been reported in *Bacillus anthracis*.[77] The riboswitch duplication is most probably important to provide a greater dynamic range for ligand binding and thus, a higher sensitivity for the control of gene regulation. A further level of complexity is shown by the SAM-B_{12}-dependent dimeric riboswitch,[76] which exhibits two aptamers that are dissimilar in their ligand binding specificity. Such a combination enables gene expression to be controlled in a Boolean NOR logic gate fashion, where the binding of any ligand produces downregulation of gene expression. Similar dimeric riboswitches responding to different ligands are also very likely to exist since a dimeric riboswitch containing a guanine aptamer as well as an additional domain with unknown ligand specificity was recently described.[76] Riboswitch-mediated gene regulation control is not only performed using metabolite-binding aptamer since other cellular signals such as temperature,[78, 79] magnesium[80] and tRNA[18] have been shown to influence RNA structure and to modulate gene expression. For instance, it is interesting to note that tandem arrangements are also observed on a higher frequency for tRNA-sensing T-boxes, which are likely to require tighter regulation control over tRNA charging.[77, 81]

Lastly, it appears that riboswitches could show even more diversity for ligand binding effectors as compelling evidences of this were recently obtained.[82, 83] For instance, using a computational pipeline, Weinberg et al. used bacterial comparative genomics and identified 22 novel candidate RNA motifs, some of which are apparently involved in transport, citric acid cycle, molybdenum cofactor biosynthesis and Vibrio cholerae natural competence.[82] Also, three other widespread candidates, which exhibit a conserved secondary structure are associated with homologous genes, and thus share most characteristics that are usually associated to typical riboswitches. However, three other candidates show a narrower distribution since they are only found in a single taxonomic order suggesting that many remaining riboswitches to be discovered may exhibit narrower phylogenetic distribution than those already known.[82]

4. Biological processes regulated

4.1 Premature transcription termination

Bacteria primarily use two mechanisms for termination of transcription. Some genes incorporate a termination signal that is dependent upon Rho protein,[84] while others make use of Rho-independent terminators (intrinsic terminators) to destabilize the transcription elongation complex.[85-87] These 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[88] and are typically located at the 3′-termini of genes or operons. However, they are also found in riboswitch expression platforms where ligand binding can either produce the formation or the disruption of the terminator, resulting in abortive transcription or elongation, respectively (Figure 1). The affinity of the aptamer-ligand complex is presumably one of the most important factors in the riboswitch gene regulation mechanism. Indeed, provided that the ligand is present at sufficient concentration, it is predicted that the aptamer-ligand complex is formed and gene regulation ensued. As long as ligand binding is fast compared to the transcription rate, riboswitches are operating under a thermodynamic regime with the equilibrium between the bound- and unbound-ligand aptamer dictating the outcome of the genetic expression.

A further notion was discovered when it was observed that for some riboswitches, the rate of ligand binding (k_{on}) is important for the regulation process.[89, 90] Indeed, compared to RNA polymerase (RNAP) transcription rate, slow ligand binding kinetics implies that formation of the aptamer-ligand complex may not attain equilibrium before the expression platform is transcribed. Thus, in conditions where ligand binding is slow compared to RNAP transcription rate, ligand binding to the aptamer domain may not occur before the expression platform is transcribed, indicating that aptamer stabilization may not be achieved by the time the genetic decision is made,

which corresponds to a kinetic regime regulation control. For riboswitches that are kinetically controlled, it is expected that the concentration of ligand required to trigger transcription termination is higher than the apparent K_D value determined for the riboswitch-ligand complex, which has been observed in many cases.[34, 45, 50, 90] However, it should be noted that the apparent discrepancy between these values could also result in part from the presence of alternative RNA structures that do not produce ligand binding-competent riboswitches. In addition, because the concentration of ribonucleotide triphosphates (NTPs) modulate RNAP transcription rate, it is expected that NTPs concentration can affect the concentration of ligand required to mediate termination, which has been observed for the *ribD* FMN[90] and the *gcvT* glycine[44] riboswitches. These results are consistent with the idea that the transcription rate may influence transcription termination by allowing more or less time for ligand binding to occur. In addition, in the context of the *pbuE* adenine riboswitch, it was recently shown[66] that the presence of the expression platform negatively influences the ability of the aptamer to perform ligand binding, which is consistent with a kinetic control under certain conditions.[89]

Under conditions in which a given riboswitch operates under a kinetic regime, and therefore when the rate of RNAP transcription is determining, pausing may be a significant factor in riboswitch regulation. Because transcriptional pauses are known to decrease transcription elongation rate by as much as 10,000-fold,[91] it is very likely that they could be of vital importance for kinetically driven riboswitches by providing additional 'breathing time' for riboswitch-ligand complex formation. For instance, in the case of the FMN riboswitch, it has been found that two major pause sites present in the riboswitch sequence are characterized by lifetimes of ~10 seconds and ~1 minute.[90] In agreement with the idea that pauses sites provide more time for ligand binding, the mutation of either pause sites increased transcription termination as well as an increase in the ligand concentration required to obtain 50% of the full extent of transcription termination. Also, the addition to

transcription reactions of the RNA binding protein NusA, which is known to alter transcription termination levels,[92] resulted in increased lifetimes at the pause sites whereas the increase in NTPs concentration resulted in the decrease of pause lifetimes, as expected from their respective influences on transcription rates. Together, these results suggest that rates of ligand binding as well as transcription are in some cases more important than the affinity of the riboswitch-ligand complex.

4.2 Translation initiation

Whereas Gram-positive bacteria riboswitches mainly control transcription attenuation, Gram-negative riboswitches tend to modulate translation initiation levels.[19, 93] This functional difference may be explained by the relative prevalence of polycistronic mRNA found in both bacterial types. For instance, given that Gram-positive bacteria have larger gene clusters, they may benefit from premature transcription attenuation control.[19, 93] Among known riboswitch representatives, those responding to TPP,[29] FMN,[41] coenzyme B_{12}[36] and SAM[94] have been shown to be involved in translational control. Translation initiation is also controlled by making use of secondary structural changes in which the access to the SD and AUG sequences is modulated to control ribosome binding, and hence translation initiation. The modulated structure may be considered as a sequestering helix given that it is involved in the base pairing of the SD and AUG sequence segments. In contrast to transcriptionally-regulated mRNA, translationally-controlled transcripts allow the mRNA molecule to be full-length when ligand binding occurs. Whether or not translation-controlling riboswitches operate under various regulation modes (kinetic vs. thermodynamic) is not readily apparent. In contrast to transcription-regulating riboswitches, those involved at the translational level do not appear to need to perform their regulation during the transcription process. However, this will need to be further investigated using appropriate methods since those already developed for

riboswitches regulating transcription may not readily be transposable.

One of the first riboswitches identified in translational control is the TPP-responding riboswitch.[29] It was previously observed that a conserved cis-acting element located in the 5′-UTR of an mRNA associated with thiamine biosynthesis and transport is important for TPP-dependent regulation.[32, 95] Importantly, no protein regulatory factor was identified using experimental assays thus setting up the *E. coli thi* box element as a potential riboswitch motif. Its riboswitch-dependent regulation was established using the *lacZ* reporter gene, which showed gene repression when bacteria were grown in presence of thiamine.[29] The repression mechanism was attributed to direct binding of TPP onto the *thi* box element, which caused structural alterations in the aptamer domain. The apparent K_D value obtained for the formation of the RNA-TPP complex is ~500 nM, while other thiamine phosphorylated forms lead to lower affinities, establishing that the TPP riboswitch is highly specific for TPP. This molecular specificity was further supported using in-line probing assays which showed that the RNA is able to make specific contacts with two different regions of the ligand, which consist of the purine and phosphate moieties. More specifically, in-line probing also revealed that the SD sequence becomes more structured upon TPP addition, consistent with the formation of a helical structure involving the SD element that does not allow ribosome binding and hence prevents translation.

4.3 mRNA processing

In addition to transcription and translation mechanisms, riboswitches have also evolved to control ligand-dependent mRNA processing. The *glmS* gene, which encodes the enzyme glutamine-fructose-6-phosphate synthetase, converts fructose-6-phosphate and glutamine into GlcN6P. Importantly, upon binding to GlcN6P, the *glmS* mRNA undergoes autocatalytic cleavage that generates 5′OH and

2',3'-cyclic phosphate moieties,[26] suggesting a catalytic mechanism similar to that previously described for small nucleolytic ribozymes. The *glmS* secondary structure is quite different from other ribozymes (Figure 3A) and shows a double pseudoknot structure (P2.1 and P2.2), a feature which is also found in other ribozymes such as the hepatitis delta virus and Diels-Alderase ribozymes. [96, 97] It has been recently reported that both pseudoknots are involved in the formation of the ribozyme core as well as the ligand binding site and that a long-range tertiary interaction is formed between a GNRA tetraloop located in the P4.1 stem and the minor groove of P1 helix, all of which are important for the compact fold of the ribozyme.[27, 28]

The ligand binding specificity of the *glmS* ribozyme has recently been studied in detail using a variety of ligands.[98] In their work McCarthy et al. established that several functional groups of the ligand are important for the cleavage reaction. Among them, the amine group of the GlcN6P is crucial given that although glucose-6-phosphate (Glc6P) does not support catalytic activity, metabolites such as serinol, L-serine and ethanolamine yield to catalysis. However, the presence of the phosphate group is not as important since efficient catalytic activity is observed in presence of D-glucosamine. By examining all potent metabolites, common features were found which consist of an ethanolamine moiety as well as a vicinal hydroxyl group,[26] and because Glc6P inhibits *glmS* activity, it suggests that the presence of the amine is dispensable for ribozyme binding. The essential role of GlcN6P was recently highlighted in an in vitro selection study[99] where it was not possible to alter the range of compounds supporting catalytic activity, which is consistent with the hypothesis that GlcN6P is used as a coenzyme for *glmS* self-cleavage.[100]

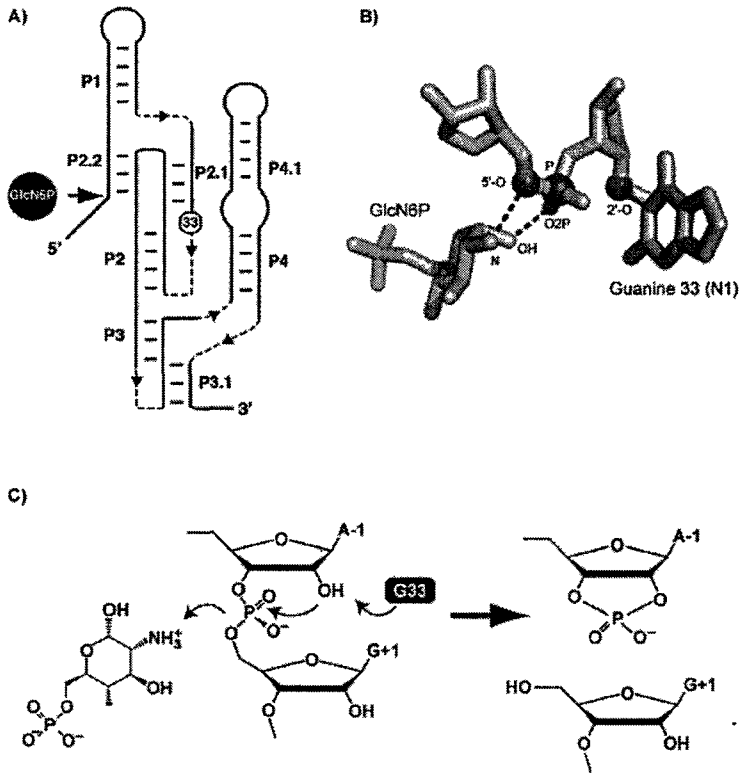


Figure 3. The *glmS* ribozyme. A) The secondary structure arrangement of the *glmS* ribozyme is shown. The cleavage site is indicated by a large arrow and the key nucleobase G33 is shown. Discontinuous lines represent zero length-linker regions. B) The cleavage site of the *glmS* ribozyme. The 2'OH (2'-O), the phosphate (P) and the 5′OH (5′-O) leaving group are positioned for an in-line attack. The N1 functional group of G33 is shown very close to the 2'OH as an activator of the catalytic reaction. The bound GlcN6P is shown to make interactions with the 5′OH and the pro-Sp (O2P) oxygen groups. The C1-OH group of GlcN6P is also shown to interact with the phosphate center to stabilize the charge developing on the scissile phosphate. C) The cleavage mechanism of the glmS ribozyme. G33 acts as a general base by deprotonating the 2'OH nucleophile and GlcN6P plays the role of a general acid by protonating the 5′-O leaving group. The cleavage reaction yields 2',3′-cyclic phosphate and 5′OH termini, which is typical of other small nucleolytic ribozymes.

Using a trans-acting form of the ribozyme, the role of GlcN6P on the *glmS* structure was investigated by using FRET analysis.[101] Interestingly, the authors found that the complex formed between the ribozyme and the substrate does not undergo a global conformational change following GlcN6P binding, which was confirmed by using terbium and enzymatic probing. Another study using hydroxyl radical protection and UV cross-linking assays showed that the riboswitch binds the ligand using a pre-folded active site pocket.[102] Together with previous studies, this suggests that GlcN6P assists *glmS* catalysis by directly participating at the chemical step. The role of metal ions in the *glmS* catalytic activity was also recently studied.[103] It was shown that significant levels of activity are supported in presence either of the exchange-inert cobalt (III) complex or of monovalent ion molar concentrations. These results suggest that metal ions are not directly involved in *glmS* catalysis but instead are used for the correct folding of the RNA structure. Supporting results were obtained from NAIM experiments indicating that phosphate oxygens, 2'-hydroxyl and particular nucleobase functional groups are essential for ribozyme catalysis.[104] In this study, it was observed that metal ion contacts are found in the catalytic core when Co(III) is used, suggesting that these metal ion-RNA interactions are not directly involved in the *glmS* catalytic reaction.

Recent crystal structures have shed light on the *glmS* riboswitch catalytic activity and the role of GlcN6P in the cleavage mechanism. [27, 28] Crystal structures obtained for a ribozyme lacking a 2'-hydroxyl at the cleavage site in presence or in absence of the competitive inhibitor Glc6P, and for the cleaved form, revealed that the ribozyme is a very rigid RNA molecule that exhibits a preformed ligand-binding site.[28] A crystal structure solved for a trans-acting form of the ribozyme having a 2'O-methyl substitution at the cleavage site in presence of GlcN6P confirmed previous findings as well as offering new insights into the catalytic mechanism.[27] For instance, as suggested by other structures,[28] this study showed that GlcN6P directly participates in the reaction as a catalytic cofactor using the ethanolamine moiety on the

glucosamine sugar moiety. This work proposes that while the C1-OH group stabilizes developing charge on the scissile phosphate, the C2-NH2 group acts as a general acid to activate the 5′-O leaving group (Figure 3B). Moreover, because there is no functional group (and no metal ion) other than the strongly conserved G33 that is sufficiently close to the nucleophile to activate the 2'-hydroxyl at the cleavage site, it was suggested that G33 could be involved as a general base (Figures 3B and 3C).[27, 28] The *glmS* ribozyme is thus not like a conventional riboswitch that experiences structural reorganization upon ligand binding, but rather uses ligand as cofactor to achieve catalytic activity, much like protein enzymes.

So what is the outcome of *glmS* processed mRNAs? Because of the ribozyme self-cleavage reaction, the *glmS* downstream transcript exhibits an unusual 5′ hydroxyl terminus which specifically targets the molecule for intracellular degradation.[105] The degradation pathway makes use of the widespread RNase J1 that has been proposed to be the functional homolog of RNase E found in *E. coli*. However, although RNase E poorly degrades mRNA transcripts exhibiting a 5′ hydroxyl group, RNase J1 specifically processes mRNA-carrying 5′OH molecules. Importantly, the presence of the 5′OH-specific RNase J1 in *B. subtilis* strongly suggests that additional gene regulation mechanisms might use similar degradation strategies.

4.4 mRNA splicing and stability

Thiamin pyrophosphate (TPP) is a coenzyme derived from thiamin, which is synthesized by most bacteria, fungus and plants.[106] It is an important cofactor in amino acid and carbohydrate metabolic routes.[37] The TPP riboswitch is the most widespread riboswitch known to date and it is so far the only riboswitch found in archaea and eukaryotes.[107] The TPP riboswitch is involved in different types of gene regulation mechanisms ranging from transcription and translation control in

bacteria[31, 32] to splicing modulation in fungi.[108, 109] Notably, it is also involved in the 3′ end processing of some plant mRNA,[110, 111] where its presence in ancient plant taxa has been speculated to indicate that this riboswitch mechanism exists since the emergence of vascular plants, some 400 million years ago.[110]

The complexity and versatility that riboswitches are able to achieve is well illustrated by the TPP-riboswitch-mediated alternative splicing of mRNA in *Neurospora crassa*.[109] It has been observed that the TPP riboswitch is flanked on one side by two 5′ splice sites and on the other side by a single 3′ splice site. When TPP concentration is low, the TPP aptamer masks the proximal 5′ splice site, resulting in the use of the distal 5′ splice site, which ultimately leads to the production of the *NMT1* mRNA. However, when TPP binds the aptamer, an mRNA conformational change takes place that sequesters both the distal 5′ splice site as well as the splicing branch site. The net effect of the conformational change is a reduction in splicing efficiency of the former mRNA species as well as an increase in an alternatively spliced mRNA that results from the use of the proximal 5′ splice site. Importantly, this TPP-dependent mRNA contains μORFs that compete with the translation of the main ORF, which thus repress *NMT1* expression. Interestingly, Breaker and coworkers have also identified a TPP riboswitch in an intron of another mRNA that interrupts the main ORF.[109] They have observed that stop codons present in the intron prevent mRNA translation. However, in presence of thiamine, the splicing of the intron is favored, suggesting that the TPP riboswitch is also acting as a positively regulating riboswitch.

TPP-dependent alternative splicing is also present in plants, where it has been observed to differentially process the mRNA 3′ end.[110, 111] This post-transcriptional regulation mechanism relies on the conditional presence of a poly A signal retained in absence of TPP and associated with the production of a stable mRNA encoding the thiamin biosynthetic gene *THIC*. However, when TPP is abundant in the cell, a

splicing reaction removes the poly A signal and produces a longer transcript carrying various poly-adenylation signals. This latter mRNA is highly unstable and is associated with a reduction of *THIC* expression. Together, these findings suggest that riboswitches are clearly important players in eukaryotic gene expression mechanisms.

4.5 Increasing complexity of riboswitch genetic regulation

Most riboswitches have been found to control gene products that are involved either in the biosynthesis or transport of the associated metabolite. However, some riboswitches have also been predicted to regulate genes that are not related to their regulating metabolite.[112] For instance, in some alpha-proteobacteria, actinobacteria, *Bacillus* species, *B. fragilis* and *T. elongatus*, the isozymes methionine synthase *MetE* and ribonucleotide reductases NrdDG and NrdAB have B_{12}-specific riboswitches located in their upstream regions (Figure 4). While these gene products operate in a B_{12}-independent fashion, a group of isozymes (MetH and NrdJ, respectively) mostly found in the same organisms are not regulated by a riboswitch-based mechanism but employ vitamin B_{12} as coenzyme. It was thus proposed that when vitamin B_{12} is present at a sufficient cellular concentration, the expression of B_{12}-independent isozymes MetE and NrdDG/NrdAB is repressed via a riboswitch regulation mechanism, and only the more efficient B_{12}-dependent isozymes are used to perform the biological reactions.[112] In addition, a slightly different situation has recently been described in *B. clausii* where the expression of MetE and MetH are also controlled in a riboswitch-dependent manner.[76] However, in this case, it has been observed that the *metH* mRNA contains a SAM-specific riboswitch while the *metE* mRNA exhibits an unusual riboswitch tandem configuration containing two complete riboswitch units responding to SAM and to B_{12}. Because both enzymes are involved in the biosynthesis of SAM, it appears logical that both mRNA can be

regulated by SAM cellular concentration. The rationale for the presence of a B_{12} riboswitch element in *metE* stems from the lower efficiency of the *metE* enzyme to catalyze the formation of homocysteine compared to *metH*, which uses a derivative of adenosylcobalamin, methylcobalamin, as a cofactor. Thus, when adenosylcobalamin is available in cells, expression of *metE* is downregulated to favor the use of the more efficient *metH* enzyme in SAM production.[76]

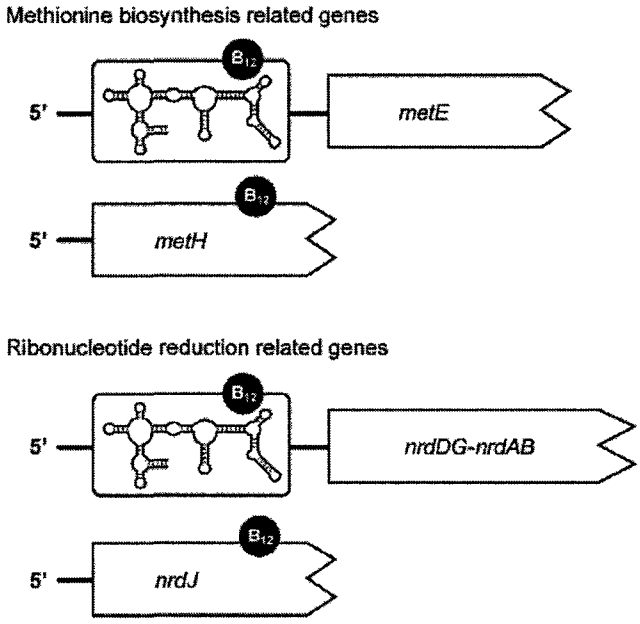


Figure 4. Riboswitches act as gene regulation elements to favor particular biochemical pathways. The MetE and NrdDG-NrdAB enzymes are regulated by B_{12}-specific riboswitches, which are expected to inhibit the expression of enzymes at elevated intracellular B_{12} concentration. However, the expresion of MetH and NrdJ B_{12}-dependent enzymes should not be affected.

Furthermore, the B_{12}-dependent riboswitch-driven regulation is not restrained to SAM and B_{12} biochemical pathways. A recent comparative genomic study in *T. denticola* has predicted two alternative pathways of glutamate utilization and identified B_{12}-specific riboswitches upstream of genes involved in glutamate

metabolism.[112] It was reported that a B_{12}-independent catabolic glutamate dehydrogenase *rocG* is controlled by a B_{12} riboswitch. Interestingly, it was observed that an ortholog of MutSL, which catalyzes the first step of the B_{12}-dependent pathway of glutamate catabolism, is also present in the same bacterium. These observations suggest that an excess of vitamin B_{12} would repress the expression of *rocG* while still allowing MutSL to catabolize glutamate. These findings suggest that B_{12}-independent isozymes are regulated by B_{12}-specific riboswitches in organisms containing both B_{12}-dependent and B_{12}-independent isozymes. Although these observations need to be experimentally verified, they nevertheless suggest that riboswitch elements can be used to add a further level of regulation complexity by favoring more efficient biochemical pathways in certain cellular conditions.

5. Naturally- and artificially-occurring aptamers

Thanks to their nucleic acid architecture, riboswitches exhibit a surprisingly high capacity and plasticity to create binding sites for targeted cellular ligands. The riboswitch plasticity, central to the ligand binding specificity, is best shown in the case of the SAM riboswitches for which four different aptamers are known.[34, 48-51] However, the plasticity of the ligand binding site can also be evaluated for other riboswitch classes given that related artificial aptamers have been selected by SELEX.[113] In general, natural riboswitches tend to be larger than their in vitro evolved counterparts.[114, 115] We refer the reader to a recent review covering artificial aptamers binding AMP, GTP, vitamin B_{12} and FMN[114] describing their structural basis of ligand binding, which are not covered here. Instead, we will focus on the XBA RNA aptamer selected using affinity chromatography on a xanthine agarose column in which xanthine was linked to the agarose through its C8 position.[113] The aptamer secondary structure is composed of a single stem-loop structure containing an asymmetric internal loop (Figure 5A). By comparing the different sequences

obtained, Yokoyama and coworkers observed that most of the internal loop is highly conserved, suggesting that it is involved in the recognition of the ligand.[113] From a survey of various metabolites, the authors found that xanthine, guanine and hypoxanthine bind with similar affinities ($K_D \sim \mu M$) but that adenine is strongly discriminated (Figure 5B). Interestingly, such a molecular discrimination is highly reminiscent of the guanine riboswitch aptamer.[42] While the asymmetric internal loop of the XBA aptamer is likely to be involved in ligand binding,[113] the *xpt* riboswitch achieves binding by using a series of stacking and hydrogen bonding interactions with guanine. Moreover, the ligand specificity is achieved by the formation of a Watson-Crick base pair between the guanine ligand and a conserved cytosine (C74) located in the aptamer core (Figure 5C). Extensive riboswitch-ligand interactions ensure that high affinity is achieved for the formation of the complex leading to dissociation constants in the low nanomolar range.[42] However, typical in vitro selection methods that make use of a ligand bound to a solid support represent a potential problem for the detection of high affinity aptamers given that it does not allow the RNA to interact with all ligand functional groups, which is not the case for natural riboswitch aptamers.

Nevertheless, other ways to select aptamers have been put forward to circumvent the use of a solid support for aptamer selection. For instance, Nomura & Yokobayashi have developed a selection scheme that allowed to re-engineer the TPP riboswitch through an in vivo SELEX procedure.[116] In this work, the authors randomized a part of the expression platform of a negatively regulating riboswitch and selected for a riboswitch that allows gene expression in presence of TPP. A very similar approach was previously employed by Lynch et al. to select for synthetic riboswitches that activate protein translation in presence of theophylline.[117] Such approaches are very powerful to harness the huge possibilities offered by riboswitches as gene regulation elements and could represent an improved way to develop artificial aptamers that mimic more efficiently nature's work.

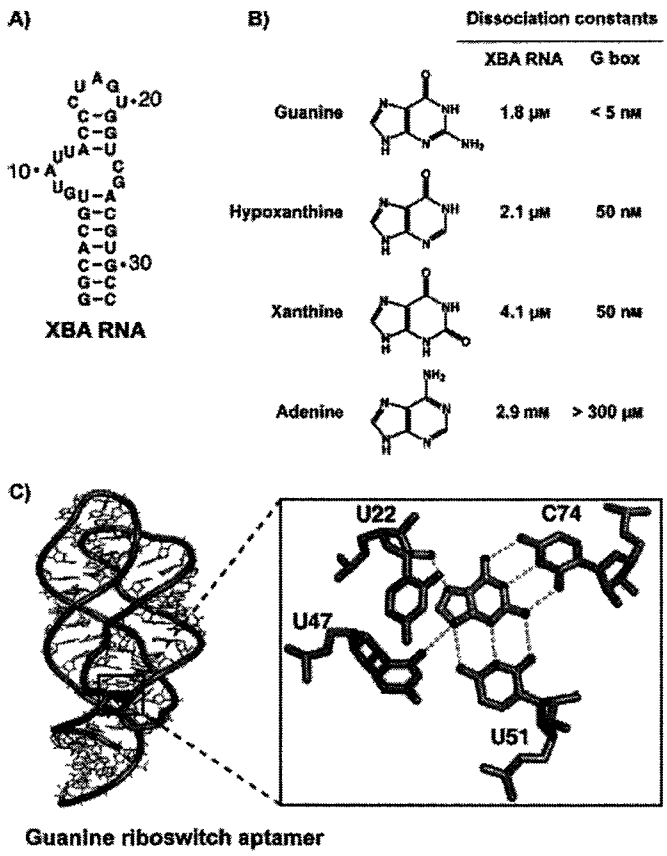


Figure 5. The XBA RNA and the guanine riboswitch aptamers. A) The secondary structure of the XBA RNA is shown and the numbering is according to Kiga *et al.* [113] B) Dissociation constants are indicated for guanine, hypoxanthine, xanthine and adenine for both the XBA RNA[113] and the guanine riboswitch aptamer.[54] Chemical structures for ligands are shown. C) The ligand binding site of the *xpt* guanine riboswitch aptamer (PDB accession code 1Y27). The guanine aptamer and the ligand binding site (inset) are shown. Nucleotides making direct H-bond interactions with the bound guanine are indicated.

The size of natural and artificial aptamers, which seems to be the most obvious difference, can also be explained by their respective aim. For instance, artificial aptamers are isolated solely by their ability to bind a particular ligand. However,

riboswitches have a more complex role to play by achieving gene regulation upon ligand binding. Riboswitch regulation is possible through the global RNA reorganization and P1 stem stabilization induced by ligand binding, which require essential RNA-ligand interactions. This may explain the high variability in size and structure of natural riboswitches, which evolved to encompass their respective ligand. The extended conformation of the TPP riboswitch represents well this ligand-riboswitch co-evolution, where the RNA conformation allows two binding pockets to be formed, which interact with the 4-amino-5-hydroxymethyl-2-methylpyrimidine and the pyrophosphate moieties of the ligand.[118] However, recently discovered preQ1[47] and metal sensing[119] riboswitches seem to head off this generally observed correlation existing between the size of riboswitches and their cognate ligand. Whilst the modified purine preQ1 binds to a single stem-loop riboswitch with great affinity, the M-box is composed of a more complex structure that binds a ligand as small as six magnesium ions, suggesting that many unexpected metabolite binding RNA structures await to be discovered, most probably displaying higher affinity than known in vitro evolved aptamers.

6. Riboswitch differential in vivo gene regulation

Each riboswitch class shows a high level of sequence and secondary structure conservation among its representatives. However, some differences are apparent in the aptamer region that could be linked to a different ligand-dependent gene regulation response as observed in two recent studies performed in *B. subtilis*.[53, 120] In one of the studies aimed at analyzing differences in gene regulation for SAM riboswitches, it was shown that expression levels can vary for each riboswitch representative as a function of concentration and time.[120] When *B. subtilis* cells were put under starvation conditions, it was observed using quantitative RT-PCR that mRNA levels of most riboswitches showed an increase after 30 min of starvation

period. However, in the case of *metK*, it was observed that gene expression initially increases during the first 30 minutes and then returns to pre-starvation levels after 60 minutes. In addition, when gene expression was monitored using *lacZ* transcriptional fusions, it was found that all but two genes are positively regulated in starvation conditions. The two candidates that do not experience riboswitch regulation, *metK* and *cysH*, showed no induction but one of them exhibited an increase of gene expression during the first part of the methionine starvation, followed by a return to its basal expression. This observation could also suggest that this particular gene might be controlled at the transcriptional level by other mechanisms. From this in vivo study, the authors conclude that 10 of the 11 SAM riboswitch variants found in *B. subtilis*, although being functional, exhibit a high variability in gene expression levels, probably related to the gene function itself.

An additional study was recently performed on guanine riboswitches in which 89 aptamer representatives were analyzed using sequence alignment, from which a representative subset was also studied biochemically.[53] Using in-line probing and 2-aminopurine fluorescence assays, it was shown that guanine aptamers exhibit a large binding affinity spectrum where variations in affinity could be as much as 800-fold, suggesting that although a high homology in secondary structure is present, subtle sequence differences in the aptamer core domain have important consequences for the aptamer ligand binding activity. More precisely, when examining binding affinities for *xpt*, *purE* and *yxjA* aptamers present in *B. subtilis*, it was observed that whilst *xpt* and *purE* exhibit similar binding affinities, *yxjA* shows a 13-fold higher binding affinity. As a way to directly monitor the *B. subtilis* riboswitch gene regulation control in a biological context, a RT-qPCR strategy was employed in which the ratio of the full length mRNA versus the prematurely truncated transcript was determined. The rationale behind these experiments is that the riboswitch mRNA should be prematurely truncated when cells are grown in presence of a sufficiently high guanine concentration, and inversely, the presence of the full length

species should be favored when cells are grown in conditions where guanine is at low concentration. When growing cells are exposed at different concentrations of guanine, *xpt*, *purE* and *yxjA* representatives show a unique gene expression profile. For instance, while the *xpt* riboswitch does not inhibit gene expression in presence of 0.25 mg/ml guanine, the *purE* representative shows a significant premature transcription termination (~20%), which is further enhanced (~80%) by incubating cells at saturating guanine conditions. Higher transcription termination efficiencies were obtained with the *yxjA* variant, which yielded ~55% and ~98% efficiencies at 0.25 mg/ml and saturating conditions, respectively. This suggests that guanine riboswitches show differential regulation in their gene control activity.

Since all three guanine riboswitch representatives are present in the same cellular environment, it may be informative to compare their thermodynamic properties and to correlate them with their gene regulation activity. When estimating the relative free energies for both terminator and antiterminator structures,[53] the tendency to adopt the terminator structure ($\Delta G^{Term}/\Delta G^{Anti}$) can be established for each riboswitch (Figure 6). In principle, this value is an indication of the propensity of premature transcription termination in conditions where the aptamer domain is fully bound, i.e., when the intracellular concentration of guanine is high. It was observed that this propensity increases in the order *xpt* < *purE* < *yxjA* suggesting that *yxjA* is the riboswitch representative that should be associated with the most efficient termination efficiency at saturating guanine concentrations. RT-qPCR data are in excellent agreement with this observation, consistent with the idea that the gene regulation process is mainly controlled by the RNA-ligand binding affinity. In addition, it appears that the affinity is directly proportional to the transcription termination propensity (Figure 6). Whether or not the ligand binding affinity is an indicator of the gene regulation efficiency remains to be determined in other systems.

Variants	2-aminopurine affinity	$\frac{\Delta G^{term}}{\Delta G^{anti}}$
xpt	7.8 μM	1.4
purE	8.2 μM	2.4
yxjA	0.59 μM	3.1

Increasing regulation efficiency

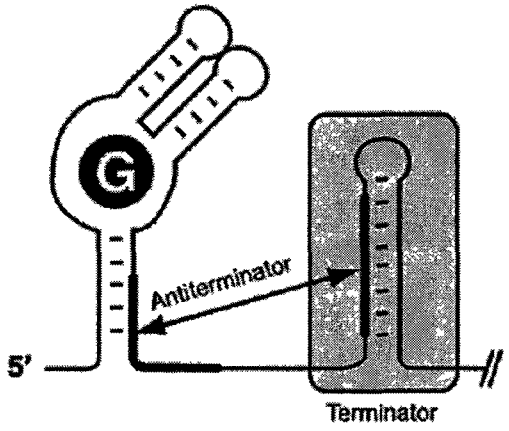


Figure 6. Gene regulation activity of selected *B. subtilis* guanine riboswitches. The ligand binding affinity of the aptamer domain as well as the propensity to terminate transcription are shown for *xpt*, *purE* and *yxjA* representatives. The increasing gene regulation control, as determined from the increase of both the ligand binding affinity and the propensity for premature transcription termination, is shown by an arrow. A secondary structure representation of a guanine riboswitch is shown where both the terminator and the anti-terminator structures are indicated. The relative free energies were calculated with the program *mfold*.[153]

Lastly, of all guanine riboswitches found in *B. subtilis*, the *yxjA* representative is the only one which is not under the control of the PurR repressor that is involved in transcription initiation regulation of several genes involved in purine synthesis, metabolism, transport and cofactor synthesis. [121, 122] The affinity of PurR for its

DNA control region is influenced by 5-phosphoribosyl-1-pyrophosphate (PRPP), which inhibits PurR-DNA binding in vitro and may thus be a transcription inducer in vivo.[123] Because *yxjA* gene regulation control does not rely on PurR transcription initiation control, it is tempting to speculate that a tighter riboswitch regulation control must be in place to ensure proper expression.

7. Riboswitch-based applications

7.1 Riboswitches as biosensor molecular tools

In principle, because the riboswitch binding site and expression platform domains are modular, one could envisage to genetically alter one domain without affecting the function of the other. Doing so could enable to alter the ligand binding specificity of the riboswitch while keeping the gene regulation control unaltered. Thus, by combining the possibility to develop novel aptamers by SELEX and the ability of riboswitches to control gene expression in a ligand-dependent manner, the generation of novel biosensors becomes readily attainable. By exploiting biological reporters such as the green fluorescent protein, β-galactosidase or luciferase, synthetic riboswitches can be used to monitor the presence of virtually any metabolites in solution thus providing molecular tools that can be tailored for very specific needs. One of the first example for such a control in vivo was provided in an eukaryotic system where a small molecule aptamer was developed against the cell-permeable Hoechst dye 33258.[124] By inserting the aptamer between the 5′-m^7G cap and the start codon, it was shown that the formation of a stable complex between the RNA and the dye precludes the ribosome scanning process required for protein translation. Since then, various studies have explored the potential of synthetic riboswitches in bacteria as well as in eukaryotes.[117, 125-134, 135 , 136]

A high-throughput compatible assay was recently developed[136] using the *glmS* ribozyme,[26] which is specifically activated in presence of glucosamine-6-phosphate (GlcN6P). In this study, Mayer and Famulok performed a model screening study using commercially available compounds to detect metabolites able to support the ribozyme self-cleavage activity, which would reveal to be a powerful approach against microbial agents carrying the *glmS* ribozyme element. For this purpose, the authors used a 5′-fluorescein labeled ribozyme and measured the fluorescence polarization (FP) in various experimental conditions. The principle of fluorescence polarization relies on the fact that low molecular weight molecules rotate and tumble at a faster rate than high molecular weight molecules.[137] It is expected that if the fluorescent ribozyme is cleaved in the presence of a competent metabolite, the 4-mer reaction product carrying the fluorescence reporter will exhibit a significant reduction in the observed FP. However, no change in polarization should be detected in presence of a metabolite unable to support catalytic activity. Of all the 88 tested compounds, none were found to support catalytic cleavage indicating that none could replace GlcN6P as a *glmS* ribozyme cofactor. Nevertheless, this study clearly demonstrates that high-throughput screening assays can be successfully adapted to identify novel antimicrobial drugs targeting riboswitches, as well as any novel metabolites used for controlling gene expression.

Gallivan and coworkers have also employed an automated screening method to identify riboswitches exhibiting better performances in the signal to noise ratio.[117] They used a previously described theophylline aptamer, [138] which was fused to a β-galactosidase reporter gene in order to control the access of the ribosomal binding site. In this study, theophylline riboswitch mutants were developed by high-throughput screening in 96 wells microplate using randomized variants of theophylline riboswitches coupled to *lacZ* reporter in *E. coli* expression system. This mutational study focused on the segment located between the aptamer and the ribosome binding site (RBS), and one mutant was found to increase the signal to

background ratio from 8- to 36-fold. One of the best clones obtained, clone 8.1, displays a predicted secondary structure for the sequestered RBS that is more stable than the one in which the RBS is unpaired ($\Delta\Delta G$ = 5.5 kcal/mol), consistent with the reduced background riboswitch activity that the authors have obtained in their study. Moreover, when determining the free energy of theophylline binding, it was found that the free energy was lowered by 9.2 kcal/mol, which supports the high activation ratio that the riboswitch displays upon ligand binding. This study shows that in vitro-selected aptamers can be readily converted into efficient riboswitches that perform gene regulation in a cellular environment.

7.2 Riboswitches as antibacterial targets

Bacterial resistance to antibiotics is a growing problem since the last 10 years, particularly in hospitals where the excessive and prolonged use of antibiotics has allowed some bacteria to acquire multiple drug resistance (MDR) following repeated antibiotic exposure. A second cause that can explain the dramatic increase of bacterial resistance is that during the last 30 years, only one new antibiotic chemical scaffold was produced,[8] leaving an opportunity for pathogenic microorganisms to circumvent the action mechanism of antibiotics. As the pharmaceutical industry chose to develop new antibiotics mostly by modifying existing ones, bacteria became more resistant to new compounds given their similarity to already existing ones. The limited number of cellular processes targeted further promotes this increase in bacterial survival.[8, 139]

Riboswitches offer a tantalizing solution to the alarming and fast growing MDR problem. Given the high specificity of riboswitches towards ligands that do not change significantly during evolution, the aptamer domain to which they bind is highly conserved. Therefore, it should be possible to design ligand analogs that bacteria cannot metabolize, thus shutting down the expression of a particular

riboswitch's regulated gene. However, mutations may be acquired that may disrupt the riboswitch gene expression control and therefore promote bacterial resistance, Precedents for this have been observed in the past for the adenine riboswitch,[122] where *B. subtilis* mutant strains found to be resistant to the mutagen 2-fluoroadenine exhibited mutations that caused overexpression of the *pbuE* gene product. The overexpressing mutations are located in the *pbuE* adenine riboswitch where portions of the terminator stem were deleted thus preventing the OFF state of the riboswitch to be adopted and favoring the constitutive expression of *pbuE*.

The compound S-(2-aminoethyl)-L-cysteine (AEC) has for a long time been known as an antimicrobial agent.[140] It was also established that AEC is incorporated into proteins, which could promote toxic effects to the cell.[141, 142] Various *E. coli* and *B. subtilis* mutant strains,[143-145] resistant to AEC, were found to cause de-repression of *lysC* expression. The mutations were located in the lysine aptamer domain,[45] suggesting that the toxic effects might be due to the repression of aspartokinase expression following binding of AEC to lysine riboswitch. A *lacZ* reporter gene construct was used to monitor the lysine riboswitch control on the expression when *B. subtilis* is grown in absence or in presence of lysine. Whilst β-galactosidase expression was inhibited in presence of lysine in wild-type bacteria, no such inhibition was observed when experiments were done using AEC resistant mutants.

In an attempt to target the lysine riboswitch for antibacterial therapy, a recent study has explored the possibility to design new compounds that bind the riboswitch and suppress lysine biosynthesis and transport genes.[146] Four compounds were found to exhibit high binding affinity and consist of L-3-[(2-aminoethyl)-sulfonyl]-alanine, L-4-oxalysine, L-homoarginine and DL-trans-2,6-diamino-4-hexenoic acid for which dissociation constants of 2.5, 13, 7, 0.97 μM were found, respectively. Compared to values of 1 μM and 30 μM for lysine and AEC, these high binding affinitiy molecules show a high potential to be used as antibiotic compounds. Of these four compounds

only three were able to inhibit bacteria growth.[146] These compounds were then tested for their ability to repress the expression of a β-galactosidase reporter and were found as efficient as lysine to inhibit gene expression. To confirm that the antibacterial effects are due to the inhibition of *lysC* by the compounds, bacteria were selected for their resistance for L-4-oxalysine by serial passage. The *B. subtilis lysC* and *yvsH* lysine riboswitches present in L-4-oxalysine-resistant bacteria were sequenced and the observed mutations in the *lysC* riboswitch domain suggest that the effects were caused by the specific interaction between the antibiotic compound and the lysine riboswitch. Moreover, two resistant clones were characterized for their resistance to L-3-[(2-aminoethyl)-sulfonyl]-alanine and DL-trans-2,6-diamino-4-hexenoic acid and for their ability to express β-gactosidase. As expected, clones were resistant to both molecules as no significant inhibition of β-galactosidase production was observed in absence or in presence of ligands. Together, these results are consistent with the idea that toxic effects are due to the inhibition of aspartokinase expression following binding of lysine analogues to the lysine riboswitch.

A second example is the finding of pyrithiamine (PT) acting mode.[147] PT is an isosteric pyrimidine analog of thiamine known to be toxic for fungi and bacteria,[148, 149] which is phosphorylated in cells to pyrithiamine pyrophosphate (PTPP).[150] PT was known to be an inhibitor of thiamine pyrophosphorylase and its product, PTPP, to inhibit enzymes that use TPP as coenzyme. [150] Nevertheless, this knowledge can not explain the toxic effect of PT and its derivative on bacteria which are able to synthesize TPP de novo and might compensate the inhibitory effect by upregulating thiamine production.[151] The TPP riboswitch binds TPP with a dissociation constant of 100 nM whereas thiamine phosphate (TP), which only differs from TPP by one missing phosphate, exhibits a constant of at least 100-fold higher [29]. The riboswitch was shown to be able to prematurely stop transcription [31] and to repress the expression of β-galactosidase reporter under the control of TPP riboswitch [29]. It was

also found that PT, and more efficiently PTPP, were able to bind the TPP riboswitch with a dissociation constant of 6 μM and 100 nM, respectively,[147] the latter being comparable to values obtained with TPP. The use of a β-galactosidase reporter allowed determining that PT is able to inhibit the expression of a gene under the control of the TPP riboswitch, thus highlighting its potential use as an antimicrobial compound. The DNA coding for TPP riboswitches from resistant clones obtained by serial passage were sequenced and mutations were located in TPP riboswitches. The introduction of these mutations in the context of a wild-type riboswitch render inactive the riboswitch to elevated ligand concentrations.

Another example was found for roseoflavin, which is an antibiotic involved in the inhibition of the riboflavin biosynthesis.[152] Because of its similarity in structure with FMN, roseoflavin might potentially target the FMN riboswitch controlling the operon involved in the riboflavin biosynthesis.[35, 41] It was recently reported that roseoflavin is able to bind FMN riboswitch in vitro.[8] As found in the other cases, mutations obtained when growing bacteria in presence of analogues where found in FMN riboswitch domains, again consistent with the idea that riboswitches can be used as antimicrobial cellular targets.

8. Outlook

The fascinating discovery that certain RNA structures found in mRNAs can recognize small metabolites using a spectrum of biochemical interactions and thus, control gene regulation at different stages, has greatly contributed to place RNA at the forefront of modern biology, highlighting its active role in many cellular processes and further expanding its apparently unlimited functional possibilities. Furthermore, because riboswitches appear to strictly rely on RNA to regulate gene expression, they have been proposed as remnants of an ancient biological environment where

RNA was the predominant functional species, before proteins emerged as main effectors. Several features of RNA switches including their evolutionary conserved architecture, probably derived from the need to recognize metabolites that have remain unchanged through evolution, together with the presence of riboswitches in diverse organisms, have led to a renaissance of the RNA World hypothesis.

Regardless of whether these structures represent relics of an ancient RNA-based world or are more recent additions to the biochemical repertoire, riboswitches constitute an emerging field with many exciting tasks to be accomplished in the future. Among these, it is crucial to elucidate in detail the interplay between RNA folding-ligand recognition and function in existing and novel discovered structures, both at post- and co-transcriptional level. It is vital to establish the mechanistic basis for riboswitch function and to determine how RNA polymerase pausing sites and perhaps, DNA template and RNA secondary structure features, can modulate whether a riboswitch is kinetically or thermodynamic controlled. Comprehensive research in this area would certainly enhance our understanding of RNA folding and function in general and will pave the way to successfully approach more complex RNA-containing structures such as the ribosome.

One of the most fundamental questions that must also be addressed is whether or not additional riboswitches are present in eukaryotes. The fact that few eukaryotic riboswitches have been discovered should not be taken as a negative answer, but rather as an indicator that they are most probably involved in the control of different molecular processes than those found in bacteria. Striking examples of this have already been reported for splicing controlling riboswitches, suggesting that even more complex structures await to be discovered in eukaryotic organisms.

Given the increasing bacterial resistance to existing antibiotics, it is also a priority to explore the feasibility of riboswitch-based therapeutics. Because of their

conservation, it may be possible to use the presence of riboswitches in pathogenic bacteria as a way to develop novel types of antibiotic compounds aimed at disrupting riboswitch cellular activity, provided that these representatives are not found in humans. Lastly, riboswitches can also be viewed as promising gene controlling elements to develop novel bimolecular tools or modules that could trigger specific cellular mechanisms in response to natural or artificial cellular metabolites.

Acknowledgements

We thank Dr Alain Lavigueur and all members of the Lafontaine laboratory for fruitful discussions and critical reading of the manuscript. This work was supported by the National Sciences and Engineering Research Council of Canada (SB and JM), the Canadian Institutes of Health Research (DAL) and the Scottish Universities Physics Alliance (JCP). DAL is a CIHR New Investigator scholar as well as a Chercheur-boursier Junior 2 from the Fonds de la recherche en Santé du Québec. We thank Dr Benoit Leblanc for the artistic representation of the cover page.

Keywords: antimicrobial agents, gene regulation mechanisms, riboswitches, ribozymes, RNA-ligand interactions

[1] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, Cell 1983, 35, 849-857.

[2] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, Cell 1982, 31, 147-157.

[3] L. Gold, B. Polisky, O. Uhlenbeck, M. Yarus, Annu. Rev. Biochem. 1995, 64, 763-797.

[4] G. F. Joyce, Angew. Chem. Int. Ed. Engl. 2007, 46, 6420-6436.

[5] T. E. Edwards, D. J. Klein, A. R. Ferre-D'Amare, Curr. Opin. Struct. Biol. 2007, 17, 273-279.

[6] A. Serganov, D. J. Patel, Nat. Rev. Genet. 2007, 8, 776-790.

[7] H. Schwalbe, J. Buck, B. Furtig, J. Noeske, J. Wohnert, Angew. Chem. Int. Ed. Engl. 2007, 46, 1212-1219.

[8] K. F. Blount, R. R. Breaker, Nat. Biotechnol. 2006, 24, 1558-1564.

[9] D. G. Sashital, S. E. Butcher, ACS Chem. Biol. 2006, 1, 341-345.

[10] W. C. Winkler, Curr. Opin. Chem. Biol. 2005, 9, 594-602.

[11] B. Krautler, Angew. Chem. Int. Ed. Engl. 2005, 44, 4288-4290.

[12] B. J. Tucker, R. R. Breaker, Curr. Opin. Struct. Biol. 2005, 15, 342-348.

[13] W. C. Winkler, Arch. Microbiol. 2005, 183, 151-159.

[14] A. Lescoute, E. Westhof, Chem. Biol. 2005, 12, 10-13.

[15] R. R. Breaker, Nature 2004, 432, 838-845.

[16] M. Mandal, R. R. Breaker, Nat. Rev. Mol. Cell. Biol. 2004, 5, 451-463.

[17] J. K. Soukup, G. A. Soukup, Curr. Opin. Struct. Biol. 2004, 14, 344-349.

[18] F. J. Grundy, T. M. Henkin, Curr. Opin. Microbiol. 2004, 7, 126-131.

[19] E. Nudler, A. S. Mironov, Trends Biochem. Sci. 2004, 29, 11-17.

[20] A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, M. S. Gelfand, Trends Genet. 2004, 20, 44-50.

[21] W. C. Winkler, R. R. Breaker, Chembiochem 2003, 4, 1024-1032.

[22] E. C. Lai, Curr. Biol. 2003, 13, R285-291.

[23] J. P. Gallivan, Curr. Opin. Chem. Biol. 2007.

[24] R. K. Montange, R. T. Batey, Annual Review of Biophysics. 2008, 37, 117-133.

[25] M. T. McManus, P. A. Sharp, Nat. Rev. Genet. 2002, 3, 737-747.

[26] W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins, R. R. Breaker, Nature 2004, 428, 281-286.

[27] J. C. Cochrane, S. V. Lipchock, S. A. Strobel, Chem. Biol. 2007, 14, 97-105.

[28] D. J. Klein, A. R. Ferre-D'Amare, Science 2006, 313, 1752-1756.

[29] W. Winkler, A. Nahvi, R. R. Breaker, Nature 2002, 419, 952-956.

[30] A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, R. R. Breaker, Chem. Biol. 2002, 9, 1043.

[31] 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.

[32] J. Miranda-Rios, M. Navarro, M. Soberon, Proc. Natl Acad. Sci. U. S. A. 2001, 98, 9736-9741.

[33] F. J. Grundy, S. C. Lehman, T. M. Henkin, Proc. Natl Acad. Sci. U. S. A. 2003, 100, 12057-12062.

[34] B. A. McDaniel, F. J. Grundy, I. Artsimovitch, T. M. Henkin, Proc. Natl Acad. Sci. U. S. A. 2003, 100, 3083-3088.

[35] M. S. Gelfand, A. A. Mironov, J. Jomantas, Y. I. Kozlov, D. A. Perumov, Trends Genet. 1999, 15, 439-442.

[36] X. Nou, R. J. Kadner, Proc. Natl Acad. Sci. U. S. A. 2000, 97, 7190-7195.

[37] G. D. Stormo, Y. Ji, Proc. Natl Acad. Sci. U. S. A. 2001, 98, 9465-9467.

[38] R. A. Kreneva, D. A. Perumov, Mol. Gen. Genet. 1990, 222, 467-469.

[39] B. Y. Wei, C. Bradbeer, R. J. Kadner, Res. Microbiol. 1992, 143, 459-466.

[40] M. Mandal, R. R. Breaker, Nature Structural & Molecular Biology 2004, 11, 29-35.

[41] W. C. Winkler, S. Cohen-Chalamish, R. R. Breaker, Proc. Natl Acad. Sci. U. S. A. 2002, 99, 15908-15913.

[42] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, Cell 2003, 113, 577-586.

[43] J. N. Kim, A. Roth, R. R. Breaker, Proc. Natl Acad. Sci. U. S. A. 2007, 104, 16092-16097.

[44] M. Mandal, M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo, R. R. Breaker, Science 2004, 306, 275-279.

[45] N. Sudarsan, J. K. Wickiser, S. Nakamura, M. S. Ebert, R. R. Breaker, Genes Dev. 2003, 17, 2688-2697.

[46] E. E. Regulski, R. H. Moy, Z. Weinberg, J. E. Barrick, Z. Yao, W. L. Ruzzo, R. R. Breaker, Mol. Microbiol. 2008.

[47] A. Roth, W. C. Winkler, E. E. Regulski, B. W. Lee, J. Lim, I. Jona, J. E. Barrick, A. Ritwik, J. N. Kim, R. Welz, D. Iwata-Reuyl, R. R. Breaker, Nature Structural & Molecular Biology 2007, 14, 308-317.

[48] V. Epshtein, A. S. Mironov, E. Nudler, Proc. Natl Acad. Sci. U. S. A. 2003, 100, 5052-5056.

[49] F. J. Grundy, T. M. Henkin, Mol. Microbiol. 1998, 30, 737-749.

[50] W. C. Winkler, A. Nahvi, N. Sudarsan, J. E. Barrick, R. R. Breaker, Nat. Struct. Biol. 2003, 10, 701-707.

[51] Z. Weinberg, E. E. Regulski, M. C. Hammond, J. E. Barrick, Z. Yao, W. L. Ruzzo, R. R. Breaker, RNA. 2008, 14, 822-828.

[52] J. X. Wang, E. R. Lee, D. R. Morales, J. Lim, R. R. Breaker, Mol. Cell. 2008, 29, 691-702.

[53] J. Mulhbacher, D. A. Lafontaine, Nucleic Acids Res. 2007, 35, 5568-5580.

[54] R. T. Batey, S. D. Gilbert, R. K. Montange, Nature 2004, 432, 411-415.

[55] A. Serganov, Y. R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A. T. Phan, C. Hobartner, R. Micura, R. R. Breaker, D. J. Patel, Chem. Biol. 2004, 11, 1729-1741.

[56] R. Rieder, K. Lang, D. Graber, R. Micura, Chembiochem 2007, 8, 896-902.

[57] S. Chowdhury, C. Maris, F. H. Allain, F. Narberhaus, EMBO J. 2006, 25, 2487-2497.

[58] J. Johansson, P. Mandin, A. Renzoni, C. Chiaruttini, M. Springer, P. Cossart, Cell 2002, 110, 551-561.

[59] M. De la Pena, S. Gago, R. Flores, EMBO J. 2003, 22, 5561-5570.

[60] A. Khvorova, A. Lescoute, E. Westhof, S. D. Jayasena, Nat. Struct. Biol. 2003, 10, 708-712.

[61] J. C. Penedo, T. J. Wilson, S. D. Jayasena, A. Khvorova, D. M. Lilley, RNA. 2004, 10, 880-888.

[62] M. Martick, W. G. Scott, Cell 2006, 126, 309-320.

[63] C. D. Stoddard, S. D. Gilbert, R. T. Batey, RNA. 2008, 14, 675-684.

[64] J. Noeske, J. Buck, B. Furtig, H. R. Nasiri, H. Schwalbe, J. Wohnert, Nucleic Acids Res. 2007, 35, 572-583.

[65] J. Noeske, H. Schwalbe, J. Wohnert, Nucleic Acids Res. 2007, 35, 5262-5273.

[66] J. F. Lemay, J. C. Penedo, R. Tremblay, D. M. Lilley, D. A. Lafontaine, Chem. Biol. 2006, 13, 857-868.

[67] S. D. Gilbert, R. P. Rambo, D. Van Tyne, R. T. Batey, Nature Structural & Molecular Biology. 2008, 15, 177-182.

[68] R. K. Montange, R. T. Batey, Nature. 2006, 441, 1172-1175.

[69] C. Lu, A. M. Smith, R. T. Fuchs, F. Ding, K. Rajashankar, T. M. Henkin, A. Ke, Nature Structural & Molecular Biology. 2008.

[70] J. Lim, W. C. Winkler, S. Nakamura, V. Scott, R. R. Breaker, Angew. Chem. Int. Ed. Engl. 2006, 45, 964-968.

[71] S. Blouin, D. A. Lafontaine, RNA 2007, 13, 1256-1267.

[72] B. Heppell, D. A. Lafontaine, Biochemistry 2008, 47, 1490-1499.

[73] N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, Science 2000, 289, 905-920.

[74] D. J. Klein, T. M. Schmeing, P. B. Moore, T. A. Steitz, EMBO J. 2001, 20, 4214-4221.

[75] M. Kwon, S. A. Strobel, RNA 2007, 14, 25-34.

[76] N. Sudarsan, M. C. Hammond, K. F. Block, R. Welz, J. E. Barrick, A. Roth, R. R. Breaker, Science 2006, 314, 300-304.

[77] R. Welz, R. R. Breaker, RNA 2007, 13, 573-582.

[78] S. Altuvia, D. Kornitzer, D. Teff, A. B. Oppenheim, J. Mol. Biol. 1989, 210, 265-280.

[79] M. T. Morita, Y. Tanaka, T. S. Kodama, Y. Kyogoku, H. Yanagi, T. Yura, Genes Dev. 1999, 13, 655-665.

[80] M. J. Cromie, Y. Shi, T. Latifi, E. A. Groisman, Cell 2006, 125, 71-84.

[81] A. Gutierrez-Preciado, R. A. Jensen, C. Yanofsky, E. Merino, Trends Genet. 2005, 21, 432-436.

[82] Z. Weinberg, J. E. Barrick, Z. Yao, A. Roth, J. N. Kim, J. Gore, J. X. Wang, E. R. Lee, K. F. Block, N. Sudarsan, S. Neph, M. Tompa, W. L. Ruzzo, R. R. Breaker, Nucleic Acids Res. 2007, 35, 4809-4819.

[83] J. E. Barrick, K. A. Corbino, W. C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J. K. Wickiser, R. R. Breaker, Proc. Natl Acad. Sci. U. S. A. 2004, 101, 6421-6426.

[84] J. P. Richardson, Biochim. Biophys. Acta 2002, 1577, 251-260.

[85] I. Gusarov, E. Nudler, Mol. Cell 1999, 3, 495-504.

[86] A. Schwartz, A. R. Rahmouni, M. Boudvillain, EMBO J. 2003, 22, 3385-3394.

[87] W. S. Yarnell, J. W. Roberts, Science 1999, 284, 611-615.

[88] F. Lillo, S. Basile, R. N. Mantegna, Bioinformatics 2002, 18, 971-979.

[89] J. K. Wickiser, M. T. Cheah, R. R. Breaker, D. M. Crothers, Biochemistry 2005, 44, 13404-13414.

[90] J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, Mol. Cell 2005, 18, 49-60.

[91] S. M. Uptain, C. M. Kane, M. J. Chamberlin, Annu. Rev. Biochem. 1997, 66, 117-172.

[92] A. V. Yakhnin, P. Babitzke, Proc. Natl Acad. Sci. U. S. A. 2002, 99, 11067-11072.

[93] W. C. Winkler, R. R. Breaker, Annu. Rev. Microbiol. 2005, 59, 487-517.

[94] R. T. Fuchs, F. J. Grundy, T. M. Henkin, Nature Structural & Molecular Biology 2006, 13, 226-233.

[95] D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, M. S. Gelfand, J. Biol. Chem. 2002, 277, 48949-48959.

[96] A. R. Ferre-D'Amare, K. Zhou, J. A. Doudna, Nature 1998, 395, 567-574.

[97] A. Serganov, S. Keiper, L. Malinina, V. Tereshko, E. Skripkin, C. Hobartner, A. Polonskaia, A. T. Phan, R. Wombacher, R. Micura, Z. Dauter, A. Jaschke, D. J. Patel, Nature Structural & Molecular Biology 2005, 12, 218-224.

[98] T. J. McCarthy, M. A. Plog, S. A. Floy, J. A. Jansen, J. K. Soukup, G. A. Soukup, Chem. Biol. 2005, 12, 1221-1226.

[99] K. H. Link, L. Guo, R. R. Breaker, Nucleic Acids Res. 2006, 34, 4968-4975.

[100] J. Lim, B. C. Grove, A. Roth, R. R. Breaker, Angew. Chem. Int. Ed. Engl. 2006, 45, 6689-6693.

[101] R. A. Tinsley, J. R. Furchak, N. G. Walter, RNA 2007, 13, 468-477.

[102] K. J. Hampel, M. M. Tinsley, Biochemistry 2006, 45, 7861-7871.

[103] A. Roth, A. Nahvi, M. Lee, I. Jona, R. R. Breaker, RNA 2006, 12, 607-619.

[104] J. A. Jansen, T. J. McCarthy, G. A. Soukup, J. K. Soukup, Nature Structural & Molecular Biology 2006, 13, 517-523.

[105] J. A. Collins, I. Irnov, S. Baker, W. C. Winkler, Genes Dev. 2007, 21, 3356-3368.

[106] R. L. White, I. D. Spencer in Cellular and Molecular Biology, Vol. 1, 2nd Ed (Eds.: F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, H. E. Umbarger, W. (Am. Soc. Microbiol., DC)), 1996, pp.680-686.

[107] J. E. Barrick, R. R. Breaker, Genome Biol. 2007, 8, R239.

[108] T. Kubodera, M. Watanabe, K. Yoshiuchi, N. Yamashita, A. Nishimura, S. Nakai, K. Gomi, H. Hanamoto, FEBS Lett. 2003, 555, 516-520.

[109] M. T. Cheah, A. Wachter, N. Sudarsan, R. R. Breaker, Nature 2007.

[110] S. Bocobza, A. Adato, T. Mandel, M. Shapira, E. Nudler, A. Aharoni, Genes Dev. 2007, 21, 2874-2879.

[111] A. Wachter, M. Tunc-Ozdemir, B. C. Grove, P. J. Green, D. K. Shintani, R. R. Breaker, Plant Cell. 2007.

[112] D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, M. S. Gelfand, J. Biol. Chem. 2003, 278, 41148-41159.

[113] D. Kiga, Y. Futamura, K. Sakamoto, S. Yokoyama, Nucleic Acids Res. 1998, 26, 1755-1760.

[114] C. A. Wakeman, W. C. Winkler, C. E. Dann, 3rd, Trends Biochem. Sci. 2007, 32, 415-424.

[115] D. J. Patel, A. K. Suri, F. Jiang, L. Jiang, P. Fan, R. A. Kumar, S. Nonin, J. Mol. Biol. 1997, 272, 645-664.

[116] Y. Nomura, Y. Yokobayashi, J. Am. Chem. Soc. 2007, 129, 13814-13815.

[117] S. A. Lynch, S. K. Desai, H. K. Sajja, J. P. Gallivan, Chem. Biol. 2007, 14, 173-184.

[118] A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, Nature 2006, 441, 1167-1171.

[119] C. E. Dann, 3rd, C. A. Wakeman, C. L. Sieling, S. C. Baker, I. Irnov, W. C. Winkler, Cell 2007, 130, 878-892.

[120] J. Tomsic, B. A. McDaniel, F. J. Grundy, T. M. Henkin, J. Bacteriol. 2008, 190, 823-833.

[121] A. K. Bera, J. Zhu, H. Zalkin, J. L. Smith, J. Bacteriol. 2003, 185, 4099-4109.

[122] L. E. Johansen, P. Nygaard, C. Lassen, Y. Agerso, H. H. Saxild, J. Bacteriol. 2003, 185, 5200-5209.

[123] M. Weng, P. L. Nagy, H. Zalkin, Proc. Natl Acad. Sci. U. S. A. 1995, 92, 7455-7459.

[124] G. Werstuck, M. R. Green, Science 1998, 282, 296-298.

[125] S. K. Desai, J. P. Gallivan, J. Am. Chem. Soc. 2004, 126, 13247-13254.

[126] T. S. Bayer, C. D. Smolke, Nat. Biotechnol. 2005, 23, 337-343.

[127] A. R. Buskirk, A. Landrigan, D. R. Liu, Chem Biol. 2004, 11, 1157-1163.

[128] D. Grate, C. Wilson, Bioorg. Med. Chem. 2001, 9, 2565-2570.

[129] I. Harvey, P. Garneau, J. Pelletier, RNA 2002, 8, 452-463.

[130] F. J. Isaacs, J. J. Collins, Nat. Biotechnol. 2005, 23, 306-307.

[131] F. J. Isaacs, D. J. Dwyer, J. J. Collins, Nat. Biotechnol. 2006, 24, 545-554.

[132] F. J. Isaacs, D. J. Dwyer, C. Ding, D. D. Pervouchine, C. R. Cantor, J. J. Collins, Nat. Biotechnol. 2004, 22, 841-847.

[133] D. S. Kim, V. Gusti, S. G. Pillai, R. K. Gaur, RNA 2005, 11, 1667-1677.

[134] B. Suess, B. Fink, C. Berens, R. Stentz, W. Hillen, Nucleic Acids Res. 2004, 32, 1610-1614.

[135] B. Suess, S. Hanson, C. Berens, B. Fink, R. Schroeder, W. Hillen, Nucleic Acids Res. 2003, 31, 1853-1858.

[136] G. Mayer, M. Famulok, Chembiochem 2006, 7, 602-604.

[137] T. Ha, Methods 2001, 25, 78-86.

[138] R. D. Jenison, S. C. Gill, A. Pardi, B. Polisky, Science 1994, 263, 1425-1429.

[139] U. Theuretzbacher, J. H. Toney, Curr. Opin. Investig. Drugs 2006, 7, 158-166.

[140] T. Shiota, J. E. Folk, F. Tietze, Arch. Biochem. Biophys. 1958, 77, 372-377.

[141] I. N. Hirshfield, P. C. Zamecnik, Biochim. Biophys. Acta 1972, 259, 330-343.

[142] M. Di Girolamo, V. Busiello, R. Coccia, C. Foppoli, Physiol. Chem. Phys. Med. NMR 1990, 22, 241-245.

[143] B. Vold, J. Szulmajster, A. Carbone, J. Bacteriol. 1975, 121, 970-974.

[144] Y. Lu, T. N. Shevtchenko, H. Paulus, FEMS Microbiol. Lett. 1992, 71, 23-27.

[145] J. C. Patte, M. Akrim, V. Mejean, FEMS Microbiol. Lett. 1998, 169, 165-170.

[146] K. F. Blount, J. X. Wang, J. Lim, N. Sudarsan, R. R. Breaker, Nat. Chem. Biol. 2007, 3, 44-49.

[147] N. Sudarsan, S. Cohen-Chalamish, S. Nakamura, G. M. Emilsson, R. R. Breaker, Chem. Biol. 2005, 12, 1325-1335.

[148] W. J. Robbins, Proc. Natl Acad. Sci. U. S. A. 1941, 27, 419-422.

[149] D. W. Woolley, J. Biol. Chem. 1951, 191, 43-54.

[150] A. Iwashima, Y. Wakabayashi, Y. Nose, J. Biochem. 1976, 79, 845-847.

[151] T. Kubodera, N. Yamashita, A. Nishimura, Biosci. Biotechnol. Biochem. 2000, 64, 1416-1421.

[152] K. Matsui, H.-C. Wang, T. Hirota, H. Matsukawa, S. Kasai, K. Shinagawa, S. Otani, Agric. Biol. Chem. 1982, 46, 2003-2008.

[153] M. Zuker, Nucleic Acids Res. 2003, 31, 3406-3415.

BIBLIOGRAPHIE

Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., et Strobel, S. A. (2004). Crystal structure of a self-splicing group I intron with both exons. Nature *430*, 45-50.

Allison, L. A., Romaniuk, P. J., et Bakken, A. H. (1991). RNA-protein interactions of stored 5S RNA with TFIIIA and ribosomal protein L5 during Xenopus oogenesis. Dev Biol *144*, 129-144.

Ames, T. D., et Breaker, R. R. (2011). Bacterial aptamers that selectively bind glutamine. RNA Biol *8*,

Ames, T. D., Rodionov, D. A., Weinberg, Z., et Breaker, R. R. (2010). A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate. Chem Biol *17*, 681-685.

Antonioli, A. H., Cochrane, J. C., Lipchock, S. V., et Strobel, S. A. (2010). Plasticity of the RNA kink turn structural motif. RNA 1-8.

Artsimovitch, I., et Henkin, T. M. (2009). In vitro approaches to analysis of transcription termination. Methods *47*, 37-43.

Artsimovitch, I., et Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc Natl Acad Sci USA *97*, 7090-7095.

Ataide, S. F., Wilson, S. N., Dang, S., Rogers, T. E., Roy, B., Banerjee, R., Henkin, T. M., et Ibba, M. (2007). Mechanisms of resistance to an amino acid antibiotic that targets translation. ACS Chem Biol *2*, 819-827.

Babitzke, P. (2004). Regulation of transcription attenuation and translation initiation by allosteric control of an RNA-binding protein: the Bacillus subtilis TRAP protein. Curr Opin Microbiol *7*, 132-139.

Babitzke, P., Baker, C., et Romeo, T. (2009). Regulation of Translation Initiation by RNA Binding Proteins. Annu Rev Microbiol

Bailey, M. J., Hughes, C., et Koronakis, V. (1997). RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. Mol Microbiol *26*, 845-851.

Baird, N. J., Kulshina, N., et Ferré-D'Amaré, A. R. (2010). Riboswitch function: Flipping the switch or tuning the dimmer? RNA Biol *7*,

Ban, N., Nissen, P., Hansen, J., Moore, P. B., et Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science *289*, 905-920.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A. E., Rafikov, R., Mustaev, A., et Nudler, E. (2005). A ratchet mechanism of transcription elongation and its control. Cell *120*, 183-193.

Bar-Nahum, G., et Nudler, E. (2001). Isolation and characterization of sigma(70)-retaining transcription elongation complexes from Escherichia coli. Cell *106*, 443-451.

Barrick, J. E., et Breaker, R. R. (2007). The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biol *8*, R239.

Barrick, J. E., Corbino, K. A., Winkler, W. C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J. K., et Breaker, R. R. (2004). New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc Natl Acad Sci USA *101*, 6421-6426.

Batey, R. T., Gilbert, S. D., et Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. Nature *432*, 411-415.

Belogurov, G. A., Mooney, R. A., Svetlov, V., Landick, R., et Artsimovitch, I. (2009). Functional specialization of transcription elongation factors. EMBO J *28*, 112-122.

Blouin, S., Chinnappan, R., et Lafontaine, D. A. (2010). Folding of the lysine riboswitch: importance of peripheral elements for transcriptional regulation. Nucleic Acids Research

Blouin, S., et Lafontaine, D. A. (2007). A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. RNA *13*, 1256-1267.

Blouin, S., Mulhbacher, J., Penedo, J. C., et Lafontaine, D. A. (2009). Riboswitches: ancient and promising genetic regulators. Chembiochem *10*, 400-416.

Blount, K. F., et Breaker, R. R. (2006). Riboswitches as antibacterial drug targets. Nat Biotechnol *24*, 1558-1564.

Blount, K. F., Wang, J. X., Lim, J., Sudarsan, N., et Breaker, R. R. (2007). Antibacterial lysine analogs that target lysine riboswitches. Nat Chem Biol *3*, 44-49.

Borsuk, P., Dzikowska, A., Empel, J., Grzelak, A., Grześkowiak, R., et Weglenski, P. (1999). Structure of the arginase coding gene and its transcript in Aspergillus nidulans. Acta Biochim Pol *46*, 391-403.

Borsuk, P., Przykorska, A., Blachnio, K., Koper, M., Pawlowicz, J. M., Pekala, M., et Weglenski, P. (2007). L-arginine influences the structure and function of arginase mRNA in Aspergillus nidulans. Biol Chem *388*, 135-144.

Borukhov, S., et Nudler, E. (2008). RNA polymerase: the vehicle of transcription. Trends Microbiol *16*, 126-134.

Branch, A. D., Benenfeld, B. J., et Robertson, H. D. (1985). Ultraviolet light-induced crosslinking reveals a unique region of local tertiary structure in potato spindle tuber viroid and HeLa 5S RNA. Proc Natl Acad Sci USA *82*, 6590-6594.

Breaker, R. R. (2010). Riboswitches and the RNA World. Cold Spring Harbor Perspectives in Biology

Browning, D. F., et Busby, S. J. (2004). The regulation of bacterial transcription initiation. Nat Rev Microbiol *2*, 57-65.

Burgess, R. R., Travers, A. A., Dunn, J. J., et Bautz, E. K. (1969). Factor stimulating transcription by RNA polymerase. Nature *221*, 43-46.

Burns, C. M., et Richardson, J. P. (1995). NusG is required to overcome a kinetic limitation to Rho function at an intragenic terminator. Proc Natl Acad Sci USA *92*, 4738-4742.

Burns, C. M., Richardson, L. V., et Richardson, J. P. (1998). Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. J Mol Biol *278*, 307-316.

Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E., et Nudler, E. (2008). Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. Science *320*, 935-938.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., et Doudna, J. A. (1996a). Crystal structure of a group I ribozyme domain: principles of RNA packing. Science *273*, 1678-1685.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R., et Doudna, J. A. (1996b). RNA tertiary structure mediation by adenosine platforms. Science *273*, 1696-1699.

Chan, C. L., Wang, D., et Landick, R. (1997). Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3' end spacing distinguishes pause and termination pathways. J Mol Biol *268*, 54-68.

Cheah, M. T., Wachter, A., Sudarsan, N., et Breaker, R. R. (2007). Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature *447*, 497-500.

Chen, N. Y., Hu, F. M., et Paulus, H. (1987). Nucleotide sequence of the overlapping genes for the subunits of Bacillus subtilis aspartokinase II and their control regions. J Biol Chem *262*, 8787-8798.

Cho, I.-M., Lai, L. B., Susanti, D., Mukhopadhyay, B., et Gopalan, V. (2010). Ribosomal protein L7Ae is a subunit of archaeal RNase P. Proc Natl Acad Sci USA

Chowdhury, S., Maris, C., Allain, F. H.-T., et Narberhaus, F. (2006). Molecular basis for temperature sensing by an RNA thermometer. The EMBO Journal *25*, 2487-2497.

Ciampi, M. S. (2006). Rho-dependent terminators and transcription termination. Microbiology *152*, 2515-2528.

Cochrane, J. C., Lipchock, S. V., et Strobel, S. A. (2007). Structural investigation of the GlmS ribozyme bound to Its catalytic cofactor. Chem Biol *14*, 97-105.

Collins, J. A., Irnov, I., Baker, S., et Winkler, W. C. (2007). Mechanism of mRNA destabilization by the glmS ribozyme. Genes Dev *21*, 3356-3368.

Corbino, K. A., Barrick, J. E., Lim, J., Welz, R., Tucker, B. J., Puskarz, I., Mandal, M., Rudnick, N. D., et Breaker, R. R. (2005). Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. Genome Biol *6*, R70.

Cromie, M. J., Shi, Y., Latifi, T., et Groisman, E. A. (2006). An RNA sensor for intracellular Mg(2+). Cell *125*, 71-84.

Darst, S. A. (2001). Bacterial RNA polymerase. Curr Opin Struct Biol *11*, 155-162.

de la Peña, M., Dufour, D., et Gallego, J. (2009). Three-way RNA junctions with remote tertiary contacts: A recurrent and highly versatile fold. RNA

Dennis, P. P., et Omer, A. (2005). Small non-coding RNAs in Archaea. Curr Opin Microbiol *8*, 685-694.

Edwards, A. L., et Batey, R. T. (2009). A structural basis for the recognition of 2'-deoxyguanosine by the purine riboswitch. J Mol Biol *385*, 938-948.

Edwards, A. L., Reyes, F. E., Héroux, A., et Batey, R. T. (2010). Structural basis for recognition of S-adenosylhomocysteine by riboswitches. RNA (New York, NY)

Epshtein, V., Cardinale, C. J., Ruckenstein, A. E., Borukhov, S., et Nudler, E. (2007). An allosteric path to transcription termination. Mol Cell *28*, 991-1001.

Epshtein, V., Dutta, D., Wade, J., et Nudler, E. (2010). An allosteric mechanism of Rho-dependent transcription termination. Nature *463*, 245-249.

Epshtein, V., Mironov, A. S., et Nudler, E. (2003). The riboswitch-mediated control of sulfur metabolism in bacteria. Proc Natl Acad Sci USA *100*, 5052-5056.

Franklund, C. V., et Kadner, R. J. (1997). Multiple transcribed elements control expression of the Escherichia coli btuB gene. J Bacteriol *179*, 4039-4042.

Fuchs, R. T., Grundy, F. J., et Henkin, T. M. (2006). The S(MK) box is a new SAM-binding RNA for translational regulation of SAM synthetase. Nat Struct Mol Biol *13*, 226-233.

Garst, Héroux, A., Rambo, R. P., et Batey, R. T. (2008). Crystal structure of the lysine riboswitch regulatory mRNA element. J Biol Chem *283*, 22347-22351.

Garst, A. D., Edwards, A. L., et Batey, R. T. (2010). Riboswitches: Structures and Mechanisms. Cold Spring Harbor Perspectives in Biology 1-14.

Garst, A. D., Héroux, A., Rambo, R. P., et Batey, R. T. (2008). Crystal structure of the lysine riboswitch regulatory mRNA element. J Biol Chem *283*, 22347-22351.

Gelfand, M. S., Mironov, A. A., Jomantas, J., Kozlov, Y. I., et Perumov, D. A. (1999). A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. Trends Genet *15*, 439-442.

Goldman, S. R., Ebright, R. H., et Nickels, B. E. (2009). Direct detection of abortive RNA transcripts in vivo. Science *324*, 927-928.

Gottesman, S., et Storz, G. (2010). Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. Cold Spring Harbor Perspectives in Biology 1-17.

Grandgenett, D. P., et Stahly, D. P. (1971). Repression of diaminopimelate decarboxylase by L-lysine in different Bacillus species. J Bacteriol *105*, 1211-1212.

Green, N. J., Grundy, F. J., et Henkin, T. M. (2010). The T box mechanism: tRNA as a regulatory molecule. FEBS Lett *584*, 318-324.

Grundy, Lehman, S. C., et Henkin, T. M. (2003). The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. Proc Natl Acad Sci U S A *100*, 12057-12062.

Grundy, F. J., et Henkin, T. M. (2003). The T box and S box transcription termination control systems. Front Biosci *8*, d20-31.

Grundy, F. J., et Henkin, T. M. (2004). Kinetic analysis of tRNA-directed transcription antitermination of the Bacillus subtilis glyQS gene in vitro. J Bacteriol *186*, 5392-5399.

Grundy, F. J., Winkler, W. C., et Henkin, T. M. (2002). tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. Proc Natl Acad Sci USA *99*, 11121-11126.

Gusarov, I., et Nudler, E. (1999). The mechanism of intrinsic transcription termination. Mol Cell *3*, 495-504.

Haddrick, M., Lear, A. L., Cann, A. J., et Heaphy, S. (1996). Evidence that a kissing loop structure facilitates genomic RNA dimerisation in HIV-1. J Mol Biol *259*, 58-68.

Hampel, K. J., et Tinsley, M. M. (2006). Evidence for preorganization of the glmS ribozyme ligand binding pocket. Biochemistry *45*, 7861-7871.

Helmann, J. D., et Chamberlin, M. J. (1988). Structure and function of bacterial sigma factors. Annu Rev Biochem *57*, 839-872.

Hengge, R. (2010). Cyclic-di-GMP Reaches Out into the Bacterial RNA World. Science Signaling *3*, pe44-pe44.

Henkin, T. M. (2008). Riboswitch RNAs: using RNA to sense cellular metabolism. Genes Dev *22*, 3383-3390.

Heppell, B., Blouin, S., Dussault, A.-M., Mulhbacher, J., Ennifar, E., Penedo, J. C., et Lafontaine, D. A. (2011). Molecular insinghts into the ligand-controlled organization of the SAM-I riboswitch. Nat Chem Biol

Heppell, B., et Lafontaine, D. A. (2008). Folding of the SAM aptamer is determined by the formation of a K-turn-dependent pseudoknot. Biochemistry *47*, 1490-1499.

Herbert, K. M., Zhou, J., Mooney, R. A., Porta, A. L., Landick, R., et Block, S. M. (2010). E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. J Mol Biol *399*, 17-30.

Hsu, L. M., Vo, N. V., Kane, C. M., et Chamberlin, M. J. (2003). In vitro studies of transcript initiation by Escherichia coli RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters. Biochemistry *42*, 3777-3786.

Jaeger, L., Michel, F., et Westhof, E. (1994). Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. J Mol Biol *236*, 1271-1276.

Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M., et Cossart, P. (2002). An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes. Cell *110*, 551-561.

Jorth, P., et Whiteley, M. (2010). Characterization of a Novel Riboswitch-Regulated Lysine Transporter in Aggregatibacter actinomycetemcomitans. J Bacteriol *192*, 6240-6250.

Juang, Y. L., et Helmann, J. D. (1994). A promoter melting region in the primary sigma factor of Bacillus subtilis. Identification of functionally important aromatic amino acids. J Mol Biol *235*, 1470-1488.

Kim, J., Blount, K., Lim, J., Link, K., et Breaker, R. (2009). Design and Antimicrobial Action of Purine Analogs that Bind Guanine Riboswitches. ACS Chem Biol

Kim, J. N., Roth, A., et Breaker, R. R. (2007). Guanine riboswitch variants from Mesoplasma florum selectively recognize 2'-deoxyguanosine. Proc Natl Acad Sci USA *104*, 16092-16097.

Kireeva, M. L., et Kashlev, M. (2009). Mechanism of sequence-specific pausing of bacterial RNA polymerase. Proc Natl Acad Sci USA *106*, 8900-8905.

Klein, D. J., Edwards, T. E., et Ferré-D'Amaré, A. R. (2009). Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. Nat Struct Mol Biol *16*, 343-344.

Klein, D. J., et Ferré-D'Amaré, A. R. (2006). Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. Science *313*, 1752-1756.

Klein, D. J., Schmeing, T. M., Moore, P. B., et Steitz, T. A. (2001). The kink-turn: a new RNA secondary structure motif. EMBO J *20*, 4214-4221.

Kochhar, S., et Paulus, H. (1996). Lysine-induced premature transcription termination in the lysC operon of Bacillus subtilis. Microbiology (Reading, Engl) *142 (Pt 7)*, 1635-1639.

Kubodera, T., Watanabe, M., Yoshiuchi, K., Yamashita, N., Nishimura, A., Nakai, S., Gomi, K., et Hanamoto, H. (2003). Thiamine-regulated gene expression of Aspergillus oryzae thiA requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR. FEBS Lett *555*, 516-520.

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem Soc Trans *34*, 1062-1066.

Landick, R. (2009). Transcriptional pausing without backtracking. Proc Natl Acad Sci USA *106*, 8797-8798.

Laptenko, O., Lee, J., Lomakin, I., et Borukhov, S. (2003). Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. The EMBO Journal *22*, 6322-6334.

Larson, M. H., Greenleaf, W. J., Landick, R., et Block, S. M. (2008). Applied force reveals mechanistic and energetic details of transcription termination. Cell *132*, 971-982.

Lauhon, C. T., et Szostak, J. W. (1995). RNA aptamers that bind flavin and nicotinamide redox cofactors. J Am Chem Soc *117*, 1246-1257.

Lee, E., Blount, K., et Breaker, R. (2009). Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. RNA Biol *6*,

Lemay, J.-F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Massé, E., et Lafontaine, D. A. (2011). Comparative Study between Transcriptionally- and

Translationally-Acting Adenine Riboswitches Reveals Key Differences in Riboswitch Regulatory Mechanisms. PLoS Genet *7*, e1001278.

Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M. J., et Lafontaine, D. A. (2006). Folding of the adenine riboswitch. Chem Biol *13*, 857-868.

Leontis, N. B., et Westhof, E. (1998). A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. J Mol Biol *283*, 571-583.

Lescoute, A., Leontis, N. B., Massire, C., et Westhof, E. (2005). Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. Nucleic Acids Research *33*, 2395-2409.

Li, J., Horwitz, R., McCracken, S., et Greenblatt, J. (1992). NusG, a new Escherichia coli elongation factor involved in transcriptional antitermination by the N protein of phage lambda. J Biol Chem *267*, 6012-6019.

Li, J., Mason, S. W., et Greenblatt, J. (1993). Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. Genes Dev *7*, 161-172.

Liao, H. H., et Hseu, T. H. (1998). Analysis of the regulatory region of the lysC gene of Escherichia coli. FEMS Microbiol Lett *168*, 31-36.

Lillo, F., Basile, S., et Mantegna, R. N. (2002). Comparative genomics study of inverted repeats in bacteria. Bioinformatics *18*, 971-979.

Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., et Johansson, J. (2010). A trans-Acting Riboswitch Controls Expression of the Virulence Regulator PrfA in Listeria monocytogenes. Cell *139*, 770-779.

Lu, Y., Chen, N. Y., et Paulus, H. (1991). Identification of aecA mutations in Bacillus subtilis as nucleotide substitutions in the untranslated leader region of the aspartokinase II operon. J Gen Microbiol *137*, 1135-1143.

Lu, Y., Shevtchenko, T. N., et Paulus, H. (1992). Fine-structure mapping of cis-acting control sites in the lysC operon of Bacillus subtilis. FEMS Microbiol Lett *71*, 23-27.

Lundrigan, M. D., Köster, W., et Kadner, R. J. (1991). Transcribed sequences of the Escherichia coli btuB gene control its expression and regulation by vitamin B12. Proc Natl Acad Sci USA *88*, 1479-1483.

Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C., et Breaker, R. R. (2003). Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. Cell *113*, 577-586.

Mandal, M., et Breaker, R. R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat Struct Mol Biol *11*, 29-35.

Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L., et Breaker, R. R. (2004). A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science *306*, 275-279.

Martick, M., et Scott, W. G. (2006). Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell *126*, 309-320.

Matera, A. G., Terns, R. M., et Terns, M. P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol *8*, 209-220.

Mattick, J. S. (2004). RNA regulation: a new genetics? Nat Rev Genet *5*, 316-323.

Mattick, J. S., et Makunin, I. V. (2006). Non-coding RNA. Hum Mol Genet *15 Spec No 1*, R17-29.

McCarthy, T. J., Plog, M. A., Floy, S. A., Jansen, J. A., Soukup, J. K., et Soukup, G. A. (2005). Ligand requirements for glmS ribozyme self-cleavage. Chem Biol *12*, 1221-1226.

McDaniel, B. A., Grundy, F. J., et Henkin, T. M. (2005). A tertiary structural element in S box leader RNAs is required for S-adenosylmethionine-directed transcription termination. Mol Microbiol *57*, 1008-1021.

McDaniel, B. A. M., Grundy, F. J., Artsimovitch, I., et Henkin, T. M. (2003). Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. Proc Natl Acad Sci USA *100*, 3083-3088.

Miranda-Ríos, J., Navarro, M., et Soberón, M. (2001). A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. Proc Natl Acad Sci USA *98*, 9736-9741.

Mironov, A., Epshtein, V., et Nudler, E. (2009). Transcriptional approaches to riboswitch studies. Methods Mol Biol *540*, 39-51.

Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., Perumov, D. A., et Nudler, E. (2002). Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. Cell *111*, 747-756.

Montange, R. K., et Batey, R. T. (2006). Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. Nature *441*, 1172-1175.

Mooney, R. A., Davis, S. E., Peters, J. M., Rowland, J. L., Ansari, A. Z., et Landick, R. (2009). Regulator trafficking on bacterial transcription units in vivo. Mol Cell *33*, 97-108.

Mooney, R. A., et Landick, R. (2003). Tethering sigma70 to RNA polymerase reveals high in vivo activity of sigma factors and sigma70-dependent pausing at promoter-distal locations. Genes Dev *17*, 2839-2851.

Mulhbacher, Brouillette, E., Allard, M., Fortier, L.-C., Malouin, F., et Lafontaine, D. A. (2010). Novel riboswitch ligand analogs as selective inhibitors of Guanine-related metabolic pathways. PLoS Pathog *6*, e1000865.

Mulhbacher, J., Brouillette, E., Allard, M., Fortier, L.-C., Malouin, F., et Lafontaine, D. A. (2010). Novel riboswitch ligand analogs as selective inhibitors of Guanine-related metabolic pathways. PLoS Pathog *6*, e1000865.

Nagaswamy, U., et Fox, G. E. (2002). Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. RNA (New York, NY) *8*, 1112-1119.

Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., et Breaker, R. R. (2002). Genetic control by a metabolite binding mRNA. Chem Biol *9*, 1043.

Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E., et Altuvia, S. (2009). A pH-responsive riboregulator. Genes Dev *23*, 2650-2662.

Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J., et Block, S. M. (2003). Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. Cell *115*, 437-447.

Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., et Steitz, T. A. (2001). RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. Proc Natl Acad Sci USA *98*, 4899-4903.

Nou, X., et Kadner, R. J. (1998). Coupled changes in translation and transcription during cobalamin-dependent regulation of btuB expression in Escherichia coli. J Bacteriol *180*, 6719-6728.

Nou, X., et Kadner, R. J. (2000). Adenosylcobalamin inhibits ribosome binding to btuB RNA. Proc Natl Acad Sci USA *97*, 7190-7195.

Nudler, E. (2009). RNA polymerase active center: the molecular engine of transcription. Annu Rev Biochem *78*, 335-361.

Nudler, E., et Gottesman, M. E. (2002). Transcription termination and anti-termination in E. coli. Genes Cells *7*, 755-768.

Nudler, E., et Gusarov, I. (2003). Analysis of the intrinsic transcription termination mechanism and its control. Biophysical, Chemical, and Functional Probes of RNA Structure, Interactions and Folding: Part B *371*, 369-382.

Nudler, E., Gusarov, I., et Bar-Nahum, G. (2003). Methods of walking with the RNA polymerase. Biophysical, Chemical, and Functional Probes of RNA Structure, Interactions and Folding: Part B *371*, 160-169.

Nudler, E., et Mironov, A. S. (2004). The riboswitch control of bacterial metabolism. Trends Biochem Sci *29*, 11-17.

Paget, M. S. B., et Helmann, J. D. (2003). The sigma70 family of sigma factors. Genome Biol *4*, 203.

Pan, T., Artsimovitch, I., Fang, X. W., Landick, R., et Sosnick, T. R. (1999). Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. Proc Natl Acad Sci USA *96*, 9545-9550.

Park, J.-S., Marr, M. T., et Roberts, J. W. (2002). E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. Cell *109*, 757-767.

Patte, J. C., Akrim, M., et Méjean, V. (1998). The leader sequence of the Escherichia coli lysC gene is involved in the regulation of LysC synthesis. FEMS Microbiol Lett *169*, 165-170.

Peters, J. M., Mooney, R. A., Kuan, P. F., Rowland, J. L., Keles, S., et Landick, R. (2009). Rho directs widespread termination of intragenic and stable RNA transcription. Proc Natl Acad Sci USA *106*, 15406-15411.

Phan, T. T. P., et Schumann, W. (2009). Transcriptional Analysis of the Lysine-Responsive and Riboswitch-Regulated lysC Gene of Bacillus subtilis. Curr Microbiol 1-6.

Poiata, E., Meyer, M. M., Ames, T. D., et Breaker, R. R. (2009). A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. RNA 1-12.

Rana, T. M. (2007). Illuminating the silence: understanding the structure and function of small RNAs. Nat Rev Mol Cell Biol *8*, 23-36.

Ravnum, S., et Andersson, D. I. (1997). Vitamin B12 repression of the btuB gene in Salmonella typhimurium is mediated via a translational control which requires leader and coding sequences. Mol Microbiol *23*, 35-42.

Regulski, E. E., Moy, R. H., Weinberg, Z., Barrick, J. E., Yao, Z., Ruzzo, W. L., et Breaker, R. R. (2008). A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. Mol Microbiol *68*, 918-932.

Reppas, N. B., Wade, J. T., Church, G. M., et Struhl, K. (2006). The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. Mol Cell *24*, 747-757.

Richter-Dahlfors, A. A., Ravnum, S., et Andersson, D. I. (1994). Vitamin B12 repression of the cob operon in Salmonella typhimurium: translational control of the cbiA gene. Mol Microbiol *13*, 541-553.

Rieder, U., Lang, K., Kreutz, C., Polacek, N., et Micura, R. (2009). Evidence for pseudoknot formation of class I preQ1 riboswitch aptamers. Chembiochem *10*, 1141-1144.

Ring, B. Z., Yarnell, W. S., et Roberts, J. W. (1996). Function of E. coli RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. Cell *86*, 485-493.

Roberts, J. W., Shankar, S., et Filter, J. J. (2008). RNA polymerase elongation factors. Annu Rev Microbiol *62*, 211-233.

Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., et Gelfand, M. S. (2003). Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? Nucleic Acids Res *31*, 6748-6757.

Rosner, A. (1975). Control of lysine biosynthesis in Bacillus subtilis: inhibition of diaminopimelate decarboxylase by lysine. J Bacteriol *121*, 20-28.

Roth, A., et Breaker, R. R. (2009). The structural and functional diversity of metabolite-binding riboswitches. Annu Rev Biochem *78*, 305-334.

Roth, A., Winkler, W. C., Regulski, E. E., Lee, B. W. K., Lim, J., Jona, I., Barrick, J. E., Ritwik, A., Kim, J. N., Welz, R., Iwata-Reuyl, D., et Breaker, R. R. (2007). A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. Nat Struct Mol Biol *14*, 308-317.

Schmidt, M. C., et Chamberlin, M. J. (1984). Binding of rho factor to Escherichia coli RNA polymerase mediated by nusA protein. J Biol Chem *259*, 15000-15002.

Schwartz, A., Rahmouni, A. R., et Boudvillain, M. (2003). The functional anatomy of an intrinsic transcription terminator. The EMBO Journal *22*, 3385-3394.

Serganov, A. (2009). The long and the short of riboswitches. Curr Opin Struct Biol

Serganov, A., Huang, L., et Patel, D. J. (2008). Structural insights into amino acid binding and gene control by a lysine riboswitch. Nature *455*, 1263-1267.

Serganov, A., Yuan, Y.-R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R., et Patel, D. J. (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. Chem Biol *11*, 1729-1741.

Sevostyanova, A., et Artsimovitch, I. (2010). Functional analysis of Thermus thermophilus transcription factor NusG. Nucleic Acids Research *38*, 7432-7445.

Sevostyanova, A., Svetlov, V., Vassylyev, D. G., et Artsimovitch, I. (2008). The elongation factor RfaH and the initiation factor sigma bind to the same site on the transcription elongation complex. Proc Natl Acad Sci USA *105*, 865-870.

Shaevitz, J. W., Abbondanzieri, E. A., Landick, R., et Block, S. M. (2003). Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. Nature *426*, 684-687.

Smith, A. M., Fuchs, R. T., Grundy, F. J., et Henkin, T. M. (2010a). Riboswitch RNAs: regulation of gene expression by direct monitoring of a physiological signal. RNA Biol *7*, 104-110.

Smith, A. M., Fuchs, R. T., Grundy, F. J., et Henkin, T. M. (2010b). The SAM-responsive SMK box is a reversible riboswitch. Mol Microbiol *78*, 1393-1402.

Srisawat, C., et Engelke, D. R. (2001). Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. RNA *7*, 632-641.

Sudarsan, N., Barrick, J. E., et Breaker, R. R. (2003a). Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA *9*, 644-647.

Sudarsan, N., Hammond, M. C., Block, K. F., Welz, R., Barrick, J. E., Roth, A., et Breaker, R. R. (2006). Tandem riboswitch architectures exhibit complex gene control functions. Science *314*, 300-304.

Sudarsan, N., Lee, E. R., Weinberg, Z., Moy, R. H., Kim, J. N., Link, K. H., et Breaker, R. R. (2008). Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science *321*, 411-413.

Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S., et Breaker, R. R. (2003b). An mRNA structure in bacteria that controls gene expression by binding lysine. Genes Dev *17*, 2688-2697.

Sullivan, S. L., et Gottesman, M. E. (1992). Requirement for E. coli NusG protein in factor-dependent transcription termination. Cell *68*, 989-994.

Szewczak, A. A., et Moore, P. B. (1995). The sarcin/ricin loop, a modular RNA. J Mol Biol *247*, 81-98.

Tavormina, P. L., Reznikoff, W. S., et Gross, C. A. (1996). Identifying interacting regions in the beta subunit of Escherichia coli RNA polymerase. J Mol Biol *258*, 213-223.

Tinsley, R. A., Furchak, J. R. W., et Walter, N. G. (2007). Trans-acting glmS catalytic riboswitch: locked and loaded. RNA *13*, 468-477.

Tomsic, J., McDaniel, B. A., Grundy, F. J., et Henkin, T. M. (2008). Natural variability in S-adenosylmethionine (SAM)-dependent riboswitches: S-box elements in bacillus subtilis exhibit differential sensitivity to SAM In vivo and in vitro. J Bacteriol *190*, 823-833.

Torres, M., Balada, J.-M., Zellars, M., Squires, C., et Squires, C. L. (2004). In vivo effect of NusB and NusG on rRNA transcription antitermination. J Bacteriol *186*, 1304-1310.

Tucker, B. J., et Breaker, R. R. (2005). Riboswitches as versatile gene control elements. Curr Opin Struct Biol *15*, 342-348.

Tuerk, C., et Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-510.

Turner, B., Melcher, S. E., Wilson, T. J., Norman, D. G., et Lilley, D. M. J. (2005). Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. RNA *11*, 1192-1200.

Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., et Gelfand, M. S. (2004). Riboswitches: the oldest mechanism for the regulation of gene expression? Trends Genet *20*, 44-50.

Vold, B., Szulmajster, J., et Carbone, A. (1975). Regulation of dihydrodipicolinate synthase and aspartate kinase in Bacillus subtilis. J Bacteriol *121*, 970-974.

von Hippel, P. H., et Yager, T. D. (1992). The elongation-termination decision in transcription. Science *255*, 809-812.

Wachter, A., Tunc-Ozdemir, M., Grove, B. C., Green, P. J., Shintani, D. K., et Breaker, R. R. (2007). Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. Plant Cell *19*, 3437-3450.

Walter, G., Zillig, W., Palm, P., et Fuchs, E. (1967). Initiation of DNA-dependent RNA synthesis and the effect of heparin on RNA polymerase. Eur J Biochem *3*, 194-201.

Wang, J. X., Lee, E. R., Morales, D. R., Lim, J., et Breaker, R. R. (2008). Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. Mol Cell *29*, 691-702.

Wei, B. Y., Bradbeer, C., et Kadner, R. J. (1992). Conserved structural and regulatory regions in the Salmonella typhimurium btuB gene for the outer membrane vitamin B12 transport protein. Res Microbiol *143*, 459-466.

Weinberg, Z., Barrick, J. E., Yao, Z., Roth, A., Kim, J. N., Gore, J., Wang, J. X., Lee, E. R., Block, K. F., Sudarsan, N., Neph, S., Tompa, M., Ruzzo, W. L., et Breaker, R. R. (2007). Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. Nucleic Acids Research *35*, 4809-4819.

Weinberg, Z., Regulski, E. E., Hammond, M. C., Barrick, J. E., Yao, Z., Ruzzo, W. L., et Breaker, R. R. (2008). The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. RNA (New York, NY) *14*, 822-828.

Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K., Moy, R. H., et Breaker, R. R. (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea and their metagenomes. Genome Biol *11*, R31.

Welz, R., et Breaker, R. R. (2007). Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. RNA *13*, 573-582.

Werstuck, G., et Green, M. R. (1998). Controlling gene expression in living cells through small molecule-RNA interactions. Science *282*, 296-298.

Wickiser, J. K., Cheah, M. T., Breaker, R. R., et Crothers, D. M. (2005a). The kinetics of ligand binding by an adenine-sensing riboswitch. Biochemistry *44*, 13404-13414.

Wickiser, J. K., Winkler, W. C., Breaker, R. R., et Crothers, D. M. (2005b). The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. Mol Cell *18*, 49-60.

Wilkinson, S. R., et Been, M. D. (2005). A pseudoknot in the 3' non-core region of the glmS ribozyme enhances self-cleavage activity. RNA (New York, NY) *11*, 1788-1794.

Winkler, Cohen-Chalamish, S., et Breaker, R. R. (2002a). An mRNA structure that controls gene expression by binding FMN. Proc Natl Acad Sci USA *99*, 15908-15913.

Winkler, Nahvi, A., et Breaker, R. R. (2002b). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature *419*, 952-956.

Winkler, M. E., et Yanofsky, C. (1981). Pausing of RNA polymerase during in vitro transcription of the tryptophan operon leader region. Biochemistry *20*, 3738-3744.

Winkler, W. C. (2005a). Metabolic monitoring by bacterial mRNAs. Arch Microbiol *183*, 151-159.

Winkler, W. C. (2005b). Riboswitches and the role of noncoding RNAs in bacterial metabolic control. Current opinion in chemical biology *9*, 594-602.

Winkler, W. C., et Breaker, R. R. (2003). Genetic control by metabolite-binding riboswitches. Chembiochem *4*, 1024-1032.

Winkler, W. C., et Breaker, R. R. (2005). Regulation of bacterial gene expression by riboswitches. Annu Rev Microbiol *59*, 487-517.

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

Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A., et Breaker, R. R. (2004). Control of gene expression by a natural metabolite-responsive ribozyme. Nature *428*, 281-286.

Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E., et Breaker, R. R. (2003). An mRNA structure that controls gene expression by binding S-adenosylmethionine. Nat Struct Biol *10*, 701-707.

Witkin, E. M. (1994). Mutation frequency decline revisited. Bioessays *16*, 437-444.

Woese, C. R., Winker, S., et Gutell, R. R. (1990). Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". Proc Natl Acad Sci USA *87*, 8467-8471.

Wösten, M. M. (1998). Eubacterial sigma-factors. FEMS Microbiol Rev *22*, 127-150.

Yakhnin, A. V., Yakhnin, H., et Babitzke, P. (2008). Function of the Bacillus subtilis transcription elongation factor NusG in hairpin-dependent RNA polymerase pausing in the trp leader. Proc Natl Acad Sci USA *105*, 16131-16136.

Yarnell, W. S., et Roberts, J. W. (1999). Mechanism of intrinsic transcription termination and antitermination. Science *284*, 611-615.

Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G., et Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. Mol Microbiol *50*, 1111-1124.